TIERS Information Technology Journal
Vol.7, No.1, June 2026, pp. 01~16
P-ISSN: 2723-4533 / E-ISSN: 2723-4541
DOI: https://doi.org/10.38043/tiers.v7i1.7521

Optimized Mobile SE-CNN for Pneumonia Detection Using Chest X-Ray Images

Baiq Anggita Arsya Rahmatin[1], I Gede Pasek Suta Wijaya [2], Ario Yudo Husodo [3], Murizah Kassim [4]

baiqanggitaarsya@gmail.com[1], gpsutawijaya@unram.ac.id[2], ario@unram.ac.id[3], murizah@uitm.edu.my[4]

[1] Master of Information Technology, University of Mataram, NTB, Indonesia
[2,3] Department of Informatics Engineering, University of Mataram, NTB, Indonesia
[4] Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA,Shah Alam, Selangor, Malaysia and Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia.

ABSTRACT

Pneumonia remains one of the leading causes of morbidity and mortality worldwide, particularly in regions with limited access to diagnostic facilities. Chest X-ray (CXR) imaging is widely used for pneumonia detection; however, manual interpretation can be time-consuming and prone to variability among radiologists. This study proposes an optimized Mobile SE-CNN architecture that integrates Mobile Inverted Bottleneck Convolution (MBConv) and Squeeze-and-Excitation (SE) mechanisms to improve feature representation while maintaining computational efficiency. The model was trained and evaluated using the COVID-19 Radiography Database consisting of four classes: COVID-19, Lung Opacity, Viral Pneumonia, and Normal. Experimental results show that the proposed model achieved a test accuracy of 93.58% with a macro-average F1-score of 94.17%. Compared with the baseline CNN model, the proposed architecture improves classification accuracy by 3.69% while reducing the number of parameters by approximately 99.62%, using only 40,606 parameters and a total size of approximately 0.15 MB. These results demonstrate that the proposed Mobile SE-CNN achieves an effective balance between diagnostic performance and computational efficiency, making it suitable for deployment in mobile or embedded medical diagnostic systems.

Keywords: Pneumonia detection; Chest X-ray; Convolutional Neural Network; MBConv; Squeeze-and-Excitation; Deep learning; Medical image classification

Article Info

Received	:	25-01-2026
Revised	:	11-03-2026
Accepted	:	11-05-2026

Correspondence Author:
Baiq Anggita Arsya Rahmatin
Master of Information Technology,
University of Mataram,
Jl. Majapahit No.62, Gomong, Kota Mataram, NTB, 83115, Indonesia.
Email: baiqanggitaarsya@gmail.com

1. INTRODUCTION

Pneumonia continues to pose a significant global health burden and is frequently described as a “forgotten pandemic” due to its persistently high rates of morbidity and mortality, particularly in low- and middle-income countries [1]. Reports from the World Health Organization (WHO) indicate that approximately one hundred children die every hour worldwide as a result of pneumonia-related complications [2]. Consequently, timely and precise diagnosis plays a crucial role in reducing mortality risk. Chest X-ray (CXR) imaging remains one of the most commonly employed diagnostic tools for pneumonia screening because it is non-invasive, relatively fast, and economically accessible. Nevertheless, the interpretation of CXR images largely depends on the experience of radiologists and may lead to variability among observers. In addition, diagnostic performance in terms of sensitivity and specificity can be influenced by disease progression and variations in image quality [3]. These limitations motivate the development of automated and objective diagnostic approaches.

Recent developments in artificial intelligence, particularly in the field of Convolutional Neural Networks (CNN), have substantially advanced the capability of medical image analysis systems. CNN-based approaches enable automatic extraction of hierarchical spatial representations from imaging data, facilitating accurate disease identification. Various studies have demonstrated promising results in pneumonia classification by utilizing transfer learning strategies and deep architectures such as VGG, ResNet, DenseNet, and Inception variants [4], [5]. Despite achieving high predictive performance, many of these models rely on complex network structures with considerable computational demands, which limits their applicability in real-time deployment on mobile or embedded platforms.

To address computational efficiency, several lightweight CNN architectures have recently been proposed for pneumonia and COVID-19 detection tasks. These include parameter-efficient models designed to minimize resource consumption [6], multi-scale architectures for CT scan analysis [7], fine-tuned EfficientNet variants [8], ultralight models augmented with generative adversarial networks [9], and lightweight binary classification models that achieve high accuracy on large-scale datasets [10]. Although these approaches successfully reduce computational complexity, many lack adaptive mechanisms to emphasize diagnostically important features. Consequently, subtle pathological patterns, such as ground-glass opacities in COVID-19 pneumonia, may not be effectively captured. This limitation underscores the importance of developing lightweight architectures that balance efficiency with intelligent feature prioritization [11].

In our earlier study, a lightweight CNN model optimized using RMSprop and Stochastic Gradient Descent (SGD) achieved competitive classification performance, with SGD reaching an accuracy of 93.19%. However, that investigation primarily focused on comparing optimization algorithms rather than improving architectural efficiency. Furthermore, the experimental setup involved a relatively limited dataset and employed data duplication techniques to address class imbalance, which could potentially introduce evaluation bias. The model architecture was also not specifically tailored for deployment on resource-constrained devices. Therefore, the present research does not simply replicate the previous approach but instead redesigns the network architecture by integrating Mobile Inverted Bottleneck Convolution (MBConv) and Squeeze-and-Excitation (SE) modules. This redesign aims to produce a more efficient and scalable CNN framework capable of maintaining strong classification performance while reducing computational requirements and model size.

This research proposes an enhanced Mobile SE-CNN architecture that combines MBConv blocks with channel attention through SE modules. MBConv improves efficiency using depthwise separable convolution and expansion–projection operations, while SE adaptively recalibrates channel features to emphasize relevant information. Unlike previous lightweight approaches that mainly focus on parameter reduction or multi-scale extraction, the proposed model incorporates adaptive attention to prioritize clinically significant patterns. By integrating spatial feature learning, computational efficiency, and channel-wise recalibration, the proposed architecture aims to achieve accurate, lightweight, and scalable pneumonia detection suitable for real-time mobile and embedded healthcare applications.

2. RESEARCH METHOD

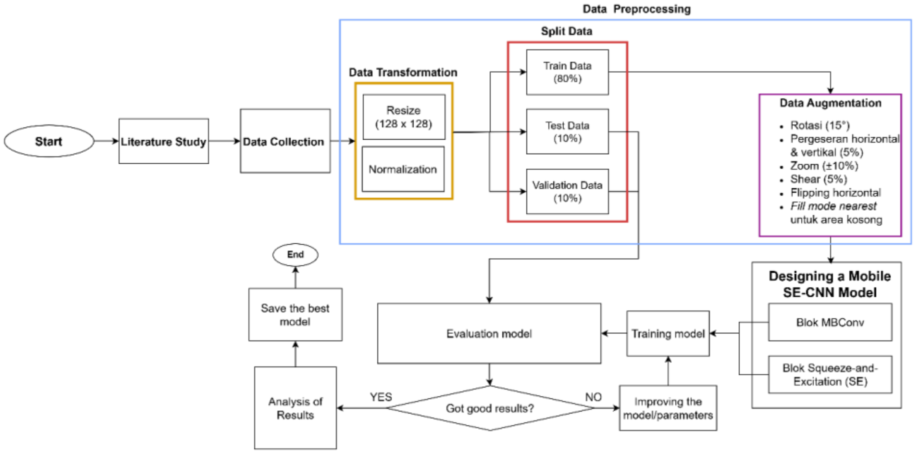

Figure 1. Research Flow

The research workflow illustrated in Figure 1 describes a structured pipeline covering: data acquisition from the COVID-19 Radiography Database, preprocessing (resizing to 128×128, normalization, augmentation), dataset splitting (80/10/10), Mobile SE-CNN architecture design integrating MBConv and SE modules, iterative training with class-weighted loss, and evaluation using accuracy, precision, recall, and F1-score.

.

2.1. Experimental Environment

All experiments were conducted using Python 3.10 and the TensorFlow 2.15 deep learning framework. Model training was performed in a cloud-based environment equipped with an NVIDIA GeForce RTX 3060 GPU with 12 GB VRAM. Supporting libraries, including NumPy, Scikit-learn, and Matplotlib, were utilized for preprocessing, model evaluation, and visualization.

2.2. Data Collection

The dataset used in this study was obtained from the publicly available COVID-19 Radiography Database [12] (https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database?select=COVID-19_Radiography_Dataset), consisting of 21,165 chest X-ray images divided into four classes: COVID-19 (3,616 images), Viral Pneumonia (1,345 images), Lung Opacity (non-COVID Pneumonia) (6,012 images), and Normal (10,192 images). The inclusion of multiple pneumonia-related categories enables a comprehensive evaluation of the model's capability to distinguish between normal lung conditions and different pathological patterns. The dataset exhibits significant class imbalance, with Normal cases comprising 48.2% of the total data while Viral Pneumonia accounts for only 6.4%. To address this imbalance, three complementary strategies were employed. First, class weights inversely proportional to class frequency were applied to the categorical cross-entropy loss function during training, increasing the penalization for misclassification of minority classes. Second, stratified random sampling was used for all data splits to ensure each subset maintains the same class distribution as the original dataset. Third, macro-average precision, recall, and F1-score were reported as primary metrics, assigning equal weight to each class and providing a balanced performance assessment independent of class frequency. These measures collectively ensure that the reported macro-average F1-score of 94.17% reflects robust performance across all classes, including the underrepresented Viral Pneumonia category.

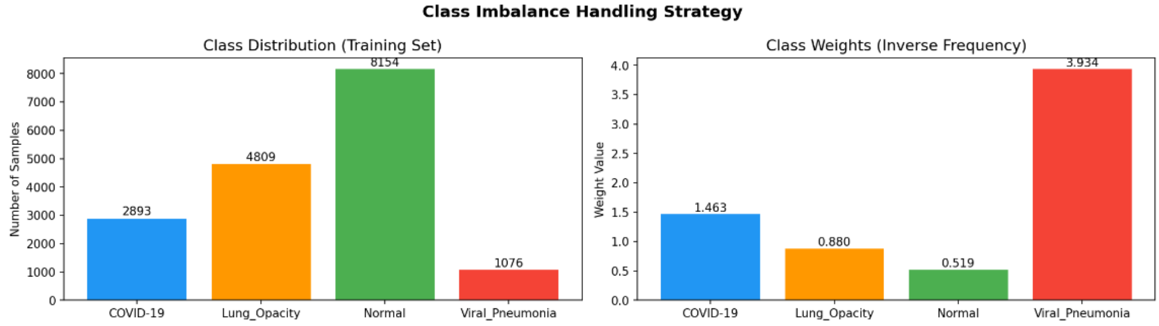

Figure 2. Class Distribution Weights

2.3. Data preprocessing

Before model training, all images were preprocessed to ensure consistency and improve learning performance. The preprocessing steps included:

2.3.1. Resizing

Each image was resized to a dimension of 128 × 128 pixels [13] ensuring uniform input size for the network and reducing computational complexity.

2.3.2. Normalization

Pixel values were scaled to a range of 0–1 by dividing by 255 [14], which accelerates convergence during training and stabilizes gradient updates.

2.4. Data Splitting

The preprocessed dataset was divided into three subsets, 80% for training, 10% for validation, and 10% for testing [15]. This split ensures sufficient data for model learning while preserving portions for unbiased validation and final evaluation. The class distribution for each subset is presented in Table 1.

Table 1. Dataset Distribution Across Splits

Class	Total	Train (80%)	Validation (10%)	Test (10%)
COVID-19	3,616	2,893	362	362
Viral Pneumonia	1,345	1,076	135	134
Lung Opacity	6,012	4,810	602	602
Normal	10,192	8,154	1,019	1,019
Total	**21,165**	**16,933**	**2,116**	**2,117**

To ensure that the reported performance is not dependent on a specific data split, a 5-fold cross-validation experiment was conducted on the entire dataset. In this approach, the dataset is divided into five equal subsets. In each iteration, four subsets are used for training while the remaining subset is used for validation. This process is repeated five times so that each subset is used as validation once. The final performance is reported as the average accuracy across the five folds.

2.5. Data Augmentation

To increase data diversity and enhance model generalization, augmentation techniques were applied exclusively to the training set. These transformations included rotation up to 15°, horizontal and vertical shifts of 5%, zoom variations of ±10%, shear transformation of 5%, and horizontal flipping. The nearest fill mode was used to handle empty regions generated during transformation, while validation and testing datasets remained unchanged to preserve evaluation integrity.

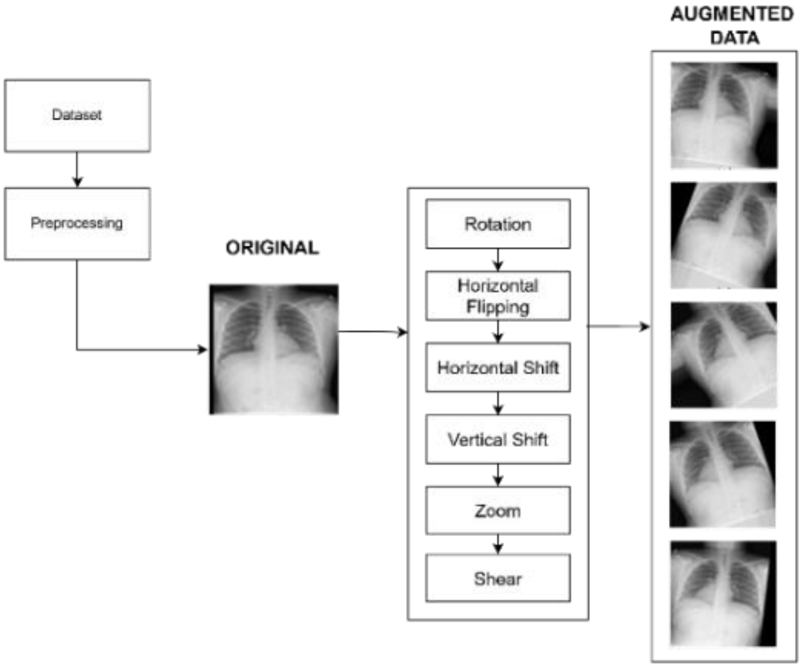

Figure 3. Data Augmentation

2.6. Model Development

The model architecture was designed to achieve a balance between classification performance and computational efficiency by integrating MBConv blocks and SE modules. This combination enables efficient spatial feature extraction and adaptive channel-wise feature recalibration, making the architecture suitable for deployment on mobile and embedded devices.

.

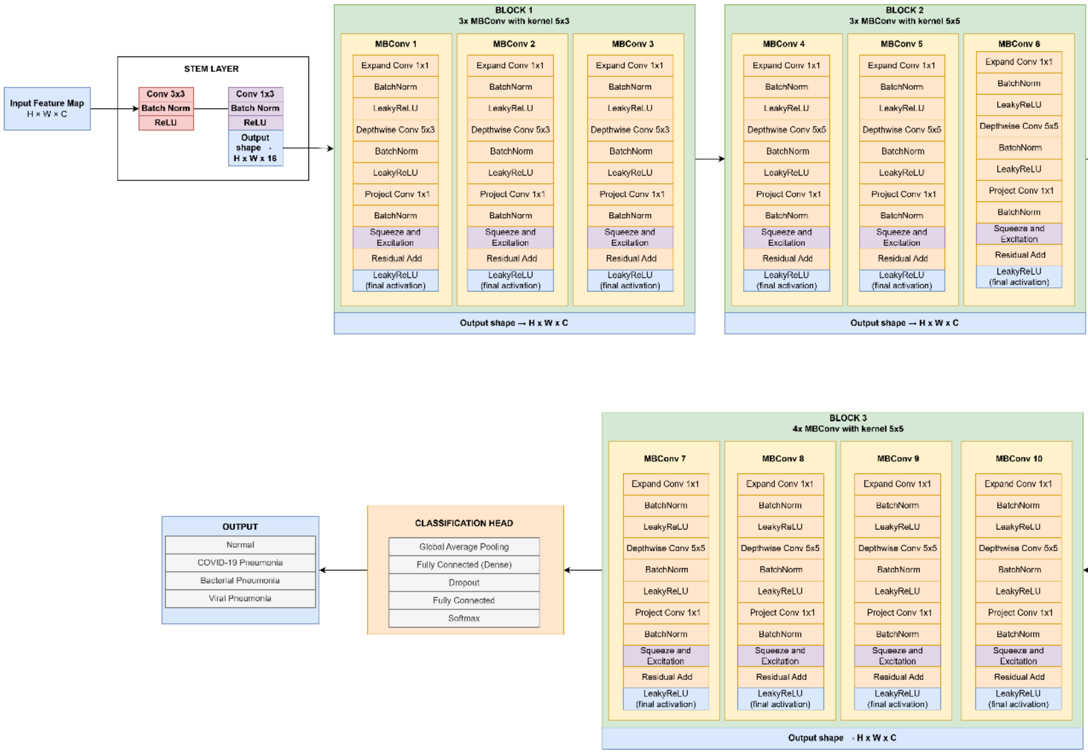

Figure 4. Proposed Model

The proposed Mobile SE-CNN Architecture is designed to maintain high performance while reducing computational complexity through the integration of Mobile Inverted Bottleneck Convolution (MBConv) and Squeeze-and-Excitation (SE) modules. MBConv provides structural efficiency using separate convolutions based on depth, while the SE module adaptively strengthens important feature channels.

Table 2. Detailed Architecture of the Proposed Mobile SE-CNN Model

Layer Type	Kernel Size/ Operation	Output Channels	Activation Function	Additional Components
Input Layer	-	(128 × 128 × 1)	-	Grayscale input
Conv2D	3 × 3	16	ReLU	Batch Normalization
MBConv Block 1	Depthwise 3 × 3 + Pointwise 1 × 1	24	LeakyReLU (0.1)	Expansion ratio 4
MBConv Block 2	Depthwise 3 × 3 + Pointwise 1 × 1	32	LeakyReLU (0.1)	Expansion ratio 4
MBConv Block 3 + SE Module	Depthwise 3 × 3 + Pointwise 1 × 1	64	LeakyReLU (0.1)	Squeeze ratio 0.25
MBConv Block 4 + SE Module	Depthwise 3 × 3 + Pointwise 1 × 1	96	LeakyReLU (0.1)	Squeeze ratio 0.25
Conv2D (Projection)	1 × 1	128	ReLU	Batch Normalization
GlobalAveragePooling2D	-	128	-	-
Dense	-	128	ReLU	Dropout (0.3)
Output (Dense)	-	4	Softmax	Classification Layer

As seen in Table 2, the architecture combines convolutional, bottleneck, and attention layers to achieve a compact yet effective structure suitable for medical image classification.

2.6.1. Expansion Phase

Each MBConv block begins with a 1×1 pointwise convolution that expands the number of feature channels . This expansion phase increases the representational capacity of the network by projecting the input into a higher-dimensional space before spatial filtering is performed. By enlarging the channel dimension, the model is able to capture richer feature interactions and model more complex non-linear relationships [16].

Following the expansion convolution, Batch Normalization is applied to stabilize feature distribution and accelerate convergence. A LeakyReLU activation function (negative slope = 0.1) is then applied to introduce non-linearity while mitigating the dying ReLU problem. Unlike standard ReLU, which produces zero gradients for negative inputs and can permanently deactivate neurons during training, LeakyReLU preserves a small gradient in the negative region, thereby improving gradient flow and enabling more stable convergence in deep network training. This property is particularly beneficial in medical image classification, where subtle pathological features may produce low activation values that standard ReLU would suppress.

2.6.2. Depthwise Convolution

After channel expansion, a depthwise convolution is applied independently to each feature channel using a 3×3 or 5×5 kernel, depending on the block configuration. Unlike standard convolution, which simultaneously filters and combines channels, depthwise convolution performs spatial filtering per channel, significantly reducing the number of parameters and computational operations [17], [18].

This design maintains the ability to extract meaningful local spatial features such as edges, textures, and opacity patterns in chest X-ray images, while preserving computational efficiency. Batch Normalization and LeakyReLU (negative slope = 0.1) are again applied to ensure stable activation distributions and effective feature learning across all MBConv blocks.

2.6.3. Projection Phase

The projection phase uses another 1×1 convolution to compress the expanded feature channels back to a lower-dimensional representation. This step controls model complexity and ensures that the block remains lightweight [19]. Consistent with the inverted residual principle, Batch Normalization is applied without a non-linear activation function after projection. Omitting the activation prevents distortion of information before residual addition and preserves linear feature relationships.

2.6.4. Squeeze-and-Excitation Block

To enhance feature discrimination, the projected features are passed through a Squeeze-and-Excitation (SE) block. The SE mechanism adaptively recalibrates channel-wise feature responses by modeling interdependencies between channels [20].

First, Global Average Pooling performs the squeeze operation, aggregating spatial information from each channel into a compact descriptor. Next, two fully connected layers perform the excitation step to learn channel importance weights. Finally, the learned weights rescale the original feature maps (scale operation), amplifying informative channels while suppressing less relevant ones [21].

In the context of pneumonia detection, this mechanism improves sensitivity to pathological regions such as consolidations and lung opacities while reducing background noise.

2.6.5. Residual Connection

When the input and output feature dimensions are identical, a residual (skip) connection is applied by adding the input tensor to the block output. This residual learning strategy facilitates gradient propagation, enhances convergence stability, and mitigates performance degradation in deeper architectures. A final LeakyReLU activation is applied before forwarding the features to the next block [22].

2.6.6. Hierarchical Block Design

The Mobile SE-CNN architecture is structured into three progressive stages to enable hierarchical feature learning. The first block employs smaller kernels (3×3) with limited repetitions to capture low-level features such as edges and basic structural patterns, while the second block increases network depth to learn mid-level representations, including texture variations and opacity distributions [23]. The third block utilizes larger kernels (5×5) with deeper repetitions to model high-level pathological patterns associated with pneumonia. This staged design allows the network to progressively construct feature representations from simple to complex patterns while maintaining computational efficiency [24].

2.6.7. Classification Head

After the final MBConv + SE block, the extracted high-level feature maps are processed by a lightweight classification head consisting of Global Average Pooling (GAP), Fully Connected (FC) layers, and Softmax activation.

2.6.8. Global Average Pooling

Global Average Pooling reduces the spatial dimensions of the feature map (H×W×C) into a 1×1×C vector by averaging each channel [25]. Compared to flatten-based fully connected layers, GAP significantly reduces the number of trainable parameters and lowers the risk of overfitting. Moreover, GAP preserves the correspondence between feature channels and class categories, improving generalization performance, especially in lightweight architectures.

2.6.9. Fully Connected Layer

The output of the GAP layer is passed to one or more fully connected layers that learn non-linear mappings between extracted features and target classes. These layers integrate global feature information and refine discriminative representations [26]. Dropout regularization is applied during training to randomly deactivate neurons, preventing co-adaptation and improving robustness against overfitting.

2.6.10. Softmax Activation

The final layer applies Softmax activation to produce a probability distribution over four classes: Normal, Lung Opacity, Viral Pneumonia, and COVID-19 Pneumonia. The class with the highest predicted probability is selected as the final output. This probabilistic interpretation also enables further analysis such as ROC-AUC evaluation and confidence-based assessment.

2.7. Baseline CNN Architecture

As a reference for comparative evaluation, a conventional CNN architecture (hereinafter referred to as the Baseline CNN) was implemented and trained under identical experimental conditions. The Baseline CNN comprises three sequential convolutional blocks, each consisting of a Conv2D layer (filters: 32, 64, and 128 respectively; kernel size: 3×3; activation: ReLU), followed by MaxPooling2D (pool size: 2×2) and Dropout (rate: 0.25). The convolutional output is then flattened and passed to two fully connected Dense layers (256 units and 128 units, both with ReLU activation and Dropout rate of 0.5), concluding with a Softmax output layer for 4-class classification. The Baseline CNN contains a total of 10,743,172 trainable parameters and has a model size of approximately 40.98 MB. The high parameter count stems primarily from the Dense layers following the convolutional blocks, where the 18×18×128 flattened feature map generates over 10.6 million parameters. This architecture was trained using the same dataset, optimizer (Adam, learning rate = 0.0001), loss function (Categorical Cross-Entropy), batch size (32), and augmentation pipeline as the proposed Mobile SE-CNN, ensuring that performance differences reflect architectural design choices rather than experimental conditions.

2.8. Training Process

The Mobile SE-CNN model was trained using the training dataset, with hyperparameters adjusted iteratively to achieve optimal performance. Techniques such as learning rate adjustment and batch optimization were employed to enhance convergence. The training process was monitored through validation accuracy and loss metrics to prevent overfitting.

The training parameters for the proposed Mobile SE-CNN models are summarized in Table 3.

Table 3. Mobile SE-CNN Model Training Parameters

Component	Parameter	Configuration
Input Data	Image Size	128 × 128 × 1
	Data Type	Chest X-ray images
Optimizer	Optimizer	Adam
	Learning Rate	0.0001
Loss Function	Loss Function	Categorical Cross-Entropy
Activation Function	Hidden Layer	LeakyReLU
	Output Layer	Softmax
Regularization	Dropout	0.3
Training Strategy	Batch Size	32
Maximum Epochs	Maximum Epoch	100
	Early Stopping	Yes
	Patience	10 Epochs
Dataset Split	Train : Validation : Test	80% : 10% : 10%
Data Augmentation	Augmentation	Rotation 15°, Zoom ±10%, Shift 0.05, Shear 5%, Horizontal Flip

2.9. Model Evaluation

2.9.1. Classification Metrics

The proposed Mobile SE-CNN was evaluated using Accuracy, Precision, Recall, and F1-Score. Since the task involves four classes (Normal, Lung Opacity, Viral Pneumonia, and COVID-19 Pneumonia), macro-averaged Precision, Recall, and F1-Score were reported to ensure balanced evaluation across classes.

A confusion matrix was generated to analyze misclassification patterns. In addition, ROC curves with a One-vs-Rest approach were used to compute AUC values for each class. Training and validation accuracy/loss curves were monitored to assess convergence and overfitting.

2.9.2. Efficiency Metrics

Model efficiency was evaluated through the number of trainable parameters, model size, and inference latency. Inference time was measured per sample (batch size = 1) on the test set using the same NVIDIA GeForce RTX 3060 GPU (12 GB VRAM) with TensorFlow 2.15, simulating real-world single-image clinical deployment. All benchmark models were evaluated under identical hardware and software conditions to ensure fair comparison. This evaluation ensures that the proposed architecture achieves competitive classification performance while maintaining low computational complexity suitable for lightweight deployment.

2.10. Model Optimization and Improvement

Performance thresholds were defined a priori: minimum validation accuracy of 85%, macro-average F1-score ≥85%, and test loss <0.5, based on benchmarks from comparable lightweight CNN architectures. Refinements including hyperparameter adjustment, block reconfiguration, and additional regularization were applied iteratively until all criteria were satisfied or after 10 tuning cycles.

2.11. Result Analysis and Conclusion

The final stage involved analyzing the model's performance based on the evaluation metrics and visualization results, including confusion matrices and accuracy-loss graphs. The findings were discussed to highlight the model's strengths, limitations, and potential future improvements. The best model was then stored for future deployment or conversion to other formats such as TensorFlow Lite.

3. RESULTS AND DISCUSSION

3.1. Model Training Result

The training of the proposed Mobile SE-CNN model, which integrates Mobile Inverted Bottleneck Convolution (MBConv) and Squeeze-and-Excitation (SE) blocks, was conducted using the COVID-19 Radiography Database from Kaggle, consisting of four classes: COVID-19, Lung Opacity, Normal, and Viral Pneumonia. The total number of parameters in the proposed model is 40,606, as summarized in Table 4. This relatively small number shows that the model is lightweight and computationally efficient, making it suitable for medical image classification tasks that require low memory consumption and fast inference.

During the training process, the model demonstrated stable convergence and effective learning behavior. The final model achieved a test accuracy of 93.58% with a test loss of 0.2128, indicating that the model generalizes well without significant overfitting.

Table 4. Parameters of the Proposed Mobile SE-CNN Model

Parameter Type	Count	Memory Size
Total Parameter	40,606	158.62 KB
Trainable Parameters	38,222	149.30 KB
Non-trainable Parameters	2,384	9.31 KB

The small parameter size (only 0.15 MB) signifies that the model is highly efficient compared to conventional CNN architectures that often exceed 100,000 parameters. This efficiency makes the Mobile SE-CNN model easier to deploy in environments with limited computational resources, such as mobile health diagnostic tools or embedded medical systems.

3.2. 5-Fold Cross-Validation

Table 5. The 5-Fold Cross-Validation Result

Fold	Accuracy
Fold 1	89.06%
Fold 2	92.83%
Fold 3	91.36%
Fold 4	91.00%
Fold 5	91.02%
Mean ± Std	**91.05% ± 1.21%**

.

The proposed Mobile SE-CNN achieved a mean accuracy of 91.05% with a standard deviation of 1.21% across the five folds. The relatively small standard deviation indicates that the model demonstrates stable performance across different data partitions. The proposed Mobile SE-CNN achieved a mean accuracy of 91.05% with a standard deviation of 1.21% across the five folds, with a 95% confidence interval of [88.69%, 93.42%]. The relatively small gap of 2.53 percentage points between the hold-out test accuracy (93.58%) and the 5-fold CV mean accuracy (91.05%) indicates good generalization consistency. We explicitly confirm the absence of data leakage: the hold-out test set was partitioned prior to all preprocessing and augmentation steps and was accessed exclusively for final performance reporting. The cross-validation mean accuracy of 91.05% ± 1.21% provides a robust estimate of the model's generalization capability, while the hold-out accuracy of 93.58% confirms strong performance on the independent test set.

3.3. Ablation Study

Table 6. Ablation Study

Scenario	Architecture	Accuracy	Params	FLOPs
S1	–SE +MBConv +GAP	79.12%	19,236	297M
S2	+SE –GAP +MBConv	48.13%	1,500,482	303M
S3	+SE +GAP –Depthwise	86.02%	88,962	1.10B
S4	–SE +GAP –Depthwise	88.47%	85,348	1.10B
S5	–SE –GAP +Depthwise	48.13%	1,496,868	300M
S6	+SE –GAP –Depthwise	48.13%	1,566,594	1.11B
S7	Light CNN	28.43%	109,313	376M
S8	Full Mobile SE-CNN	93.58%	40,606	89M

Table 6 presents the results of the ablation study conducted to analyze the contribution of each architectural component in the proposed Mobile SE-CNN model. Eight experimental scenarios were evaluated by selectively removing or modifying key components including the SE-Block, Global Average Pooling (GAP), and depthwise separable convolution within the MBConv structure.

The results indicate that the removal of Global Average Pooling significantly degrades model performance, as observed in Scenarios S2, S5, and S6 where the classification accuracy dropped to approximately 48%. This demonstrates that GAP plays a crucial role in aggregating spatial information before the classification layer.

Furthermore, replacing depthwise separable convolution with standard convolution (Scenarios S3 and S4) increases the computational cost substantially, raising the FLOPs from approximately 300 million to over 1.1 billion operations. Although these configurations achieve slightly higher accuracy, they significantly reduce computational efficiency.

The full Mobile SE-CNN model (S8) achieves a balanced trade-off between accuracy and efficiency, reaching 93.58% classification accuracy with only 40,606 parameters and approximately 89 million FLOPs. This demonstrates that the combination of MBConv, SE-Block, and GAP provides an effective architecture for efficient medical image classification.

3.4. Model Evaluation

The proposed model achieved high and consistent performance across all classes, as shown in Table 7. The overall test accuracy reached 93.58%, with macro-average precision of 92.97%, recall of 95.51%, and F1-score of 94.17%. These results indicate that the model not only performs well globally but also maintains balanced detection capability across all categories.

Table 7. Classification Performance of Proposed Model

Class	Precision	Recall	F1-Score	Support
COVID-19	0.9491	0.9779	0.9633	362
Lung Opacity	0.9039	0.9219	0.9128	602
Normal	0.9542	0.9205	0.9371	1019
Viral Pneumonia	0.9116	1.0000	0.9537	134
Accuracy			**0.9358**	2117
Macro Average	**0.9297**	**0.9551**	**0.9417**	
Weighted Average	**0.9358**	**0.9358**	**0.9357**	

The model performs particularly well in identifying COVID-19 and Viral Pneumonia (F1-scores exceeding 95%). In contrast, the Normal and Lung Opacity categories exhibit relatively lower recall values of 92.05% and 92.19%, respectively. This performance difference may be attributed to the subtle and variable radiographic characteristics of Lung Opacity cases, which can overlap with normal anatomical structures and make them more difficult to distinguish from adjacent classes.

Previous studies [27] have shown that non-COVID pneumonia presents a wide range of lesion patterns on chest X-ray, varying from lobar to segmental and even patchy opacities that often overlap with normal vascular and bronchial structures. These studies found that only the shape of the lesion demonstrated significant differences, with no notable variation in lesion area, distribution, or zonation between COVID-19 and non-COVID pneumonia. Such morphological diversity makes it difficult to accurately differentiate Lung Opacity cases, as the visual patterns are not consistent enough to establish strong diagnostic cues, which leads to a higher rate of missed detection.

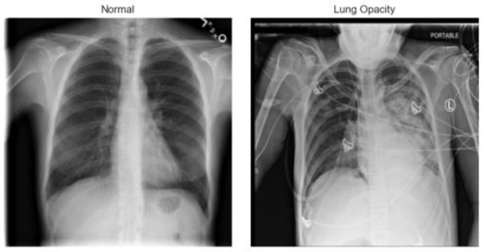

Figure 5. Normal and Lung Opacity

As illustrated in Figure 5, the Lung Opacity image shows diffuse and irregular hazy densities in both lower lung fields that obscure the underlying vascular markings and produce non-uniform opacification. In contrast, the Normal chest X-ray shown in Figure 5 presents clear lung fields with visible bronchovascular structures and sharply defined diaphragmatic borders. The visual similarity between mild Lung Opacity patterns and normal lung appearances may contribute to occasional misclassification, especially in cases with low-contrast or limited abnormal regions.

3.4.1. Accuracy and Loss Graph

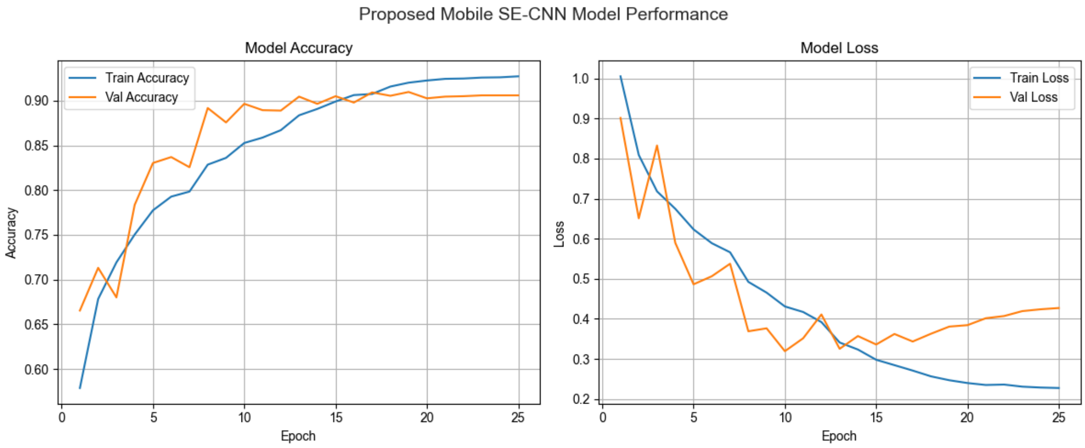

Figure 6. Accuracy and Loss Graph

Figure 6 presents the training and validation accuracy and loss curves of the proposed Mobile SE-CNN model over 25 epochs. The accuracy graph shows a rapid increase during the initial training phase (epochs 1–5), indicating effective feature learning from chest X-ray images, followed by gradual stabilization above 0.90 for both training and validation accuracy. The close alignment between the two curves throughout training suggests strong generalization capability without significant overfitting. Similarly, the loss curves demonstrate consistent optimization behavior, where training loss decreases steadily from approximately 1.0 to around 0.23, while validation loss declines sharply in the early epochs and reaches its minimum around epochs 10–12 before exhibiting a slight upward trend. This mild divergence in later epochs indicates the onset of limited overfitting; however, the gap remains moderate, confirming stable convergence and a balanced learning process supported by regularization and data augmentation strategies.

.

3.5. Confusion Matrix Analysis

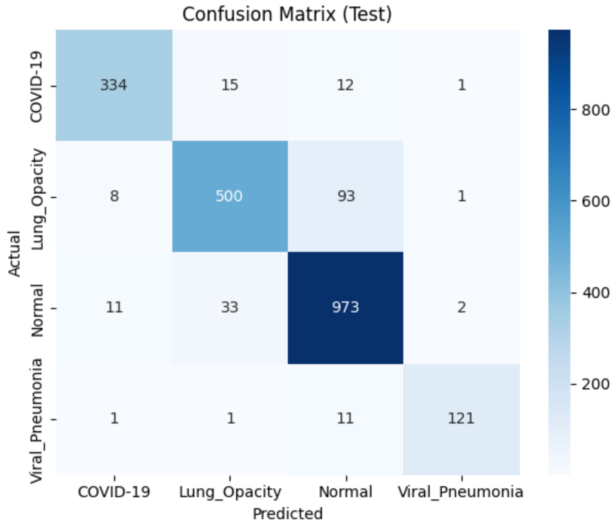

Figure 7. Confusion Matrix

The confusion matrix in Figure 7 illustrates the distribution of predicted and actual labels on the test dataset. Most predictions fall correctly along the diagonal line, demonstrating that the model effectively distinguishes between the four chest X-ray categories.

The model shows a small number of misclassifications, primarily between Lung Opacity and Normal categories, which is understandable since these two classes share similar radiographic features. However, the strong diagonal dominance proves that the Mobile SE-CNN model has successfully learned discriminative representations for each pulmonary condition.

3.6. Misclassified Sample

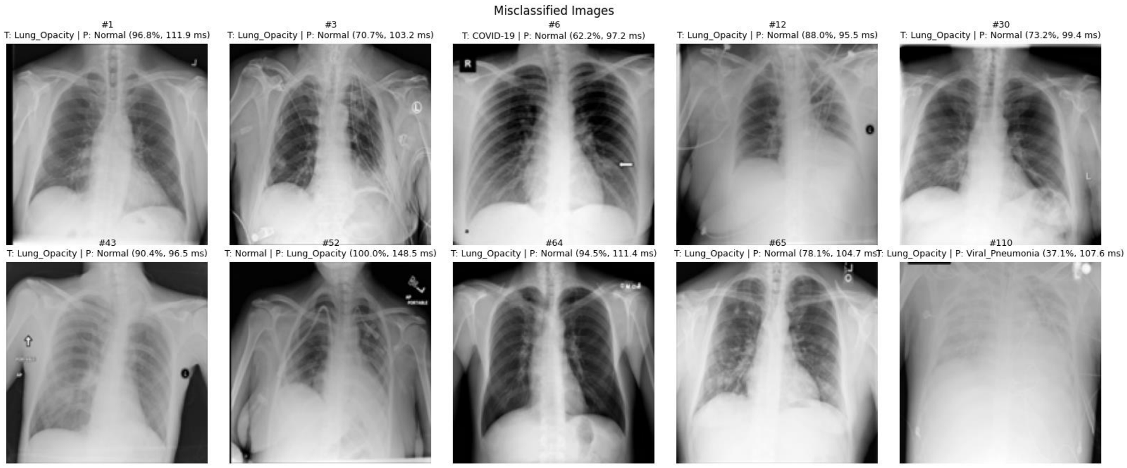

Figure 8. Misclassified Sample

Figure 8 shows several misclassified chest X-ray samples from the test set, where most errors occur between the Lung Opacity and Normal classes. This indicates that subtle and heterogeneous opacity patterns often resemble normal lung structures, making discrimination challenging. Some confusion is also observed between Lung Opacity and Viral Pneumonia due to overlapping radiographic characteristics. The relatively high confidence in certain incorrect predictions suggests intrinsic visual ambiguity rather than unstable model behavior. Overall, these findings confirm that Lung Opacity is the most challenging class, while the limited number of errors demonstrates that the proposed Mobile SE-CNN maintains strong classification performance with efficient architecture.

3.7. Prediction Visualization

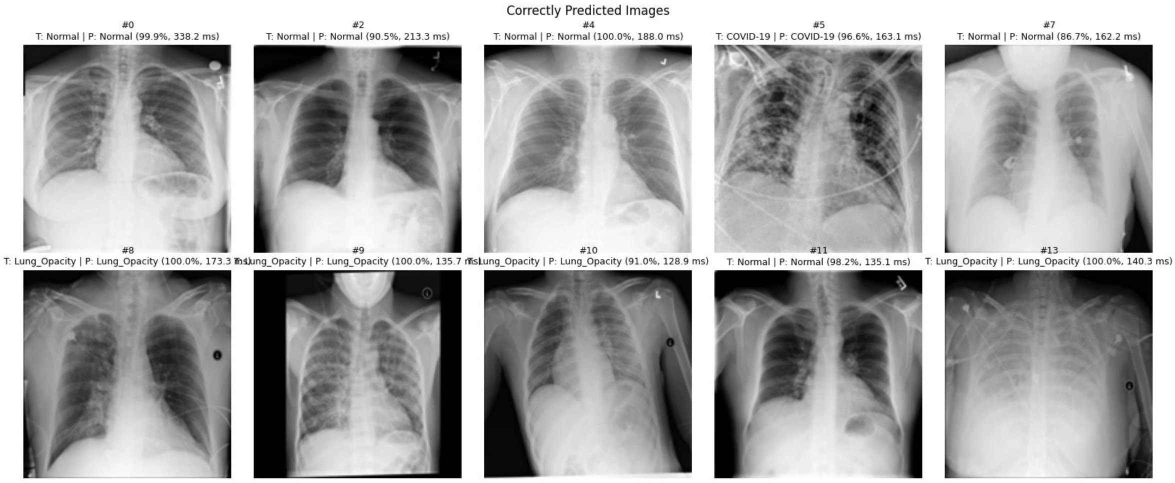

Figure 9. Correctly Predicted

Figure 9 presents several correctly classified chest X-ray images from the test dataset. The model demonstrates high confidence in distinguishing Normal, Lung Opacity, and COVID-19 cases, with probability scores generally exceeding 90%. Clear lung fields with well-defined bronchovascular structures are accurately identified as Normal, while diffuse opacities and infiltrative patterns are correctly recognized as pathological conditions. The consistent alignment between true and predicted labels indicates that the proposed Mobile SE-CNN effectively captures discriminative radiographic features across different classes. These results further confirm the model's strong generalization capability and its ability to maintain reliable performance despite its lightweight architectural design.

3.8. Grad-CAM Visualization

To assess the interpretability of the proposed model, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to representative samples from each class in the test set. Grad-CAM generates heatmaps by computing the gradient of the class score with respect to the final convolutional feature maps, highlighting the spatial regions most influential to the model's decision. The activation maps demonstrate that the model consistently focuses on anatomically and pathologically relevant regions. For COVID-19 cases, high-intensity activations are predominantly located in peripheral bilateral regions corresponding to ground-glass opacities. For Lung Opacity cases, activations concentrate in lower-lobe areas with diffuse haziness. For Viral Pneumonia, the model highlights interstitial and perihilar regions. In Normal cases, activations are diffusely distributed along the tracheal and mediastinal regions rather than concentrating on peripheral lung parenchyma, consistent with the absence of peripheral pathological patterns. This indicates that the model correctly avoids attributing pathological significance to normal lung tissue. These results demonstrate that the Mobile SE-CNN produces clinically interpretable predictions, supporting its reliability for medical diagnostic applications and addressing the concern that deep learning models may function as uninterpretable "black boxes" in clinical settings.

.

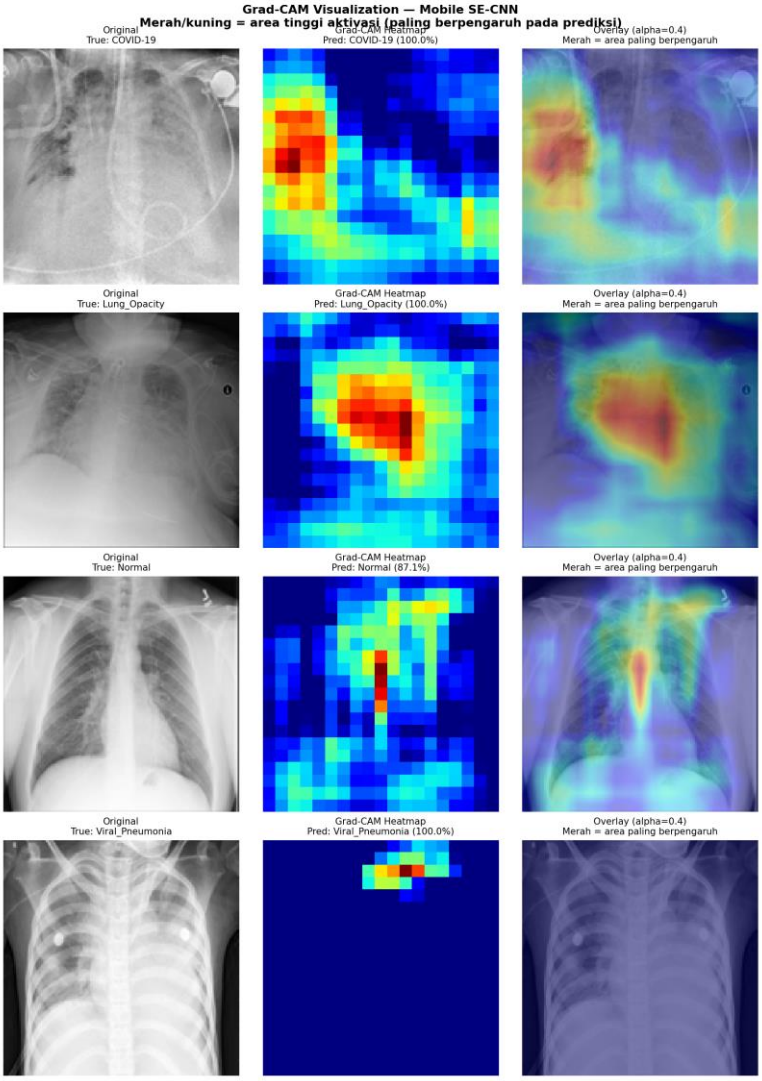

Figure 10. GradCAM Visualization

3.9. Discussion of Model Performance

The synergy between MBConv and SE mechanisms is central to the model's performance: MBConv efficiently compresses and expands feature channels while SE adaptively recalibrates channel weights, enabling the network to focus on diagnostically relevant features. This combination achieves strong classification accuracy with only 40,606 parameters (~0.15 MB), making it practical for resource-limited clinical environments.

3.10. Comparative Performance Analysis of Evaluated CNN Architectures

Table 8 presents the performance comparison between the previous CNN model and the proposed Mobile SE-CNN model, along with additional benchmark architectures including MobileNetV2, EfficientNetB0, and SqueezeNet.

All models were trained and evaluated on the same dataset under identical experimental settings to ensure fair comparison.

Table 8. Comparative Performance Analysis of Evaluated CNN Architectures (All models evaluated on NVIDIA GeForce RTX 3060 GPU, 12 GB VRAM; TensorFlow 2.15; batch size = 1 per image)

Model	Parameters	FLOPs	Model Size (MB)	Inference Time (ms/image)	Test Accuracy (%)	Test Loss
Proposed Mobile SE-CNN	40,606	89,360,412	0.15	68.30	93.58	0.2128
Baseline CNN Model	10,743,172	537,145,828	40.98	62.25	89.89	0.3071
MobileNetV2	2,265,670	299,414,756	9.05	23.23	79.66	0.6792
EfficientNetB0	4,057,257	395,006,475	27.41	23.23	81.34	1.9583
SqueezeNet	121,958	227,836,484	1.50	22.19	83.32	1.0905

The baseline CNN model contains 10,743,172 parameters and requires approximately 537 million FLOPs. It achieved a test accuracy of 89.89%, and exhibited a test loss of 0.3071, indicating slightly less stable confidence calibration. Additionally, the model size (40.98 MB) is substantially larger than that of the proposed architecture. The large parameter count stems primarily from the fully connected layers following the convolutional blocks, which connect a high-dimensional flattened feature map to Dense units without the channel compression mechanism present in MBConv.

In contrast, the proposed Mobile SE-CNN significantly reduces model complexity, using only 40,606 parameters and 89 million FLOPs. This corresponds to a 99.62% reduction in parameters and an 83% reduction in computational cost compared to the baseline CNN model. Despite this reduction, the proposed model achieves a higher test accuracy of 93.58% and a substantially lower test loss (0.2128), suggesting improved predictive stability. The model size is reduced to approximately 0.15 MB, making it considerably more suitable for deployment in memory-constrained environments. Although the inference time per image is slightly higher (68.30 ms vs. 62.25 ms), this difference is negligible in clinical deployment scenarios and is outweighed by the dramatic reduction in model size and parameter count. The proposed Mobile SE-CNN improves classification accuracy by 3.69% compared with the baseline CNN model (89.89%), while simultaneously reducing the number of parameters by approximately 99.62% and lowering computational cost from 537M FLOPs to 89M FLOPs.

When compared to other lightweight architectures, such as MobileNetV2 and SqueezeNet, the proposed model maintains competitive or superior classification performance while requiring fewer parameters and lower computational cost. EfficientNetB0, although powerful in general-purpose image classification tasks, requires substantially higher computational resources and model size in this configuration.

Overall, the results demonstrate that the proposed Mobile SE-CNN achieves a more favorable balance between accuracy and computational efficiency than the previous CNN model. Because all models were trained on the same dataset under identical conditions, the observed improvements can be attributed to architectural optimization rather than dataset bias. These findings indicate that integrating depthwise convolution and squeeze-and-excitation mechanisms provides an effective strategy for designing lightweight yet accurate medical image classification models.

4. CONCLUSION

This study proposes an optimized Mobile SE-CNN architecture for pneumonia detection from chest X-ray images by integrating Mobile Inverted Bottleneck Convolution (MBConv) and Squeeze-and-Excitation (SE) mechanisms. The experimental results demonstrate that the proposed architecture achieves a strong balance between classification performance and computational efficiency. The model achieved a test accuracy of 93.58% with a macro-average F1-score of 94.17%, while maintaining an extremely lightweight structure with only 40,606 parameters and a total model size of approximately 0.15 MB.

Compared with the baseline CNN architecture, the proposed Mobile SE-CNN improves classification accuracy by 3.69% (from 89.89% to 93.58%) while reducing the number of parameters by approximately 99.62% and lowering computational cost significantly. The ablation study confirms that MBConv, depthwise separable convolution, and channel-wise attention each an essential role in the model’s efficiency and accuracy.

Although the model demonstrates strong performance across four classes, the Lung Opacity category remains the most challenging due to its subtle and heterogeneous radiographic patterns that often resemble normal lung structures. This finding highlights the inherent difficulty of differentiating non-COVID pneumonia from normal cases using chest X-ray images.

Overall, the proposed Mobile SE-CNN provides an efficient and accurate solution for automated pneumonia detection while maintaining a compact architecture suitable for deployment in mobile or embedded medical diagnostic systems. Future work may explore the integration of explainable AI techniques and validation on multi-center clinical datasets to further improve model interpretability and generalization capability.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Master of Information Technology Program, Faculty of Engineering, University of Mataram, for providing academic guidance and research facilities throughout this study. The authors would also like to thank the provider of the pneumonia dataset, namely Kaggle, for making the COVID-19 Radiography Database publicly available for academic and research purposes. Special appreciation is extended to the Institute for Research and Community Service LPPM, University of Mataram, for financial support under the Professor Research Grant Scheme, which significantly contributed to the completion of this research.

.

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

REFERENCES

[1] K. Thomas, “Global Burden of Pneumonia,” International Journal of Infectious Diseases, vol. 45, p. 1, Apr. 2016, doi: 10.1016/j.ijid.2016.02.027.

[2] World Health Organization (WHO), “Global Health Estimates: Leading Causes of Death and Disability, 2000–2019,” https://www.who.int/news-room/fact-sheets/detail/pneumonia.

[3] M. Ayasrah, R. Al Hrout, and K. Rawahneh, “Quality Assessment of Plain Chest X-Rays of the Expatriates in Jordan: Multi-Centres Study,” Risk Manag. Healthc. Policy, vol. 18, pp. 933–945, Mar. 2025, doi: 10.2147/rmhp.s507002.

[4] H. S. Salve, “Transfer Learning for Medical Imaging: An Empirical Evaluation of CNN Architectures on Chest Radiographs,” medRxiv, preprint, Jan. 08, 2026, doi: 10.64898/2026.01.07.26343591.

[5] D. R. Rochmawati and L. Maryani, “Deep Learning-Based ResNet-50 Transfer Learning Approaches for Pneumonia Detection from Chest X-Ray Images: With and Without Fine-Tuning,” Indonesian Journal of Health Research and Development, vol. 3, no. 3, pp. 146–153, Nov. 2025, doi: 10.58723/ijhrd.v3i3.507.

[6] A. Howard et al., “Searching for MobileNetV3,” in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Oct. 2019, pp. 1314–1324, doi: 10.1109/iccv.2019.00140.

[7] X. Ouyang et al., “Dual-Sampling Attention Network for Diagnosis of COVID-19 from Community Acquired Pneumonia,” IEEE Trans. Med. Imaging, vol. 39, no. 8, pp. 2595–2605, Aug. 2020, doi: 10.1109/tmi.2020.2995508.

[8] H. Nasiri and S. Hasani, “Automated Detection of COVID-19 Cases from Chest X-Ray Images Using Deep Neural Network and XGBoost,” Radiography, vol. 28, no. 3, pp. 732–738, Aug. 2022, doi: 10.1016/j.radi.2022.03.011.

[9] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, “CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved COVID-19 Detection,” IEEE Access, vol. 8, pp. 91916–91923, 2020, doi: 10.1109/access.2020.2994762.

[10] M. Rahimzadeh and A. Attar, “A Modified Deep Convolutional Neural Network for Detecting COVID-19 and Pneumonia from Chest X-Ray Images Based on the Concatenation of Xception and ResNet50V2,” Inform. Med. Unlocked, vol. 19, p. 100360, 2020, doi: 10.1016/j.imu.2020.100360.

[11] C. Sitaula and M. B. Hossain, “Attention-Based VGG-16 Model for COVID-19 Chest X-Ray Image Classification,” Applied Intelligence, vol. 51, no. 5, pp. 2850–2863, May 2021, doi: 10.1007/s10489-020-02055-x.

[12] T. Rahman et al., “Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-Ray,” Applied Sciences (Switzerland), vol. 10, no. 9, 2020, doi: 10.3390/app10093233.

[13] B. Omarov and I. Bazarkulova, “A Deep Learning Model for Pneumonia Detection from X-Ray Images,” Scientific Journal of Astana IT University, pp. 91–103, Jun. 2023, doi: 10.37943/14aswx8861.

[14] M. A. M. Abueed, D. Md Nor, N. Ibrahim, and J.-M. Ogier, “Pneumonia Detection Using Transfer Learning: A Systematic Literature Review,” International Journal of Advanced Computer Science and Applications, vol. 16, no. 2, 2025, doi: 10.14569/IJACSA.2025.01602102.

[15] L. J. Muhammad et al., “Deep Learning Models for Predicting COVID-19 Using Chest X-Ray Images,” 2022, pp. 127–144, doi: 10.1007/978-3-030-75945-2_6.

[16] T. B. Alakus and I. Turkoglu, “Comparison of Deep Learning Approaches to Predict COVID-19 Infection,” Chaos Solitons Fractals, vol. 140, p. 110120, Nov. 2020, doi: 10.1016/j.chaos.2020.110120.

[17] J. L. Gayathri, B. Abraham, M. S. Sujarani, and M. S. Nair, “A Computer-Aided Diagnosis System for the Classification of COVID-19 and Non-COVID-19 Pneumonia on Chest X-Ray Images by Integrating CNN with Sparse Autoencoder and Feed Forward Neural Network,” Comput. Biol. Med., vol. 141, Feb. 2022, doi: 10.1016/j.compbiomed.2021.105134.

[18] J. Shang, K. Zhang, Z. Zhang, C. Li, and H. Liu, “A High-Performance Convolution Block Oriented Accelerator for MBConv-Based CNNs,” Integration, vol. 88, pp. 298–312, Jan. 2023, doi: 10.1016/j.vlsi.2022.10.012.

[19] M. Tan and Q. V. Le, “EfficientNetV2: Smaller Models and Faster Training,” Jun. 2021. [Online]. Available: http://arxiv.org/abs/2104.00298

[20] W. Meng et al., “A Lightweight CNN and Transformer Hybrid Model for Mental Retardation Screening Among Children from Spontaneous Speech,” Comput. Biol. Med., vol. 151, Dec. 2022, doi: 10.1016/j.compbiomed.2022.106281.

[21] Y. Liang, M. Cao, and S. Zhang, “NeuroPred-ResSE: Predicting Neuropeptides by Integrating Residual Block and Squeeze-Excitation Attention Mechanism,” Anal. Biochem., vol. 695, Dec. 2024, doi: 10.1016/j.ab.2024.115648.

[22] W. H. Tarn, C. H. Chong, L. Wang, C. F. Kuo, and J. Chen, “A Lightweight Image Segmentation Network Leveraging Inception and Squeeze-Excitation Modules for Efficient Skin Lesion Analysis,” Eng. Appl. Artif. Intell., vol. 159, Nov. 2025, doi: 10.1016/j.engappai.2025.111541.

[23] S. Burman, N. Pandey, C. Srivastava, and M. R. Prusty, “MBConvNet-CNS: A Lightweight Convolutional Neural Network (CNN) Architecture for Plant Disease Detection,” in 2024 12th International Conference on Intelligent Systems and Embedded Design (ISED), IEEE, Dec. 2024, pp. 1–6, doi: 10.1109/ISED63599.2024.10957590.

[24] S. Kiziloluk, E. Sert, M. Hammad, R. Tadeusiewicz, and P. Pławiak, “EO-CNN: Equilibrium Optimization-Based Hyperparameter Tuning for Enhanced Pneumonia and COVID-19 Detection Using AlexNet and DarkNet19,” Biocybern. Biomed. Eng., vol. 44, no. 3, pp. 635–650, Jul. 2024, doi: 10.1016/j.bbe.2024.06.006.

[25] M. Hayat, “Squeeze & Excitation Joint with Combined Channel and Spatial Attention for Pathology Image Super-Resolution,” Franklin Open, vol. 8, Sep. 2024, doi: 10.1016/j.fraope.2024.100170.

[26] R. Raju and T. M. Thasleema, “BO-CNN: A Deep Learning Framework with Bayesian Hyperparameter Tuning for Nutrient Stress Classification in Piper nigrum,” Franklin Open, vol. 13, Dec. 2025, doi: 10.1016/j.fraope.2025.100444.

[27] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, “Pneumonia Detection Using CNN Based Feature Extraction,” Proceedings of 2019 3rd IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT 2019, 2019, doi: 10.1109/icecct.2019.8869364.

[28] M. Roberts et al., “Common Pitfalls and Recommendations for Using Machine Learning to Detect and Prognosticate for COVID-19 Using Chest Radiographs and CT Scans,” Nat. Mach. Intell., vol. 3, no. 3, pp. 199–217, Mar. 2021, doi: 10.1038/s42256-021-00307-0.

BIOGRAPHIES OF AUTHORS

Baiq Anggita Arsya Rahmatin is currently pursuing a Master's degree in Information Technology at the Faculty of Engineering, University of Mataram, Indonesia. Her academic and research interests include artificial intelligence, computer vision, and information systems. She has actively participated in academic conferences, including The 3rd MIMSE and ICEEIE, where she received the Best Speaker Award for her research presentation. Her current research focuses on the development of efficient deep learning models for medical image classification, reflecting her commitment to advancing intelligent and practical digital solutions. Through her academic involvement and research activities, she aims to contribute to the development of innovative technologies that support real-world applications. She can be contacted at: baiqanggitaarsya@gmail.com

I Gede Pasek Suta Wijaya is a Professor in the Informatics Engineering Study Program at Universitas Mataram, Indonesia. He obtained his Master's degree in Information Engineering from Universitas Gadjah Mada and earned his Ph.D. from Kumamoto University, Japan, in 2001 and 2010, respectively. His research interests include artificial intelligence, pattern recognition, image processing, computer vision, and intelligent systems. He has published numerous scientific articles in reputable international and national journals, with research contributions spanning machine learning–based image classification, deep learning for surveillance systems, smart agriculture applications, and decision support systems. In addition to research and teaching, He is actively involved in academic leadership, journal editorial activities, and community engagement programs. His work focuses on developing practical AI-based solutions that address real-world problems while advancing research and innovation in information technology at both national and international levels. He can be contacted at email: gpsutawijaya@unram.ac.id

Ario Yudo Husodo is an Associate Professor in Informatics Engineering at the Faculty of Engineering, University of Mataram, Indonesia. He received his Doctorate in Computer Science from Universitas Indonesia with a Cum Laude distinction, achieving the highest GPA in his doctoral cohort. His research interests include optimization methods, swarm intelligence, computational intelligence, and intelligent decision-support systems, particularly in addressing complex and multi-agent problems. His doctoral research focused on optimization-based strategies for intelligent drone swarm coordination and resulted in publications in reputable Q1-indexed international journals. He has been actively involved in research and collaborative projects related to optimization-driven intelligent systems, multi-agent coordination, and strategic decision-making models, especially within defense and advanced technology domains. He regularly contributes to international conferences and scientific forums in computational intelligence and artificial intelligence. In this study, he serves as an academic supervisor, providing guidance on the conceptual framework, refinement of the optimization model, and critical evaluation of methodological and analytical aspects to ensure scientific rigor. He also holds professional certifications in Artificial Intelligence and related technologies, reflecting his continued commitment to advancing intelligent systems research. He can be contacted at email: ario@unram.ac.id

Murizah Kassim is an Associate Professor at Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia, serving under the School of Electrical Engineering, College of Engineering. She is also actively involved with the Institute for Big Data Analytics and Artificial Intelligence (IBDAAI) at UiTM. She obtained her Ph.D. in Electronic, Electrical and Systems Engineering from Universiti Kebangsaan Malaysia in 2016. Her research interests include computer networking, Internet of Things (IoT), mobile and web-based systems, artificial intelligence applications, and intelligent data analytics. Dr. Murizah has authored and co-authored numerous Scopus-indexed publications covering areas such as convolutional neural networks for image classification, IoT network performance optimization, smart agriculture systems, augmented reality applications, and energy-efficient network architectures. With extensive academic and technical experience, she contributes actively to research innovation, postgraduate supervision, and international scholarly collaborations in information and communication technology.
She can be contacted at: murizah@uitm.edu.my

.