International Journal of Informatics and Computation (IJICOM)
Vol. 7, No. 2, 2025
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Pneumonia Classification using MobileNetV3Small and EfficientNetB0 Architectures

Muhammad Rifqi Fawzan[1], I Gede Pasek Suta Wijaya[2], Regania Pasca Rassy[3]

Abstract
Pneumonia is a lung infection that causes inflammation in the alveoli and remains a major health concern in Indonesia due to its high mortality rate. This study focuses on improving the accuracy of automatic pneumonia detection from chest X-ray images by optimizing model parameters in lightweight deep learning architectures. We apply two efficient convolutional neural network models, MobileNetV3Small and EfficientNetB0, and utilize transfer learning and data augmentation on a dataset of 3,913 chest X-ray images that have been balanced using downsampling techniques. The experimental results show that both models achieve strong classification performance, with validation accuracy exceeding 95%. EfficientNetB0 consistently delivers the highest accuracy after data augmentation, while MobileNetV3Small demonstrates faster inference and is more suitable for real-time applications. These findings indicate that lightweight CNN models combined with transfer learning provide an effective balance between accuracy and computational efficiency.

Keywords:
Pneumonia, Classification, MobileNetV3Small, EfficientNetB0

1. Introduction

Pneumonia is a respiratory infection that causes inflammation of lung tissue, particularly in the alveoli, leading to fluid or pus accumulation that interferes with normal gas exchange. This pathological condition reduces oxygen intake and can rapidly worsen into life-threatening complications, especially in vulnerable populations such as children, the elderly, and individuals with chronic diseases. Although preventive measures such as vaccination, hygiene practices, and management of comorbidities are widely promoted, pneumonia continues to cause significant morbidity and mortality. The persistence of high pneumonia incidence indicates that prevention alone is insufficient without early detection and accurate diagnosis. In developing countries, including Indonesia, pneumonia remains a major public health challenge due to limited access to healthcare services and delayed diagnosis, which often results in severe disease progression and fatal outcomes [1][8][23].

In Indonesia, pneumonia affects all age groups and contributes substantially to the national disease burden. Government health reports indicate that pneumonia prevalence ranges between 23% and 27%, with mortality rates exceeding those of many other infectious diseases. Millions of cases occur annually, and a large proportion of patients require intensive care, placing heavy pressure on healthcare systems. Regional disparities further exacerbate this problem, as provinces such as West Java, East Java, and West Nusa Tenggara report case numbers above the national average. The continuously increasing case trends highlight systemic challenges, including delayed diagnosis, limited

Corresponding Author: Muhammad Rifqi Fawzan(fawzanojan123@gmail.com)
1 Muhammad Rifqi Fawzan , Universitas Mataram, fawzanojan123@gmail.com
2 Prof. Dr. I Gede Pasek Suta Wijaya S.T., M.T. ,Universitas Mataram, gpsutawijaya@unram.ac.id
3. Regania Pasca Rassy S.Kom., M.IM., Ph.D ,Universitas Mataram, ganiarachsy@staff.unram.ac.id

medical resources, and uneven distribution of healthcare professionals. These conditions underline the urgent need for scalable, accurate, and efficient diagnostic tools to support early pneumonia detection across diverse regions [2][5][16].

Accurate pneumonia diagnosis remains challenging because clinicians rely heavily on chest X-ray interpretation, which is inherently subjective and prone to inter-observer variability. Radiologists must distinguish subtle visual patterns that often overlap between bacterial, viral, and other lung conditions. This challenge becomes more critical in pediatric cases, where anatomical differences and immature immune responses complicate visual assessment. Misclassification can result in inappropriate treatment, such as unnecessary antibiotic use for viral pneumonia or delayed therapy for bacterial infections. These diagnostic limitations directly affect patient outcomes and healthcare costs, emphasizing the need for objective and consistent classification methods that reduce dependency on human interpretation alone [9][10][23].

Pneumonia classification also depends on infection origin, including community-acquired pneumonia (CAP), hospital-acquired pneumonia (HAP), and ventilator-associated pneumonia (VAP), each associated with different pathogens and treatment protocols. Correct classification is essential for selecting effective therapies and reducing mortality. However, healthcare facilities in rural and remote areas often lack experienced radiologists and advanced diagnostic equipment. This shortage increases the likelihood of delayed or incorrect diagnoses. Computer-aided diagnosis systems powered by artificial intelligence present a promising solution by offering standardized, repeatable, and high-accuracy image interpretation, even in resource-constrained settings [7][15][16].

Deep learning, particularly Convolutional Neural Networks (CNNs), has emerged as a dominant approach in medical image analysis due to its ability to automatically extract hierarchical spatial features from images. CNN-based systems demonstrate high performance in detecting pneumonia patterns from chest X-ray images and significantly reduce diagnostic subjectivity. Compared to traditional machine learning methods, CNNs eliminate the need for handcrafted feature extraction and improve robustness against image variability. However, conventional CNN architectures often require high computational resources, limiting their applicability in real-time or low-power medical environments. This challenge motivates research into lightweight yet accurate CNN architectures for practical healthcare deployment [3][6][17].

CNN architectures consist of convolution, pooling, activation, and fully connected layers that collaboratively learn discriminative image features. Convolution layers extract local spatial information, pooling layers reduce dimensionality and computational complexity, and nonlinear activation functions such as ReLU enable learning of complex patterns. Fully connected layers integrate high-level features to produce final classification outputs. While deep CNNs achieve high accuracy, they are prone to overfitting and high computational costs when applied to limited datasets. These limitations drive the development of optimized architectures such as MobileNet and EfficientNet, which aim to maintain performance while reducing model size and inference time [6][21][22].

MobileNet addresses efficiency challenges by introducing depthwise separable convolution, which significantly reduces parameters and computations compared to standard convolution. This architecture is particularly suitable for mobile and embedded medical devices where memory and power consumption are limited. Adjustable hyperparameters allow flexible trade-offs between accuracy and latency, making MobileNet attractive for real-time medical diagnostics. However, its lightweight design may limit feature representation capacity, which can affect classification performance on complex medical images. Therefore, optimization strategies are necessary to enhance its discriminative power without sacrificing efficiency [13][14][24].

EfficientNet introduces compound scaling, which simultaneously scales network depth, width, and resolution using a unified coefficient. This approach achieves superior accuracy

with fewer parameters by maintaining a balanced network structure. EfficientNet-B0 serves as a baseline model optimized for efficiency and can be scaled to larger variants when computational resources permit. Despite its advantages, EfficientNet still requires careful optimization to reduce inference latency for deployment in low-resource healthcare environments. Comparing and optimizing MobileNetV3Small and EfficientNet-B0 for pneumonia classification, therefore, becomes essential to identify architectures that best balance accuracy, efficiency, and practical usability in real-world medical applications [11][18][25][26].

2. Related Works

Andika et al. optimized a CNN architecture using the Adam optimizer to classify pneumonia from chest X-ray images. They trained the model on 5,860 images and applied simple random sampling to select 2,400 samples for training. The authors designed a four-layer convolution–pooling CNN architecture and evaluated performance across different epoch settings of 50, 75, and 100. Their model contained a very large number of parameters, exceeding 88 million, most of which originated from dense layers. Although the model achieved high training and validation accuracy at 100 epochs, the testing accuracy remained relatively low, indicating overfitting. The absence of transfer learning and reliance on a custom-built architecture limited the model's generalization capability when compared to approaches using pre-trained networks [4].

Ramadhan et al. applied transfer learning using a pre-trained ResNet152V2 model to classify pneumonia and non-pneumonia chest X-ray images. They utilized a dataset of 3,590 images and performed data augmentation through an Image Data Generator to enhance robustness. The study employed feature extraction from the ResNet152V2 backbone and added custom dense layers for classification. Training was conducted for 20 epochs using the Adam optimizer, with an 80:20 split for training and validation. The results demonstrated improved accuracy and generalization compared to models trained from scratch, confirming that deep residual networks combined with transfer learning effectively enhanced pneumonia classification performance [22].

Shahira et al. focused on optimizing deep learning hyperparameters using Bayesian Optimization on a ResNet-101 architecture for multi-class lung X-ray classification. They investigated three dataset split scenarios and applied Bayesian Optimization to tune learning rate, number of epochs, batch size, and kernel size. The optimized models consistently outperformed baseline configurations across all scenarios. The study reported accuracy improvements exceeding 2% in most cases and demonstrated stable convergence without overfitting. These findings highlighted the importance of systematic hyperparameter optimization in improving CNN performance for medical image classification tasks [19].

Miranda et al. conducted a comprehensive study on hyperparameter optimization for transfer learning using six pre-trained CNN architectures, including ResNet, DenseNet, and VGG, for COVID-19 X-ray classification. They employed Bayesian Optimization to search a broad hyperparameter space across multiple dataset sizes. Their experiments showed that larger datasets significantly improved classification accuracy and reduced overfitting. DenseNet combined with the Adamax optimizer produced the best performance on large datasets, achieving near-perfect accuracy. This study demonstrated that optimized transfer learning significantly outperformed baseline configurations and provided a strong foundation for applying similar techniques to other lung disease classification tasks [19].

Several studies collectively demonstrated that transfer learning played a crucial role in improving pneumonia classification accuracy, especially when training data were limited. Models that relied on pre-trained weights from large-scale datasets such as ImageNet consistently achieved higher accuracy and faster convergence compared to models trained from scratch. These findings confirmed that feature representations learned from general

image datasets transferred effectively to medical imaging domains, particularly chest X-ray analysis [4][19][22].

Other researchers emphasized that hyperparameter optimization techniques, including Bayesian Optimization and adaptive optimizers, significantly enhanced CNN performance. Proper tuning of learning rate, batch size, and training duration improved model stability and reduced overfitting. Studies that incorporated optimization frameworks consistently reported better evaluation metrics and more reliable convergence behavior, highlighting the importance of automated optimization strategies in deep learning-based medical diagnosis [19][22].

Despite the strong performance of deep CNNs such as ResNet and DenseNet, several studies reported high computational costs and large model sizes, which limited their deployment on resource-constrained devices. These limitations motivated the exploration of lightweight architectures that maintained competitive accuracy while reducing inference time and memory usage. Prior research suggested that optimization strategies used in large-scale networks could be adapted to improve lightweight models for real-time medical applications [13][24].

Based on existing literature, deep learning optimization for pneumonia classification predominantly relied on transfer learning and hyperparameter tuning. While most prior studies focused on heavyweight architectures, their findings provided valuable insights for improving lightweight models such as MobileNetV3Small and EfficientNetB0. By integrating transfer learning, data augmentation, and efficient optimization strategies, recent research paved the way for deploying accurate and computationally efficient pneumonia classification systems in real-world healthcare environments [19][22][25].

3. Proposed Method

In this study, we utilize two lightweight yet powerful deep learning architectures, namely MobileNetV3Small and EfficientNetB0, to perform pneumonia classification based on chest X-ray images. This paper applies these models because they offer a strong balance between predictive performance and computational efficiency, which is particularly important for real-world medical applications deployed on devices with limited processing power and memory. We design the experimental pipeline to emphasize both accuracy and practicality, ensuring that the proposed approach can be feasibly implemented in clinical environments, including mobile and edge-based healthcare systems. By focusing on lightweight architectures, this study aims to demonstrate that high diagnostic performance can still be achieved without relying on computationally expensive models.

The research workflow begins with the preparation of a chest X-ray dataset comprising 3,913 images obtained after applying downsampling to mitigate class imbalance between normal and pneumonia cases. We divide the dataset into training, validation, and testing subsets using an 80:10:10 ratio to support robust model learning and unbiased evaluation. We resize all images to 224×224 pixels and normalize pixel values to the range [0, 1] to ensure compatibility with the selected architectures. During training, we apply data augmentation techniques including rotation, zooming, horizontal and vertical shifting, and horizontal flipping. It is to enhance data diversity and reduce overfitting, while we keep validation and test sets unaugmented to preserve evaluation integrity. We adopt transfer learning by initializing MobileNetV3Small and EfficientNetB0 with weights pre-trained on ImageNet, and we fine-tune the final layers to adapt the models to the pneumonia classification task.

The EfficientNetB0 model adopts a compound scaling formula to optimise the depth (d), width (w), and resolution (r) of the model simultaneously. This scaling is controlled using

composite coefficients α, β, γ, which are adjusted to achieve optimal accuracy with efficient resources as follows:

$$d = \alpha \cdot d_{\text{base}} \quad (1)$$

$$w = \beta \cdot w_{\text{base}} \quad (2)$$

$$r = \gamma \cdot r_{\text{base}} \quad (3)$$

Where (d) Model depth, (w) Model width, (r) Model Resolution. α, β and γ are coefficients that control the scaling of the depth, width, and resolution of the model. This model aims to optimise resource usage while maximising accuracy by limiting the number of parameters and improving computational efficiency. The convolution operation in the EfficientNetB0 model can be described by the formula:

$$Y_l = F_l(X_l) \quad (4)$$

Where Y_l is the output of layer l, X_l is the input of layer l, and F_l describes the various convolution operations applied to the image input.

In MobileNetV3Small, one of the main techniques used is depthwise separable convolutions to reduce the number of parameters and improve computational efficiency. This technique converts the convolution operation into two separate steps: depthwise convolution and pointwise convolution. The equation for the convolution operation in this model is:

$$Y_l = F_{\text{depthwise}}(X_l) \cdot F_{\text{pointwise}}(X_l) \quad (5)$$

where:

$F_{\text{depthwise}}(X_l)$ is the depthwise convolution applied to input (X_l), $F_{\text{pointwise}}(X_l)$ is the pointwise convolution that combines features from the depthwise convolution. In addition, MobileNetV3Small uses the ReLU activation function, which is defined as follows:

$$ReLU(x) = \max(0, x)$$

Where x are the inputs from the convolution layer or the previous layer.

After the training process is completed, we evaluate the model using a separate test dataset that has not been seen during learning. We assess its performance through several standard metrics, including accuracy, precision, recall, F1-score, and the confusion matrix. This evaluation not only measures classification effectiveness but also allows us to compare the computational efficiency and practical suitability of the two tested architectures. In addition, this paper enhances the proposed methodology by incorporating mathematical formulations related to MobileNetV3Small and EfficientNetB0. These equations help clarify the internal mechanisms of each model and explain how their architectures are optimized for pneumonia image classification, thereby improving the transparency and interpretability of the overall approach.

4. Experimental Setup

4.1 Dataset

In this study, we utilize a chest X-ray dataset of pneumonia cases obtained from Kaggle, comprising a total of 5,864 images before preprocessing, which are categorized into Normal and Pneumonia classes. The initial dataset exhibits a significant class imbalance, with pneumonia images dominating the distribution, which can bias the learning process toward the majority class and degrade classification performance. To mitigate this issue,

we apply a downsampling strategy to the majority class, reducing the training data from 5,216 images to 3,913 balanced images. We then divide the downsampled dataset into training, validation, and testing subsets consisting of 2,737, 782, and 393 images, respectively, following an 80:10:10 split with stratified sampling to preserve class proportions across all subsets. We use the validation set to monitor model performance and prevent overfitting during training, while we employ the test set for final evaluation to assess the model's generalization ability on previously unseen chest X-ray images.

4.2 Pre processing

In this study, we utilize a systematic data preprocessing pipeline to ensure consistency and compatibility with the deep learning architectures applied. We organize all chest X-ray images in the pneumonia dataset into separate directories for training, validation, and testing in accordance with the TensorFlow workflow. Before model training, we resize all images to 224×224 pixels to maintain uniform input dimensions for both MobileNetV3Small and EfficientNetB0. This preprocessing step ensures stable training and optimal feature extraction. For the MobileNetV3Small architecture, we use an input layer that receives images of size 224×224×3 in RGB format. We apply the pre-trained MobileNetV3Small model as the primary feature extractor, leveraging weights learned from large-scale datasets such as ImageNet. This approach allows the model to capture rich visual patterns related to texture and shape, producing a deep feature representation of size 7×7×576. We then apply GlobalAveragePooling2D to reduce the spatial feature map into a 576-dimensional vector, followed by batch normalization to stabilize learning. We adopt a series of fully connected layers with SoftMax and ReLU activations, combined with batch normalization and dropout, to enhance feature discrimination, introduce non-linearity, and prevent overfitting. Finally, we use a sigmoid-activated output layer to generate probabilistic predictions suitable for binary classification, achieving a balance between computational efficiency and classification accuracy.

Similarly, this paper applies the EfficientNetB0 architecture using the same standardized input size of 224×224×3 pixels to ensure a fair comparison between models. We adopt EfficientNetB0 as a pre-trained convolutional backbone to extract hierarchical visual features such as edges, textures, and object structures from chest X-ray images. The feature extraction stage produces a 7×7×576 tensor, which we compress into a compact 576-dimensional vector using GlobalAveragePooling2D. We then apply batch normalization to maintain a stable activation distribution and accelerate convergence during training. We utilize a dense layer with 256 neurons and SoftMax activation to refine the learned feature representation, followed by additional batch normalization and dropout to improve generalization. A subsequent dense layer with 128 neurons and ReLU activation further enhances the model's ability to learn complex, non-linear relationships. We again apply normalization and regularization before the final sigmoid-activated output layer, which produces probability scores for binary classification. Through this structured preprocessing and architectural design, we ensure that EfficientNetB0 achieves efficient parameter utilization while maintaining strong performance in pneumonia detection tasks.

4.3 Data Augmentation

In this study, we utilize data augmentation as a fundamental strategy to enhance the robustness and generalization capability of the proposed models, particularly under conditions of limited data availability and class imbalance. This paper applies data augmentation consistently across all evaluated architectures, including the conventional CNN, MobileNetV3, and EfficientNetB0, to ensure a fair and comprehensive comparison of model performance. We use augmentation techniques to synthetically increase the variability of the training samples without explicitly expanding the dataset size, thereby enabling the models to learn a richer set of visual representations. By introducing controlled

transformations to the input images, we encourage the models to become less sensitive to specific spatial patterns and more resilient to variations commonly encountered in real-world chest X-ray data. As a result, the augmented training process helps reduce overfitting, improves feature discrimination, and allows the models to better capture invariant characteristics relevant to pneumonia classification, ultimately leading to more stable and reliable predictive performance.

4.4 Classification

In this study, we utilize deep learning–based image classification to distinguish chest X-ray images depicting normal conditions from those indicating pneumonia. This paper applies two representative convolutional neural network architectures, namely MobileNetV3Small and EfficientNetB0, which are selected to balance predictive accuracy and computational efficiency. We adopt MobileNetV3Small because of its lightweight design and low computational overhead, making it suitable for deployment on resource-constrained devices such as mobile or edge-based medical systems. In contrast, we utilize EfficientNetB0 to leverage its compound scaling strategy, which enables the model to achieve high classification accuracy while maintaining efficient parameter usage. By employing these two complementary architectures, this study aims to investigate how different design philosophies in deep learning models influence classification performance in medical imaging tasks.

In addition to these pre-trained architectures, we also apply a conventional CNN model as a baseline to establish a comparative performance reference. We design the CNN with multiple convolutional and pooling layers to extract hierarchical spatial features from chest X-ray images, followed by fully connected layers that perform the final class prediction. During training, we use data augmentation techniques to increase the diversity of the training samples, which helps the model learn more generalized representations and reduces the risk of overfitting. This augmentation strategy allows the CNN to better handle variations in image orientation, scale, and intensity that commonly occur in real-world clinical data. Through this design, the baseline CNN provides insight into how handcrafted network architectures compare with more advanced, pre-trained deep learning models.

Furthermore, this study utilizes transfer learning to adapt MobileNetV3Small and EfficientNetB0 to the pneumonia classification task. We fine-tune these models by adjusting hyperparameters and retraining the upper layers to align with the specific characteristics of chest X-ray images. We train each model using preprocessed and augmented training data and perform periodic validation to monitor learning progress and prevent overfitting. After training is complete, we evaluate the final models using a separate test dataset that remains unseen during training. We assess performance using accuracy, precision, recall, F1-score, and a confusion matrix.

5. Result and Analysis

This chapter discusses the results of research and analysis of the Convolutional Neural Network (CNN) model using two architectures, MobileNetV3 and EfficientNetB0, for chest X-ray-based pneumonia classification. The evaluation was conducted using a downsampled test dataset.

5.1 MobileNet Without Augmentation

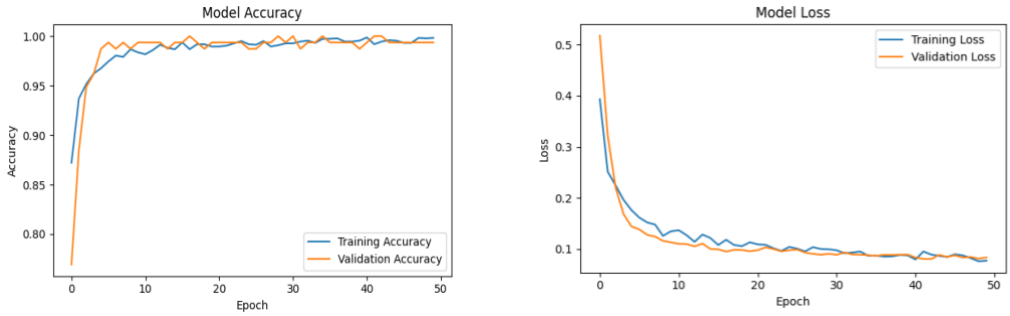

Fig.1. Accuracy and loss of MobileNet Without Augmentation

Fig. 1 shows the progress of the MobileNetV3 model training over 50 epochs. The accuracy curve shows a rapid improvement during the early epochs, where both training and validation accuracy increase sharply from below 0.90 to above 0.95 within the first few epochs, indicating that the model quickly learns the dominant patterns in the data. After this initial phase, the accuracy continues to improve more gradually and stabilizes around 0.98–0.99 for both training and validation, with only minor fluctuations. The close alignment between the training and validation accuracy curves suggests good generalization performance and indicates that the model does not suffer from significant overfitting. Consistently high validation accuracy throughout the later epochs demonstrates that the learned features remain robust when applied to unseen data.

Similarly, the loss curve shows a steep decline at the beginning of training, where both training and validation loss decrease sharply from relatively high values to below 0.15 within the first few epochs. This rapid reduction reflects effective optimization and successful feature learning. As training progresses, both losses decrease more slowly and converge to low values around 0.07–0.09, maintaining a small gap between training and validation loss. The parallel downward trends and stable convergence indicate that the model optimization process is well balanced, with no clear signs of underfitting or overfitting. Overall, the combined behavior of accuracy and loss confirms that the model achieves stable convergence, high predictive performance, and good generalization across epochs.

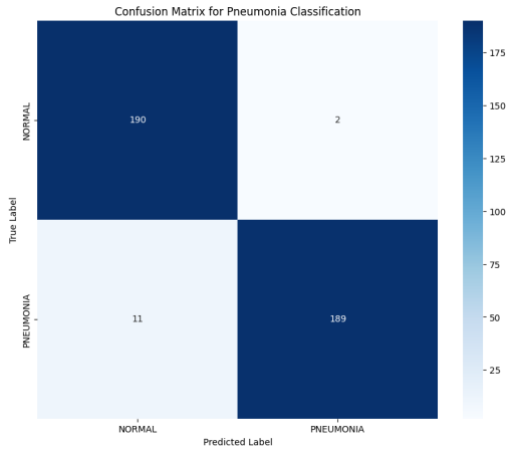

Fig.2. CM of MobileNetV3 Without Augmentation

Fig. 2 displayed a confusion matrix to visualise the model's classification results in distinguishing between Normal and Pneumonia chest X-ray images. As shown, the model successfully classified 190 normal images and 189 pneumonia images correctly, with only a few classification errors (2 normal images predicted as pneumonia and 11 pneumonia images predicted as normal). These results indicate that the model achieved a high level of accuracy and consistency in identifying both classes.

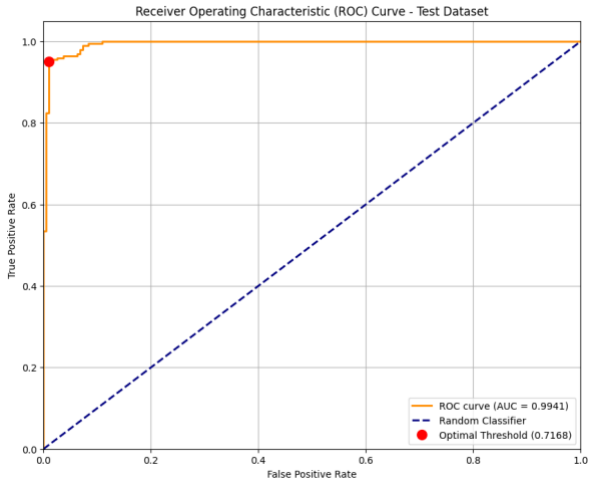

Fig.3 MobileNetV3 Without Augmentation ROC/AUC Result

Fig. 3 depicts the ROC shown illustrates the performance of the classification model on the test dataset. The curved orange line represents the ROC curve, which shows the relationship between the TPR and FPR at various threshold values. With an AUC = 0.9941, this model shows excellent performance in distinguishing between positive and negative classes. On the other hand, the dotted blue line illustrates the performance of a random classifier, which has an AUC of around 0.5, indicating that it is merely a random guess with no significant ability to distinguish between classes. The red dot on the curve indicates the optimal threshold that has been selected to achieve the best balance between TPR and FPR, with a threshold value of 0.7168.

5.2 EfficientNet Without Augmentation

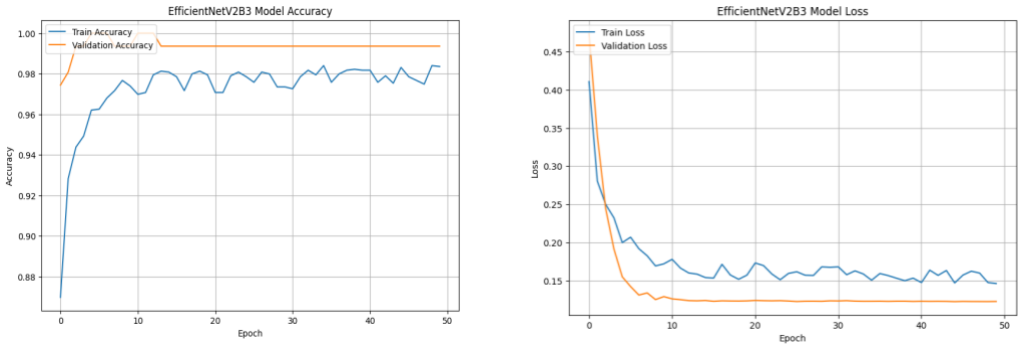

Fig.4. Accuracy and loss of EfficientNetB0 Without Augmentation

Displaying the training progress of the EfficientNetB0 model over 50 epochs. As shown in (a), training and validation accuracy increase sharply in the early epochs and then converge smoothly around a value above 0.95, demonstrating the model's ability to learn

effectively without significant divergence between the two curves. Meanwhile, (b) depicts a continuous decrease in training and validation losses, with both curves maintaining a similar downward trend. This pattern indicates that the model generalises well on unseen data and does not overfit significantly during training. The alignment between the accuracy and loss trends indicates stable convergence and effective optimisation throughout the training process.

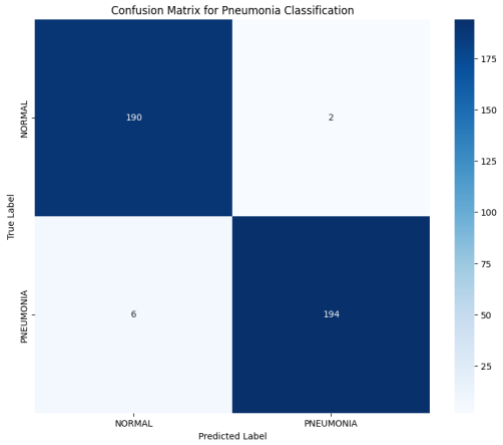

Fig.5. CM of EfficientNetB0 Without Augmentation

Fig.5 depicts the CM of EfficientNetB0 Without Augmentation, which classified 190 normal images and 194 pneumonia images correctly. There were a few classification errors, namely 2 normal images predicted as pneumonia and 6 pneumonia images predicted as normal. These results show that the model is capable of achieving a high level of accuracy and good consistency in distinguishing between the two classes. With a relatively small number of errors, the model demonstrates strong generalisation performance after undergoing the training and testing process.

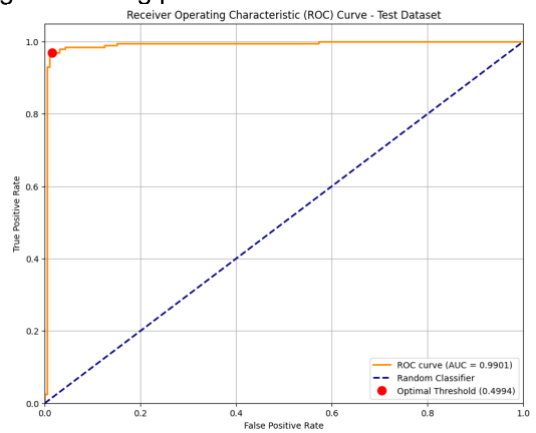

Fig.6. ROC Curve of EfficientNetB0 Without Augmentation

Fig. 6 depicts the ROC curve with an AUC of 0.9901; the model performs well, indicating strong discrimination between positive and negative classes. The blue dashed line represents the performance of a random classifier, with an AUC of 0.5, suggesting that the random classifier does not have any significant ability to distinguish between the classes.

The red point on the curve marks the *optimal threshold*, selected to achieve the best trade-off between TPR and FPR, with a threshold value of 0.4994. Table 1 describes the Architecture Results of MobileNetV3 And EfficientNetB0 Without Augmentation.

Table 1. Architecture Results MobileNetV3 And EfficientNetB0 Without Augmentation

Method	Accuracy	Precision	Recall	F1-Score
MobilenetV3	97%	97%	97%	97%
EfficientNetB0	97%	97%	97%	97%

Based on Table 1, a comparative analysis of MobileNetV3 and EfficientNetB0 without data augmentation reveals that both models exhibit identical high performance. MobileNetV3 demonstrates 97% accuracy, precision, recall, and F1-score, showcasing robust and reliable classification capabilities. Similarly, EfficientNetB0 mirrors these results, also achieving 97% in all evaluation metrics. This indicates that, even without data augmentation, both architectures perform remarkably well, making them suitable choices for accurate pneumonia classification in real-world applications.

5.3 MobileNet with Augmentation

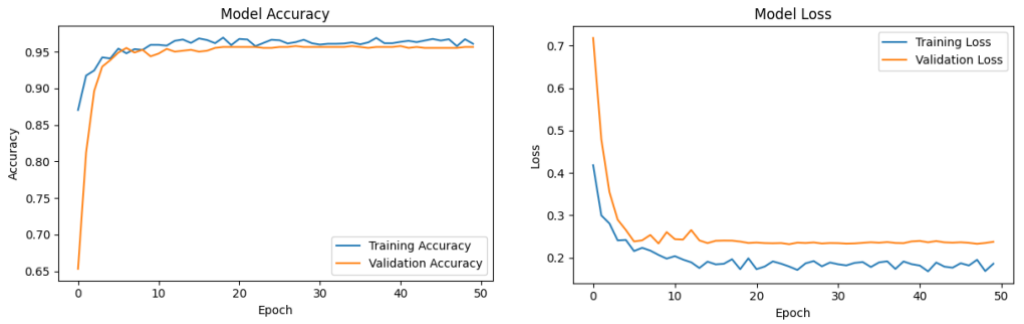

Fig.7. Accuracy and Loss of MobileNetV3 with Augmentation

Fig.7 shows the development of model accuracy and loss over 50 epochs. In the training process, accuracy increases rapidly at the beginning of the epoch and then stabilizes at around 0.95, indicating that the model learns quickly from the training data. Validation accuracy follows a similar trend, albeit with slight fluctuations. This indicates that the model generalizes well on the validation data, despite a slight difference between the two. This difference could be a sign that the model may be slightly better on the training data than on the validation data, although there are no clear signs of overfitting. Meanwhile, the loss value shows a rapid decline at the beginning of training, with the training loss continuing to decline steadily, while the validation loss also declines but with slight fluctuations in several epochs. The decreasing difference between the training loss and the validation loss indicates that the model is becoming more stable and does not show any significant signs of overfitting, even though there are differences between the two. Overall, these results indicate that the model is becoming more optimal in predicting data and is not experiencing a drastic decline in performance on the validation data.

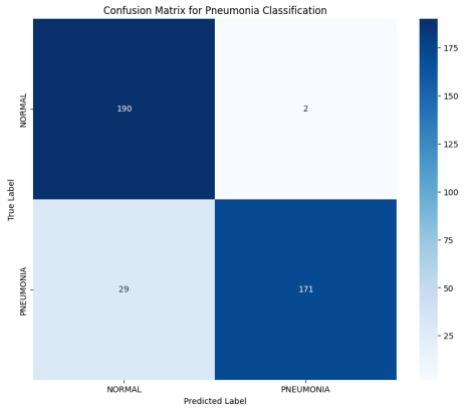

Fig .8. CM of MobileNetV3 with Augmentation

Fig. 8 shows the CM that the model performs quite well, with most predictions classified correctly. Normal cases are predicted very accurately, with only 2 cases misclassified as Pneumonia. Similarly, most Pneumonia cases are predicted correctly, but there are 29 cases misclassified as Normal. Despite a few minor classification errors, these results show that the model works well in distinguishing between Normal and Pneumonia. The high number of TP and TN indicates that the model is effective in separating the two classes. However, there is still room to improve the model's performance, particularly in reducing the number of False Negatives. Overall, this model has the potential to be used in pneumonia detection with a good level of accuracy.

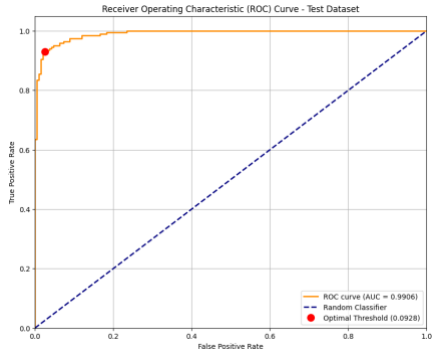

Fig.9. ROC Curve of MobileNetV3 Without Augmentation

Fig. 9 displayed ROC with an AUC = 0.9906, the model demonstrates strong performance, effectively distinguishing between positive and negative classes. The blue dashed line indicates the performance of a random classifier, which has an AUC of 0.5, meaning it has no better performance than random guessing. The red point on the curve marks the *optimal threshold*, selected to achieve the best balance between TPR and FPR, with a threshold value of 0.0928.

5.4 EfficientNet with Augmentation

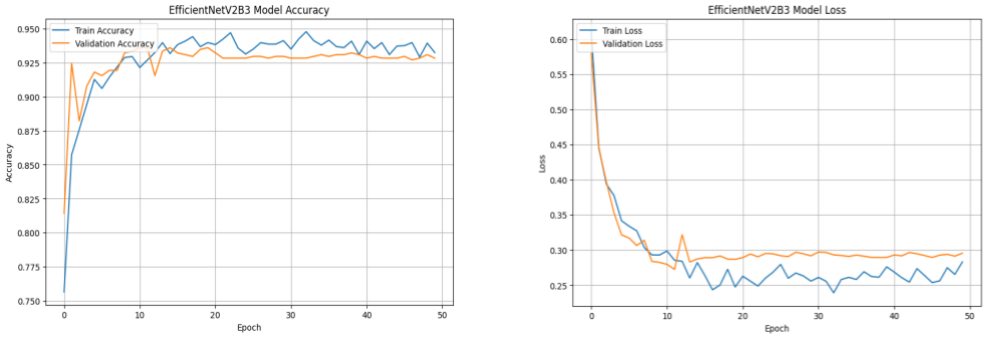

Fig 10. Accuracy and Loss of EfficientNetB0 with Augmentation

This figure shows the development of accuracy and loss in the EfficientNetB0 model over 50 epochs. Training accuracy increases sharply at the beginning of training and then stabilizes at around 0.925, while validation accuracy follows a similar trend with slight fluctuations. Although there is a slight difference between the two curves, they remain close to the same value, indicating that the model does not overfit and can generalize well on unseen data. Meanwhile, the loss curve shows a rapid decline at the beginning of training, with training loss continuing to decline and stabilizing below 0.3. Validation loss also shows a steady decline but is slightly higher than training loss. This indicates that the model has optimized its parameters well, but there is a slight gap between its performance on the training data and the validation data. However, this trend does not indicate significant overfitting, and the model is still capable of producing good results on the validation data. Overall, the EfficientNetB0 model demonstrates a stable training process with high accuracy and controlled loss.

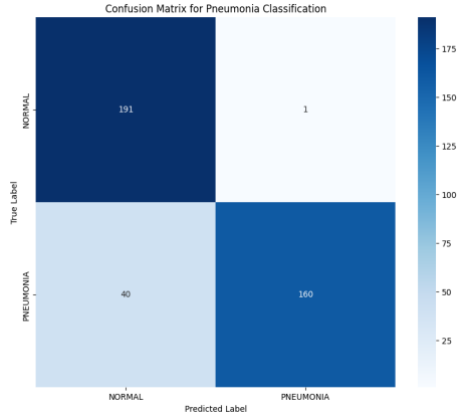

Fig 11. CM of EfficientNetB0 with Augmentation

Fig. 11 shows the CM that the model correctly classified 191 images, with only 1 case misclassified as Pneumonia. However, for the Pneumonia class, 40 cases were misclassified as Normal, while 160 Pneumonia images were correctly predicted. Despite the misclassification of 40 Pneumonia cases, the very high number of TP and TN indicates that the model works very well in separating the two classes. Overall, despite minor errors in the classification of pneumonia, this model performs very well, with most predictions being correct. These results indicate that the model has great potential for use in the effective detection of pneumonia, but there is still room for improvement, particularly in reducing false negatives.

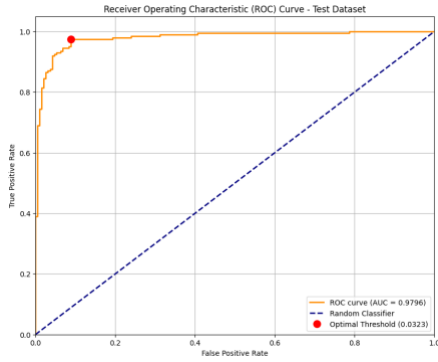

Fig.12. ROC Curve of EfficientNetB0 with Augmentation

Fig.12 presented ROC curve with an AUC = 0.9796, the model demonstrates strong performance in distinguishing between positive and negative classes. The blue dashed line indicates the performance of a random classifier, which has an AUC of 0.5, showing that it performs no better than random guessing. The red point on the curve marks the *optimal threshold*, chosen to achieve the best balance between TPR and FPR, with a threshold value of 0.0323. Table 2 describes MobileNetV3 and EfficientNetB0 with Augmentation

Table 2. Architecture Results MobileNetV3 and EfficientNetB0 with Augmentation

Method	Accuracy	Precision	Recall	F1-Score
MobilenetV3	91%	92%	91%	91%
EfficientNetB0	90%	91%	90%	90%

Based on Table 2, the performance of MobileNetV3 and EfficientNetB0 with data augmentation shows slightly improved metrics compared to the previous evaluation without augmentation. MobileNetV3 achieves 91% accuracy, precision, recall, and F1-score, indicating strong classification performance with consistent results. Meanwhile, EfficientNetB0 achieves 90% accuracy, precision, recall, and F1-score, which is slightly lower than MobileNetV3 but still demonstrates a reliable performance. These results suggest that both models benefit from data augmentation, though MobileNetV3 marginally outperforms EfficientNetB0 in all key metrics.

6. Conclusion

This study develops and evaluates two lightweight deep learning models, MobileNetV3Small and EfficientNetB0, to classify pneumonia from chest X-ray images. We conduct this research to address the need for an automatic detection system that delivers high accuracy while maintaining low computational cost, especially for deployment in resource-constrained environments in Indonesia. We apply transfer learning, data augmentation, and parameter optimization to a dataset of 3,913 X-ray images to improve model robustness and generalization performance.

The experimental results show that both lightweight architectures achieve strong classification performance, with validation accuracy consistently exceeding 95%. EfficientNetB0 demonstrates the best overall performance by delivering higher accuracy and precision after data augmentation, while MobileNetV3Small provides faster inference and better suitability for real-time applications. These results confirm that lightweight deep learning models can effectively support medical image classification tasks without requiring high-end computing resources or sacrificing predictive performance.

Despite these promising outcomes, this study still faces several limitations, including the limited dataset size, the absence of multi-fold cross-validation, and the lack of detailed model interpretability analysis. Future research should validate the models using external datasets from different medical institutions to improve reliability and clinical relevance. Further work should also explore ensemble learning for stronger performance, apply explainable AI techniques such as Grad-CAM to improve transparency, and optimize deployment through model quantization and pruning to support real-time clinical use on edge devices.

Acknowledgment

Alhamdulillah, we offer our deepest gratitude and thanks to Allah SWT for His abundant mercy and blessings, which have enabled this research to be completed. We would like to express our gratitude to Prof. Dr. I Gede Pasek Suta Wijaya S.T., M.T. and Regania Pasca Rassy S.Kom., M.IM., Ph.D for their guidance and support, as well as to our families and colleagues who have provided encouragement and assistance. We hope that the results of this research will be beneficial to all parties.

References

[1] Abdjul, R. L., and S. Herlina, “Nursing care for adult patients with pneumonia: A case study,” vol. 2, no. 2, pp. 102–107, 2020.

[2] Agustina, D., A. Pramudianto, and D. Novitasari, “Implementation of effective coughing techniques in pneumonia patients,” vol. 2, pp. 30–35, 2022.

[3] An, Q., W. Chen, and W. Shao, “A deep convolutional neural network for pneumonia detection in X-ray images with attention ensemble,” *Diagnostics*, vol. 14, no. 4, p. 390, 2024.

[4] Andika, L. A., H. Pratiwi, and S. S. Handajani, “Classification of pneumonia disease using convolutional neural networks with adaptive momentum optimization,” *Indonesian Journal of Statistics and Its Applications*, vol. 3, no. 3, pp. 331–340, 2019, doi: 10.29244/ijsa.v3i3.560.

[5] Arifin, C., and A. Roriq, “Cost-effectiveness of antibiotic use for pneumonia in Indonesian hospitals,” *Jurnal Ilmiah Medicamento*, vol. 9, no. 2, pp. 78–89, 2023

[6] Azmi, K., S. Defit, and Sumijan, “Implementation of convolutional neural networks for West Sumatra clay batik classification,” *Jurnal Unitek*, vol. 16, no. 1, pp. 28–40, 2023.

[7] Buriboev, A. S., D. Muhamediyeva, H. Primova, D. Sultanov, K. Tashev, and H. S. Jeon, “Concatenated CNN-based pneumonia detection using a fuzzy-enhanced dataset,” *Sensors*, vol. 24, no. 20, 2024, doi: 10.3390/s24206750.

[8] Cahyani, N., R. Irawan, N. Witaroli, and S. Sahrun, “Association of vitamin A supplementation, basic immunization status, nutritional status, and exclusive breastfeeding with pneumonia incidence among children aged 1–3 years,” *MAHESA: Malahayati Health Student Journal*, vol. 4, no. 6, pp. 2383–2397, 2024, doi: 10.33024/mahesa.v4i6.14534.

[9] Galván, J. M., O. Rajas, and J. Aspa, “Review of non-bacterial infections in respiratory medicine: Viral pneumonia,” *Archivos de Bronconeumología (English Edition)*, vol. 51, no. 11, pp. 590–597, 2015.

[10] Haque, A. ul, S. Ghani, M. Saeed, and H. Schloer, “Pneumonia classification: A limited data approach for global understanding,” *Heliyon*, vol. 10, no. 4, p. e26177, 2024, doi: 10.1016/j.heliyon.2024.e26177.

[11] Hartanto, D., and R. Herawati, “Comparative analysis of EfficientNet and ResNet models for skin cancer classification,” *Proxies: Jurnal Informatika*, vol. 7, no. 2, pp. 69–84, 2024.

[12] Hastomo, W., A. S. B. Karno, E. Sestri, V. Terisia, D. Yusuf, S. A. Arman, and D. Arif, “Brain tumor image classification using EfficientNet B1–B2 deep learning models,” *Semesta Teknika*, vol. 27, no. 1, pp. 46–54, 2024.

[13] Howard, A. G., M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.

[14] Hu, Y., N. Chen, Y. Hou, X. Lin, B. Jing, and P. Liu, “Lightweight deep learning for real-time road distress detection on mobile devices,” *Nature Communications*, vol. 16, no. 1, p. 4212, 2025.

[15] Husain, N. P., H. Arfandy, and R. M. W. Ramli, “Automated medical image processing for lung pneumonia diagnosis based on LS-SVM,” *Journal of System and Computer Engineering*, vol. 6, no. 1, pp. 45–51, 2025, doi: 10.61628/jsce.v6i1.1597.
[16] Khan, M. A., A. Bajwa, and S. T. Hussain, “Pneumonia: Recent updates on diagnosis and treatment,” *Microorganisms*, vol. 13, no. 3, pp. 1–12, 2025, doi: 10.3390/microorganisms13030522.
[17] Lamichhane, B. R., G. Srijuntongsiri, and T. Horanont, “CNN-based 2D object detection techniques: A review,” *Frontiers in Computer Science*, vol. 7, p. 1437664, 2025.
[18] Lin, C., P. Yang, Q. Wang, Z. Qiu, W. Lv, and Z. Wang, “Efficient and accurate compound scaling for convolutional neural networks,” *Neural Networks*, vol. 167, pp. 787–797, 2023, doi: 10.1016/j.neunet.2023.08.053.
[19] Miranda, M., K. Valeriano, and J. Sulla-Torres, “A detailed study on hyperparameter selection for transfer learning in COVID-19 image datasets using Bayesian optimization,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, pp. 327–335, 2021, doi: 10.14569/IJACSA.2021.0120441.
[20] Pamungkas, Y., M. R. N. Ramadani, and E. N. Njoto, “Effectiveness of CNN architectures and SMOTE for handling imbalanced X-ray data in childhood pneumonia detection,” *Journal of Robotics and Control*, vol. 5, no. 3, pp. 775–785, 2024, doi: 10.18196/jrc.v5i3.21494.
[21] Raitoharju, J., “Convolutional neural networks,” in *Deep Learning for Robot Perception and Cognition*, A. Iosifidis and A. Tefas, Eds. Academic Press, 2022, pp. 35–69, doi: 10.1016/B978-0-32-385787-1.00008-7.
[22] Ramadhan, M., D. I. Mulyana, and M. B. Yel, “Optimization of CNN algorithms using transfer learning for pneumonia and non-pneumonia lung X-ray image classification,” *Jurnal Teknik Informatika Kaputama*, vol. 6, no. 2, pp. 670–679, 2022.
[23] Shoushtari, A. H., and K. Nugent, “Diagnosis and treatment of adults with community-acquired pneumonia: An official clinical practice guideline,” *Southwest Respiratory and Critical Care Chronicles*, vol. 8, no. 33, pp. 1–6, 2020, doi: 10.12746/swrccc.v8i33.625.
[24] Sun, K., X. Wang, X. Miao, and Q. Zhao, “A review of AI edge devices and lightweight CNN and LLM deployment,” *Neurocomputing*, vol. 614, p. 128791, 2025, doi: 10.1016/j.neucom.2024.128791.
[25] Tan, M., and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *Proceedings of the International Conference on Machine Learning*, 2019, pp. 6105–6114.
[26] Zuowen, X., “Research on flower classification based on improved EfficientNetB7,” *International Journal of Science and Engineering Applications*, vol. 14, no. 4, pp. 42–52, 2025, doi: 10.7753/ijsea1404.1007.