

Deep Learning-Based Gender Prediction from Eye Images: A Comparative Study Using EfficientNet and MobileNet

Khovita Zulkarimah¹, Yunnifa Nur Lailli¹, Abigail Pormes¹, Isnaini Zayyana Fitri¹, Rizky Amelia¹

¹Department of Informatics, Institut Teknologi Kalimantan, Balikpapan, Indonesia

Corresponding author: Rizky Amelia (rizky.amelia@lecturer.itk.ac.id)

To cite this article: K. Zulkarimah, Y. N. Lailli, A. Pormes, I. Z. Fitri, R. Amelia, "Deep Learning-Based Gender Prediction from Eye Images: A Comparative Study Using EfficientNet and MobileNet" *Innovative Informatics and Artificial Intelligence Research*, vol. 2, issue 1, 2026. [Online]. Available: <https://doi.org/10.35718/iiair.v2i1.8481446>

Abstract

Eye images offer a unique and rich sources of biometric information, with consistent patterns across the eyebrows, eyes, and eyelids, making them ideal for tasks like gender prediction. Gender classification through eye images is an approach for identification without the need for physical contact, making it suitable for device such as on resource constrained device like CCTV or smartphones. This study shows a comparative analysis of two lightweight deep learning architectures, EfficientNet and MobileNet, to identify which one is the optimal and best model. In the evaluation, we used diverse datasets consisting of 13.024 images from Kaggle and additional data from manual documentation. The experimental process in this study consist of several stages, namely data preprocessing, augmentation, model training using adjusted hyperparameters, and evaluation using accuracy, precision, recall, and F1-score. The results showed that EfficientNetB0 had better performance, reaching the highest test accuracy of 96.55%. EfficientNetB0 also showed good balance in classification with F1-scores of 0,97 for males and 0.96 for females, indicating that EfficientNetB0 performs better than MobileNetV3 because it is able to balance computational efficiency with predictive accuracy in classifying gender based on eye image.

Keywords: gender-eye prediction; image classification; machine learning; deep learning

1. Introduction

An image-based recognition and identification system is one of the technologies that is currently developing and has become an important component in many security systems [1].

Biometric technology can be used to automatically identify or recognize face based on different characteristics [2].

Generally, cloud based technologies for face verification, recognition, and attribute classification have become more advance in a significant way and are now widely applied in real-world biometric systems [3]. In this context, biometrics technology focuses on identifying through their unique characteristics, such as fingerprint, eyes, and facial features. Among these various biometric methods, eye recognition is considered to have higher accuracy and does not require physical contact, is non-invasive, and is more stable with consistent performance [4].

Gender classification has long been a common topic of research, but as time progresses, various new methods continue to emerge in efforts to enhance its accuracy [5]. Although there are lots of significant progress that has been made in gender classification to improve the accuracy of prediction result [6], further improvements are still necessary.

Facial expression and eye image recognition have become increasingly relevant with the advancement of technologies in devices such as computers, smartphones, and robots [7]. The advancement of deep learning technology through Convolutional Neural Networks (CNNs) has significantly improved the accuracy and efficiency of image classification systems. MobileNet is a model which suitable for real time mobile application due to its lightweight design and fast execution. Meanwhile, EfficientNetB0 is a model that use a combined scalability method that balances network depth, width, and resolution, which make it able to perform with low computational cost while achieving high accuracy [3][8][9].

The strong urgency of this research is due to the needs for precise and efficient automatic identification systems, especially on low resources device like CCTV cameras or smartphones. MobileNet was chosen because of its lightweight architecture, which is suitable for implementation

on portable devices. On the other hand, EfficientNet was selected for its ability to keep the accuracy and efficiency in balance way through compound scaling. As a new contribution to biometric system research, this research focuses on analyzing the performance of both model architectures in predicting gender based on eye images.

In this context, this study aims to develop a gender prediction system based on eye images by utilizing EfficientNet and MobileNet, efficient deep learning architectures. The eye images were chosen as the main object because they provide extensive biometric details and are compatible with portable and non-invasive systems. The gender prediction system that examined in this study, developed using two architectures, namely EfficientNetB0 and MobileNetV3. These two models are considered more lightweight compared to InceptionV3 and MobileNetV2 based on their parameters [10], and also refer to the study utilized by Listyo [11] which may lead to imbalances in terms of model complexity, computational efficiency, and prediction accuracy.

Therefore, this study utilizes these two lighter methods to provide a more balanced evaluation. Additionally, it also proposes a comparison between two efficient and lightweight models that have similar numbers of parameters and complexity levels [12]. The application of gender classification based on eye patterns in this study has numerous of real-world benefits, such as in the field of marketing. For example, targeted advertising can be achieved by automatically identify a customer's gender, age, or other characteristics, which allows business to offer more personalized products or service [13] based on their gender. This approach can also be applied to real world scenarios, such a place of worship, where the need for an accurate and efficient gender prediction system is important to help maintain security and order in the worship environment.

This study introduces several main contributions through the implementation of a gender classification system using eye images, as follows:

1. Developing a gender classification system based on eye images by implementing two lightweight deep learning architectures, namely EfficientNetB0 and MobileNetV3.
2. Comparing the performance of both models based on accuracy, inference time, and memory usage efficiency.
3. Showing the potential of eye patterns as an effective alternative for gender classification, especially in environment with limited computational resources.
4. Providing some insight into difference in the characteristics of eye shapes between males and females identified by the models, which can be utilized as a foundation for further biometrics research.
5. Testing the models ability to maintain accurate predictions when faced with varied eye images, such as those with low lighting, low resolution, or the differences between the right and left eyes, to ensure its robustness in practical scenarios.

2. Related Works

A Deep learning models was developed by Nguyen et al [14] to predict age and gender from facial images using UTKFace Dataset. Three architectures that were tested are Unet, MobileNets, dan EfficientNets. Among them, Unet achieve the highest accuracy in gender prediction at 97.22%, while MobileNets achieved the best performance in age prediction with the lowest Mean Absolute Error (MAE) of 2.448. The results show that when deep learning models are supported by suitable pre-processing and model configuration, the models are capable of producing accurate predictions despite data imbalance. Meanwhile, Muddaluru et al [15] do comparisons between a few of deep learning models performances such as InceptionNetV3, EfficientNetB0, DenseNet121, and MobileNetV3, showing that the gaussian filter produced the best results, achieving an accuracy of 96%. The gaussian filter significantly improved the quality of retinal image quality for diagnostic purposes, while the greyscale and Gabor filters performance was comparatively lower.

Firdaus et al [16] did research on gender classification through eye image analysis using Convolutional Neural Networks (CNN). The study used a datasets of 11,5255 eye images from Kaggle, and the process involved data preprocessing, model training, and performance evaluation through a confusion matrix. As a result, the CNN model achieved an accuracy of 94%, with precision of 97% for females and 90% for males, along with F1-scores of 94% for females and 93% for males. Meanwhile, Gilang [17] implemented the same method using 11.000 of eye images dataset from Kaggle, but the result showed an accuracy 94%, with 97% for females and 90% for males. Both of the studies showed better performance compared to K-Nearest Neighbor and Naive Bayes used in previous research.

Gender recognition from facial features using a remote camera was examined by Choirina et al. [18], who implemented the Haar Cascade Classifier approach. The detected facial features included the eyes, nose, and mouth, from several distances were measured to determine the gender. The study compared two classification methods that used, such as Euclidean Distance and the C4.5 algorithm. The results showed that C4.5 provided higher accuracy, particularly at a distance of 100-200 cm, achieving a maximum accuracy of 100%, demonstrating the methods effectiveness for long range applications. While Asmara et al. [19] employed Naive Bayes by utilizing facial features such as the eyes, nose, and mouth, which were also detected using the Haar Cascade Classifier technique but proceeded through Principal Component Analysis (PCA). The study used a dataset of 61 facial image consist of 25 females and 36 males, which were processed through preprocessing, feature extraction, and classification stages. The method achieved an accuracy of 80%, this was lower compared to the results obtained by the study of Choirina et al [18].

Gender classification was studied by Rosiani et al. [20] through the application of facial landmark detection to measure specific facial features, including the distance between eyes, nose width, and lip height. The datasets processed using Support Vector Machine (SVM) methods and test result showed that the systems were functional, although

the accuracy rate reached only about 60% because of the similarities in facial features between males and females. Further improvements are suggested, either by using more unique features or adopting different deep learning techniques. Meanwhile, Sani et al. [21] developed a real-time application for gender classification using Support Vector Machine (SVM), which analyze facial patterns and image texture through GLCM and SFTA techniques. The application achieved a testing accuracy of 92% and indicate effectiveness in identifying gender accurately, provided the subject was directly facing the camera under the sufficient lighting conditions.

Kumar and Bansal [22] implemented a modified MobileNetV2 model for real time face mask detection on image and video captures using a transfer learning approach via the Caffe framework. The datasets included both static images and video captures from a webcam to evaluate the models performance. The results showed that the proposed method achieved a high accuracy of 99.64%. The model was also successfully implemented in real-time video surveillance by detecting facial location and efficiently classifying mask usage. Meanwhile, the development of real-time facial expression recognition system was studied by Ulhaq et al. [23] on an android device using the SSD method and the MobileNetV2 architecture. The model was trained using TensorFlow and TFLite Model Make and achieved an average accuracy of 51,2% and it was capable of processing image at speed of 30 frames per second (fps). Although the accuracy is not optimal yet, but the processing speed demonstrates the potential for implementing a lightweight and responsive facial expression recognition system on mobile device.

Our study employed an experimental approach by comparing two deep learning architectures, namely EfficientNetB0 and MobileNetV3, in the task of gender classification based on eye images. Hyperparameter evaluation was conducted for both models using validation data. The adjustment of hyperparameters which are number of epoch, learning rate, and batch size was performed using a simple grid search approach and stepwise experiments to find the optimal hyperparameter combination to minimize loss and maximize accuracy. In addition, the internal structure of the classifier was also evaluated for each model, such as the number of dense layers units and dropout rate.

3. EfficientNetB0

EfficientNetB0 is a Convolutional Neural Network (CNN) architecture specifically designed to achieve an optimal balance between accuracy and computational efficiency. This model was designed as a lightweight baseline network using Neural Architecture Search (NAS), utilizing the compound scaling approach to balance depth, width, and resolution. It features MBConv (Mobile Inverted Bottleneck Convolutional) blocks and integrates the Squeeze and Excitation (SE) mechanism [24]. In the development of CNN, models are generally scaled up using three aspects such as depth, width, and resolution to improve accuracy. However, increasing only one of these aspects often makes the performance of the models inefficient, so EfficientNet was introduced to address this issue. EfficientNet employs the Compound Scaling approach by simultaneously increasing the depth, width, and

resolution in a balanced and systematic manner, aiming to maximize the accuracy while maintaining the computational efficiency [25]. Compound Scaling addressed this issue using an exponential scaling method controlled by a single scaling coefficient ϕ formulated as follows equation 1, 2, 3, 4, 5:

$$\text{depth:} \quad d = \alpha^\phi \quad (1)$$

$$\text{width:} \quad w = \beta^\phi \quad (2)$$

$$\text{resolution:} \quad r = \gamma^\phi \quad (3)$$

$$\text{s. t. } \alpha \cdot \beta \cdot \gamma \approx 2 \quad (4)$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1 \quad (5)$$

EfficientNetB0 introduced as the baseline model ($\phi = 0$), with later variants such as EfficientNetB1 ($\phi = 1$), EfficientNetB2 ($\phi = 2$), and up to EfficientNetB7, were developed by progressively increasing the scaling coefficient ϕ , as shown in Figure 1.

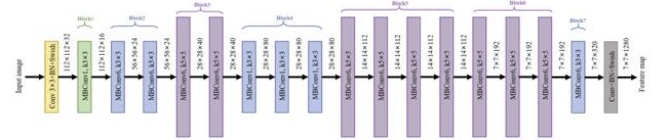


Figure 1: EfficientNetB0 Architecture [24]

The EfficientNetB0 architecture begins with a convolutional layer, which receives input in the form of an RGB image (3 channels) and applies an initial convolution to extract basic features such as edges, textures, and color patterns. After the initial convolution process, the image is further processed through the core efficiency component of the EfficientNetB0 architecture, namely MBConv (Mobile Inverted Bottleneck Convolution), as illustrated in Figure 2. MBConv consists of several main stages, starting with a 1×1 convolution that functions to expand the features from the input in order to recognize more patterns or information within the data. This process is followed by normalization through Batch Normalization and Swish activation, which is a non-monotonic activation function formulated as:

$$f(x) = x \cdot \sigma(x) = \frac{x}{1 + e^{-x}} \quad (6)$$

After the features are expanded, the data is processed using depthwise convolution, a type of convolution that applies a single filter to each image channel, making the process lighter and faster compared to standard convolution. The result is then normalized and passed through the Swish activation function to add non-linearity.

4. MobileNetV3

MobileNetV3 is a convolutional neural network model designed to give better performance even when used on limited devices like smartphones or IoT devices. The architecture combines and improves key ideas from both MobileNetV1, which uses depthwise separable convolution, and MobileNetV2, which uses inverted residual blocks and a linear bottleneck [24]. MobileNetV3 also adds a few new components, such as Squeeze-and-Excitation (SE) modules and the hard-swish activation function, which allow a

significant reduction in the number of parameters and computation without sacrificing feature representation quality [25].



Figure 2: The Structure of the MBConv block [25]

Our study uses two variants of the MobileNetV3 architecture, MobileNetV3-Large and MobileNetV3-Small. Each of the variants is optimized for a different goal, the large variant is optimized for higher accuracy, while the small variant prioritizes efficiency for resource constrained environments [26]. The architecture of MobileNetV3-Large consists of several main parts, including the first block (initial layers), the bottleneck blocks (core of the architecture), and the final block (classifier head), as illustrated in Figure 3.

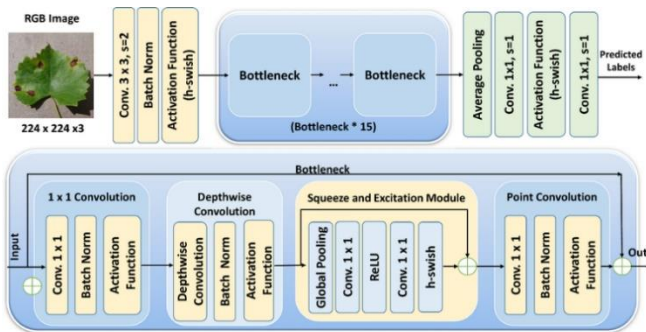


Figure 3: MobileNet-V3 Large Architecture [27]

In the initial block stage, the model receives an input in the form of an RGB image sized $224 \times 224 \times 3$. This image is then processed using an initial 3×3 convolution with a stride of 2, followed by Batch Normalization and the h-swish activation function. The purpose of this stage is to reduce spatial resolution while increasing the number of channels for feature extraction. After the initial stage, the architecture continues with 15 bottleneck blocks. Each bottleneck block consists of three main stages: Expand (1×1 convolution) to increase the number of input channels, Depthwise Convolution (3×3) to perform convolution independently on each channel for computational efficiency, and Project (1×1). This block is called an inverted residual because the channel expansion is done at the beginning (expand). If the input and output dimensions are the same, a shortcut connection is applied, which is the addition of the original input with the final output of the block:

$$output = F(x) + x \tag{7}$$

Where $F(x)$ is the transformation function of the bottleneck block, and x is the original input to the block. This approach allows the model to carry identity information unchanged when the transformation of $F(x)$ does not provide significant contribution. This strategy helps prevent accuracy degradation in very deep networks, as the model is not forced to learn everything from scratch but only the necessary changes from the existing information. To maintain the efficiency, MobileNetV3-Large uses a linear bottleneck, which is the final stage of the bottleneck block (1×1

convolution) that does not apply an activation function. This helps to preserve important information from being lost due to the effects of non-linear activation functions, especially when the feature data has already been reduced in size.

Some blocks in this architecture are equipped with a Squeeze and Excitation (SE) module, such as an attention mechanism that adaptively adjusts the importance of each feature channel. This module will help the network to focus on the most relevant information by enhancing important features and suppressing the less useful ones. The main activation function used in MobileNetV3-Large is hard-swish (h-swish), a lightweight version of the swish function designed for computational efficiency. This function is defined as:

$$h-swish(x) = x \cdot \frac{ReLU6(x+3)}{6} \tag{8}$$

$$ReLU6(x) = \min(\max(0, x), 6) \tag{9}$$

This function is continuous and has gradients across almost on its entire domain, making the training process more stable and efficient compared to the standard ReLU, while still being more computationally efficient than the original swish function. After passing through all the bottleneck blocks, the extracted features are summarized using global average pooling into a smaller dimension, then processed by a 1×1 convolutional layer with h-swish activation, and finally by a classification layer to produce prediction scores for each class.

Compared to MobileNetV3-Small, the MobileNetV3-Large version has a higher number of parameters and computational complexity, with approximately 5.4 million parameters and 219 million FLOPs. This allows it to achieve better performance in recognizing more complex visual patterns [25].

5. Dataset

The dataset used in this study is derived from two sources which contain 13,024 data of eye images. The first source is from Kaggle, the Gender Classification Based on Eye Image Dataset, which contains 5,202 images of female eyes and 6,323 images of male eyes. In addition, the second source was collected manually, consisting of 1,499 eye images by cropping from our personal documentation with every individual's permission. This additional data was carried out to increase the amount of data training process and to enhance the diversity of eye features, especially for the Asian. Label 0 was given to female and 1 to male.

Each person has their own unique and relatively consistent eye pattern. In facial recognition, the eye region such as the eyebrows, eyes, and eyelids become a key focus point for detecting a person's gender, especially in situations where full facial recognition is difficult. By using eye patterns, accurate and rapid gender classification can be achieved even under challenging conditions [5]. Each eye image includes several areas from the top of the eyebrows to the bottom of the eye bags. This variation helps the model learn patterns related to gender characteristics based on eye images.

6. Experimental Settings

All the processes were performed on Google Collab using python 3.12.11, PyTorch 2.8.0+cu126 and TensorFlow 2.19.0. The hardware environment was utilized by Tesla T4 GPU with 15 GB of memory and has a total of 13.6 GB of RAM. For all the models tested in this study, we used pretrained models. We configured the model's hyperparameters, including the number of epochs, learning rate, batch size, and type of optimizer, as shown in Table 1. For comparison purposes, we used the best model weights obtained at the epoch with the highest validation accuracy to maximize each model's potential.

Table 1: Experimental Settings

Hyperparameter	Nilai
Epoch	50
Optimizer	Adam
Learning Rate	0.001
Batch Size	16

6.1 Data Preprocessing

At this stage, several initial steps were carried out to prepare the images before being used by the deep learning model. The goal is to ensure that the data used has a consistent format and quality so that the model can learn more effectively. The first step we did was to load the image dataset, then labeled it, took five sample images from each class of female and male, and display them in a grid format. This was done to ensure that the data to be used is appropriate and visually diverse. The results of the data preprocessing are shown in the following Figure 4.

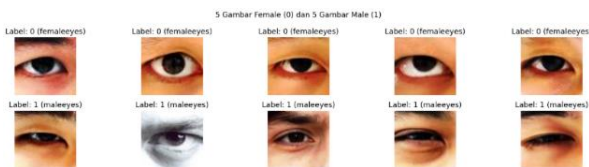


Figure 4: Data Preprocessing

6.2 Data Transformation

The next step is to perform image transformation to ensure a uniform size before being used in the training process. Each image was resized to 224×224 pixels to match the requirements of the architecture. This transformation was carried out by resizing the images to standardize their dimensions.

6.3 Data Splitting

After the data was successfully processed, we divided the dataset into three main parts, 70% for training data, 20% for validation data, and 10% for testing data, as summarized in Table 2. This step was done randomly, but still used seeds to make the result stay consistent and continue to be reproduced. The main goal is to make sure that the model does not learn only from the training data, but also to be tested on new data that had never been seen or studied before.

Table 2: Results of Data Splitting

Data Type	Total Samples
Training	9116

Data Type	Total Samples
Validation	2604
Test	1304
Total Data	13024

6.4 Data Augmentation

At this step, we applied data augmentation to enhance the diversity of eye images without adding manual data. This process was carried out to enable the model to recognize patterns from various image variations that may occur in real-world scenarios. Several augmentation techniques used in this process include random resized crop, horizontal flip, rotation and light translation, brightness and contrast adjustment, as well as Gaussian blur. These techniques allow the model to learn from a broader range of variations. Each technique was applied randomly to the eye images, which helps improve the model's performance on unseen data. The Figure 5 illustrates the results of the data augmentation process.



Figure 5: Data Augmentation

7. Results and Discussions

In this section, we will show the results of the training and testing processes for four models, namely MobileNetV3 Small, MobileNetV3-Large, EfficientNetB0, and EfficientNetB1. Each model was evaluated to compare the accuracy and loss values on both validation and test data, as well as to analyze whether the model experienced overfitting or underfitting during the training process.

7.1 Training Model

In this process, we pay attention to two important key indicators namely, highest validation accuracy and final accuracy. Highest validation accuracy refers to the highest accuracy value that the model has ever reached on the validation data during training. Meanwhile, final accuracy is the accuracy value of the last epoch that the model achieves after all the training is completed.

Based on Figure 6, MobileNetV3 Small achieved the highest accuracy on the validation data at 96.66%, with a final accuracy of 95.89% and the training data accuracy reached 98.61%. Meanwhile, the loss value on the validation data decreased to 0.1954 and the training data loss was recorded at 0.0373.

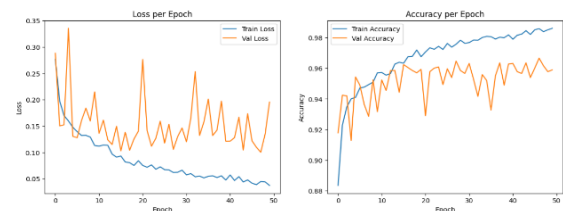


Figure 6: Training MobileNetV3 Small Model

As shown in Figure 7, MobileNetV3 Large achieved the highest validation accuracy at 96.70%, with a final accuracy of 96.12% and the training accuracy reached 98.39%.

Meanwhile, the validation loss decreased to 0.1404 and the training loss was recorded at 0.0426. Comparing the accuracy and loss on both training and validation data, the model does not show signs of overfitting.

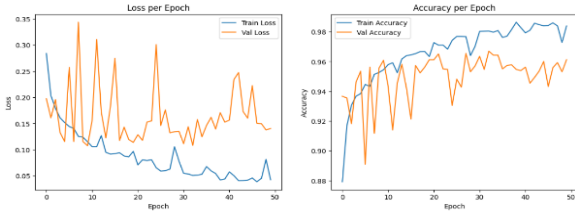


Figure 7: Training MobileNet Large

Based on Figure 8, EfficientNetB0 achieved the highest validation accuracy at 97.04%, with a final accuracy of 96.70% and the training accuracy reached 98.88%. Meanwhile, the validation loss decreased to 0.1185 and the training loss was recorded at 0.0323.

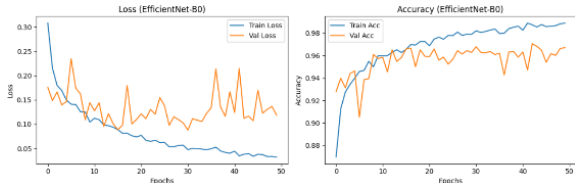


Figure 8: Training EfficientNet-B0

Based on Figure 9, EfficientNetB1 achieved the highest validation accuracy at 97.16%, with a final accuracy of 96.47% and the training accuracy reached 99.00%. Meanwhile, the validation loss decreased to 0.1123 and the training loss was recorded at 0.0287.

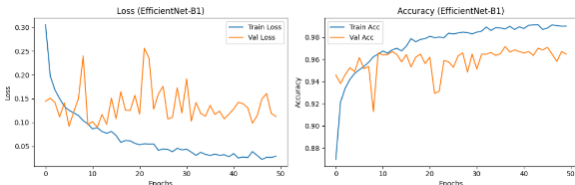


Figure 9: Training EfficientNet-B1

Here are the results of each model training process.

Table 3: Training Models Results

Model	Val Acc	Accuracy	Train Acc	Parameter
MobileNetV3 Small	96,66%	95.78%	98,61%	2.9 M
MobileNetV3 Large	96,70%	96.01%	98,39%	5.4 M
EfficientNet B0	97,04%	96.55%	98,88%	5.3 M
EfficientNet B1	97,16%	96.17%	99,00%	7.8 M

Overall, all four models demonstrated stable performance without any indication of overfitting or underfitting, as seen from the differences in accuracy and loss between the training and validation data, as presented in Table 3. EfficientNetB1 recorded the highest validation accuracy at 97.16%, outperforming EfficientNetB0 which reached 97.04%. Although the highest validation accuracy of MobileNetV3 Small and Large was a bit lower, but both models still showed good and consistent training patterns. These results indicate that the EfficientNetB1 architecture delivered the best

performance in gender classification based on eye images in this study.

7.2 Testing Model

Based from the table and the confusion matrix below, MobileNetV3-Small model show very good performance on the test data with an accuracy of 95.78%. Based on the confusion matrix in Figure 10, the model correctly identified 531 images labeled as female, while 47 female images were misclassified as male. For the male class, the model correctly recognized 718 images, with only 8 misclassifications. As shown in Table 4, the precision for the female class was 0.99 and the recall was 0.92, resulting in an F1-score of 0.95. For the male class, the precision was 0.94 and the recall reached 0.99, yielding an F1-score of 0.96. These results indicate that the model performs well and is relatively balanced in recognizing male and female images, although its performance on the male class is a bit superior based on the F1-score.

Table 4: Classification Report MobileNetV3-Small

	Precision	Recall	f1-score	support
Female	0.99	0.92	0.95	578
Male	0.94	0.99	0.96	726
accuracy			0.96	1304
macro avg	0.96	0.95	0.96	1304
weighted avg	0.96	0.96	0.96	1304

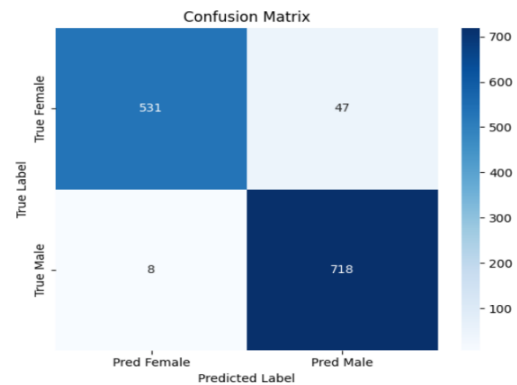


Figure 10: Confusion Matrix MobileNetV3-Small

Meanwhile, MobileNetV3-Large model demonstrated very good performance on the test data with an accuracy of 96.01%. Based on the confusion matrix in Figure 11, the model correctly identified 552 images labeled as female, while 26 female images were misclassified as male. For the male class, the model correctly recognized 700 images, with 26 misclassifications. As shown in Table 5, the precision for the female class was 0.96 and the recall was 0.96, resulting in an F1-score of 0.96. Similarly, the male class also had a precision and recall of 0.96, with an F1-score of 0.96. These results indicate that the model has an ability to give balanced classification to classify between male and female images.

Table 5: Classification Report MobileNetV3-Large

	Precision	Recall	f1-score	support
Female	0.96	0.96	0.96	578

	Precision	Recall	f1-score	support
Male	0.96	0.96	0.96	726
accuracy			0.96	1304
macro avg	0.96	0.96	0.96	1304
weighted avg	0.96	0.96	0.96	1304

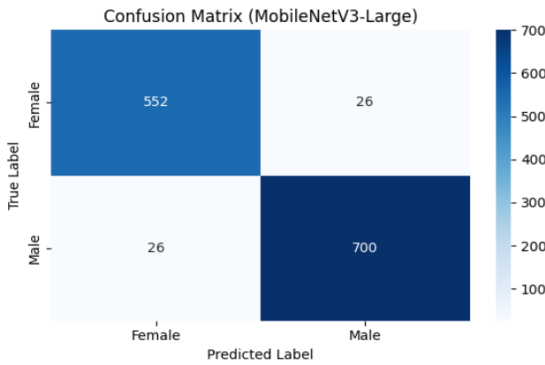


Figure 11: Confusion Matrix MobileNetV3-Large

The EfficientNetB0 model showed very good performance on the test data with an accuracy of 96.55%. Based on the confusion matrix in Figure 12, the model correctly identified 552 images labeled as female, while 26 female images were misclassified as male. For the male class, the model correctly recognized 707 images, with 19 misclassifications. As shown in Table 6, the precision for the female class was 0.97 and the recall reached 0.96, resulting in an F1-score of 0.96. Meanwhile, the male class had a precision of 0.96 and a recall of 0.97, with an F1-score of 0.97.

Table 6: Classification Report EfficientNetB0

	Precision	Recall	f1-score	support
Female	0.97	0.96	0.96	578
Male	0.96	0.97	0.97	726
accuracy			0.97	1304
macro avg	0.97	0.96	0.96	1304
weighted avg	0.97	0.97	0.97	1304

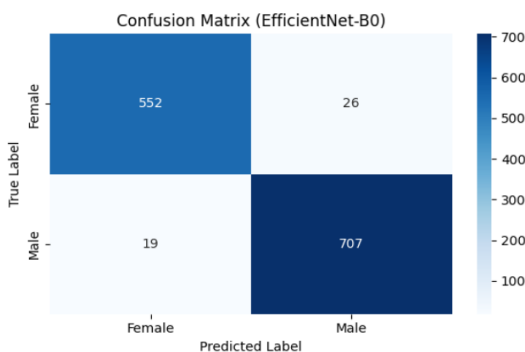


Figure 12: Confusion Matrix EfficientNetB0

Meanwhile, the EfficientNetB1 model showed very good performance on the test data, achieving an accuracy of 96.17%. Based on the confusion matrix in Figure 13, the model correctly classified 552 images labeled as female,

while 26 female images were incorrectly classified as male. For the male class, the model correctly identified 702 images, with 24 misclassifications. As shown in Table 7, the precision for the female class was 0.96, and the recall reached 0.96, resulting in an F1-score of 0.96. Meanwhile, the male class achieved a precision of 0.96 and a recall of 0.97, with an F1-score of 0.97.

Table 7: Classification Report EfficientNetB1

	Precision	Recall	f1-score	support
Female	0.96	0.96	0.96	578
Male	0.96	0.97	0.97	726
accuracy			0.96	1304
macro avg	0.96	0.96	0.96	1304
weighted avg	0.96	0.97	0.96	1304

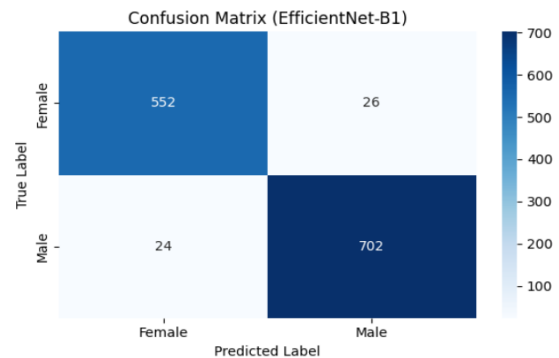


Figure 13: Confusion Matrix EfficientNetB1

Here are the results of each model testing process, including the score of the accuracy, precision, recall, and f1-score.

Table 8: Classification Report.

Model	Accu racy	Precision(Female)	Reca ll (Fem ale)	f1- score (Fem ale)	Preci sion (Mal e)	Rec all (M ale)	f1- sco re (M ale)
Mobile NetV3 Small	0.9578	0.99	0.92	0.95	0.94	0.99	0.96
Mobile NetV3 Large	0.9601	0.96	0.96	0.96	0.96	0.96	0.96
Efficien tNet B0	0.9655	0.97	0.96	0.96	0.96	0.97	0.96
Efficien tNet B1	0.9617	0.96	0.96	0.96	0.96	0.97	0.96

As we can see from the table above from the evaluation results of the four tested models presented in Table 8, EfficientNetB0 demonstrated the best performance with the highest accuracy of 96.55%. The highest F1-score for the male class, 0.97, was achieved by both EfficientNet B0 and B1. Meanwhile, the highest F1-score for the female class was 0.96, achieved by MobileNetV3-Large, EfficientNetB0, and EfficientNetB1. These results indicate that the models are more accurate in recognizing male eye images and confirm that the EfficientNetB0 architecture has an advantage in

gender classification based on eye images compared to MobileNetV3.

8. Conclusions

This study compared four lightweight models which are MobileNetV3-Small, MobileNetV3-Large, EfficientNetB0, and EfficientNetB1. After all models were evaluated for gender eye image classification, EfficientNetB0 was determined to be the most effective architecture achieving the highest testing accuracy of 96.55% and demonstrating the most balanced performance in recognizing male and female eye images, as shown by the F1-score of 0.97 for male and 0.96 for female. These results clearly show the advantage of EfficientNetB0 over MobileNetV3 and EfficientNetB1.

For future research, it is hoped that this study can be further developed by improving the applied methodology and focusing on practical applications in real time. The model also needs to be validated under more challenging conditions beyond data augmentation, such as dealing with partial occlusion or variations in real-world camera quality. Furthermore, exploring modern architectures like EfficientNetV2 or Vision Transformer with a multimodal approach that combines eye image analysis with other facial features such as shape of the nose, mouth, or facial texture can create a more reliable and accurate identification system. This technology could be applied in cybersecurity, for instance through automated monitoring systems in public areas such as malls or stations, to enhance security and enable more effective crowd management.

References

- [1] A. Arifandi. "Identifikasi Dan Prediksi Umur Serta Jenis Kelamin Berdasarkan Citra Wajah Menggunakan Algoritma Convolutional Neural Network (CNN)", *Jurnal Sains dan Teknologi-Universitas PGRI Kanjuruhan Malang*, vol. 4, no. 2, p. 2022, doi: [10.21067/jtst.v4i2.6985](https://doi.org/10.21067/jtst.v4i2.6985).
- [2] S. Hosgurmath, V. V. Mallappa, N. B. Patil, and V. Petli, "A face recognition system using convolutional feature extraction with linear collaborative discriminant regression classification," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 2, p. 1468, 2022, doi: [10.11591/ijece.v12i2.pp1468-1476](https://doi.org/10.11591/ijece.v12i2.pp1468-1476).
- [3] Mozhddeh Rouhsedaghat, Yifan Wang, Xiou Ge, Shuowen Hu, Suya You, and C.-C. Jay Kuo, "FaceHop: A Light-Weight Low-Resolution Face Gender Classification Method" 2020, doi: [10.1007/978-3-030-68793-9_12](https://doi.org/10.1007/978-3-030-68793-9_12).
- [4] Qingqiao Hua, Siyang Yina, Huiyang Ni, Yisiyuan Huang, "An End to End Deep Neural Network for Iris Recognition", 2020, doi: [10.1016/j.procs.2020.06.118](https://doi.org/10.1016/j.procs.2020.06.118).
- [5] T. Cut Al-Saidina Zulkhaidi, E. Maria, P. "Pengenalan Pola Bentuk Wajah dengan OpenCV," *Studi Teknologi Rekayasa Perangkat Lunak*, and P. Pertanian Negeri Samarinda (JURTI), vol. 3, no. 2, 2019, doi: [10.30872/jurti.v3i2.4033](https://doi.org/10.30872/jurti.v3i2.4033).
- [6] K. Mujib, A. Hidayatno, and D. T. Prakoso, "Pengenalan Wajah Menggunakan Local Binary Pattern (LBP) Dan Support Vector Machine (SVM)", 2018.
- [7] A. Wulan, "Deep Learning Untuk Deteksi Wajah Yang Berhijab Menggunakan Algoritma Cnn Dengan Tensorflow", 2020.
- [8] M. Naurah, "Pengenalan Wajah Dengan Latar Belakang Bebas Menggunakan Template Matching Dan Local Binary Pattern (LBP)", 2019.
- [9] Anggiratih, Endang, et al. "Klasifikasi Penyakit Tanaman Padi Menggunakan Model Deep Learning Efficientnet B3 dengan Transfer Learning." *Jurnal Ilmiah SINUS* 19.1(2021): 75-83, doi : [10.30646/sinus.v19i1.526](https://doi.org/10.30646/sinus.v19i1.526).
- [10] <https://keras.io/2/api/>.
- [11] S. W. P. Listio, "Performance of deep learning Inception model and MobileNet model on gender prediction through eye image," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 7, no. 4, pp. 11887, Oct. 2022, doi: [10.33395/sinkron.v7i4.11887](https://doi.org/10.33395/sinkron.v7i4.11887), doi: [0000-0001-8199-3312](https://doi.org/0000-0001-8199-3312).
- [12] H. Pu and K. Yi, "A Comparative Analysis of EfficientNet and MobileNet Models' Performance on Limited Datasets: An Example of American Sign Language Alphabet Detection," *Highlights in Science, Engineering and Technology*, vol. 94, 2024, CMLAI 2024.
- [13] Fernando Alonso-Fernandez, Kevin Hernandez Diaz, Silvia Ramis, Francisco J. Perales, Josef Bigun, "Facial Masks and Soft-Biometrics: Leveraging Face Recognition CNNs for Age and Gender Prediction on Mobile Ocular image", 2021, doi: [10.1049/bme2.12046Digital](https://doi.org/10.1049/bme2.12046Digital).
- [14] H. T. Nguyen, L. T. T. Pham, D. T. Dang, S. N. Huynh, P. H. Dang, dan Q. T. H. Nguyen, "Age Prediction from Facial Images Using Deep Learning Architecture," *Applied Computer Systems*, vol. 29, no. 2, pp. 22–29, Dec. 2024.
- [15] R. V. Muddaluru, S. R. Thoguluva, S. Prabha, T. K. Reddy, and S. Palaniswamy, "A Comparative Study of Filters and Deep Learning Models to Predict Diabetic Retinopathy," *arXiv preprint arXiv:2309.15216*, ver. 3, Jan. 9, 2024, doi: [10.48550/arXiv.2309.15216](https://doi.org/10.48550/arXiv.2309.15216).
- [16] Rahmad Firdaus, Joni Satria, Baidarus, "Klasifikasi Jenis Kelamin Berdasarkan Gambar Mata dengan Menggunakan Algoritma Convolutional Neural Network (CNN), 2022, doi: [10.37859/coscitech.v3i3.4360](https://doi.org/10.37859/coscitech.v3i3.4360).
- [17] Gilang Syah Amadha K, "Klasifikasi Jenis Kelamin Berdasarkan Citra Mata dengan Menggunakan Algoritma K-Nearest Neighbor, 2023.
- [18] Priska Choirina, Rosa Andrie Asmara, "Deteksi Jenis Kelamin Berdasarkan Citra Wajah Jarak Jauh dengan Metode Haar Cascade Classifier", 2016, doi: [10.33795/JIP.V2I4.77](https://doi.org/10.33795/JIP.V2I4.77).
- [19] Rosa Andrie Asmara, Bella Sita Andjani, Ulla Delfana Rosiani, Priska Choirina, "Klasifikasi Jenis Kelamin pada Citra Wajah Menggunakan Metode Naive Bayes", 2018, doi: [10.33795/jip.v4i3.209](https://doi.org/10.33795/jip.v4i3.209).

- [20] Ulla Delfana Rosiani, Rosa Andrie Asmara, Nadhifatul Laeily, "Penerapan Facial Landmark Point untuk Kasifikasi Jenis Kelamin Berdasarkan Citra Wajah", 2019, doi: [10.33795/jip.v6i1.328](https://doi.org/10.33795/jip.v6i1.328).
- [21] Khairul Sani, Inung Wijayanto S.T.,M.T., Eko usatio S.T.,M.T. "Implementasi Aplikasi Pengenalan Jenis Kelamin Berdasarkan Citra Wajah dengan Metode Support Vector Machine secara Real Time, 2016.
- [22] B. A. Kumar and M. Bansal, "Face mask detection on photo and real-time video images using Caffe-MobileNetV2 transfer learning," *Applied Sciences*, vol. 13, no. 2, p. 935, 2023, doi: [10.3390/app13020935](https://doi.org/10.3390/app13020935).
- [23] M. R. D. Ulhaq, D. Firdaus, and M. A. Zaidan, "Pengenalan ekspresi wajah secara real-time menggunakan metode SSD MobileNet berbasis Android," *Journal of Technology and Informatics (JoTI)*, vol. 5, no. 1, Oct. 2023, doi: [10.48550/arXiv.1905.11946](https://doi.org/10.48550/arXiv.1905.11946).
- [24] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, version 5, Sep. 2020, doi: [10.48550/arXiv.1905.11946](https://doi.org/10.48550/arXiv.1905.11946).
- [25] Zhou, A., Ma, Y., Ji, W. et al. Multi-head attention-based two-stream EfficientNet for action recognition. *Multimedia Systems* 29, 487–498 (2023), doi: [10.1007/s00530-022-00961-3](https://doi.org/10.1007/s00530-022-00961-3).
- [26] T. Shahriar, "Comparative Analysis of Lightweight Deep Learning Models for Memory-Constrained Devices," *arXiv preprint arXiv:2505.03303*, 2025, doi: [10.48550/arXiv.2505.03303](https://doi.org/10.48550/arXiv.2505.03303).
- [27] Karim, Md & Goni, Md & Nahiduzzaman, Md & Ahsan, Mominul & Haider, Julfikar & Kowalski, Marcin, "Enhancing agriculture through real-time grape leaf disease classification via an edge device with a lightweight CNN architecture and Grad-CAM", 2024, doi: [10.1038/s41598-024-66989-9](https://doi.org/10.1038/s41598-024-66989-9).