Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 9 No. 6 (2025) 1291 - 1305 e-ISSN: 2580-0760

The Impact of Squeeze-and-Excitation Blocks on CNN Models and Transfer Learning for Pneumonia Classification Using Chest X-ray Images

Muhammad Yunan[1], Aris Marjuni[2], Affandy[3], Mochamad Arief Soeleman[4], Iqbal Firdaus[5]
[1]Master of Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, and Institut Bisnis dan Teknologi Kalimantan, Banjarmasin, Indonesia
[2,3,4]Master of Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia
[5]Information Technology, Faculty of Computer Science, Institut Bisnis dan Teknologi Kalimantan, Banjarmasin, Indonesia

[1]myunan12@gmail.com, [2]aris.marjuni@dsn.dinus.ac.id, [3]affandy@dsn.dinus.ac.id, [4]m.arief.soeleman@dsn.dinus.ac.id, [5]iqbalfirdaus@ibitek.ac.id

Abstract

Pneumonia is one of the leading causes of death due to respiratory tract infections, especially in children and the elderly. Early detection using chest X-ray images is crucial to accelerate diagnosis and treatment, but manual interpretation is often subjective and error-prone. This study evaluates the effect of Squeeze-and-Excitation (SE) Block integration on the performance of a custom Convolutional Neural Network (CNN) model and three popular transfer learning architectures: MobileNetV2, VGG16, and InceptionV3 in X-ray image-based pneumonia classification. A dataset of 5,856 images, taken from Chest X-ray Images (Pneumonia) on Kaggle, was processed through preprocessing, undersampling, and augmentation. Each model was tested in two configurations: without and with SE Block. Evaluation was performed using accuracy, precision, recall, F1-score, and test loss metrics. The results show that SE Block integration improves the performance of most models. The accuracy of the custom CNN increased from 95.17% to 95.88%, MobileNetV2 from 97.18% to 97.59%, and VGG16 from 96.88% to 97.69%. InceptionV3 also saw an accuracy increase from 94.06% to 94.16%, although accompanied by an increase in test loss. SE Block proved effective in strengthening the model's emphasis on important features through an inter-channel recalibration mechanism, especially on efficient architectures like MobileNetV2 and complex models like VGG16. These findings support the development of a more accurate, efficient, and adaptive deep learning-based pneumonia diagnosis system, especially for implementation in healthcare facilities with limited resources.

Keywords: Chest X-ray imaging; Deep CNN; pneumonia classification; squeeze-and-excitation block; transfer learning

How to Cite: Y. Muhammad, A. Marjuni, A. Affandy, M. A. Soeleman, and I. Firdaus, "The Impact of Squeeze-and-Excitation Blocks on CNN Models and Transfer Learning for Pneumonia Classification Using Chest X-ray Images", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 9, no. 6, pp. 1291 - 1305, Nov. 2025.
Permalink/DOI: https://doi.org/10.29207/resti.v9i6.6693

Received: May 29, 2025
Accepted: October 18, 2025
Available Online: November 29. 2025

This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia

1. Introduction

Pneumonia is a significant acute respiratory infection that contributes substantially to global morbidity and mortality, particularly among children and the elderly. The World Health Organization (WHO) reports that pneumonia is responsible for approximately 14% of all deaths in children under five years of age worldwide, leading to more than 700,000 fatalities each year [1]. In Indonesia, pneumonia continues to be a leading cause of mortality among children under five, according to the 2023 Indonesian Health Profile published by the Ministry of Health [2]. Early detection of pneumonia is crucial to reduce mortality rates and improve clinical outcomes. Chest radiography is a primary diagnostic tool that allows direct visualization of lung structures to identify infection. However, manual interpretation of chest X-ray images depends heavily on the radiologist's expertise and is prone to subjective errors, especially in regions with limited medical personnel and technology [3]. Therefore, the development of Artificial Intelligence based computer-assisted diagnostic systems presents a promising solution to help improve diagnostic accuracy, efficiency, and accessibility to quality healthcare.

In recent years, Convolutional Neural Networks (CNNs) have shown excellent performance in medical image classification, including pneumonia detection from chest X-rays [4]. Their ability to automatically

extract spatial and textural features without manual engineering makes them highly effective for complex classification tasks.

Previous studies have shown that transfer learning with pretrained models like MobileNetV2 [5], VGG16 [6], and InceptionV3 [7], fine-tuned on medical imaging datasets, can be highly effective. These models were chosen because of their proven ability to perform well in image classification tasks. MobileNetV2 offers lightweight design and computational efficiency, which enables its use on devices with limited computational resources, as supported by Gao et al. [8] in lung tumor classification via CT scans. VGG16 is recognized for its deep and stable architecture, capable of extracting detailed spatial features; Sharma and Guleria [9] reported 95.4% accuracy in pneumonia detection using VGG16. InceptionV3, with its ability to capture multiscale features through inception modules and factorized convolutions, has shown strong performance in detecting pneumonia and retinal diseases [10], [11]. The diverse characteristics of these models allow for a comprehensive evaluation of SE Block integration across different CNN structures. Similarly, In their study, Neshat et al. [12] proposed a Hybrid Inception-ResNet design aimed at combining the strengths of both architectures. outperforming seven other CNN models by up to 16.1% in accuracy over EfficientNetV2-Small and 2.6% over InceptionV3. Gu and Lee [3] proposed a real-world feature transfer learning approach with fine-tuned ResNet and DenseNet models on grayscale X-ray images, achieving improved classification accuracy. Ali et al. [13] evaluated six pretrained CNN architectures and found EfficientNetV2L attained the highest accuracy of 94.02%. Tamilkodi et al. [14] compared VGG-16, ResNet-50, and GoogLeNet, confirming that fine-tuning pretrained networks significantly reduces classification errors and accelerates training convergence.

To address generalization limitations of single models, ensemble learning has gained popularity. Mabrouk et al. [15] proposed a probabilistic voting ensemble integrating DenseNet169, MobileNetV2, and Vision Transformer (ViT), achieving around 93.9% accuracy and 93.9% F1-score. surpassing individual models and conventional methods. Despite these successes, most conventional CNNs neglect inter-channel feature dependencies during feature extraction. The SE Block, an attention mechanism that adaptively recalibrates feature responses based on channels, was introduced to address this limitation.

The application of Squeeze-and-Excitation (SE) Blocks in recent years has contributed to significant performance improvements in deep learning models in medical image analysis, including classification and segmentation. Zhou et al. [16] proposed the scSE-NL V-Net, which combines spatial–channel SE and non-local attention, achieving a Dice coefficient between 0.65 and 0.82 on the BraTS 2020 brain tumor dataset. Jiang et al. [17] developed SSA-UNet, a 2.5D U-Net model with SE Blocks and self-attention, which significantly improved brain segmentation accuracy. Similarly, Prasanna et al. [18] incorporated SE Blocks into Attention U-Net (SEEA-UNet), demonstrating better training convergence and segmentation accuracy. Kim et al. [19] introduced E-HRNet, an enhanced HRNet architecture using SE Blocks for urban scene segmentation (Cityscapes and PASCAL Context), showing performance gains without adding significant complexity. These findings support the effectiveness of SE Blocks in adaptively recalibrating feature channels and enhancing spatial representation, making them well-suited for CNN and transfer learning applications in tasks such as pneumonia classification using chest X-ray images.

Huang et al. [20] demonstrated that integrating SE Block into ResNet50V2 increased brain tumor classification accuracy from 92.6% to 98.4% and AUC from 0.987 to 0.999 on MRI data. However, systematic exploration of SE Block in pneumonia classification using chest X-rays remains limited.

Addressing this research gap, this study aims to comprehensively evaluate the impact of integrating SE Blocks into custom CNN architectures and three popular transfer learning models MobileNetV2 [5], VGG16 [6], and InceptionV3 [7] for pneumonia classification from chest X-ray images. The research proceeds in two phases: training baseline models without SE Blocks, followed by training identical architectures with SE Block integration. All datasets undergo preprocessing, including image normalization, class balancing via undersampling, and data augmentation to enhance feature diversity [21], [22]. Model performance was assessed using four main evaluation metrics: accuracy, precision, recall, and F1 score. test loss to provide a thorough analysis of SE Block effectiveness within CNN and transfer learning frameworks [21]–[23].

This study is expected to contribute significantly to the development of deep learning-based diagnostic systems that are accurate, adaptive to critical features in medical images, and resilient in data-limited environments [21], [23]. Given the relatively small dataset (5,856 images), strategies such as undersampling, data augmentation, and transfer learning were employed to optimize model performance. The integration of Squeeze-and-Excitation (SE) Blocks proved effective in enhancing feature representation, even in lightweight models like MobileNetV2 and custom CNNs. These results demonstrate the proposed approach's robustness in scenarios with limited data and resources.

Through a comparative evaluation of model performance with and without SE Block integration, this study provides meaningful insights into the impact of adaptive recalibration within CNN architectures. Custom CNN models were used as controlled baselines to enable precise architectural tuning, such as layer depth, kernel size, and SE Block positioning. As supported by Borvornvitchotikarn and Yooyativong

[24], SE integration in simple CNNs significantly improves multi-label classification, while Roy et al. [23] emphasized the importance of customizable CNNs for optimizing performance with limited medical data.

Although previous research has demonstrated the strong capabilities of pretrained models for pneumonia classification, this study adds value by systematically evaluating the impact of SE Blocks across both custom and transfer learning architectures. Evaluations were conducted under controlled conditions, ensuring objective assessment of SE Blocks' contribution to feature representation. Empirical results indicate consistent improvements in metrics such as accuracy, recall, and F1-score particularly for VGG16 (97.69%) and MobileNetV2 (97.59%).

In terms of adaptability, SE Blocks maintained high performance across varying architectural complexities. However, this study is limited by its focus on a single data type (chest X-rays) and a single algorithm class (CNN). Future research should expand to other medical image modalities and explore alternative architectures such as transformers or hybrid CNN-RNN models to further assess the robustness of this approach. Overall, the findings reinforce the potential of SE-enhanced CNNs in medical image diagnosis and highlight opportunities for advancing robustness through broader testing and architectural diversification.

2. Methods

This study aims to categorize chest X-ray images into pneumonia and normal classes using two main strategies. The first strategy uses a custom Convolutional Neural Network (CNN) enhanced with Squeeze and Excitation Blocks to improve spatial and channel-specific feature extraction. The second strategy uses a transfer learning approach, relying on well-known pre-trained models such as MobileNetV2, VGG16, and InceptionV3, which have demonstrated strong performance in medical image classification [25], [26]. The overall procedure, illustrated in Figure 1, includes several stages: data collection, pre-processing, class balancing, data augmentation, dataset splitting, model training, and model performance evaluation.

2.1 Data Collection

The dataset employed in this study is the Chest X-ray (Pneumonia) dataset, available on the Kaggle platform [27], and consists of 5,856 chest radiographs labeled as pneumonia or normal. The image resolutions vary and reflect a variety of clinical conditions.

As an initial preprocessing step, All chest X-ray images were standardized to 128×128 pixels. using linear interpolation to ensure consistent input dimensions for the CNN model. This resolution was selected to balance the preservation of important image details with processing efficiency. Prior studies have shown that image resizing significantly impacts deep learning model performance and training time [28].

2.2 Research Flowchart

Figure 1 presents the systematic process stages, starting from data collection, image preprocessing, class distribution balancing through undersampling techniques, data augmentation, dataset sharing, model training, and performance evaluation. This study was conducted using two primary configurations for each model architecture, one with and one without SE Block integration, to analyze how the attention mechanism affects the performance of pneumonia classification using X-ray images.

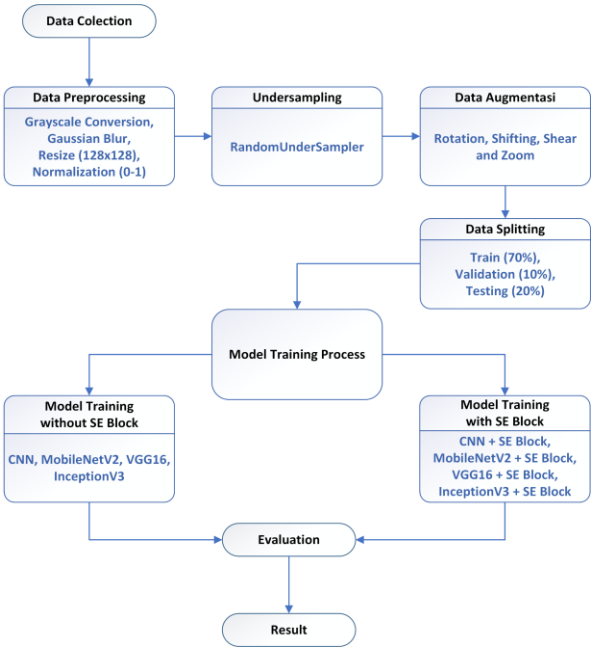


Figure 1. Methodology Diagram: Training of CNN and Transfer Learning Models With and Without SE Block

2.3. Data Preprocessing

Before the model can be trained, the images in the dataset undergo a sequential preprocessing process to ensure format consistency and input quality. The preprocessing steps implemented in this study consist of four main steps: grayscale conversion, Gaussian blur application, image resizing, and pixel value normalization.

In the first stage, each image in the dataset is converted to grayscale to reduce the data dimensionality and eliminate reliance on color information. This process does not sacrifice information important for medical image analysis, such as textural features or patterns relevant for disease detection [29]. Second, Gaussian blur is applied to reduce noise and enhance the sharpness of important features. This method convolves images using a Gaussian function, which averages pixel values within a specified radius, thereby smoothing the image and reducing high-frequency components that could interfere with model training [30]. In the third step, image resizing was carried out to 128×128 pixels to maintain consistent input size across the dataset, a compromise between image detail and processing efficiency [28]. Finally, normalization is performed,

where the pixel values in each image are normalized in the range [0, 1] by dividing each pixel value by 255. This normalization step aims to speed up the training process by reducing the scale variability between features, which can speed up convergence and increase model stability [30].

Figure 2 illustrates the results of each preprocessing stage, from the original image to normalized input.

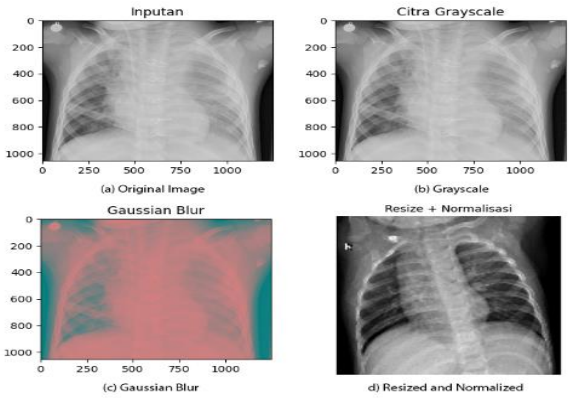


Figure 2. Sample Preprocessing Results: (a) Original Image, (b) Grayscale, (c) Gaussian Blur, (d) Resized and Normalized

The image preprocessing stages are explained in the following sub-figures. (a) The original image is a chest radiograph in RGB format with varying resolutions depending on the dataset. (b) Grayscale converts the image into a single grayscale channel to reduce data dimensionality and computational load, without losing key diagnostic information. (c) Gaussian blur is applied to reduce high-frequency noise so that the lung anatomy pattern becomes clearer and is not disturbed by artifacts. (d) Resizing and normalization are performed by resizing the image to 128×128 pixels for uniformity, then the pixel intensity is Pixel values were scaled to the [0,1] range to speed up model convergence and enhance training stability.

2.4 Data Balancing (Undersampling)

The dataset exhibited a disproportion in the number of samples between pneumonia and normal categories. To mitigate this issue, an undersampling approach was utilized on the majority class (pneumonia) with the help of the RandomUnderSampler tool available in the imbalanced-learn package. This procedure aimed to equalize class representation and minimize potential bias of the model toward the more frequent category. The primary reason for choosing undersampling over oversampling was to avoid data duplication that could lead to overfitting, as the original medical data (minority class) is more valuable in terms of clinical information than the synthetic data resulting from oversampling. Although undersampling can result in some data loss, this technique is offset by data augmentation and model performance evaluation using the F1-Score metric, which considers class imbalance and offers a more balanced representation of the model's performance across both categories.

This technique has been applied in studies involving COVID-19 detection [31], clinical predictions [32], and physiological signal classification [33]. While effective, undersampling must be used cautiously to avoid accuracy degradation or overestimation. Figure 3 shows the class distribution before and after applying undersampling.

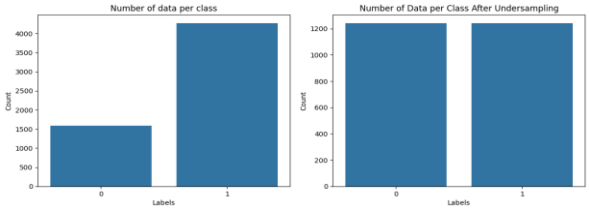


Figure 3. Class Distribution Before and After Undersampling

2.5 Data Augmentation

In this research, the augmentation of image data was carried out with the help of TensorFlow's ImageDataGenerator function, aiming to enhance the variability of training samples and mitigate overfitting issues. Within the field of medical image processing, such augmentation strategies are commonly employed to enrich dataset diversity without generating new raw images, which in turn enhances the model's ability to adapt to previously unseen inputs [34].

In this study, the augmentation process was carried out through several transformations. First, a random rotation was applied with a range of ±3 degrees to help the model recognize small variations in image orientation. Next, horizontal and vertical shifts of up to ±2% of the image dimensions were applied to ensure the model could recognize objects despite varying positions. Furthermore, light shearing and zooming processes were applied to simulate changes in scale and shape without damaging or distorting important anatomical structures in the image.

Figures 4 and 5 depict examples of the images prior to augmentation and once the process has been applied.

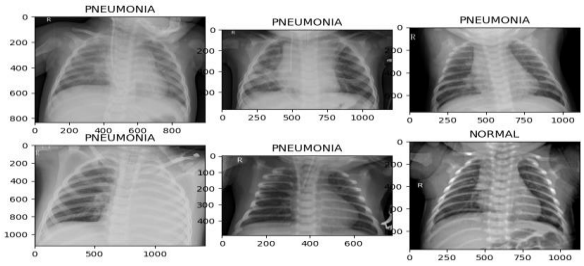


Figure 4. Image before Augmentation process

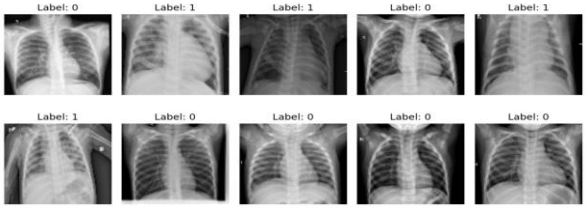


Figure 5. Augmented X-ray Images with Rotation, Shifting, Shearing, and Zoom

2.6 *Dataset Splitting*

After applying data augmentation to the undersampled dataset, a total of 4,968 labeled chest X-ray images was obtained. The dataset was then divided into three subsets following a commonly used ratio in machine learning: 70% for training, 10% for validation, and 20% for testing, corresponding to 3,477, 497, and 994 images, respectively. This partitioning was designed to provide the model with a representative training set while maintaining independent data for validation and testing.

The validation subset was used during training to monitor learning progress and mitigate overfitting when a model fits too closely to training data and fails to generalize. In contrast, the test subset was employed only after training to objectively assess the model's performance on unseen data. This data-splitting strategy is widely adopted in machine learning research to ensure fair, reliable, and unbiased model evaluation [35].

2.7 Model Training

This study employs two primary model development strategies as shown in Figures 6 and 7:

The custom CNN architecture developed in this study consists of multiple convolutional layers, followed by an integrated Squeeze-and-Excitation (SE) block. The SE Block serves to improve the representation of important features by focusing more on relevant feature channels and improving the efficiency of those feature representations [20], [24]. This model was developed in two stages: in the first stage, the custom CNN model was built without the SE Block integration, and in the second stage, the SE Block was added to evaluate the model's performance improvement. The custom CNN model used in this study has a structure that begins with an input layer, which receives a 128×128×3 image after preprocessing. This stage includes image size and intensity normalization, class balancing using undersampling techniques, and augmentation to increase feature diversity. Next, the feature extraction process is carried out via several stacked convolutional layers configured with a 3×3 kernel and a ReLU activation function. Squeeze-and-Excitation (SE) Blocks are applied to several convolutional layers to improve the quality of the learned features by focusing more on relevant feature channels and reducing the influence of unimportant channels. Next, a MaxPooling layer is used to reduce the feature dimensionality and minimize the possibility of overfitting without losing important extracted information. The obtained features are then integrated through a fully connected layer to produce optimal classification results. Finally, an output layer with two neurons and a softmax activation function is used to classify the images into two categories: pneumonia and normal.

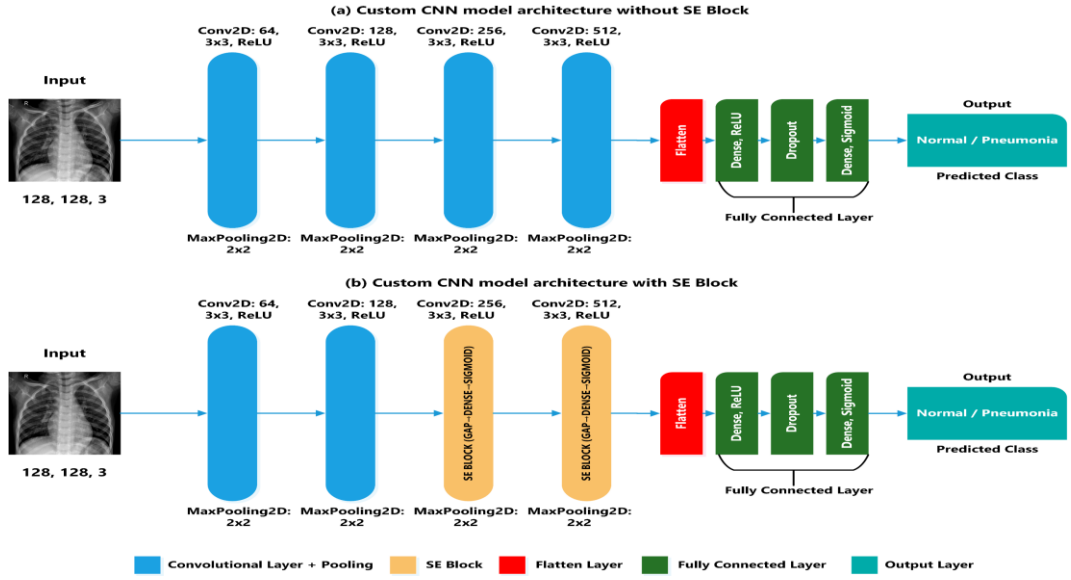
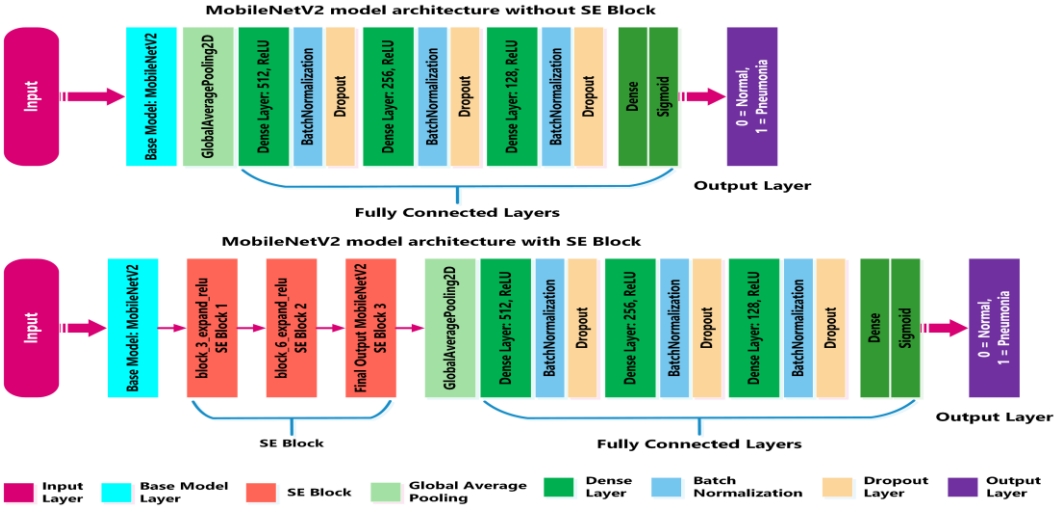
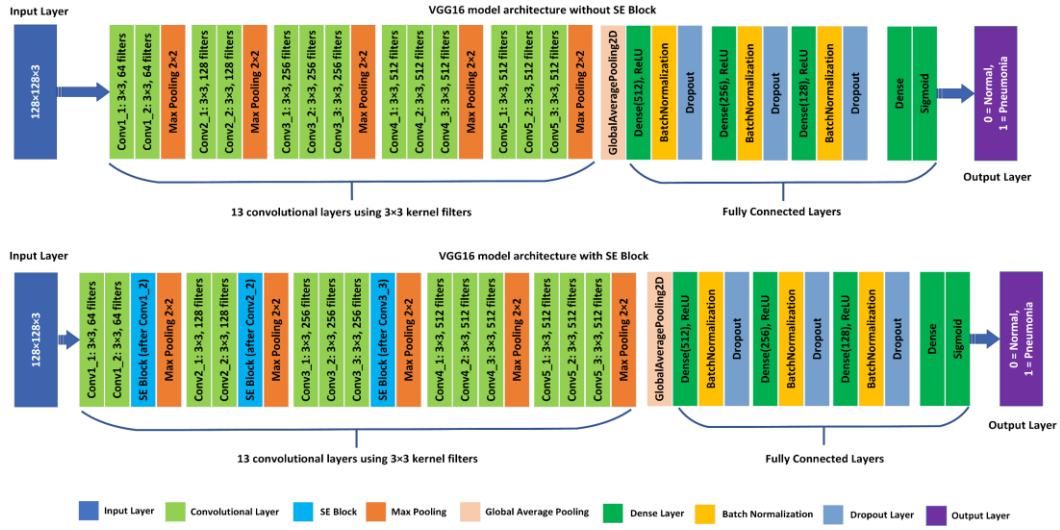


Figure 2. Architecture of the custom CNN model: (a) without the SE Block and (b) with the SE Block.

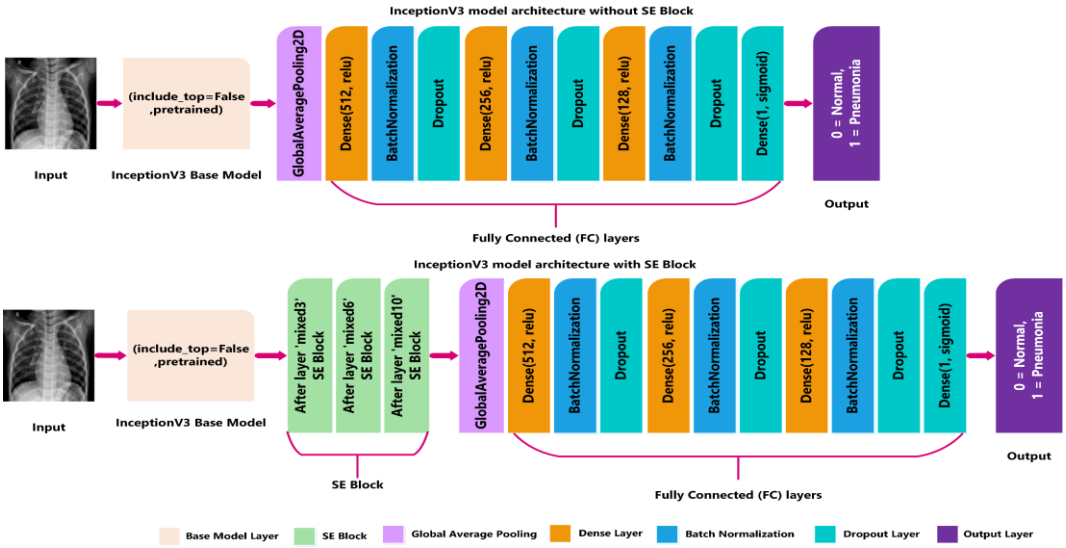
In addition to the custom CNN models, this study utilized a transfer learning approach by adopting pre-trained networks that were previously trained on large-scale datasets such as ImageNet. The selected models MobileNetV2, VGG16, and InceptionV3 each offer unique strengths in computational efficiency, architectural complexity, and feature extraction for medical imaging tasks. MobileNetV2, in particular, is a lightweight CNN architecture optimized for efficiency, making it suitable for applications on devices with limited computational capacity. The model implements depthwise separable convolution and inverted residuals techniques, which significantly minimize the parameter count and computational burden without sacrificing model performance without sacrificing model performance. It has been proven effective in real-time applications and various medical imaging tasks [8], [36], [37].



(a) MobileNetV2 model architecture with and without SE Block



(b) VGG16 model architecture with and without SE Block



(c) InceptionV3 model architecture with and without SE Block

Figure 7. Transfer learning model architectures: (a) MobileNetV2, (b) VGG16, and (c) InceptionV3 each illustrated in two configurations: without and with integrated SE Blocks.

The VGG16 model is a classic CNN architecture with consisting of 13 convolutional layers and 3 fully connected layers, which is recognized for spatial feature extraction using 3×3 recurrent filters and 2×2 max pooling. Despite its high computational demands, this method is still widely used for medical image classification [38], [39]. The InceptionV3 model utilizes the Inception module, which combines 1×1, 3×3, and 5×5 filters to capture multiscale features, along with factored convolutions and additional classifiers to reduce computation and improve training efficiency. It is effective in complex image classification tasks [10], [11]. In this study, pre-trained models were implemented using a feature extraction approach, In this setup, the weights of all convolutional layers were kept fixed, while only the final classification layers were appended and trained. This strategy takes advantage of the deep feature representations acquired from large-scale datasets, like ImageNet without retraining the base model parameters. Each model was evaluated under two configuration: one without SE Block integration, and the other with SE Blocks added to relevant convolutional layers to enhance feature representation.

2.8 Model Evaluation

The model's performance was evaluated using four primary metrics: accuracy, precision, recall, and F1-score. These metrics were applied to assess how well the model distinguishes between pneumonia and normal chest X-ray images. Besides the numerical evaluations, a confusion matrix was also used to visualize the classification outcomes, which include true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This matrix provides insights into error patterns and helps improve model interpretability and refinement [40]. The formula for each evaluation metric is shown in Formulas 1 to 4.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall \text{ (Sensitivy)} = \frac{TP}{TP+FN} \quad (3)$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

In these equations, TP (True Positive) refers to pneumonia images correctly identified by the model, while TN (True Negative) denotes normal images accurately recognized as non-pneumonia. FP (False Positive) indicates normal images incorrectly classified as pneumonia, and FN (False Negative) represents pneumonia cases that the model failed to detect.

These four metrics together provide a comprehensive overview of classification performance. Accuracy measures the overall correctness of predictions, precision reflects the reliability of positive classifications, recall indicates the model's sensitivity in detecting all positive cases, and the F1-score acts as a balanced indicator that integrates both precision and recall.

2.9 Implementation and Testing

After completing the training phase, the model was evaluated using test data that had never been used in training or validation, ensuring an objective assessment of its generalization to new samples. Two custom CNN configurations were examined, one without SE Blocks and one with their integration.

The version without SE Blocks served as a baseline to establish the performance of the custom CNN reference, while the inclusion of SE Blocks served as an attention mechanism to highlight important features. The performance of both configurations was then compared with the commonly used transfer learning models MobileNetV2, VGG16, and InceptionV3 to evaluate the contribution of SE Blocks and the overall effectiveness of the proposed CNN in pneumonia classification from chest X-ray images.

3. Results and Discussions

3.1 Visualization of Model Training Results

Training performance was analyzed by plotting the training loss, validation loss, training accuracy, and validation accuracy curves for each model. Figures 8 through 15 present the learning dynamics and performance trends of the CNN architectures, MobileNetV2, VGG16, and InceptionV3, both in their standard form and with the Squeeze-and-Excitation Block integrated. Each figure is discussed individually to provide a detailed interpretation of the learning progression and convergence behavior of the models during training.

Figure 8 shows the loss and accuracy curves of the CNN model without the Squeeze-and-Excitation Block. The model achieved a final accuracy of 95.17% and a test loss of 0.1236. The plotted curves show a stable learning pattern with only slight variations in the validation loss, indicating that the model successfully captured relevant training features while still leaving room for improvement in its generalization to unseen data.

Figure 9 illustrates the training performance of the CNN architecture after incorporating the SE Block. The inclusion of this module led to an increase in accuracy, reaching 95.88%, along with a decrease in test loss to 0.1169. The obtained training and validation curves show more stable and uniform convergence behavior, indicating that SE Blocks help the network concentrate on more meaningful features, which in turn improves its ability to generalize.

Figure 10 shows the training and validation performance curves of the MobileNetV2 architecture before SE Block integration. The model achieved 97.18% accuracy with a test loss of 0.0966. The stable curves indicate that MobileNetV2, as a lightweight model, is efficient during training and capable of delivering high performance with a relatively small number of parameters. The stability of the curves also

reflects the architecture's ability to learn important patterns without overfitting.

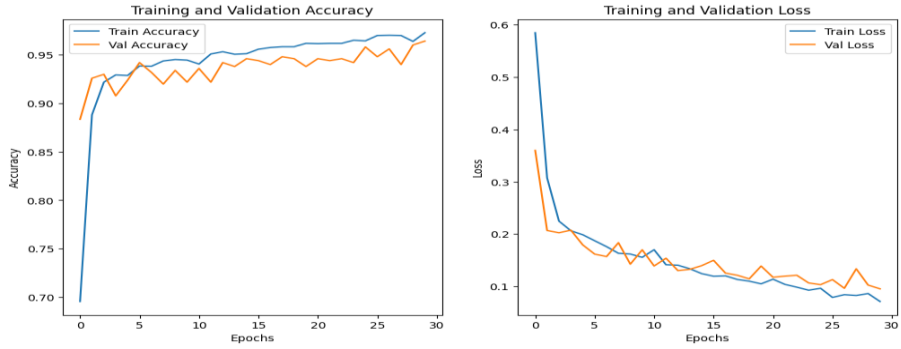


Figure 8. Loss and accuracy curves of the CNN model without SE Block

Figure 11 illustrates the performance of MobileNetV2 after integrating the SE Block. The model shows an accuracy improvement to 97.59% and a reduction in test loss to 0.0882. The accuracy and loss curves appear smoother and more stable, indicating that the SE Block enhances feature selection through channel recalibration. This enhancement indicates that the integration of SE Block within the MobileNetV2 architecture is highly suitable for efficient implementation in environments with limited computational resources.

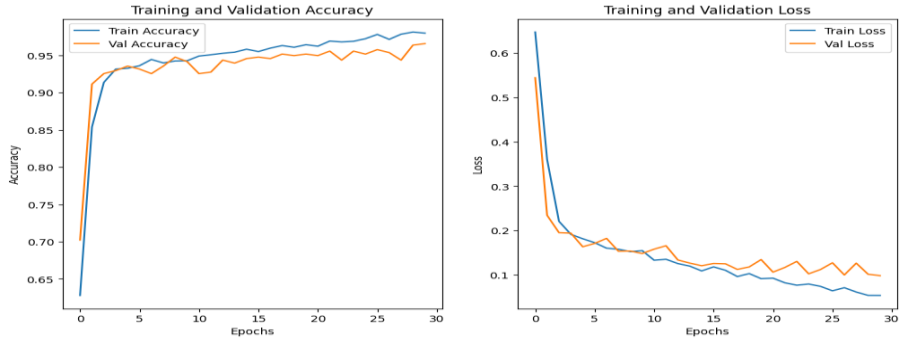


Figure 3. Training curve for CNN model after adding SE Block

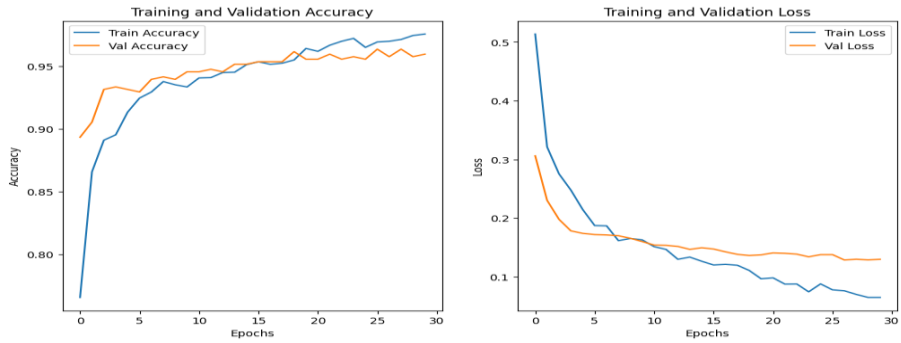


Figure 4. Training curve for MobileNetV2 without SE Block

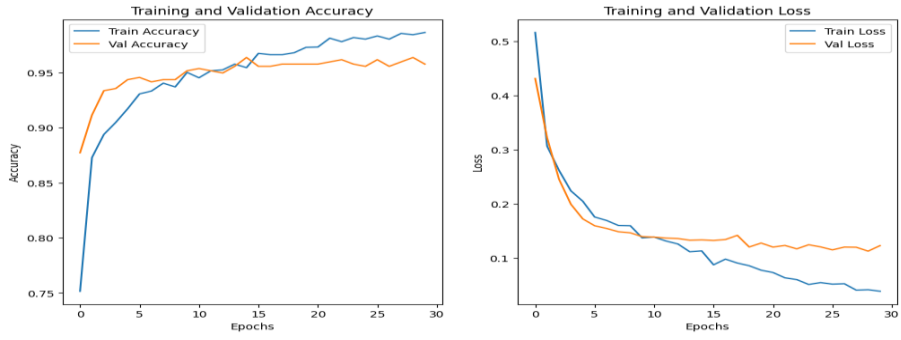


Figure 5. MobileNetV2 training curve after adding SE Block

Figure 12 illustrates the training results of the VGG16 architecture before the SE Block integration. The model achieved 96.88% accuracy with a test loss of 0.0841. The loss and accuracy trends exhibit strong stability during training, indicating VGG16's solid capacity to hierarchically extract complex spatial representations. Even without the SE Block, the model demonstrates competitive results.

Figure 13 depicts the VGG16 architecture enhanced with SE Block integration. The inclusion of the SE module leads to an accuracy improvement to 97.69% and a slight reduction in test loss to 0.0827. The resulting curves show smoother and more stable convergence compared to the baseline model. This finding suggests that the SE Block substantially enhances VGG16's learning performance by introducing channel-wise attention, enabling the network to better concentrate on key spatial regions and thereby improving its effectiveness in medical image classification.

Figure 14 illustrates the training results of the InceptionV3 model without SE Block integration. This model achieved an accuracy of 94.06% and a test loss value of 0.1535. The loss curve exhibits more noticeable fluctuations compared to previous models, indicating that complex architectures like InceptionV3 are more sensitive to data variability and training configurations. This instability suggests a need for more refined tuning or regularization strategies to improve convergence and generalization.

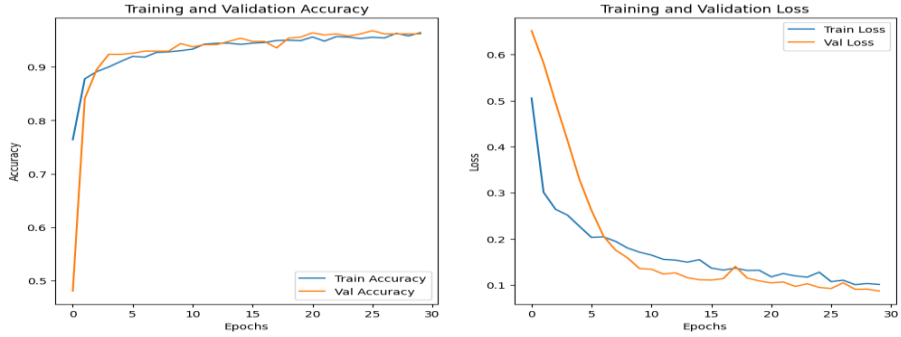


Figure 6. VGG16 training curve without SE Block

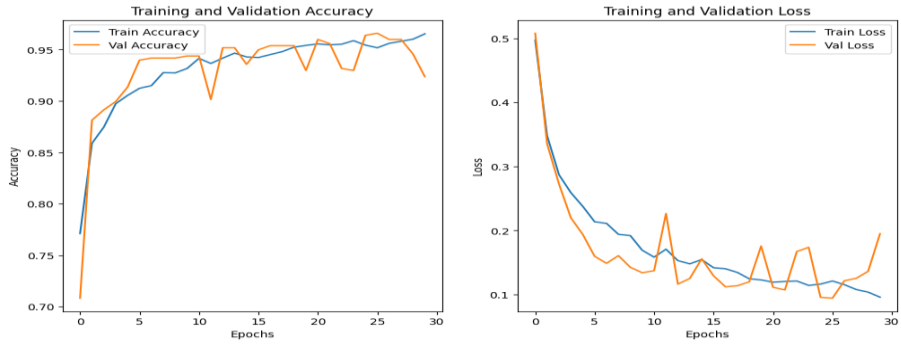


Figure 13. VGG16 training curve after SE Block integration

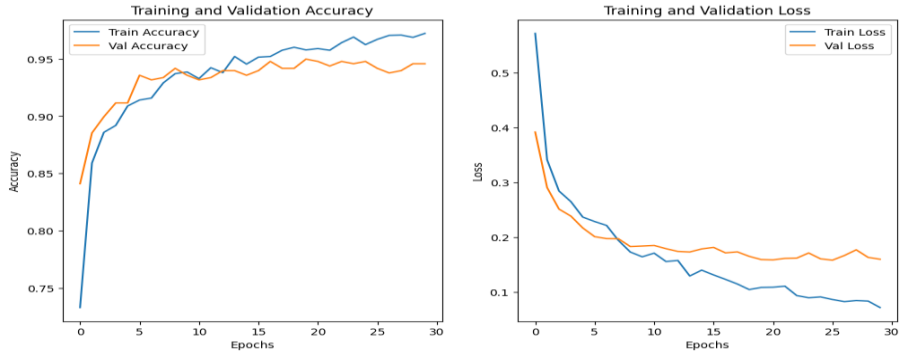


Figure 14. InceptionV3 training curve without SE Block

Figure 15 presents the InceptionV3 model's performance following the integration of the SE Block. The addition of the SE Block resulted in a minor improvement in accuracy, increasing from 94.06% to 94.16%. However, the test loss also rose from 0.1535 to 0.1922, suggesting that the model exhibited some instability during training. This outcome indicates that incorporating the SE Block did not yield substantial performance gains for this particular architecture. From an architectural perspective, InceptionV3 utilizes an

Inception module designed to process visual features through several parallel convolutional paths with different kernel dimensions (1×1, 3×3, and 5×5), combined with pooling operations to capture multiscale image information.

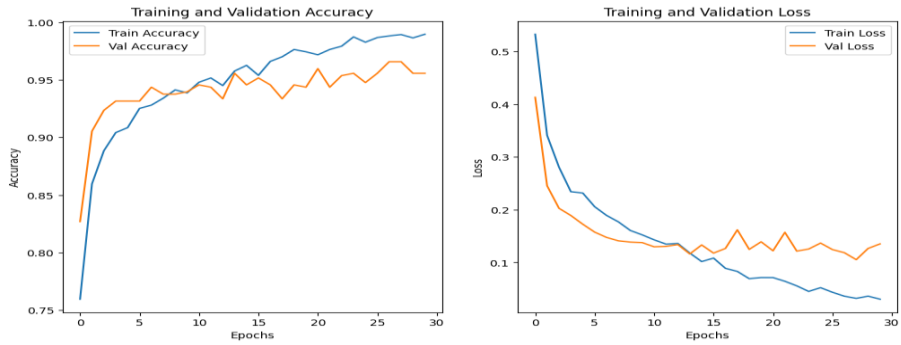


Figure 15. InceptionV3 training curve after SE Block integration

Channel-specific attention processes in SE Block have the potential to disproportionately suppress or magnify the contribution of certain features. thus disrupting the balance between parallel paths. These results emphasize that the effectiveness of the SE Block is not always uniform across all architectures, and in complex structures like InceptionV3, channel recalibration can actually reduce the model's ability to maintain a harmonious multiscale representation.

3.2 Model Performance Evaluation

To quantitatively assess model performance, this study employed a confusion matrix along with five primary evaluation metrics: accuracy, precision, recall, F1-score, and test loss. These indicators were selected to offer a comprehensive understanding of the model's ability to consistently and reliably distinguish pneumonia cases from normal chest X-ray images.

All performance measurements were carried out using the test dataset, which was entirely separate from both the training and validation data. This approach ensures that the evaluation outcomes accurately represent the model's generalization capability when encountering unseen samples. Such evaluation is crucial not only for measuring predictive accuracy but also for examining the model's sensitivity in detecting pneumonia cases and its precision in minimizing classification errors.

The confusion matrix is used to display the counts of accurate predictions (TP and TN) and incorrect predictions (FP and FN) for each model.. Visualization of the confusion matrices for all models is presented in Figures 16 to 19. The evaluation results indicate that models integrated with the SE Block tend to produce more accurate and consistent predictions compared to their standard counterparts.

Figure 16 shows the confusion matrices for both the baseline CNN and the CNN with SE Block. The standard CNN model exhibits several misclassifications, particularly in detecting pneumonia cases (false negatives). After incorporating SE Block, the incidence of misclassified samples showed a significant decrease. These results indicate that SE Block increases the model's sensitivity to pneumonia, which is often invisible and more difficult to identify in X-ray images.

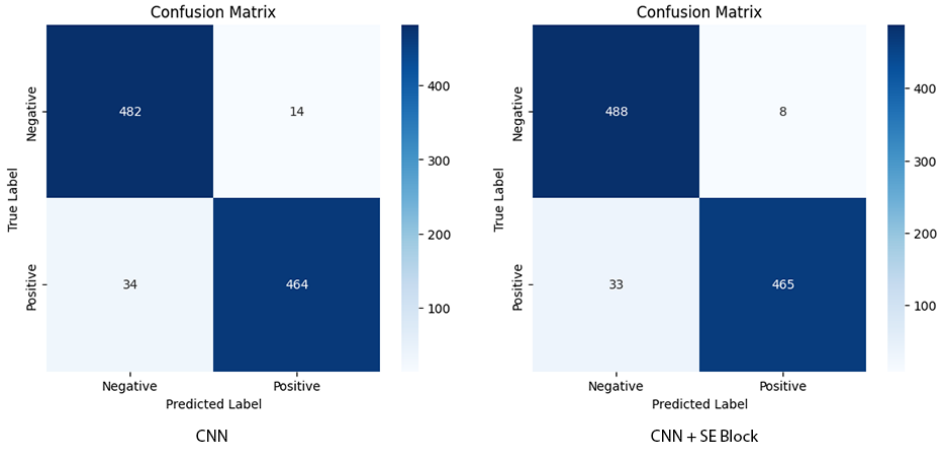


Figure 16. Confusion Matrix – CNN and CNN (SE Block)

Figure 17 displays the confusion matrix of the MobileNetV2 architecture before and after the integration of SE blocks. In its original form, the model already achieved high accuracy, but after the SE blocks were implemented, the balance between true positive and true negative predictions improved significantly.

SE blocks contributed to a stronger ability to identify the minority class (pneumonia) while maintaining accuracy for the majority class (normal). This improvement was further supported by higher recall and F1-score values, indicating increased sensitivity and a more balanced overall performance across both categories.

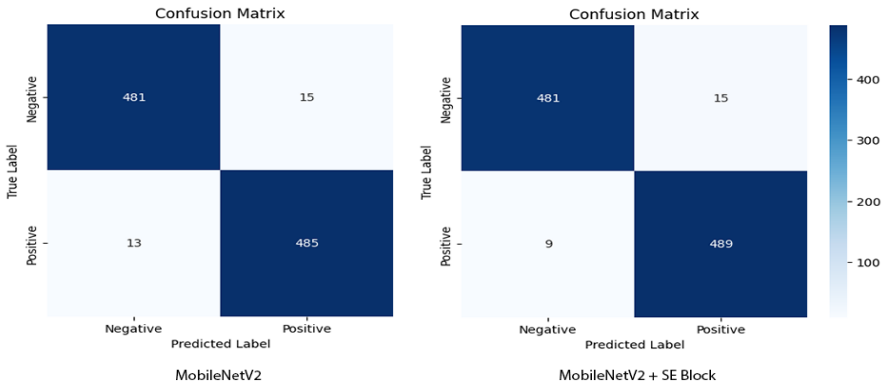


Figure 17. Confusion Matrix – MobileNetV2 and MobileNetV2 (SE Block)

Figure 18 presents the confusion matrix of the VGG16 architecture, comparing the standard configuration with the enhanced version by SE Block. The VGG16 model integrated with SE Block exhibits very few misclassifications, achieving very high accuracy and precision. These findings indicate that SE Block effectively strengthens feature representation within the robust VGG16 spatial framework, resulting in significant improvements in overall classification performance.

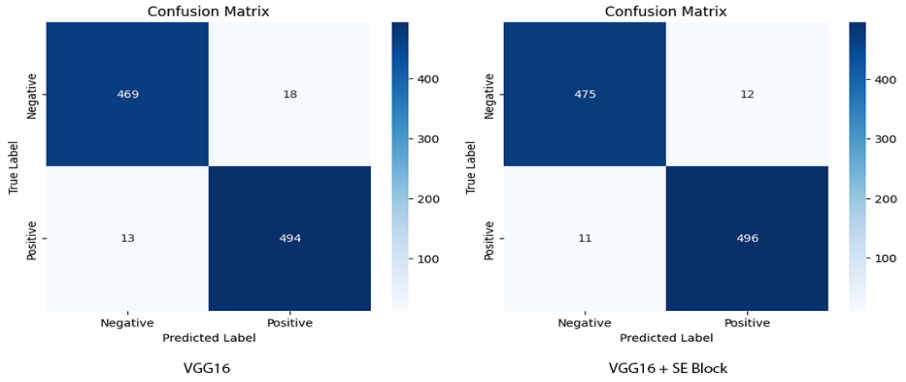


Figure 18. Confusion Matrix – VGG16 and VGG16 (SE Block)

Figure 19 presents the confusion matrices for the InceptionV3 model, both in its original form and with SE Block integration. While a slight reduction in classification errors is observed, the improvement is not as pronounced as in other models. This supports previous findings that the InceptionV3 architecture may not be fully compatible with SE Block in the context of this dataset. The addition of SE Block does not yield significant performance gains and may even hinder the model's ability to generalize.

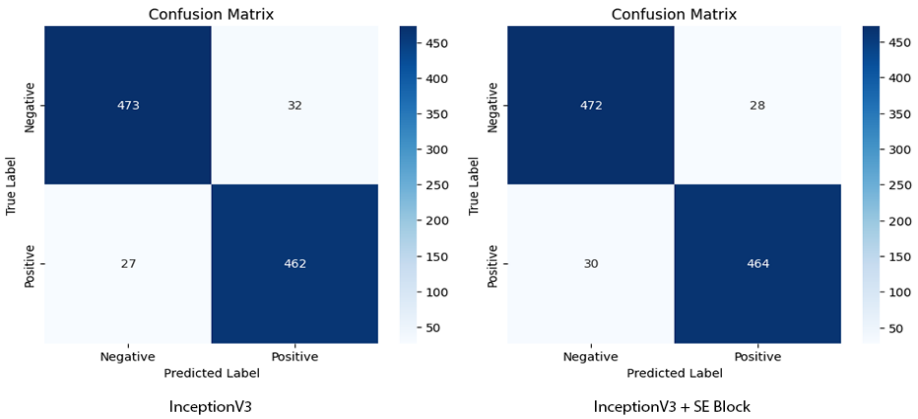


Figure 19. Confusion Matrix – InceptionV3 and InceptionV3 (SE Block)

Table 1 provides a detailed comparison between the baseline model (without SE Block) and the proposed model (with SE Block), evaluated using five performance metrics: accuracy, precision, recall, F1 score, and test loss. This table is designed to help readers clearly observe the impact of SE Block integration on the performance of each architecture.

Table 1 shows that the integration of the SE Block generally improves classification performance on CNNs. MobileNetV2, and VGG16 models. CNN improved from 95.17% to 95.88%, MobileNetV2 from 97.18% to 97.59%, and VGG16 from 96.88% to 97.69%, with a consistent decrease in test loss across all three. This indicates that the SE Block is effective in emphasizing important features in linear and sequential architectures.

MobileNetV2 performed best in terms of balancing accuracy and efficiency. Its lightweight structure based on depthwise separable convolution makes it highly compatible with the SE Block, enabling it to enhance feature attention without significantly increasing the computational burden. These characteristics make MobileNetV2 superior for implementation on constrained systems such as edge computing.

Table 1. Performance Comparison of Models

Model	Test Loss	Accuracy	Precision	Recall	F1-Score
CNN	0.1236	0.9517	0.9524	0.9517	0.9517
CNN (SE Block)	0.1169	0.9588	0.9599	0.9588	0.9587
MobileNetV2	0.0966	0.9718	0.9718	0.9718	0.9718
MobileNetV2 (SE Block)	0.0882	0.9759	0.9759	0.9759	0.9759
VGG16	0.0841	0.9688	0.9689	0.9688	0.9688
VGG16 (SE Block)	0.0827	0.9769	0.9769	0.9769	0.9769
InceptionV3	0.1535	0.9406	0.9407	0.9406	0.9406
InceptionV3 (SE Block)	0.1922	0.9416	0.9417	0.9416	0.9416

The custom CNN model also proved more stable than InceptionV3, primarily due to its simpler architecture, which makes it more adaptable to limited datasets. In contrast, InceptionV3 experienced a performance decline after the addition of the SE Block, with the test loss increasing from 0.1535 to 0.1922 due to a mismatch between the channel-level attention process implemented in the SE Block and the complex parallel architecture of InceptionV3. This mechanism reweights the contribution of each feature channel after global aggregation. In linear models like CNN, MobileNetV2, and VGG16, this consistently enhances important features. However, in InceptionV3, the generated features are derived from multi-scale parallel paths (1×1, 3×3, 5×5, and pooling). Channel recalibration can disproportionately suppress or magnify certain paths, disrupting the balance between parallel features. As a result, redundancy increases and generalization ability decreases. This suggests that the effectiveness of the SE Block is architecture-specific: optimal on linear structures, but less compatible on parallel architectures like InceptionV3. Overall, the effectiveness of the SE Block is highly dependent on the underlying structure of the model. Simple architectures like MobileNetV2 and CNN benefit significantly, while complex architectures require a more precise attention approach.

Beyond classification performance, this study also evaluates computational complexity by analyzing the number of parameters and estimated Floating Point Operations (FLOPs) for each model. This assessment is crucial for determining deployment feasibility on resource-constrained edge devices, such as Raspberry Pi, which face limitations in memory, processing power, and energy consumption [25], [37]. The following table presents the estimated model sizes and computational loads for all tested architectures, with and without SE Block integration. Estimates are based on standard CNN configurations with 128×128×3 input dimensions, using architecture definitions from Keras Applications and FLOPs formulas for convolutional layers [20], [36].

The results in Table 2 show that MobileNetV2 + SE Block has the lowest computational complexity among the high-performance models, with only around 2.5 million parameters and 330 million FLOPs. In comparison, VGG16 + SE Block requires approximately 11,800 million FLOPs, while InceptionV3 + SE Block requires 13,600 million FLOPs. Thus, MobileNetV2 + SE Block proves to be significantly more computationally efficient and well-suited for implementation on resource-constrained devices, without sacrificing classification accuracy.

Table 2. Comparison of Parameter Counts and Computational Complexity (FLOPs)

(Estimation based on 128×128×3 image input with standard configuration for each model)

Model	Parameters (M)	FLOPs (M)
CNN	1.2 M	180 M
CNN (SE Block)	1.4 M	210 M
MobileNetV2	2.3 M	300 M
MobileNetV2 (SE Block)	2.5 M	330 M
VGG16	15.0 M	11,200 M
VGG16 (SE Block)	16.0 M	11,800 M
InceptionV3	23.0 M	13,000 M
InceptionV3 (SE Block)	25.0 M	13,600 M

Although the SE Block is not a Transformer-based attention module like Multi-Head Self-Attention (MHSA), it employs a lightweight channel-wise attention mechanism. SE Blocks operate by recalibrating feature responses across channels rather than spatial positions, which significantly reduces computational cost.

In comparison, attention mechanisms derived from Transformer architectures such as Multi Head Self Attention (MHSA) analyze the relationships between all spatial elements within an input, leading to computational complexity that scales quadratically with the input dimensions. Therefore, SE Block achieves an effective compromise between improving model performance and maintaining computational efficiency, particularly in compact architectures like MobileNetV2 and custom CNNs.

The evaluation of model performance encompasses both accuracy and computational performance, which becomes critical in low-resource operational contexts. Therefore, total training time and inference time per model were also analyzed.

From the results shown in Table 3, MobileNetV2 and MobileNetV2 + SE Block exhibit the fastest training and inference times among all evaluated models, confirming their suitability for use in environments where computational resources are scarce. The addition of the SE Block only slightly increased the inference time (by approximately ±0.77 seconds), while still yielding noticeable improvements in classification accuracy. Interestingly, the custom CNN model, despite achieving high accuracy, recorded the longest total training time (±252 minutes), although it maintained a relatively fast inference speed. On the other hand, VGG16 and InceptionV3 reported the highest inference times, and while the SE Block integration did not significantly extend training durations, it did introduce additional computational overhead.

Overall, MobileNetV2 integrated with the SE Block achieves an effective balance among accuracy, training efficiency, and inference speed, thereby positioning it as a strong candidate for clinical deployment and real-time healthcare applications requiring prompt and accurate diagnostic decisions. Interestingly, the training time for InceptionV3 with SE Block (74.07 minutes) was significantly shorter than that of InceptionV3 without SE Block (96.33 minutes). This difference is likely due to non-deterministic factors during training, such as variations in the GPU optimization path, faster early stopping due to earlier convergence, or differences in batch processing due to changes in network structure. The addition of SE Block may also affect the gradient distribution, allowing the model to reach convergence faster despite theoretically adding a small amount of computational complexity. Therefore, this difference is more likely a training effect than a structural characteristic of SE Block. This phenomenon requires further verification through multiple runs to ensure consistency of the results and determine whether this training speedup represents a stable pattern or simply an anomaly in a particular experiment.

Table 3. Comparison of Training Time and Inference Time for Each Model

Model	Training Time (minutes)	Inference Time (seconds)
CNN	252.71	22.95
CNN (SE Block)	252.26	23.14
MobileNetV2	81.67	16.14
MobileNetV2 (SE Block)	81.80	16.91
VGG16	113.50	27.78
VGG16 (SE Block)	111.79	27.98
InceptionV3	96.33	30.61
InceptionV3 (SE Block)	74.07	30.61

3.3 Results and Discussion

According to the experimental outcomes, integrating the SE Block leads to consistent enhancements in the model's performance for pneumonia detection on chest X-ray data. The most pronounced improvements appear in recall and F1-score, two key metrics in medical imaging that indicate the accuracy of identifying true pneumonia cases.

Among all architectures evaluated, the VGG16 model integrated with SE Block demonstrated the highest performance, achieving an accuracy of 97.69%, an F1 score of 0.9769, and the lowest loss value of 0.0827. These results suggest that deep convolutional architectures such as VGG16 remain highly effective for medical image classification tasks, particularly when enhanced with adaptive attention mechanisms like the SE Block.

In addition, the MobileNetV2 model integrated with the SE Block achieved notable outcomes, reaching an accuracy of 97.59% and an F1-score of 0.9759. These results indicate that MobileNetV2 enhanced by the SE Block serves as an effective solution for environments with limited computational capacity, such as mobile and edge-based systems, because it successfully maintains an optimal trade-off between model accuracy and computational efficiency.

Compared to the other models, InceptionV3 displayed relatively lower performance than the other models. Although the integration of the SE Block slightly improved its accuracy from 94.06% to 94.16%, the loss value increased from 0.1535 to 0.1922. This finding implies a reduction in the model's generalization ability, possibly due to the high architectural complexity of InceptionV3, which may not be optimally suited for the dataset scale and distribution used in this study.

Overall, these results confirm that integrating the SE Block into CNN architectures enhances feature extraction effectiveness and strengthens the model's attention to relevant visual information. This approach holds strong potential for application in deep learning-based clinical decision support systems, particularly for computer-assisted diagnosis of pneumonia utilizing chest X-ray data with high accuracy and sensitivity.

3.4 Research Limitations

Several limitations were identified in this study. The first limitation is that this experiment only used a single Chest X-ray Pneumonia dataset with a limited amount of data. This could restrict the generalizability of the results, as the model might not completely represent the variability present in diverse clinical cases. Second, this study did not validate using a larger external medical dataset or one from a different institution. Consequently, the model's performance cannot be guaranteed to be consistent in real-world scenarios with varying image quality, radiology devices, or more heterogeneous patient populations. Third, although the integration of SE Blocks demonstrated improvements in some architectures, this study did not evaluate further optimization strategies, such as combining them with other attention mechanisms or selective placement of SE Blocks.

These limitations open up opportunities for future research to test the model on larger, cross-institutional datasets and explore hybrid attention approaches to improve robustness and generalizability.

4. Conclusions

The results of this study show that the application of the Squeeze and Excitation Block consistently improves pneumonia classification in chest X-ray images. This improvement was observed across custom CNN architectures and transfer learning models, including MobileNetV2 and VGG16. In nearly all models evaluated, the addition of the SE Block resulted in improved results across all key performance metrics.

The VGG16 model enhanced with the SE Block achieved the highest classification performance, attaining an accuracy of 97.69%, an F1-score of 0.9769, and a test loss of 0.0827. These findings suggest that the deep architecture of VGG16 synergizes effectively with the channel recalibration mechanism provided by the SE Block. In comparison, MobileNetV2, also integrated with the SE Block, delivered similar outcomes, with an accuracy of 97.59%, an F1-score of 0.9759, and a test loss of 0.0882. In addition, MobileNetV2 offers superior computational efficiency, utilizing only about 2.5 million parameters and 330 million floating-point operations, making it well-suited for deployment in resource-constrained environments.

In contrast, the InceptionV3 model, both with and without SE Block implementation, showed lower performance. Although there was a slight increase in accuracy from 94.06% to 94.16%, the test loss rose from 0.1535 to 0.1922, indicating reduced generalization capability. This outcome implies that the effectiveness of SE Blocks varies depending on the model architecture. In models with complex structures such as InceptionV3, the channel-wise attention mechanism implemented by SE Blocks may not be able to provide optimal results.

Therefore, MobileNetV2 integrated with the SE Block can be regarded as a well-balanced model that achieves optimal accuracy with efficient computational performance. It represents a strong potential candidate for use in CNN-based diagnostic frameworks incorporating SE mechanisms. This approach is particularly suitable for deployment in edge computing environments, including Raspberry Pi, Internet of Things (IoT)-based diagnostic tools, or mobile health systems. In summary, this research enhances the effectiveness of medical image classification models and fosters the creation of diagnostic frameworks that are intelligent, flexible, and efficient. These systems are consistent with the principles of system design and are especially beneficial in clinical contexts with limited technological resources, such as those commonly found in Indonesia.

References

[1] World Health Organization, "Pneumonia." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia

[2] Kementerian Kesehatan Republik Indonesia, "Profil Kesehatan Indonesia 2023." [Online]. Available: https://kemkes.go.id/app_asset/file_content_download/172231123666a86244b83fd8.51637104.pdf

[3] C. Gu and M. Lee, "Deep Transfer Learning Using Real-World Image Features for Medical Image Classification, with a Case Study on Pneumonia X-ray Images," *Bioengineering*, vol. 11, no. 4, p. 406, Apr. 2024, doi: 10.3390/bioengineering11040406.

[4] I. Mwendo, K. Gikunda, and A. Maina, "Deep transfer learning for detecting Covid-19, Pneumonia and Tuberculosis using CXR images -- A Review," Mar. 26, 2023, *arXiv*: arXiv:2303.16754. doi: 10.48550/arXiv.2303.16754.

[5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT: IEEE, Jun. 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.

[6] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Apr. 10, 2015, *arXiv*: arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556.

[7] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," Dec. 11, 2015, *arXiv*: arXiv:1512.00567. doi: 10.48550/arXiv.1512.00567.

[8] Z. Gao, Y. Tian, S.-C. Lin, and J. Lin, "A CT Image Classification Network Framework for Lung Tumors Based on Pre-trained MobileNetV2 Model and Transfer learning, And Its Application and Market Analysis in the Medical Field," *ACE*, vol. 133, no. 1, pp. 90–96, Jan. 2025, doi: 10.54254/2755-2721/2025.20605.

[9] S. Sharma and K. Guleria, "A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks," *Procedia Computer Science*, vol. 218, pp. 357–366, 2023, doi: 10.1016/j.procs.2023.01.018.

[10] M. Mujahid, F. Rustam, R. Álvarez, J. Luis Vidal Mazón, I. D. L. T. Díez, and I. Ashraf, "Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network," *Diagnostics*, vol. 12, no. 5, p. 1280, May 2022, doi: 10.3390/diagnostics12051280.

[11] Y. Pan *et al.*, "Fundus image classification using Inception V3 and ResNet-50 for the early diagnostics of fundus diseases," *Front. Physiol.*, vol. 14, p. 1126780, Feb. 2023, doi: 10.3389/fphys.2023.1126780.

[12] M. Neshat, M. Ahmed, H. Askari, M. Thilakaratne, and S. Mirjalili, "Hybrid Inception Architecture with Residual Connection: Fine-tuned Inception-ResNet Deep Learning Model for Lung Inflammation Diagnosis from Chest Radiographs," *Procedia Computer Science*, vol. 235, pp. 1841–1850, 2024, doi: 10.1016/j.procs.2024.04.175.

[13] M. Ali *et al.*, "Pneumonia Detection Using Chest Radiographs With Novel EfficientNetV2L Model," *IEEE Access*, vol. 12, pp. 34691–34707, 2024, doi: 10.1109/ACCESS.2024.3372588.

[14] R. Tamilkodi *et al.*, "Automated Medical Image Classification for Disease Prognosis," in *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, vol. 112, K. R. Madhavi, P. Subba Rao, J. Avanija, I. L. Manikyamba, and B. Unhelkar, Eds., in Advances in Computer Science Research, vol. 112. , Dordrecht: Atlantis Press International BV, 2024, pp. 1325–1334. doi: 10.2991/978-94-6463-471-6_127.

[15] A. Mabrouk, R. P. D. Redondo, A. Dahou, M. A. Elaziz, and M. Kayed, "Pneumonia Detection on chest X-ray images Using Ensemble of Deep Convolutional Neural Networks," *Applied Sciences*, vol. 12, no. 13, p. 6448, Jun. 2022, doi: 10.3390/app12136448.

[16] J. Zhou *et al.*, "scSE-NL V-Net: A Brain Tumor Automatic Segmentation Method Based on Spatial and Channel

'Squeeze-and-Excitation' Network With Non-local Block," *Front. Neurosci.*, vol. 16, p. 916818, May 2022, doi: 10.3389/fnins.2022.916818.

[17] S. Jiang, X. Chen, and C. Yi, "SSA-UNet: Whole brain segmentation by U-Net with squeeze-and-excitation block and self-attention block from the 2.5D slice image," *IET Image Processing*, vol. 18, no. 6, pp. 1598–1612, May 2024, doi: 10.1049/ipr2.13052.

[18] G. Prasanna, J. R. Ernest, G. Lalitha, and S. Narayanan, "Squeeze Excitation Embedded Attention U-Net for Brain Tumor Segmentation," in *Emerging Electronics and Automation*, vol. 1088, M. Gabbouj, S. S. Pandey, H. K. Garg, and R. Hazra, Eds., in Lecture Notes in Electrical Engineering, vol. 1088. , Singapore: Springer Nature Singapore, 2024, pp. 107–117. doi: 10.1007/978-981-99-6855-8_9.

[19] J.-S. Kim *et al.*, "E-HRNet: Enhanced Semantic Segmentation Using Squeeze and Excitation," *Electronics*, vol. 12, no. 17, p. 3619, Aug. 2023, doi: 10.3390/electronics12173619.

[20] K. A. Huang, A. Alkadri, and N. Prakash, "Employing Squeeze-and-Excitation Architecture in a Fine-Tuned Convolutional Neural Network for Magnetic Resonance Imaging Tumor Classification," *Cureus*, vol. 17, no. 3, pp. 1–9, Mar. 2025, doi: 10.7759/cureus.80084.

[21] Md. R. Hasan, S. M. Azmat Ullah, and S. Md. Rabiul Islam, "Recent advancement of deep learning techniques for pneumonia prediction from chest X-ray image," *Medical Reports*, vol. 7, p. 100106, Oct. 2024, doi: 10.1016/j.hmedic.2024.100106.

[22] R. Siddiqi and S. Javaid, "Deep Learning for Pneumonia Detection in Chest X-ray Images: A Comprehensive Survey," *J. Imaging*, vol. 10, no. 8, p. 176, Jul. 2024, doi: 10.3390/jimaging10080176.

[23] A. Roy, A. Bhattacharjee, D. Oliva, O. Ramos-Soto, F. J. Alvarez-Padilla, and R. Sarkar, "FA-Net: A Fuzzy Attention-aided Deep Neural Network for Pneumonia Detection in Chest X-Rays," Jun. 21, 2024, *arXiv*: arXiv:2406.15117. doi: 10.48550/arXiv.2406.15117.

[24] T. Borvornvitchotikarn and T. Yooyativong, "Combining convolutional neural networks and spatial-channel 'squeeze and excitation' block for multiple-label image classification," *IJ-AI*, vol. 13, no. 1, p. 368, Mar. 2024, doi: 10.11591/ijai.v13.i1.pp368-374.

[25] 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.

[26] T. Sanida, M. V. Sanida, A. Sideris, and M. Dasygenis, "Optimizing Lung Condition Categorization through a Deep Learning Approach to Chest X-ray Image Analysis," *BioMedInformatics*, vol. 4, no. 3, pp. 2002–2021, Sep. 2024, doi: 10.3390/biomedinformatics4030109.

[27] M. Paul, "Chest X-Ray Images (Pneumonia)," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/data

[28] S. Saponara and A. Elhanashi, "Impact of Image Resizing on Deep Learning Detectors for Training Time and Model Performance," in *Applications in Electronics Pervading Industry, Environment and Society*, vol. 866, S. Saponara and A. De Gloria, Eds., in Lecture Notes in Electrical Engineering, vol. 866. , Cham: Springer International Publishing, 2022, pp. 10–17. doi: 10.1007/978-3-030-95498-7_2.

[29] R. Buettner, C. Mai, and P. Penava, "Improvement of Deep Learning Models Using Retinal Filter: A Systematic Evaluation of the Effect of Gaussian Filtering With a Focus on Industrial Inspection Data," *IEEE Access*, vol. 13, pp. 43201–43217, 2025, doi: 10.1109/ACCESS.2025.3549271.

[30] C. Mai, P. Penava, and R. Buettner, "A Novel Deep Learning-Based Approach for Defect Detection of Synthetic Leather Using Gaussian Filtering," *IEEE Access*, vol. 12, pp. 196702–196714, 2024, doi: 10.1109/ACCESS.2024.3521497.

[31] P. Ormeño-Arriagada, G. Márquez, D. Araya, C. Rimassa, and C. Taramasco, "Applying Machine Learning Sampling Techniques to Address Data Imbalance in a Chilean COVID-19 Symptoms and Comorbidities Dataset," *Applied Sciences*, vol. 15, no. 3, p. 1132, Jan. 2025, doi: 10.3390/app15031132.

[32] C. Yang, E. A. Fridgeirsson, J. A. Kors, J. M. Reps, and P. R. Rijnbeek, "Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data," *J Big Data*, vol. 11, no. 1, p. 7, Jan. 2024, doi: 10.1186/s40537-023-00857-7.

[33] R. S. Abdulsadig and E. Rodriguez-Villegas, "A comparative study in class imbalance mitigation when working with physiological signals," *Front. Digit. Health*, vol. 6, p. 1377165, Mar. 2024, doi: 10.3389/fdgth.2024.1377165.

[34] C. Yashvi, U. Vivek, B. Yamini, L. Nidhi, and R. Ravindra, "Data Augmentation Techniques applied to Medical Images," *International Journal of Research Publication and Reviews*, vol. 5, no. 7, pp. 483–501, Jul. 2024.

[35] M. Rohan, "Data Splitting (Train-Test-Validation) in Machine Learning." [Online]. Available: https://medium.com/@rohanmistry231/data-splitting-train-test-validation-in-machine-learning-2d5d1927fa69

[36] A. Surya, A. K. Shah, J. Kabore, and S. T. Sasikumar, "Enhanced Breast Cancer Tumor Classification using MobileNetV2: A Detailed Exploration on Image Intensity, Error Mitigation, and Streamlit-driven Real-time Deployment," Dec. 2023, *arXiv*: arXiv:2312.03020. doi: 10.48550/arXiv.2312.03020.

[37] M. Ilham, A. Prihantoro, I. K. Perdana, R. Magdalena, and S. Saidah, "Experimenting with the Hyperparameter of Six Models for Glaucoma Classification," *J. Ilm. Tek. Elektro Komput. Dan Inform*, vol. 9, no. 3, pp. 571–584, Jul. 2023, doi: 10.26555/jiteki.v9i3.26331.

[38] H. H. Zargar, S. H. Zargar, R. Mehri, and F. Tajdini, "Using VGG16 Algorithms for classification of lung cancer in CT scans Image," May 2023, *arXiv*: arXiv:2305.18367. doi: 10.48550/arXiv.2305.18367.

[39] A. Sharma and S. Arora, "Brain Tumor Detection using CNN and VGG-16 Model," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 3, pp. 4203–4212, Mar. 2024.

[40] A. Tharwat, "Classification assessment methods," *ACI*, vol. 17, no. 1, pp. 168–192, Jan. 2021, doi: 10.1016/j.aci.2018.08.003.