Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics e-ISSN: 2656-8624
Homepage: https://ijeeemi.org/; Vol. 8, No. 3, pp. 365-379, August 2026

RESEARCH PAPER OPEN ACCESS

Comparative Analysis of Image Augmentation and Class Weighting on ResNet50-CBAM for Pneumonia Detection from Chest X-Ray Images

Kafilah Akhmad Fatahillah, Triando Hamonangan Saragih, Dwi Kartini, Fatma Indriani, and Irwan Budiman

Department of Computer Science, Faculty of Mathematics and Natural Science, Lambung Mangkurat University, Banjarbaru, Indonesia

Corresponding author: Triando Hamonangan Saragih (e-mail: triando.saragih@ulm.ac.id), Author(s) Email: Kafilah Akhmad Fatahillah (e-mail: 2211016210021@mhs.ulm.ac.id); Dwi Kartini (e-mail: dwikartini@ulm.ac.id); Fatma Indriani (e-mail: f.indriani@ulm.ac.id); Irwan Budiman (e-mail: irwan.budiman@ulm.ac.id)

Abstract Pneumonia remains one of the leading causes of morbidity and mortality worldwide, particularly among children, older adults, and immunocompromised individuals. Although chest X-ray (CXR) imaging is widely used to diagnose pneumonia, manual interpretation can be time-consuming, subjective, and dependent on radiologist expertise. Deep learning has shown promising performance for automated pneumonia classification; however, class imbalance may introduce prediction bias and affect classification performance. This study systematically evaluates image augmentation and class weighting for handling class imbalance in pneumonia classification using a ResNet50-CBAM framework. Four experimental scenarios were evaluated: Baseline, Augmentation Only, Class Weighting Only, and Hybrid, using identical experimental settings. We conducted experiments on the publicly available Chest X-Ray Images (Pneumonia) dataset from Kaggle, comprising 1,583 Normal and 4,273 Pneumonia images. We divided the dataset into 80% training, 10% validation, and 10% testing subsets. Pneumonia was defined as the positive class, and the optimal classification threshold for each scenario was determined using Youden's J-Statistic on the validation set before final evaluation on the test set. Model performance was assessed using Accuracy, ROC-AUC, Precision, Recall, Macro F1-score, and confusion matrix analysis. The Hybrid model achieved the best overall performance, with an Accuracy of 95.74%, ROC-AUC of 98.83%, Macro Precision of 96.01%, Macro Recall of 93.13%, and Macro F1-score of 94.44%. It also reduced false negative predictions from 25 cases in the Baseline model to 5 cases. These findings indicate that combining image augmentation and class weighting provides complementary benefits for handling class imbalance within the evaluated ResNet50-CBAM framework. The findings are limited to internal validation on the evaluated dataset and require further validation using independent clinical datasets.

Keywords: Pneumonia; Image Augmentation; Class Weighting; ResNet50; CBAM;

I. Introduction

Pneumonia remains a major global public health challenge and continues to be one of the leading causes of respiratory morbidity and mortality worldwide, particularly among children under five years of age, older adults, and immunocompromised individuals [1] [2]. Chest X-ray (CXR) imaging is widely used to detect pneumonia because of its accessibility, affordability, and effectiveness in identifying pulmonary abnormalities. However, manual interpretation of chest radiographs remains time-consuming and highly dependent on radiologist expertise [3]. Furthermore, inter-observer variability and the visual similarity between pneumonia and other thoracic abnormalities may lead to inconsistent diagnoses, potentially delaying clinical intervention and affecting patient outcomes [4]. Despite the rapid adoption of artificial intelligence in medical imaging, class imbalance remains a major challenge for automated pneumonia classification [5]. Publicly available chest X-ray datasets often contain substantially more Pneumonia images than Normal images, which may cause deep learning models to favor the majority class [6]. Consequently, minority-class samples may be misclassified, potentially reducing the reliability of computer-aided diagnosis [7] [8].

Recent advances in Artificial Intelligence (AI), particularly Deep Learning (DL), have significantly improved automated medical image analysis [9] [10]. Transfer learning based on Convolutional Neural Networks (CNNs) has become widely adopted because it enables extraction of discriminative image features without manual feature engineering [11] [12] [13]. Among various CNN architectures, ResNet50 has demonstrated strong feature-extraction capability through residual learning and stable optimization of deep networks [14]. Previous studies have reported ResNet50-based models achieving accuracies of up to 98.14% and an AUC of 99.71% for pneumonia detection using chest X-ray images [15] [16]. To improve feature representation, attention mechanisms have also been incorporated into CNN architectures. The Convolutional Block Attention

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

Module (CBAM), for example, sequentially applies channel and spatial attention to refine feature maps and emphasize diagnostically relevant information [12]. Previous studies integrating CBAM with residual networks reported accuracies of up to 98.41% and F1-scores of 0.98 [9]. In addition to model architecture, selecting a classification threshold is important for imbalanced datasets because the conventional threshold of 0.5 may not provide an appropriate balance between sensitivity and specificity [6]. Youden's J-Statistic provides a systematic approach for selecting an optimal threshold based on sensitivity and specificity [17] [18].

Image augmentation and class weighting represent two different approaches to handling class imbalance. Image augmentation operates at the data level by introducing additional variations into training images, whereas class weighting operates at the optimization level by assigning greater importance to underrepresented classes. Previous studies have shown the effectiveness of these approaches; however, they often evaluate their individual and combined effects under different experimental settings. Differences in datasets, preprocessing procedures, architectures, training configurations, and evaluation protocols also make direct comparisons difficult. Therefore, this study presents a controlled comparative investigation of image augmentation and class weighting for pneumonia classification using chest X-ray images. A single ResNet50-CBAM architecture was used as the common backbone across four experimental scenarios: Baseline, Augmentation Only, Class Weighting Only, and Hybrid. Youden's J-Statistic was additionally used to determine the optimal classification threshold from validation data. Rather than proposing a new deep learning architecture, this study's primary contribution is a controlled quantitative comparison of data-level and algorithm-level imbalance-handling strategies under consistent experimental conditions.

The study aims to determine the individual and combined effects of image augmentation and class weighting while maintaining the same dataset, preprocessing procedure, ResNet50-CBAM architecture, training configuration, and evaluation protocol across all scenarios. This design enables a more consistent examination of each imbalance-handling strategy's relative contribution within the evaluated experimental setting. Overall, this study makes four main contributions: (1) a systematic comparison of four imbalance-handling scenarios for chest X-ray pneumonia classification; (2) a quantitative evaluation of the individual and combined effects of image augmentation and class weighting within the ResNet50-CBAM framework; (3) an analysis of the performance trade-offs associated with different imbalance-handling strategies; and (4) the use of Youden's J-Statistic to determine the optimal classification threshold from validation data.

The remainder of this paper is organized as follows. Section II describes the dataset, preprocessing procedures, ResNet50-CBAM architecture, experimental scenarios, and evaluation methods. Section III presents the experimental results, including individual model performance, comparative analysis, and confusion matrix evaluation. Section IV discusses the findings, compares them with related studies, and outlines limitations and implications. Finally, Section V concludes the study and outlines directions for future research.

II. Materials and Method

This study adopts a systematic framework for pneumonia classification using chest X-ray images. The workflow begins with data preparation and preprocessing, followed by four experimental scenarios involving image augmentation, class weighting, and their combination. The processed data are then classified using the proposed ResNet50-CBAM architecture. Model performance is evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix analysis. Additionally, Youden's J-Statistic is applied to determine the optimal classification threshold. Finally, the performance of all scenarios is compared to identify the most effective strategy for handling class imbalance in pneumonia detection. Fig. 1 illustrates the overall research workflow employed in this study. The process begins with collecting and preparing chest X-ray images, followed by partitioning the dataset into training, validation, and testing subsets. Subsequently, image preprocessing is performed through resizing and pixel normalization. To address class imbalance , image augmentation and class weighting strategies are applied according to the predefined experimental scenarios. The processed images are then fed into the proposed ResNet50-CBAM architecture, which combines deep feature extraction and attention mechanisms for pneumonia classification. During training, the model parameters are optimized using transfer learning and fine-tuning techniques [19]. The classification performance is evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix analysis. Furthermore, Youden's J-Statistic is employed to determine the optimal classification threshold, enabling a more balanced evaluation on the imbalanced dataset. Finally, the performance of the four experimental scenarios is compared to identify the most effective strategy for handling data imbalance in pneumonia detection.

A. Data Collection

Table 1. Distribution of chest X-ray images across Normal and Pneumonia classes used in this study.

Class	Number of Images
Normal	1,583
Pneumonia	4,273
Total	5,856

This study utilized the Chest X-Ray Images (Pneumonia) dataset developed by Paul Mooney and publicly available on Kaggle platform. The dataset has been widely adopted as a benchmark in pneumonia classification research because of its accessibility and well-defined class labels. It consists of chest radiographs categorized into two

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics e-ISSN: 2656-8624
Homepage: https://ijeeemi.org/; Vol. 8, No. 3, pp. 365-379, August 2026

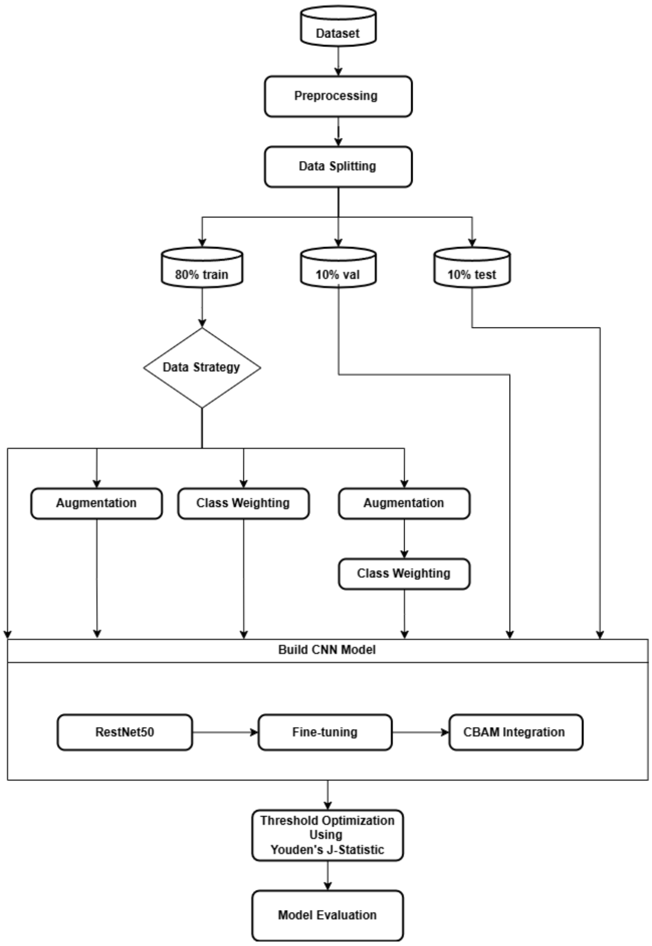

Fig. 1. Flowchart of the proposed study, including data preprocessing, imbalance handling, ResNet50-CBAM training, threshold optimization, and performance evaluation.

classes, namely Normal and Pneumonia, representing healthy lungs and lungs affected by pneumonia, respectively [7]. A total of 5,856 chest X-ray images were used in this study, comprising 1,583 Normal images and 4,273 Pneumonia images. The distribution indicates an imbalanced dataset with an approximate ratio of 1:2.7, where the Pneumonia class constitutes the majority class. Such class imbalance may introduce prediction bias toward the dominant class and potentially reduce the model's ability to correctly identify minority-class samples. Table 1 presents the distribution of chest X-ray images used in this study.

Table 2. Algorithm for image preprocessing through resizing and normalization of chest X-ray images.

Image preprocessing
1. **BEGIN**
2. Dataset = load_images(DATASET_PATH)
3. IMG_SIZE = 224
4. FOR each Image in Dataset DO
5. Image = resize(Image, IMG_SIZE, IMG_SIZE)
6. Image = rescale(Image, 1/255)
7. Add Image to Preprocessed_Dataset
8. END FOR
9. OUTPUT Preprocessed_Dataset
10. **END**

B. Image preprocessing

Image preprocessing was performed to standardize chest X-ray images and reduce visual inconsistencies before model training. Because the original radiographs varied in image resolution and pixel intensity, preprocessing was necessary to ensure uniform input for the deep learning model [4]. First, all images were resized to 224 × 224 pixels to meet the input requirements of the pre-trained ResNet50 architecture and reduce computational complexity [20]. Subsequently, pixel values were normalized by rescaling the intensity range to 0-1, reducing intensity variations among images, stabilizing the training process, and facilitating faster model convergence [3]. The resulting preprocessed images were then used as the input for the data splitting, augmentation, and classification stages. The overall preprocessing procedure is summarized in Table 2.

C. Data Splitting

Subsequent to the preprocessing phase, the radiographic dataset was partitioned into three distinct subsets for

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

training, validation, and testing using a stratified image-level splitting strategy with an 80:10:10 allocation ratio. Specifically, 80% of the images were allocated for model training, while the remaining images were equally divided between the validation and testing sets, each comprising 10% of the dataset [11]. Stratified splitting was employed to preserve the original class distribution of Normal and Pneumonia images across all subsets, thereby minimizing potential sampling bias and ensuring each subset adequately represented the overall dataset distribution. Stratified partitioning process was performed independently for each class to maintain proportional class distributions across the training, validation, and testing subsets. To guarantee experimental reproducibility and maintain strict consistency across all four experimental scenarios, the partitioning process was implemented in Python using a fixed random seed of 42. The dataset was automatically shuffled prior to splitting, after which the images were assigned to the training, validation, and testing subsets according to the

Table 3. Distribution of chest X-ray images after stratified dataset splitting.

Subset	Normal	Pneumonia	Total
Training	1.266	3.418	4.684
Validation	158	427	585
Testing	159	428	587
Total	1.583	4.273	5.856

predefined ratio. The resulting image distribution for each subset is presented in Table 3.

During the network's learning phase, the training set was exclusively used to optimize the model parameters [1]. The validation set served as an independent subset for monitoring model convergence, tuning hyperparameters, and determining the optimal classification threshold using Youden's J-Statistic. After model development was completed, the testing set was used only once to evaluate final classification performance, ensuring no test samples were used during model optimization or threshold selection [3]. Furthermore, strict separation was maintained between the training, validation, and testing subsets throughout the entire experimental process to prevent data leakage. No images from the testing set were used during model training, hyperparameter tuning, or threshold optimization. This partitioning strategy minimizes the risk of information leakage between subsets and enables a fair and objective comparison among the investigated imbalance-handling strategies [21].

The publicly available Chest X-Ray Images (Pneumonia) dataset does not provide patient identifiers. Consequently, the dataset was partitioned at the image level rather than the patient level. Therefore, the reported performance represents internal validation based on image-level splitting, and the generalizability of the proposed framework should be further investigated using independent external datasets.

D. Image Augmentation

Image augmentation was applied only to the training set to increase data diversity and reduce overfitting while preserving the original class labels [22]. In medical image analysis, geometric augmentation can enrich the spatial variability of training samples and provide additional variations for learning under imbalanced data conditions [23] [24]. Three geometric transformations were applied using Keras ImageDataGenerator: random rotation within ±5°, horizontal and vertical shifting of 0.05, and zooming of 0.05 [21]. These conservative transformations were selected to preserve the anatomical characteristics of chest X-ray images. Horizontal flipping was disabled because lateral inversion may alter the natural anatomical orientation of thoracic structures [23]. The transformations were generated stochastically and on the fly during training. No explicit per-transformation probability parameter was specified in the implementation. Therefore, a numerical augmentation probability is not reported [25]. The augmented images were generated dynamically rather than stored as additional files, providing varied training samples without permanently expanding the dataset [26]. Fig. 2 shows representative examples of the applied transformations.

E. Class Weighting

To further mitigate the effect of class imbalance, this study employed class weighting as an algorithm-level strategy by assigning larger penalties to misclassified samples from the minority class during model optimization. Unlike data-level approaches that increase the number of training samples, class weighting modifies the Binary Cross-Entropy loss function to reduce prediction bias toward the majority class and encourage a more balanced

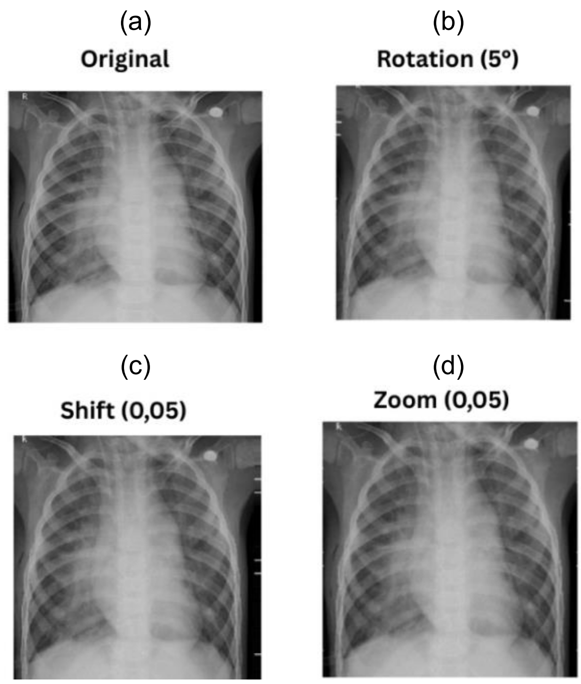

Fig. 2. Representative chest X-ray images before and after augmentation: (a) original, (b) rotation ±5°, (c) width and height shifting 0.05, and (d) zooming 0.05.

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

decision boundary [6] [17]. The initial class weights were automatically computed using the balanced weighting strategy provided by the Scikit-learn library, as expressed in Eq. (1).

$$\mathrm{w_i} = \frac{\mathrm{N}}{\mathrm{C} \times n_i} \quad (1)$$

where w_i is the class weight for class i, N is the total number of training samples, C denotes the number of classes, and n_i represents the number of training samples belonging to class i. To prevent excessively large differences between class weights, a soft-scaling transformation was subsequently applied according to Eq. (2).

$$w_i{}' = 1 + \alpha(w_i - 1) \quad (2)$$

where w_i' denotes the adjusted class weight after soft scaling, w_i is the original class weight obtained from Eq. (1), and α is the scaling factor. In this study, $\alpha = 0.5$ was adopted to moderate the difference between class weights while preserving greater emphasis on the minority class. Previous methodological studies have shown that moderated weighting strategies provide a better balance between minority-class learning and optimization stability than excessively aggressive weighting schemes, thereby reducing the risk of unstable optimization during training [22] [27].

Based on the class distribution of the training set, the resulting class weights were 1.85 for the Normal class and 0.69 for the Pneumonia class. These weights were consistently applied throughout all class-weighting experiments to encourage balanced learning while

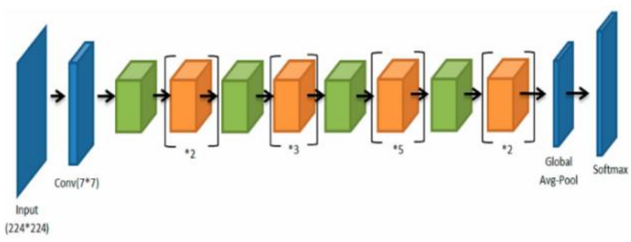

Fig. 3. Architecture of the pre-trained ResNet50 model adopted for hierarchical feature extraction from chest X-ray images, reproduced from [15].

maintaining stable model convergence. The complete procedure is presented in Table 4.

Table 4. Algorithm for computing and applying class weights to address class imbalance during model training.

Class Weighting	
1.	**BEGIN**
2.	y_train = extract_class_labels(Training_Set)
3.	Original_Weights = compute_class_weight(balanced, y_train)
4.	soft_factor = 0.5

5.	FOR each Weight in Original_Weights DO
	Soft_Weight = 1 + soft_factor × (Weight − 1)
	END FOR

6.	Class_Weights = collect(Soft_Weight)
7.	Apply Class_Weights during model training
8.	OUTPUT Class_Weights
9.	**END**

F. CNN and ResNet50 Architecture

Convolutional Neural Networks (CNNs) have emerged as a foundational framework for medical image classification, primarily due to their intrinsic capacity to autonomously learn hierarchical feature representations directly from raw pixel data [12] [28]. Structurally, a standard CNN operates through a sequential combination of convolutional, pooling, activation, and fully connected layers, which collaboratively extract and abstract complex visual patterns [3]. By bypassing the subjective and tedious process of manual feature engineering required in conventional machine learning, CNNs achieve significantly superior and more robust diagnostic performance across various medical imaging tasks [12] [29]. For the specific task of pneumonia detection, this study employs ResNet50 as the primary feature-extraction backbone. As a deep architecture, ResNet50 distinguishes itself by integrating identity shortcut connections that effectively route information across the network [8]. This residual learning mechanism is crucial for mitigating the vanishing gradient problem typically encountered in exceedingly deep networks, thereby allowing the model to safely increase depth and capture highly discriminative pathological features from chest radiographs [5]. To accelerate computational convergence and elevate feature extraction capabilities, we leveraged a transfer learning paradigm utilizing a ResNet50 model pre-trained on the large-scale ImageNet dataset [19]. During the fine-tuning phase, the final 50 layers of the network were unfrozen and optimized specifically on the chest X-ray dataset, while the earlier layers maintained their pre-trained parameters. This strategic adaptation ensures that the network successfully transitions from recognizing generic shapes to identifying pneumonia-specific clinical patterns without discarding the fundamental visual knowledge acquired from the massive source domain [22]. Following the deep feature extraction phase, the feature maps generated by the ResNet50 backbone are subsequently routed into a Convolutional Block Attention Module (CBAM). This attention mechanism refines the final feature representation by adaptively emphasizing diagnostically salient regions spatially and across channels while suppressing irrelevant background noise [12]. Fig. 3[15] visually summarizes the overall structural pipeline of the proposed architecture.

G. Convolutional Block Attention Module (CBAM)

To improve feature representation, the proposed model integrates the Convolutional Block Attention Module

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics e-ISSN: 2656-8624
Homepage: https://ijeeemi.org/; Vol. 8, No. 3, pp. 365-379, August 2026

(CBAM) into the ResNet50 backbone. The CBAM module is inserted after the final feature extraction stage of ResNet50 to refine the extracted feature maps before the classification process. As illustrated in Fig. 4, CBAM sequentially applies the Channel Attention Module (CAM) and Spatial Attention Module (SAM) to emphasize informative features while suppressing less relevant responses, thereby improving feature discrimination for pneumonia classification [12]. The Channel Attention Module (CAM) computes channel-wise attention using both Global Average Pooling and Global Max Pooling, followed by a shared multilayer perceptron with a reduction ratio of 8 and ReLU activation. The generated attention weights are normalized using a sigmoid activation and multiplied element-wise with the input feature map. Subsequently, the Spatial Attention Module (SAM) aggregates spatial information through average pooling and maximum pooling across the channel dimension, followed by a 7 × 7 convolution and sigmoid activation to generate the spatial attention map, which is multiplied with the intermediate feature map to produce refined feature representations [12].

The refined feature map is subsequently processed through a Global Average Pooling layer, followed by a fully connected layer with 512 neurons using ReLU activation, a Dropout layer (0.5), and a final sigmoid output layer for binary classification of Normal and Pneumonia images. Representative feature maps before and after applying the CBAM module are presented in Fig. 5, illustrating the refinement of intermediate feature responses after sequential channel and spatial attention.

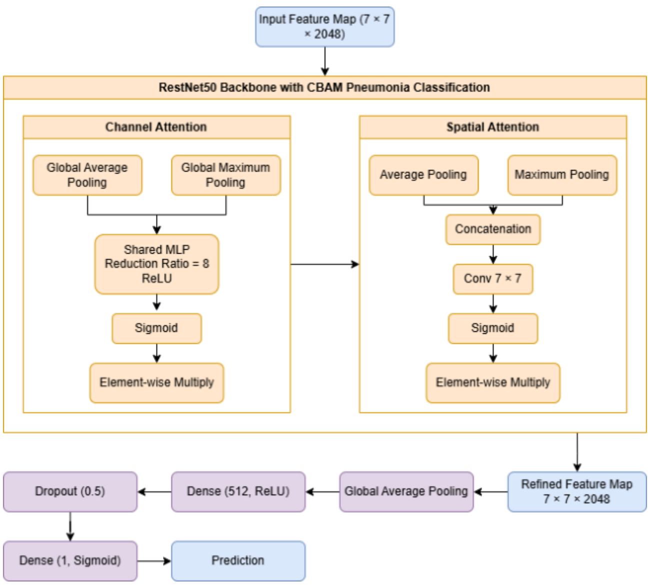

Fig. 4. Architecture of the proposed ResNet50-CBAM framework for pneumonia classification.

H. Training Configuration

The ResNet50-CBAM models were implemented using Python 3.11.9, TensorFlow 2.16.1, Keras 3.7.0, and Scikit-learn 1.7.0. ResNet50 was initialized with ImageNet-pretrained weights, using 224×224 input images and a batch size of 16. The final 50 ResNet50 layers were fine-tuned, followed by CBAM, global average pooling, a 512-unit dense layer, Dropout (0.5), and a sigmoid output. The resulting model contained 25,688,291 parameters, with 25,635,171 trainable and 53,120 non-trainable parameters.

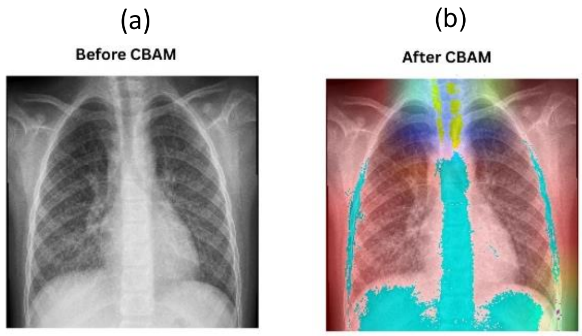

Fig. 5. Visualization of chest X-ray images before and after applying CBAM for feature refinement in pneumonia classification: (a) before CBAM, (b) after CBAM.

Training used Adam with a learning rate of 5×10^{-6} and Binary Crossentropy for a maximum of 50 epochs. EarlyStopping used validation accuracy with a patience of 15, while ModelCheckpoint retained the best validation model, and ReduceLROnPlateau used a factor of 0.3, patience of 2, and a minimum learning rate of 1×10^{-6}. Images were rescaled by 1/255 and augmented during training using rotation (±5°), width/height shifts (0.05), and zoom (0.05), without horizontal flipping. The data were split at the image level into 80% training, 10% validation, and 10% testing using random seed 42. For class-weighting scenarios, balanced class weights were adjusted using a 0.5 soft-scaling factor.

I. Evaluation

The performance of the proposed model was evaluated using several classification metrics derived from the confusion matrix, including Accuracy (ACC), Precision (PRE), Recall (REC), F1-Score (F1), Macro Precision, Macro Recall, and Macro F1-Score, as well as Receiver Operating Characteristic (ROC) analysis and the Area Under the Curve (AUC) [3] [4]. These metrics were selected to provide a comprehensive assessment of classification performance under different imbalance-handling scenarios. In this study, Pneumonia was treated as the positive class, whereas Normal was considered the negative class. Accuracy (ACC) measures the proportion of correctly classified samples among all evaluated samples. Higher accuracy indicates better overall classification performance [3]. The computation of accuracy is presented in Eq. (3) [3].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where TP, TN, FP, and FN denote the numbers of True Positive, True Negative, False Positive, and False Negative samples, respectively.

Precision (PRE) measures the proportion of correctly predicted positive samples among all samples predicted as positive. Higher precision indicates fewer false-positive predictions [3]. The computation of precision is presented in Eq. (4) [3].

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

Table 5. Training configuration of the proposed ResNet50-CBAM model.

Parameter	Configuration
Software	Python 3.11.9; TensorFlow 2.16.1; Keras 3.7.0; Scikit-learn 1.7.0
Pretrained weights	ImageNet
Input size	224 × 224 × 3
Batch size	16
Maximum epochs	50
Optimizer	Adam
Learning rate	5×10^{-6}
Loss function	Binary Crossentropy
Fine-tuning	Final 50 ResNet50 layers
Dropout	0.5
EarlyStopping	Val. accuracy; patience = 15
ReduceLROnPlateau	Factor = 0.3; patience = 2
Minimum learning rate	1×10^{-6}
Class-weight scaling	0.5
Data split	80:10:10
Split level	Image level
Random seed	42
Total parameters	25,688,291
Trainable parameters	25,635,171
Non-trainable parameters	53,120
Hardware	AMD Ryzen 5 5625U; 8 GB RAM
Training time	Approximately 3 min/epoch

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

where TP represents correctly classified Pneumonia images, and FP represents Normal images incorrectly classified as Pneumonia. Recall (REC), also called sensitivity, measures the proportion of actual positive samples the model correctly identifies. This metric is particularly important in pneumonia detection because it reflects the model's ability to identify disease cases [3]. The computation of recall is presented in Eq. (5) [3].

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

where TP represents correctly classified Pneumonia images, and FN represents Pneumonia images incorrectly classified as Normal. F1-Score (F1) is the harmonic mean of Precision and Recall, providing a balanced evaluation by simultaneously considering false positive and false negative predictions [3]. The computation of the F1-score is presented in Eq. (6) [3].

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

where Precision and Recall are calculated using Eq. (4) and Eq. (5), respectively Since the dataset is imbalanced, Macro Precision, Macro Recall, and Macro F1-Score were additionally reported to evaluate the average performance across all classes by assigning equal importance to each class regardless of its sample size [3].

$$\text{Macro Precision} = \frac{\sum_{i=1}^{C} \text{Precision}_i}{C} \tag{7}$$

where Precision_i denotes the Precision of class i, and C represents the total number of classes.

$$\text{Macro Recall} = \frac{\sum_{i=1}^{C} \text{Recall}_i}{C} \tag{8}$$

where Recall_i denotes the Recall of class i, and C represents the total number of classes.

$$\text{Macro F1} = \frac{\sum_{i=1}^{C} \text{F1}_i}{C} \tag{9}$$

where F1_i denotes the F1-Score of class i, and C represents the total number of classes. In addition to confusion matrix-based metrics, ROC-AUC analysis was performed to evaluate the model's ability to distinguish between Normal and Pneumonia classes across different classification thresholds. Unlike single-threshold metrics, ROC-AUC provides a threshold-independent assessment of classification performance, making it suitable for imbalanced datasets. A higher AUC value indicates better discriminative capability [6]. To determine the optimal classification threshold, this study employed Youden's J-Statistic using the True Positive Rate (TPR) and False Positive Rate (FPR) as follows [18].

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{10}$$

where TP denotes true positive predictions, and FN denotes false negative predictions. TPR represents the proportion of positive-class samples correctly identified by the model and corresponds to sensitivity or recall.

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{11}$$

where FP denotes false positive predictions, and TN denotes true negative predictions. FPR represents the proportion of negative-class samples incorrectly classified as positive.

Youden's J-Statistic was then calculated as [18]:

$$\text{J} = \text{TPR} - \text{FPR} \tag{12}$$

where J denotes Youden's J-Statistic, TPR represents the true positive rate, and FPR represents the false positive rate. The optimal threshold was determined by selecting

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

Table 6. Algorithm for selecting the optimal classification threshold using Youden's J-Statistic.

Youden's J-Statistic
1. **BEGIN**
2. Input = Validation_Set Model = Trained ResNet50-CBAM
3. Predictions = model.predict(Validation_Set)
4. FOR each Candidate_Threshold t DO Convert Predictions to Binary Labels Calculate Sensitivity(t) Calculate Specificity(t) J(t) = Sensitivity(t) + Specificity(t) − 1 END FOR
5. Optimal_Threshold = threshold with maximum J(t)
6. Apply Optimal_Threshold to test predictions
7. OUTPUT Optimal_Threshold
8. **END**

the threshold that maximized J using the validation set only. The selected threshold was subsequently applied to the independent test set for final performance evaluation, thereby preventing optimistic performance estimation [18]. The procedure is summarized in Table 6 [18].

III. RESULTS

A. Overview of Experimental Results

This study evaluated four imbalance-handling strategies within the ResNet50-CBAM framework: Baseline (M1), Augmentation Only (M2), Class Weighting Only (M3), and Hybrid (M4). The results showed that the effects of the strategies varied across evaluation metrics. The Hybrid model achieved the highest overall performance, with an Accuracy of 95.74%, ROC-AUC of 98.83%, Macro Precision of 96.01%, and Macro F1-Score of 94.44%. Compared with the Baseline model, the Hybrid strategy increased Accuracy from 93.70% to 95.74%, corresponding to an absolute improvement of 2.04 percentage points and a relative improvement of 2.18%. Macro F1-Score increased from 92.22% to 94.44%, representing an absolute gain of 2.22 percentage points and a relative improvement of 2.41%. ROC-AUC also

Table 7. Performance evaluation results of the Baseline ResNet50-CBAM model without imbalance handling strategies.

Class	Precision	Recall	F1-Score
Normal	0.85465	0.92453	0.88822
Pneumonia	0.97108	0.94159	0.95611
Accuracy			0.93697
ROC-AUC			0.98384

Table 8. Performance evaluation results of the ResNet50-CBAM model using image augmentation only.

Class	Precision	Recall	F1-Score
Normal	0.90385	0.88679	0.89524
Pneumonia	0.95824	0.96495	0.96158
Accuracy			0.94378
ROC-AUC			0.98257

increased from 98.38% to 98.83%, an absolute gain of 0.45 percentage points. These results indicate that the combined strategy provided the strongest overall performance, although improvements were not uniform across all metrics.

B. Performance of Individual Models

To evaluate each imbalance-handling strategy in detail, the performance of the Baseline, Augmentation Only, Class Weighting Only, and Hybrid models was analyzed individually. The corresponding evaluation results are summarized in Tables 7-10. The Baseline model achieved an Accuracy of 93.70% and a ROC-AUC of 98.38%, providing a reference point for evaluating the effects of the imbalance-handling strategies. The model achieved a Macro Precision of 91.29%, Macro Recall of 93.31%, and Macro F1-Score of 92.22%. The detailed class-wise results are presented in Table 7. The Augmentation Only model achieved an Accuracy of 94.38%, representing an absolute improvement of 0.68 percentage points and a relative improvement of 0.73% over the Baseline. Its Macro Precision and Macro F1-Score also increased to 93.10% and 92.84%, respectively. However, ROC-AUC decreased slightly from 98.38% to 98.26%, indicating that the augmentation strategy improved several threshold-dependent metrics but did not improve overall ranking discrimination. The Class Weighting Only model achieved an Accuracy of 94.21%, corresponding to an absolute improvement of 0.51 percentage points and a relative improvement of 0.55% compared with the Baseline. Its Macro Recall increased from 93.31% to 94.84%, an absolute gain of 1.54 percentage points, while the Macro F1-Score increased from 92.22% to 92.96%. ROC-AUC also increased to 98.80%, indicating improved discrimination compared with the Baseline. The detailed performance metrics of this model are presented in Table 9. The Hybrid model achieved the highest performance among the four scenarios, with an Accuracy of 95.74%, ROC-AUC of 98.83%, Macro Precision of 96.01%, Macro Recall of 93.13%, and Macro F1-Score of 94.44%. Compared with the Baseline, the Hybrid model improved Accuracy by 2.04 percentage points (2.18%) and Macro F1-Score by 2.22 percentage points (2.41%). ROC-AUC increased by 0.45 percentage points, while Macro Precision increased by 4.72 percentage points. The complete evaluation results are presented in Table 10.

C. Threshold Optimization

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

Table 9. Performance evaluation results of the ResNet50-CBAM model using class weighting only.

Class	Precision	Recall	F1-Score
Normal	0.84530	0.96226	0.90000
Pneumonia	0.98522	0.93458	0.95923
Accuracy			0.94208
ROC-AUC			0.98798

Table 10. Performance evaluation results of the Hybrid ResNet50-CBAM model using image augmentation and class weighting.

Class	Precision	Recall	F1-Score
Normal	0.96528	0.87421	0.91749
Pneumonia	0.95485	0.98832	0.97130
Accuracy			0.95741
ROC-AUC			0.98830

To obtain a more balanced decision boundary, the optimal probability threshold for each experimental scenario was determined using Youden's J-Statistic on the validation set, as described in the previous section. The threshold that maximized Youden's J value was subsequently applied during the evaluation of the independent test set. This procedure ensured the test data were not involved in threshold selection, preventing optimistic performance estimates. As presented in Table 11, the optimal thresholds differed across the four experimental scenarios, indicating that each imbalance-handling strategy produced a distinct probability distribution during classification. The Class Weighting Only model yielded the lowest optimal threshold (0.65492), suggesting that a lower decision boundary was required to balance sensitivity and specificity after incorporating class-dependent penalties during training. In contrast, the Hybrid model produced the highest optimal threshold (0.90341), indicating stronger confidence in positive predictions after combining image augmentation and class weighting. These findings demonstrate that the optimal decision threshold is model-dependent and should be determined individually rather than adopting the default threshold of 0.5 for all classification scenarios.

Table 11. Optimal probability thresholds obtained using Youden's J-Statistic.

Model	Optimal Threshold
Baseline (M1)	0.84328
Augmentation Only (M2)	0.82250
Class Weighting Only (M3)	0.65492
Hybrid (M4)	0.9034

D. Comparative Performance Analysis

Fig. 6 compares the performance of the four experimental scenarios across Accuracy, ROC-AUC, Macro Precision, Macro Recall, and Macro F1-Score. The Hybrid model achieved the highest Accuracy, ROC-AUC, Macro Precision, and Macro F1-Score, whereas the Class Weighting Only model achieved the highest Macro Recall. Relative to the Baseline, the Hybrid model increased Accuracy from 93.70% to 95.74%, corresponding to an

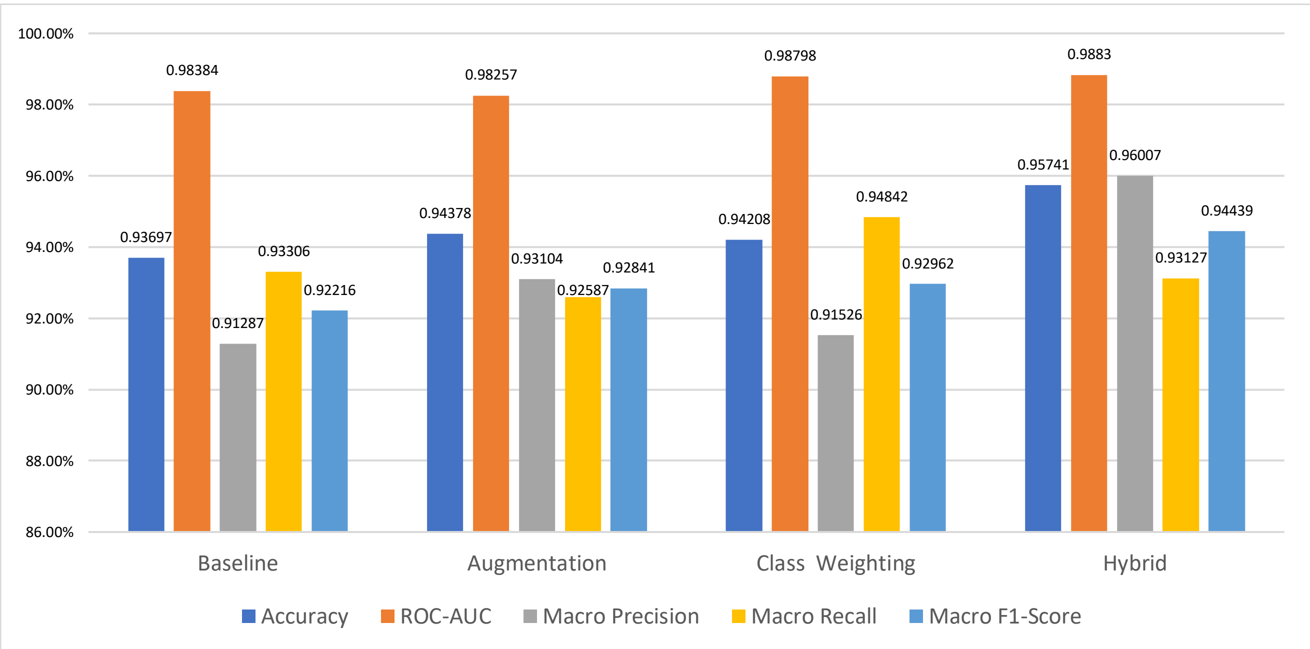

Fig. 6. Comparative performance of four experimental models: Baseline (M1), Augmentation Only (M2), Class Weighting Only (M3), and Hybrid (M4) across multiple evaluation metrics.

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics e-ISSN: 2656-8624
Homepage: https://ijeeemi.org/; Vol. 8, No. 3, pp. 365-379, August 2026

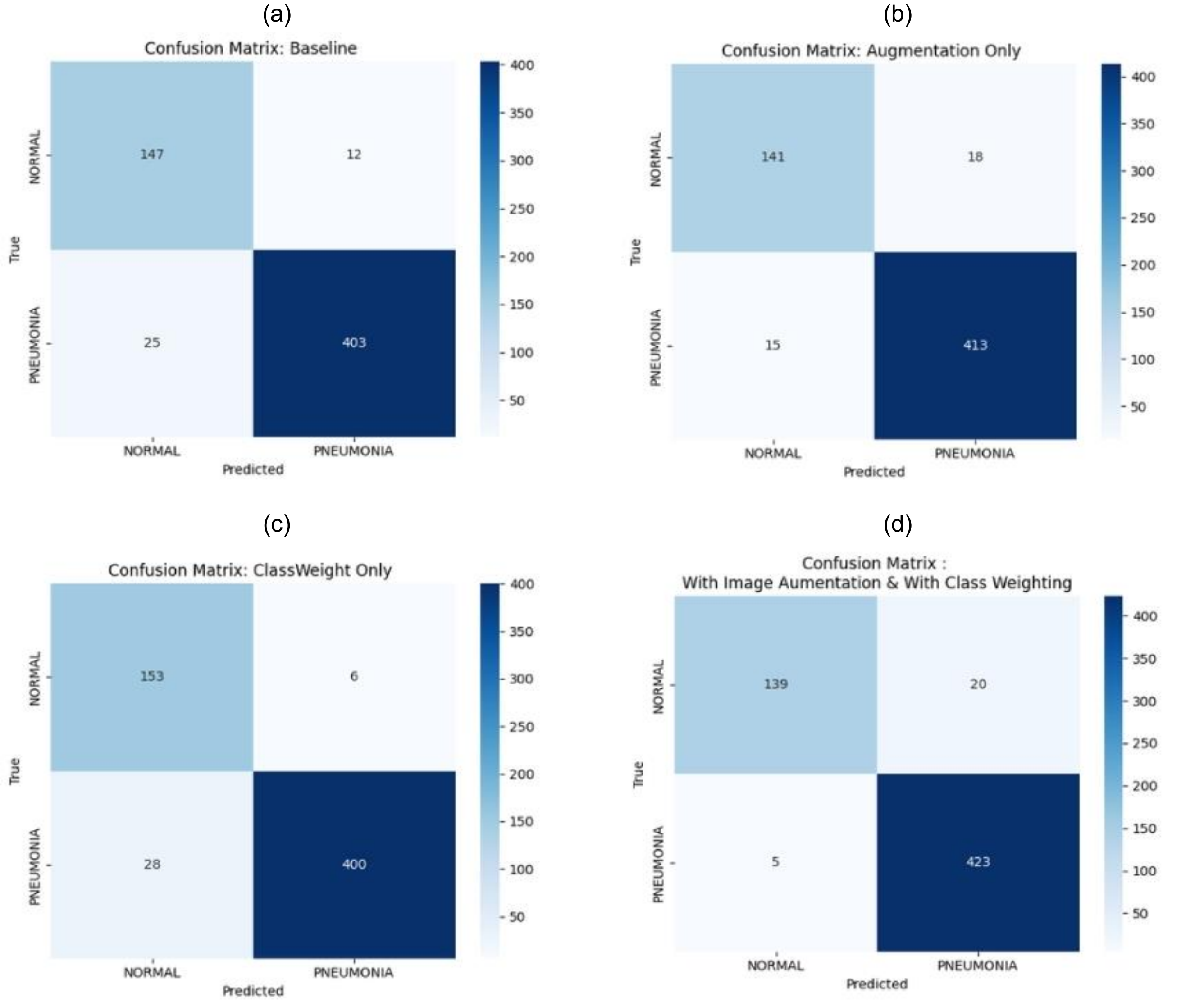

Fig. 7. Confusion matrices of the four ResNet50-CBAM models: (a) Baseline (M1), (b) Augmentation Only (M2), (c) Class Weighting Only (M3), and (d) Hybrid (M4).

absolute gain of 2.04 percentage points and a relative improvement of 2.18%. Macro F1-Score increased from 92.22% to 94.44%, yielding an absolute gain of 2.22 percentage points and a relative improvement of 2.41%. Macro Precision showed the largest increase, from 91.29% to 96.01%, corresponding to an absolute gain of 4.72 percentage points and a relative improvement of 5.17%. The improvement was not consistent across every metric. The Augmentation Only model increased Accuracy by 0.68 percentage points, but its ROC-AUC decreased by 0.13 percentage points relative to the Baseline. In contrast, Class Weighting Only increased ROC-AUC by 0.41 percentage points and achieved the highest Macro Recall (94.84%). These differences indicate that augmentation and class weighting affected different aspects of classification performance rather than producing uniform improvements across all metrics.

E. Confusion Matrix Analysis

The confusion matrices in Fig. 7 provide a detailed comparison of the prediction outcomes obtained by the four experimental models. Overall, the effects of the imbalance-handling strategies on Pneumonia detection varied across the experimental models. Augmentation Only and Hybrid reduced false negative predictions compared with the Baseline, whereas Class Weighting Only increased the number of false negatives. However, these improvements in disease detection came with different trade-offs in false positive predictions. The Baseline (M1) model correctly classified 147 Normal and 403 Pneumonia images, while producing 12 false positives and 25 false negatives. The relatively high number of false negative cases indicates that several Pneumonia images were incorrectly classified as Normal, which may reduce the effectiveness of automated screening because missed Pneumonia cases could delay further clinical assessment. The Class Weighting Only (M3) model reduced the number of false positives from 12 to 6, resulting in improved recognition of the Normal class. However, false negatives slightly increased to 28, indicating that emphasizing balanced learning alone was insufficient to improve Pneumonia detection.

The Augmentation Only (M2) model substantially

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

reduced false negative predictions to 15, demonstrating that increased image diversity improved the model's ability to recognize Pneumonia patterns. Nevertheless, this improvement was accompanied by an increase in false positives (18), indicating a greater tendency to classify Normal images as Pneumonia. Among all evaluated models, the Hybrid (M4) model achieved the lowest number of false negatives (5), representing an absolute reduction of 20 cases and a relative reduction of 80.00% compared with the Baseline model. However, false positives increased from 12 to 20, representing an absolute increase of 8 cases or 66.67%. Thus, the Hybrid strategy substantially reduced missed Pneumonia cases while producing more false-positive predictions. From a clinical perspective, these results reflect a sensitivity-specificity trade-off. Reducing false negatives is particularly important because missed Pneumonia cases may delay further clinical assessment. Conversely, increased false positives may result in additional examinations or referrals. Therefore, within the evaluated test set, the Hybrid model demonstrated the most favorable trade-off for Pneumonia detection, particularly through its substantial reduction in false-negative predictions. These findings should nevertheless be interpreted as internal validation results and should not be considered evidence of clinical effectiveness without further external and prospective validation.

IV. Discussion

The results demonstrate that image augmentation and class weighting produced different effects across the evaluated metrics. The Hybrid model achieved the best overall performance, with 95.74% Accuracy, 98.83% ROC-AUC, 96.01% Macro Precision, and 94.44% Macro F1-Score. Compared with the Baseline, Accuracy and Macro F1-Score increased by 2.04 and 2.22 percentage points, respectively. However, improvements were not consistent across all metrics, as the Augmentation Only model showed a slightly lower ROC-AUC than the Baseline, while Class Weighting Only achieved the highest Macro Recall. The confusion matrices further demonstrate the trade-off between false negatives and false positives. The Hybrid model reduced false negatives from 25 to 5 cases (an 80.00% reduction) but increased false positives from 12 to 20 cases. In contrast, Class Weighting Only reduced false positives to 6 cases but increased false negatives to 28. These results indicate that the Hybrid strategy primarily improved Pneumonia detection, whereas class weighting alone was more effective in reducing false-positive predictions.

As shown in Table 12, the proposed method achieved higher Accuracy than Usman et al. [11] (90.50%), Al Reshan et al. [30] (92.05%), Sharma et al. [3](92.15%), and Azhar et al. [19] (94.00%), while remaining slightly below Yanar et al. [21] (96.00%). The proposed method also achieved the highest reported AUC in the table (98.83%), exceeding Al Reshan et al. [30] (98.00%) and Sharma et al. [3] (97.40%). For Sensitivity, the proposed method (93.12%) was lower than Usman et al. [11] (99.05%) and Azhar et al. [19] (95.00%), but higher than Al Reshan et al. [30] (91.73%) and Yanar et al. [21] (91.00%), and was comparable to Sharma et al. [3] (93.08%).

These comparisons should be interpreted cautiously because the reported studies may differ in data splitting, augmentation, threshold selection, and evaluation protocols. In this study, the dataset was divided at the image level, the threshold was optimized using the validation set, and the selected threshold was subsequently applied to the test set. Patient-level independence could not be guaranteed because patient identifiers were unavailable. Therefore, the results represent internal validation under the present experimental protocol rather than a direct ranking of

Table 12. Comparison of the proposed method with previous studies on pneumonia classification using chest X-ray images

Study	Dataset	Method	Accuracy(%)	AUC(%)	Sensitivity(%)
Usman et al. [11]	Chest X-Ray Images (Pneumonia) (Kaggle)	EfficientNet-based Deep Transfer Learning	90,5%	72,53%	99,05%
Reshan et al. [30]	Chest X-Ray Images (Pneumonia) (Kaggle)	MobileNet	92,05%	98%	91.73%
Sharma et al. [3]	Chest X-Ray Images (Pneumonia) (Kaggle)	VGG-16 & Neural Network	92,15%	97,4%	93,08%
Azhar et al. [19]	Chest X-Ray Images (Pneumonia) (Kaggle)	ResNet50v2 (Pre-trained dari ImageNet)	94,0%	-	95%
Yanar et al. [21]	Chest X-Ray Images (Pneumonia) (Kaggle)	PELM Ensemble (InceptionV3, VGG16, ResNet50, ViT)	96,0%	91%	91%
Proposed method	**Kaggle**	**Image Augmentation + Class Weighting, ResNet50-CBAM**	**95,74%**	**98,83%**	**93,12%**

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

model performance. The proposed framework also differs in computational design. Yanar et al. [21] used an ensemble of InceptionV3, VGG16, ResNet50, and ViT, whereas this study uses a single ResNet50-CBAM backbone with augmentation and class weighting. Thus, the slightly lower Accuracy of the proposed method should be considered alongside its simpler single-backbone design. The main contribution of this study is therefore not architectural novelty, but the controlled comparison of augmentation and class weighting under consistent experimental conditions. Several limitations remain. The study used a single public dataset and an image-level split with one fixed random seed; therefore, it did not assess external validity or run-to-run variability. In addition, the model was evaluated only for binary Normal-Pneumonia classification. Although the Hybrid model substantially reduced false negatives, its clinical utility cannot yet be established because external, prospective, patient-level, and radiologist validation were not performed.

Overall, the findings indicate that combining image augmentation and class weighting provided the strongest performance among the evaluated strategies, achieving 95.74% Accuracy, 98.83% ROC-AUC, 94.44% Macro F1-Score, and only five false negatives. These findings are limited to the internal validation setting and require further validation on independent, clinically representative datasets. Overall, this study provides both methodological and practical implications. Methodologically, the findings demonstrate that integrating data-level and algorithm-level imbalance-handling strategies within an attention-based convolutional neural network effectively improves pneumonia classification under imbalanced data conditions. More importantly, the controlled comparative framework provides clearer evidence regarding the individual contribution of each imbalance-handling strategy while minimizing the influence of differences in network architecture and training configuration. From a practical perspective, the proposed ResNet50-CBAM framework offers a computationally efficient alternative for automated pneumonia screening. The substantial reduction in false negative predictions achieved by the Hybrid model indicates improved disease detection capability while maintaining competitive overall classification performance, suggesting its potential to support radiologists as a computer-aided diagnostic tool rather than replacing clinical decision-making.

V. Conclusion

This study systematically evaluated the individual and combined effects of image augmentation and class weighting for addressing class imbalance in pneumonia classification using chest X-ray images within a ResNet50-CBAM framework. Four experimental scenarios, namely Baseline, Augmentation Only, Class Weighting Only, and Hybrid, were evaluated under consistent datasets, preprocessing procedures, network architecture, training configuration, and evaluation protocols to enable a controlled comparison. The Hybrid model achieved the best overall performance among the evaluated scenarios, obtaining an Accuracy of 95.74%, ROC-AUC of 98.83%, and Macro F1-score of 94.44%. It also achieved a Macro Precision of 96.01% and Macro Recall of 93.13%. Notably, the number of false negative predictions decreased from 25 cases in the Baseline model to 5 cases in the Hybrid model, representing an 80.00% reduction. However, this improvement was accompanied by an increase in false positive predictions, indicating a trade-off between detecting Pneumonia cases and avoiding false-positive classifications. The main contribution of this study is the controlled quantitative comparison of image augmentation and class weighting within the same ResNet50-CBAM framework. The findings indicate that combining these data-level and algorithm-level strategies provided the strongest overall performance under the evaluated experimental conditions. However, the results should be interpreted as internal validation findings and should not be considered evidence of clinical effectiveness or generalizability beyond the evaluated dataset. Future research should validate the proposed approach using independent external datasets and patient-level data splitting, as well as investigate repeated experimental runs, additional imbalance-handling strategies, and explainability methods such as Grad-CAM. Such validation is necessary to determine the reproducibility, generalizability, and potential clinical utility of the proposed framework.

Acknowledgement

The authors gratefully acknowledge the Department of Computer Science, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University, Banjarbaru, Indonesia, for its academic assistance and for providing the facilities and resources necessary to conduct this research. The authors also extend their appreciation for the continuous support and encouragement provided by the department throughout the completion of this study.

Funding

This research was conducted without receiving any specific financial support from public, commercial, or non-profit funding organizations.

Data Availability

The dataset used in this study was obtained from the publicly available Chest X-Ray Images (Pneumonia) dataset provided by Paul Mooney through Kaggle, which is accessible at: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia. No additional datasets were created for this research, and all experiments were performed exclusively using the images provided in the original dataset.

Author Contribution

Kafilah Akhmad Fatahillah was responsible for the conceptualization and methodological development of the study, carried out the experimental implementation and data analysis, and prepared the initial manuscript. Triando

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

Hamonangan Saragih provided research supervision and methodological direction, as well as critical assessment of the experimental design and findings. Dwi Kartini contributed to methodological validation and assessment of the experimental outcomes. Fatma Indriani contributed to the critical examination and refinement of the manuscript and provided constructive input regarding the methodology and presentation of the study. Irwan Budiman contributed to research supervision and critical evaluation of both the study and the manuscript. All authors participated in the revision of the manuscript, reviewed and approved its final version, and accepted responsibility for the integrity and accuracy of the work.

Funding

Ethical Approval

Ethical approval was not necessary for this research, as it relied on a publicly accessible secondary dataset containing de-identified chest X-ray images and involved no direct interaction, recruitment, or intervention with human participants.

Consent for Publication

Consent for publication was not required because the study employed publicly available, de-identified chest X-ray data without any personally identifiable information.

Competing Interests

The authors report no conflicts of interest or competing interests related to this research.

References

[1] I. S. Masad, A. Alqudah, A. M. Alqudah, and S. Almashaqbeh, “A hybrid deep learning approach towards building an intelligent system for pneumonia detection in chest x-ray images,” *Int. J. Electr. Comput. Eng.*, vol. 11, no. 6, pp. 5530–5540, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5530-5540.

[2] J. Poovieng, B. Sakboonyarat, and W. Nasomsong, “Bacterial etiology and mortality rate in community-acquired pneumonia, healthcare-associated pneumonia and hospital-acquired pneumonia in Thai university hospital,” *Sci. Rep.*, vol. 12, no. 1, article 9004, 2022, doi: 10.1038/s41598-022-12904-z.

[3] 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 Comput. Sci.*, vol. 218, pp. 357–366, 2023, doi: 10.1016/j.procs.2023.01.018.

[4] M. Pal *et al.*, “A comparative analysis of the binary and multiclass classified chest X-ray images of pneumonia and COVID-19 with ML and DL models,” *Open Med.*, vol. 20, no. 1, article 20241110, 2025, doi: 10.1515/med-2024-1110.

[5] N. Sebastian and B. Ankayarkanni, “Enhanced ResNet-50 with Multi-Feature Fusion for Robust Detection of Pneumonia in Chest X-Ray Images,” *Diagnostics*, vol. 15, no. 16, article 2041, 2025, doi: 10.3390/diagnostics15162041.

[6] E. Chamseddine, N. Mansouri, M. Soui, and M. Abed, “Handling class imbalance in COVID-19 chest X-ray images classification: Using SMOTE and weighted loss,” *Appl. Soft Comput.*, vol. 129, article 109588, 2022, doi: 10.1016/j.asoc.2022.109588.

[7] M. Mujahid, F. Rustam, R. Álvarez, J. L. Vidal Mazón, I. de la Torre Díez, and I. Ashraf, “Pneumonia classification from X-ray images with Inception-V3 and convolutional neural network,” *Diagnostics*, vol. 12, no. 5, article 1280, 2022, doi: 10.3390/diagnostics12051280.

[8] D. Zhang, F. Ren, Y. Li, L. Na, and Y. Ma, “Pneumonia detection from chest x-ray images based on convolutional neural network,” *Electronics*, vol. 10, no. 13, article 1512, 2021, doi: 10.3390/electronics10131512.

[9] Nisha, “Deep learning-based medical image classification using CNN and attention mechanisms for disease detection and diagnosis,” *Int. J. Comput. Artif. Intell.*, vol. 6, no. 1, pp. 187–194, 2025, doi: 10.33545/27076571.2025.v6.i1c.153.

[10] C. Shorten, T. M. Khoshgoftaar, and B. Furht, “Deep Learning applications for COVID-19,” *J. Big Data*, vol. 8, article 18, 2021, doi: 10.1186/s40537-020-00392-9.

[11] O. L. Usman, O. O. Adeusi, M. A. Kareem, A. A. Owoade, and R. C. Muniyandi, “Qualitative Study on Impact of EfficientNet-Based Deep Transfer Learning Model for Pneumonia Detection with Explainable Artificial Intelligence using Chest Radiographs,” *Zamfara Int. J. Educ.*, vol. 3, no. 1, pp. 45–56, 2026, doi: 10.64348/zije.2026344.

[12] Y. Liu, F. Liu, S. Tu, S. Liu, and B. Han, “Attention enhanced residual network for automatic pulmonary tuberculosis detection on chest radiographs images,” *Digit. Signal Process.*, vol. 159, article 104975, 2025, doi: 10.1016/j.dsp.2024.104975.

[13] Y. Wang, L. Liu, and C. Wang, “Trends in using deep learning algorithms in biomedical prediction systems,” *Front. Neurosci.*, vol. 17, article 1256351, 2023, doi: 10.3389/fnins.2023.1256351.

[14] M. A. Saleem *et al.*, “Enhancing stroke risk prediction through class balancing and data augmentation with CBDA-ResNet50,” *Sci. Rep.*, vol. 15, no. 1, article 24553, 2025, doi: 10.1038/s41598-025-07350-6.

[15] M. F. Hashmi, S. Katiyar, A. W. Hashmi, and A. G. Keskar, “Pneumonia detection in chest X-ray images using compound scaled deep learning model,” *Automatika*, vol. 62, no. 3–4, pp. 397–406, 2021, doi: 10.1080/00051144.2021.1973297.

[16] R. Kundu, R. Das, Z. W. Geem, G. T. Han, and R. Sarkar, “Pneumonia detection in chest X-ray images using an ensemble of deep learning models,” *PLoS One*, vol. 16, no. 9, article e0256630, 2021, doi: 10.1371/journal.pone.0256630.

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

[17] Y. Liu *et al.*, "CI-TransCNN: A class imbalance hybrid CNN-Transformer Network for facial attribute recognition," *Image Vis. Comput.*, vol. 165, article 105823, 2026, doi: 10.1016/j.imavis.2025.105823.

[18] S. Liu, Q. Tian, Y. Liu, and P. Li, "Joint Statistical Inference for the Area under the ROC Curve and Youden Index under a Density Ratio Model," *Mathematics*, vol. 12, no. 13, article 2118, 2024, doi: 10.3390/math12132118.

[19] Y. Azhar, W. Priyo Wicaksono, and Z. Sari, "Pneumonia Diagnosis Through Deep Learning: ResNet50V2 Model Implementation," *J. Nas. Pendidik. Tek. Inform.*, vol. 13, no. 2, pp. 349–358, 2024, doi: 10.23887/v13i2.72068.

[20] A. Kareem, H. Liu, and V. Velisavljevic, "A federated learning framework for pneumonia image detection using distributed data," *Healthcare Anal.*, vol. 4, article 100204, 2023, doi: 10.1016/j.health.2023.100204.

[21] E. Yanar, F. Hardalaç, and K. Ayturan, "PELM: A Deep Learning Model for Early Detection of Pneumonia in Chest Radiography," *Appl. Sci.*, vol. 15, no. 12, article 6487, 2025, doi: 10.3390/app15126487.

[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, article 176, 2024, doi: 10.3390/jimaging10080176.

[23] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A Comprehensive Survey of Image Augmentation Techniques for Deep Learning," *Pattern Recognit.*, vol. 137, article 109347, 2023, doi: 10.1016/j.patcog.2023.109347.

[24] A. S. Buriboev, D. Muhamediyeva, H. Primova, D. Sultanov, K. Tashev, and H. S. Jeon, "Concatenated CNN-Based Pneumonia Detection Using a Fuzzy-Enhanced Dataset," *Sensors*, vol. 24, no. 20, article 6750, 2024, doi: 10.3390/s24206750.

[25] Y. Chen *et al.*, "Generative Adversarial Networks in Medical Image augmentation: A review," *Comput. Biol. Med.*, vol. 144, article 105382, 2022, doi: 10.1016/j.compbiomed.2022.105382.

[26] M. M. Musthafa M, T. R. Mahesh, V. Kumar V, and S. Guluwadi, "Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50," *BMC Med. Imaging*, vol. 24, no. 1, article 107, 2024, doi: 10.1186/s12880-024-01292-7.

[27] B. Bakirarar and I. S. Yilmaz, "A New Class-Weighting Formulation for the Class Imbalance Problem: A Methodological Research," *Turkiye Klinikleri J Biostat.*, vol. 15, no. 2, pp. 79–90, 2023, doi: 10.5336/biostatic.2023-96293.

[28] E. Sogancioglu, E. Çallı, B. van Ginneken, K. G. van Leeuwen, and K. Murphy, "Deep Learning for Chest X-ray Analysis: A Survey," *Med. Image Anal.*, vol. 72, article 102125, 2021, doi: 10.1016/j.media.2021.102125.

[29] A. Kareem, H. Liu, and P. Sant, "Review on Pneumonia Image Detection: A Machine Learning Approach," *Hum.-Cent. Intell. Syst.*, vol. 2, pp. 31–43, 2022, doi: 10.1007/s44230-022-00002-2.

[30] M. S. Al Reshan *et al.*, "Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model," *Healthcare*, vol. 11, no. 11, article 1561, 2023, doi: 10.3390/healthcare11111561.

Author Biography

Kafilah Akhmad Fatahillah is an undergraduate student in the Department of Computer Science, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University, Indonesia. His research interests include artificial intelligence, deep learning, computer vision, medical image analysis, and intelligent healthcare applications. His current research focuses on pneumonia classification from chest X-ray images using a ResNet50-CBAM architecture integrated with image augmentation and class weighting techniques. Through this work, he investigates effective strategies for handling class imbalance and improving classification performance. He is committed to developing reliable artificial intelligence solutions for medical image analysis and healthcare applications. He can be contacted at email: 2211016210021@mhs.ulm.ac.id

Triando Hamonangan Saragih is a lecturer in the Computer Science Department, Faculty of Mathematics and Natural Science, Lambung Mangkurat University in Banjarbaru, where he is actively engaged in academic and research activities, with a strong focus on the broad field of Data Science. He earned his bachelor's degree in Informatics from Brawijaya University, Malang, in 2016, which laid the foundation for his subsequent academic trajectory. He later pursued and completed a master's degree in Computer Science at the same institution in 2018, further strengthening his expertise. His current research interests remain centered on Data Science and its associated analytical methodologies. He can be contacted at triando.saragih@ulm.ac.id.

Dwi Kartini earned her Bachelor's and Master's degrees in Computer Science from the Faculty of Computer Science at Putra Indonesia "YPTK" in Padang, Indonesia. She is also a lecturer in the Department of Computer Science, where she teaches various subjects including linear algebra, discrete mathematics, research methods, and others. Her research interests focus on applications of Artificial Intelligence and Data Mining. Currently, she serves as an assistant professor in the Department of Computer Science, Faculty of Mathematics and Natural Sciences at Lambung

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375

Mangkurat University in Banjarbaru, Indonesia. She heads the Computer Science study program. She can be contacted at email: dwikartini@ulm.ac.id

Fatma Indriani is a lecturer in the Department of Computer Science at Lambung Mangkurat University, with a strong research interest in Data Science. Before pursuing an academic career, she completed her undergraduate studies in the Informatics Department at the Bandung Institute of Technology. In 2008, she began her career as a lecturer at Lambung Mangkurat University, contributing to the field of Computer Science through teaching and research. To further expand her expertise, she pursued a master's degree at Monash University, Australia, which she successfully completed in 2012. She continued her academic journey with a doctorate in Bioinformatics from Kanazawa University, Japan, which she completed in 2022. With a focus on both Data Science and Bioinformatics, she actively engages in research, exploring innovative ways to leverage data-driven technologies for scientific advancement. Her dedication to academia and research allows her to contribute significantly to knowledge development in her field, while also mentoring students and collaborating on interdisciplinary projects. She can be contacted via email at: f.indriani@ulm.ac.id.

Irwan Budiman is an esteemed academic and lecturer at Lambung Mangkurat University. He laid the foundation for his education by earning a Bachelor's Degree in Informatics Engineering from Universitas Islam Indonesia in Yogyakarta. He further advanced his expertise by completing a Master's degree in Information Systems at Diponegoro University, Semarang. His comprehensive research portfolio spans several critical domains, including data mining, human-computer interaction, applied business intelligence, and e-government system development. Through his scholarly work, he actively contributes to the evolution of information technology. He can be reached for academic collaboration via email at: irwan.budiman@ulm.ac.id

Manuscript received 13 June 2026; Revised 10 July 2026; Accepted 30 July 2026; Available online 27 August 2026
DOI: https://doi.org/10.35882/ijeeemi.v8i3.375