Journal of Artificial Intelligence and Technology, (Ahead of Print)
https://doi.org/10.37965/jait.2026.1236

ISTP
RESEARCH ARTICLE

AI-Based Approach for Detecting Pediatric Pneumonia from Chest X-rays

Saeed Hamouda,[1] Ayman Mohamed,[2] Ayman Elshenawy,[1] Abdelrahman Elsayed,[3] and Mohamed M. Reda Ali[3,4]

[1]Department of Networks and Cybersecurity, Faculty of Information Technology, Al-Ahliyya Amman University, Amman, Jordan

[2]Department of Data Science and Artificial Intelligence, Faculty of Information Technology, Al-Ahliyya Amman University, Amman 19328, Jordan

[3]Department of Computer Science, Faculty of Information Technology, Isra University, Amman, Jordan

[4]Climate Change Information Center and Expert Systems, Agricultural Research Center (ARC), Cairo, Egypt

(Received 02 March 2026; Revised 14 March 2026; Accepted 23 April 2026; Published online 20 June 2026)

Abstract: Pneumonia remains a significant health challenge for children, particularly in regions with limited diagnostic resources. Early and accurate detection is essential for timely treatment and improved clinical outcomes. This paper presents an automated deep learning (DL)-based framework for pediatric pneumonia detection (PPD) using chest X-ray images, designed to support clinical decision-making and reduce diagnostic workload. Five convolutional neural networks (CNNs) architectures: DenseNet121, MobileNetV2, VGG16, InceptionV3, and ResNet50 are trained and evaluated through extensive experimentation. Their performance is assessed using accuracy, precision, recall, and F1-score, alongside confusion matrices. Results demonstrate that DenseNet121 and MobileNetV2 achieved superior performance, with an accuracy of 93.75%, F1-scores of nearly 95.1 %, outperforming the other models in balancing sensitivity and specificity. VGG16 achieves competitive performance with high sensitivity, whereas InceptionV3 and ResNet50 exhibit limitations, particularly in terms of generalization and specificity. The system also demonstrates scalability potential, with MobileNetV2 offering lightweight deployment capabilities for low-resource environments. The findings confirm the clinical value of DL systems in supporting PPD and highlight the promise of DenseNet121 and MobileNetV2 as practical solutions for real-world healthcare applications.

Keywords: chest X-rays; deep learning; MobileNetV2; PPD; ResNet50; VGG16

I. INTRODUCTION

Pneumonia is a lung infection caused by viral, bacterial, or fungal pathogens transmitted through airborne droplets [1,2]. It is considered the leading cause of childhood mortality [3]. It causes coughing, fever, shortness of breath, chest pain, chills, fatigue, or sweating, with the highest risk in young children and adults over 65 years of age [4]. Its diagnosis is done by analyzing X-rays, CT scans, and magnetic resonance imaging (MRI). Chest X-rays are especially important for identifying lung infiltrates, consolidations, and complications like abscesses [5]. Pediatric pneumonia detection (PPD) is a challenge due to vague symptoms and limited resources, while misdiagnosis, high costs, overcrowding, and long waits have increased interest in automated diagnostic systems [6–10].

Previous methods for PPD rely on capturing radiographic patterns using texture analysis, histogram-based, and wavelet transforms and classifying the output using traditional machine learning (ML), such as support vector machines (SVMs), decision trees (DTs), or K-nearest neighbors (KNNs) [6,11,12]. These methods suffered from limited generalizability due to their dependency on manually engineered features that often failed to capture the complexity and variability of medical images [13,14]. Despite incremental improvements, these methods lacked robustness and adaptability to real-world scenarios [15–17].

Deep learning (DL), particularly CNNs, plays a pivotal role in medical image analysis because it can automatically learn complex hierarchical features from imaging data such as X-rays, CT scans, and MRIs, thereby enabling more accurate and reliable disease diagnosis compared to traditional methods, ultimately supporting clinicians in early intervention, reducing diagnostic errors, and improving patient outcomes [12,18–24].

Architectures such as VGG, ResNet, and DenseNet have demonstrated strong diagnostic performance in clinical practice [6,24]. However, many prior studies were limited by methodological shortcomings, including reliance on mixed datasets of adults and children, and inconsistent evaluation metrics, which hinder fair benchmarking and generalizability [25–27]. This highlights the necessity for systematic evaluation of state-of-the-art CNN models on standardized pediatric datasets to assess their true diagnostic performance.

This paper proposes a framework for PPD by systematically evaluating five state-of-the-art CNN architectures (MobileNetV2,

Corresponding author: Ayman Elshenawy (e-mail: a.elshenawy@ammanu.edu.jo).

DenseNet121, InceptionV3, VGG16, and ResNet50) on a chest X-ray dataset. Leveraging transfer learning, the pretrained networks were adapted by replacing their top layers with a custom binary classifier and further optimized through fine-tuning, early stopping, and dropout regularization. Their performance is assessed using accuracy and F1-score metrics, ensuring reliable and clinically relevant benchmarking of DL approaches for PPD. Findings indicate that DenseNet121 and MobileNetV2 are the most clinically reliable models, achieving 93.75%, 95.1% for accuracy, and F1-score, respectively, making them the preferred choice while requiring maximum diagnostic precision. Overall, this study not only benchmarks the performance of leading CNNs for PPD but also establishes a reproducible and clinically relevant methodology to support future AI-driven diagnostic applications in PPD.

The main contributions of this paper are summarized as follows:

i. A systematic and fair comparison of five widely used DL architectures for PPD using chest X-ray images within a unified experimental framework.

ii. A consistent training pipeline in which identical preprocessing, data augmentation, and transfer learning strategies are applied across all models to ensure a reliable and unbiased comparison.

iii. An analysis of practical deployment considerations, identifying DenseNet121 as the most diagnostically reliable model while highlighting MobileNetV2 as a lightweight alternative suitable for deployment in resource-constrained environments.

The paper is structured as follows: Section II reviews related work, Section III outlines the methodology, Section IV presents results, Section V discusses findings, and Section VI concludes with implications.

II. RELATED WORK

AI-based diagnostic tools, such as CNN, have achieved performance levels comparable to, and in some cases exceeding, those of human experts in medical diagnosis [27,33,34]. For example, CheXNet, a 121-layer CNN trained on more than 100,000 frontal chest X-rays from the ChestX-ray dataset, outperforms the average diagnostic accuracy of four experienced radiologists [35,36]. CNN-based systems have demonstrated consistently strong results in both metrics [37,38]. Furthermore, in terms of average weighted error rates, CNN models have outperformed individual radiologists [39].

Table I summarizes key related works, highlighting different models, datasets, image types, evaluation metrics, and performance. It highlights the growing effectiveness of DL approaches in PPD. These prior studies have reported high accuracy across a range of CNN architectures and datasets, highlighting the promise of DL for PPD. Advanced architectures are DenseNet201[40], CMixNet [41], and DCDD_Net [42]. Classical CNNs [28,29] as well as more advanced architectures such as Enhanced CNNs [30], Attention-based DenseNets [31], and Inception V3 [23] have achieved high accuracy, often exceeding 90%. These models benefit from deeper representations, attention mechanisms, or a hybrid design strategy that enables them to capture complex and subtle patterns in radiographic data.

Despite these advances, several weaknesses remain apparent. Many studies rely heavily on accuracy as the primary evaluation metric [28,29], with limited attention to other clinical measures such as sensitivity and specificity, which are crucial in medical diagnosis. The datasets employed are often limited in size [30], which may result in overfitting and poor generalizability to real-world hospital settings. This paper addresses these weaknesses by systematically comparing five state-of-the-art CNN models on a pediatric chest X-ray dataset to identify the most effective model for automated PPD.

III. METHODOLOGY

As shown in Fig. 1, the methodology adopted in this paper follows a structured pipeline to ensure reliable evaluation of the selected DL models. It begins with preparing the dataset, followed by

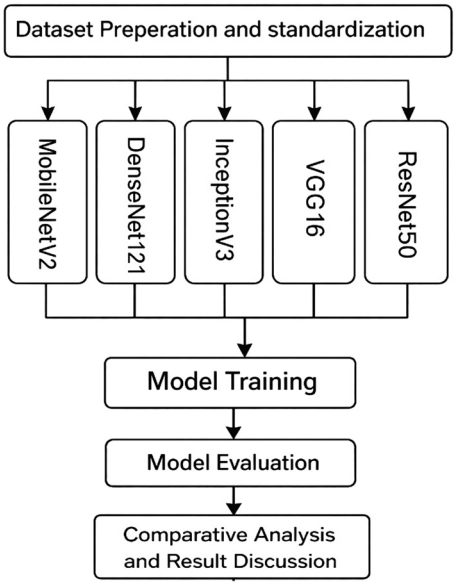

Fig. 1. Model development and evaluation pipeline.

Table I. Related work results

Ref.	Model	Dataset	Type of image	No. of images	Evaluation measures	Accuracy (%)
[30]	CNN	Kaggle dataset ()	X-ray	5840	Accuracy	92.31
[31]	Enhanced CNN model	Chest X-ray	X-ray	5863	Precision, Recall, Accuracy	92.4
[32]	Attention-based DenseNet	Chest X-ray	X-ray	5857	Precision, Recall, F1-score, Accuracy	92.8
[23]	Inception V3	Optical Coherence Tomography dataset (OCT)/(X-ray images)	X-ray	207,130	Sensitivity, Specificity, Accuracy	92.8% for chest X-rays

applying transfer learning techniques to adapt pretrained CNNs for the classification task, leveraging existing knowledge from image datasets to the chest X-ray domain. To further enhance performance, the models undergo fine-tuning to optimize their feature extraction capabilities for medical imaging. The final stage uses multiple performance metrics to evaluate the suggested models, providing a well-rounded assessment of diagnostic effectiveness.

A. THE DATASET

The proposed models are evaluated using a publicly available pediatric chest X-ray dataset introduced by Kermany *et al.* [23], which was originally collected from Guangzhou Women and Children's Medical Center, Guangzhou, China. It is composed of 5,856 X-ray images of anterior–posterior chest views, carefully selected from retrospective pediatric patients aged 1 to 5 years distributed across two categories: Normal or Pneumonia. Figure 2 displays a sample image from the dataset, and the statistics of the dataset are shown in Fig. 3.

The distribution indicates that pneumonia cases in the training set provide the models with a strong representation of infection patterns. The dataset thus provides a comprehensive and well-structured foundation for assessing DL models, combining diversity in disease types with practical challenges in class balance that reflect real-world diagnostic scenarios.

The dataset was divided into training, validation, and testing subsets to ensure reliable evaluation of the proposed models. Specifically, 5,232 images (approximately 90%) were used for training and validation, while 624 images (approximately 10%) were reserved for testing, following the dataset's predefined split. From the training portion, a validation subset (10%) was further separated and used for model monitoring during training. In addition, the dataset split was performed in a patient-independent manner, ensuring that images from the same patient did not appear in both training and testing sets. This strategy prevents data leakage and allows for a more realistic evaluation of model generalization.

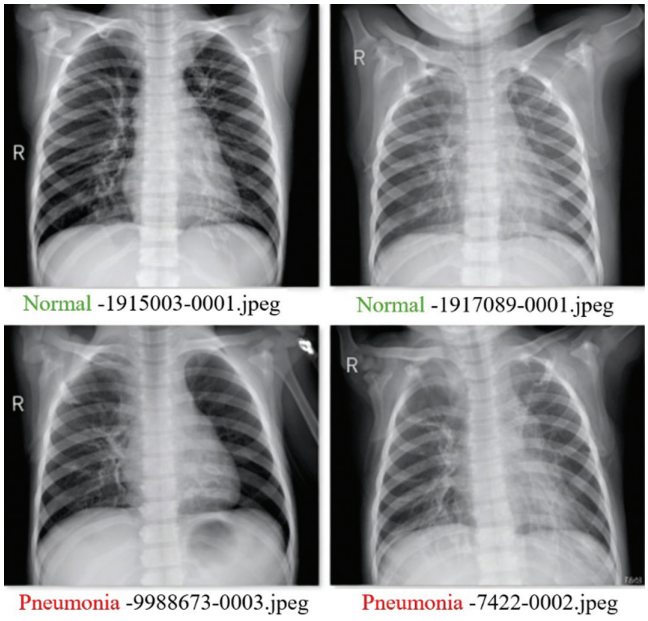

Fig. 2. Sample images from the dataset.

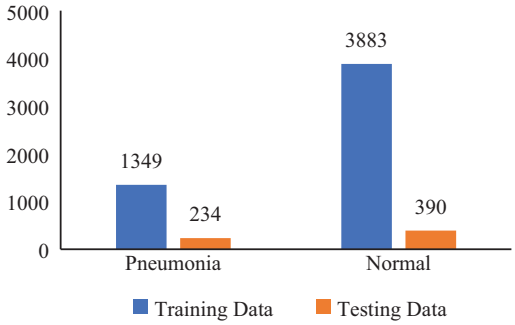

Fig. 3. The datasets involved (training and testing).

B. DATA PREPROCESSING

To ensure compatibility with the pretrained CNN architectures, all images were resized to a uniform dimension matching the selected model's standard input size and converted to the RGB color format. This step standardizes the input size and format, which is critical for models trained on similar specifications, such as those in the ResNet or VGG families.

The pixel intensity values of each image were normalized to the [0, 1] range by scaling by 1/255. This normalization step helps stabilize the training process by ensuring that the input features fall within a consistent numerical range. By reducing the variance in pixel values, the model can converge more efficiently during training, improving both performance and reliability.

C. Model Architecture

As shown in Figure 4, the overall architecture adopted in this study provides a standardized pipeline for comparative analysis of five state-of-the-art CNN models for PPD from chest X-ray images. The pipeline consists of the following stages:

i. Input Stage: Standardized chest X-ray images of size 224×224 pixels for ResNet50, VGG16, InceptionV3, and MobileNetV2, while DenseNet121 uses 256×256 pixels to leverage its capacity for richer feature extraction. These images are passed through pretrained CNN architectures originally trained on the ImageNet dataset. To adapt these models to the medical imaging domain, the top classification layers of the original architectures are removed, preserving only the convolutional base as a feature extractor.

- Data Augmentation: To improve generalization, augmentation is applied differently depending on the selected architecture. (i) A slightly richer augmentation is applied in the case of DenseNet121, including small rotations (±4°), slight contrast adjustments (±10%), and minor translations (±2%), to leverage the higher input resolution. (ii) For other architectures (ResNet50, VGG16, InceptionV3, and MobileNetV2), a lighter augmentation is applied, with only small rotations (±2°) and contrast change of (±5%) to maintain stability with standard 224×224 inputs.
- Feature Extraction: A custom classifier head is appended to each model to enable binary classification between pneumonia and normal cases. This head includes:
  i. A Flatten layer to convert feature maps into a one-dimensional vector.

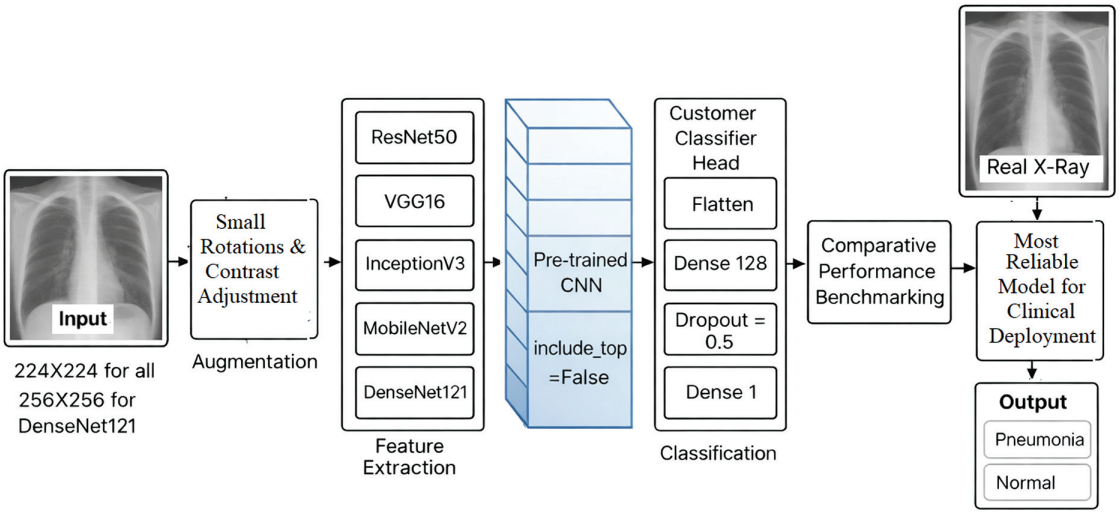

Fig. 4. Global model architecture for PPD.

ii. A Dense layer with 128 neurons and ReLU activation to introduce nonlinearity and enhance learning capacity.

iii. A Dropout layer with a rate of 0.5 to mitigate overfitting.

iv. A Dense output layer with a single neuron and sigmoid activation to produce probability scores for binary classification.

D. MODEL TRAINING

The proposed architecture employs transfer learning by leveraging convolutional bases from pretrained models (ResNet50, VGG16, InceptionV3, DenseNet121, and MobileNetV2) that were originally trained on the large-scale ImageNet dataset. These layers capture hierarchical features, from basic edges to complex shapes, that generalize well to chest X-rays. Initially, the convolutional base is frozen to preserve these representations, and only the custom classifier layers are trained on the pediatric pneumonia dataset. This staged approach speeds up convergence, lowers computational cost, and mitigates overfitting, providing a strong performance baseline.

Since the dataset exhibits class imbalance, with pneumonia cases, a class weighting strategy is applied during model training to reduce bias toward the majority class. Class weights were computed based on the inverse frequency of each class and incorporated into the loss function during training. This approach assigns a higher penalty to misclassification of the minority class, encouraging the model to pay greater attention to underrepresented samples. In addition, evaluation metrics such as precision, recall, and F1-score were used alongside accuracy to provide a more balanced assessment of model performance under imbalanced data conditions.

All evaluated CNN architectures are trained under a unified configuration to ensure fair and reproducible comparison. Transfer learning was applied by removing the original top classification layers of the pretrained networks (include_top = False) and retaining the convolutional base as a feature extractor. A custom classifier head was then added for binary classification. Training used the Adam optimizer with binary cross-entropy loss, an initial learning rate of 0.001, and a batch size of 32. The convolutional base was initially frozen so that only the newly added classifier layers were trained.

After this stage, fine-tuning was performed by unfreezing the last 50 layers of the network and retraining them using a reduced learning rate of 0.0001 to enable stable adaptation of the learned features to the pediatric chest X-ray dataset. Training was conducted for up to 25 epochs, with early stopping based on validation AUC to prevent overfitting. In addition, mixed-precision training (mixed_float16) was enabled to improve computational efficiency.

IV. EXPERIMENTAL RESULTS

The experiments are conducted on Google Colab Pro+, a cloud-based environment optimized for DL. Python 3.9 with TensorFlow 2.x and Keras is used for model development, training, and evaluation. Data preprocessing and augmentation are performed with OpenCV and TensorFlow utilities, while evaluation metrics (accuracy, precision, recall, and F1-score) were computed using scikit-learn.

Figure 5 presents the confusion matrices of the five implemented DL for classifying chest X-ray images as Normal or Pneumonia. DenseNet121 achieves 386 true positives (TP) and 199 true negatives (TN), with only 4 false negatives (FN) and 35 false positives (FP), making it highly reliable due to its very low missed pneumonia cases. MobileNetV2 records 374 TP and 211 TN, with 16 FN and 35 FP, showing slightly lower sensitivity but stronger specificity and good efficiency for resource-limited settings. In contrast, ResNet50 achieves 382 TP and 124 TN but produces 110 FP and 8 FN, indicating poor specificity despite a low FN rate. VGG16 correctly classifies most cases with 4 FN and 49 FP, maintaining balanced sensitivity, while InceptionV3 shows 20 FN and 49 FP, reducing its reliability compared to DenseNet121 and VGG16. Overall, DenseNet121 provides the best balance between sensitivity and specificity among the evaluated models.

As a result, DenseNet121 achieves high accuracy in PPD while keeping FN very low, making it highly dependable for diagnostic purposes. InceptionV3, while accurate overall, has a noticeably higher FN count (20), which reduces its sensitivity and makes it less reliable when compared to VGG16 or DenseNet121. This suggests that although InceptionV3 is effective in feature

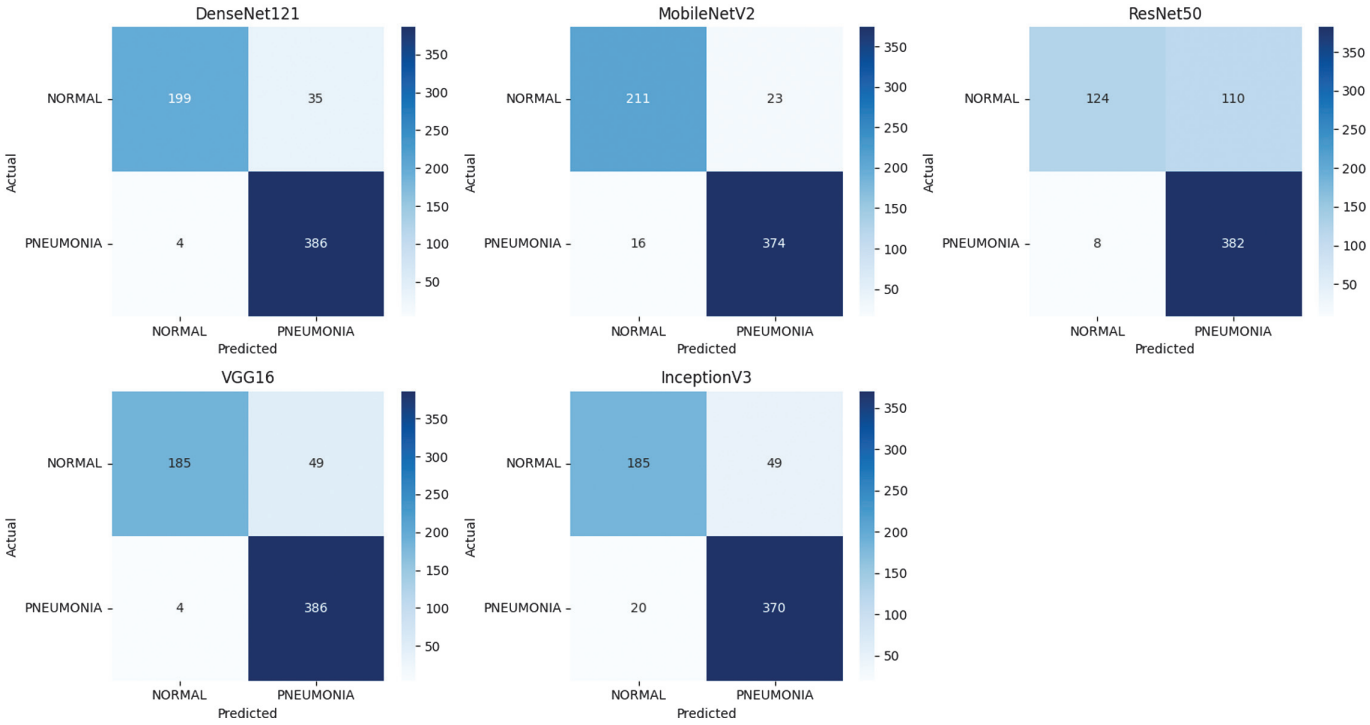

Fig. 5. Confusion matrix for five CNN architectures.

extraction, it may require fine-tuning for this specific medical application.

Table II and Fig. 6 provide a clear and practical comparison of the DL models evaluated for PPD detection, showing noticeable differences in both performance and real-world applicability. The comparative analysis of the evaluated DL models for pneumonia detection reveals noticeable differences in overall classification performance. Among all models, DenseNet121 achieved the strongest overall results, obtaining the highest F1-score (95.19%) with excellent accuracy (93.75%), high precision (91.69%), and outstanding recall (98.97%), indicating its strong capability to correctly detect pneumonia cases while maintaining balanced predictions.

MobileNetV2 delivered nearly equivalent performance, matching the highest accuracy (93.75%) and achieving the best precision (94.20%) with a slightly lower recall (95.90%), making it particularly attractive for practical or resource-constrained deployments due to its efficiency and competitive accuracy. VGG16 also performed well, achieving 91.51% accuracy and an F1-score of 93.58%, supported by very high recall (98.97%) but comparatively lower precision, suggesting a higher false-positive rate.

Table II. Performance comparison of DL models for PPD

Model	Accuracy %	Precision %	Recall %	F1-score %
ResNet50	81.09	77.64	97.95	86.62
InceptionV3	88.94	88.31	94.87	91.47
VGG16	91.51	88.74	98.97	93.58
MobileNetV2	93.75	94.20	95.90	95.05
DenseNet121	93.75	91.69	98.97	95.19

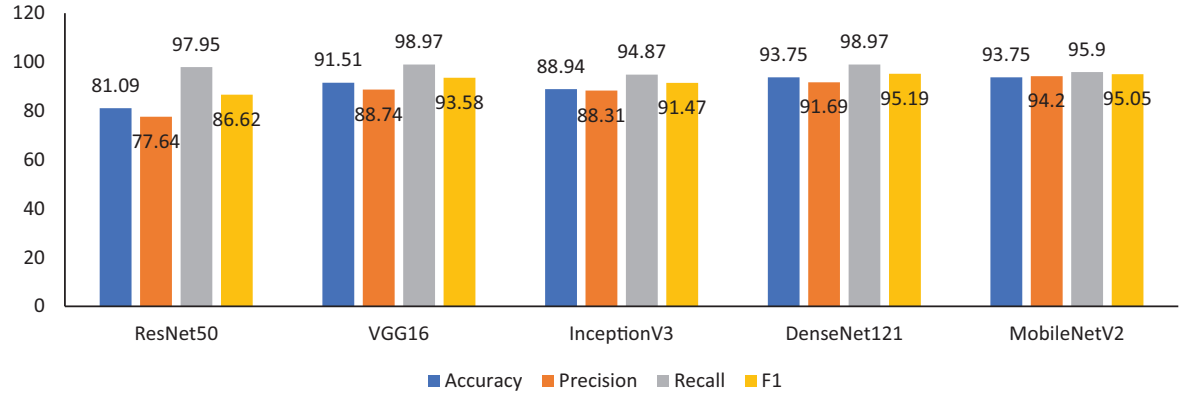

Fig. 6. Performance metrics of evaluated CNN models.

Table III. Performance comparison between DenseNet121 and other models

Ref.	Model	Dataset	Type of image	Accuracy (%)
[29]	ResNet50	Chest X-ray + COVID19	X-ray	91.39
[30]	CNN	Kaggle dataset ()	X-ray	92.31
[31]	Enhanced CNN model	Chest X-ray	X-ray	92.4
[32]	Attention-based DenseNet	Chest X-ray	X-ray	92.8
[23]	Inception V3	Optical tomography dataset	X-ray	92.8%
This paper	DenseNet121	Chest X-ray	X-ray	93.75%

InceptionV3 showed moderately strong and balanced results (accuracy 88.94% and F1-score 91.47%), though it did not surpass the leading models. In contrast, ResNet50 demonstrated the weakest overall performance, with the lowest accuracy (81.09%), precision (77.64%), and F1-score (86.62%), despite maintaining high recall (97.95%), indicating a tendency toward increased FP. Overall, the findings suggest that DenseNet121 and MobileNetV2 provide the most reliable and balanced performance for PPD, with DenseNet121 showing a slight advantage in overall diagnostic effectiveness.

In terms of deployment, MobileNetV2's lightweight design makes it well suited for mobile or resource-limited environments, while DenseNet121, though more computationally intensive, fits well within hospital systems equipped with GPU resources. The modular structure of the system also allows for easy retraining and fine-tuning with new data, supporting scalability and adaptability across different clinical settings without sacrificing reliability.

Table III and Fig. 7 provide a straightforward comparison between the DenseNet121 model proposed in this study and previously published approaches, making the performance differences easy to see. Earlier methods reported accuracies of 91.39% with ResNet50 [29], 92.31% with a standard CNN [30], and 92.4% with an enhanced CNN model [31], while both attention-based DenseNet and InceptionV3 reached 92.8% [23,32].

In comparison, the DenseNet121 model in this work achieved 93.75% accuracy, clearly surpassing the results reported in prior studies. The accompanying chart visually emphasizes this improvement, showing a noticeable increase in accuracy for the proposed approach. Overall, these findings suggest that the chosen architecture and optimization strategy contributed meaningfully to the performance gain, supporting the reliability of DenseNet121 for chest X-ray-based PPD detection.

Despite these competitive results, DenseNet121 achieved an accuracy of 93.75% on the evaluated dataset. While this result is competitive with previously reported studies, direct performance comparisons should be interpreted with caution because the referenced works were evaluated on different datasets and experimental settings. This performance boost illustrates the strength of DenseNet121's densely connected architecture, which enhances feature reuse, mitigates the vanishing gradient problem, and improves overall generalization. The bar chart in Fig. 7 clearly demonstrates this improvement, as DenseNet121 stands above the rest, reflecting its potential as a robust framework for PPD and similar medical imaging applications.

To assess how well the trained models generalize beyond the training data, the final DenseNet121 model was applied to unseen images from the test set. Figure 8 illustrates several sample predictions made by DenseNet121 on randomly selected chest X-rays. As shown, the model successfully distinguishes between pneumonia and normal cases, assigning each image a classification label.

Figure 9 extends this visual assessment by presenting predictions from the remaining five CNN architectures, MobileNetV2, ResNet50, VGG16, InceptionV3, and DenseNet121, on a similar set of unseen X-ray images (with the true class displayed above each image).

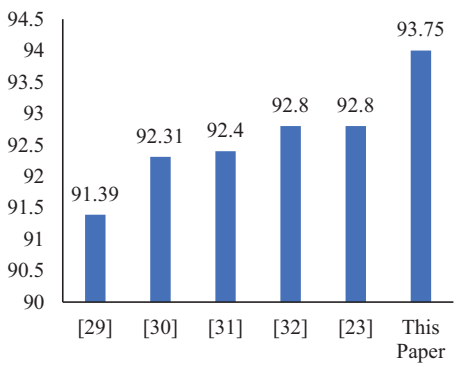

Fig. 7. Comparison between DenseNet121 accuracy and other models. Testing and validation.

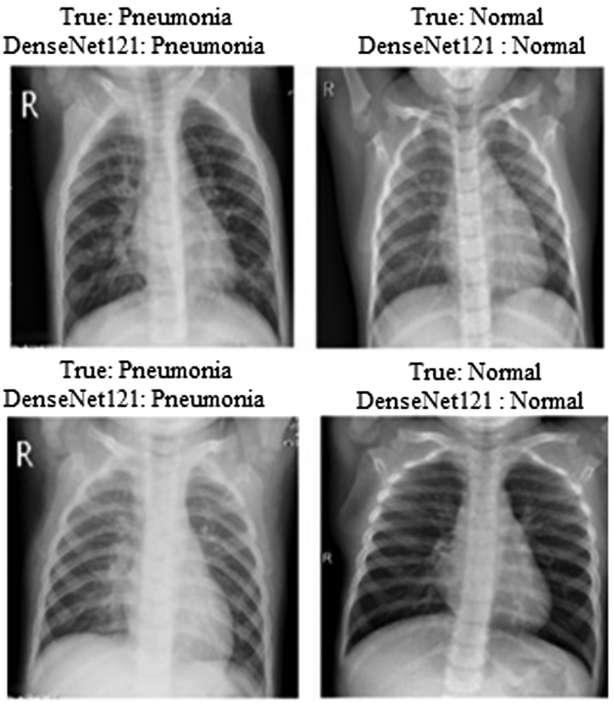

Fig. 8. Using DenseNet121 for PPD.

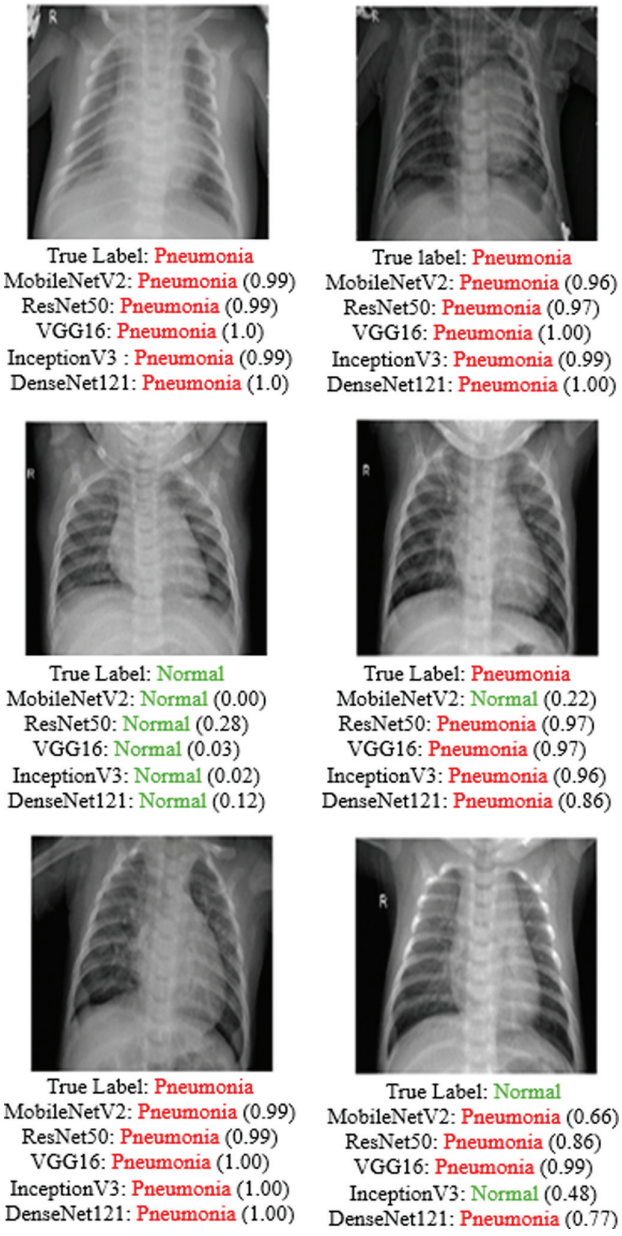

Fig. 9. Visualization of model predictions for selected chest X-ray images. The true diagnosis (Pneumonia or Normal) is shown at the top of each panel, followed by the predicted class and confidence value from each evaluated CNN architecture.

Correctly classified pneumonia examples show that models such as DenseNet121 and MobileNetV2 consistently assign very high confidence scores (often close to or equal to 1.00), demonstrating strong sensitivity in detecting infected cases. Similarly, in correctly identified normal cases, most models assign the NORMAL class a high probability, though some variation in confidence can be observed, particularly with ResNet50 and occasionally with InceptionV3.

The figure also highlights subtle differences in prediction certainty across models, with DenseNet121 generally producing the most confident and stable outputs across both classes. Overall, the visual results support the quantitative findings, confirming that DenseNet121 and MobileNetV2 provide more consistent and reliable predictions compared to the other evaluated architectures.

V. DISCUSSION

The evaluation of five DL architectures on chest X-ray images highlighted clear differences in classification performance. Confusion matrices showed that DenseNet121 achieved the best balance, with only 4 FN and 35 FP, effectively minimizing missed pneumonia cases while maintaining strong specificity. MobileNetV2 performed nearly as well, with slightly more FN but fewer FP, making it highly competitive and well suited for resource-constrained settings. In contrast, ResNet50 produced 110 FP, resulting in very low specificity, while VGG16 and InceptionV3 offered intermediate results, with VGG16 showing higher sensitivity and InceptionV3 yielding more FN.

Quantitative metrics reinforced these findings. DenseNet121 and MobileNetV2 achieved the highest accuracy (93.75%) with excellent sensitivity (98.97% and 95.90%) and strong specificity (85.04% and 90.17%). VGG16 followed with 91.51% accuracy and very high sensitivity (98.97%), though its specificity was moderate (79.06%). InceptionV3 showed weaker results (88.94% accuracy, 94.87% sensitivity, and 79.06% specificity), reflecting more frequent pneumonia misclassifications. ResNet50 performed the worst, with 81.09% accuracy and low specificity (52.99%), despite maintaining high sensitivity (97.95%), indicating a strong bias toward predicting pneumonia. These results confirm DenseNet121 and MobileNetV2 as the most clinically reliable models due to their balance of sensitivity and specificity.

Additional metrics further validated these outcomes. DenseNet121 and MobileNetV2 both achieved 0.94 accuracy and 0.95 F1-score, demonstrating not only strong accuracy but also stable decision boundaries. VGG16 remained competitive (0.92 accuracy and 0.94 F1-score), excelled in recall (0.99), while InceptionV3 lagged slightly with 0.89 accuracy and 0.92 F1-score. ResNet50 underperformed across all metrics (0.82 accuracy and 0.87 F1-score), consistent with its high false-positive rate. Learning curves further supported these observations: DenseNet121 showed the most stable progression with low risk of overfitting, while MobileNetV2 also displayed smooth training dynamics. VGG16 and InceptionV3 achieved acceptable accuracy but exhibited unstable validation loss during training, while ResNet50 showed the poorest generalization performance, characterized by a significant gap between training and validation results.

In summary, DenseNet121 proved to be the most effective and reliable model for PPD, combining high sensitivity, specificity, and training stability. MobileNetV2 emerged as an equally strong option with the added advantage of computational efficiency, making it ideal for low-resource clinical environments. VGG16 offered dependable performance, particularly in sensitivity, while InceptionV3 remained competitive but less robust. ResNet50, however, showed poor generalization and specificity, limiting its clinical suitability. These findings highlight the importance of balancing sensitivity and specificity in medical AI, with DenseNet121 and MobileNetV2 standing out as the most practical solutions for real-world deployment.

Table III provides a summary of representative results reported in previous studies for pneumonia detection. However, these results were obtained using different datasets, preprocessing strategies, and experimental protocols. Therefore, the comparison should be interpreted as a general overview of the performance range reported in the literature rather than a direct benchmark

comparison. Within the controlled experimental setup used in this study, DenseNet121 demonstrated the strongest performance among the evaluated models.

VI. CONCLUSION

This paper evaluated the performance of five state-of-the-art DL architectures (DenseNet121, MobileNetV2, VGG16, InceptionV3, and ResNet50) for PPD using chest X-ray images. The results show that DenseNet121 and MobileNetV2 achieved the best overall performance, obtaining the highest accuracy and F1-score values. VGG16 also demonstrated strong reliability, particularly with high sensitivity, making it suitable for scenarios where detecting all pneumonia cases is essential. In contrast, InceptionV3 achieved moderate performance, while ResNet50 showed lower specificity, which may limit its clinical suitability. Overall, the findings suggest that the proposed system can effectively support automated PPD. However, additional clinical validation is necessary before real-world deployment. Future work will focus on improving the system's practicality by incorporating explainable AI for better interpretability, optimizing lightweight models such as MobileNetV2 for deployment on mobile and edge devices in low-resource settings, and integrating the system with hospital information systems.

CONFLICT OF INTEREST STATEMENT

The author(s) declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

[1] G. Shih *et al.*, "Augmenting the national institutes of health chest radiograph dataset with expert annotations of possible pneumonia," *Radiol. Artif. Intell.*, vol. 1, no. 1, p. e180041, 2019.

[2] D. K. Smith, D. P. Kuckel, and A. M. Recidoro, "Community-acquired pneumonia in children: Rapid evidence review," *Am. Fam. Physician*, vol. 104, no. 6, pp. 618–625, 2021.

[3] N. Miyashita, "Atypical pneumonia: Pathophysiology, diagnosis, and treatment," *Respir. Investing.*, vol. 60, no. 1, pp. 56–67, 2022.

[4] A. N. Csep *et al.*, "Guide to diagnosis and treatment of acute pneumonia in children," *Arch. Pharm. Pract.*, vol. 15, no. 1-2024, pp. 17–22, 2024

[5] R. Kundu *et al.*, "Pneumonia detection in chest X-ray images using an ensemble of deep learning models," *PloS One*, vol. 16, no. 9, p. e0256630, 2021.

[6] O. Stephen *et al.*, "An efficient deep learning approach to pneumonia classification in healthcare," *J. Healthc. Eng.*, vol. 2019, no. 1, p. 4180949, 2019.

[7] J. Scott *et al.*, "The definition of pneumonia, the assessment of severity, and clinical standardization in the Pneumonia Etiology Research for Child Health study," *Clin. Infect. Dis.*, vol. 54, no. suppl_2, pp. S109–S116, 2012.

[8] A. B. Gupta *et al.*, "Inappropriate diagnosis of pneumonia among hospitalized adults," *JAMA Intern. Med.*, vol. 184, no. 5, pp. 548–556, 2024.

[9] I. Rudan *et al.*, "Epidemiology and etiology of childhood pneumonia in 2010: Estimates of incidence, severe morbidity, mortality, underlying risk factors and causative pathogens for 192 countries," *J. Glob. Health*, vol. 3, no. 1, p. 010401, 2013.

[10] R. Izadnegahdar *et al.*, "Childhood pneumonia in developing countries," *Lancet Respir. Med.*, vol. 1, no. 7, pp. 574–584, 2013.

[11] F. Alshanketi *et al.*, "Pneumonia detection from chest X-Ray images using deep learning and transfer learning for imbalanced datasets," *J. Imaging Inf. Med.*, vol. 38, no. 4, pp. 2021–2040, Nov. 2024, doi: 10.1007/s10278-024-01334-0.

[12] H. A. Owida *et al.*, "Deep learning algorithms to improve COVID-19 classification based on CT images," *Bull. Electr. Engineering Inf.*, vol. 11, no. 5, pp. 2876–2885, 2022

[13] N. Barakat, M. Awad, and B. A. Abu-Nabah, "A machine learning approach on chest X-rays for pediatric pneumonia detection," *Digit Health*, vol. 9, 20552076231180008, 2023.

[14] G. Huang *et al.*, "Densely connected convolutional networks," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708. 2017.

[15] F. Hussein *et al.*, "Hybrid clahe-cnn deep neural networks for classifying lung diseases from x-ray acquisitions," *Electronics*, vol. 11, no. 19, p. 3075, 2022.

[16] S. Iqbal *et al.*, "On the analyses of medical images using traditional machine learning techniques and convolutional neural networks," *Arch. Comput. Methods Eng.*, vol. 30, no. 5, pp. 3173–3233, Jun. 2023, DOI: 10.1007/s11831-023-09899-9.

[17] W. Khan, N. Zaki, and L. Ali, "Intelligent pneumonia identification from chest x-rays: A systematic literature review," *IEEE Access*, vol. 9, pp. 51747–51771, 2021.

[18] R. Tarek *et al.*, "Automated diagnosis of dental diseases using deep learning on radiographic images," *SN Comput. Sci.*, vol. 6, no. 6, p. 751, 2025.

[19] Y. W. Jin *et al.*, "Integrative data augmentation with U-Net segmentation masks improves detection of lymph node metastases in breast cancer patients," *Cancers (Basel)*, vol. 12, p. 2934, 2020, DOI: 10.3390/cancers12102934.

[20] H. A. Owida *et al.*, "Improved deep learning architecture for skin cancer classification," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 36, no. 1, pp. 501–508, 2024.

[21] A. Elshenawy, A. Mohammed, and S. Hamouda, "The evolution of deep learning: Models, applications, and future directions," in *2025 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), IEEE*, Sep. 2025, pp. 1–8. DOI: 10.1109/MIUCC66482.2025.11196834.

[22] E. Donelly, "Very deep convolutional networks for large-scale image recognition," *Int. J. Artif. Intell. Mach. Learn.*, vol. 2, no. 1, 2012.

[23] 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.

[24] D. Varshni *et al.*, "Pneumonia detection using CNN based feature extraction," In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pp. 1–7. IEEE, 2019.

[25] J. S. Park *et al.*, "Artificial intelligence models for pediatric lung sound analysis: Systematic review and meta-analysis," *J. Med. Internet Res.*, vol. 27, p. e66491, Apr. 2025, DOI: 10.2196/66491.

[26] 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.

[27] P. Daniel *et al.*, "Adults miscoded and misdiagnosed as having pneumonia: Results from the British Thoracic Society pneumonia audit," *Thorax*, vol. 72, no. 4, pp. 376–379, 2017.

[28] K. Wang *et al.*, "Attention-based DenseNet for pneumonia classification," *IRBM*, vol. 43, no. 5, pp. 479–485, 2022.

[29] N. S. Kavya, N. Veeranjaneyulu, and D. Divya Priya, "Detecting Covid19 and pneumonia from chest X-ray images using deep

convolutional neural networks," *Mater. Today Proc.*, vol. 64, pp. 737–743, 2022.

[30] 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.

[31] H. Malik *et al.*, "Deep learning-based classification of chest diseases using X-rays, CT scans, and cough sound images," *Diagnostics (Basel)*, vol. 13, no. 17, p. 2772, 2023.

[32] N. Nasrullah *et al.*, "Automated lung nodule detection and classification using deep learning combined with multiple strategies," *Sensors*, vol. 19, no. 17, p. 3722, 2019.

[33] H. A. Younis *et al.*, "A systematic review and meta-analysis of artificial intelligence tools in medicine and healthcare: applications, considerations, limitations, motivation and challenges," *Diagnostics (Basel)*, vol. 14, no. 1, p. 109, 2024.

[34] R. Al-Dmour *et al.*, "Impact of AI and big data analytics on healthcare outcomes: An empirical study in Jordanian healthcare institutions," *Digit. Health*, vol. 11, pp. 1–15, Jan. 2025.

[35] H. Bairwa and R. Jangid, "Pneumonia Detection from Chest X-rays Using the CheXNet Deep Learning Algorithm," Jul. 01, 2024.

[36] Y. Kumar *et al.*, "Artificial intelligence in disease diagnosis: A systematic literature review, synthesizing framework and future research agenda," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 7, pp. 8459–8486, 2023.

[37] M. Chavoshi, S. Zamani, and S. A. Mirshahvalad, "Diagnostic performance of deep learning models versus radiologists in COVID-19 pneumonia: A systematic review and meta-analysis," *Clin. Imaging*, vol. 107, p. 110092, Mar. 2024, DOI: 10.1016/j.clinimag.2024.110092.

[38] M. Nagendran *et al.*, "Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies," *BMJ*, vol. 368, pp. 1–11, 2020.

[39] X. Liu *et al.*, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis," *Lancet Digit. Health*, vol. 1, no. 6, pp. e271–e297, 2019.

[40] A. Bhandary *et al.*, "Deep-learning framework to detect lung abnormality–A study with chest X-Ray and lung CT scan images," *Pattern Recognit. Lett.*, vol. 129, pp. 271–278, 2020.

[41] T. Rahman *et al.*, "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray," *Appl. Sci.*, vol. 10, no. 9, p. 3233, 2020.

[42] M. Anthimopoulos *et al.*, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.