ILKOM Jurnal Ilmiah Vol. 18, No. 2, August 2026, pp. 302-319
Accredited 2nd by KEMDIKTISAINTEK No. 10/C/C3/DT.05.00/2025; E-ISSN 2548-7779 | P-ISSN 2087-1716

Lightweight Deep Learning Models for Lung Disease Classification Using Chest X-ray Images

Abdullah Sholum[a,1]; Aji Prasetya Wibawa[a,2*]; Ardhana Putra Agustavada[a,3]; Dafa Fadhilah Hilmi[a,4]; Felix Andika Dwiyanto[b,5]
[a]*Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Malang, 65145, Indonesian*
[b] *Faculty of Computer Science, AGH University of Krak´ow, al. Adama Mickiewicza 30, Kraków 30-059, Poland*
[1]*abdullah.sholum.2205356@students.um.ac.id;* [2*]*aji.prasetya.ft@um.ac.id;* [3]*ardha.ardhana.2205356@students.um.ac.id;*
[4]*dafa.fadhilah.2205356@students.um.ac.id;* [5]*dwiyanto@agh.edu.pl*

Article history: Received March 06, 2026; Revised March 12, 2026; Accepted June 04, 2026; Available online August 10, 2026

Abstract

Lung disease classification from chest X-ray images using deep learning has attracted significant attention due to its potential to support rapid and automated medical diagnosis. However, many deep learning models require high computational resources, limiting their applicability in resource-constrained environments. This study evaluates the effectiveness of lightweight deep learning architectures for lung disease classification using chest X-ray images consisting of COVID-19, Normal, and Pneumonia classes. Three architectures were comparatively analyzed, including a baseline Convolutional Neural Network (CNN), EfficientNetV2, and MobileNetV2. All models were trained and evaluated under identical preprocessing and experimental conditions using a dataset of 697 chest X-ray images with a 60:20:20 split ratio for training, validation, and testing. Experimental results show that EfficientNetV2 and MobileNetV2 achieved classification accuracy up to 0.99 with AUC-ROC values approaching 1.0. In terms of computational efficiency, both lightweight architectures reduced trainable parameters by approximately 99.3% compared to the baseline CNN, decreasing from 576,851 to 3,843 trainable parameters through transfer learning with frozen pretrained layers. MobileNetV2 achieved this performance with 0.613 GFLOPs and a model size of 9,435 KB, while EfficientNetV2 required 0.781 GFLOPs and 15,993 KB. Although EfficientNetV2 demonstrated slightly more stable performance, MobileNetV2 provided the most effective trade-off between classification accuracy and computational efficiency, making it more suitable for deployment in resource-constrained healthcare environments. These findings demonstrate the feasibility of lightweight deep learning architectures for efficient and reliable lung disease classification from chest X-ray images.

Keywords: Chest X-ray; CNN; Deep learning; Lung Disease; Lightweight Architectures; Medical Image Analysis

Introduction

Lung diseases remain one of the leading causes of morbidity and mortality worldwide, with conditions such as pneumonia and COVID-19 significantly impacting global public health systems [1], [2]. Early detection plays a crucial role in reducing complications and improving treatment outcomes [3]. Chest X-ray imaging is widely used as a primary diagnostic tool due to its accessibility, cost-effectiveness, and rapid acquisition time [4]. However, manual interpretation of chest X-ray images requires expert radiological assessment and may be affected by inter-observer variability and increasing clinical workload [5]. These challenges highlight the need for supportive diagnostic technologies that can enhance accuracy, consistency, and efficiency in clinical decision-making.

Recent advancements in artificial intelligence, particularly deep learning, have demonstrated promising results in medical image classification tasks [6]. Convolutional Neural Networks (CNNs) have been widely adopted for automated analysis of chest X-ray images due to their capability in extracting hierarchical visual features [7]. Various deep architectures, including ResNet, DenseNet, and EfficientNet, have achieved high classification accuracy in lung disease detection [8], [9] Despite these achievements, many existing approaches rely on computationally intensive models that require substantial hardware resources and large-scale datasets.

In practical healthcare environments, especially in developing regions or primary care facilities, computational resources may be limited [10]. Deploying deep learning systems in such environments necessitates models that not only deliver reliable classification performance but also maintain computational efficiency. Lightweight architectures, such as MobileNet and EfficientNet variants, have been proposed to reduce model complexity through

techniques such as depthwise separable convolutions and compound scaling [11], [12]. These architectures offer a potential balance between predictive performance and resource consumption.

Several previous studies have primarily focused on maximizing classification accuracy using deep architectures such as ResNet and EfficientNet [13], [14]. While these approaches demonstrate strong predictive performance, they often rely on computationally intensive models and large-scale datasets, which limit their applicability in resource-constrained environments. Furthermore, studies such as [15] and [16] discuss the trade-off between model complexity and performance; however, the analysis remains general and lacks empirical evaluation under constrained dataset conditions. In addition, the use of highly curated datasets in prior works [17] may not adequately represent real-world scenarios where data availability is limited and heterogeneous.

Although recent studies have introduced lightweight and edge-oriented architectures, many existing approaches still prioritize predictive accuracy over deployability in practical healthcare systems [18]. In real-world medical environments, especially in developing regions and low-resource healthcare facilities, limited access to high-performance hardware may restrict the implementation of computationally intensive deep learning models [19]. As a result, models designed for medical image classification should not only achieve high predictive performance but also maintain computational efficiency, low memory consumption, and stable inference capability under constrained conditions [20].

In addition to these limitations, the practical deployment of deep learning models in healthcare environments requires consideration of edge computing scenarios, where computational resources, memory capacity, and power consumption are inherently constrained [21]. In such settings, models must be capable of performing efficient inference on low-power devices, such as embedded systems or portable diagnostic tools, without sacrificing classification reliability [22]. This highlights the importance of designing and evaluating lightweight architectures that are not only accurate but also feasible for real-time and on-device implementation [23].

Recent developments in edge artificial intelligence, model compression, and explainable artificial intelligence have further enabled the deployment of deep learning systems in resource-constrained environments [24]. Techniques such as pruning, quantization, and knowledge distillation are commonly employed to reduce model complexity and computational requirements while maintaining classification performance [25]. In addition, explainable artificial intelligence methods are increasingly important in medical applications to improve model transparency, interpretability, and clinical trust during diagnostic decision-making [26].

To further clarify the research gap and position of this study, a comparison of several relevant previous works is presented in **Table 1**.

Table 1. Comparison of Previous Studies on Lung Disease Classification

Model	Dataset Scale	Focus	Computational Analysis
ResNet [27]	Large	Accuracy optimization	Not discussed
EfficientNet [28]	Large	Accuracy improvement	Limited
Various ML models [29]	Medium	Accuracy vs complexity	General analysis
Deep Neural Networks [30]	Large	Robustness and accuracy	Not focused on efficiency
CNN, MobileNetV2, EfficientNetV2 [This Study]	Limited	Accuracy and efficiency trade-off	Detailed parameter and convergence analysis

As shown in Table 1, most previous studies emphasize predictive performance, while limited attention is given to computational efficiency and deployability under constrained conditions [31]. In addition, many of these studies rely on large-scale and well-curated datasets, which may not adequately reflect real-world clinical scenarios characterized by limited and heterogeneous data availability [32]. Furthermore, the lack of explicit evaluation of model complexity, such as parameter efficiency and inference cost, makes it difficult to assess their practical feasibility for deployment on resource-constrained systems [33]. In contrast, this study systematically evaluates both

performance and efficiency aspects using controlled experimental settings, thereby providing a more comprehensive and practically relevant understanding of lightweight model behavior.

Addressing this gap, the novelty of this study lies in its systematic and controlled comparative evaluation of lightweight deep learning architectures for lung disease classification under constrained dataset conditions. The analysis emphasizes not only predictive accuracy but also parameter efficiency, convergence stability, and multi-scenario training behavior across multiple epoch configurations [34]. Unlike prior works that primarily report isolated performance metrics, this research integrates quantitative evaluation, convergence curve analysis, and class-level confusion matrix assessment under identical preprocessing and training pipelines [35].

Furthermore, computational efficiency is explicitly examined through architectural complexity and parameter considerations, enabling a clearer understanding of the trade-off between classification performance and deployability, particularly in edge-based healthcare environments [36]. By integrating predictive evaluation with efficiency-oriented analysis, this study contributes to the development of deployable and scalable artificial intelligence systems for medical image classification.

Considering these factors, the present study examines the capability of lightweight deep learning architectures for classifying lung diseases from chest X-ray images. Several architectures were systematically evaluated, including a conventional baseline CNN and multiple lightweight models, using identical preprocessing procedures and training settings to ensure fair comparison. The study aims to assess classification performance, convergence behavior during training, and the balance between predictive capability and computational efficiency. Through a comparative and efficiency-focused evaluation framework, this research seeks to support the advancement of scalable and practically deployable artificial intelligence-based diagnostic systems in the domains of computer science and health informatics.

Method

This study employs an experimental research method to evaluate the performance of lightweight deep learning models for lung disease classification using chest X-ray images, with an emphasis on computational efficiency and reliable classification accuracy. The overall research workflow of the proposed method is illustrated in **Figure 1**, which presents the sequential stages of the experimental process.

As shown in **Figure 1**, the workflow begins with the dataset preparation stage, where chest X-ray images are collected and categorized into Normal, Pneumonia, and COVID-19 classes. At this stage, label verification and validation of the number of samples for each class are conducted to ensure data consistency and reliability. The next stage involves data preprocessing, which includes relabeling, image normalization, image size adjustment, and the distribution of data into training, validation, and testing subsets.

Following preprocessing, the study proceeds to the model architecture and training stage. In this phase, three different architectures namely a conventional CNN, MobileNetV2, and EfficientNetV2 are initialized [37], [38]. Each model is trained under multiple training scenarios with varying numbers of epochs, ranging from 10 to 50 epochs, to analyze the effect of training duration on model performance. This multi-scenario training strategy enables a fair and systematic comparison of learning behavior across architectures.

Finally, the trained models are evaluated using standard classification performance metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, as summarized in the evaluation stage of **Figure 1**. These metrics are used to assess both the predictive performance and robustness of each model, providing a comprehensive evaluation of lightweight deep learning approaches for lung disease classification.

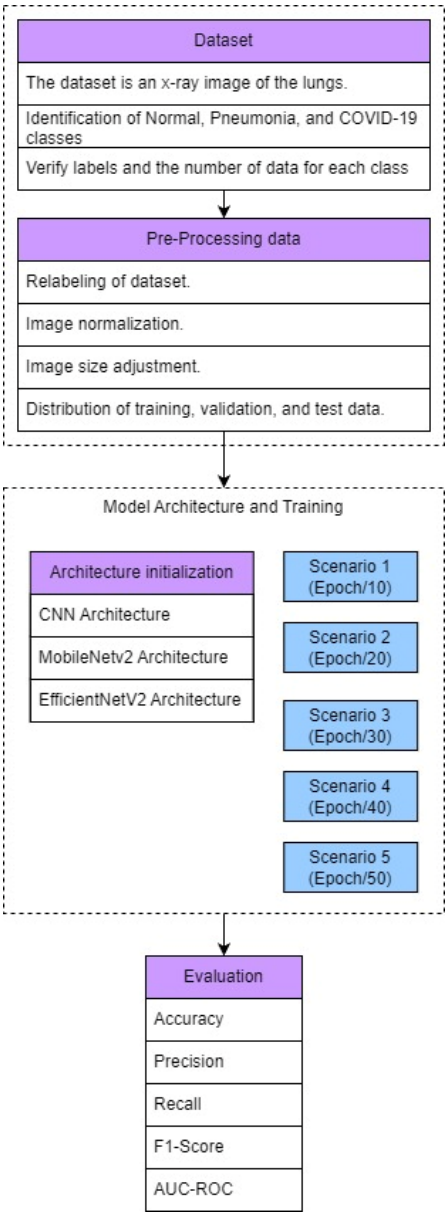

Figure 1. Research Method

A. Problem Formulation

Let $X = \{x_1, x_2, \dots, x_N\}$denote a set of chest X-ray images, where each image $x_i \in \mathbb{R}^{H \times W \times C}$represents an input image with height H, width W, and channel C. Each image is associated with a corresponding ground-truth label $y_i \in \mathcal{Y}$, where $\mathcal{Y} = \{1,2, \dots, K\}$denotes the set of lung disease classes.

The lung disease classification task can be formulated as a supervised multi-class classification problem. The objective is to learn a mapping function $f(\cdot;\theta)$, parameterized by θ, that predicts the class label $\hat{y}_i$for a given input image x_i, as expressed in Equation (1).

$$\hat{y}_i = f(x_i;\theta) \qquad (1)$$

During training, the model parameters $\theta$are optimized by minimizing the categorical cross-entropy loss function, which measures the discrepancy between the predicted class probability distribution $\hat{p}_i$and the true label y_i. The loss function is defined in Equation (2):

$$\mathcal{L}(\theta) = -\frac{1}{N}\sum_{i=1}^{N} \sum_{k=1}^{K} y_{i,k} \log(\hat{p}_{i,k}) \qquad (2)$$

Where Ndenotes the number of training samples, Krepresents the total number of classes, $y_{i,k} \in \{0,1\}$is the one-hot encoded ground-truth label for class k, and $\hat{p}_{i,k}$is the predicted probability of class $k$generated by the softmax layer of the network.

The optimization process aims to find an optimal parameter set θ^*that minimizes the loss function, as expressed in Equation (3):

$$\theta^* = \arg\min_{\theta} \mathcal{L}(\theta) \quad (3)$$

These formulations provide a mathematical foundation for the proposed lung disease classification model. By minimizing the categorical cross-entropy loss, the model is trained to maximize the probability of correctly classifying chest X-ray images into their respective lung disease categories while maintaining computational efficiency through the use of lightweight deep learning architectures.

B. Dataset and Preprocessing

The dataset used in this study consists of chest X-ray images categorized into three classes: COVID-19, Normal, and Pneumonia. A detailed summary of the dataset composition is presented in **Table 2**. The COVID-19 class contains 257 images representing pulmonary infection caused by SARS-CoV-2, characterized by radiographic features such as ground-glass opacity and bilateral lung involvement. The Pneumonia class consists of 250 images showing lung infections with visible infiltrates or consolidation on X-ray images. The Normal class includes 190 chest X-ray images without observable signs of infection or pathological abnormalities.

Table 2. Dataset Information

Class	Number	Description	Image Example
COVID-19	257	Pulmonary infection caused by SARS-CoV-2 with ground-glass opacity and bilateral [39]	
Normal	190	Lungs without signs of infection or pathological abnormalities visible on X-ray images	
Pneumonia	250	Lung infection with infiltrates or consolidation visible on X-ray images	

Prior to model training, all images are resized to a fixed resolution of 224 × 224 pixels to ensure uniform input dimensions across different convolutional neural network architectures. Pixel value normalization is applied by scaling image intensities to the range 0–1, which enhances numerical stability and facilitates model convergence during optimization. Class labels are assigned according to the dataset structure and encoded into categorical format

to support multi-class classification. No data augmentation techniques were applied in this study in order to preserve the original characteristics of the chest X-ray images and maintain controlled experimental conditions for comparative evaluation across different lightweight architectures.

Following preprocessing, the dataset was randomly shuffled and divided into training, validation, and testing subsets using an approximate ratio of 60 percent 20 percent and 20 percent. A total of 417 images were allocated for training, while 140 images were used for validation and 140 images were reserved for testing. This partitioning strategy enables supervised learning, hyperparameter tuning, and unbiased final performance evaluation under mutually exclusive subsets [40].

The validation subset was employed to monitor training progress and support model selection during optimization, while the testing subset was used exclusively for final performance assessment. This approach allows consistent comparison across different model architectures while preserving the original class distribution of the dataset.

K-fold cross-validation was not employed in this study due to several practical considerations. First, the research focuses on comparative evaluation of multiple lightweight deep learning architectures under various epoch configurations, which already requires substantial computational resources and repeated training procedures. Second, the relatively limited dataset size and controlled experimental setting were considered sufficient for conducting systematic model comparison under consistent conditions. Finally, maintaining a fixed data partitioning strategy ensures reproducibility and consistency during efficiency-oriented evaluation across all experimental scenarios.

The dataset used in this study is relatively limited in size and exhibits homogeneous characteristics across classes [41]. While this condition enables a more controlled experimental evaluation and reduces variability during model training, it may not fully represent the complexity and diversity of real-world clinical data. In practical settings, chest X-ray images often vary significantly in terms of acquisition conditions, patient demographics, noise levels, and disease manifestations [42]. Therefore, the use of a relatively uniform dataset may lead to optimistic performance estimates and limit the generalization capability of the models when applied to more heterogeneous data [42]. Nevertheless, this controlled setting remains valuable for systematically analyzing model behavior and comparing the efficiency and performance of different architectures under consistent conditions.

C. Lightweight Deep Learning Architecture

The architecture of the deep learning models used in this study is illustrated in **Figure 2**. As shown in the figure, three model configurations are considered, including a baseline convolutional neural network (CNN), EfficientNetV2, and MobileNetV2. Each architecture is designed to extract discriminative features from chest X-ray images while maintaining computational efficiency.

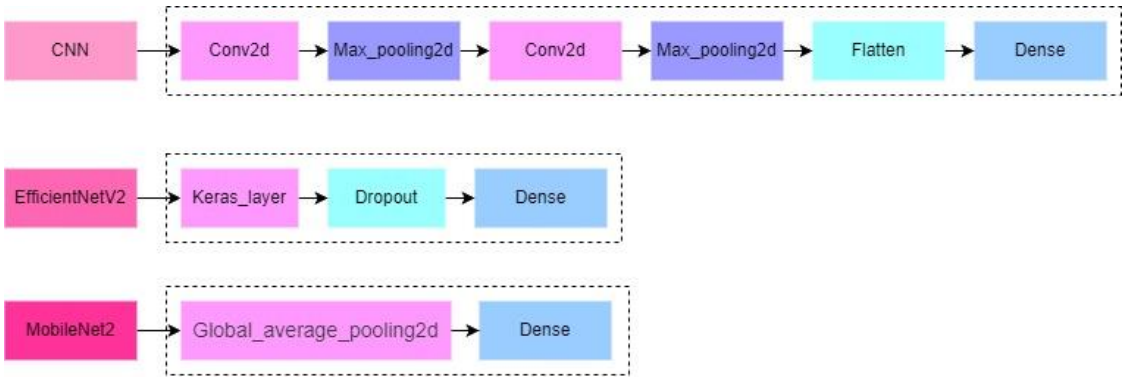

Figure 2. Architecture Model.

The baseline CNN architecture consists of sequential convolutional and max-pooling layers, followed by a flatten operation and fully connected layers, as depicted in **Figure 2**. The convolutional layers are responsible for learning low- to mid-level spatial features from input images, while max-pooling layers reduce spatial dimensions and computational complexity. This baseline model is used as a reference to evaluate the effectiveness of more

advanced lightweight architectures [43]. As a conventional deep learning structure, it provides a comparative foundation for assessing architectural improvements in terms of efficiency and performance.

EfficientNetV2, as illustrated in **Figure 2**, is employed to improve feature extraction efficiency through a modern lightweight design. The architecture utilizes a pre-trained feature extractor followed by a dropout layer and a dense classification layer. The inclusion of dropout helps reduce overfitting, particularly when training on a limited dataset. EfficientNetV2 is selected due to its compound scaling strategy, which balances network depth, width, and resolution to achieve strong performance with relatively low computational cost [44]. This design makes EfficientNetV2 an appropriate candidate for evaluating performance optimization under efficiency-oriented architectural scaling.

MobileNetV2 adopts an even more compact architecture, as shown in **Error! Reference source not found.**, by incorporating a global average pooling layer followed by a dense output layer. This design eliminates the need for large fully connected layers and significantly reduces the number of trainable parameters. MobileNetV2 leverages depthwise separable convolutions to achieve efficient computation, making it suitable for deployment in resource-constrained environments [45]. Such architectural characteristics highlight its relevance for applications requiring lightweight yet reliable medical image classification models. Overall, the architectures presented in Figure 2 demonstrate different design strategies for achieving efficient lung disease classification from chest X-ray images. By comparing these models, this study aims to analyze the trade-off between classification performance and computational efficiency in lightweight deep learning architectures.

D. Training Strategy

Model training is conducted using supervised learning. The categorical cross-entropy loss function is used to measure prediction errors, while the Adam optimizer is employed to update model parameters with an initial learning rate of 0.001. A batch size of 32 is used during training to balance computational efficiency and model convergence. The training process incorporates early stopping with a patience of 5 epochs and learning rate scheduling to prevent overfitting and ensure stable convergence. To analyze the effect of training duration on model performance, several training scenarios are defined using different numbers of epochs, as illustrated in **Figure 1**.

The dataset is divided into training, validation, and testing subsets to allow objective evaluation of model performance. The validation set is used to monitor training progress and optimize hyperparameters, while the testing set is reserved exclusively for final performance assessment. This experimental setup enables a systematic evaluation of model performance under varying training conditions.

All experiments were implemented using Python 3.11.5 with TensorFlow versions 2.15.0 and 2.20.0 together with Keras version 3.12.0 on the Windows 10 operating system. The experiments were conducted on a laptop equipped with an AMD Ryzen 5 processor, 16 GB of system memory, and an NVIDIA GeForce GTX 1650 GPU with 4 GB of VRAM. GPU acceleration was utilized to improve computational efficiency during training. On average, each model required approximately 60 to 460 seconds to complete a single training scenario depending on the model architecture and number of epochs. This configuration represents a mid-range computational environment suitable for evaluating lightweight deep learning models under resource-constrained conditions.

E. Performance Evaluation

The performance of the proposed lightweight deep learning models is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. These metrics are commonly employed in multi-class classification problems to provide a comprehensive assessment of model effectiveness [46]. The evaluation is conducted on the testing dataset, which is kept independent from the training and validation processes.

Accuracy measures the overall correctness of the classification results and is defined as the ratio of correctly classified samples to the total number of samples, as shown in Equation (4).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (4)$$

Precision evaluates the reliability of positive predictions by measuring the proportion of correctly predicted positive samples among all samples predicted as positive. Precision is defined in Equation (5).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

Recall, also known as sensitivity, measures the ability of the model to correctly identify positive samples. It is defined as the ratio of true positive predictions to the total number of actual positive samples, as expressed in Equation (6).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

The F1-score provides a balanced measure between precision and recall by computing their harmonic mean. This metric is particularly useful when dealing with class imbalance. The F1-score is defined in Equation (7).

$$\text{F1-score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (7)$$

In addition to these metrics, confusion matrix analysis is employed to examine class-wise prediction performance and identify common misclassification patterns among lung disease categories. This analysis provides further insight into the strengths and limitations of the evaluated models beyond aggregated performance scores.

Results and Discussion

This section describes the experimental findings derived from the assessment of three deep learning architectures for lung disease classification based on chest X-ray imagery, including a baseline Convolutional Neural Network (CNN), EfficientNetV2, and MobileNetV2. All architectures were trained and tested using the same preprocessing procedures and experimental conditions, with epoch variations ranging from 10 to 50. The evaluation emphasizes both predictive performance and computational efficiency by employing quantitative indicators such as accuracy, precision, recall, F1-score, and AUC-ROC. Furthermore, patterns of training and validation convergence, together with confusion matrix evaluation, were analyzed to obtain a broader understanding of model learning behavior and category-specific prediction outcomes. The experimental findings were subsequently examined to identify the balance between classification capability and computational efficiency across the evaluated lightweight deep learning models.

A. Result

1) CNN

Table 3. CNN Classification Performance Report

Scenario	Accuracy	Precision	Recall	F1-score	AUC-ROC
Epoch 10	0.98	0.98	0.98	0.98	0.9986
Epoch 20	0.97	0.97	0.97	0.97	0.9984
Epoch 30	0.93	0.94	0.93	0.93	0.9944
Epoch 40	0.98	0.98	0.98	0.98	0.9986
Epoch 50	0.97	0.97	0.97	0.97	0.9991

Table 3 presents the classification performance of the baseline CNN model across multiple epoch configurations ranging from 10 to 50 epochs. The findings demonstrate that the CNN architecture achieved consistently high classification performance under all experimental conditions, with accuracy values ranging from 0.93 to 0.98. The best performance was recorded in the epoch 10 and epoch 40 configurations, where accuracy, precision, recall, and F1-score each reached 0.98.

In addition, the AUC-ROC scores remained consistently close to 1.0 throughout all training scenarios, indicating strong discriminative capability among the COVID-19, Normal, and Pneumonia categories. Despite these results, performance inconsistencies were observed in several epoch configurations, particularly in the epoch 30 setting, suggesting that the baseline CNN exhibited greater sensitivity to training duration and less stable learning behavior compared with lightweight transfer learning architectures. Figure 3 illustrates the training and validation accuracy and loss curves obtained from the CNN model.

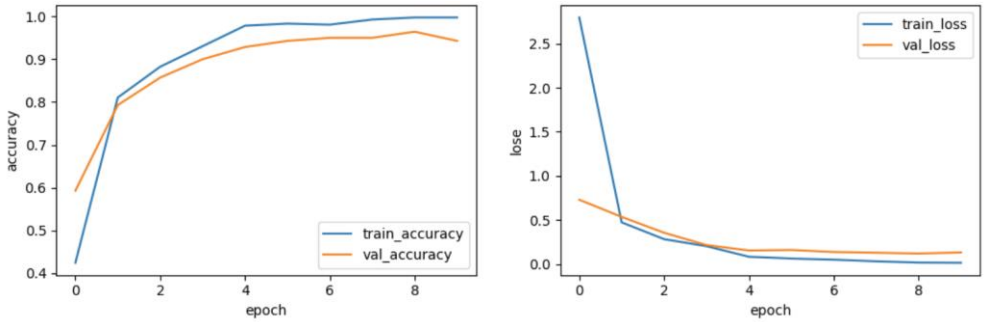

Figure 3. Training and validation accuracy and loss curves of the CNN model under the Epoch 10 scenario

Figure 3 illustrates the training and validation accuracy and loss patterns of the CNN model in the epoch 10 configuration. The accuracy trend demonstrates substantial improvement during the initial training phase. Training accuracy increased rapidly from approximately 0.42 in the first epoch and approached 1.00 by the final observed epoch. Validation accuracy also showed a steady upward trend and achieved a high-performance level, although a slight decline was observed toward the end of the training process.

The loss curves indicate a continuous reduction in both training and validation loss values. Training loss decreased sharply during the early epochs and continued to decline until reaching a minimal value. Similarly, validation loss showed a reduction in the initial stage before stabilizing after several epochs. The training procedure was terminated at epoch 8 through the implementation of an early stopping mechanism, which halted the process once validation performance no longer exhibited significant improvement. This pattern suggests that the CNN model attained stable convergence prior to reaching the maximum epoch limit, while the early stopping strategy contributed to minimizing unnecessary training and limiting the possibility of overfitting.

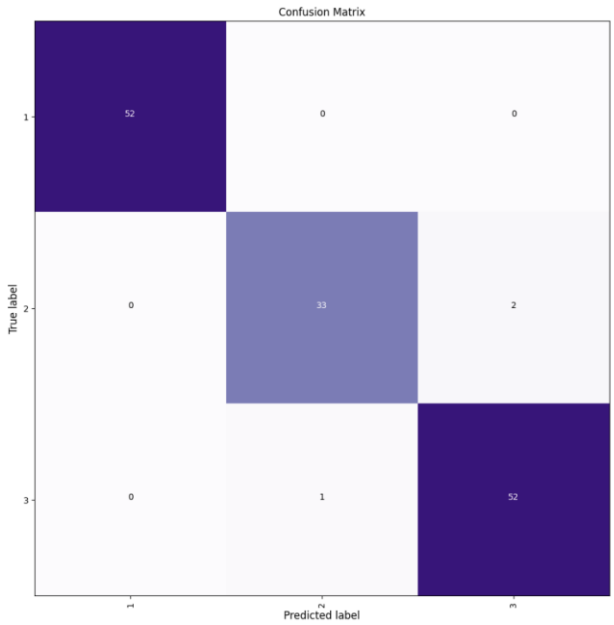

Figure 4. Confusion matrix of the CNN model under the Epoch 10 scenario.

Figure 4 illustrates the confusion matrix generated by the CNN model in the 10-epoch experiment. The distribution of predictions demonstrates that the majority of samples were assigned to their correct categories, as reflected by the dominant values positioned along the main diagonal of the matrix. The model successfully classified 52 samples belonging to class 1, 33 samples from class 2, and 52 samples associated with class 3. These findings indicate that the CNN architecture achieved strong classification performance at the class level across the three lung disease categories.

Only a limited number of prediction errors were identified in this configuration. Specifically, two samples from class 2 were incorrectly predicted as class 3, whereas one sample from class 3 was classified as class 2. No incorrect

predictions were observed for class 1, demonstrating perfect classification performance for this category within the testing dataset. Overall, the confusion matrix indicates that the CNN model produced reliable classification outcomes in the 10-epoch scenario, with misclassification occurring only minimally between class 2 and class 3.

2) EfficientNetV2

Table 4. EfficientNetV2 Classification Performance Report

Scenario	Accuracy	Precision	Recall	F1-score	AUC-ROC
Epoch 10	0.98	0.98	0.98	0.98	0.9965
Epoch 20	0.98	0.98	0.98	0.98	0.9974
Epoch 30	0.99	0.99	0.99	0.99	0.9982
Epoch 40	0.99	0.99	0.99	0.99	0.9993
Epoch 50	0.98	0.98	0.98	0.98	0.9995

Table 4 reports the classification performance of the EfficientNetV2 model across several training configurations. The model produced consistently strong results in all epoch settings, with accuracy values ranging from 0.98 to 0.99. Precision, recall, and F1-score also remained stable, indicating balanced classification ability across the COVID-19, Normal, and Pneumonia categories. The best results were achieved in the epoch 30 and epoch 40 configurations, where accuracy, precision, recall, and F1-score each reached 0.99. The AUC-ROC values also remained higher than 0.996 in all experiments, with the maximum value of 0.9995 recorded in the epoch 50 configuration.

These findings suggest that EfficientNetV2 delivered highly accurate predictions and demonstrated strong capability in distinguishing among the three classes throughout the training process. Overall, the evaluation metrics indicate that EfficientNetV2 maintained stable classification performance across different epoch configurations and provided high predictive effectiveness for lung disease classification using chest X-ray images. To further analyze its learning behavior, the training and validation accuracy and loss curves for the epoch 40 configuration are shown in the following figure.

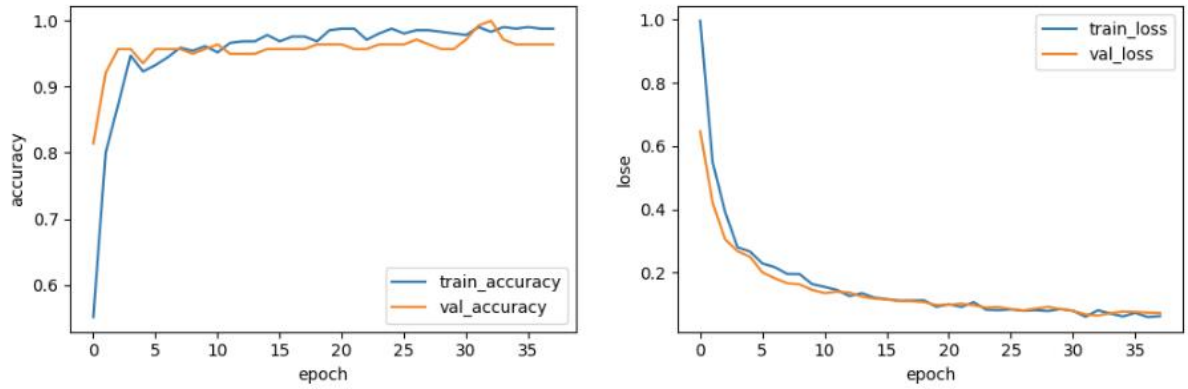

Figure 5. Training and validation accuracy and loss curves of the EfficientNetV2 model under the Epoch 40 scenario

Figure 5 illustrates the training and validation accuracy and loss curves of the EfficientNetV2 model in the epoch 40 configuration. The accuracy curve indicates a rapid improvement during the initial training stage, particularly within the first several epochs. Training accuracy rose sharply from a low starting point and then progressively stabilized at approximately 0.98. Validation accuracy also reached a high level early in the training process and remained relatively consistent, with only slight fluctuations.

The loss curve demonstrates a steady decline in both training and validation loss. Training loss decreased substantially at the beginning and continued to drop gradually until the final recorded epoch. Validation loss showed a similar trajectory and remained close to the training loss curve. The narrow gap between training and validation loss suggests effective learning with no indication of severe overfitting. Although the experiment was initially set to run for epoch 40, training was stopped at epoch 35 by the early stopping mechanism because validation performance no longer showed substantial improvement.

This pattern indicates that EfficientNetV2 achieved stable convergence before reaching the maximum epoch limit. Overall, the curves demonstrate that EfficientNetV2 maintained consistent learning behavior on both training and validation data under the epoch 40 configuration. To examine class-level classification performance in greater detail, the confusion matrix of the EfficientNetV2 model for the epoch 40 scenario is presented in the following figure.

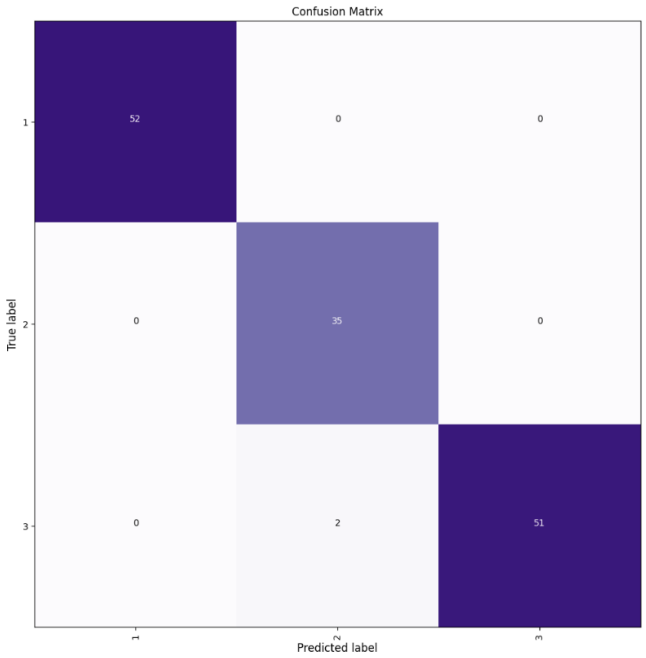

Figure 6. Confusion matrix of the EfficientNetV2 model in the Epoch 40 scenario.

Figure 6 presents the confusion matrix generated by the EfficientNetV2 model in the epoch 40 experiment. The distribution within the matrix demonstrates that the majority of samples were assigned to their correct classes, as reflected by the dominant values located along the principal diagonal. The model accurately classified 52 samples belonging to class 1, 35 samples from class 2, and 51 samples from class 3. These findings indicate that EfficientNetV2 achieved strong classification capability across the three lung disease categories.

Only a limited number of prediction errors were identified in this configuration. In particular, two samples from class 3 were incorrectly predicted as class 2, while no misclassification was observed for class 1 or class 2. This outcome suggests that the model possessed strong class-specific prediction performance, with only minimal confusion occurring between two categories. Overall, the confusion matrix confirms that EfficientNetV2 delivered reliable classification results under the epoch 40 configuration, supported by a high proportion of correct predictions and a very low classification error rate.

3) MobileNetV2

Table 5. MobileNetV2 Classification Performance Report

Scenario	Accuracy	Precision	Recall	F1-score	AUC-ROC
Epoch 10	0.97	0.97	0.97	0.97	0.9982
Epoch 20	0.99	0.99	0.99	0.99	0.9996
Epoch 30	0.97	0.97	0.97	0.97	0.9990
Epoch 40	0.99	0.99	0.99	0.99	0.9996
Epoch 50	0.99	0.99	0.99	0.99	0.9995

Table 5 presents the classification performance of the MobileNetV2 model under multiple epoch configurations ranging from 10 to 50 epochs. The evaluation results show that MobileNetV2 achieved consistently high classification performance across all experimental scenarios, with accuracy values ranging from 0.97 to 0.99. The

highest accuracy of 0.99 was obtained in the Epoch 20, Epoch 40, and Epoch 50 scenarios, while precision, recall, and F1-score also reached 0.99 under the same configurations.

In addition, the AUC-ROC values consistently remained above 0.998, with the highest value reaching 0.9996, indicating excellent class separability among COVID-19, Normal, and Pneumonia categories. These results demonstrate that MobileNetV2 maintains highly competitive classification capability while utilizing an efficient lightweight architecture. Furthermore, the balanced precision and recall values across all scenarios indicate stable and unbiased prediction performance across all classes. Figure 7 presents the training and validation accuracy and loss curves of the MobileNetV2.

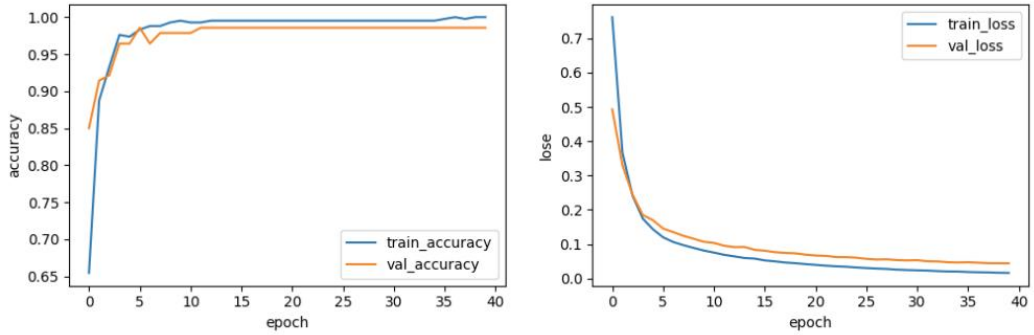

Figure 7. Training and validation accuracy and loss curves of the MobileNetV2 model under the Epoch 40 scenario.

Figure 7 presents the training and validation accuracy and loss curves of the MobileNetV2 model under the Epoch 40 scenario. The accuracy curve shows a sharp increase during the early training phase. Training accuracy rose from approximately 0.65 at the initial epoch and reached a high value within the first few epochs. Validation accuracy followed a similar pattern, increasing rapidly and remaining close to the training accuracy curve throughout the training process. After several epochs, both curves became stable, indicating consistent learning performance.

The loss curve shows a steady decrease in both training and validation loss. Training loss declined sharply during the early epochs and continued to decrease gradually until the final epoch. Validation loss also decreased consistently, although it remained slightly higher than training loss. The small gap between the two loss curves indicates that the model maintained good generalization performance without severe overfitting. Overall, these curves show that MobileNetV2 achieved stable convergence under the Epoch 40 configuration and maintained consistent performance on both training and validation data.

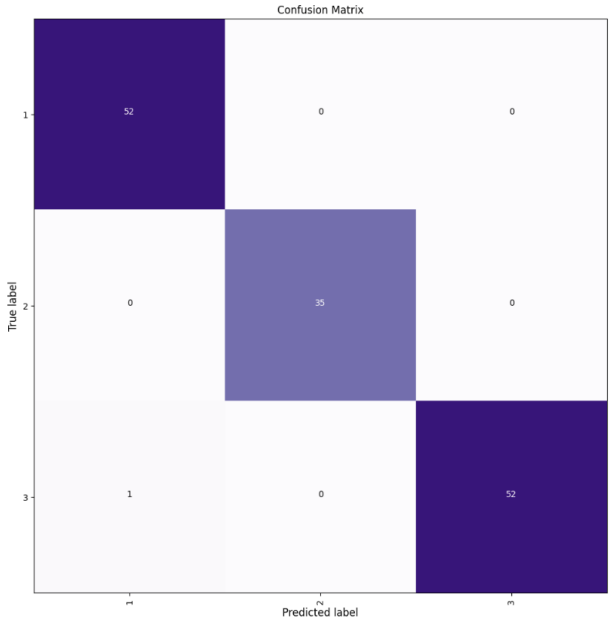

Figure 8. Confusion matrix of the MobileNetV2 model in the Epoch 40 scenario.

Figure 8 shows the confusion matrix of the MobileNetV2 model under the Epoch 40 scenario. The matrix indicates that almost all samples were correctly classified, as shown by the dominant values on the main diagonal. The model correctly identified 52 samples from class 1, 35 samples from class 2, and 52 samples from class 3. These results show that MobileNetV2 achieved strong class-level classification performance across the three lung disease categories.

Only one misclassification was observed in this scenario, where one sample from class 3 was classified as class 1. No misclassification occurred in class 1 and class 2, indicating that the model classified these categories accurately in the testing set. Overall, the confusion matrix shows that MobileNetV2 produced reliable classification results under the Epoch 40 configuration, with a very low classification error and strong predictive consistency across all classes.

Table 6. Comparative Performance and Computational Efficiency Analysis of CNN, MobileNetV2, and EfficientNetV2 Models

	CNN	**MobileNetV2**	**EfficientNetV2**
Epoch	10	40	40
Accuracy	0.98	0.99	0.99
Precision	0.98	0.99	0.99
F1-Score	0.98	0.99	0.99
AUC-Roc	0.9986	0.9993	0.9996
Total Parameter	576,851	2,261,827	4,053,407
Model Size (MB)	6.79 MB	9.44 MB	15.99 MB
Training Time	92s	262s	406s
GFLOPs	0.463 GFLOPs	0.613 GFLOPs	0.781 GFLOPs
Testing Time	1s	2s	3s

Table 6 presents a comparative assessment of both predictive performance and computational efficiency among the CNN, MobileNetV2, and EfficientNetV2 architectures using their optimal epoch settings. The evaluation results show that MobileNetV2 and EfficientNetV2 achieved the best classification performance, obtaining accuracy, precision, and F1-score values of 0.99, surpassing the baseline CNN model, which consistently reached 0.98 for these metrics. In terms of discriminative capability, EfficientNetV2 produced the highest AUC-ROC score of 0.9996, followed closely by MobileNetV2 with 0.9993, indicating excellent separability across the evaluated lung disease classes. From the computational perspective, the CNN architecture demonstrated the lowest computational burden, including smaller model size, lower FLOPs, and shorter training and inference times. However, the transfer learning-based architectures required additional computational resources due to their more complex network structures. Despite this, MobileNetV2 exhibited a more favorable trade-off between accuracy and efficiency compared to EfficientNetV2 by maintaining comparable predictive performance while requiring lower FLOPs, reduced model size, shorter training duration, and faster inference speed. Overall, these findings highlight MobileNetV2 as the most balanced architecture for lightweight medical image classification tasks in resource-constrained healthcare environments.

B. Discussion

The experimental findings indicate that lightweight deep learning architectures are able to deliver highly competitive performance in lung disease classification while preserving computational efficiency. EfficientNetV2 and MobileNetV2 consistently achieved accuracy scores of up to 0.99, with AUC-ROC values close to 1.0 across different epoch configurations. These results reflect strong discrimination among the COVID-19, Normal, and Pneumonia categories. The baseline CNN model also showed strong classification results, with accuracy reaching 0.98. Nevertheless, its performance varied more across epoch settings, suggesting less stable learning behavior when trained on a limited dataset.

Among the assessed models, EfficientNetV2 showed the most stable convergence pattern, as indicated by smooth training and validation curves and consistently high evaluation scores across multiple training scenarios. The compound scaling mechanism used in EfficientNetV2 appears to facilitate efficient feature extraction and stable

optimization while retaining computational efficiency. MobileNetV2 also produced highly competitive results despite its compact structure and lower number of trainable parameters. Its depthwise separable convolution design supported efficient feature learning and maintained balanced prediction performance across classes, with only minor misclassification between categories with similar visual characteristics.

Overall, the results suggest that lightweight transfer learning models offer a stronger balance between classification performance and computational efficiency than the conventional CNN architecture. The combination of high accuracy, stable convergence, and reduced parameter requirements indicates the suitability of EfficientNetV2 and MobileNetV2 for resource-limited healthcare settings and lightweight medical image analysis applications.

Table 7. Comparison with Previous Studies on Lightweight Lung Disease Classification

Study	Method	Acc (%)	Pre	Rec	F1-Sc
Kajal Kansal et al. 2025 [47]	EfficientNetB7	99.42%	99.42%	99.42%	99.42%
Theodora Sanida et al. 2024 [48]	Lightweight CNN (FL)	98.56%	98.56%	98.56%	98.56%
Chih-Ta Yen et al. 2024 [49]	Custom CNN	97.03%	97.05%	97.03%	97.03%
Proposed Model	MobileNetV2	99.00%	99.00%	99.00%	99.00%

Table 7 compares the proposed MobileNetV2 model with several earlier studies on lightweight lung disease classification using chest X-ray images. The experimental results confirm that lightweight deep learning architectures, especially EfficientNetV2 and MobileNetV2, are capable of providing reliable lung disease classification performance while requiring relatively modest computational resources and fewer trainable parameters. These advantages indicate strong applicability in healthcare systems with limited hardware capacity, such as portable medical diagnostic devices, embedded healthcare technologies, and edge AI-based clinical applications. In particular, MobileNetV2 demonstrated superior computational efficiency through lower memory requirements, faster inference processing, and easier deployment feasibility while still maintaining competitive classification accuracy. In addition, the study highlights the growing significance of efficiency-driven deep learning methods in medical imaging, where lightweight neural network models can maintain a balanced trade-off between predictive effectiveness and computational practicality for real-world healthcare implementation.

The findings of this study demonstrate that lightweight deep learning models, particularly EfficientNetV2 and MobileNetV2, can achieve high-performance lung disease classification while maintaining relatively low computational demands and minimizing trainable parameters. Such characteristics highlight their potential for deployment in resource-constrained healthcare environments, including portable screening devices, embedded medical systems, and edge AI-oriented diagnostic applications. Among the evaluated architectures, MobileNetV2 showed notable efficiency advantages by enabling faster inference speed, reduced memory consumption, and more feasible real-world implementation without substantial degradation in predictive performance. Furthermore, these results reinforce the importance of efficiency-oriented deep learning approaches in medical imaging, where compact neural network architectures can offer an effective compromise between classification capability and computational sustainability for practical healthcare applications.

Despite achieving encouraging classification performance, this study still presents several limitations that should be carefully considered. The experimental evaluation relied on a relatively limited chest X-ray dataset, which may reduce the generalizability of the developed models when applied to more diverse clinical environments and imaging variations. Furthermore, several important validation aspects were not incorporated in the current work, including the use of external datasets, k-fold cross-validation, explainable artificial intelligence approaches, and real-world deployment analysis on edge devices. As a result, practical factors such as inference latency and hardware resource utilization were not comprehensively assessed. Future research is therefore recommended to utilize larger and more heterogeneous datasets, integrate explainable AI methods to enhance clinical transparency, and perform more rigorous statistical and external validation procedures. In addition, further investigation into lightweight optimization strategies and real-device deployment evaluation is necessary to improve the robustness, reliability, and practical implementation of medical AI systems.

Conclusion

This study investigated the performance of lightweight deep learning architectures for lung disease classification using chest X-ray images in environments with limited data availability and computational resources. The experimental findings showed that EfficientNetV2 and MobileNetV2 achieved the best classification results, obtaining accuracy values up to 0.99 and AUC-ROC scores exceeding 0.999 under several training configurations, while significantly reducing trainable parameters compared to the conventional CNN model. The comparative evaluation also confirmed that lightweight architectures can preserve high predictive capability while enhancing computational efficiency and deployment practicality. Although EfficientNetV2 demonstrated slightly superior stability and discriminative performance, MobileNetV2 offered the most balanced trade-off between classification accuracy and computational cost due to its lower FLOPs and lighter architectural complexity. These characteristics make MobileNetV2 more appropriate for deployment in lightweight medical AI applications and resource-limited healthcare settings. However, the study is constrained by the relatively small and homogeneous dataset used during experimentation. Future studies should therefore consider larger and more diverse datasets, more robust validation approaches such as cross-validation and statistical testing, the integration of explainable AI methods, and real-world edge deployment evaluations to improve model reliability, interpretability, and clinical applicability.

Acknowledgement

The authors would like to express their sincere appreciation to the Department of Electrical Engineering and Informatics, Universitas Negeri Malang, for providing academic support and research facilities that contributed to the completion of this study. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] J. Oh *et al.*, "Global, regional, and national burden of chronic respiratory diseases and impact of the COVID-19 pandemic, 1990–2023: a Global Burden of Disease study," *Nat. Med.*, vol. 32, no. 1, pp. 197–223, Jan. 2026, doi: 10.1038/s41591-025-04077-9.

[2] H. Gan *et al.*, "Smoking: a leading factor for the death of chronic respiratory diseases derived from Global Burden of Disease Study 2019," *BMC Pulm. Med.*, vol. 22, no. 1, p. 149, Dec. 2022, doi: 10.1186/s12890-022-01944-w.

[3] I. Ahmad *et al.*, "New paradigms to break barriers in early cancer detection for improved prognosis and treatment outcomes," *J. Gene Med.*, vol. 26, no. 8, Aug. 2024, doi: 10.1002/jgm.3730.

[4] D. Raval *et al.*, "Cost-effectiveness analysis of AI-assisted chest X-ray interpretation tools for TB screening: a rapid HTA," *Front. Digit. Health*, vol. 7, Dec. 2025, doi: 10.3389/fdgth.2025.1629127.

[5] H. H. Pham, H. Q. Nguyen, H. T. Nguyen, L. T. Le, and L. Khanh, "An Accurate and Explainable Deep Learning System Improves Interobserver Agreement in the Interpretation of Chest Radiograph," *IEEE Access*, vol. 10, pp. 104512–104531, 2022, doi: 10.1109/ACCESS.2022.3210468.

[6] T. Mahmood, A. Rehman, T. Saba, L. Nadeem, and S. A. O. Bahaj, "Recent Advancements and Future Prospects in Active Deep Learning for Medical Image Segmentation and Classification," *IEEE Access*, vol. 11, pp. 113623–113652, 2023, doi: 10.1109/ACCESS.2023.3313977.

[7] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, and K. Wang, "Automated detection of COVID-19 through convolutional neural network using chest x-ray images," *PLoS One*, vol. 17, no. 1, p. e0262052, Jan. 2022, doi: 10.1371/journal.pone.0262052.

[8] V. Ravi, V. Acharya, and M. Alazab, "A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images," *Cluster Comput.*, vol. 26, no. 2, pp. 1181–1203, Apr. 2023, doi: 10.1007/s10586-022-03664-6.

[9] G. A. Sandag and D. T. Kabo, "Comparative Analysis of Lung Cancer Classification Models Using EfficientNet and ResNet on CT-Scan Lung Images," *CogITo Smart Journal*, vol. 10, no. 1, pp. 259–270, Jun. 2024, doi: 10.31154/cogito.v10i1.706.680-691.

[10] M. N. Akhtar, A. Haleem, and M. Javaid, "Scope of health care system in rural areas under Medical 4.0 environment," *Intelligent Pharmacy*, vol. 1, no. 4, pp. 217–223, Dec. 2023, doi: 10.1016/j.ipha.2023.07.003.

[11] V.-T. Hoang and K.-H. Jo, "Practical Analysis on Architecture of EfficientNet," in *2021 14th International Conference on Human System Interaction (HSI)*, IEEE, Jul. 2021, pp. 1–4. doi: 10.1109/HSI52170.2021.9538782.

[12] S. Muhammad Raza, S. Murtaza Hussain Abidi, M. Masuduzzaman, and S. Y. Shin, "DSConvNet: A Lightweight Architecture for Extracting Image Features From Depthwise Separable Convolution Network for Edge Devices," *IEEE Access*, vol. 13, pp. 210102–210116, 2025, doi: 10.1109/ACCESS.2025.3639410.

[13] M. Chowdhury, S. R. Shrima, and M. S. Islam, "Comparative analysis of deep learning architectures for multi-class mineral classification: a study using EfficientNet and ResNet models," *Earth Sci. Inform.*, vol. 18, no. 3, p. 485, Sep. 2025, doi: 10.1007/s12145-025-01983-x.

[14] K. Kansal, T. B. Chandra, and A. Singh, "ResNet-50 vs. EfficientNet-B0: Multi-Centric Classification of Various Lung Abnormalities Using Deep Learning," *Procedia Comput. Sci.*, vol. 235, pp. 70–80, 2024, doi: 10.1016/j.procs.2024.04.007.

[15] H. Louati, A. Louati, K. Mansour, and E. Kariri, "Achieving Faster and Smarter Chest X-Ray Classification With Optimized CNNs," *IEEE Access*, vol. 13, pp. 10070–10082, 2025, doi: 10.1109/ACCESS.2025.3529206.

[16] A. S. Nour, C. Raymond, D. Zewdneh, and U. Anazodo, "Artificial intelligence-enabled pediatric radiology in low-resource settings: addressing resource constraints in the African healthcare system," *Pediatr. Radiol.*, Jan. 2026, doi: 10.1007/s00247-025-06504-y.

[17] F. Alshanketi *et al.*, "Pneumonia Detection from Chest X-Ray Images Using Deep Learning and Transfer Learning for Imbalanced Datasets," *Journal of Imaging Informatics in Medicine*, vol. 38, no. 4, pp. 2021–2040, Nov. 2024, doi: 10.1007/s10278-024-01334-0.

[18] "Deployable and interpretable deep learning architectures: navigating clinical challenges in medical image segmentation," 2024. doi: 10.14264/98c85ec.

[19] G. Kornaros, "Hardware-Assisted Machine Learning in Resource-Constrained IoT Environments for Security: Review and Future Prospective," *IEEE Access*, vol. 10, pp. 58603–58622, 2022, doi: 10.1109/ACCESS.2022.3179047.

[20] H. Abdel-Jaber, D. Devassy, A. Al Salam, L. Hidaytallah, and M. EL-Amir, "A Review of Deep Learning Algorithms and Their Applications in Healthcare," *Algorithms*, vol. 15, no. 2, p. 71, Feb. 2022, doi: 10.3390/a15020071.

[21] P. Arroba, R. Buyya, R. Cárdenas, J. L. Risco-Martín, and J. M. Moya, "Sustainable edge computing: Challenges and future directions," *Softw. Pract. Exp.*, vol. 54, no. 11, pp. 2272–2296, Nov. 2024, doi: 10.1002/spe.3340.

[22] A. Biglari and W. Tang, "A Review of Embedded Machine Learning Based on Hardware, Application, and Sensing Scheme," *Sensors*, vol. 23, no. 4, p. 2131, Feb. 2023, doi: 10.3390/s23042131.

[23] A. Sivanathan *et al.*, "Real-Time and Trustworthy Classification of IoT Traffic Using Lightweight Deep Learning," *IEEE Trans. Netw. Sci. Eng.*, vol. 13, pp. 3256–3273, 2026, doi: 10.1109/TNSE.2025.3628913.

[24] C. C. Yang, "Explainable Artificial Intelligence for Predictive Modeling in Healthcare," *J. Healthc. Inform. Res.*, vol. 6, no. 2, pp. 228–239, Jun. 2022, doi: 10.1007/s41666-022-00114-1.

[25] J. Kim, "Quantization Robust Pruning With Knowledge Distillation," *IEEE Access*, vol. 11, pp. 26419–26426, 2023, doi: 10.1109/ACCESS.2023.3257864.

[26] Adewale Abayomi Adeniran, Amaka Peace Onebunne, and Paul William, "Explainable AI (XAI) in healthcare: Enhancing trust and transparency in critical decision-making," *World Journal of Advanced Research and Reviews*, vol. 23, no. 3, pp. 2447–2658, Sep. 2024, doi: 10.30574/wjarr.2024.23.3.2936.

[27] E. Hassan, M. S. Hossain, A. Saber, S. Elmougy, A. Ghoneim, and G. Muhammad, "A quantum convolutional network and ResNet (50)-based classification architecture for the MNIST medical dataset," *Biomed. Signal Process. Control*, vol. 87, p. 105560, Jan. 2024, doi: 10.1016/j.bspc.2023.105560.

[28] P. Sharma and V. Sharma, "Classification of COVID-19 Utilizing CT Scan Images Employing the EfficientNet Model," in *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*, IEEE, Sep. 2024, pp. 1–7. doi: 10.1109/ACROSET62108.2024.10743990.

[29] A. Assis, J. Dantas, and E. Andrade, "The performance-interpretability trade-off: a comparative study of machine learning models," *J. Reliab. Intell. Environ.*, vol. 11, no. 1, p. 1, Mar. 2025, doi: 10.1007/s40860-024-00240-0.

[30] M. A. Abdou, "Literature review: efficient deep neural networks techniques for medical image analysis," *Neural Comput. Appl.*, vol. 34, no. 8, pp. 5791–5812, Apr. 2022, doi: 10.1007/s00521-022-06960-9.

[31] E. Ostrowski and M. Shafique, "J-Net: A Low-Resolution Lightweight Neural Network for Semantic Segmentation in the Medical Field for Embedded Deployment," 2025, pp. 480–492. doi: 10.1007/978-3-031-77392-1_36.

[32] M. Karki *et al.*, "Generalization Challenges in Drug-Resistant Tuberculosis Detection from Chest X-rays," *Diagnostics*, vol. 12, no. 1, p. 188, Jan. 2022, doi: 10.3390/diagnostics12010188.

[33] S. Somvanshi *et al.*, "From Tiny Machine Learning to Tiny Deep Learning: A Survey," *ACM Comput. Surv.*, vol. 58, no. 7, pp. 1–33, May 2026, doi: 10.1145/3776588.

[34] F. Ma, Z. Ma, B. Sun, and S. Li, "TA-CNN," in *Proceedings of the 30th ACM International Conference on Multimedia*, New York, NY, USA: ACM, Oct. 2022, pp. 7099–7103. doi: 10.1145/3503161.3551587.

[35] J. Li, H. Sun, and J. Li, "Beyond confusion matrix: learning from multiple annotators with awareness of instance features," *Mach. Learn.*, vol. 112, no. 3, pp. 1053–1075, Mar. 2023, doi: 10.1007/s10994-022-06211-x.

[36] A. Assis, J. Dantas, and E. Andrade, "The performance-interpretability trade-off: a comparative study of machine learning models," *J. Reliab. Intell. Environ.*, vol. 11, no. 1, p. 1, Mar. 2025, doi: 10.1007/s40860-024-00240-0.

[37] S. Riyadi, R. Mulya, and A. Nabila Realisti, "Comparison of MobilenetV2 and EfficiennetB3 Method to Classify Diseases on Corn Leaves," *E3S Web of Conferences*, vol. 595, p. 02006, Nov. 2024, doi: 10.1051/e3sconf/202459502006.

[38] M. Rajkumar, S. R. Siriyala, A. Sharma, and D. N. S. Kumar, "Comparative Analysis of ResNet, EfficientNet, and MobileNetV2 for Detecting Skin Conditions in Deep Learning Models," in *2024 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, IEEE, Dec. 2024, pp. 1487–1492. doi: 10.1109/ICSCNA63714.2024.10864030.

[39] C. Sorino, D. Ceriani, and S. Agati, "Acute respiratory failure in a patient with diffuse ground-glass opacities during the COVID-19 pandemic," in *Rare and Interstitial Lung Diseases*, Elsevier, 2025, pp. 245–256. doi: 10.1016/B978-0-323-93522-7.00030-6.

[40] W. Cardoso, T. Barros, G. Gonçalves, C. Premebida, and U. J. Nunes, "Multispectral Image Segmentation in Agriculture: Evaluating Deep Learning Models with Train-Test Split and Cross-Validation Strategies," in *2024 7th Iberian Robotics Conference (ROBOT)*, IEEE, Nov. 2024, pp. 1–6. doi: 10.1109/ROBOT61475.2024.10797395.

[41] B. Billot, C. Magdamo, Y. Cheng, S. E. Arnold, S. Das, and J. E. Iglesias, "Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain MRI datasets," *Proceedings of the National Academy of Sciences*, vol. 120, no. 9, Feb. 2023, doi: 10.1073/pnas.2216399120.

[42] A. J. Abdi *et al.*, "Standardisation and Optimisation of Chest and Pelvis X-Ray Imaging Protocols Across Multiple Radiography Systems in a Radiology Department," *Diagnostics*, vol. 15, no. 12, p. 1450, Jun. 2025, doi: 10.3390/diagnostics15121450.

[43] M. Sahlabadi, R. C. Muniyandi, Z. Shukur, and F. Qamar, "Lightweight Software Architecture Evaluation for Industry: A Comprehensive Review," *Sensors*, vol. 22, no. 3, p. 1252, Feb. 2022, doi: 10.3390/s22031252.

[44] J. Padhi, L. Korada, A. Dash, P. K. Sethy, S. K. Behera, and A. Nanthaamornphong, "Paddy Leaf Disease Classification Using EfficientNet B4 With Compound Scaling and Swish Activation: A Deep Learning Approach," *IEEE Access*, vol. 12, pp. 126426–126437, 2024, doi: 10.1109/ACCESS.2024.3451557.

[45] S. Mehta and A. Singh, "Harnessing MobileNetV2 for Real-Time Detection of Citrus Diseases in Resource-Constrained Environments," in *2025 International Conference on Computing Technologies (ICOCT)*, IEEE, Jun. 2025, pp. 1–5. doi: 10.1109/ICOCT64433.2025.11118771.

[46] P. Del Moral, S. Nowaczyk, and S. Pashami, "Why Is Multiclass Classification Hard?," *IEEE Access*, vol. 10, pp. 80448–80462, 2022, doi: 10.1109/ACCESS.2022.3192514.

[47] K. Kansal, T. B. Chandra, and A. Singh, "Deep Learning Approaches for Multi-Class Lung Disease Classification Utilizing EfficientNet Architectures and CXR Images," *Procedia Comput. Sci.*, vol. 260, pp. 175–181, 2025, doi: 10.1016/j.procs.2025.03.191.

[48] T. Sanida and M. Dasygenis, "A novel lightweight CNN for chest X-ray-based lung disease identification on heterogeneous embedded system," *Applied Intelligence*, vol. 54, no. 6, pp. 4756–4780, Mar. 2024, doi: 10.1007/s10489-024-05420-2.

[49] C.-T. Yen and C.-Y. Tsao, "Lightweight convolutional neural network for chest X-ray images classification," *Sci. Rep.*, vol. 14, no. 1, p. 29759, Nov. 2024, doi: 10.1038/s41598-024-80826-z.

Supplementary Material

Supplementary material that may be helpful in the review process should be prepared and provided as a separate electronic file. That file can then be transformed into PDF format and submitted along with the manuscript and graphic files to the appropriate editorial office.