

## RetMobileNet: A New Deep Learning Approach for Multi-Class Eye Disease Identification

Anjar Wanto<sup>1,2\*</sup>, Yuhandri Yuhandri<sup>1</sup>, Okfalisa Okfalisa<sup>3</sup>

<sup>1</sup> Department of Computer Science, Universitas Putra Indonesia YPTK Padang, Padang 25221, Indonesia

<sup>2</sup> STIKOM Tunas Bangsa, Pematangsiantar 21127, Indonesia

<sup>3</sup> Department of Informatics Engineering, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru 28293, Indonesia

Corresponding Author Email: [anjarwanto@ieee.org](mailto:anjarwanto@ieee.org)



Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ria.380401>

### ABSTRACT

**Received:** 28 December 2023

**Revised:** 2 May 2024

**Accepted:** 7 June 2024

**Available online:** 23 August 2024

#### Keywords:

*deep learning, Convolutional Neural Networks, retinal fundus images, eye disease, identification, new approach, RetMobileNet*

Image processing plays a vital role in disease identification, particularly in analyzing retinal fundus images to detect various eye conditions, such as Diabetic Retinopathy, Drusen, and Central Retinal Vein Occlusion. This study aims to develop a CNN for multi-class eye disease identification using retinal fundus images. The MuReD dataset combines 1855 images across 20 classes from ARIA, STARE, and RFMiD datasets. Eye specialists validated 1122 images, representing 12 disease classes. All images were standardized to 224×224 pixels. Data was split 80%:10%:10% for training, validation, and testing, respectively, using K-Fold Cross-validation (k=10). Seven CNN models (VGG19, InceptionV4, Resnet50, MobileNetV2, MobileNetV1, MobileNetV3 Large, and MobileNetV3 Small) were tested, with MobileNetV1 and MobileNetV2 showing the most promise. We innovatively enhanced the architectures of MobileNetV1 and MobileNetV2 through proper CLAHE configuration, optimization of training parameters (Batch size, learning rate, optimizer), and layer modification (skip connection, Dense layer, convolution layer, batch normalization, and dropout), culminating in a novel approach named Retina MobileNet (RetMobileNet). RetMobileNet consistently outperformed other models, achieving an average accuracy, recall, precision, and F1-score of over 90%. Our developed model, RetMobileNet, presents a significant advancement in the accurate identification of eye diseases from fundus images, with improved efficiency.

## 1. INTRODUCTION

Human eye is usually called as the window to the soul, it is an intricate organ responsible for the sense of sight [1]. However, its complexity makes it susceptible to a myriad of diseases that can compromise vision [2-4]. From conditions like glaucoma and macular degeneration to diabetic retinopathy [5], the spectrum of eye diseases is vast. The early and accurate identification of these diseases is crucial, as timely intervention can prevent irreversible damage and preserve vision [6-8]. Traditional diagnostic methods, while effective, often rely on the expertise of ophthalmologists and can be time-consuming [9]. Moreover, with the increasing global prevalence of eye diseases, there is a pressing need for more efficient diagnostic tools [10]. The global prevalence of eye diseases such as cataract, glaucoma, and dry eye disease is a major medical, social, and economic problem, especially among the elderly population [11]. The burden of vision loss due to eye diseases has been steadily increasing, especially among those aged 50 years and above [12]. Resource-poor countries bear a disproportionate burden of eye diseases, with blindness due to age-related conditions such as cataract and glaucoma increasing [13]. The global population growth and aging process are contributing to the increasingly severe situation of blindness worldwide [14].

Recent years have witnessed groundbreaking advancements at the intersection of technology and medicine, with deep learning emerging as a transformative tool for medical image analysis [15]. Deep learning, a subset of artificial intelligence, harnesses neural networks to process vast datasets, making it adept at tasks like image recognition [16]. Deep learning is a subset of artificial intelligence (AI) focused on training neural networks to learn from vast amounts of data [17]. These networks consist of layers of interconnected nodes, akin to neurons in the human brain [18]. Through exposure to extensive datasets, deep learning algorithms can autonomously identify patterns and make decisions without explicit programming [19]. This capability renders deep learning particularly effective in tasks such as image and speech recognition, natural language processing, and medical diagnosis [20]. Deep learning is the foundation of CNN, which is a multi-layer network structure simulating the operation mechanism of the biological vision system [21]. Deep learning, with its powerful learning ability to abstract data at a higher level, forms the basis for CNN, allowing it to obtain more accurate data and feature descriptions from original data [22]. CNN is a widely used technique in deep learning for better feature extraction from large datasets, especially in imaging tasks [23].

Convolutional Neural Networks (CNNs) have established

themselves as the primary choice for image classification due to their capacity to autonomously and systematically acquire features from images [24]. In the realm of ophthalmology, the application of CNNs to retinal fundus images offers a promising avenue for the automated identification of eye diseases [25]. Retinal fundus images, that capture the back of the eye, including the retina, optic disc, and blood vessels, provide a wealth of information about the health of the eye. Given the intricate details present in these images, CNNs, with their ability to detect subtle patterns, are well-suited for this task [26].

In this study, we conducted a comprehensive assessment of seven prominent CNN models, including VGG19, InceptionV4, Resnet50, MobileNetV1, MobileNetV2, MobileNetV3 Small, and MobileNetV3 Large, to analyze retinal fundus images and determine the most effective model for disease identification. The objective was to determine the most effective model for analyzing retinal fundus images. The results were illuminating. Among the seven models, MobileNetV1 and MobileNetV2 emerged as the superior models, outperforming their counterparts in terms of accuracy and efficiency.

MobileNet architectures, known for their compactness and efficiency, are designed for mobile and embedded vision applications [27]. MobileNetV1 introduced depth-wise separable convolutions that optimize computational efficiency without sacrificing performance [28]. MobileNetV2 further enhances this design with the introduction of the inverted residual structure, ensuring the faster processing times and improved accuracy [29]. Given the stellar performance of MobileNetV1 and MobileNetV2, we developed RetMobileNet, merging their strengths into a powerful tool tailored for multi-class eye disease identification. RetMobileNet 's development is meticulous, incorporating depth-wise convolutional layers for efficient retinal image feature extraction. These layers process each input channel separately to capture intricate details crucial for disease identification. Additionally, we improve the model's performance by implementing batch normalization to stabilize training and speed up convergence. We also introduce pivotal enhancements, such as integrating residual blocks with skip connections. Deep learning models, especially those with numerous layers, can sometimes face the challenge of the vanishing gradient problem, which can impede the model's training [30]. Residual blocks with skip connections address this issue, allowing for more effective training of the deeper layers of the network.

RetMobileNet utilizes five dense layers for final predictions, synthesizing features for accurate identification of eye diseases. The research significantly advances

ophthalmology academically by employing deep learning techniques, particularly CNNs, enhancing our understanding of the field. This study thoroughly assesses seven different CNN models to determine their effectiveness in analyzing retinal fundus images, thereby shedding light on the strengths and weaknesses of these models in the context of medical image classification. Furthermore, this research introduces a novel deep learning architecture called RetMobileNet, which combines MobileNetV1 and MobileNetV2 for multi-class eye disease identification. It represents a significant advancement in the design of deep learning models for medical image analysis. Additionally, this study addresses a common issue in deep neural networks: residual blocks with skip connections. By doing this study, it contributes valuable insights to the improvement of deep learning model training. On the practical side, this research has resulted in the development of RetMobileNet, an efficient tool for the automatic identification of eye diseases using retinal fundus images. This streamlines diagnostics has the potential to lessen dependence on human experts. Through optimized feature extraction and improved training, RetMobileNet achieves higher disease identification accuracy, providing substantial advantages for early intervention and vision preservation in real-world healthcare settings.

## 2. RELATED WORK

The research related to technology and image processing techniques utilizing Deep Learning, especially the Convolutional Neural Network (CNN) algorithm, has expanded and rapidly developed in recent years [31]. CNNs have the capability to automatically extract features, allowing them to learn relevant patterns and features from images without the need for complex manual feature extraction. In the medical field, it enables a more adaptive and efficient approach to disease identification [32]. Will explore fundus retina images for the identification of different eye health conditions (ocular diseases), including DR (Diabetic Retinopathy), Myopia, Drusen (DN), Optic Disc Cupping (ODC), Central Retinal Vein Occlusion (CRVO), and various other diseases, utilizing the CNN algorithm.

Generally, computer systems used to assist the diagnosis of eye diseases based on fundus images employ conventional methods. However, conventional detection methods can only be specifically used for one type of disease, such as glaucoma or other retinal diseases because of, the different signs of abnormalities in each disease [33]. Several studies have been developed to address eye diseases using fundus images, as shown in Table 1.

**Table 1.** Related research (previous research)

| Ref. No. | Propose Method  | Dataset  | Excess  | Lack  |
|----------|---|--|---|---|
| [34]     | AlexNet-SVDD modification                             | DIARETDB1 and Navid-Didegan (each consisting of 94 retinal Images) | The proposed algorithm achieved more than 98% precision and sensitivity for two-class classification, indicating its effectiveness in diabetic retinopathy screening              | This study does not provide detailed information about the data set used, such as size, diversity, or potential bias. Lack of Comparative Analysis with Other CNN Architectures |
| [35]     | Model Residual Dense Connection Based Unet (RDC-UNet) | EyeQ dataset consists of 1000 fundus images                        | An innovative and detailed approach to enhancing fundus images, with comprehensive quality evaluation and the use of individually trained RDC-UNet models for various distortions | Evaluation of the refinement results uses only one specific model (MvRCNN), may not provide a complete quality assessment   |

|                 |  |   |  |   |
|-----------------|--|---|--|---|
| [36]            | Algoritma Autoregressive-Henry Gas Sailfish Optimization (Ar-HGSO)   | Database Indian Diabetic Retinopathy Image Dataset (IDRID) and Dataset for Diabetic Retinopathy (DDR)   | This article introduces the innovative Ar-HGSO model, with in-depth comparative analysis and strong statistical evidence, demonstrating its effectiveness in the detection and classification of the severity of diabetic retinopathy. The AMF-RCNN method achieves a recall rate of 99.2% and an f-measure of 96.5%, effective in detecting eye diseases with high precision and treating blood vessel disorders. | The high complexity of the Ar-HGSO model may make it difficult to replicate or adapt it in different contexts. This study only discusses diabetic retinopathy.  |
| [37]            | Modification of Faster RCNN to AMF-RCNNs (anchor-free modified faster region-based CNNs)                       | 400 Fundus Images obtained from the STARE (Structured Analysis of the Retina) database  | MiniNet model reduces the number of parameters, shortens training time, and achieves high accuracy, especially with small data set.  | Reliance on large datasets for high accuracy, possible limitations in dealing with complex retinal data variations, and the need for significant computational resources.   |
| [38]            | MiniNet: Dense squeeze with depthwise separable convolutions   | Dataset CIFAR-10 and Flower   | This research provides a potential solution for large-scale screening programs, requiring minimal effort from doctors and enhancing the feasibility of identifying individuals at risk of developing AMD.  | The accuracy of the proposed model is slightly lower than other models used as comparison.  |
| [39]            | Original VGG16 with Age-Related Macular Degeneration   | AMDLesions (980 images), ADAM (400 images), ARIA (143 images), and STARE (397 images)   | MFgcForest model surpasses traditional machine learning methods and original deep forest approaches in classifying diabetic retinal images.  | This research employs lesion labels at the image level for model training, that may potentially lack detailed information about specific lesions, and the limited availability of data may impact the model's performance and generalization ability. |
| [40]            | Improved Deep Forest Model yang disebut MFgcForest   | The left and right eye image data set is 8408 images obtained from the Kaggle website   | The MBSaNet model outperforms other convolutional networks with fewer parameters. It exhibits superior global modeling capabilities, enabling more effective extraction of pathological features from various lesions.   | Lack of comparison with other deep learning models and limited evaluation on a single dataset. The research does not provide detailed information regarding the feature extraction process.   |
| [41]            | The MBSaNet model is a combination of the Block Convolutional CNN mechanism and the SA (Self Attention) Module | 5000 Fundus Ocular Disease Intelligent Recognition (ODIR-2019) from the Peking University International Competition   | This research leverages deep learning and multimodal data to accurately predict cardiovascular risk, providing a potential tool for early detection and prevention.  | The number of images in some categories is limited, which may affect model performance. Additionally the use of different camera equipment in different environmental conditions results in high variability in fundus images.                        |
| [42]            | Multimodal Method (Combination of Dense-Net169, U-Net and DNN)   | Fundus images for CVD (Cardiovascular Disease) and non-CVD from Samsung Medical Center  | The strength of this study lies in achieving a high level of accuracy in classifying diabetic retinopathy using the proposed approach.   | It relies on limited data sets and may not fully capture the complexity and diversity of cardiovascular disease, thereby limiting its generalizability.   |
| [43]            | LuNet-LightGBM   | Messidor (1200 Fundus Images), APTOS 2019 (1928 images), and IDRiD (516)  | RetMobileNet has high accuracy (>90%), surpassing other CNN models. The model can automatically identify multi-class eye diseases (12 classes). The developed model shows potential to enhance the accuracy of eye disease identification based on fundus images with more efficient compatibility.  | Limited information on methodology and lack of comparison with state-of-the-art techniques.   |
| Proposed Method | Modification MobileNetV1 and MobileNetV2 (RetMobileNet)  | ARIA, STARE, and RFMiD database. Image is 1855 consisting of 20 classes. This dataset has been validated by eye experts from SMEC Eye Hospital Medan, Indonesia |  | The level of accuracy can still be improved. The model developed is not yet capable of detecting all types of eye diseases, it is still limited to automatically identifying 12 classes of eye diseases.  |

Karsaz [34] introduced AlexNet-SVDD to classify Diabetic Retinopathy (DR) images. Raj et al. [35] proposed RDC-UNet to enhance fundus image quality. Elwin et al. [36] presented Ar-HGSO for DR detection and severity classification. Joseph [37] suggested AMF-RCNNs for predicting five eye disease classes. Tseng et al. [38] proposed MiniNet, outperforming existing models like DenseNet and MobileNet. Morano et al. [39] utilized VGG 16 for identifying Age-Related Macular Degeneration (AMD) and retinal lesions. Qin et al. [40] introduced MFgcForest for Diabetic Retinopathy

classification, improving accuracy compared to traditional methods. Wang et al. [41] proposed MBSaNet for classifying eight eye diseases with fewer parameters. Lee et al. [42] suggested a Multimodal Method for predicting cardiovascular diseases using retinal abnormalities and risk factors. Bapatla and Harikiran [43] developed LuNet-LightGBM for DR lesion segmentation and grading, enhancing classification with the Aquila Optimizer.

Based on the analysis and considerations from previous studies, it can be concluded that while some studies

demonstrate high accuracy, the corresponding time requirements tend to be lengthy, and vice versa. Moreover, a number of these studies have lack comparisons between the proposed CNN model and other state-of-the-art CNN architectures. Additionally, several studies utilized devices with relatively high specifications, and the scope of eye diseases discussed, classified, and identified was confined to only one or two diseases. There is still a scarcity of research aimed at identifying multi-class eye diseases. Therefore, it is imperative to conduct research that automatically classifies and identifies various eye diseases based on fundus images, rather than focusing solely on one or a few conditions. Furthermore, the consideration of using more compact and

efficient devices capable of effectively carrying out classification and identification should also be taken into account.

### 3. PROPOSED METHODOLOGY

#### 3.1 Research framework

The research framework (Figure 1) consists of six main stages, namely: Input Fundus Image, Data Preprocessing, Initial Training Phase, Accuracy Optimization, Model Development, and Evaluation and Identification.

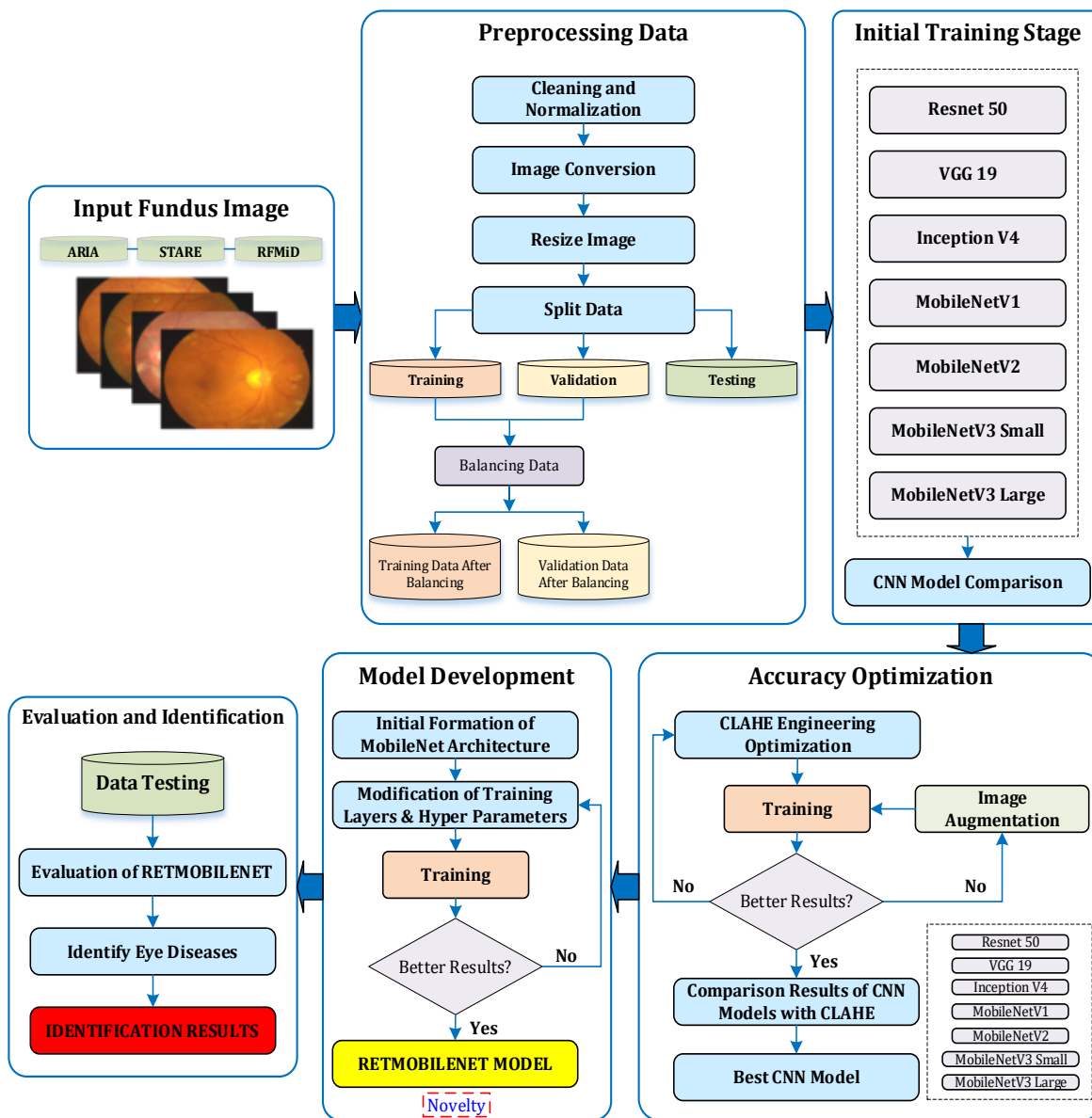


Figure 1. Research framework

In the Input Fundus Image stage, image data was obtained from sources such as public databases. After collecting image data, the next stage involved Data Preprocessing. Preprocessing is a pre-processing process aimed at preparing data before analysis. It involved processes like altering image formats, standardizing pixel intensities, cropping images, or eliminating noise or artifacts from the data. Subsequently, the Initial Training Phase entails the initial training or comparison of seven Convolutional Neural Network (CNN) models

(VGG19, InceptionV4, Resnet50, MobileNetV2, MobileNetV1, MobileNetV3 Large, and MobileNetV3 Small) to determine the optimal model. Development or adjustments will then be performed based on the chosen model. After the initial training, the next stage was to optimize accuracy using CLAHE (Contrast Limited Adaptive Histogram Equalization). The use of CLAHE aimed to enhance the contrast and clarity of images with the purpose of improving accuracy in object analysis or recognition. CLAHE enhances local contrast more

effectively, making details in various areas of the fundus image more visible. The implementation of CLAHE involves dividing the image into small, non-overlapping areas or tiles, then applying Histogram Equalization to each tile. Contrast is then limited: the histogram is clipped, meaning the maximum value allowed for the histogram bin is set at a predefined threshold to prevent any bin from dominating the histogram, thereby controlling noise amplification. The tiles are then combined using bilinear interpolation to eliminate artificial boundaries between them, resulting in a smooth overall image. By limiting contrast, CLAHE reduces the risk of noise amplification, which is crucial in medical images where clarity and accuracy are paramount. This method facilitates the identification and diagnosis of image conditions by enhancing the visibility of important structures such as blood vessels and the optic disc.

Subsequently, the Model Development Stage was performed that was a further optimization of the selected best CNN model. We found-tunes parameters, such as selecting the optimal learning rate, adding regularization, or applying other optimization techniques. After the development process was complete, we tested the model's performance on a separate testing dataset to obtain an objective evaluation. The final stage was Evaluation and Identification to assess the performance of the created model. Evaluation is conducted to determine the model that can identify or recognize patterns from previously unseen data. By evaluating the model, strengths and weaknesses can be identified, aiding in refining the model for better identification results.

The primary goals of model evaluation involved assessing accuracy, enhancing performance, optimizing the model, and determining its readiness for use. Various evaluation metrics, including accuracy, precision, recall, F1-score, ROC curve, and AUC, can be utilized based on the specific problem type and model goals. Expert verification will be conducted to assess the model's effectiveness in identifying diseases in fundus images.

### 3.2 Input fundus image

The Fundus Image input is from the MuReD (Multi-Label Retinal Disease) that combines the STARE dataset, RFMiD dataset, and ARIA dataset. The total number of images was 2208, comprising 20 classes [44]. Upon further observation, it was found that there were many missing data, resulting in a total of 1855 available images from the 20 classes. The dataset of fundus images utilized in this research had undergone validation by an eye specialist at SMEC Eye Hospital in Medan, Indonesia.

### 3.3 Preprocessing data

The goal of Data Preprocessing was to prepare the data for analysis or the creation of Deep Learning models. Proper Data Preprocessing improved the accuracy and reliability of Deep Learning models while preventing errors that may arise from using raw, unprepared data.

#### 3.3.1 Cleaning and normalization

In the previous discussion (Fundus Image Input), after joint observation with fundus image experts, out of the 1855 images available across 20 classes, there were 497 images diagnosed with multiple diseases in one image, and 64 duplicated images. Consequently, these images need to be removed to avoid disrupting the research process. Additionally, out of the 20

classes, we eliminated 8 classes with 94 images due to a very small and imbalanced number of images (as seen in Figure 2).

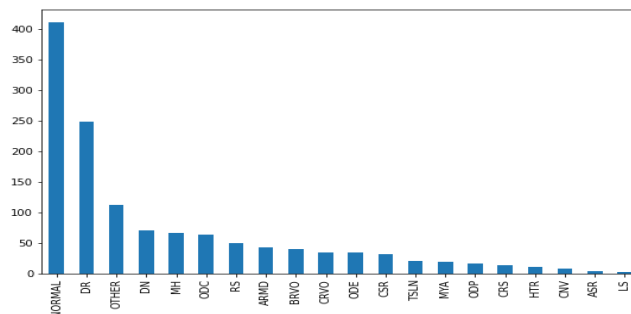


Figure 2. Graph of number of images (type of disease)

Based on Figure 2, it can be seen that the top 12 disease classes with the most images include Normal Retina (NORMAL), DN, DR, ODC, Media Haze (MH), Age-Related Macular Degeneration (ARMD), Retinitis (RS), Branch Retinal Vein Occlusion (BRVO), CRVO, Optic Disc Edema (ODE), Central Serous Retinopathy (CSR), and Other Diseases (OTHER). These 12 diseases can be used as the research dataset, while the remaining 8 disease classes with a total of 94 images were not used due to their insufficient quantity. Upon further analysis, it was found that 78 images were blurry and contained light reflections or shadows (lens flare artifacts) during image capture, rendering them unsuitable for the research process as they are considered damaged. After completing the data cleaning process, the dataset was available for research consists of 1122 images.

#### 3.3.2 Image conversion

Based on 1122 images, there were two types of image formats (\*.TIFF and \*.PNG). Therefore, all images in the \*.TIFF format will be converted to \*.PNG to ensure uniformity across all images. The conversion from \*.TIFF to \*.PNG in this research was conducted using the Python programming.

#### 3.3.3 Resize image

Image resizing was carried out due to varying resolutions, resulting in our images being reduced to 224 x 224 pixels. It is done considering the specifications of the research device, that was a notebook with an AMD RYZEN 7 2.49 GHz processor, 8 GB RAM, and a NVIDIA GFORCE GTX 1660Ti 6GB GPU.

#### 3.3.4 Split data

The data split technique consisted of dividing the dataset into three parts, comprising Training, Validation, and Testing data. The Training and Validation data were used as input images during the training process, while the Testing data were separated from the beginning to treat the images used as if they were new data in the identification testing, it is also known as blind testing. The ideal data split selection was performed using the K-Fold Cross-Validation method (K=10) [45]. The Training data split were divided into three parts: 60%, 70%, and 80% of the total image data, while the Validation data were 30%, 20%, and 10%.

#### 3.3.5 Balancing data

The data balancing technique was employed to make the sample count in each group proportional. In this phase, the technique used was oversampling with image augmentation methods on the minority class. The training data after data balancing for each class, utilizing augmentation techniques

initially, amounts to 500 images, resulting in a total of 6,000 training data after balancing. The validation data after data balancing for each class, utilizing augmentation techniques, amounts to 50 images. It resulted a total of 600 validation data after balancing.

### 3.4 Initial training stage

This stage is conducted to assess the initial performance of the CNN model after the data splitting and balancing stages. In the initial training phase, the capabilities of several standard CNN models, namely ResNet50, VGG19, InceptionV4, MobileNetV1, MobileNetV2, MobileNetV3Small, and MobileNetV3Large, will be evaluated using fundus image inputs. This stage resulted comparison of each CNN model's performance.

### 3.5 Accuracy optimization

The previously trained models based on fundus retina images using 7 CNN architecture models that was optimized for their accuracy using CLAHE. CLAHE, an image processing method was, utilized to improve image contrast while maintaining enhanced details in regions with low contrast levels [46]. It was incorporated as part of the process to enhance accuracy optimization [47]. The CLAHE process consists of three stages, namely: dividing the original image into  $M \times N$ -sized sub-images, calculating the histogram for each sub-image, and applying clipped histogram on each image. The CLAHE technique involved computing the average number of pixels at each grayscale level, as expressed in Eq. (1) [48], and applying the histogram clip limit, detailed in Eq. (2) [49].

$$N_{avg} = \frac{N_{CR-XP} \times N_{CR-YP}}{N_{gray}} \quad (1)$$

$$\beta = \frac{M}{N} \left( 1 + \frac{\alpha}{100} (S_{max} - 1) \right) \quad (2)$$

where  $N_{avg}$  described the average number of pixels at each grayscale level,  $N_{gray}$  represents the count of gray-level values within the sub-image,  $N_{CR-XP}$  signifies the pixel quantity along the X dimension of the sub-image, and  $N_{CR-YP}$  was the number of pixels in the Y dimension of the sub-image.  $\beta$  was the calculated clip limit, M was the number of pixels in the histogram block, N was the number of intensity intervals (bins) in the histogram,  $S_{max}$  was the maximum possible intensity, and  $\alpha$  is a factor influencing the extent to which contrast limitation is permitted.

CLAHE was chosen for optimization due to the following key aspects: Local Contrast Enhancement (Enhances contrast in local areas, crucial for medical imaging like fundus images, making small details such as blood vessels and the optic disc more visible). Noise Control (Limits excessive contrast enhancement, reducing noise amplification and maintaining image clarity). Visual Consistency (Uses bilinear interpolation to combine tiles, producing smoother images without visible boundaries). Discussion of alternative techniques for accuracy optimization: Histogram Equalization (Enhances contrast globally but can cause over-enhancement and excessive noise). Adaptive Histogram Equalization (Enhances local contrast but amplifies noise in homogeneous areas). Gaussian Filtering (Reduces noise but sacrifices sharpness and

important details). Median Filtering (Removes impulsive noise but may not preserve sharp edges). Wavelet Transform (Enhances contrast and reduces noise but is complex and computationally intensive). CLAHE is preferred because: Efficiency and Effectiveness (Enhances local details and controls noise efficiently, suitable for precise medical image analysis). Optimal Balance (Balances detail enhancement and noise control without significant added complexity). Proven Performance (Widely tested and used in medical imaging, consistently improving image quality for diagnostics).

Images subjected to CLAHE will undergo retraining. If the outcomes surpass the initial training, they can be compared across different architecture models. If there was no improvement, continuous data augmentation can be employed until superior results were achieved. Improved accuracy will lead to a comparison of the 7 CNN models based on CLAHE-optimized images, identifying MobileNetV1 and MobileNetV2 as the best models for further development.

### 3.6 Model development (proposed model)

This stage was undertaken to develop the best CNN models previously described, namely, to refine the MobileNetV1 and MobileNetV2 models as the basis for the proposed new model (RetMobileNet). The best CNN models (MobileNetV1 and MobileNetV2) can be enhanced, starting with Depth-Wise Convolution to prevent overfitting and boost computational efficiency while maintaining accuracy. These models undergone the retraining with various parameters like Data Split Ratio, Data Balancing, Image Enhancement, and advanced training hyperparameters (Epoch, Batch\_size, Image\_size, Weights, Learning rate, Activation Function, Optimizer, Metrics, and Callbacks) until optimal results were achieved. If results remained unsatisfactory, iterations with different parameters and training hyperparameters will be conducted. Additionally, Residual Blocks with Skip Connections will be introduced to help the model learn more concise representations. Further layer modifications, including the addition of 3 Dense Layers and 3 batch normalizations, will aim for an optimal model. It will be followed by the incorporation of an extra Dense Layer and Flatten. Finally, more Dense Layers will be added for making predictions. In total, 5 Dense Layers and 3 Batch Normalizations have been added, resulting in the creation of a new model called RetMobileNet. This model can be reevaluated against the previous 5 CNN models to determine if performance improvements have occurred.

Based on Figure 3, it can be observed that the proposed CNN model, RetMobileNet was constructed from two CNN architectures: MobileNetV1 and MobileNetV2. As a result, RetMobileNet had two inputs during the feature extraction process. The first input in the RetMobileNet model used MobileNetV1, augmented with skip connections to help enhance accuracy and address issues like vanishing or exploding gradients. These skip connections with residual blocks operated by connecting the output from one depthwise separable convolution to the input of the next depthwise separable convolution. In the second input, there was a convolutional layer with 128 filters and a (2,2) stride before being used as input for MobileNetV2. The output from each model included an average pooling layer that was then concatenated using the concatenate technique. It ensured that the features obtained from the outputs of MobileNetV1 and MobileNetV2 were weighted in the fully connected layers, consisting of Dense Layer, batch normalization, and Flatten

layer. In the final layer, an additional Dense Layer was added to make predictions (identification) of eye diseases using fundus images.

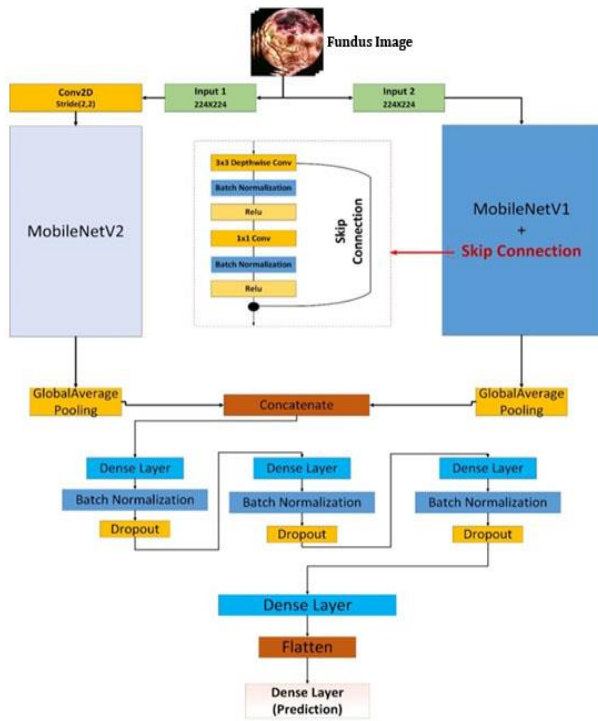


Figure 3. RetMobileNet (proposed model)

### 3.7 Evaluation and identification

The evaluation of the RetMobileNet model in identifying retinal fundus image diseases involved assessing the model's performance in classifying retinal fundus images as normal or identifying the type of disease present. The evaluation was conducted using validation or test data. The evaluation metrics included accuracy (1), precision (2), recall (3), and F1 score (4) [50], to gauge the extent to which the model can accurately recognize and classify retinal fundus images. The way to assess the model's ability to identify images was by conducting tests on a selection of image samples (test data).

$$Accuracy = \frac{(AB) - (AC)}{(AB + AD + AC + AE)} \quad (3)$$

$$Precision = \frac{AB}{(AB + AE)} \quad (4)$$

$$Recall = \frac{AB}{(AB + AD)} \quad (5)$$

$$F1\ Score = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (6)$$

where AB: true positive, AC: true negative, AD: false negative, and AE: false positive.

## 4. RESULT ANALYSIS AND DISCUSSION

### 4.1 Class diseases fundus image

After the cleaning and normalization of data, the dataset

available for research consisted of 1122 images (see Table 2). Discussion about the diversity of the dataset is crucial to ensure that the trained model can perform well on various types of fundus images. To create an effective model for real-world applications, the dataset must reflect the distribution of the global population. The dataset should include a range of eye medical conditions, from common to rare, to ensure the model can accurately detect and diagnose these conditions. The dataset should also have variations in image quality, including images with noise, poor lighting, and other artifacts, to ensure the model is robust against suboptimal imaging conditions.

Table 2. Research dataset after cleaning

| Class       | Diagnosis                        | Abbreviation | Amount |
|-------------|----------------------------------|--------------|--------|
| 1           | Normal Retina                    | NORMAL       | 384    |
| 2           | Diabetic Retinopathy             | DR           | 231    |
| 3           | Other Diseases                   | OTHER        | 106    |
| 4           | Drusen                           | DN           | 63     |
| 5           | Media Haze                       | MH           | 52     |
| 6           | Optic Disc Cupping               | ODC          | 62     |
| 7           | Retinitis                        | RS           | 46     |
| 8           | Age-Related Macular Degeneration | ARM D        | 42     |
| 9           | Branch Retinal Vein Occlusion    | BRVO         | 40     |
| 10          | Central Retinal Vein Occlusion   | CRVO         | 31     |
| 11          | Optic Disc Edema                 | ODE          | 34     |
| 12          | Central Serous Retinopathy       | CSR          | 31     |
| Total Image |                                  |              | 1122   |

Bias in the dataset can affect the performance and applicability of the model. If the dataset is dominated by images from one group or one type of medical condition, the trained model tends to perform well on that group but poorly on others. If images of certain medical conditions are underrepresented, the model may fail to correctly detect those conditions. In the medical field, dataset bias can have serious implications for diagnosis and patient care. Misdiagnosis due to a biased model can lead to adverse health outcomes. By ensuring the diversity and representativeness of the dataset and addressing bias, we can improve the model's performance and reliability in real-world applications, particularly in the medical field, which requires high levels of accuracy and trust.

### 4.2 Initial CNN model training

Based on the training of the 7 CNN models, MobileNetV1 and MobileNetV2 outperform the other models. MobileNetV1 excels in the 70:20:10% and 80:10:10% data split ratios, while MobileNetV2 performed better in the 60:30:10% data split ratio. However, overall, the accuracy of MobileNetV1 and MobileNetV2 did not differ significantly.

Based on the training results presented in Table 3, the highest accuracy was obtained using the 80:10:10% split ratio with MobileNetV1=57.86% and MobileNetV2=56.43%. It is the reason why MobileNetV1 and MobileNetV2 were chosen to be developed into RetMobileNet. Additionally, both models had a shorter training time compared to other models. Based on the experimental results, the split ratio significantly impacts the CNN model's initial training accuracy, with a larger training set portion enhancing the model's ability to learn patterns and improve accuracy. Selecting the right split ratio is crucial for effective learning and accurate generalization to new data, a balanced split ratio of 80:10:10 (training set:validation set: test set) ensures sufficient data for reliable model training, validation, and testing.

**Table 3.** Initial training results with the CNN model

| No. | Model             | Split Ratio / Training Accuracy |              |              |
|-----|-------------------|---------------------------------|--------------|--------------|
|     |                   | 60:30:10%                       | 70:20:10%    | 80:10:10%    |
| 1   | ResNet50          | 25.14                           | 26.07        | 44.71        |
| 2   | InceptionV4       | 35.47                           | 34.66        | 44.05        |
| 3   | VGG19             | 44.42                           | 45.76        | 49.05        |
| 4   | MobileNetV1       | 49.38                           | <b>52.57</b> | <b>57.86</b> |
| 5   | MobileNetV2       | <b>54.75</b>                    | 52.36        | 56.43        |
| 6   | MobileNetV3_Small | 28.03                           | 24.71        | 35.71        |
| 7   | MobileNetV3_Large | 23.97                           | 24.40        | 29.52        |

#### 4.3 Utilization of the CLAHE technique

To assess the impact of CLAHE on accuracy improvement, testing was conducted with various parameters to determine the most ideal values for tilegridsize and cliplimit on the images. Based on the test results, summarized in Table 4, the most suitable parameters were identified with the adjustment of Clip Limit = 5 and tileGridSize = (12,12).

Table 5 displays image quality samples post-testing with the CLAHE technique using tilegridsize and cliplimit.

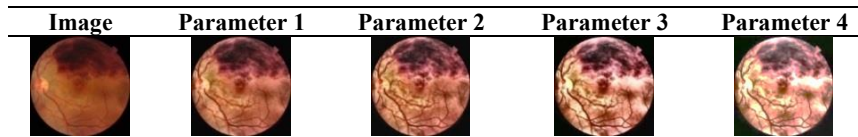
Data were augmented by doubling samples to prevent overfitting, enhance training data, improve model robustness, boost performance, and prevent resampling bias. Post-

augmentation, training data increased to 1000 images per class (totaling 12000 images), while validation and testing data remained at 600 images each. CNN model comparison resulted with CLAHE, using images subjected to advance augmentation as presented in Table 6.

Based on Table 6, it can be observed that the accuracy of each CNN model using K-Fold Cross Validation K=10 had increased after advanced augmentation. The MobileNetV1 model with K-Fold Cross Validation K=5 performed the best with an accuracy of 86.66%. Both MobileNetV1 and MobileNetV2 models achieved better accuracy compared to other models.

**Table 4.** CLAHE results with several parameters

| No. | Parameter                              | Results      |              |              |
|-----|--|--------------|--------------|--------------|
|     |  | Acc          | Prec         | Rec          |
| 1   | Clip Limit = 3, tileGridSize = (8,8)   | 69.33        | 71.23        | 67.67        |
| 2   | Clip Limit = 4, tileGridSize = (10,10) | 70.83        | 72.89        | 67.67        |
| 3   | Clip Limit = 5, tileGridSize = (12,12) | <b>70.67</b> | <b>73.29</b> | <b>69.50</b> |
| 4   | Clip Limit = 6, tileGridSize = (14,14) | 70.33        | 73.26        | 68.50        |

**Table 5.** Sample image quality with CLAHE parameters**Table 6.** CNN model results after advanced augmentation

| No | Model             | K-Fold Cross Validation (K=10) |       |       |       |       |       |       |       |       |       |
|----|-------------------|--------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|    |                   | K=1                            | K=2   | K=3   | K=4   | K=5   | K=6   | K=7   | K=8   | K=9   | K=10  |
| 1  | ResNet50          | 23.13                          | 28.76 | 34.15 | 28.77 | 47.72 | 27.66 | 26.50 | 16.47 | 47.16 | 31.83 |
| 2  | InceptionV4       | 18.30                          | 38.33 | 43.88 | 38.57 | 18.83 | 48.13 | 38.70 | 37.68 | 33.30 | 19.08 |
| 3  | VGG19             | 29.33                          | 41.11 | 38.77 | 48.61 | 37.44 | 40.22 | 48.53 | 38.34 | 48.54 | 48.33 |
| 4  | MobileNetV1       | 83.99                          | 80.33 | 82.33 | 63.49 | 86.66 | 80.66 | 86.50 | 70.49 | 83.16 | 83.33 |
| 5  | MobileNetV2       | 55.00                          | 64.33 | 74.00 | 76.33 | 70.99 | 80.16 | 56.49 | 37.50 | 66.00 | 52.16 |
| 6  | MobileNetV3_Small | 38.15                          | 48.79 | 18.93 | 19.66 | 23.57 | 38.35 | 38.12 | 18.24 | 28.90 | 20.78 |
| 7  | MobileNetV3_Large | 8.33                           | 8.33  | 8.33  | 8.33  | 8.33  | 8.33  | 8.33  | 8.33  | 8.33  | 8.33  |

#### 4.4 Forerunner of RedMobileNet

RetMobileNet is built based on MobileNetV1 and MobileNetV2 that outperformed several CNN models trained in the previous stage. The development's inception started with the use of training hyperparameters: Epoch = 100, Batch\_size = 16, Image\_size = 224×224 pixels, 3 channels, Weights = ImageNet, Learning rate = 0.0001, Activation Function = Softmax, Optimizer = Adam (learning\_rate = learning\_rate, beta\_1 = 0.9, beta\_2 = 0.999, amsgrad = False), Metrics = (Accuracy, precision, recall), Callbacks = Learning rate scheduler: learning\_rate \* (0.1 \*\* int(epoch / 10)), Early Stopping: EarlyStopping(monitor = 'loss', min\_delta = 0.001,

patience = 10, verbose = 1, mode = "min"). The results can be seen in Table 7.

#### 4.5 Modification layer and hyperparameter

The initial modification is performed by adding batch normalization to reduce overfitting through normalizing the input given to each layer of the neural network. Meanwhile, dropout aids in reducing dependencies between units within the layer and prevents overfitting. The enhancements made by incorporating batch normalization are detailed in Table 8, illustrating the developmental impact on the model.

The development model by adding batch normalization to



the output layers significantly improved accuracy. The evaluation results based on accuracy metrics on training, validation, and testing data were obtained as follows: 99.92%, 88.17%, and 88.00%, respectively. Meanwhile, based on precision metrics on Training, validation, and Testing data, the results were as follows: 99.93%, 89.38%, and 89.42%, respectively. In terms of Recall/Sensitivity metrics, the evaluations on training, validation, and testing data were 99.92%, 88.00%, and 87.33%, respectively. Furthermore, the addition of Residual Blocks with skip connections was implemented to enhance the model's ability. The use of residual blocks and skip connections allowed information to

skip several blocks between convolution layers, helping the model learn from more concise and less specific representations in the Training data. The impact of adding these residual blocks is detailed in Table 9, which shows the progression and improvements in the model's architecture.

Adding a residual block improved training percentage, precision, and recall compared to using batch normalization alone. Subsequent development includes adding a dense layer, altering output dimensions through matrix-vector multiplication. Detailed analyses for model development can be seen in Table 10.

**Table 7.** Development of initial model of RetMobileNet architecture

| Development Model | Accuracy |            |         | Precision |            |         | Recall   |            |         |
|-------------------|----------|------------|---------|-----------|------------|---------|----------|------------|---------|
|                   | Training | Validation | Testing | Training  | Validation | Testing | Training | Validation | Testing |
| Base Model        | 99.96    | 85.50      | 85.17   | 99.96     | 86.56      | 87.35   | 99.96    | 84.83      | 84.00   |

**Table 8.** Development model by adding batch normalization

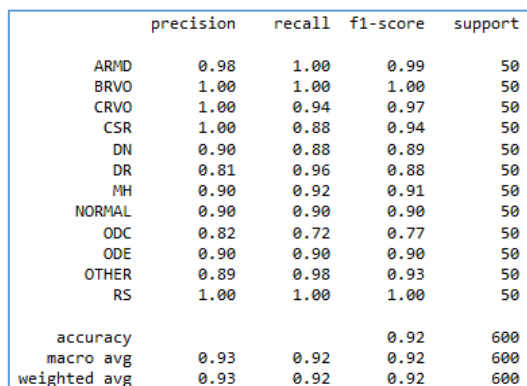
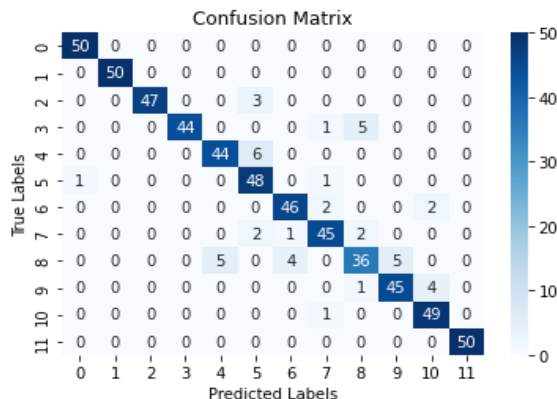
| Development Model   | Accuracy |            |         | Precision |            |         | Recall   |            |         |
|---------------------|----------|------------|---------|-----------|------------|---------|----------|------------|---------|
|                     | Training | Validation | Testing | Training  | Validation | Testing | Training | Validation | Testing |
| Batch Normalization | 99.92    | 88.17      | 88.00   | 99.93     | 89.38      | 89.42   | 99.92    | 87.00      | 87.33   |

**Table 9.** Development model by adding residual blocks

| Development Model | Accuracy |            |         | Precision |            |         | Recall   |            |         |
|-------------------|----------|------------|---------|-----------|------------|---------|----------|------------|---------|
|                   | Training | Validation | Testing | Training  | Validation | Testing | Training | Validation | Testing |
| Residual Blocks   | 99.92    | 89.83      | 90.33   | 99.92     | 90.73      | 90.59   | 99.92    | 89.67      | 89.83   |

**Table 10.** Development model by adding dense layer, batch normalization, dropout, and flatten

| No. | Development Model                                    | Parameters | Times | Training (Acc, Prec, Rec) | Validation (Acc, Prec, Rec) | Testing (Acc, Prec, Rec) |
|-----|--|------------|-------|---------------------------|-----------------------------|--------------------------|
| 1   | Dense 512 + Batch Normalization + Dropout 0.2 +      | 6,597,260  | 194ms | 99.92                     | 93.00                       | 90.17                    |
|     | Dense 256 + Batch Normalization + Dropout (0.2) +    |            |       | 99.92                     | 93.46                       | 90.25                    |
|     | Dense 128 + Flatten.                                 |            |       | 99.92                     | 92.83                       | 89.50                    |
| 2   | Dense 1280 + Dense 1280 + Batch Normalization +      | 8,519,948  | 190ms | 99.92                     | 94.33                       | 90.00                    |
|     | Dense 512 + Dense 512 + Batch Normalization + Dense  |            |       | 99.92                     | 94.33                       | 90.12                    |
|     | 128 + Dense 128 + Flatten                            |            |       | 99.92                     | 94.33                       | 89.67                    |
| 3   | Dense 512 + Dense 512 + Batch Normalization + Dense  | 6,622,028  | 189ms | 99.92                     | 92.50                       | 90.50                    |
|     | 256 + Dense 256 + Batch Normalization + Dense 128 +  |            |       | 99.92                     | 92.79                       | 90.74                    |
|     | Dense 128 + Flatten                                  |            |       | 99.92                     | 92.17                       | 89.83                    |
| 4   | Dense 1280 + Batch Normalization + Dense 512 + Batch | 8,188,940  | 195ms | 99.92                     | 93.00                       | 90.83                    |
|     | Normalization + Dense 128 + Batch Normalization +    |            |       | 99.92                     | 93.28                       | 91.43                    |
|     | Dense 64 + Flatten                                   |            |       | 99.92                     | 92.50                       | 90.67                    |
| 5   | Dense 1280 + Batch Normalization + Dropout (0.2) +   | 8,188,940  | 193ms | 99.92                     | 93.50                       | 91.67                    |
|     | Dense 512 + Batch Normalization + Dropout (0.2) +    |            |       | 99.92                     | 93.94                       | 91.76                    |
|     | Dense 128 + Batch Normalization + Dropout (0.2) +    |            |       | 99.92                     | 93.00                       | 91.00                    |
|     | Dense 64 + Flatten                                   |            |       |                           |                             |                          |



**Figure 4.** Confusion matrix RetMobileNet on testing data

**Figure 5.** Classification report RetMobileNet on testing data

In Table 10, it can be observed that the development of model number 5 achieved the highest accuracy of 99.92% in the Training data, 93.94% in the Validation data, and 91.76% in the Testing data. The model with the best parameters and hyperparameter settings obtained in our experiments is named RetMobileNet.

#### 4.6 Evaluation and identification

The RetMobileNet model can be assessed on 600 Retina Fundus images across 12 disease classes in the testing data. Model accuracy, evaluated using the confusion matrix and classification report, is depicted in Figures 4 and 5.

The RetMobileNet model can detected all types of diseases with >90% accuracy. However, there is one type of disease, ODC, that the model struggles to predict well, it achieved more than 82% accuracy. This limitation is due to the model's ability to predict only 36 images, while the other 14 images were predicted as different diseases (5 as DN, 4 as MH, and 5 as ODE).

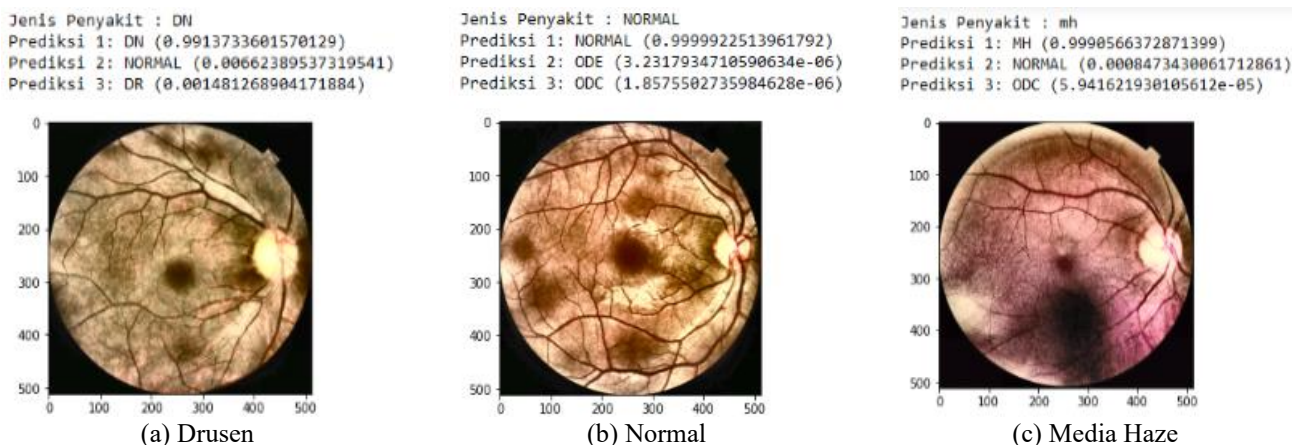
Overall, the RetMobileNet model had an accuracy of 0.92 (92%), precision of 0.93 (93%), recall of 0.92 (92%), and an

f1-score of 0.92 (92%). Subsequently, further testing was conducted with seven other models to assess the capability of this model in classifying retina fundus images.

RetMobileNet exhibits superior capabilities compared to other CNN models (Table 11). Figure 6 represents a sample test for the automatic identification of eye diseases based on this model, displaying three possible disease predictions.

**Table 11.** Comparison of CNN model testing

| Model               | Accuracy   | Precision  | Recall     | F1-Score   |
|---------------------|------------|------------|------------|------------|
| ResNet50            | 75%        | 78%        | 75%        | 75%        |
| InceptionV4         | 80%        | 82%        | 80%        | 80%        |
| VGG19               | 75%        | 78%        | 75%        | 76%        |
| MobileNetV1         | 86%        | 88%        | 86%        | 85%        |
| MobileNetV2         | 86%        | 86%        | 86%        | 85%        |
| MobileNetV3_Small   | 53%        | 55%        | 53%        | 52%        |
| MobileNetV3_Large   | 57%        | 59%        | 57%        | 57%        |
| <b>RetMobileNet</b> | <b>92%</b> | <b>93%</b> | <b>92%</b> | <b>92%</b> |



**Figure 6.** Identification with RetMobileNet Model (Sample)

Sample (a) was originated from an image with DN (Drusen) eye condition, and the highest prediction showed DN with a confidence score of 99.13%. Sample (b) was taken from an image with NORMAL eye condition, and the top prediction aligns with NORMAL eye condition with a confidence score of 99.99%. Sample (c) was sourced from an image with MH (Media Haze) eye condition, and the leading prediction came with a confidence score of 99.90%.

#### 5. CONCLUSIONS

The generalization of RetMobileNet to other datasets and populations may vary, as the obtained results could be specific to the dataset used in this research. Significant advancements in medical imaging, particularly in identifying common eye diseases from retina fundus images, have been achieved through computational models. The Retina MobileNet (RetMobileNet), which combines the strengths of MobileNetV1 and MobileNetV2, is noteworthy for its transformative potential. With several enhancements such as Depthwise Convolutional layers, Batch Normalization, Residual Blocks with Skip Connections, and the inclusion of 4 Dense Layers, this model has proven to be a robust tool for retina image identification. Its impressive performance

metrics, including accuracy, precision, recall, and f1 score, consistently demonstrate scores of over 90%, indicating statistical significance and clinical relevance. The high performance of this model ensures its crucial reliability for early and accurate diagnosis of eye diseases, given their immutable nature. The efficiency and compatibility of the RetMobileNet model make it suitable for real-world clinical settings, especially when rapid and precise diagnosis is paramount. Future research could explore further optimization by integrating RetMobileNet with emerging computational models and implementing real-time feedback mechanisms for continuous refinement based on practical usage. Collaborative studies involving technology developers and ophthalmologists can provide valuable insights to meet evolving clinical needs. Additionally, evaluating the model's application in various medical imaging scenarios can enhance its usability and impact.

#### ACKNOWLEDGMENT

Thank you to the all academic communities of Universitas Putra Indonesia YPTK Padang for their assistance in conducting this research. Special thanks are extended to Dr. Andrea Radotma Silitonga, Sp. M, and the SMEC Medan Eye

Hospital, Indonesia, for their help in validating retinal fundus image data, it enables the research to be carried out successfully.

## REFERENCES

- [1] Mekjavic, P.J., Tipton, M.J., Mekjavic, I.B. (2021). The eye in extreme environments. *Experimental Physiology*, 106(1): 52-64. <https://doi.org/10.1113/EP088594>
- [2] Stepp, M.A., Menko, A.S. (2021). Immune responses to injury and their links to eye disease. *Translational Research*, 236: 52-71. <https://doi.org/10.1016/j.trsl.2021.05.005>
- [3] Zang, P., Hormel, T.T., Hwang, T.S., Bailey, S.T., Huang, D., Jia, Y. (2023). Deep-learning-aided diagnosis of diabetic retinopathy, age-related macular degeneration, and glaucoma based on structural and angiographic OCT. *Ophthalmology Science*, 3(1): 100245. <https://doi.org/10.1016/j.xops.2022.100245>
- [4] Ong, J., Stewart, C., Smith, R., Roberts, D., Fisher, S., Shaw, T., Lee, A. (2022). Terrestrial health applications of visual assessment technology and machine learning in spaceflight associated neuro-ocular syndrome. *NPJ Microgravity*, 8(1): 37. <https://doi.org/10.1038/s41526-022-00222-7>
- [5] Coan, L.J., Williams, B.M., Adithya, V.K., Upadhyaya, S., Alkafri, A., Czanner, S., Venkatesh, R., Willoughby, C.E., Kavitha, S., Czanner, G. (2023). Automatic detection of glaucoma via fundus imaging and artificial intelligence: A review. *Survey of Ophthalmology*, 68(1): 17-41. <https://doi.org/10.1016/j.survophthal.2022.08.005>
- [6] Bennett, J.L., Costello, F., Chen, J.J., Petzold, A., Biousse, V., Newman, N.J., Galetta, S.L. (2023). Optic neuritis and autoimmune optic neuropathies: Advances in diagnosis and treatment. *The Lancet Neurology*, 22(1): 89-100. [https://doi.org/10.1016/S1474-4422\(22\)00187-9](https://doi.org/10.1016/S1474-4422(22)00187-9)
- [7] Bispo, P.J., Belanger, N., Li, A., Liu, R., Susarla, G., Chan, W.L., Chodosh, J., Gilmore, M.S., Sobrin, L. (2023). An all-in-one highly multiplexed diagnostic assay for rapid, sensitive, and comprehensive detection of intraocular pathogens. *American Journal of Ophthalmology*, 250: 82-94. <https://doi.org/10.1016/j.ajo.2023.01.021>
- [8] Sarki, R., Ahmed, K., Wang, H., Zhang, Y., Ma, J., Wang, K. (2021). Image preprocessing in classification and identification of diabetic eye diseases. *Data Science and Engineering*, 6(4): 455-471. <https://doi.org/10.1007/s41019-021-00167-z>
- [9] Singh, L.K., Khanna, M., Thawkar, S., Singh, R. (2022). Collaboration of features optimization techniques for the effective diagnosis of glaucoma in retinal fundus images. *Advances in Engineering Software*, 173: 103283. <https://doi.org/10.1016/j.advengsoft.2022.103283>
- [10] Liu, H., Li, R., Zhang, Y., Zhang, K., Yusufu, M., Liu, Y., Mou, D., Chen, X., Tian, J., Li, H., Fan, S., Tang, J., Wang, N. (2023). Economic evaluation of combined population-based screening for multiple blindness-causing eye diseases in China: A cost-effectiveness analysis. *The Lancet Global Health*, 11(3): e456-e465. [https://doi.org/10.1016/S2214-109X\(22\)00554-X](https://doi.org/10.1016/S2214-109X(22)00554-X)
- [11] Konyaev, D.A., Popova, E.B., Titov, A.A., Agarkov, N.M., Aksenov, V.V. (2021). The prevalence of eye diseases in the elderly population is a global problem of modernity. *Zdravookhranenie Rossiiskoi Federatsii/Ministerstvo zdravookhraneniia RSFSR*, 65(1): 62-68. <https://doi.org/10.47470/0044-197X-2021-65-1-62-68>
- [12] Bourne, R., Steinmetz, J. D., Flaxman, S., et al. (2021). Trends in prevalence of blindness and distance and near vision impairment over 30 years: An analysis for the Global Burden of Disease Study. *The Lancet Global Health*, 9(2): e130-e143. [https://doi.org/10.1016/S2214-109X\(20\)30425-3](https://doi.org/10.1016/S2214-109X(20)30425-3)
- [13] Beare, N.A.V., Bastawrous, A. (2023). 71-Ophthalmology in the tropics and sub-tropics. In *Manson's Tropical Diseases (Twenty fourth Edition)*, pp. 954-992. <https://doi.org/10.1016/B978-0-7020-7959-7.00071-3>
- [14] Gao, H., Chen, X.N., Shi, W.Y. (2019). Analysis of the prevalence of blindness and major blinding diseases in China. *Chinese Journal of Ophthalmology*, 55(8): 625-628. <https://doi.org/10.3760/cma.j.issn.0412-4081.2019.08.016>
- [15] Wong, K.K., Ayoub, M., Cao, Z., Chen, C., Chen, W., Ghista, D.N., Zhang, C.W. (2023). The synergy of cybernetical intelligence with medical image analysis for deep medicine: A methodological perspective. *Computer Methods and Programs in Biomedicine*, 240: 107677. <https://doi.org/10.1016/j.cmpb.2023.107677>
- [16] Deperlioglu, O., Kose, U., Gupta, D., Khanna, A., Giampaolo, F., Fortino, G. (2022). Explainable framework for Glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation. *Future Generation Computer Systems*, 129: 152-169. <https://doi.org/10.1016/j.future.2021.11.018>
- [17] Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R.K., Kumar, P. (2021). Artificial intelligence to deep learning: Machine intelligence approach for drug discovery. *Molecular Diversity*, 25: 1315-1360. <https://doi.org/10.1007/s11030-021-10217-3>
- [18] Zhang, Y., Farrugia, N., Bellec, P. (2022). Deep learning models of cognitive processes constrained by human brain connectomes. *Medical Image Analysis*, 80: 102507. <https://doi.org/10.1016/j.media.2022.102507>
- [19] Touretzky, D., Gardner-McCune, C., Seehorn, D. (2023). Machine learning and the five big ideas in AI. *International Journal of Artificial Intelligence in Education*, 33(2): 233-266. <https://doi.org/10.1007/s40593-022-00314-1>
- [20] Khan, W., Daud, A., Khan, K., Muhammad, S., Haq, R. (2023). Exploring the frontiers of deep learning and natural language processing: A comprehensive overview of key challenges and emerging trends. *Natural Language Processing Journal*, 4: 100026. <https://doi.org/10.1016/j.nlp.2023.100026>
- [21] Amin, R., Al Ghamdi, M.A., Almotiri, S.H., Alruily, M. (2021). Healthcare techniques through deep learning: Issues, challenges and opportunities. *IEEE Access*, 9: 98523-98541. <https://doi.org/10.1109/ACCESS.2021.3095312>
- [22] Singh, N., Sabrol, H. (2021). Convolutional neural networks-an extensive arena of deep learning. A comprehensive study. *Archives of Computational Methods in Engineering*, 28(7): 4755-4780. <https://doi.org/10.1007/s11831-021-09551-4>
- [23] Li, X., Li, Z., Qiu, H., Hou, G., Fan, P. (2023). An overview of hyperspectral image feature extraction,

- classification methods and the methods based on small samples. *Applied Spectroscopy Reviews*, 58(6): 367-400. <https://doi.org/10.1080/05704928.2021.1999252>
- [24] Upadhyay, P.K., Rastogi, S., Kumar, K.V. (2022). Coherent convolution neural network based retinal disease detection using optical coherence tomographic images. *Journal of King Saud University - Computer and Information Sciences*, 34(10): 9688-9695. <https://doi.org/10.1016/j.jksuci.2021.12.002>
- [25] Nakayama, L.F., Ribeiro, L.Z., Dychiao, R.G., Zamora, Y.F., Regatieri, C.V.S., Celi L.A., Silva, P., Sobrin, L., Jr, R.B. (2023). Artificial intelligence in uveitis: A comprehensive review. *Survey of Ophthalmology*, 68(4): 669-677. <https://doi.org/10.1016/j.survophthal.2023.02.007>
- [26] Wang, J.Z., Lu, N.H., Du, W.C., Liu, K.Y., Hsu, S.Y., Wang, C.Y., Chen, Y.J., Chang, L.C., Twan, W.H., Chen, T.B., Huang, Y.H. (2023). Classification of color fundus photographs using fusion extracted features and customized CNN models. *Healthcare*, 11(15): 2228. <https://doi.org/10.3390/healthcare11152228>
- [27] Rodriguez-Conde, I., Campos, C., Fdez-Riverola, F. (2022). Optimized convolutional neural network architectures for efficient on-device vision-based object detection. *Neural Computing and Applications*, 34(13): 10469-10501. <https://doi.org/10.1007/s00521-021-06830-w>
- [28] Prasetyo, E., Purbaningtyas, R., Adityo, R.D., Suciati, N., Fatichah, C. (2022). Combining MobileNetV1 and depthwise separable convolution bottleneck with expansion for classifying the freshness of fish eyes. *Information Processing in Agriculture*, 9(4): 485-496. <https://doi.org/10.1016/j.inpa.2022.01.002>
- [29] Togatorop, P.R., Pratama, Y., Sianturi, A.M., Pasaribu, M.S., Sinaga, P.S. (2023). Image preprocessing and hyperparameter optimization on pretrained model MobileNetV2 in white blood cell image classification. *Image*, 12(3): 1210-1223. <https://doi.org/10.11591/ijai.v12.i3.pp1210-1223>
- [30] Lou, J., Xu, J., Zhang, Y., Sun, Y., Fang, A., Liu, J., Mur, L.A.J., Ji, B. (2022). PPsNet: An improved deep learning model for microsatellite instability high prediction in colorectal cancer from whole slide images. *Computer Methods and Programs in Biomedicine*, 225: 107095. <https://doi.org/10.1016/j.cmpb.2022.107095>
- [31] Entezari, A., Aslani, A., Zahedi, R., Noorollahi, Y. (2023). Artificial intelligence and machine learning in energy systems: A bibliographic perspective. *Energy Strategy Reviews*, 45: 101017. <https://doi.org/10.1016/j.esr.2022.101017>
- [32] Aytac, U.C., Güneş, A., Ajlouni, N. (2022). A novel adaptive momentum method for medical image classification using convolutional neural network. *BMC Medical Imaging*, 22(1): 34. <https://doi.org/10.1186/s12880-022-00755-z>
- [33] Aydınođan, G., Kavaklı, K., Şahin, A., Artal, P., Ürey, H. (2021). Applications of augmented reality in ophthalmology. *Biomedical Optics Express*, 12(1): 511-538. <https://doi.org/10.1364/boe.405026>
- [34] Karsaz, A. (2022). A modified convolutional neural network architecture for diabetic retinopathy screening using SVDD. *Applied Soft Computing*, 125: 109102. <https://doi.org/10.1016/j.asoc.2022.109102>
- [35] Raj, A., Shah, N.A., Tiwari, A.K. (2022). A novel approach for fundus image enhancement. *Biomedical Signal Processing and Control*, 71: 103208. <https://doi.org/10.1016/j.bspc.2021.103208>
- [36] Elwin, J.G.R., Mandala, J., Maram, B., Kumar, R.R. (2022). Ar-HGSO: Autoregressive-Henry Gas Sailfish Optimization enabled deep learning model for diabetic retinopathy detection and severity level classification. *Biomedical Signal Processing and Control*, 77: 103712. <https://doi.org/10.1016/j.bspc.2022.103712>
- [37] Joseph, S.S.A. (2022). Identification of retinal disease using anchor-free modified faster region. *International Journal of Advanced Computer Science and Applications*, 13(9): 490-499. <https://doi.org/10.14569/IJACSA.2022.0130956>
- [38] Tseng, F.H., Yeh, K.H., Kao, F.Y., Chen, C.Y. (2023). MiniNet: Dense squeeze with depthwise separable convolutions for image classification in resource-constrained autonomous systems. *ISA Transactions*, 132: 120-130. <https://doi.org/10.1016/j.isatra.2022.07.030>
- [39] Morano, J., Hervella, Á.S., Rouco, J., Novo, J., Fernández-Vigo, J.I., Ortega, M. (2023). Weakly-supervised detection of AMD-related lesions in color fundus images using explainable deep learning. *Computer Methods and Programs in Biomedicine*, 229: 107296. <https://doi.org/10.1016/j.cmpb.2022.107296>
- [40] Qin, X., Chen, D., Zhan, Y., Yin, D. (2023). Classification of diabetic retinopathy based on improved deep forest model. *Biomedical Signal Processing and Control*, 79: 104020. <https://doi.org/10.1016/j.bspc.2022.104020>
- [41] Wang, K., Xu, C., Li, G., Zhang, Y., Zheng, Y., Sun, C. (2023). Combining convolutional neural networks and self-attention for fundus diseases identification. *Scientific Reports*, 13(1): 76. <https://doi.org/10.1038/s41598-022-27358-6>
- [42] Lee, Y.C., Cha, J., Shim, I., Park, W.Y., Kang, S.W., Lim, D.H., Won, H.H. (2023). Multimodal deep learning of fundus abnormalities and traditional risk factors for cardiovascular risk prediction. *NPJ Digital Medicine*, 6(1): 14. <https://doi.org/10.1038/s41746-023-00748-4>
- [43] Bapatla, S., Harikiran, J. (2023). LuNet-LightGBM: An effective hybrid approach for lesion segmentation and DR grading. *Computer Systems Science and Engineering*, 46(1): 597-617. <https://doi.org/10.32604/csse.2023.034998>
- [44] Rodríguez, M.A., AlMarzouqi, H., Liatsis, P. (2022). Multi-label retinal disease classification using transformers. *IEEE Journal of Biomedical and Health Informatics*, 27(6): 2739-2750. <https://doi.org/10.1109/JBHI.2022.3214086>
- [45] Zhang, X., Liu, C.A. (2023). Model averaging prediction by K-fold cross-validation. *Journal of Econometrics*, 235(1): 280-301. <https://doi.org/10.1016/j.jeconom.2022.04.007>
- [46] Ji, W., Peng, J., Xu, B., Zhang, T. (2023). Real-time detection of underwater river crab based on multi-scale pyramid fusion image enhancement and MobileCenterNet model. *Computers and Electronics in Agriculture*, 204: 107522. <https://doi.org/10.1016/j.compag.2022.107522>
- [47] Sheet, S.S.M., Tan, T.S., As'ari, M.A., Hitam, W.H.W., Sia, J.S.Y. (2022). Retinal disease identification using upgraded CLAHE filter and transfer convolution neural network. *ICT Express*, 8(1): 142-150. <https://doi.org/10.1016/j.icte.2021.05.002>
- [48] Alom, M., Ali, M. Y., Islam, M. T., Uddin, A.H.,

- Rahman, W. (2023). Species classification of brassica napus based on flowers, leaves, and packets using deep neural networks. *Journal of Agriculture and Food Research*, 14: 100658. <https://doi.org/10.1016/j.jafr.2023.100658>
- [49] Saberi, Z., Hashim, N., Ali, A., Boursier, P., Abdullah, J., Embi, Z.C. (2023). Adaptive contrast enhancement of satellite images based on histogram and non-linear transfer function methods. *IAENG International Journal of Applied Mathematics*, 53(1): 1-9.
- [50] Chamid, A.A., Widowati, Kusumaningrum, R. (2023). Multi-label text classification on indonesian user reviews using semi-supervised graph neural networks. *ICIC Express Letters*, 17(10): 1075-1084. <https://doi.org/10.24507/icicel.17.10.1075>