Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 12 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i12.55953

PAPER

An Attention-Enhanced Hybrid Deep Learning Model Based on VGG16 and VGG19 for Pneumonia Detection from Chest X-ray Images

Thi Thoa Mac[1], Xuan Thuan Nguyen[1], Huy Anh Bui[2], Thanh Hung Nguyen[1], Hoang Hiep Ly[1] (✉)

[1]School of Mechanical Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam

[2]School of Mechanical and Automotive Engineering, Hanoi University of Industry, Hanoi, Vietnam

hiep.lyhoang@hust.edu.vn

ABSTRACT

Pneumonia is one of the most dangerous respiratory diseases and could be life-threatening if not promptly diagnosed and treated. In addition, pneumonia is an infectious disease, and missing a case poses a significant risk to the community. Conventionally, doctors rely on chest X-ray (CXR) images to examine the lungs and detect abnormalities associated with pneumonia. The development of artificial intelligence (AI), especially deep learning (DL) algorithms, can assist doctors in diagnosing the disease more quickly and accurately. This study proposes a hybrid DL model that combines two convolutional neural networks (CNNs), VGG16 and VGG19, with an attention mechanism to enhance pneumonia detection from CXR images. By integrating the lightweight structure of VGG16 with the deeper feature extraction of VGG19 and directing focus to key pathological regions through attention, the model achieves improved diagnostic performance. Evaluated on a public pediatric CXR dataset, the proposed model outperforms VGG16, VGG19, DenseNet121, and InceptionV3 in all major metrics: 89.10% accuracy, 86.42% precision, 91.82% F1-score, and 97.95% recall. The high recall rate is particularly significant in minimizing false negatives, which is critical in clinical contexts to prevent missed pneumonia cases. Despite having the highest parameter count among the compared models, it maintains a fast inference time of 33.48 ms per image, supporting real-time clinical application.

KEYWORDS

pneumonia detection, chest X-ray (CXR), deep learning (DL), hybrid model, attention mechanism, VGG 16, VGG 19

1 INTRODUCTION

Pneumonia is a widespread infectious disease affecting people globally [1]. It is a lung infection caused by bacteria, viruses, or fungi, leading to inflammation in the alveoli. This inflammation results in the accumulation of fluid or pus in the air sacs [2].

Mac, T. T., Nguyen, X. T., Bui, H. A., Nguyen, T. H., Ly, H. H. (2025). An Attention-Enhanced Hybrid Deep Learning Model Based on VGG16 and VGG19 for Pneumonia Detection from Chest X-ray Images. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(12), pp. 63–83. https://doi.org/10.3991/ijoe.v21i12.55953

Article submitted 2025-04-22. Revision uploaded 2025-07-08. Final acceptance 2025-07-08.

Pneumonia affects about 7% of the global population annually, with about four million of those affected facing a potentially fatal outcome [3]. Pneumonia poses serious health risks, especially to the elderly and children. According to the World Health Organization (WHO) statistics, pneumonia accounts for around 1.4 million deaths each year among children under five, representing approximately 18% of fatalities in this age group [4].

Early detection and timely treatment of pneumonia play an important role in reducing the risk of complications, increasing recovery rates, and significantly reducing patient mortality [5]. Chest radiography (CR) is the most commonly used method for early detection of pneumonia because it can identify areas of lung opacity [6]. CR can be used to detect pneumonia even before clear clinical symptoms occur. Chest imaging using computed tomography (CT) or magnetic resonance imaging (MRI) provides more accurate results but is considerably more time-consuming and expensive than chest X-ray (CXR) [6]–[7]. Most cases of severe pneumonia occur in underdeveloped countries where patients cannot afford expensive diagnostic tests [8]. Therefore, accurate, simple, rapid, and cost-effective diagnostic methods are preferred. Currently, the diagnosis of pneumonia using CXR imaging meets most of these requirements. However, accurate diagnosis through evaluation of a patient's CXR image is time-consuming and relies significantly on detailed evaluation by doctors [9].

Recent advances in artificial intelligence (AI), particularly in deep learning (DL), have demonstrated immense potential in transforming medical diagnostics. These technologies not only offer the potential to increase diagnostic accuracy but also significantly reduce the time required for image interpretation [9]–[10]. DL-based approaches, especially those utilizing convolutional neural networks (CNNs), have proven highly effective in various medical imaging applications, including tumor classification, pathology detection, and more [11]–[15]. Compared to traditional diagnostic techniques, CNN-based systems can automatically learn complex patterns in imaging data and deliver high-accuracy results with minimal human supervision [16]. Their ability to automate the feature extraction and classification process makes them suitable candidates for deployment in real-time clinical environments. In pneumonia diagnosis using CXR images, employing CNN architectures to automatically extract and learn important features from medical images has become increasingly popular. However, many CNN architectures are based on single-model designs, which eliminate their ability to capture deeper and more detailed patterns within the data or images [17]. Moreover, single models often lack scalability and generalization capabilities, which are crucial for practical applications [18]. Thus, it is necessary to apply advanced AI techniques, such as pre-training, transfer learning, attention mechanisms, or the combination of multiple CNN models to enhance the performance of detection models in pneumonia diagnosis using CXR images [19].

This study proposes a hybrid model that combines VGG16 and VGG19 for pneumonia detection using CXR images. VGG16 [20]–[21] offers a lightweight structure with lower computational demand and faster processing, making it suitable for real-time clinical applications. In contrast, VGG19 [21], with its deeper architecture, achieves better performance on complex image features but requires more computational resources. By integrating VGG16 and VGG19, the model leverages the speed and efficiency of VGG16 alongside the higher representational power of VGG19, thereby enhancing both accuracy and robustness. Additionally, an attention mechanism is incorporated to help the model focus on diagnostically important regions in the lungs, further improving detection performance. The contributions of this paper are listed as follows:

- Design of a hybrid model that combines VGG16 and VGG19 to balance lightweight inference with deep semantic learning, thereby improving diagnostic performance.

- Integration of an attention mechanism to enhance spatial feature learning and guide the model toward critical pathological areas in CXR images.
- Comprehensive evaluation of a public CXR dataset, demonstrating the model's improved accuracy, robustness, and generalizability compared to single-model baselines.

The subsequent sections are arranged as follows: Section 2 provides a review of related work on pneumonia detection using CNN-based, pre-trained, and ensemble DL architectures. In Section 3, the proposed approach is described, with emphasis on the model architecture, attention mechanism, and training procedure. Section 4 describes the experimental settings, and the evaluation results of the proposed model are presented and compared with those of other DL models. Section 5 discusses the model's performance, limitations, and potential improvements. Finally, Section 6 concludes the paper and outlines directions for future research.

2 RELATED WORKS

In pneumonia diagnosis using CXR images, many DL models have been proposed [22]. GM et al. [23] evaluated fifteen CNN models and proposed a lightweight design suitable for clinical use. Aljawarneh and Al-Quraan [24] developed an efficient CNN designed for real-time diagnosis with computational efficiency. Sharma et al. [25] built a CNN from scratch for feature extraction. Zhang et al. [26] applied a basic VGG-style CNN, while Yen and Tsao [27] introduced a model optimized for low-resource image classification. Additionally, Singh et al. [28] proposed a quaternion CNN for improved spatial learning, while Jakhar et al. [29] introduced a Keras-based light-weight CNN for small datasets, and Saraiva et al. [30] presented a shallow CNN for pediatric pneumonia detection. Although these models are easy to implement and computationally efficient, their limited depth often restricts feature representation and generalization, reducing their effectiveness in complex clinical settings.

To overcome the aforementioned limitations of the simple CNN architectures, recent studies have increasingly adopted pre-trained models with transfer learning techniques. Pre-training leverages models trained on large-scale datasets to acquire general features, which are then fine-tuned on smaller domain-specific datasets, such as CXR images, thereby enhancing the output model performance [31]–[32]. Ho et al. applied the pre-trained DenseNet-121 with feature integration to detect chest diseases, achieving an accuracy of 84.62% [33]; El et al. used pre-trained Xception, VGG16, and VGG19 [34] for pneumonia classification based on CXR images, achieving accuracies of 83.14%, 86.26%, and 85.94%, respectively. Hammoudi et al. [35] utilized image pre-processing and data augmentation to increase data availability and applied tools including machine learning (ML)-based label binaries to encode CXR images, thereby improving model accuracy for pneumonia diagnosis. They thus made a significant gain in sensitivity (95.92%) and accuracy (91.69%). In these methods, pre-trained CNN models were independently evaluated. Although pre-trained models offer improved performance, each architecture differs in learning capacity, feature representation, and computational efficiency. Therefore, various researchers have turned to ensemble strategies that combine multiple models to exploit the advantage of foundation models and minimize their disadvantages in pneumonia diagnosis from CXR images. In [36], an ensemble of Mask-RCNN with ResNet50 and ResNet101 improved detection accuracy but suffered from high computational cost and overfitting despite regularization efforts. In [37], a combination of RetinaNet

and Mask R-CNN was highly ranked in a Kaggle competition but required a large dataset and intensive processing. The transfer learning method in [38] utilized residual structures and dilated convolutions to address model degradation, but deeper layers introduced slower inference and reduced practical efficiency. A hybrid LDA-SVM model in [39] achieved nearly 93% accuracy but required many iterations, leading to a high computational cost. In [40], a tri-model feature extractor combining AlexNet, VGG16, and VGG19 showed moderate performance but suffered from unstable validation accuracy. Chouhan et al. [41] applied a transfer learning approach leveraging five pre-trained models, which improved classification but introduced ensemble complexity. In [42], EfficientNetB4 and vision transformer (ViT) were combined with a hybrid CNN-transformer framework for pneumonia classification, achieving improved accuracy through local-global feature fusion, though at the cost of increased training time and higher computational demands. Kundu et al. [43] proposed an ensemble of DenseNet121, ResNet18, and GoogLeNet, performing well on the Kermany dataset but exhibiting a drop in accuracy (86.85%) on RSNA data. Nneji et al. [44] addressed image quality issues by fusing features extracted from shallow CNNs and MobileNet-V3, though it required careful parameter calibration. Rishav et al. [45] combined ResNet50 with adaptive particle swarm optimization for better feature selection, but this introduced an additional computational burden. Furthermore, Yaseliani et al. [46] proposed a VGG-based hybrid model using multiple classifiers, where the KNN classifier performed best, though the approach demanded a large dataset and significant training time. The ensemble and hybrid models for pneumonia diagnosis using CXR images are summarized in Table 1.

Table 1. Ensemble and hybrid models for pneumonia diagnosis using CXR images

Essemble Models	Strengths	Limitations
Mask-RCNN + ResNet50/101 [36]	High accuracy, post-processing step	High computational cost, overfitting
RetinaNet + Mask R-CNN [37]	Top 3% in Kaggle challenge	Requires large dataset, complex
Residual + Dilated CNN [38]	Addresses degradation issues	Slow inference, low efficiency
LDA + SVM [39]	93% accuracy	High computation time
AlexNet + VGG16/19 [40]	Multi-model feature extraction	Unstable validation accuracy
Transfer Learning (5 CNNs) [41]	Effective ensemble classification	Requires ensemble of five models
DenseNet121 + ResNet18 + GoogLeNet [43]	High accuracy (98.8%)	Fails on RSNA dataset
Shallow CNN + InceptionV3 + MobileNetV3 [44]	Multi-channel image processing	Complex fusion strategy
ResNet50 + AAPSO [45]	Improved feature selection	Computation for AAPSO
VGG16/19 + ML classifiers [46]	Hybrid DL + ML model	Needs large dataset, high computation

Recently, attention mechanisms have been used to guide CNN-based models to focus on key features, such as lung opacities, thereby enhancing pneumonia detection and improving both accuracy and interpretability. An et al. [47] integrated attention mechanisms into an ensemble of seven CNNs to improve pneumonia localization. Similarly, Afifi et al. [48] employed global-local attention modules with DenseNet161, enhancing classification robustness. In [49], Khater et al. proposed AttCDCNet, an attention-augmented DenseNet121 that improved classification accuracy while reducing computational cost through depthwise separable convolutions. However, the aforementioned models mainly apply simple CNN architectures, while the combination of multiple different CNN models has not been widely considered.

3 MATERIALS AND METHODS

3.1 Data preparation

The dataset used in this study consists of 5,840 CXR images of pediatric patients aged 1 to 5 years in China and was provided by Kermany et al. [50]. Figure 1 shows several sample images in the dataset. The dataset comprises 4265 images from pneumonia patients ("pneumonia" class, as shown in Figure 1b) and 1575 images from non-pneumonia patients ("normal" class, as shown in Figure 1a).

To satisfy the input criteria of the VGG16 and VGG19 architectures, all photos in the dataset are scaled to a resolution of 224 × 224 pixels. Moreover, standardized to the range [0, 1] are the gray levels of the photographs, which first span 0 to 255. These preprocessing actions are crucial to guarantee compatibility between two architectures and to raise performance.

Two sets of data were split in order to create and verify the model. Training employed the first data set, which accounted for 79% of all the data. The second data set, which comprised 11% of all the data, was used as the testing data to assess the model's performance. The training dataset was further split: 80% of the training data was utilized for developing the first model and 20% as the validation set for assessing and fine-tuning the model to prevent overfitting or underfitting. Table 2 shows details on the dataset used in this investigation.

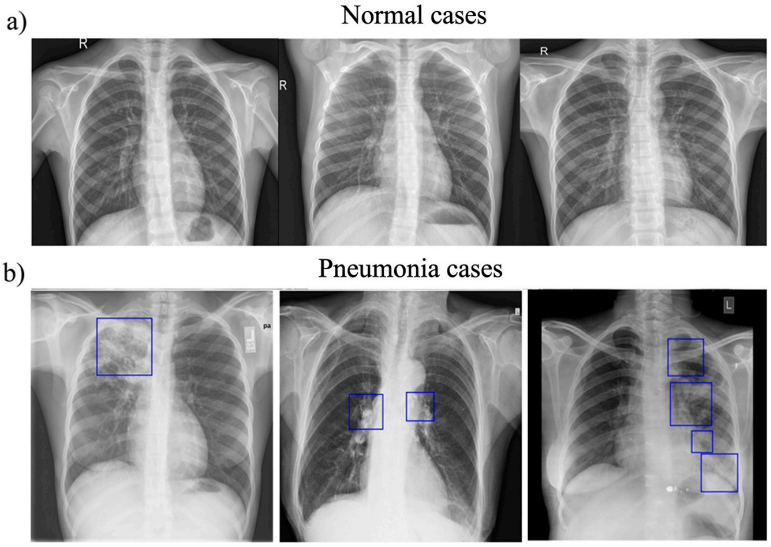

Fig. 1. Sample chest X-ray images from the Kermany dataset: a) Chest X-ray images of normal patients b) Chest X-ray images of patients with pneumonia

Table 2. Dataset distribution

Category	Training Stage		Testing Stage
	Training Set	Validation Set	
Pneumonia case	3099	776	390
Normal case	1073	268	234
Sum	4172	1044	624
Proportion (%)	71%	18%	11%

3.2 Proposed hybrid deep learning model

A hybrid DL model including two pre-trained models, VGG16 and VGG19, depicted in Figure 2, was proposed. VGG16 is a CNN architecture with 16 layers, including 13 convolutional layers and 3 fully connected layers, as shown in Figure 3a. The input image size of the VGG16 model is 224×224. While preserving the salient characteristics of the input picture, the convolutional layers help to reduce computational complexity. The convolutional layers in VGG16 are organized into five blocks: block 1 comprises two convolutional layers with 64 filters; block 2 comprises two convolutional layers with 128 filters; block 3 comprises three convolutional layers with 256 filters; and blocks 4 and 5 comprise three convolutional layers with 512 filters each. The convolutional operation for a layer is defined as follows:

$$y_{i,j,k} = \sum_{m,n,c} W_{m,n,c,k} \times x_{i+m,j+n,c} + b_k \tag{1}$$

where $x_{i+m,j+n,c}$ is the input feature map value at position $(i + m, j + n)$ in channel c; $W_{m,n,c,k}$ denotes the weight of the kernel for filter k at position (m, n) in input channel c; b_k is the bias for filter k; $y_{i,j,k}$ is the output feature map value at position (i, j) for the filter k.

Every block end with a maximum pooling layer (kernel size 2×2, stride 2) added to reduce the model's size and the max-pooling operation is described as follows:

$$y_{i,j,k} = \max_{m,n \in [0,1]} x_{2i+m,2j+n,k} \tag{2}$$

The RELU function, as an activation function, is applied after each convolutional layer in Blocks 1–5 to introduce nonlinearity, transforming the output as follows:

$$Y'_{i,j,k} = \mathrm{ReLU}(Y_{i,j,k}) = \max(0, Y_{i,j,k}) \tag{3}$$

Almost exactly in structure, VGG-19 is a deeper variation of the VGG model. Figure 3b illustrates that VGG-19 adds three convolutional layers: one with 256 filters in block 3, one with 512 filters in block 4, and one with 512 filters in block 5, using the same convolutional, max-pooling operation and activation function as in Equations (1)–(3).

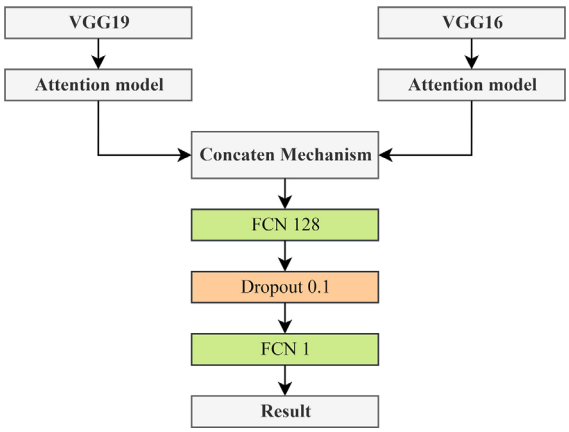

Fig. 2. Diagram of the proposed model

In the proposed model, the attention mechanism processes features from VGG16 and VGG19 to enable the model to ignore undesired characteristics and concentrate on highly discriminative and important patterns. Figure 3c demonstrates the

architecture of the attention network used in this study. It begins with batch normalization to stabilize training, which is defined as follows:

$$\hat{X}_{i,j,k} = \frac{X_{i,j,k} - \mu_k}{\sqrt{\sigma_k^2 + \epsilon}} \cdot \gamma_k + \beta_k \tag{4}$$

where μ_k and σ_k^2 are the mean and variance of the k channel, γ_k and β_k are learnable parameters, and ϵ is a small constant. To lower the dimensionality of the features and lighten the model, two convolutional layers with 1×1 kernels containing 64 and 16 filters respectively, subsequently create attention maps. After another 1×1 convolution (512 filters), a local concatenation convolution layer is used to generate local concatenation and highlight significant areas, hence compressing the features. Finally, the global average pooling (GAP) layer generates a vector expressing the significance of the feature as the following equation:

$$Y_k = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{i,j,k} \tag{5}$$

where H and W are the height and width of the feature map. A multiply operation focuses on the relevant information and is then used to eliminate noise from the input feature map (acquired after batch normalization) and the map of interest, obtained before compressing the features, as the following equation:

$$Z_{i,j,k} = X_{i,j,k} \times A_{i,j,k} \tag{6}$$

where A is the attention map. Subsequently, the GAP layer compresses the data to lower its dimensionality. The Lambda layer applies a standardization transformation, which is described as follows:

$$Z_{i,j,k} = \frac{X_{i,j,k} - \mu}{\sigma} \tag{7}$$

where μ and σ are the mean and standard deviation of the input features. The Reshape layer flattens the feature map into a vector, as the following equation:

$$Z = \text{Reshape}(X, N) \tag{8}$$

where $N = H \times W \times C$ for channels C. The fully connected layers process the reshaped features. The 1024-node layer applies a linear transformation with ReLU activation, as follows:

$$Y = \max(0, W \times X + b) \tag{9}$$

where W is the weight matrix, X is the input vector, and b is the bias. The 512-node layer incorporates dropout with probabilities 0.5 and 0.1 to prevent overfitting, as follows:

$$Y_i = \begin{cases} 0 & \text{with probability } p \\ \dfrac{X_i}{1-p} & \text{otherwise} \end{cases} \tag{10}$$

where p is the dropout rate. The 128-node layer with ReLU and 0.1 dropout precedes a single-node layer with sigmoid activation, as follows:

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}} \tag{11}$$

where $z = W \times X + b$. The outputs from VGG16 and VGG19 are concatenated to make use of their respective benefits, as the following equation:

$$Z = \text{Concat}(F_{VGG16}, F_{VGG19}) \tag{12}$$

where F_{VGG16} and F_{VGG19} are the feature vectors from VGG16 and VGG19. The model is optimized using binary cross-entropy loss for pneumonia detection as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{13}$$

where N is the number of samples, $y_i \in \{0, 1\}$ is the true label (0 for negative, 1 for positive), and $\hat{y}_i \in [0,1]$ is the predicted probability. This loss penalizes incorrect predictions, ensuring effective learning. This hybrid model leverages the depth of VGG16 and VGG19, enhanced by an attention mechanism, to focus on relevant features while reducing noise, making it well-suited for accurate pneumonia detection.

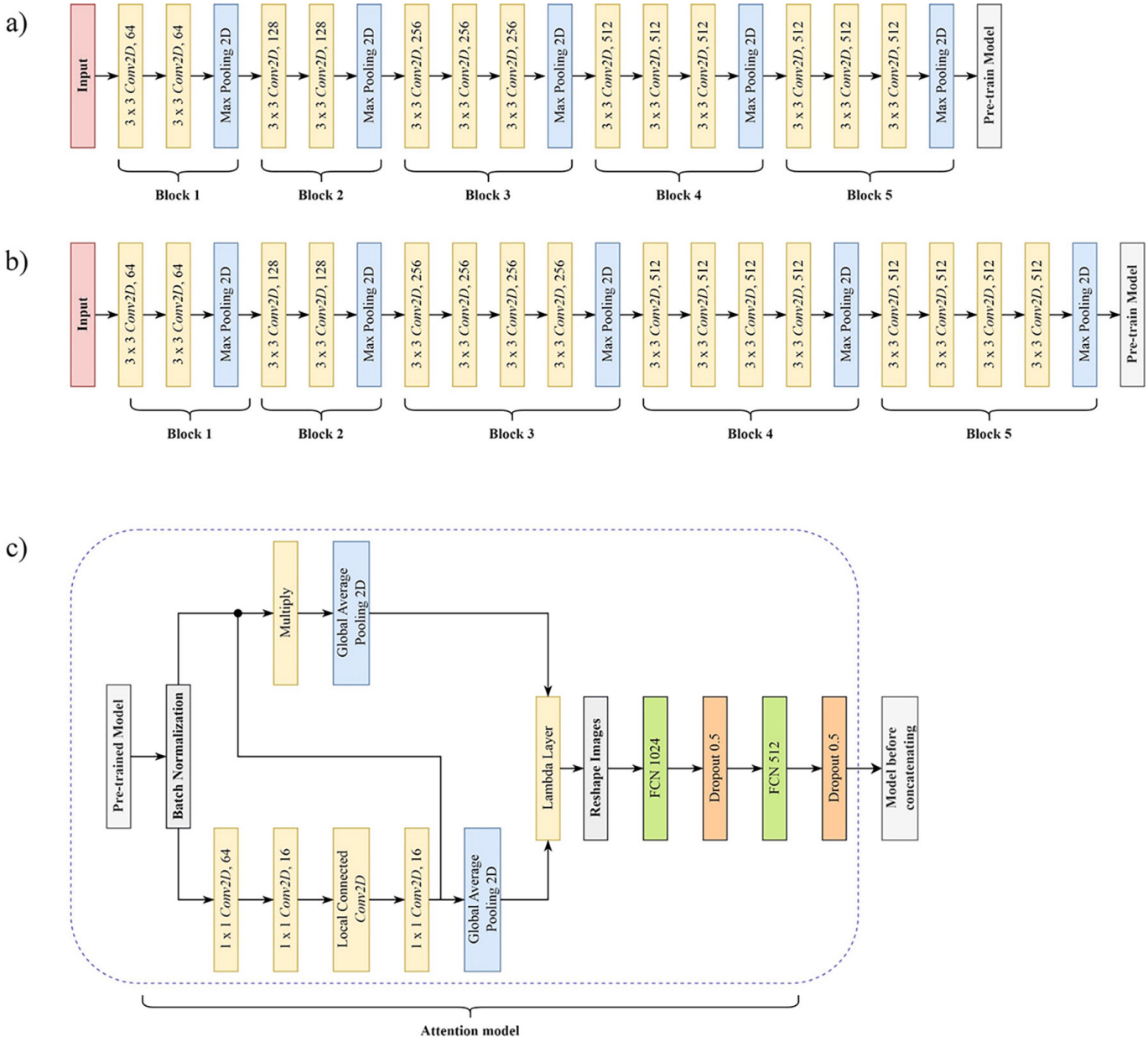

Fig. 3. Diagram of the VGG16, VGG 19, and Attention models [21]: a) VGG16 model; b) VGG19 mode; c) Attention model

4 RESULTS

4.1 Training and validation results

The performance of the proposed model was compared with that of VGG16, VGG19, DenseNet121, and InceptionV3. The models were trained and compiled using Google Colab and a computer with an Intel Xeon Gold 6128 CPU, 32 GB RAM, and NVIDIA GeForce GTX 4080 (16 GB) GPU. The model training is described as in Table 3.

Table 3. Training parameters of the models

Model	Proposed Model	VGG16	VGG19	DenseNet121	InceptionV3
Initial Learning Rate	0.00005	0.00005	0.00005	0.00005	0.00005
Optimization (Optimizer)	Adam	Adam	Adam	Adam	Adam
Epochs Trained	50	50	50	50	50
Training Time per Epoch (after initial epochs) (seconds/epoch)	118–125	73–85	76–89	71–131	116–130

Table 4. The model complexity and efficiency metrics

Model	Proposed Model	VGG16	VGG19	DenseNet121	InceptionV3
Total Params	35,467,459	14,780,481	20,090,177	7,168,833	22,065,185
Trainable Params	726,467	65,793	65,793	131,329	262,401
Non-trainable Params	34,740,992	14,714,688	20,024,384	7,037,504	21,802,784
Inference Time per Image (average on 50 samples)	33.48 ms	40.23 ms	42.72 ms	58.79 ms	59.33 ms
FLOPs (Floating Point Operations)	69.64 GFLOPs	15.5 GFLOPs	19.6 GFLOPs	2.87 GFLOPs	GFLOPs

All models were trained using the same initial learning rate of 0.00005, using the Adam optimizer, and were trained for 50 epochs. However, the training time per epoch varied among the models. DenseNet121 exhibited the highest variability in training time per epoch (71–131 seconds), while VGG16 had the shortest duration (73–85 seconds). The proposed model requires a relatively stable but longer training time (118–125 seconds), similar to InceptionV3. The training phase enabled fine-tuning of model parameters and hyperparameters. The complexity and efficiency metrics across five models are presented in Table 4. The proposed model has the highest total number of parameters (35.47 million), but only a small portion is trainable (726,467), with the majority being non-trainable. In contrast, VGG16 and VGG19 have fewer total parameters and significantly fewer trainable parameters (65,793 each). DenseNet121 has the lowest total parameter count (7.17 million) and also the fewest floating point operations (FLOPs) (2.87 GFLOPs), making it computationally light. The proposed model achieves the fastest inference time per image at 33.48 ms, while InceptionV3 and DenseNet121 are the slowest, at 59.33 ms and 58.79 ms, respectively. Additionally, the proposed model has the highest computational complexity at 69.64 GFLOPs.

Figure 4 displays the models' performance in the validation and training phases. The accuracy of the proposed model remained stable throughout 50 epochs for both training and validation. The difference between training and validation accuracy and loss was minimal in accuracy and loss. Table 5 presents the validation and training accuracy/loss for the five models. Every evaluated model achieved a training accuracy above 91% and a validation loss below 0.25. Moreover, all models attained a validation accuracy exceeding 90%. Among the evaluated models, VGG19 exhibited the lowest performance, with a training accuracy of 91.11%, training loss of 0.2487, validation accuracy of 90.13%, and validation loss of 0.9262 after 50 epochs. In contrast, the proposed hybrid model (VGG16 + VGG19) achieved the best performance, with a training accuracy of 97.14%, training loss of 0.0722, validation accuracy of 96.17%, and validation loss of 0.1058. Comparatively, the other models, including DenseNet121, InceptionV3, and VGG16, demonstrated slightly lower results in both accuracy and loss.

Table 5. Performance metrics on the training and validation sets of the proposed model compared to other models

Model	Training Accuracy (%)	Validation Accuracy (%)	Training Loss	Validation Loss
Proposed model	**97.14**	**96.17**	**0.0722**	**0.1058**
VGG16	92.62	90.13	0.2487	0.9262
VGG19	91.11	90.13	0.2245	0.2442
InceptionV3	95.59	95.21	0.1234	0.1277
DenseNet121	96.16	93.77	0.1075	0.1507

This paper conducted a five-fold cross-validation to evaluate the performance of the models. Tables 6–9 present the training accuracy, training loss, validation accuracy, and validation loss for each fold across the five models. The average values and standard deviations across the five folds indicate that all models achieved consistently high accuracy and low loss, with relatively small standard deviations (less than 0.51% for accuracy and less than 0.89% for loss). The proposed model (VGG16 + VGG19) achieved the highest average accuracy of 96.45%, with a standard deviation of only 0.34%. Similarly, on the validation data, the proposed model outperformed the others, achieving an average accuracy of 94.71% ± 0.53%.

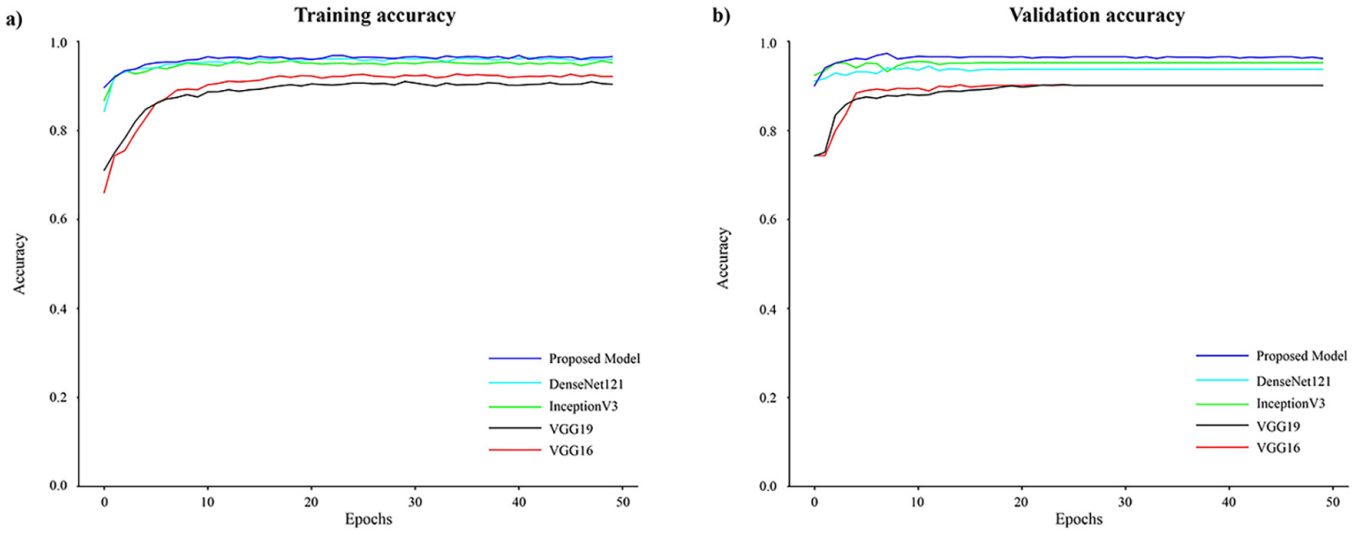

Fig. 4. *(Continued)*

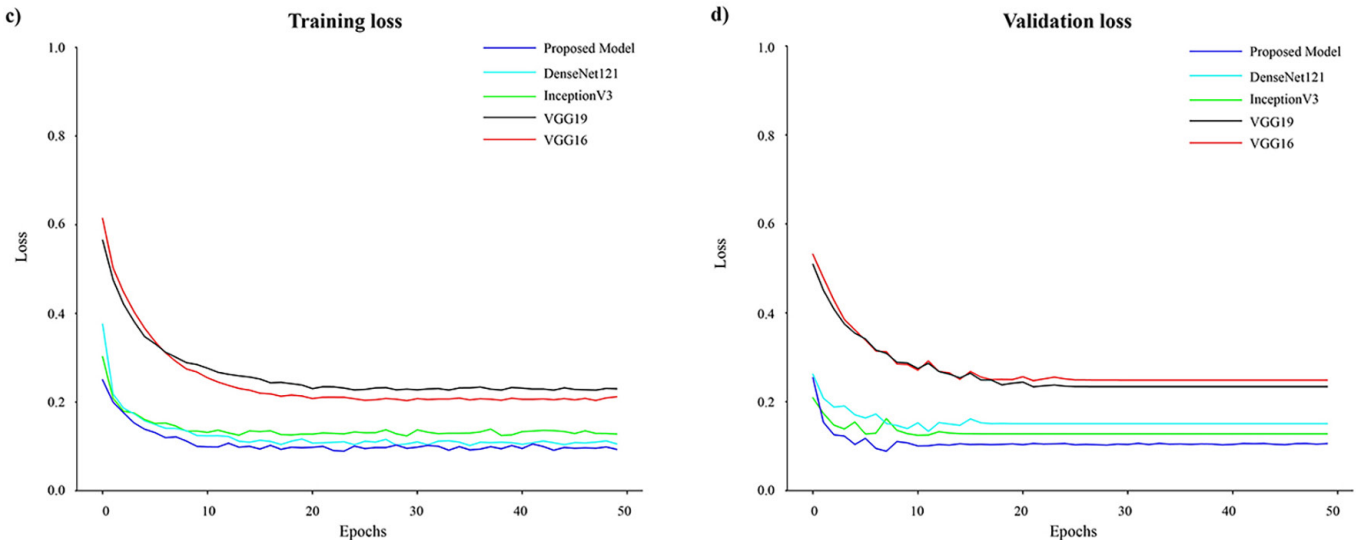

Fig. 4. Training and validation performance of the proposed model and the other evaluated models: a) Training accuracy b) Validation accuracy c) Training loss d) Validation loss

Table 6. Training accuracy of five evaluated models over five-fold cross-validation

K-Fold	Training Accuracy				
	VGG16	VGG19	InceptionV3	DenseNet121	Proposed Model
1	92.94	90.41	94.54	95.24	**96.66**
2	92.47	91.27	95.11	95.46	**96.51**
3	93.21	90.56	94.61	95.14	**96.74**
4	93.09	91.41	94.10	95.29	**95.80**
5	92.57	91.74	94.13	95.59	**96.55**
Avg ± std.dev	92.86 ± 0.29	91.08 ± 0.51	94.50 ± 0.37	95.34 ± 0.16	**96.45 ± 0.34**

Table 7. Training loss of the five evaluated models across five-fold cross-validation

K-Fold	Training Loss				
	VGG16	VGG19	InceptionV3	DenseNet121	Proposed Model
1	18.75	22.87	14.03	12.73	**9.38**
2	19.68	22.23	13.36	12.03	**10.08**
3	18.48	22.89	14.37	12.93	**8.83**
4	18.66	22.29	15.41	13.04	**10.88**
5	19.85	21.65	15.79	12.43	**10.31**
Avg ± std.dev	19.08 ± 0.57	22.39 ± 0.46	14.59 ± 0.89	12.63 ± 0.37	**9.90 ± 0.72**

Table 8. Validation accuracy of the five evaluated models across five-fold cross-validation

K-Fold	Validation Accuracy				
	VGG16	VGG19	InceptionV3	DenseNet121	Proposed Model
1	92.89	92.04	92.80	94.69	**95.63**
2	90.49	90.24	93.66	94.26	**94.52**
3	91.86	90.84	95.21	94.94	**94.60**
4	92.04	90.84	93.41	93.75	**94.78**
5	91.27	91.01	93.66	93.57	**94.00**
Avg ± std.dev	91.71 ± 0.80	90.99 ± 0.58	93.75 ± 0.80	94.24 ± 0.53	**94.71 ± 0.53**

Table 9. Validation loss of the five evaluated models across five-fold cross-validation

K-Fold	Validation Loss				
	VGG16	VGG19	InceptionV3	DenseNet121	Proposed Model
1	20.29	21.2	18.49	14.02	**12.22**
2	24.13	24.66	17.53	15.89	**17.55**
3	21.01	22.39	12.82	13.33	**13.71**
4	21.15	23.33	16.56	14.66	**13.66**
5	23.07	24.79	16.17	16.79	**16.77**
Avg ± std.dev	21.93 ± 1.43	23.27 ± 1.36	16.31 ± 1.92	14.94 ± 1.25	**14.78 ± 2.03**

4.2 Test results

Following training, the model's performance was evaluated on the test dataset using five metrics: accuracy, precision, recall, F1-score, and area under the curve (AUC) score. The formulas for calculating these metrics are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{16}$$

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

where: true positive (TP) represents the number of pneumonia cases correctly identified as positive, while false positive (FP) refers to normal cases that were incorrectly classified as pneumonia. true negative (TN) indicates the number of normal cases accurately predicted as negative, and false negative (FN) counts pneumonia cases that were mistakenly classified as normal. To assess the model's performance at various thresholds, the AUC score is used. This score is determined by plotting the true positive rate (TPR) against the false positive rate (FPR) at all various thresholds,

with the AUC representing the area under the receiver operating characteristic (ROC) curve. The formulas for TPR and FPR are given as follows:

$$\begin{cases} TPR = \dfrac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \\ FPR = \dfrac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} \end{cases} \tag{18}$$

The confusion matrices obtained after applying the models to the test dataset are shown in Figure 5. Based on these confusion matrices, the evaluation metrics for each model were calculated and summarized in Table 10. Overall, the proposed model outperformed the others in most evaluation metrics, including accuracy, precision, recall, and F1-score, with the exception of the AUC score. The accuracy of the proposed model reaching 89.10%, followed by DenseNet121, VGG16, InceptionV3, and VGG19 with accuracies of 87.66%, 85.90%, 82.69%, and 79.33%, respectively. Recall values were relatively high across all models, with the lowest observed in VGG16 at 93.07% on the test set. The proposed model achieved the highest recall at 97.95%, indicating a strong ability to detect pneumonia cases with minimal false negatives. Moreover, the proposed model's precision of 86.42% and F1-score of 91.82% (also the highest overall) demonstrate a strong balance between detecting pneumonia and minimizing false alarms. While DenseNet121 achieved a slightly higher AUC score (0.956) compared to the proposed model's (0.933), its other evaluation metrics, such as precision (85.01%), recall (97.43%), and F1- score (90.8%), were all lower than those of the proposed model. VGG16 exhibited the second-highest precision (85.61%) along with a solid F1-score of 89.49%, indicating good class balance. However, it had the lowest recall (93.07%) among the five evaluated models. In contrast, VGG19 and InceptionV3 performed worse than the other three models. Despite relatively high recall values (95.13% for VGG19 and 96.41% for InceptionV3), both models exhibited low precision (77.13% and 80.00%, respectively), which negatively impacted their overall balance. This is reflected in their lower F1-scores (85.19% and 87.44%) and lower overall accuracies (79.33% and 82.69%, respectively).

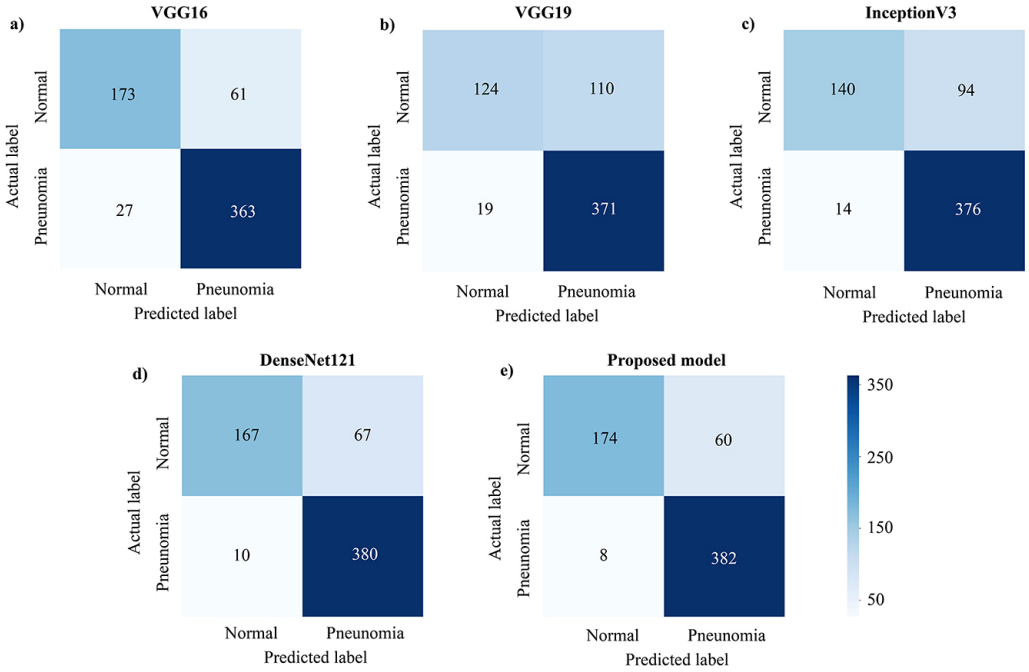

Fig. 5. Confusion matrices over the test dataset: a) VGG16 model; b) VGG19 model c) InceptionV3 model; d) DenseNet model; e) Proposed model

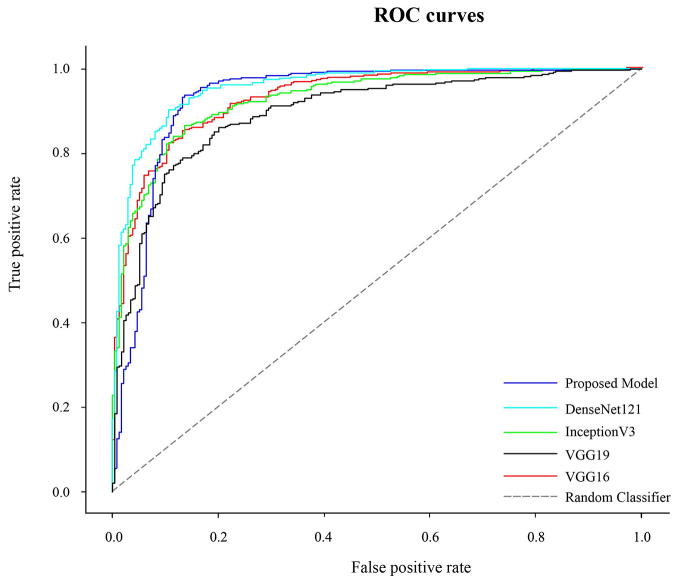

Fig. 6. Receiver operating characteristic curves for five CNN models evaluated on the test dataset

Figure 6 presents the ROC curves of the five DL models assessed after training. The highest AUC score, 0.956, was observed for the Densenet121 model, while the proposed model followed closely, with an AUC score of only 0.032 lower (0.933). These were followed by VGG16, InceptionV3, and VGG19, respectively.

Table 10. Evaluation metrics on the test dataset for all evaluated models

Model	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	AUC-Score
Proposed model	**86.42**	**97.95**	**91.82**	**89.10**	**0.933**
VGG16	85.61	93.07	89.19	85.90	0.930
VGG19	77.13	95.13	85.19	79.33	0.896
InceptionV3	80.00	96.41	87.44	82.69	0.929
Densenet121	85.01	97.43	90.80	87.66	0.956

5 DISCUSSIONS

Regarding training configuration, all models were trained for 50 epochs using the same learning rate and optimizer settings. Despite requiring a longer training time per epoch (118–125 seconds), the proposed model demonstrated superior stability during training. In terms of model complexity, it contains approximately 35.5 million total parameters, of which only 726,467 are trainable, due to the use of frozen pre-trained layers. This balance allows the model to maintain high accuracy while reducing the risk of overfitting. Although the proposed model has the highest number of FLOPs among the evaluated models (69.64 GFLOPs), it achieves the fastest inference time per image (33.48 ms), indicating excellent computational efficiency for real-time applications. These results suggest that the model's architectural design—including attention mechanisms and parallel feature fusion from VGG16 and VGG19—contributes to both computational scalability and strong deployment potential in clinical environments.

Table 11. Comparison of the proposed model to the other existing models

Model	No. of Images	Image Size	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	AUC
Proposed model	**5,840**	**224 × 224**	**86.42**	**97.95**	**91.82**	**89.10**	**0.933**
RetinaNet + Mask RCNN [37]	25684	512 × 512	75.8	79.3	77.5	–	–
Residual + Dilated CNN [38]	5856	150 × 150	89.1	96.7	92.7	90.5	0.953
AttCDCNet (DenseNet121 + Attention) [49]	21265	–	95.14	94.53	–	94.94%	0.929

The training results show that, in pneumonia image identification, the proposed model attained the lowest training accuracy (97.14%) and the lowest training loss (0.0722). Moreover, the validation accuracy (96.17%) and validation loss (0.1058) closely matched the training results, indicating the suggested model was well-tuned with minimal overfitting or underfitting. Although showing slight fluctuations, the training and validation accuracy/loss curves remained stable throughout training. Moreover, five-fold cross-valuation was employed to validate the model. With low standard deviations of 0.34% (training) and 0.53% (validation), the proposed model attained the highest average training accuracy (96.45%) and validation accuracy (94.71%). This consistency shows the model's capacity to generalize despite moderate data variation. Based on the training results, InceptionV3 and DenseNet121 performed slightly worse than the proposed model. When used individually, VGG16 and VGG19 produced lower performance in pneumonia detection tasks. With a high recall of 97.95%, the test results confirm that the proposed model achieved the best overall performance in pneumonia detection among all evaluated models. Similarly, in comparison to the existing ensemble models (Table 11), the proposed model demonstrates a high recall of 97.95%, outperforming other models, including RetinaNet + Mask R-CNN (79.3%) and even the Residual + Dilated CNN (96.7%). The high recall value highlights the model's reliability in identifying pneumonia cases, effectively reducing false negatives and minimizing missed diagnoses. In medical diagnostics, this is critically important, as missing pneumonia cases can lead to severe consequences for patients. The confusion matrix given in Figure 5 clearly reflects this effectiveness, as only 8 pneumonia cases out of 390 were misclassified. For infectious diseases such as pneumonia, where missing cases might contribute to community transmission, this is particularly crucial. Further proving the model's robustness and effectiveness in balancing sensitivity with general detection performance are its highest precision (86.42%), F1-score (91.82%), and accuracy (89.10%) among the evaluated models.

The DenseNet121 model achieved the highest AUC score (0.965), slightly higher than that of the proposed model (0.933), indicating strong discriminatory power. However, its F1-score, recall, and accuracy were all lower than those of the proposed model, while its precision was also lower than that of both the VGG16 model and the proposed model. This suggests that although the DenseNet121 model performs well in terms of AUC, it exhibits less balanced overall performance.

The precision of VGG16 reached the second-highest value (86.51%); however, it had the lowest recall (93.97%) among all evaluated models. Although the VGG19 model has a higher recall (95.13%), its precision was very low (77,13%). These results indicate that neither VGG16 nor VGG19 alone offers a well-balanced classification between normal and pneumonia cases. In contrast, the proposed hybrid model,

which combines the VGG16 and VGG19, achieved consistently high performance across all key metrics. This demonstrates that the hybrid architecture successfully leverages the complementary strengths of the two individual CNN models.

The high recall and performance consistency of the proposed model suggest that the hybrid integration of VGG16 and VGG19 with an attention mechanism contributes effectively to pneumonia detection in CXR images. The use of pre-trained models enables knowledge transfer, thereby improving generalization and reducing the need for extensive training data. The attention mechanism further enhances feature discrimination by focusing on disease-relevant regions, which also contributes to model interpretability. However, the model has some limitations. First, although the recall is high, both accuracy and precision could be further improved. Enhancing data diversity through augmentation or incorporating advanced regularization techniques may help boost overall classification performance. Additionally, experimenting with alternative classifiers, fine-tuning optimizers, or adopting more sophisticated ensemble strategies could improve model robustness—though such enhancements may also increase complexity, computational cost, and training time. Second, the model was developed and validated solely on the Kermany dataset [39], which, although widely used as a benchmark, contains a limited number of samples and lacks demographic diversity. This restricts the model's generalizability to broader clinical settings. Future studies should incorporate additional public or institutionally sourced datasets to validate the model's performance across diverse populations and imaging conditions. Lastly, although the model shows promise for real-world deployment, its clinical utility remains to be validated in hospital environments. Collaborations with healthcare institutions will be essential to evaluate the model's effectiveness in practice and ensure it reduces diagnostic workload while minimizing missed pneumonia cases, particularly those with high transmission potential. Addressing these limitations in future work would enhance the model's clinical relevance and support its adoption in public health systems.

6 CONCLUSIONS

This paper proposes a DL model for detecting pneumonia cases using CXR images. The proposed model is established by combining pre-trained architectures, VGG16 and VGG19, and integrating an attention mechanism into the transfer learning process to maximize the strengths of each component. Test results confirm that the proposed model outperforms other compared models across most evaluation metrics, achieving the highest accuracy, precision, recall, and F1-score when compared to standalone models, such as DenseNet121, InceptionV3, VGG16, and VGG19. The proposed model achieves a recall of 97.75%, missing only eight pneumonia cases, indicating its effectiveness in minimizing missed pneumonia cases. This is particularly critical for preventing potential community transmission, a concern, especially given the severe consequences witnessed during the COVID-19 pandemic. Furthermore, reducing the number of missed positive cases enables timely medical treatment, thereby lowering the risk of severe conditions and life-threatening complications. In future research, the model will be improved by incorporating additional training data, either from other public datasets or through data augmentation. The model is intended for deployment in hospitals and clinics to assist doctors in diagnosing and treating diseases more efficiently.

7 ACKNOWLEDGMENTS

This research is funded by Hanoi University of Science and Technology (HUST) under project number T2024-PC-020.

8 REFERENCES

[1] S. Akter, S. M. Shamsuzzaman, and F. Jahan, "Community acquired bacterial pneumonia: Aetiology, laboratory detection and antibiotic susceptibility pattern," *Malaysian Journal of Pathology*, vol. 36, no. 2, pp. 97–103, 2014.

[2] M. Kolditz and S. Ewig, "Community-acquired pneumonia in adults," *Deutsches Ärzteblatt International*, vol. 114, no. 49, pp. 838–845, 2017. https://doi.org/10.3238/arztebl.2017.0838

[3] Y. Shen *et al.*, "Impact of pneumonia and lung cancer on mortality of women with hypertension," *Scientific Reports*, vol. 6, p. 20, 2016. https://doi.org/10.1038/s41598-016-0023-2

[4] World Health Organization, "The top 10 causes of death, Geneva, Switzerland," 2017. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [Accessed: Nov. 10, 2019].

[5] T. M. File Jr and T. J. Marrie, "Burden of community-acquired pneumonia in North American adults," *Postgraduate Medicine*, vol. 122, no. 2, pp. 130–141, 2010. https://doi.org/10.3810/pgm.2010.03.2130

[6] W. H. Self, D. M. Courtney, C. D. McNaughton, R. G. Wunderink, and J. A. Kline, "High discordance of chest X-ray and computed tomography for detection of pulmonary opacities in ED patients: Implications for diagnosing pneumonia," *American Journal of Emergency Medicine*, vol. 31, no. 2, pp. 401–405, 2013. https://doi.org/10.1016/j.ajem.2012.08.041

[7] Radiological Society of North America, "Pneumonia," RadiologyInfo.org, 2023. [Online]. Available: https://www.radiologyinfo.org/en/info/pneumonia [Accessed: Dec. 31, 2019].

[8] V. Chouhan *et al.*, "A novel transfer learning-based approach for pneumonia detection in chest X-ray images," *Applied Sciences*, vol. 10, no. 2, p. 559, 2020. https://doi.org/10.3390/app10020559

[9] M. Lavine, "The early clinical X-ray in the United States: Patient experiences and public perceptions," *Journal of the History of Medicine and Allied Sciences*, vol. 67, no. 4, pp. 587–625, 2012. https://doi.org/10.1093/jhmas/jrr047

[10] S. Rajaraman, S. Candemir, G. Thoma, and S. Antani, "Visualizing and explaining deep learning predictions for pneumonia detection in pediatric chest radiographs," in *Proceedings of Medical Imaging 2019: Computer-Aided Diagnosis*, vol. 10950, 2019, pp. 200–211. https://doi.org/10.1117/12.2512752

[11] S. T. Ahmed and S. M. Kadhem, "Using machine learning via deep learning algorithms to diagnose the lung disease based on chest imaging: A survey," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 15, no. 16, pp. 95–112, 2021. https://doi.org/10.3991/ijim.v15i16.24191

[12] P. Rajesh, A. Murugan, B. Murugamantham, and S. Ganesh, "Lung cancer diagnosis and treatment using AI and mobile applications," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 14, no. 17, pp. 189–203, 2020. https://doi.org/10.3991/ijim.v14i17.16607

[13] M. A. Mohammed, B. Obeng, S. Alornyo, M. Asante, and B. Obo Essah, "ResFCNET: A skin lesion segmentation method based on a deep residual fully convolutional neural network," *IETI Transactions on Data Analysis and Forecasting (iTDAF)*, vol. 1, no. 1, pp. 4–19, 2023. https://doi.org/10.3991/itdaf.v1i1.35723

[14] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological Physics and Technology*, vol. 10, pp. 257–273, 2017. https://doi.org/10.1007/s12194-017-0406-5

[15] D. Shen, G. Wu, and H. I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017. https://doi.org/10.1038/s41377-022-00743-6

[16] M. Akçakaya, B. Yaman, H. Chung, and J. C. Ye, "Unsupervised deep learning methods for biological image reconstruction and enhancement: An overview from a signal processing perspective," *IEEE Signal Processing Magazine*, vol. 39, no. 2, pp. 28–44, 2022. https://doi.org/10.1109/MSP.2021.3119273

[17] M. Mujahid, F. Rustam, R. Álvarez, J. Luis Vidal Mazón, I. D. L. T. Díez, and I. Ashraf, "Pneumonia classification from X-ray images with inception-V3 and convolutional neural network," *Diagnostics*, vol. 12, no. 5, p. 1280, 2022. https://doi.org/10.3390/diagnostics12051280

[18] N. S. Kavya, N. Veeranjaneyulu, and D. D. Priya, "Detecting Covid19 and pneumonia from chest X-ray images using deep convolutional neural networks," *Materials Today: Proceedings*, vol. 64, pp. 737–743, 2022. https://doi.org/10.1016/j.matpr.2022.05.199

[19] P. K. Wong *et al.*, "Automatic detection of multiple types of pneumonia: Open dataset and a multi-scale attention network," *Biomedical Signal Processing and Control*, vol. 73, p. 103415, 2022. https://doi.org/10.1016/j.bspc.2021.103415

[20] K. El Asnaoui, "Design ensemble deep learning model for pneumonia disease classification," *Int. J. Multimed. Inf. Retr.*, vol. 10, pp. 55–68, 2021. https://doi.org/10.1007/s13735-021-00204-7

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. https://doi.org/10.48550/arXiv.1409.1556

[22] M. H. Al-Adhaileh *et al.*, "Deep learning algorithms for detection and classification of gastrointestinal diseases," *Complexity*, vol. 2021, no. 1, p. 6170416, 2021. https://doi.org/10.1155/2021/6170416

[23] H. G. M., M. K. Gourisaria, S. S. Rautaray, and M. Pandey, "Pneumonia detection using CNN through chest X-ray," *Journal of Engineering Science and Technology (JESTEC)*, vol. 16, no. 1, pp. 861–876, 2021. https://jestec.taylors.edu.my/Vol%2016%20issue%201%20February%202021/16_1_61.pdf

[24] S. A. Aljawarneh and R. Al-Quraan, "Pneumonia detection using enhanced convolutional neural network model on chest X-ray images," *Big Data*, vol. 13, no. 1, pp. 16–29, 2025. https://doi.org/10.1089/big.2022.0261

[25] H. Sharma, J. S. Jain, P. Bansal, and S. Gupta, "Feature extraction and classification of chest X-ray images using CNN to detect pneumonia," in *2020 10th Int. Conf. Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2020, pp. 227–231. https://doi.org/10.1109/Confluence47617.2020.9057809

[26] D. Zhang, T. Zhang, J. Zhang, and X. Li, "Pneumonia detection from chest X-ray images based on convolutional neural network," *Electronics*, vol. 10, no. 13, p. 1512, 2021. https://doi.org/10.3390/electronics10131512

[27] C.-T. Yen and C.-Y. Tsao, "Lightweight convolutional neural network for chest X-ray images classification," *Scientific Reports*, vol. 14, p. 29759, 2024. https://doi.org/10.1038/s41598-024-80826-z

[28] S. Singh and B. K. Tripathi, "Pneumonia classification using quaternion deep learning," *Multimedia Tools and Applications*, vol. 81, pp. 1743–1764, 2022. https://doi.org/10.1007/s11042-021-11409-7

[29] K. Jakhar and N. Hooda, "Big data deep learning framework using Keras: A case study of pneumonia prediction," in *9th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, 2018, pp. 1–5. https://doi.org/10.1109/ICCCNT.2018.8493963

[30] A. A. Saraiva *et al.*, "Classification of images of childhood pneumonia using convolutional neural networks," in *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2019) – BIOIMAGING*, 2019, pp. 112–119. https://doi.org/10.5220/0007404301120119

[31] S. Lee, W. J. Kim, J. Chang, and J. C. Ye, "LLM-CXR: Instruction-finetuned LLM for CXR image understanding and generation," *arXiv preprint arXiv:2305.11490*, 2023. https://doi.org/10.48550/arXiv.2305.11490

[32] R. Tadokoro, R. Yamada, K. Nakashima, R. Nakamura, and H. Kataoka, "Primitive geometry segment pre-training for 3D medical image segmentation," *arXiv preprint arXiv:2401.03665*, 2024. https://doi.org/10.48550/arXiv.2401.03665

[33] G. Vrbančič and V. Podgorelec, "Transfer learning with adaptive fine-tuning," *IEEE Access*, vol. 8, pp. 196197–196211, 2020. https://doi.org/10.1109/ACCESS.2020.3034343

[34] T. K. Ho and J. Gwak, "Multiple feature integration for classification of thoracic disease in chest radiography," *Appl. Sci.*, vol. 9, no. 19, p. 4130, 2019. https://doi.org/10.3390/app9194130

[35] K. El Asnaoui, Y. Chawki, and A. Idri, "Automated methods for detection and classification of pneumonia based on X-ray images using deep learning," in *Artificial Intelligence and Blockchain for Future Cybersecurity Applications. Studies in Big Data*, Y. Maleh, Y. Baddi, M. Alazab, L. Tawalbeh, I. Romdhani, Eds., vol. 90, 2021, pp. 257–284. https://doi.org/10.1007/978-3-030-74575-2_14

[36] A. K. Jaiswal, P. Tiwari, S. Kumar, D. Gupta, A. Khanna, and J. J. P. C. Rodrigues, "Identifying pneumonia in chest X-rays: A deep learning approach," *Measurement*, vol. 145, pp. 511–518, 2019. https://doi.org/10.1016/j.measurement.2019.05.076

[37] I. Sirazitdinov, M. Kholiavchenko, T. Mustafaev, Y. Yixuan, R. Kuleev, and B. Ibragimov, "Deep neural network ensemble for pneumonia localization from a large-scale chest X-ray database," *Comput. Electr. Eng.*, vol. 78, pp. 388–399, 2019. https://doi.org/10.1016/j.compeleceng.2019.08.004

[38] G. Liang and L. Zheng, "A transfer learning method with deep residual network for pediatric pneumonia diagnosis," *Comput. Methods Programs Biomed.*, vol. 187, p. 104964, 2020. https://doi.org/10.1016/j.cmpb.2019.06.023

[39] G. Ling and C. Cao, "Atomatic detection and diagnosis of severe viral pneumonia CT images based on LDA-SVM," *IEEE Sens. J.*, vol. 20, no. 20, pp. 11927–11934, 2020. https://doi.org/10.1109/JSEN.2019.2959617

[40] M. Toğaçar, B. Ergen, Z. Cömert, and F. Özyurt, "A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models," *IRBM*, vol. 41, no. 4, pp. 212–222, 2020. https://doi.org/10.1016/j.irbm.2019.10.006

[41] V. Chouhan *et al.*, "A novel transfer learning based approach for pneumonia detection in chest X-ray images," *Appl. Sci.*, vol. 10, no. 2, p. 559, 2020. https://doi.org/10.3390/app10020559

[42] A. Mabrouk, R. P. Díaz Redondo, A. Dahou, M. Abd Elaziz, and M. Kayed, "Pneumonia detection on chest X-ray images using ensemble of deep convolutional neural networks," *Appl. Sci.*, vol. 12, no. 13, p. 6448, 2022. https://doi.org/10.3390/app12136448

[43] 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, 2021. https://doi.org/10.1371/journal.pone.0256630

[44] G. U. Nneji, J. Cai, J. Deng, H. N. Monday, E. C. James, and C. C. Ukwuoma, "Multi-channel based image processing scheme for pneumonia identification," *Diagnostics*, vol. 12, no. 2, p. 325, 2022. https://doi.org/10.3390/diagnostics12020325

[45] R. Pramanik, S. Sarkar, and R. Sarkar, "An adaptive and altruistic PSO-based deep feature selection method for pneumonia detection from Chest X-rays," *Appl. Soft Comput.*, vol. 128, p. 109464, 2022. https://doi.org/10.1016/j.asoc.2022.109464

[46] M. Yaseliani, A. Z. Hamadani, A. I. Maghsoodi, and A. Mosavi, "Pneumonia detection proposing a hybrid deep convolutional neural network based on two parallel visual geometry group architectures and machine learning classifiers," *IEEE Access*, vol. 10, pp. 62110–62128, 2022. https://doi.org/10.1109/ACCESS.2022.3182498

[47] Q. An, W. Chen, and W. Shao, "A deep convolutional neural network for pneumonia detection in X-ray images with attention ensemble," *Diagnostics*, vol. 14, no. 4, p. 390, 2024. https://doi.org/10.3390/diagnostics14040390

[48] A. Afifi, N. E. Hafsa, M. A. S. Ali, A. Alhumam, and S. Alsalman, "An ensemble of global and local-attention based convolutional neural networks for COVID-19 diagnosis on chest X-ray images," *Symmetry*, vol. 13, no. 1, p. 113, 2021. https://doi.org/10.3390/sym13010113

[49] O. H. Khater, A. S. Shuaib, S. Ul Haq, and A. J. Siddiqui, "AttCDCNet: Attention-enhanced Chest disease classification using X-ray images," in *2025 IEEE 22nd International Multi-Conference on Systems, Signals & Devices (SSD)*, 2024, pp. 891–896. https://doi.org/10.1109/SSD64182.2025.10989974

[50] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018. https://doi.org/10.1016/j.cell.2018.02.010

9 AUTHORS

Thi Thoa Mac serves as an Associate Professor at Hanoi University of Science and Technology in Vietnam. She received the B.Eng. degree in mechatronics engineering from Hanoi University of Science and Technology, Vietnam, in 2006. She obtained her master degree in 2009 at National Taiwan University of Science and Technology in Faculty of Engineering, Taiwan and obtained her Ph.D. degree in 2018 at Ghent University, Belgium. Her study interests include intelligent mechatronic and robotic systems covering the use of smart sensing systems, trajectory planning, SLAM, motion control, machine vision, intelligent control, optimization, and artificial intelligence.

Xuan Thuan Nguyen is currently serving as a Lecturer at the Department of Mechatronics, School of Mechanical Engineering, Hanoi University of Science and Technology. He received his Bachelor's degree in Mechatronics from Hanoi University of Science and Technology in 2012, completed his Master's program in Mechatronics there in 2014, and later obtained a Master's degree (2016) and a Ph.D. (2019) in Mechanical Engineering from Kyoto Institute of Technology, Japan. His study interests focus on vibration, control, robotics, and AI, aiming to develop intelligent robotic systems and advanced control methods to serve automation and smart manufacturing.

Huy Anh Bui is currently working as Lecturer at the Department of Mechatronics, School of Mechanical and Automotive Engineering, Hanoi University of Industry. He received his Bachelor's degree in Electrical–Electronic Engineering from Ho Chi Minh City University of Technology (HCMUT) in 2018, and his Master's degree in Electrical Engineering from Hanoi University of Science and Technology (HUST) in 2020. His study focuses on AI, computer vision, image processing, and robotics, with an emphasis on developing intelligent robotic systems and programming models for human–robot interaction.

Thanh Hung Nguyen serves as an Associate Professor at Hanoi University of Science and Technology in Vietnam. He earned his Ph.D. in Mechanical and Electrical Engineering from the National Taipei University of Technology in Taiwan. Thanh Hung's expertise extends across many fields, with a primary focus on machine vision, robotics, and engineering applications of AI. His commitment to advancing knowledge and contributing to these cutting-edge areas underscores his valuable contributions to academia and research.

Hoang Hiep Ly is currently a Lecturer at the School of Mechanical Engineering, Hanoi University of Science and Technology (HUST). He graduated with a Bachelor's degree from the Talented Program at HUST in 2015 and completed his Master's and PhD degrees at Nagoya Institute of Technology, Japan. He was a research fellowship and a researcher at Japan Society for the Promotion of Science, and the University of Tokyo, respectively. His main research focuses on robotics and assistive devices for upper-limb rehabilitation in post-stroke patients, haptic glove technology combined with virtual reality, and the development of intelligent medical devices applying artificial intelligence (E-mail: hiep.lyhoang@hust.edu.vn).