Applied Medical Informatics
Original Research Vol. 47, No.4/2025, pp -

Enhanced Pneumonia Detection from Chest X-Rays via VGG-16 and Self-Attention Mechanisms

Priyankar BISWAS[1], Sourav SANA[1], Anindya NAG[2], Sagar KUNDU[1]

[1] Department of Electrical and Electronic Engineering, Gopalganj Science and Technology University, Gopalganj-8105, Bangladesh.
[2] Department of Computer Science & Engineering, Northern University of Business and Technology Khulna, Khulna-9100, Bangladesh.
E-mail(s): priyankarbiswas03@gmail.com; souravsana1321@gmail.com; anindyanag@ieee.org; siddharthsagar257@gmail.com

* Author to whom correspondence should be addressed: https://orcid.org/0009-0001-9165-8101

Received: 18 August 2025/Accepted: 15 November 2025/ Published online: 12 December 2025

Abstract
Early and accurate detection of pneumonia on chest X-rays is critical for effective treatment, especially in resource-constrained healthcare settings. Manual diagnosis is time-consuming and prone to variations, underscoring the need for robust automated approaches. This study addressed the challenge of improving the diagnostic accuracy and interpretability of deep learning models for pneumonia detection using chest radiographs. The method proposes a novel deep-learning framework that combines transfer learning using a pre-trained VGG16 model with a self-attention-enhanced convolutional architecture. The VGG16 backbone extracts low-level visual features, whereas the self-attention mechanism highlights clinically relevant lung regions, improving spatial focus during classification. The proposed model leverages the VGG16 backbone to extract low-level visual features, whereas a self-attention mechanism enhances the spatial focus by emphasizing clinically significant lung regions. The VGG16 model, guided by attention, achieved a 97% accuracy, precision, and recall for pneumonia detection. In addition, Grad-CAM (Gradient-weighted Class Activation Mapping) visualizations improved interpretability and model performance compared with baseline CNNs (Convolutional Neural Networks) and pre-trained architectures. The integration of a self-attention mechanism into a transfer learning framework significantly improves both the performance and interpretability of pneumonia detection models using chest X-rays. This approach closely replicates the spatial reasoning of human experts and offers a scalable solution for clinical deployment. The results indicate that attention-enhanced deep learning architectures are well-suited for medical imaging tasks, particularly in resource-constrained settings where diagnostic expertise may be limited.

Keywords: Pneumonia Detection; Chest X-ray Images; VGG16 Architecture; Self-Attention Mechanism.

Introduction

Pneumonia, a severe inflammatory condition of the lungs, continues to pose a significant global health burden, particularly in resource-constrained settings. According to the World Health Organization, it remains a leading cause of mortality among children under the age of five and the elderly, particularly in low- and middle-income countries [1, 2]. Timely and accurate diagnosis is vital to reduce complications and improve clinical outcomes [3]. Chest X-ray imaging is widely employed as a non-invasive and cost-effective diagnostic modality; however, its reliability often hinges on the availability of skilled radiologists and is susceptible to inter-observer variability [4, 5]. Now, advances in artificial intelligence (AI) [6], particularly in deep learning (DL) [7], have significantly

improved medical image analysis by enabling automated high-accuracy diagnostic support. CNNs, known for their ability to learn hierarchical features from complex imaging data, have demonstrated considerable success in interpreting chest radiographs [8]. Furthermore, transfer learning, especially with pre-trained architectures such as VGG16 has become an essential strategy in medical imaging, as it allows for effective model training on limited annotated datasets while significantly reducing computational time and resource demands [9]. Despite its effectiveness, VGG16 has demonstrated strong performance in medical image classification tasks, and its standalone architecture often falls short of capturing the intricate spatial dependencies necessary for accurately localizing pneumonia-related abnormalities on chest radiographs. Traditional CNNs tend to apply uniform attention across the entire image, potentially overlooking critical pathological regions and thereby diminishing diagnostic precision [10, 11]. To address this challenge, Attention mechanisms, particularly self-attention, have emerged as powerful tools that enable neural networks to dynamically prioritize clinically significant features while attenuating irrelevant information. However, robust models that seamlessly integrate VGG16 with self-attention in the context of pneumonia detection while also offering interpretability for clinical decision-making are lacking [12, 13].

Our study presents an enhanced deep learning framework that integrates a fine-tuned VGG-16 backbone with a self-attention module to enhance both classification accuracy and model transparency. To further support explainability, Grad-CAM was employed, offering visual insight into the lung regions that most influence the model's diagnostic decisions, thereby reinforcing trust and facilitating clinical validation. The proposed method is evaluated on a publicly available chest X-ray dataset and benchmarked against leading models such as DenseNet169, MobileNet, InceptionV3, and Xception using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

Materials and Methods

The applied methodology was structured into six sequential stages: dataset preparation, backbone model configuration, attention mechanism design, model training, performance evaluation, and comparative benchmarking.

Dataset Preparation and Class Balancing

We utilized the publicly available CXRs Pneumonia dataset from Kaggle, which comprises 5,863 grayscale chest radiograph images categorized into two classes: NORMAL (1,583 images) and PNEUMONIA (4,280 images). The dataset presents two key challenges: a significant class imbalance (approximately 37% normal vs. 63% pneumonia), and an inadequately small validation set containing only 16 images. To address these issues, we adopted a systematic data restructuring approach. The dataset was re-partitioned into training (70%), validation (10%), and test (20%) subsets, while maintaining the original class distribution. This stratified split yielded 3,504 training images (1,499 normal and 2,005 pneumonia images), 501 validation images (215 normal and 286 pneumonia images), and 858 test images (368 normal and 490 pneumonia images). To counteract this imbalance in the training set, we implemented a duplication-based oversampling technique for the NORMAL class, increasing its sample size to match that of the PNEUMONIA class (2,005 images each). The validation and test sets were imbalanced to reflect the clinical prevalence of pneumonia in real-world settings. To enhance the generalizability of the model, data augmentation was applied to the training set using a medically informed ImageDataGenerator. The augmentation pipeline included ±10° rotation, ±10% zoom, ±10% width and height shifts, and horizontal flipping parameters chosen to introduce anatomically realistic variability while preserving critical diagnostic features. All images were normalized to the range [0, 1]. Augmentation was applied only during training to ensure an unbiased evaluation of the validation and test sets.

Transfer Learning Framework with VGG-16 Backbone

This adopted a transfer learning approach using the VGG-16 architecture, pre-trained on ImageNet, as the foundational model. VGG-16 was selected for its strong feature extraction capabilities and has been widely validated in medical imaging tasks. The model is initialized with include_top=False and an input_shape of (224 ×

224 × 3), aligned with the resolution of the chest X-ray images. A selective layer-freezing strategy was applied to balance computational efficiency and domain-specific learning. The first 10 convolutional layers were frozen to retain low-level; general-purpose feature detectors, whereas the remaining layers were unfrozen and fine-tuned to capture task-specific patterns relevant to pneumonia detection. The VGG-16 backbone processes the input through five sequential convolutional blocks, each reducing the spatial resolution and increasing the depth of feature representations. The output of the final convolutional layer, block5_conv3, yielded a feature map of size 7×7×512 pixels. This output was subsequently fed into a custom self-attention module, which enhanced the model's ability to focus on diagnostically significant regions. The integration of hierarchical convolutional features with spatial attention enables a more effective localization and interpretation of pneumonia-related abnormalities on chest radiographs.

Self-Attention Mechanism Implementation

The self-attention module was applied to the 512-channel feature map output by the final convolutional block of VGG-16. It begins with three parallel 1×1 convolutional layers that project the input features into three distinct embeddings: query and key (each with 64 channels), and value (with 512 channels). These projections capture spatial dependencies while preserving the channel-wise information. To compute attention, the query and key projections undergo matrix multiplication to generate a similarity matrix representing the relationships between all spatial positions, resulting in a 49×49 attention map (corresponding to the 7×7 spatial grid). This matrix is then normalized using a softmax function to produce attention weights. The value projection is then aggregated using these weights through weighted summation, producing context-enriched feature representations that emphasize spatial regions with high diagnostic relevance. A residual connection was added to retain the original features, modulated by a learnable scaling parameter γ, which was initialized to zero. This allows the model to progressively integrate attention-based refinement during training without destabilizing the early feature learning. By enabling the network to dynamically focus on critical areas, such as pulmonary infiltrates, the self-attention mechanism enhances interpretability and diagnostic accuracy while preserving the spatial structure of the chest radiograph.

Model Architecture and Training Protocol

The attention-refined features are passed through a carefully structured classification head:

- Global Average Pooling (GAP) to reduce spatial redundancy;
- Dense layer (256 neurons, ReLU) for nonlinear transformations;
- Dropout (rate = 0.5) to mitigate overfitting;
- Sigmoid output neuron for binary classification (*NORMAL* vs. *PNEUMONIA*);
- The model contains 16.9M total parameters, with ~4.7M trainable due to partial freezing. The model was trained using: Optimizer: Adam (learning rate = 1e-4), Loss: Binary cross-entropy, Epochs: 10, Batch size: 8 (to accommodate GPU memory constraints).

Performance Evaluation and Explainability

The model's performance was comprehensively assessed on a held-out test set containing 858 chest X-ray images; using a combination of quantitative evaluation and interpretability techniques. Quantitative analysis includes standard classification metrics, such as accuracy, precision, recall, and F1-score, which together offer a well-rounded view of the model's diagnostic performance. A confusion matrix was also utilized to analyze class-specific behavior, highlighting the distribution of true positives, true negatives, false positives, and false negatives. In addition, a Receiver Operating Characteristic (ROC) curve was generated, with the Area Under the Curve (AUC) providing a threshold-independent measure of the model's discriminative ability. To enhance clinical transparency, Gradient-weighted Class Activation Mapping (Grad-CAM) is applied to the block5_conv3 layer of the VGG-16 backbone. This technique produces visual heat maps that highlight the most influential regions in the radiograph contributing to each prediction, offering interpretability aligned with clinical diagnostic reasoning. For a more detailed understanding of the model's decision-making under different levels of certainty, test cases were grouped into four confidence intervals: 0–25%, 25–50%, 50–75%, and 75–100%. Representative examples from each interval illustrate how the attention of the model shifts with a varying prediction confidence. This integrated

evaluation framework provides rigorous performance validation and delivers clinically meaningful visual explanations. By aligning technical robustness with diagnostic transparency, this approach ensures that the model meets the critical requirements for reliable deployment in real-world medical settings.

The following performance metrics were reported:

- The accuracy of the model is defined as the proportion of correctly classified samples out of the total number of samples. It shows how often the model predicts the correct label.

Accuracy = (Number of Correct Predictions)/(Total Number of Predictions) (1)

- Precision measures how many of the samples predicted as a specific class were actually of that class. It focuses on the correctness of positive predictions.

Precision = (True Positives)/(True Positives + False Positives) (2)

High precision means that when the model predicts a sample as a certain class, it is more likely to be correct.

- Recall (also known as sensitivity) measures how many of the actual positive samples were correctly identified. It focuses on the model's ability to capture all relevant samples for a particular class. High recall means that the model is identifying most of the relevant samples. It is crucial in cases like medical diagnosis, where missing a positive case (false negative) can have serious consequences.

Recall = (True Positives)/(True Positives + False Negatives) (3)

- The F1-Score is the harmonic mean of precision and recall. It provides a balance between precision and recall, especially useful when the class distribution is imbalanced.

F1-Score= (Precision*Recall)/(Precision+Recall) (4)

If precision and recall are both high, the F1-Score will also be high.

Comparative Model Evaluation

To benchmark the proposed self-attention-enhanced VGG-16 model, its performance was compared with that of six established deep learning models: DenseNet-169, MobileNet, Inception-v3, Xception, EfficientNetB5, and ConvNextLarge. Each of these models was implemented under the same preprocessing, training, and evaluation conditions for fairness. The comparison included both quantitative performance metrics and qualitative Grad-CAM visualizations, highlighting the strengths and limitations of each architecture in pneumonia classification tasks. The proposed model demonstrated superior performance across multiple metrics and enhanced interpretability, validating its effectiveness in real-world diagnostic applications. Figure 1 illustrates a block diagram of the proposed pneumonia detection framework based on transfer learning with VGG-16 and a self-attention mechanism.

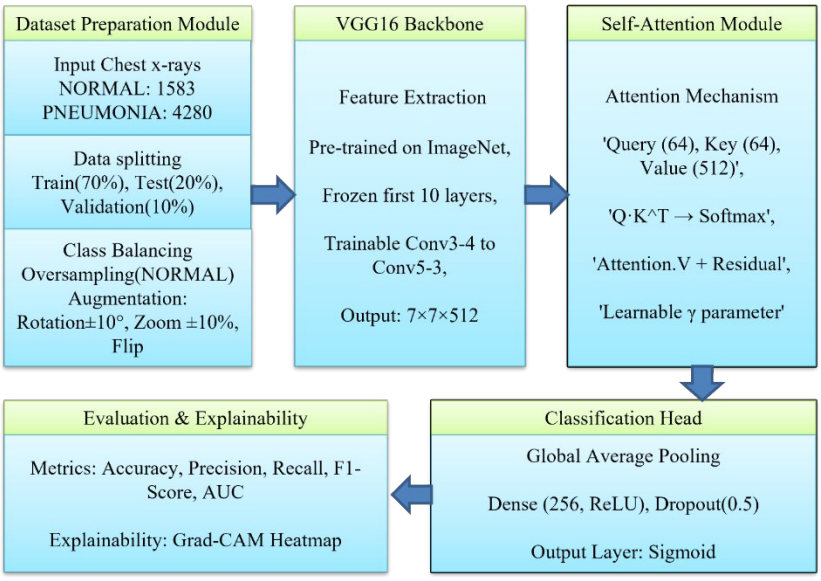


Figure 1. Overall Block Diagram of *VGG-16*-Based Model with Self-Attention for Pneumonia Classification.

Results

Table 1 summarizes the classification performances of all models, showing better performance of the VGG-16.

Table 1. Class-wise performance for all applied models.

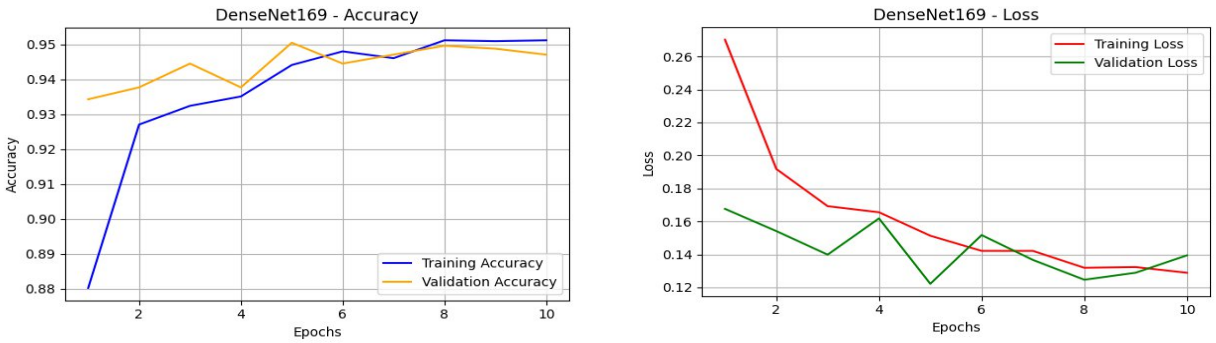
Applied Model	Class	Precision	Recall	F1-score	Average Accuracy
DenseNet-169	*NORMAL*	0.8418	0.9430	0.8896	0.9368
	PNEUMONIA	0.9779	0.9344	0.9557	
MobileNet	*NORMAL*	0.8793	0.9684	0.9217	0.9556
	PNEUMONIA	0.9878	0.9508	0.9690	
Inceptio-v3	*NORMAL*	0.8667	0.9051	0.8854	0.9368
	PNEUMONIA	0.9643	0.9485	0.9563	
Xception	*NORMAL*	0.8343	0.9241	0.8769	0.9299
	PNEUMONIA	0.9707	0.9321	0.9510	
EfficientNetB5	*NORMAL*	0.86	0.99	0.92	0.91
	PNEUMONIA	0.98	0.84	0.91	
ConvNextLarge	*NORMAL*	0.89	0.95	0.92	0.92
	PNEUMONIA	0.94	0.89	0.91	
VGG-16	*NORMAL*	0.93	0.96	0.94	0.97
	PNEUMONIA	0.98	0.97	0.98	

A comparative overview of the classification outcomes for each applied model, in terms of accuracy, precision, recall, and F1-score, is presented in Table 2.

Figure 2 presents the training and validation performance of all implemented deep learning models through accuracy and loss curves over 10 epochs.

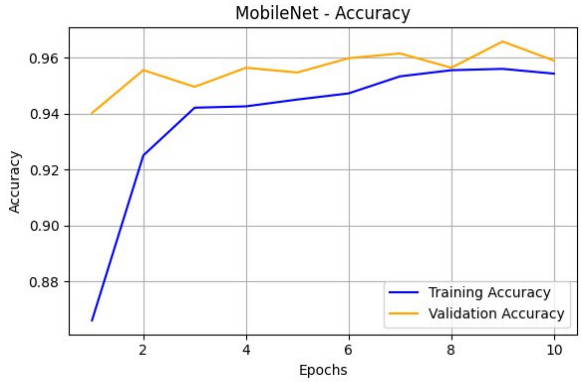
Table 2. Performance for all applied models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DenseNet-169	93.68	94.12	93.68	93.78
MobileNet	95.56	95.85	95.56	95.62
Inception-v3	93.68	93.79	93.68	93.72
Xception	92.99	93.39	92.99	93.10
EfficientNetB5	91.00	92.00	91.00	91.00
ConvNeXtLarge	92.00	92.00	92.00	92.00
VGG-16	**97.00**	**97.00**	**97.00**	**97.00**

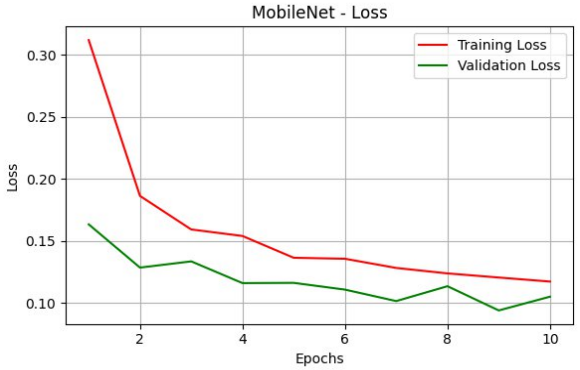


(a): Accuracy curve of DenseNet-169 (b): Loss curve of DenseNet-169

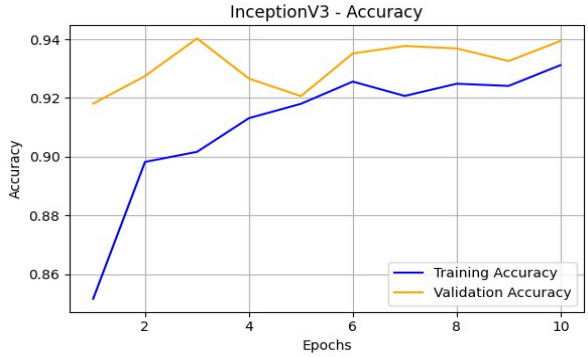
Figure 2. Accuracy and loss curve of all applied models.



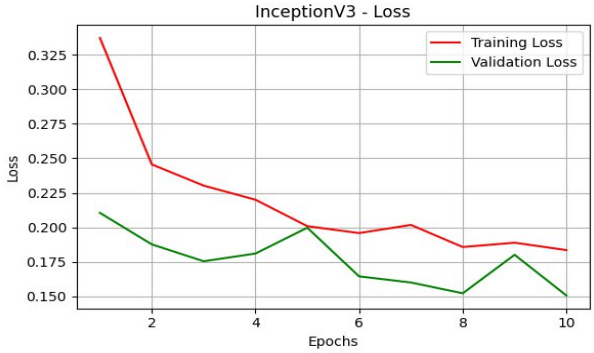
(c): Accuracy curve of MobileNet



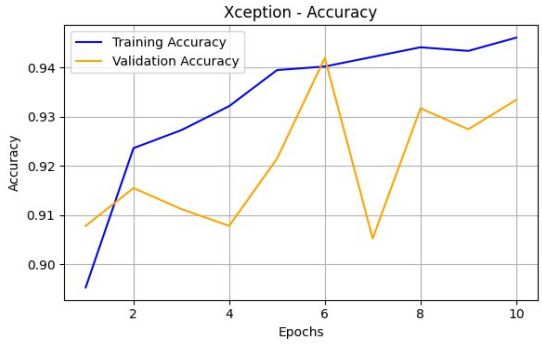
(d): Loss curve of MobileNet



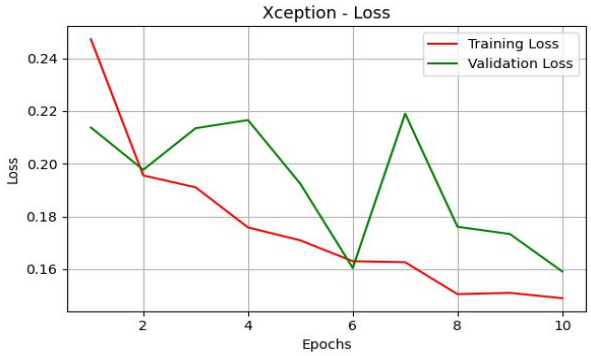
(e): Accuracy curve of Inception-v3



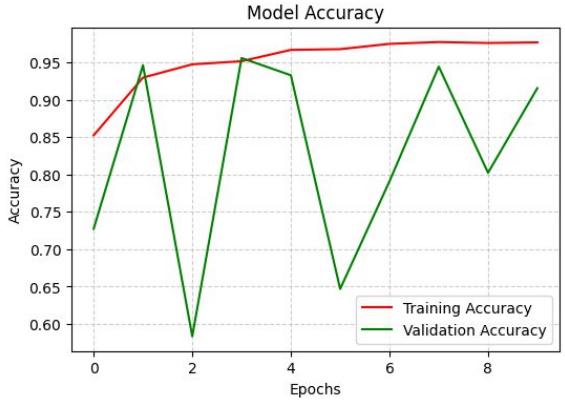
(f): Loss curve of Inception-v3



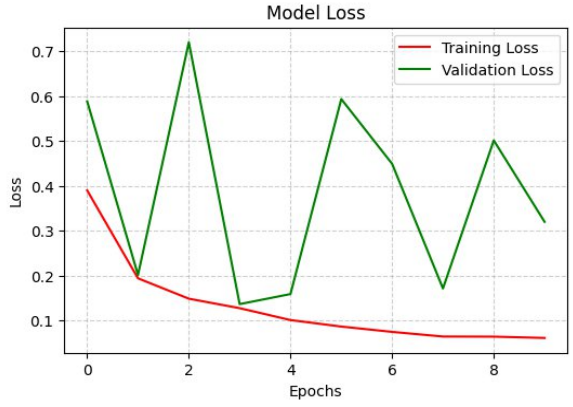
(g): Accuracy curve of Xception



(h): Loss curve of Xception

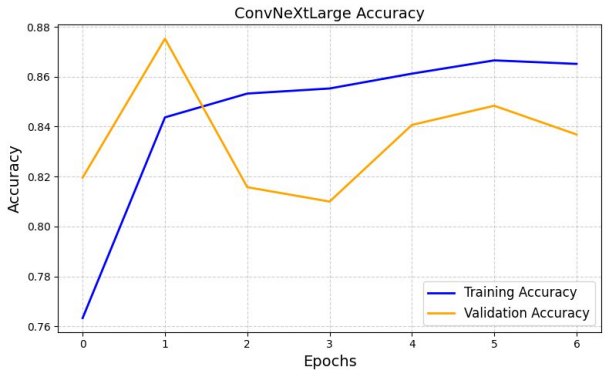


(i): Accuracy curve of EfficientNetB5

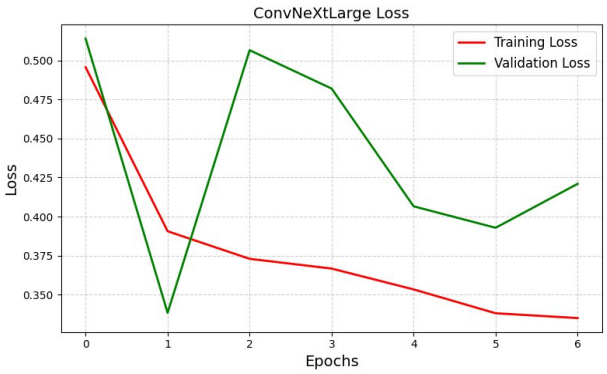


(j): Loss curve of EfficientNetB5

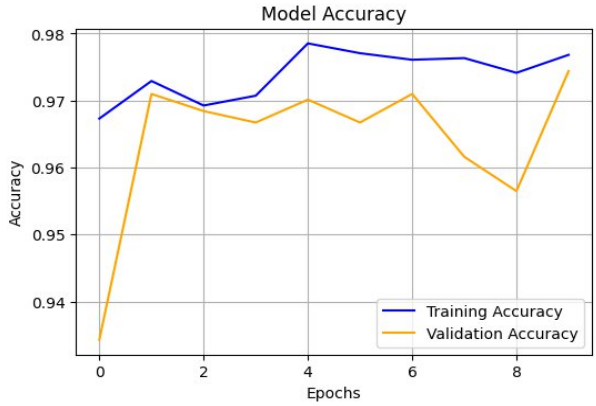
Figure 2. (continuation) Accuracy and loss curve of all applied models.



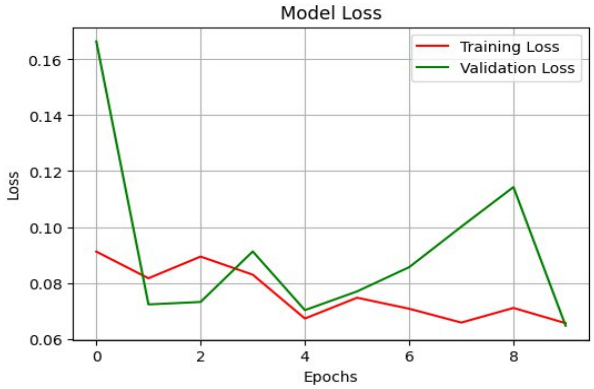
(k): Accuracy curve of ConvNextLarge



(l): Loss curve of ConvNextLarge



(m): Accuracy curve of *VGG-16*



(n): Loss curve of *VGG-16*

Figure 2. (continuation) Accuracy and loss curve of all applied models.

Figure 3 presents the Receiver Operating Characteristic (ROC) curve for the proposed model, which serves as a crucial metric for assessing the effectiveness of binary classification systems.

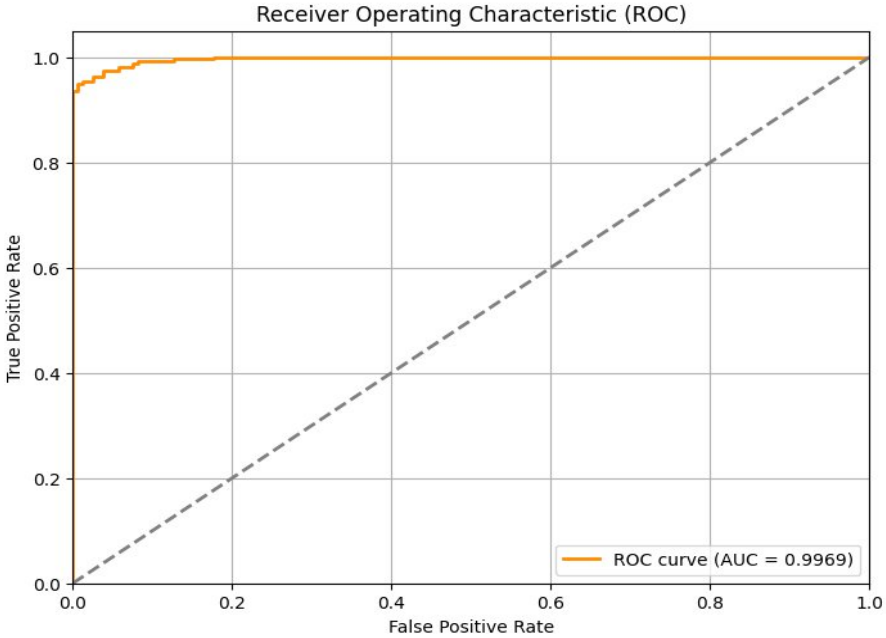
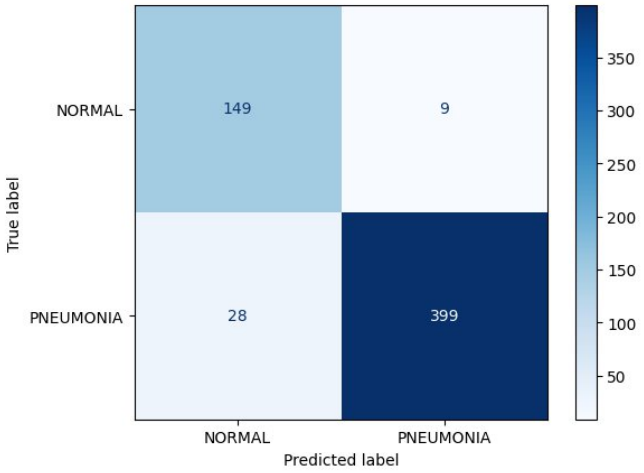


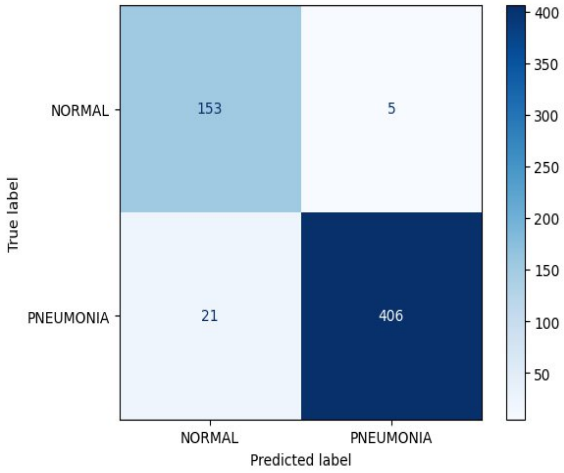
Figure 3. ROC curve of the proposed model.

The curve plots the true positive rate (TPR), or sensitivity, against the false positive rate (FPR), equivalent to 1 minus the specificity, across a range of classification thresholds. Each point on the curve represents a different balance between detecting true positives (pneumonia cases) and minimizing false positives (misclassified normal cases). The dashed diagonal line indicates the baseline performance of the random classifier. By contrast, the orange ROC curve for the proposed model lies significantly above this line, highlighting its superior performance. The AUC was calculated to be 0.9969, reflecting a near-perfect discriminative ability. This high AUC underscores the model's exceptional capacity to distinguish between pneumonia and normal cases with minimal errors.

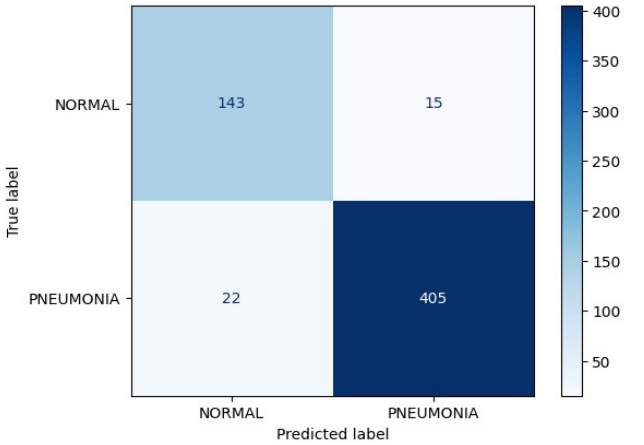
Figure 4 presents the confusion matrices for five deep learning models: DenseNet-169, MobileNet, Inception-v3, Xception, and the proposed *VGG-16*, evaluated for pneumonia detection using chest X-ray images.



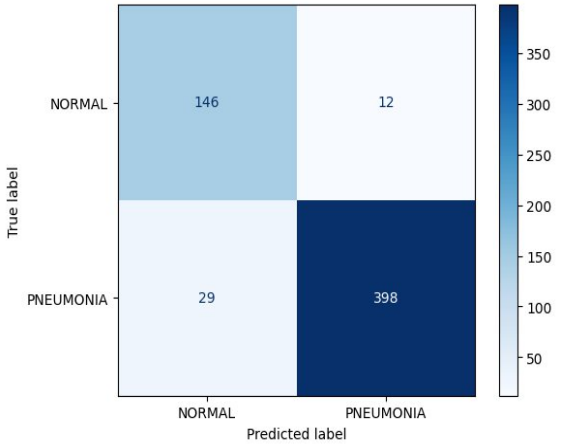
(a): Confusion matrix of DenseNet-169



(b): Confusion matrix of MobileNet

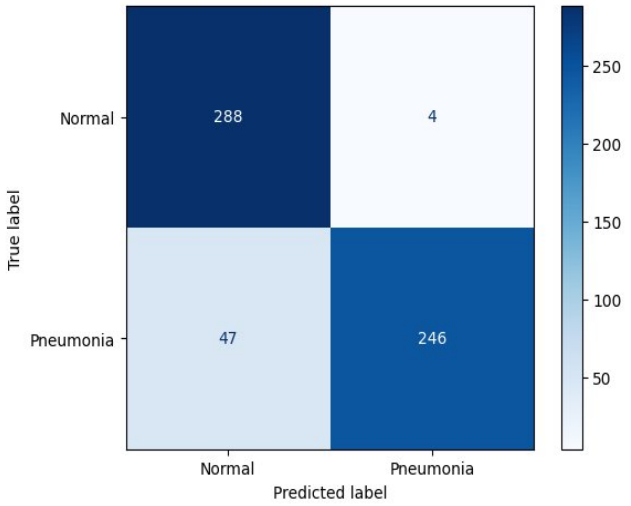


(c): Confusion matrix of Inception-v3

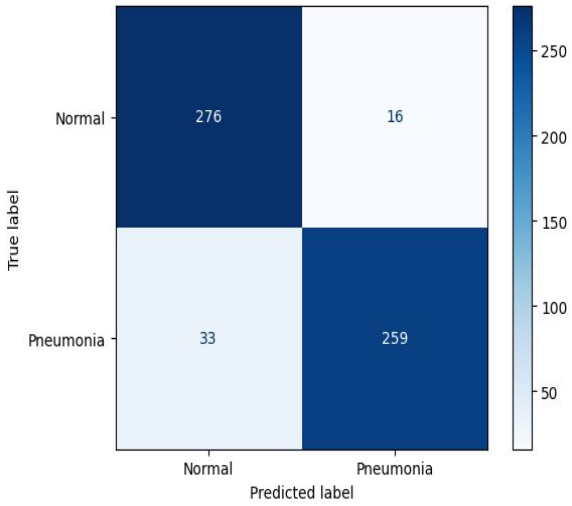


(d): Confusion matrix of Xception

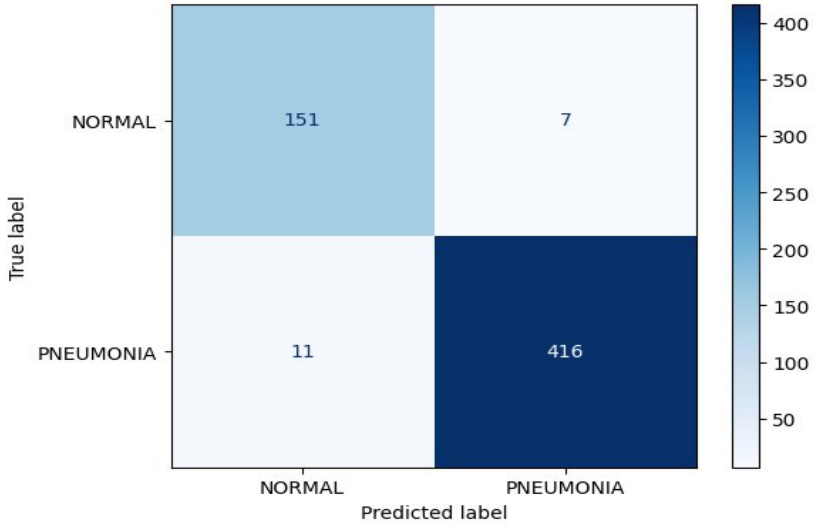
Figure 4. (continuation) Confusion matrix of all applied models.



(e): Confusion matrix of EfficientNetB5



(f): Confusion matrix of ConvNextLarge



(g): Confusion matrix of *VGG-16*

Figure 4. (continuation) Confusion matrix of all applied models.

The Grad-CAM visualizations highlighting the attention of the proposed VGG-16 model attention on five representative chest X-ray cases for pneumonia detection are shown in Figure 5.

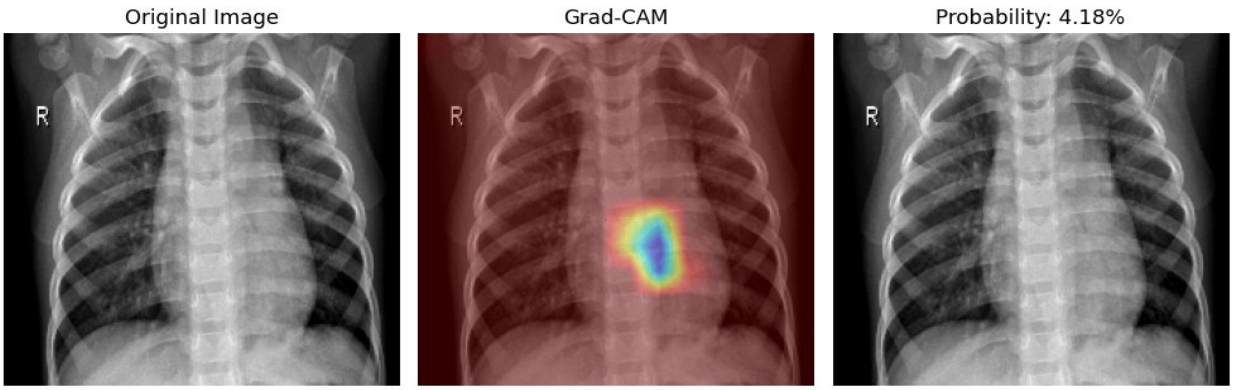


Figure 5. (continuation) *Grad-CAM* Visualizations of the Proposed Model's Attention for Pneumonia Detection.

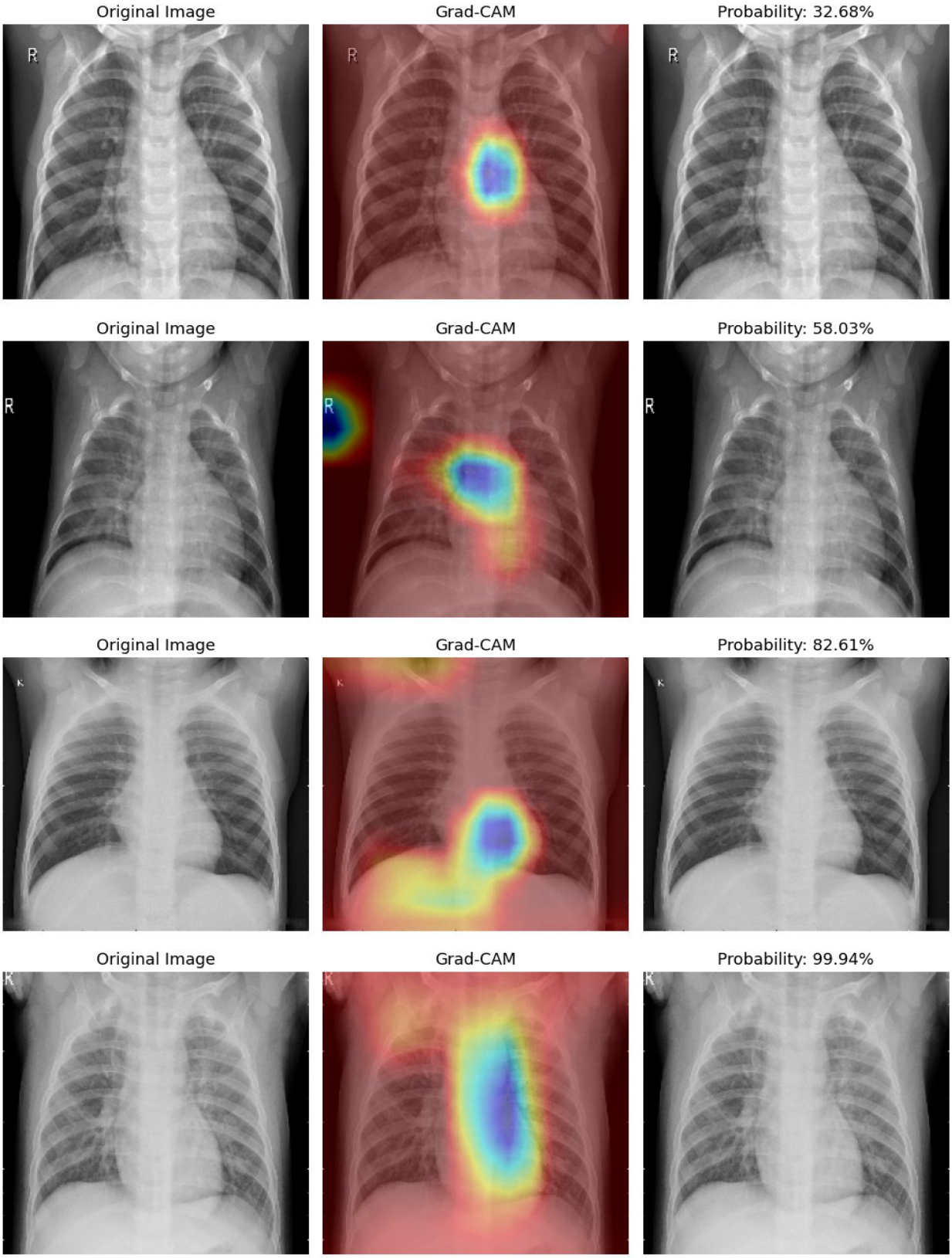


Figure 5. (continuation) *Grad-CAM* Visualizations of the Proposed Model's Attention for Pneumonia Detection.

For the first instance, in the top row, the chest X-ray appeared normal, and the corresponding Grad-CAM heatmap exhibited only faint and localized activation, primarily around the mediastinum, with negligible focus in

the lung regions. The predominance of blue and green hues indicates low attention intensity. With a predicted pneumonia probability of only 4.18%, the model correctly classified this case as non-pneumonic, showing minimal concern for abnormal features. In the second instance, the second row, Grad-CAM reveals slightly more noticeable activation centered around the heart and mediastinum, with minor extension into the lower lung zones. The heatmap remained largely blue and green, with a yellow touch at the center. The model assigned a 32.68% probability of pneumonia, suggesting that while some irregularities may exist, the case is still interpreted as normal, reflecting the model's caution in the absence of strong pathological signals. The third instance, the third row, presents a more extensive and focused activation, especially within the lower to mid-lung area on the left side (right side of the image). The heatmap shifted to warmer tones (yellow and orange), signifying elevated attention. With a predicted probability of 58.03%, the model classified this instance as positive for pneumonia. Localized attention correlates well with potential radiographic abnormalities, indicating early or mild infection. In the fourth bottom row of the fourth instance, the Grad-CAM heatmap highlights broad and intense activation throughout both lungs, particularly in the lower and central lung zones. The vivid red and yellow areas reflect high confidence in identifying the abnormal features. With a pneumonia probability of 82.61%, the model firmly classifies this as a positive case, and the widespread activation is consistent with patterns typically observed in moderate-to-severe pneumonia. Finally, the fifth instance showed the most confident prediction, with a pneumonia probability of 99.94%. The heatmap reveals a concentrated, elongated region of intense activation, marked in bright red and yellow, spanning a substantial portion of the central lung fields. This precise and vivid focus closely aligns with common pneumonia-affected areas, clearly demonstrating the model's capability to localize disease regions with high accuracy and diagnostic confidence.

Discussion

The results of our study demonstrate that the proposed model performed better than several widely used networks, including DenseNet169, MobileNet, InceptionV3, EfficientNetB5, and ConvNeXtLarge. It achieved an accuracy of 97% and an AUC of 0.9969, indicating strong and reliable performance.

The results presented in Table 1 demonstrate that VGG-16 stands out as the top-performing model, achieving the highest average accuracy of 97%, with strong F1-scores for both NORMAL 0.94 and PNEUMONIA 0.98 classes. MobileNet also performed exceptionally well, with an average accuracy of 95.56%, indicating its efficiency, despite its lightweight architecture. DenseNet-169 and Inception-v3 showed comparable results, both reaching 93.68% accuracy, whereas Xception and ConvNextLarge followed closely at 92%. EfficientNetB5 achieved 91% accuracy. These findings suggest that while all models are capable of accurate pneumonia detection, VGG-16 offers the most balanced and reliable performance across both classes in this study.

DenseNet-169 produced 93.68% accuracy with 94.12% precision, 93.68% recall, and 93.78% F1-score (see Table 2). MobileNet demonstrated superior performance with an accuracy of 95.56% and corresponding precision, recall, and F1-score values above 95%. With somewhat lower average metrics, InceptionV3 and Xception produced accuracies of 93.68% and 92.99%, respectively. The ConvNeXtLarge model demonstrated stable but relatively poorer generalization performance, achieving 92.00% accuracy across all metrics, while the EfficientNetB5 model achieved 91.00% accuracy. With all four important metrics equal to 97.00%, the suggested *VGG-16* model performed the best. These results demonstrate that VGG16 is the best model for this binary classification task, although its architecture is comparatively older. It can outperform deeper and more update models with the help of transfer learning and appropriate fine-tuning.

Figures 2(a) and 3(b) illustrate the DenseNet1-69 model's learning behavior, where both the accuracy and loss trends indicate effective training with convergence above 94% and a steady reduction in loss, suggesting strong generalization. In subfigures 2(c) and 2(d), MobileNet demonstrates a consistent performance, achieving nearly 96% accuracy with smoothly declining loss curves, reflecting its robustness. Inception-v3, as shown in subfigures 2(e) and 2(f), also exhibits stable learning with gradual accuracy improvements and a consistent decrease in loss. Subfigures 2(g) and 2(h) depict the performance of Xception, where the training accuracy improved steadily, but the validation accuracy showed fluctuations, indicating possible overfitting. The loss curves further support this, with visible variance. Finally, subfigures 2(i)–2(l) showed that EfficientNetB5 achieved high training accuracy but

fluctuating validation performance, suggesting overfitting, whereas ConvNeXtLarge demonstrated steady improvement in accuracy and consistent loss reduction, indicating stable and balanced learning. Finally, subfigures 2(m) and 2(n) highlight the proposed VGG-16-based model, which achieves the highest and most consistent accuracy of approximately 97%, accompanied by minimal loss. This performance underscores the model's superior learning capacity and reliability for pneumonia detection. Overall, Figure 2 demonstrates the advantages of the proposed model over other architectures in terms of both training stability and validation accuracy.

Figure 4(a) shows that DenseNet-169 recorded 149 true negatives (TN), nine false positives (FP), 28 false negatives (FN), and 399 true positives (TP), reflecting solid performance but with a moderately high rate of missed pneumonia cases. As shown in Figure 4(b), MobileNet offers improved diagnostic accuracy with 153 TN, 5 FP, 21 FN, and 406 TP, reducing both false positives and false negatives compared to DenseNet169. In addition, Figure 4(c) displays the results for InceptionV3, which yielded 143 TN, 15 FP, 22 FN, and 405 TP; While its true positive rate was comparable, it exhibited the highest number of false positives, leading to more incorrect pneumonia predictions. As shown in Figure 4(d), Xception achieved 146 TN, 12 FP, 29 FN, and 398 TP, indicating the lowest sensitivity among all models owing to the highest false negative count and the lowest true positive detection. Figures 4(e) and 4(f) show the confusion matrices for EfficientNetB5 and ConvNeXtLarge, respectively. EfficientNetB5 recorded 288 TN, 4 FP, 47 FN, and 246 TP, indicating lower sensitivity due to many missed cases, while ConvNeXtLarge achieved 276 TN, 16 FP, 33 FN, and 259 TP, reflecting a better balance and higher sensitivity with slightly more false positives. Finally, the proposed VGG16 model, depicted in Figure 4(g), delivered the most favorable results: 151 TN, 7 FP, 11 FN, and the highest TP of 416, demonstrating superior diagnostic precision and reliability. By achieving the highest true positive rate, while simultaneously minimizing false negatives and maintaining a low false positive count, the VGG16 model outperformed DenseNet169, MobileNet, InceptionV3, Xception, EfficientNetB5, and ConvNextLarge. These results, as illustrated in Figures 4(a)–4(g), represent the proposed VGG16 model as the most accurate and clinically effective among the evaluated architectures.

To better understand the model's reasoning, Grad-CAM visualizations were applied. The highlighted regions aligned well with areas typically affected by pneumonia, reinforcing the clinical relevance of the results. These findings suggest that explainable AI can offer practical support to clinicians, particularly in hospitals or regions where access to experienced radiologists may be limited. Despite the promising outcomes, this work has some limitations. The model was trained and tested on a single publicly available dataset, which may not fully represent the diversity of real clinical environments. Additionally, the evaluation was carried out under controlled experimental settings. Its performance in day-to-day clinical operations remains to be verified. Future research should therefore involve datasets from multiple healthcare institutions and real-world testing. Integrating the system into clinical workflows will also be important to ensure that the approach is robust, scalable, and genuinely useful for medical practice.

Table 3 shows a summary of deep-learning-based pneumonia detection reported in the scientific literature. Our proposed approach exhibits high accuracy and high AUC, comparable to other performance metrics reported in the scientific literature.

Table 3. Summary of deep learning-based pneumonia detection.

Ref	Contributions / Methodology	Key Findings	Limitations	Future Directions
[14]	Investigate CNN-based classification of pneumonia from chest X-rays	Accuracy of 91.18%	Emphasis on explainable AI and lightweight architecture	developing Explainable AI models for transparent and interpretable decision
[15]	Develop the Inception-v3 architecture and potentially reduce misdiagnosis rates	Accuracy of 90.48%	Classification spectrum, enhancing the generalizability with multicenter trials, and clinical implementation.	classification spectrum, multicenter trials, and clinical implementation

Ref	Contributions / Methodology	Key Findings	Limitations	Future Directions
[16]	Proposes LDDNet using a hybrid CNN/Transformer with attention; trained on augmented X-rays/CTs; supports multiclass classification	>96% accuracy; 3–8% better sensitivity vs. VGG-16, EfficientNet; <0.1s inference time	Dataset diversity, clinical trial needs, GPU reliance, and excludes non-infectious diseases	Clinical validation; extension to broader pathologies
[17]	Compares 20 CNNs for pneumonia detection; identifies EfficientNet-B0 as top performer	94.13% accuracy; 93.5% precision, 93.14% F1-score	Binary only; lacks pediatric/adult segmentation; real-world validation pending	Expand to multiclass; diverse age-specific datasets
[18]	Uses Xception, *VGG-16*/VGG-19 with transfer learning for 3-class CXR classification (COVID vs pneumonia vs normal)	98% accuracy; >95% in precision/recall/F1; robust differentiation between COVID and regular pneumonia	Dataset diversity, real-world deployment gap, and limited disease scope	Broader clinical trials; test on other pathologies
[19]	Uses *VGG-16*, Inception ResNet, and custom CNN; applies Grad-CAM; evaluates on two datasets	Custom CNN: 97% accuracy; IR/VGG16: ~96–97%	Demographic bias drops to 93% on unbalanced data	Expand datasets; cross-modality testing
[20]	Introduces CP_DeepNet with custom layers and SqueezeNet; uses synthetic augmentation	Binary: 99.32%; Multiclass: 99.62%; Outperforms RT-PCR	No cross-validation; only open-source data; unknown demographic diversity	Add fatality/recovery labels; extend to TB, cancer
[21]	First model to jointly detect COVID, TB, and pneumonia from CXRs using a custom CNN	98.72% accuracy; recall >96% for all classes; good for co-infection cases	Dataset imbalance; public dataset bias	Use SMOTE; apply transfer learning; include more diseases
[22]	Custom ResNet-50 with channel/spatial attention; improved loss function; optimized for class imbalance	98% accuracy; 4% higher than baseline; strong for imbalanced data	Vanishing gradient risks; binary-only scope	Extend to multiclass; include CT/MRI; real-world testing
[23]	Uses a hybrid CNN with bottom-up/top-down attention, interpretable maps, and limited data training	95.47% accuracy; 0.92 F1-score; better than transfer learning	Binary only; untested on multi-institutional datasets	Explore quaternion CNNs; integrate into hospitals
[24]	Hybrid VGG-16 + ResNet-50; 3-class classification; uses noise filtering and augmentation	96.5% accuracy; F1-score 95.9%; AUC 0.98	Small dataset; not tested for co-infections; generalizability unknown	Try federated learning; test on CT/MRI

Ref	Contributions / Methodology	Key Findings	Limitations	Future Directions
[25]	Yolo-v3 achieves 0.32 mAP (vs 0.25 benchmark); optimizes for low-resource settings	Outperforms Mask RCNN; fast processing; high potential	Class imbalance, GPU dependence, and limited iteration	Increase training cycles; optimize preprocessing

Conclusions

Our hybrid deep learning framework, which integrates transfer learning via VGG-16 with a custom self-attention mechanism, demonstrates good performance metrics in improving the accuracy and interpretability of pneumonia detection in chest X-ray images. By leveraging the feature extraction strengths of VGG-16 CNN-based model and enhancing spatial focus through attention, the model consistently outperformed traditional CNN-based approaches in terms of classification accuracy, sensitivity, and specificity. Furthermore, attention-based visualizations contribute to clinical explainability, making the model a valuable tool for radiological decision-making support. These promising results highlight the framework's robustness and potential for real-world deployment. Future research should aim to validate its generalizability across diverse datasets, incorporate longitudinal patient data, and expand its application to a wider spectrum of thoracic diseases.

List of Abbreviations: Not applicable.

Author Contributions: P.B.: Conceptualization, Methodology, Writing - Original Draft, Coding. S.S.: Conceptualization, Methodology, Coding, Writing—Original Draft, A.N. Original Draft, Methodology. S.K.: Software, Investigation, Coding, Data curing and editing.

Funding: This research received no external funding.

Data Availability Statement: Mooney PT. Chest X-ray images (pneumonia). Kaggle [Internet]. 2025 Jul 04 [cited 2025 Aug 18]. Available from: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia.

Conflict of Interest: The authors declare that there are no conflicts of interest related to the publication of this manuscript.

References

1. Kaushik P, Jain E, Kukreja V, Hariharan S, Krishnamoorthy M, Ahuja V, et al. Modelling radiological features fusion and explainable AI in pneumonia detection: a graph-based deep learning and transformer approach. Results Eng. 2025;26:105225. https://doi.org/10.1016/j.rineng.2025.105225
2. World Health Organization. Pneumonia in children. WHO [Internet]. 2025 Jul 04 [cited 2025 Aug 18]. Available from: https://www.who.int/news-room/fact-sheets/detail/pneumonia
3. Al-Dulaimi DS, Mahmoud AG, Hassan NM, Alkhayyat A, Majeed SA. Development of pneumonia disease detection model based on deep learning algorithm. Wirel Commun Mob Comput. 2022;2022:2951168. https://doi.org/10.1155/2022/2951168.
4. Rana N, Marwaha H. Pneumonia detection from X-ray images using federated learning–An unsupervised learning approach. Meas Sens. 2025;37:101410. https://doi.org/10.1016/j.measen.2024.101410.
5. Lamouadene H, EL Kassaoui M, El Yadari M, El Kenz A, Benyoussef A, El Moutaouakil A, et al. Detection of COVID-19, lung opacity, and viral pneumonia via X-ray using machine learning and deep learning. Comput Biol Med. 2025;191:110131. https://doi.org/10.1016/j.compbiomed.2025.110131.
6. Yang Y, Xing W, Liu Y, Li Y, Ta D, Song Y, Hou D, et al. Medical imaging-based artificial intelligence in pneumonia: A narrative review. Neurocomputing. 2025;630:129731. https://doi.org/10.1016/j.neucom.2025.129731.

7. Nageye AY, Jimale AD, Abdullahi MO, Ahmed YA, Addow MA. Enhancing deep learning for pneumonia detection: developing web-based solution for Dr. Sumait Hospital in Mogadishu Somalia. Discov Appl Sci. 2025;7(4):67–76. https://doi.org/10.1007/s42452-025-06735-6.
8. Kailasam R, Balasubramanian S. Deep learning for pneumonia detection: a combined CNN and YOLO approach. Hum Cent Intell Syst. 2025;5(1):44–62. https://doi.org/10.1007/s44230-025-00091-9.
9. Wajgi R, Yenurkar G, Nyangaresi VO, Wanjari B, Verma S, Deshmukh A, et al. Optimized tuberculosis classification system for chest X-ray images: fusing hyperparameter tuning with transfer learning approaches. Eng Rep. 2024;6(11):e12906. https://doi.org/10.1002/eng2.12906.
10. Yao S, Chen Y, Tian X, Jiang R. Pneumonia detection using an improved algorithm based on Faster R-CNN. Comput Math Methods Med. 2021;2021:8854892. https://doi.org/10.1155/2021/8854892.
11. Muhammad Y, Alshehri MD, Alenazy WM, Hoang TV, Alturki R. Identification of pneumonia disease applying an intelligent computational framework based on deep learning and machine learning techniques. Mob Inf Syst. 2021;2021:9989237. https://doi.org/10.1155/2021/9989237.
12. Almaslukh B. A lightweight deep learning-based pneumonia detection approach for energy-efficient medical systems. Wirel Commun Mob Comput. 2021;2021:5556635. https://doi.org/10.1155/2021/5556635.
13. Hong A, Li Z, Song Y. Electrophoretic techniques for rapid detection of bacterial pneumonia: current status and future perspectives. Int J Optoelectron Sensors. 2025;100928. https://doi.org/10.1016/j.ijoes.2025.100928.
14. Solanki J, Agrawal V. A novel approach for pneumonia detection from chest X-ray images using deep learning. Int J Intell Syst Appl Eng [Internet]. Available from: http://www.ijisae.org
15. Li S, Hu Y, Yang L, Lv B, Kong X, Qiang G. DSEception: a novel neural networks architecture for enhancing pneumonia and tuberculosis diagnosis. Front Bioeng Biotechnol. 2024;12:1454652. https://doi.org/10.3389/fbioe.2024.1454652.
16. Podder P, Das SR, Mondal MRH, Bharati S, Maliha A, Hasan MJ, et al. LDDNet: a deep learning framework for the diagnosis of infectious lung diseases. Sensors. 2023;23(1):480. https://doi.org/10.3390/s23010480.
17. Akbar W, Soomro A, Hussain A, Hussain T, Ali F, Ul Haq MI, et al. Pneumonia detection: a comprehensive study of diverse neural network architectures using chest X-rays. Int J Appl Math Comput Sci. 2024;34(4):679-699. https://doi.org/10.61822/amcs-2024-0045.
18. Jain DK, Singh T, Saurabh P, Bisen D, Sahu N, Mishra J, et al. Deep learning-aided automated pneumonia detection and classification using CXR scans. Comput Intell Neurosci. 2022;2022:7474304. https://doi.org/10.1155/2022/7474304.
19. Mohan G, Subashini MM, Balan S, Singh S. A multiclass deep learning algorithm for healthy lung, COVID-19 and pneumonia disease detection from chest X-ray images. Discov Artif Intell. 2024;4(1):20. https://doi.org/10.1007/s44163-024-00110-x.
20. Mehmood MH, Hassan F, Ur Rahman A, Khan W, Mostafa SM, Ghadi YY, et al. CP_DeepNet: a novel automated system for COVID-19 and pneumonia detection through lung X-rays. Multimed Tools Appl. 2024;83(41):88681–88698. https://doi.org/10.1007/s11042-024-18921-6.
21. Ahmed MS, Rahman A, AlGhamdi F, AlDakheel S, Hakami H, et al. Joint diagnosis of pneumonia, COVID-19, and tuberculosis from chest X-ray images: a deep learning approach. Diagnostics. 2023;13(15):2562. https://doi.org/10.3390/diagnostics13152562.
22. Li D. Attention-enhanced architecture for improved pneumonia detection in chest X-ray images. BMC Med Imaging. 2024;24(1):177. https://doi.org/10.1186/s12880-023-01177-1.
23. Singh S, Rawat SS, Gupta M, Tripathi BK, Alanzi F, Majumdar A, et al. Deep attention network for pneumonia detection using chest X-ray images. Comput Mater Continua. 2023;74(1):1673–1691. https://doi.org/10.32604/cmc.2023.032364.
24. Sarkar R, Ahamed S, Bari M, Hasan M, Shishir FH. Deep learning methods for chest X-ray imaging-based COVID-19 pneumonia detection. J Image Process Intell Remote Sens. 2024;45:55–62. https://doi.org/10.55529/jipirs.45.55.62.

25. Darapaneni N, Ranjan A, Bright D, Trivedi D, Kumar K, Kumar V, et al. Pneumonia detection in chest X-rays using neural networks. arXiv [Preprint]. 2022. https://doi.org/10.48550/arXiv.2204.03618.