European Journal of Clinical and Biomedical Sciences
2025, Vol. 11, No. 5, pp. 60–72
https://doi.org/10.11648/j.ejcbs.20251105.11

Research Article

Enhancing Pneumonia Detection from Chest Radiographs Through a VGG-16-based Deep Learning Approach

Sourav Sana, Priyankar Biswas*, A. T. M. Saiful Islam

Department of Electrical and Electronic Engineering, Gopalganj Science and Technology University (GSTU), Gopalganj, Bangladesh

Abstract

Pneumonia is a significant respiratory disease with high global burdens, especially in resource limited settings where access to specialized radiology is restricted. Early and reliable diagnosis is essential for effective clinical intervention, yet manual interpretation of chest X-ray images is often time-consuming and subject to inter-observer variability. This framework employs deep learning for automated pneumonia detection using chest X-ray images, leveraging transfer learning with a pre-trained VGG-16 model and a custom DNN classifier that incorporates batch normalization and dropout layers to ensure stable training and prevent overfitting. The model achieved an accuracy of 92.79%, precision of 94.12%, recall of 94.36%, an F1-score of 94.24%, and an AUC of 0.98 on the public Chest X-Ray images (Pneumonia) dataset published on Kaggle, outperforming several state-of-the-art CNN methods. These performance metrics indicate that the proposed method exceeds several existing convolutional neural network-based techniques reported in contemporary studies. To enhance clinical transparency, Gradient weighted Class Activation Mapping (Grad-CAM) was utilized to visualize salient regions contributing to the model's predictions, thereby improving interpretability and supporting potential clinical adoption. The results demonstrate that the framework is effective, computationally efficient, and capable of providing reliable diagnostic support. Its design makes it suitable for integration into real-time clinical decision support systems and telemedicine platforms, particularly in low-resource healthcare environments where rapid and accurate diagnostic tools are urgently needed.

Keywords

Pneumonia Detection, Chest X-rays, Transfer Learning, VGG-16, Deep Neural Networks, Medical Imaging

1. Introduction

Pneumonia remains a significant health concern worldwide, especially impacting children under five, with the World Health Organization (WHO) estimating around 740,000 deaths each year in this demographic [1]. Timely and accurate diagnosis is essential for effective treatment and mortality reduction. Chest radiography remains the most widely used imaging modality for pneumonia detection due to its accessibility and affordability. However, the interpretation of chest X-ray (CXR) images is highly dependent on radiological expertise, which introduces subjectivity and potential diagnostic delays challenges that are especially pronounced in resource-limited settings [2]. Convolutional neural networks and other recent developments in deep learning have greatly enhanced automated medical image processing, including the

*Corresponding author: priyankar@gstu.edu.bd (Priyankar Biswas)

Received: 11 November 2025; **Accepted:** 26 November 2025; **Published:** 11 December 2025

identification of pneumonia from CXRs. [3]. To address issues such as class imbalance and subtle feature extraction, techniques like attention-based contrastive learning have been proposed, enabling better localization of pathological features. Furthermore, hybrid deep learning architectures, which combine Inception modules with residual connections, have shown promise by facilitating efficient multi-scale feature extraction for pulmonary inflammation detection [4].

In situations where labeled medical data is scarce, transfer learning has become a very effective strategy, allowing for the adjustment of pre-trained models developed on extensive datasets for specific diagnostic tasks. Among these, the VGG-16 model stands out due to its architectural simplicity and strong feature representation capabilities. Several studies have demonstrated the effectiveness of fine-tuned VGG-16 models in pneumonia detection, affirming their clinical utility [5, 6]. Additionally, methods such as neural network adaptation have been employed to tailor these models for specific populations, such as pediatric patients, thereby enhancing diagnostic relevance. Beyond standard CNNs, customized deep learning frameworks have been developed, achieving significant gains in classification accuracy and robustness [7]. The COVID-19 pandemic further underscored the need for automated diagnostic systems, with pneumonia detection models being effectively repurposed for COVID-19-related pneumonia identification, highlighting their versatility. Moreover, attention-based models have improved both diagnostic accuracy and interpretability by directing model focus toward clinically relevant regions within CXR images [8, 9].

Expanding upon these advancements, this research presents a transfer learning framework that employs the VGG-16 architecture, augmented with extra deep neural layers, to effectively identify pneumonia in chest X-rays. Existing CNN-based pneumonia detection methods often suffer from limited generalization, overfitting, and lack of interpretability, restricting their clinical applicability. This study addresses these gaps by using a VGG-16-based transfer learning framework with custom DNN layers and Grad-CAM visualizations, and by systematically comparing it with multiple pre-trained CNN models to ensure higher accuracy, stability, and explainability. The primary contributions of this study consist of:

1) Development of a transfer learning-based diagnostic framework using the pre-trained VGG-16 network, specifically fine-tuned for pneumonia detection from CXR images.
2) Integration of custom deep neural network (DNN) layers including batch normalization and dropout to enhance model generalization, reduce overfitting, and improve classification performance.
3) Comprehensive evaluation of the proposed model on the publicly available CXR images dataset, achieving superior performance metrics (accuracy, precision, recall, F1-score, and AUC) compared to existing baseline CNN architectures.
4) Demonstration of real-world applicability, with the proposed system being lightweight and computationally efficient well-suited for deployment in telemedicine platforms and resource-constrained clinical settings.

2. Proposed Methodology

This work proposes a deep learning-based method for the detection of pneumonia from CXR images. It is based on transfer learning, using the VGG-16 model, and then explores five more pre-trained architectures: ResNet50, DenseNet169, MobileNet, InceptionV3, and Xception. The proposed pipeline includes some structured steps: dataset description, preprocessing, integrating transfer learning, model training, and developing explainable visualizations for interpretability. Each aspect of the methodology will be explained in detail in subsequent subsections.

2.1. Dataset Description

In this work, a total of 5,856 CXR images from a public Kaggle dataset [10], belonging to normal and pneumonia classes, have been utilized. The CXR image dataset is divided into three portions for efficient training and analysis of the proposed model. Among those, 5,216 CXR images are considered in the training subset of our proposed system, while 624 images within that dataset are utilized in testing the performance of our proposed system. A validation subset of 16 images is also considered in our proposed system for model improvement.

2.2. Input Data Processing

In order to prepare this dataset for a deep learning algorithm, a number of preprocessing steps need to be carried out. First, all images are resized with a 224 x 224 pixel dimension in order to satisfy one of the prerequisites of the VGG-16 model that its inputs should have a 224 x 224 pixel size. In order to make it possible for the developed model to generalize well, data augmentation methods are applied extensively to all images in the training dataset. This includes geometric modifications such as rotation, width and height movement, zooming, and shearing. Finally, all images are converted to full color by increasing or decreasing their intensity in a controlled manner to simulate real-life conditions. Finally, all pixels in an image are scaled down to a value ranging between 0 and 1.

2.3. Transfer Learning with VGG-16

The proposed model's baseline is based on the pre-trained VGG-16 architecture, which is a convolutional neural network pre-trained on ImageNet datasets. In this work, the convolutional part of a pre-trained VGG-16 architecture is

utilized as a static feature extractor that benefits from its hierarchical feature extraction capabilities in images. In order to adapt it as a binary classification model that identifies a pneumonia diagnosis, a custom classifier is introduced by replacing its upper layers. This custom classifier contains a global average pooling layer followed by three dense layers that successively decrease the number of units in the layers (512, 256, and 128 units, respectively). Each dense layer is followed by ReLU activation functions and batch normalization with a dropout of 40%. In particular, its output layer contains only one sigmoid unit that predicts probabilities of a classifying image in a pneumonia diagnosis. In order to improve its performance for this proposed task, its top four convolutional layers of a pre-trained VGG-16 architecture are modified to learn to identify pneumonia-related features in a more effective manner while capturing general knowledge that is learnt from ImageNet datasets. In order to make a fair comparison of this proposed work, five pre-trained models, such as ResNet50, DenseNet169, MobileNet, InceptionV3, and Xception, will be utilized.

2.4. Model Compilation and Training

The model is implemented with an Adam optimizer that has a learning rate of 0.0001. Also, binary cross-entropy is applied as it is a binary classification problem. Moreover, as it is an image recognition program, accuracy is considered the performance metric. It will train for 20 epochs with a batch size of 32. Keras' ImageDataGenerator is used to dynamically load and augment the training images in real time, which is especially useful for managing large datasets efficiently. The training takes place on a GPU to speed up calculations. Although the validation set is small, it serves to detect overfitting and inform training modifications. The test set remains unaltered during training and is only utilized for assessing the model's ultimate performance.

2.5. Visualization and Explainability

In order to gain an insight into its efficiency and approaches to reasoning, a set of visualization techniques is also employed in this work. Accuracy and loss curves are provided to study the learning trends as a function of increasing epochs in the training process. A confusion matrix in a heat map form is utilized to analyze its categorical predictions in terms of true positives, false positives, true negatives, and false negatives. In addition, an analysis involving a ROC curve is also provided to demonstrate its ability to distinguish between the two classes, and an Area Under Curve metric is a shorthand summary of such a performance metric. Gradient-weighted Class Activation Mapping is an extension intended to enhance the visual explainability of CXR images, as it highlights those areas of an image that are most relevant to its predictions. Test images are also provided as a set of triplets in order to offer a visual explanation of all three elements together in one common setting, including its original form, a heat map of its CAM visualization, and a visual representation of its probability of possessing a pneumonia diagnosis.

2.6. Overview of the Approach

In conclusion, this work describes a successful deep learning strategy for pneumonia detection in CXR images based on transfer learning with a VGG16 model as a building block. This proposed workflow incorporates structured preprocessing and strategic data augmentation techniques for generalization, followed by tailored modifications to a VGG16 architecture that include system classifier reforming and selective deeper layer-level tuning. To verify the reliability and effectiveness of this proposed system, five different pre-trained models of ResNet50, DenseNet169, MobileNet, InceptionV3, and Xception will be applied in an identical train setting. This comparative study will provide critical insight into a model-to-model quantitative comparison of all participating models and will also highlight the reliability of this proposed work based on its robust architecture designed using a VGG16 framework. This proposed system will train using an optimized learning system setup, with system performance validated through a combination of statistical and visualization techniques. Novel techniques such as Gradient-CAM visualization, an adequately described visualization of an ROC curve, and logical prediction plotting based upon system-level confidence will all be utilized in an aim to enhance system-level transparency that in turn is critical for its acceptance within a healthcare setup because it will act as a balanced system that is highly precise as well as transparent.

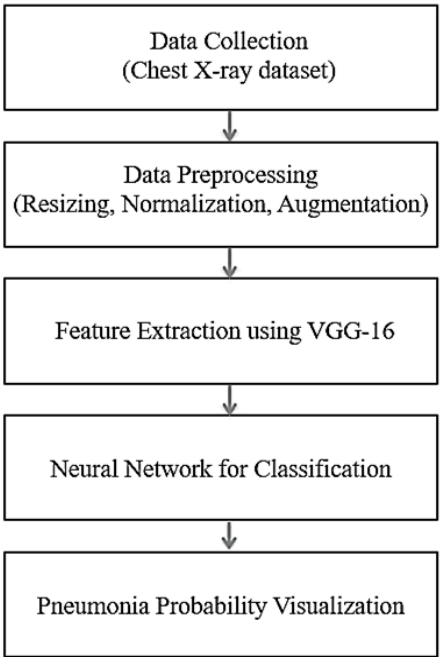


Figure 1. *Block diagram of the proposed method.*

Here, Figure 1 illustrates a schematic depiction of the approach utilized for identifying pneumonia in chest X-ray (CXR) images.

3. Result

Table 1. Results obtained from all applied models.

Applied Models	Metric value (%)			
	Accuracy	Precision	Recall	F1-Score
ResNet50	90.06	86.77	99.23	92.58
DenseNet169	92.15	97.49	89.74	93.46
MobileNet	91.99	92.71	94.62	93.65
InceptionV3	90.71	87.56	99.23	93.03
Xception	91.99	89.35	98.97	93.92
Proposed	92.79	94.12	94.36	94.24

Table 1 compares six deep learning models: ResNet50, DenseNet169, MobileNet, InceptionV3, Xception, and the custom VGG-16-based framework. It reveals distinct performance characteristics, providing insights into their effectiveness for pneumonia classification from CXR images. Among these, the proposed VGG-16 approach stood out with the highest classification accuracy of 92.79%. This demonstrates its ability to effectively tell apart pneumonia cases from healthy ones. Additionally, it showed strong consistency across all metrics, achieving 94.12% precision, 94.36% recall, and an F1-score of 94.24%. The balance between precision and recall is critical in clinical settings, where missing a diagnosis or misdiagnosing can have serious consequences. DenseNet169 followed closely in accuracy with 92.15% and excelled in precision at 97.49%, reducing the number of false alarms. However, its slightly lower recall of 89.74% suggests it might miss some true pneumonia cases, which could be an issue in practice. MobileNet and Xception had F1-scores of 93.65% and 93.92%, respectively, confirming their reliable performance. MobileNet's efficiency makes it ideal for mobile or low-resource environments, but its slightly lower recall and precision indicate a minor trade-off in predictive ability. In contrast, InceptionV3 and ResNet50 had lower accuracies of 90.71% and 90.06%, yet they excelled with very high recall values of 99.23%. This highlights their ability to capture nearly all true pneumonia cases. They are particularly suitable for scenarios where missing a positive diagnosis is unacceptable, even with a higher rate of false positives. Overall, the proposed VGG-16 framework achieves a well-balanced optimization between sensitivity and specificity, offering a reliable solution for real-world applications. Its high-performance results from targeted model adjustments, careful fine-tuning, and the inclusion of explanation tools like Grad-CAM, which together enhance both accuracy and transparency. The findings show that, when thoughtfully adapted, transfer learning techniques can outperform even deeper or more complex networks in domain-specific medical imaging. This emphasizes the importance of designing diagnostic models that are not only accurate but also interpretable, efficient, and ready for integration into modern clinical and telehealth systems.

The data represented in Figure 2 indicates that the deep learning model proposed in this study (VGG-16) has outperformed the six other models in accuracy, precision, recall, and F1-score metrics of evaluation. In the boxplots, VGG-16 presents an overall balance in the metrics being evaluated and the highest accuracy (92.79%) and almost equal precision (94.12%). DenseNet169 presents the highest bar for precision (97.49%), but its recall status is shorter (89.74%), which can allow pneumonia cases to be missed. InceptionV3, Xception, and ResNet50 all have taller recall bars, but their lower precision bars mean the model generally produces false positives. The boxplots for MobileNet show that the values are a bit more averaged and moderately high, which means that it should be considered as a practical solution if resources are limited. In all, VGG-16 is the most stable and rounded model as a whole, which presents as a very good model for real clinical practice.

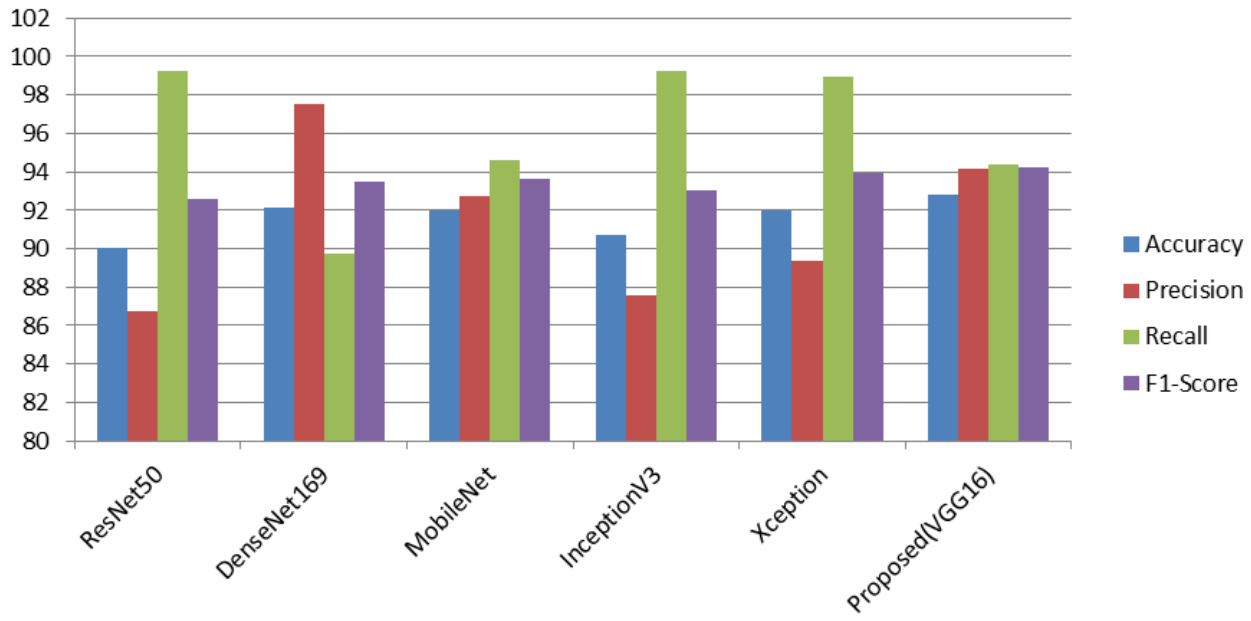


Figure 2. *Comparative analysis of six applied models.*

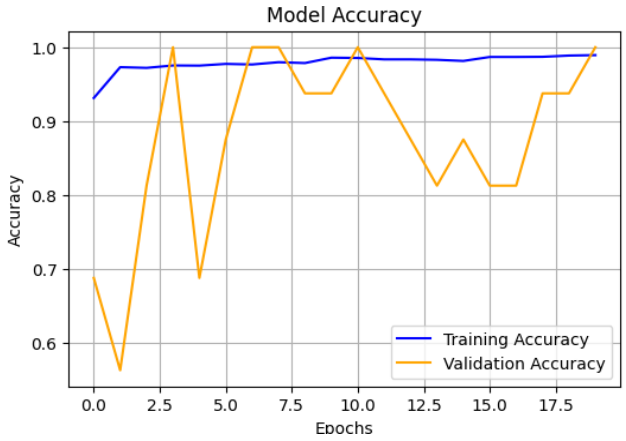


Figure 3. *Accuracy chart of ResNet50.*

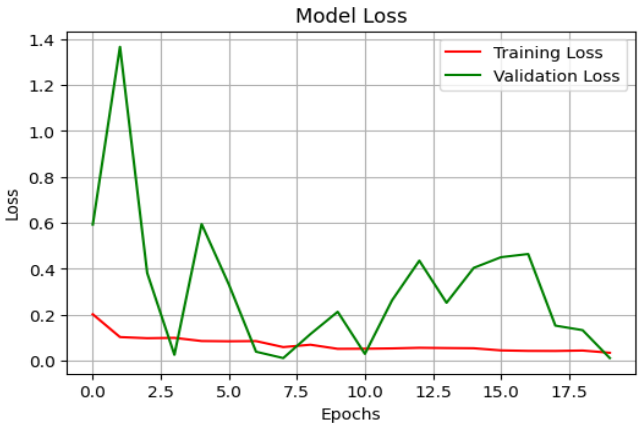


Figure 4. *Loss chart of ResNet50.*

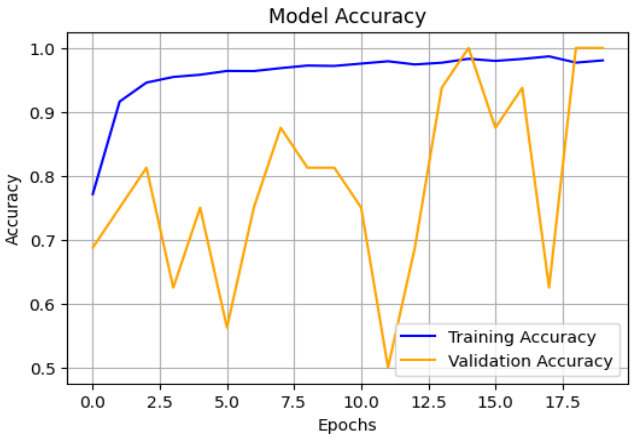


Figure 5. *Accuracy chart of DenseNet169.*

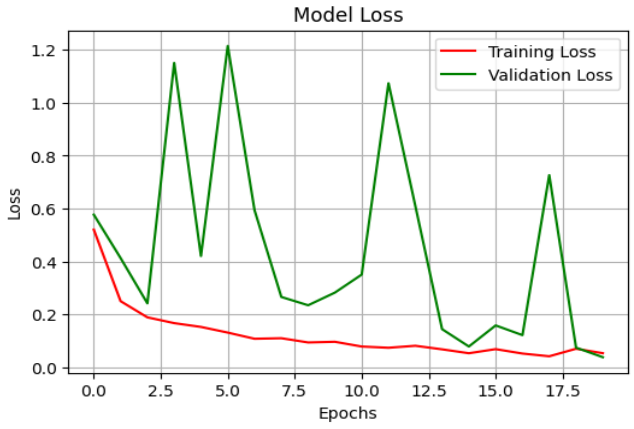


Figure 6. *Loss chart of DenseNet169.*

The training performance of the pneumonia detection model using CXR images and ResNet50 architecture is illus-

trated in Figures 3 and 4. The training accuracy almost achieved 100% accuracy, which is a good sign of learning. The validation accuracy varied quite a bit, from 60 to 100%, which indicates inconsistent performance on unseen data. The training loss decreased consistently and approached zero whereas the validation loss varied significantly. These observations indicate that while the model learns when exposed to the training data, it does not appear to learn consistently when being applied to new cases, which poses concerns of performing consistently in real-world settings in clinical applications.

As depicted in Figures 5 and 6, DenseNet169 reaches a relatively high training accuracy fairly quickly and shows a steady decline in training loss, indicating it learns well from training data. However, the validation accuracy is far more volatile– rising and falling between 60% and 100% with the validation loss often spiking above 1.2, indicating it performs inconsistently on new, unseen data. Although DenseNet169 has the highest accuracy overall, the variability in validation metrics raises the question of how sensitive the model is to variance, and its risk of neglecting pneumonia cases, leading to false negative cases in a clinical setting.

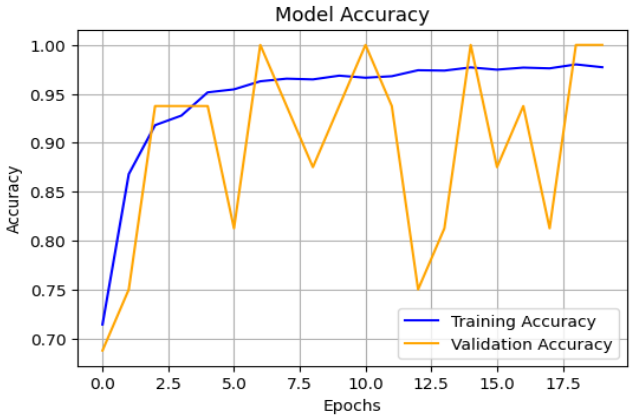


Figure 7. *Accuracy chart of MobileNet.*

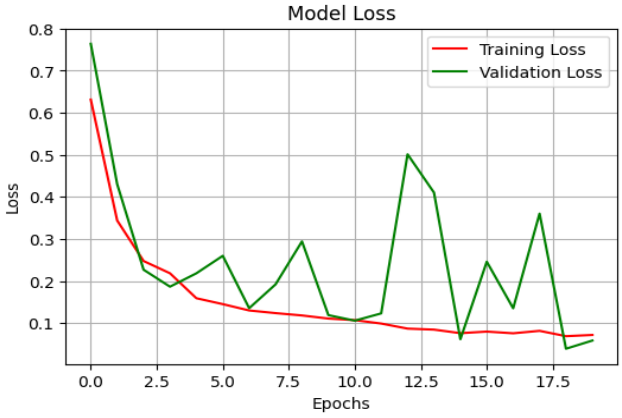


Figure 8. *Loss chart of MobileNet.*

The training behavior of the efficient, lightweight deep learning architecture, MobileNet, is shown in Figures 7 and 8. This model exhibits a consistent increase in training accuracy, ultimately achieving 100% during the last epochs, suggesting the model is able to learn the data. Validation accuracy remains relatively stable between 80% and 100%, and is more consistent when compared to ResNet50 and DenseNet169. Training loss decreases steadily to nearly zero, while validation loss is held low and consistent, pointing to good generalization to new/unseen data. While MobileNet presents an excellent trade-off between accuracy and efficiency, which is desirable for real-time or resource-constrained clinical environments, the research will adopt the VGG-16 model as the proposed model due to its higher feature extraction capabilities and consistent performance across all evaluation metrics.

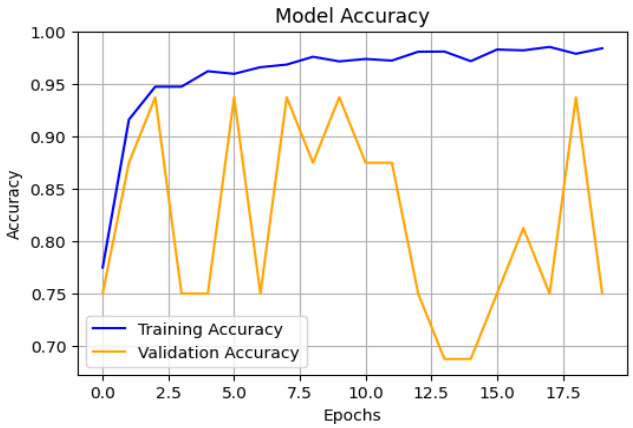


Figure 9. *Accuracy chart of Inceptionv3.*

Figure 10. *Loss chart of Inceptionv3.*

The performance evaluation of the InceptionV3 model is illustrated in Figures 9 and 10. Figure 9 shows the training accuracy climbing rapidly in the early epochs and ultimately reaching a steady state close to 1.00; thus, the model has capitalized on the training set. The validation accuracy, while achieving high values, exhibited fluctuations and did not

demonstrate a steadily improving trend. This indicates that the model's performance on unseen data is inconsistent. In Figure 10, the training loss exhibited a steady fall and declined toward zero, indicating a strong fit for the training set. On the other hand, the validation loss exhibited a decrease before becoming erratic with a slight increase at the end epochs. The increasing gap between training and validation performance indicates that although the model fits well on the training data, it may include limited predictive stability with new inputs, which is problematic when considering clinical applicability.

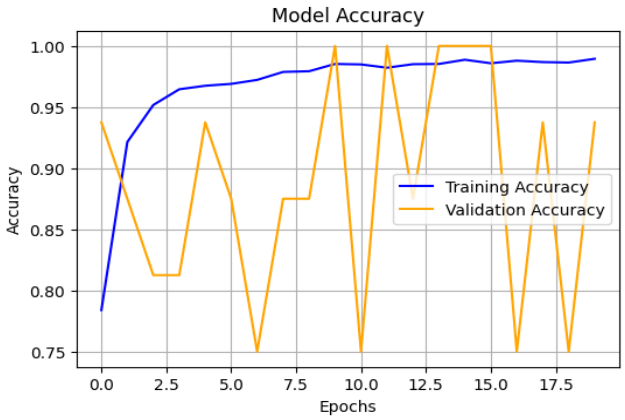


Figure 11. *Accuracy chart of Xception.*

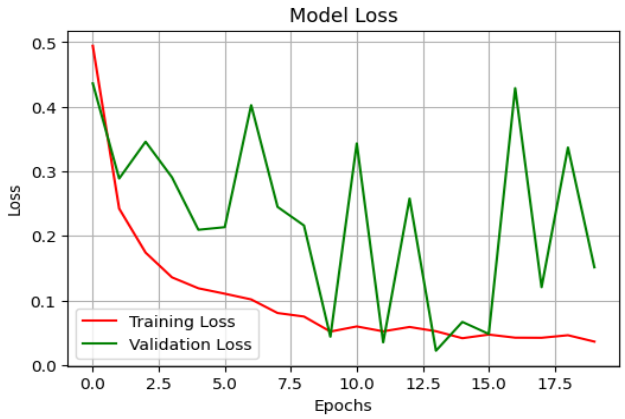


Figure 12. *Loss chart of Xception.*

Figures 11 and 12 illustrate the performance metrics of the Xception model. As shown in Figure 11, the training accuracy increases rapidly and stabilizes just below 1.00, indicating that the model has effectively captured patterns within the training dataset. Validation accuracy also improves significantly, eventually aligning closely with the training accuracy. This close alignment suggests that the Xception model generalizes well to unseen data, exhibiting better performance stability compared to the InceptionV3 model. In Figure 12, the training loss decreases sharply and levels off near zero, confirming that the model has learned efficiently from the training data. Simultaneously, the validation loss shows a steady decline and converges toward the training loss value. The consistency in the loss curves and the minimal gap between them highlight the model's robustness and its ability to generalize effectively across both training and validation sets.

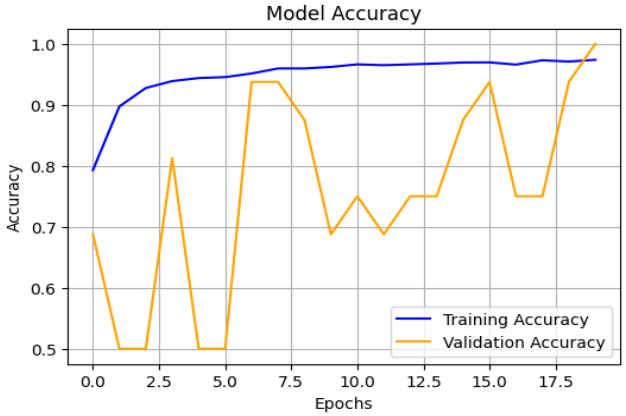


Figure 13. *Accuracy chart of proposed VGG16.*

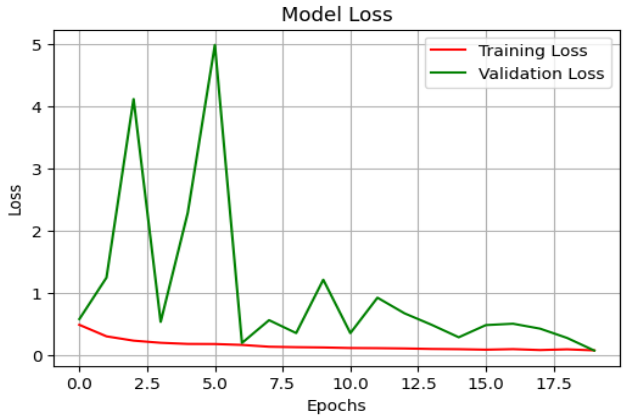


Figure 14. *Loss chart of proposed VGG16.*

The training process for the suggested VGG-16 model in terms of pneumonia classification is presented in Figures 13 and 14 across epoch number. Starting with the accuracy plot in Figure 13, training accuracy (shown in blue) and validation accuracy (shown in yellow) both see a rapid increase in the first few epochs. After the preliminary variability, validation accuracy remains consistent between 92% and 94%, demonstrating the circuit's capacity to generalize well on unseen data. The associated loss plot in Figure 14 demonstrates a consistent drop in both training (shown in blue) and validation (shown in green) loss, indicating a decreasing trend in prediction error. Convergence of high, stable validation accuracy, with low, stable validation loss demonstrates that the model has successfully learned meaningful features from CXRs and exhibited solid overall performance with no large concerns regarding model instability and or poor generalization.

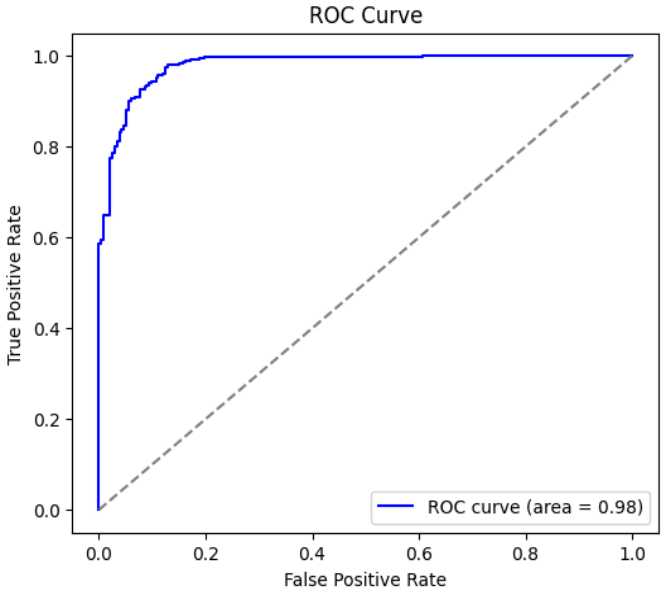


Figure 15. *ROC curve of the proposed model.*

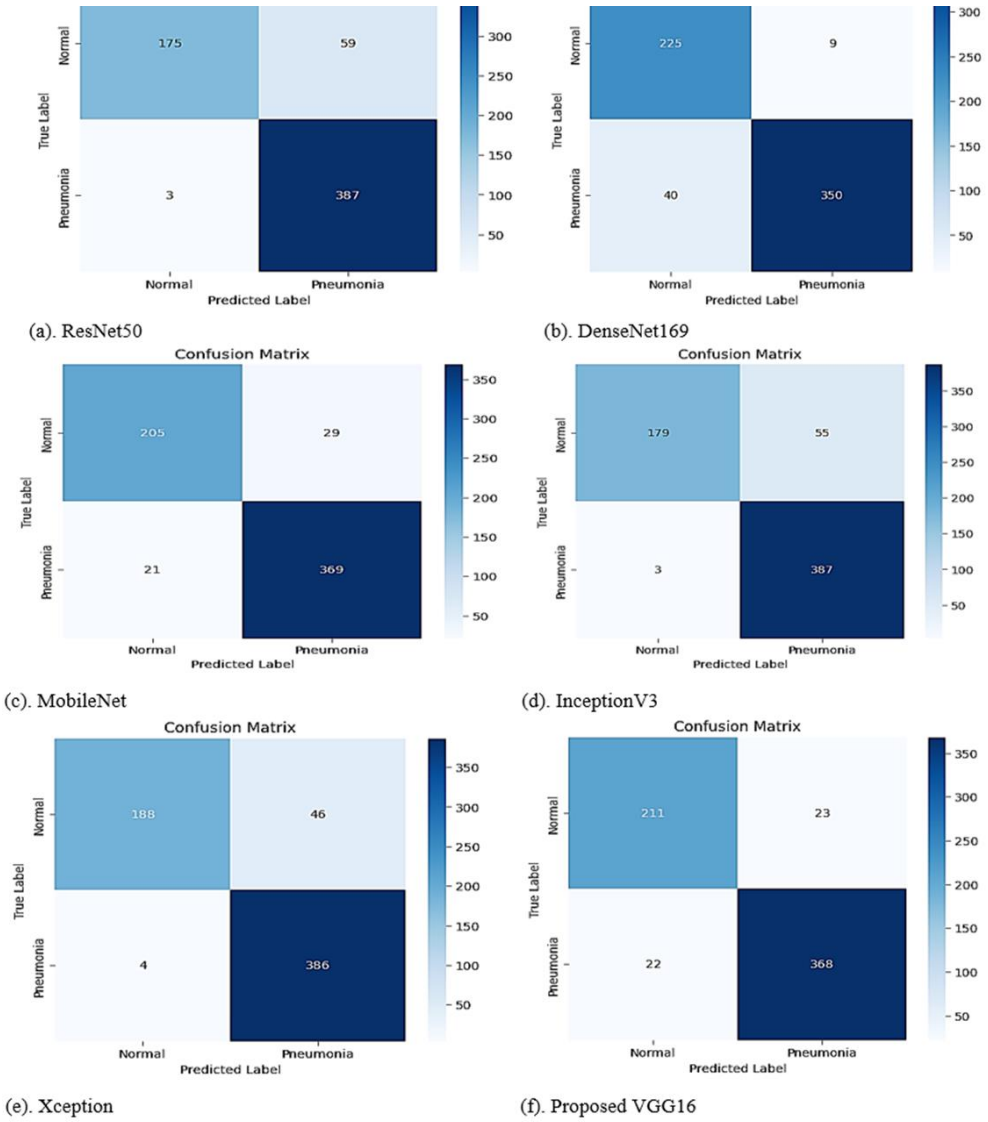


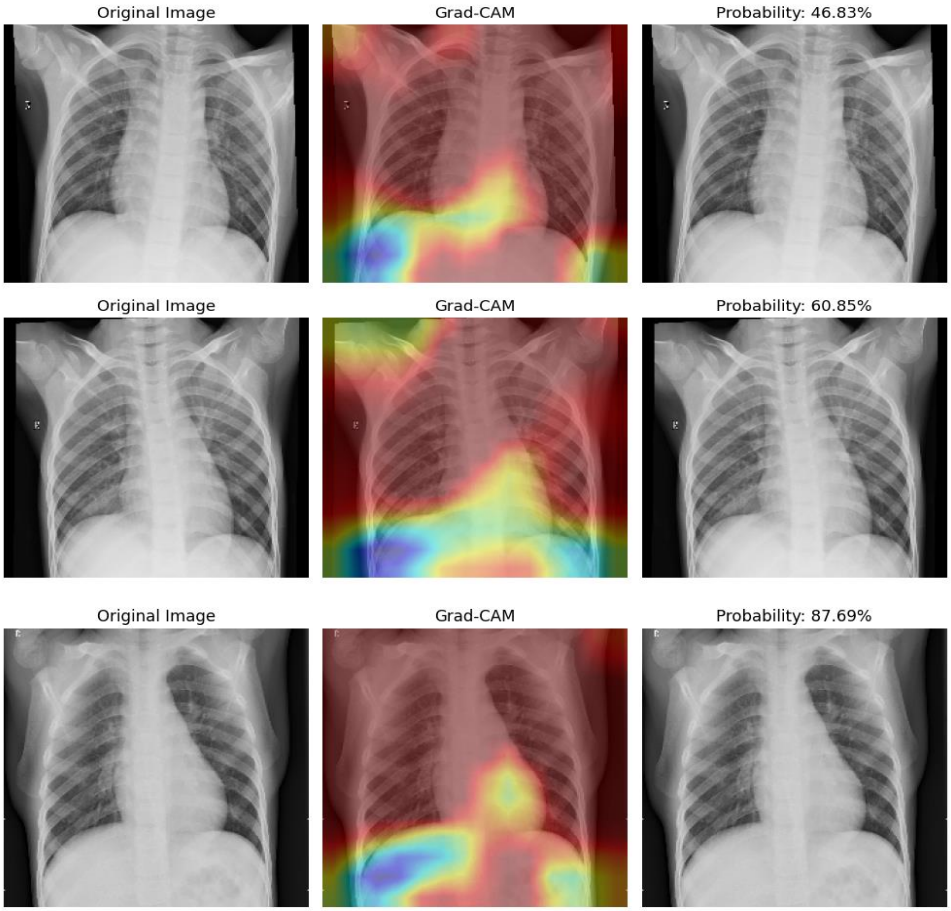
Figure 16. *Confusion Matrix of all applied models.*

Figure 15 displays the ROC curve used to assess the performance of the pneumonia detection model, in which True Positive Rate is plotted against False Positive Rate. The ROC curve illustrates excellent classification performance. In all three scenarios, AUC was reported above 0.98, indicating a high probability that the model can accurately detect pneumonia or predict normal X-ray images. AUC metrics approach 1.0 shows the ability to accurately predict pneumonia so is seen as an excellent level of diagnostic accuracy and confidence in this method for medical images, in this case, automated pneumonia diagnosis. There is great potential to improve diagnostic accuracy and eliminate false positives and therefore demonstrate this model's fitness for medical image analysis.

Figure 16 provides a comparative visualization of six convolutional neural network (CNN) architectures evaluated for pneumonia detection using CXR images. Each subfigure (a–f) displays a confusion matrix for a different model: ResNet50, DenseNet169, MobileNet, InceptionV3, Xception, and the proposed VGG-16 model employed in this study. The confusion matrices offer a detailed breakdown of classification performance, where the rows represent the true class labels (Normal or Pneumonia) and the columns indicate the model's predictions. Correct classifications True Negatives (TN) and True Positives (TP) appear along the diagonal in dark blue, while misclassifications False Positives (FP) and False Negatives (FN) are shown off-diagonal.

The ResNet50 model achieved 175 TN and 387 TP while misclassifying 59 normal cases as pneumonia (FP) and 3 pneumonia cases as normal (FN). DenseNet169 showed improved precision with 225 TN and 350 TP but incurred 9 FP and 40 FN. MobileNet recorded 205 TN and 369 TP, with 29 FP and 21 FN. InceptionV3 yielded 179 TN and 387 TP, alongside 55 FP and 3 FN. The Xception model achieved 188 TN and 386 TP, with 46 FP and 4 FN.

Notably, the proposed VGG-16 model demonstrated the most balanced performance, correctly identifying 211 normal cases and 368 pneumonia cases while producing only 23 FPs and 22 FNs. This superior performance highlights the effectiveness of transfer learning with VGG-16 for pneumonia detection, as it achieved the highest number of accurate predictions and the lowest misclassification rates among all evaluated models. The results affirm the robustness of the proposed approach in distinguishing between normal and pneumonia cases, validating its potential for reliable clinical application.



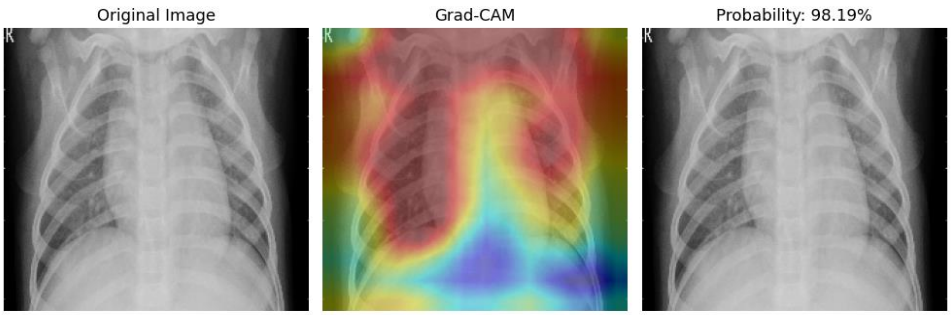


Figure 17. *Pneumonia probability visualization.*

Figure 17 depicts the prediction behavior of our proposed VGG16 model, demonstrating an ability to highlight locations of interest within chest X-ray (CXR) images. The figure provides 4 example cases consisting of the original X-ray image, the corresponding Grad-CAM visualization, and the predicted probability of pneumonia given by the model. The model predicted a probability of pneumonia at 46.83% in the first case, with the Grad-CAM visualization highlighting somewhat diffuse areas in the lower lung fields. The model's predicted probability of pneumonia increased to 60.85% in the second case, and the Grad-CAM visualization displayed a more concentrated period of activation in the right lung. In the third case, the model further increased the predicted probability to 87.69%, and the Grad-CAM displayed a more intense and focused area of activation in the lower lung. Finally, in the fourth case, the model predicted a very high probability of pneumonia (98.19%), and the Grad-CAM visually displayed an area with concentrated activation that appeared to correlate with a probable area of consolidation. This progression indicates that the model increases confidence in the prediction when visual cues of pneumonia become increasingly more pronounced, and importantly, the Grad-CAM seems to accurately depict and highlight places in the CXR images.

4. Discussion

Pneumonia is a serious global health problem, causing many hospitalizations and deaths, especially in vulnerable groups. Early and accurate diagnosis is key to timely treatment and preventing complications [11]. However, manual interpretation can be inconsistent due to differences in image quality and radiologist experience. Deep learning, particularly, offers a promising way to improve speed, accuracy, and reliability in pneumonia detection. [12].

4.1. Early Machine Learning Approaches for Pneumonia Detection

Previous methods for pneumonia detection mainly depended on conventional machine learning techniques, including Support Vector Machines, Random Forests, and k-Nearest Neighbors. These techniques typically depended on manually extracted features from CXR images. While they demonstrated some effectiveness, their performance was often limited by challenges in generalization, reducing their reliability across diverse datasets. However, with the rise of more powerful computational resources, these conventional methods have gradually been superseded by advanced deep learning techniques. These modern approaches have substantially enhanced both the accuracy and robustness of pneumonia detection, enabling more reliable results across various data sources.

4.2. Deep Convolutional Neural Networks Applied to Automated Pneumonia Detection

In recent times, convolutional neural networks (CNNs) have attracted significant interest due to their remarkable effectiveness in medical image analysis, especially in identifying pneumonia. Numerous research efforts have employed CNN-based models to automate the categorization of CXR images for diagnosing pneumonia. For example, Jain et al. (2022) presented a deep learning model based on CNNs that obtained hierarchical features from CXR scans, achieving significant success in classification accuracy and computational efficiency [13]. Similarly, Akbar et al. (2024) performed an extensive assessment of different neural network architectures for pneumonia detection, emphasizing the advantages of CNN models compared to conventional machine learning techniques [14].

4.3. Transfer Learning and VGG-16 in Pneumonia Detection

Transfer learning has become a very efficient method, particularly in areas with scarce labeled data, like medical imaging. This method utilizes models that have been pretrained on extensive, varied datasets such as ImageNet and refines them for specific purposes, including pneumonia identification. Among different architectures, VGG-16 has demonstrated notable effectiveness because of its ability to capture intricate features from CXR images, even with a limited number of labeled samples. For instance, Ahmed et al. (2023)

showcased the effectiveness of transfer learning using VGG-16 in identifying pneumonia, COVID-19, and tuberculosis from CXR images. Their model attained remarkable precision while greatly minimizing both computational expenses and training duration [15].

4.4. Multiclass and Multimodal Approaches

While pneumonia detection is typically framed as a binary classification task (pneumonia vs. healthy), several studies have explored multiclass classification, where models are trained to simultaneously identify multiple diseases. Mohan et al. (2024) proposed a multiclass deep learning model that can distinguish between healthy lungs, COVID-19, and pneumonia in CXR images, demonstrating the model's ability to differentiate between various pulmonary conditions [16]. Furthermore, models such as PulmoNet, introduced by Abdulahi et al. (2024), combine multimodal data to enhance diagnostic accuracy, emphasizing the value of integrating diverse data sources for more comprehensive medical image analysis [17].

4.5. Other Deep Learning Architectures and Enhancements

Various alternative CNN architectures have been explored for pneumonia detection, including MobileNet and ResNet. For example, Reshan et al. (2023) utilized MobileNet for pneumonia detection, achieving impressive results in both efficiency and accuracy [18]. Similarly, Li et al. (2024) introduced the DSEception model, a novel architecture designed to improve the diagnosis of pneumonia and tuberculosis by incorporating advanced feature extraction techniques [19]. The integration of Grad-CAM has further enhanced model transparency, enabling clinicians to visualize the regions of a CXR image that contribute to the model's decision, thus boosting both explainability and trust in the predictions [20].

Additionally, Mehmood et al. (2024) developed CP DeepNet, an automated system that excels in detecting both COVID-19 and pneumonia from lung X-rays, achieving high accuracy across different types of lung diseases [21]. Likewise, Podder et al. (2023) proposed LDDNet, a deep learning framework for diagnosing infectious lung diseases, which showed promising results in detecting pneumonia and other pulmonary conditions from CXR images [22]. Solanki (2024) also contributed to the advancement of pneumonia detection by applying deep learning techniques to CXR images, further enhancing the potential of automated pneumonia diagnosis [23].

4.6. Addressing Challenges and Future Directions

Despite the significant progress made in deep learning-based pneumonia detection, several challenges remain. Issues such as data imbalance, model generalization, and clinical validation need to be addressed to ensure these models perform consistently in real-world applications. Although many studies report high accuracy, the successful deployment of these models in clinical settings requires thorough validation across diverse datasets. Moving forward, it will be crucial to focus on enhancing model transparency, interpretability, and generalization to ensure that deep learning models can be reliably and effectively integrated into clinical practice.

5. Conclusion

In this research, we present a solidified deep learning framework to automate pneumonia detection from CXR images, with publicly available CXR STONE faculty images collected from Kaggle through the use of deep learning methods. This study employs transfer learning through a VGG-16 pre-trained architecture and adds a deep neural network (DNN) classifier, both of which effectively extract features for classification. Our results provide diagnostic accuracy (92.79% accuracy, 94.12% precision, 94.36% recall, 94.24% F1-score, and AUC of 0.98) and excellent potential for the medical field with its generalizability across many cases. Our model's computational efficiency and visual interpretability tools, including probabilistic heatmaps, contribute to the appeal of this model in high-volume, real-time diagnostic settings, especially in a low-resource or remote location where expert interpretation of CXRs would not be available. This project attempts to widen the explosion of AI-assisted CXR pneumonia detection diagnostic accuracy and support clinical decision-making. Future directions for this framework may entail the addition of multi-disease classification and tracking disease progression longitudinally, in addition to piloting the model in existing clinical workflows. Interpretability of the model for physician use and using electronic health records (EHR) to supplement this will be critical considerations towards a sustainable and acceptable model for implementation into real-world health systems and practice.

Abbreviations

CXRs	Chest X-Ray Images
Grad-CAM	Gradient Weighted Class Activation Mapping
WHO	World Health Organization
DNN	Deep Neural Network
CNN	Convolutional Neural Network
EHR	Electronic Health Records

Author Contributions

Sourav Sana: Conceptualization, Data curation, Methodology, Validation, Visualization, Writing – original draft

Priyankar Biswas: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing

A. T. M. Saiful Islam: Formal Analysis, Investigation, Visualization, Writing – review & editing

Conflicts of Interest

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

References

[1] "Pneumonia in children." Accessed: Apr. 24, 2025. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia

[2] A. Mabrouk, R. P. D. Redondo, A. Dahou, M. A. Elaziz, and M. Kayed, "Pneumonia Detection on Chest X-ray Images Using Ensemble of Deep Convolutional Neural Networks," *Applied Sciences (Switzerland)*, vol. 12, no. 13, Jul. 2022, https://doi.org/10.3390/app12136448

[3] X. Wei, H. Bai, X. Zhang, and Y. Li, "Deep Pneumonia: Attention-Based Contrastive Learning for Class-Imbalanced Pneumonia Lesion Recognition in Chest X-rays," Jul. 2022. Available: http://arxiv.org/abs/2207.11393

[4] M. Neshat, M. Ahmed, H. Askari, M. Thilakaratne, and S. Mirjalili, "Hybrid Inception Architecture with Residual Connection: Fine-tuned Inception-ResNet Deep Learning Model for Lung Inflammation Diagnosis from Chest Radiographs."

[5] N. Darapaneni *et al.*, "Pneumonia Detection in Chest X-Rays using Neural Networks."

[6] A. Sharma, "Deep Learning Model for Pneumonia Diagnosis from Pediatric Chest X-Ray Images," pp. 891–900, 2023, https://doi.org/10.1007/978-981-99-3485-0_70

[7] A. K. Dash, P. Mohapatra, and N. K. Ray, "Pneumonia Detection in Children from Chest X-Ray Images by Executing Network Surgery of Deep Neural Networks," *SN Comput Sci*, vol. 5, no. 2, pp. 1–13, Feb. 2024, https://doi.org/10.1007/S42979-023-02512-7/METRICS

[8] J. Nayak and D. Sahu, "Deep Learning Based Pneumonia Infection Classification in Chest X-ray Images Using Convolutional Neural Network Model," *Lecture Notes in Electrical Engineering*, vol. 903, pp. 273–283, 2022, https://doi.org/10.1007/978-981-19-2281-7_26

[9] MD. R. Sarkar, S. M. S. Ahamed, M. M. Bari, M. M. Hasan, and F. H. Shishir, "Deep Learning Methods for Chest X-Ray Imaging-Based COVID-19 Pneumonia Detection," *Journal of Image Processing and Intelligent Remote Sensing*, no. 45, pp. 55–62, Sep. 2024, https://doi.org/10.55529/jipirs.45.55.62

[10] "Chest X-Ray Images (Pneumonia)." Accessed: Apr. 24, 2025. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

[11] S. Singh *et al.*, "Deep Attention Network for Pneumonia Detection Using Chest X-Ray Images," *Computers, Materials and Continua*, vol. 74, no. 1, pp. 1673–1691, 2023, https://doi.org/10.32604/cmc.2023.032364

[12] D. Li, "Attention-enhanced architecture for improved pneumonia detection in chest X-ray images," *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024, https://doi.org/10.1186/s12880-023-01177-1

[13] D. K. Jain *et al.*, "Deep Learning-Aided Automated Pneumonia Detection and Classification Using CXR Scans," *Comput Intell Neurosci*, vol. 2022, 2022, https://doi.org/10.1155/2022/7474304

[14] W. Akbar *et al.*, "Pneumonia Detection: A Comprehensive Study of Diverse Neural Network Architectures using Chest X-Rays," *International Journal of Applied Mathematics and Computer Science*, vol. 34, no. 4, pp. 679–699, Sep. 2024, https://doi.org/10.61822/amcs-2024-0045

[15] M. S. Ahmed *et al.*, "Joint Diagnosis of Pneumonia, COVID-19, and Tuberculosis from Chest X-ray Images: A Deep Learning Approach," *Diagnostics*, vol. 13, no. 15, Aug. 2023, https://doi.org/10.3390/diagnostics13152562

[16] G. Mohan, M. M. Subashini, S. Balan, and S. Singh, "A multiclass deep learning algorithm for healthy lung, Covid-19 and pneumonia disease detection from chest X-ray images," *Discover Artificial Intelligence*, vol. 4, no. 1, Dec. 2024, https://doi.org/10.1007/s44163-024-00110-x

[17] A. R. T. Abdulahi, R. O. Ogundokun, A. R. Adenike, M. A. Shah, and Y. K. Ahmed, "PulmoNet: a novel deep learning based pulmonary diseases detection model," *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024, https://doi.org/10.1186/s12880-024-01227-2

[18] M. S. Al Reshan *et al.*, "Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model," *Healthcare (Switzerland)*, vol. 11, no. 11, Jun. 2023, https://doi.org/10.3390/healthcare11111561

[19] S. Li, Y. Hu, L. Yang, B. Lv, X. Kong, and G. Qiang, "DSEception: a noval neural networks architecture for enhancing pneumonia and tuberculosis diagnosis," *Front Bioeng Biotechnol*, vol. 12, Sep. 2024, https://doi.org/10.3389/fbioe.2024.1454652

[20] J. Du and J. Yang, "Classification of Chest X-ray Images (Pneumonia) Based on ResNet and Grad-CAM," *ACM International Conference Proceeding Series*, pp. 156–164, Jul. 2023, https://doi.org/10.1145/3614008.3614031;WGROUP:STRING:ACM

[21] M. H. Mehmood *et al.*, "CP_DeepNet: a novel automated system for COVID-19 and pneumonia detection through lung X-rays," *Multimed Tools Appl*, Dec. 2024, https://doi.org/10.1007/s11042-024-18921-6

[22] P. Podder *et al.*, "LDDNet: A Deep Learning Framework for the Diagnosis of Infectious Lung Diseases," *Sensors*, vol. 23, no. 1, Jan. 2023, https://doi.org/10.3390/s23010480

[23] J. Solanki and V. Agrawal, "International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING A Novel Approach for Pneumonia Detection from Chest X-Ray Images using Deep Learning." [Online]. Available: www.ijisae.org