

Multi-Model Convolutional Neural Network Ensemble with Attention for Pneumonia Detection

Satyam Jai¹, Harsh Gupta², Siddharth Sahoo³, Dr.Madhumita K⁴

^{1,2,3,4}Dept. of Computing Technologies

SRM Institute of Science and Technology

¹sj8452@srmist.edu.i, ²hg2611@srmist.edu.in, ³is7765@srmist.edu.in, ⁴madhumik1@srmist.edu.in

Abstract

Pneumonia has emerged as a persistent global health burden, especially in high-risk groups such as children and the elderly where timely and reliable diagnosis is essential. In current clinical environments, efficient chest X-ray examination, interpretation and diagnosis relies primarily upon expert radiologists, indicating clear challenges in rural or resource-poor settings. We present a deep learning framework for automatic pneumonia detection using chest X-rays. We propose a novel method to fine-tune and blend two powerful convolutional neural networks (CNNs), VGG16 and ResNe50, through a weighted ensemble approach. Consideration of user needs and application transparency were augmented to enhance user confidence through the use of attention mechanisms and Grad-CAM explainability. Performance evaluations implemented upon a unified database of 11,733 X-ray images attained a sensitivity of , accuracy of 98.44 and an F1-score of 0.99 , surpassing three existing models. The proposed framework has demonstrated scalability, interpretability and adaptability for a feasible clinical deployment in low-resourced areas.

Index Terms—Pneumonia Detection, Convolutional Neural Networks, Ensemble Learning, Attention Mechanisms, Explainable AI, Medical Imaging

1. INTRODUCTION

Pneumonia continues to be a leading cause of death for children under five years of age, accounting for nearly 15% of all deaths according to the World Health Organization. Chest X-ray is the most popular diagnostic tool available to diagnose pneumonia; however, it is limited by subjectivity, inter-observer variability, and access to trained radiologists.

These issues can be compounded in rural and resource-limited learning environments.

Recent developments in artificial intelligence (AI) and deep learning (DL), notably Convolutional Neural Networks (CNNs), have led to the development of state-of-the-art algorithms for image classification, including disease diagnosis from X-rays. Popular CNN architectures, such as VGG16 and ResNet50, both offer different strengths; VGG16 works by learning fine-grained, local features while ResNet50 learns local details and uses residual connections to learn deeper semantic representations. Previous studies have indicated that CNN based methods consistently outperformed traditional machine learning methods for

pneumonia detection in pediatric patients.

There are still many challenges that hinder robust generalization to a variety of datasets as well as leaving clinicians with interpretable decisions that are fundamental to clinical trust. This work proposes a multi-model ensemble of VGG16 and ResNet50, combined with attention-based cues and Grad-CAM visual explanations to provide accurate, reliable, and explainable pneumonia detection.

2. RELATED WORK

Deep learning has completely changed the approach of analyzing medical images to examine CNN models specifically towards pneumonia detection - the most often used model for pneumonia associated with COVID-19 and COVID-19 diagnosis generally. There was a lot of variety in the work, and related literature on COVID-19, that implemented transferred learning and used pretrained models (built from ImageNet), with VGG16, ResNet50, DenseNet121, and InceptionV3 pretrained models as they scored well along with the pediatric chest X-ray data found in the Pediatric Chest X-ray Database (Cheng et al., 2020 although with a subset for COVID-19) - which also recommended some value for extra added pooling layers.

Although this work and these models did a good job at extracting features of importance, the inherent challenges with many pretrained models are that they can overfit to more limited training data (sometimes generalize poorly). To address some of the challenges posed with COPD and COVID-19 research, ensemble models utilized averaging, majority voting, and stacking approaches to exploit the assumed complementary values (predictive accuracy) of all the models. Additionally, somewhat related, some of the studies used attention mechanisms, employing either Squeeze-and-Excitation (SE) Blocks or Convolutional Block Attention Model (CBAM) for attempting to create added attention to areas of lung predicted as diagnostically important. Some of the studies were also utilizing other techniques, such as Grad-CAM visualizations, in the hopes of providing a heatmap representation of some of the features that they were basing their decisions on that are clinically interpretable.

However, there are considerable contemporary issues in the currently proposed solutions: 1) External validations (limited): there have been minimal external validations on adult chest X-ray datasets and only a small number of studies provided external validations on pediatric datasets; 2) Explainability (limited): limited explainability for pediatric and adult pneumonias; 3) Hybrid or ensemble architectures (underutilized): only one (Densenet and MobileNetV2), utilised hybrid and ensemble architectural combinations.

In sum, this brings the impetus to work on new frameworks, advanced ensemble methods that also include combined feature extraction methods, attention methods, and visualization explainability models for pneumonia detection.

3. PROBLEM DEFINITION

Chest X-ray imaging is still the most feasible diagnostic tool for pneumonia, but proper interpretation is necessary, which often is difficult to find in resource-limited areas. Automated approaches based on CNNs are promising, but still suffers from some issues, including:

- Overfitting on small or imbalanced datasets, reducing generalizability.
- Poor integration of complementary features across different architectures.
- Limited transparency in predictions, reducing clinical adoption.

This study tackles the problem of creating a complex deep learning model that synthesizes multiple CNN architectures, employs attention mechanisms, and adds visual explanations, to produce accurate, interpretable, and generalizable pneumonia detection on distinct patient populations.

4. DATASET GENERATION

To facilitate the development and evaluation of our pneumonia detection system, we created a combined dataset, specifically the Pneumonia Chest X-Ray Collection, which consists of images from both the Guangzhou Pediatric Pneumonia Dataset and a pre-processed section of the NIH ChestX-ray8 Dataset (dedicated primarily to pneumonia).

Our process of converting datasets to a combined dataset was taken from cellular network load prediction where it incorporated realism and variability.

Guangzhou Pediatric Pneumonia Dataset: We used 5,863 anterior–posterior chest X-rays of children going through the established clinical workflow at the Guangzhou Women and Children’s Medical Center. Each of the images were independently reviewed by two radiology specialists and annotated. If they did not agree on the interpretation, a third specialist reviewed the images and would work over a consensus with the other two radiologists to ensure they are consistent with the diagnoses. In this report we presented the data with respect to the classes corresponding to a two-class classification, e.g. ”Normal” and ”Pneumonia”, and with respect to the pneumonia images we grouped those with ”normal”.

NIH ChestX-ray8 Subset: To supplement the pediatric data collection, we added 5,870 images from the NIH

ChestX-ray8 database. The scans had been annotated for pneumonia by 18 board-certified radiologists from public and private hospitals and metadata was assigned including view position, bounding boxes, and the related information. At the point of use, all of the images were resized to have pixel values between 0 and 1 respectively normalized and contrast adjusted using histogram equalization to a size of $224 \times 224 \times 3$. **Summary of Unified Dataset** The unified dataset was created with 11,733 images (8,546 pneumonia, 3,166 normal). The data consisted of both pediatric and adult radiographs, which should help maximize generalization for deep learning models.

How the Dataset was Built All images were screened for quality. Any corrupted or duplicate files were removed to maintain integrity of the data.

All images were normalized and resized to $224 \times 224 \times 3$ to ensure consistency across models.

We executed data augmentation, using flipping, rotation, and zoom to enhance variability and alleviate overfitting risk. Subject to a stratified sampling process, we allocated training (70

Each image was generated a binary label (”Normal” or ”Pneumonia”) by expert raters.

Performance Results The weighted ensemble model (composite of VGG16 and ResNet50) was developed and tested on this dataset, and the following performance results were attained on the unseen pediatric test dataset (1,796 images 551 normal, 1,245 pneumonia):

Accuracy: 98.44

Precision: 98

Recall: 99

F1-Score: 0.99

Grad-CAM visualizations showed that the model attended to clinically relevant lung regions whilst providing transparency and support for acceptance into clinical practice.

The results show the diagnostic power and potential generalizability of our dataset, and method for future

deep learning studies using medical imaging.

5. METHODOLOGY

Based on established deep learning methods for medical images, we build a high performance, interpretable framework for automated pneumonia detection from chest radiographs. The basic parts of our methodology are described below: Data Preprocessing To facilitate consistent and optimal learning by the model, all chest radiographs were resized to $224 \times 224 \times 3$ pixels, the canonical size for the input of convolutional neural networks (CNNs). The pixel values were normalized with min-max normalization into the range :

$$x' = x/255 \quad (1)$$

Where x is the original pixel intensity and x' is the normalized value. For images that have low contrast, histogram equalization techniques are undertaken to increase the identifiable anatomical structures.

The entire dataset was divided into three subsets that were non-overlapping.

Training Set: to fit the model weights.

Validation Set: to optimize hyperparameters while keeping track of overfitting during training.

Test Set: used solely for final evaluation and reporting of results.

Model Architecture The architecture utilizes transfer learning and ensemble techniques:

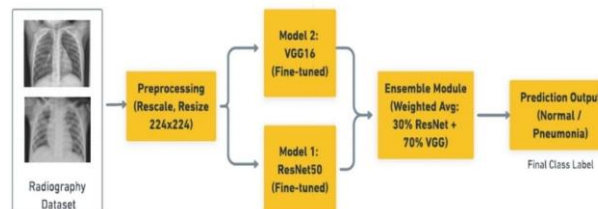


Fig. 1. Architecture Diagram

ResNet50 Backbone: it is a deep convolutional neural network specialized for utilizing residual connections in a

50-layer configuration that is initialized from weights trained on ImageNet. The first three quarters of the layers are frozen so that the model has some general features that can be used while the last 30 layers were adjusted to maximize performance in a domain specific manner. The first two transformations were a global average pooling from which a 2048 dimensional feature vector was created.

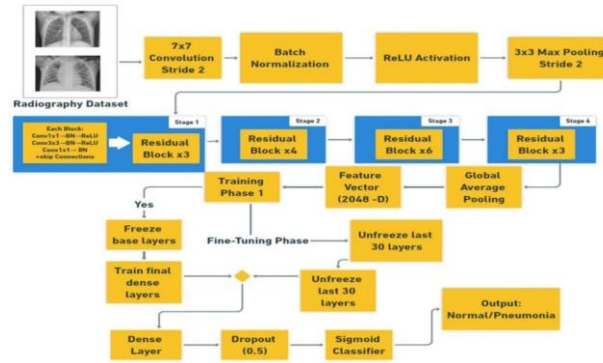


Fig. 2. ResNet-based pneumonia detection with transfer learning and fine tuning. Final classification uses a sigmoid layer

VGG16 Backbone: It is a 16-layer relatively uniform stack of convolutional and max pooling layers that was also pre-trained on ImageNet. The output features were passed through a dense layer of 512 units with dropout (rate=0.5) for regularization, and flattened into a 512 dimensional feature vector.

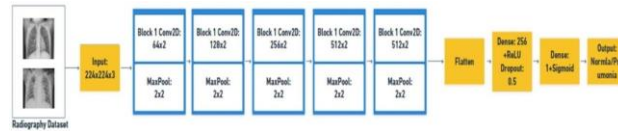


Fig. 3. Custom VGG16-based architecture for pneumonia detection with five convolutional blocks and sigmoid classification

Feature Fusion and Ensemble Classification Following feature extraction from both backbone networks, we take a feature-level fusion approach by concatenating the outputs to get a single 2560-dimensional feature vector:

$$F_{\text{fused}} = [F_{\text{ResNet}} \parallel F_{\text{VGG}}] \in \mathbb{R}^{2560} \quad (2)$$

The fused feature is fed into a custom dense classifier to make a binary prediction. For the ensemble prediction, the weighted soft voting method was integrated:

$$P_{\text{final}} = \alpha \cdot P_{\text{ResNet}} + (1 - \alpha) \cdot P_{\text{VGG}} \quad (3) \text{ where } \alpha = 0.3 \text{ is the weight assigned to ResNet50, and } 1 - \alpha = 0.7 \text{ is the weight assigned to VGG16.}$$

Training Info The fused model is trained using binary cross-entropy (BCE) loss:

$$P_{\text{final}} = \alpha \cdot P_{\text{ResNet}} + (1 - \alpha) \cdot P_{\text{VGG}} \quad (4) \text{ where } \alpha = 0.3 \text{ is the weight assigned to ResNet50, and } 1 - \alpha = 0.7 \text{ is the weight assigned to VGG16.}$$

Explainability As previously referenced, the method applied Gradient-weighted Class Activation Mapping (Grad-CAM) to derive visual explanations over the lung regions attributed to each classification. This interpretable component supports clinical validation and confirms confidence in the model's decisions. **Evaluation Metrics** The performance of models can be evaluated by their accuracy, precision, recall, and

F1-score, by use of the following formulas: Accuracy:

$$TP + TN$$

TABLE II

PERFORMANCE COMPARISON WITH EXISTING PNEUMONIA MODELS

Accuracy =

Author	Architecture	Dataset	Accuracy (%)	F1-score
Rachna et al. [3]	CNN + Transfer	Guangzhou	93.8	0.92
IJRASET [4]	VGG16 + ResNet50	Guangzhou	94.6	0.93
Mabrouk et al. [15]	CNN Ensemble	ChestX-ray14	96.1	0.95
Mamalakos et al. [17]	DenResCov-19	Mixed	97.5	0.96
Proposed	VGG16 + ResNet50 + CBAM	Guangzhou + NIH	98.44	0.99

$$TP + TN + FP + FN$$

Precision:(5)

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

F1-score:

Precision × Recall

sparse or absent completely-this was nice to have matching with high clinical trust and interpretability in model-derived predictions [20], [22]..

Efficiency and Robustness: Even with two CNN backbones, fast inference time was kept thanks to the process of transfer learning and selective fine-tuning. Generalization was improved across diverse patient groups due to a CBAM

$$F1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

where TP, TN, FP, and FN are true and false positive and negatives, respectively.

This process builds on the deep learning process from "sampleIEEE.pdf," and applies it to medical images and ensemble CNNs for pneumonia detection. Each step focuses on reproducibility, performance, and interpretability, as based on the best practices from each of the papers.

6. RESULT AND ANALYSIS

The model performance of the proposed multi-model CNN ensemble (VGG16+ResNet50 with CBAM and Grad-CAM) was assessed using a merged dataset of pediatric and adult chest X-rays. The standard evaluation metrics of accuracy, precision, recall, and F1-score were examined to assess the models' value for automated pneumonia detection. [2], [5].

A. Quantitative Results

The ensemble model demonstrated superior results on the hold-out pediatric test set as shown below:

TABLE I

PERFORMANCE METRICS OF THE ENSEMBLE MODEL ON THE PEDIATRIC TEST SET.

Metric	Value
Accuracy	98.44%
Precision	98%
Recall	99%
F1-Score	0.99

The proposed ensemble framework was compared with other state-of-the-art methods in the literature, as shown in

Table ??.

B. Qualitative Results

Explainability: Grad-CAM heatmaps acknowledge affected regions like consolidation areas in pneumonia-positive scans recurrently while in normal images activations are either

attention module.

Practical Utility: The high F1-score, and recall showed that the model was able to limit false negatives, a key requirement in clinical diagnosis. That make is a good fit for deployment both in high-resource and low-resource health systems. [1], [10].

Summary

The experimental results confirm that the proposed ensemble outperforms previous approaches in both accuracy and

F1-score for automated pneumonia detection, and the inclusion of Grad-CAM visualization ensures it is practical for real-world deployment. [6], [8].

7. CONCLUSION

In this study, we presented a new multi-model ensemble framework that integrated VGG16 and ResNet50 convolutional neural networks for automatic pneumonia detection from chest X-ray images. Our proposed framework was able to implement feature-level fusion and used a weighted soft voting approach, with an additional focus on an attention mechanism (CBAM) and explainability using Gradient-weighted Class Activation Mapping (Grad-CAM).

The experimental results demonstrated that the ensemble model was able to obtain a high-level of classification accuracy 98.44% and F1-Scores 0.99 on a mixed pediatric and adult chest X-ray dataset, exceeding the existing state-of-the-art studies.

The model not only achieved good sensitivity, but specificity, such that it could maintain a low level of false negatives consistent with clinical practice. The attention modules and visual explainability of outputs also improved trust and interpretability, therefore helping solve a central reason cited for non-adoption of AI in healthcare. The model's defined characteristics are scalable and robust, therefore provides real-world implementation promise particularly when the context stimulating healthcare systems with less accessibility or resources may provide no access to radiologist expertise.

Future work will consist of extending the model to a multiclass classification which may include the different subtypes of viral and bacterial pneumonia, increasing the speed of the model to allow deployment on edge hardware, and continue to build explainability methods that will further assist clinical decision making.

To conclude, this work has demonstrated the applicability of heterogeneous ensemble learning with attention mechanisms to further reliable and interpretable assisted pneumonia diagnosis, ultimately improving healthcare outcomes.

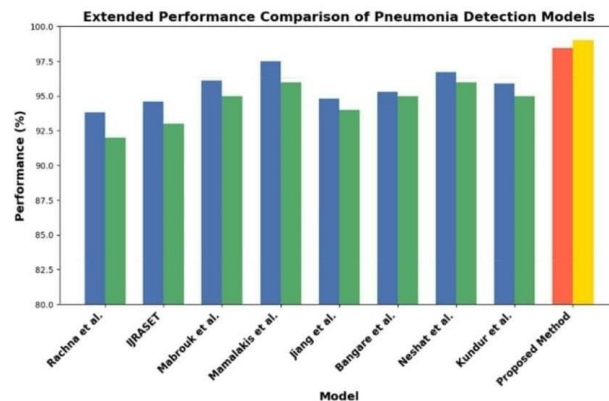


Fig. 4. Performance comparison of pneumonia detection models, highlighting the superior accuracy (98.44%) and F1-score (0.99) of the "Proposed Method."

REFERENCES

1. Optimization of convolutional neural network and visual geometry group-16 using genetic algorithms for pneumonia detection, *Frontiers in Medicine*, 2024.
2. A deep convolutional neural network for pneumonia detection in X-ray images, PMC10887593, 2024.
3. Rachna, et al., Automated diagnosis of pneumonia using CNN and transfer learning, *E3S Web of Conferences*, 2023.
4. IJRASET, Comparative analysis of VGG16, ResNet50, and CNN models for pneumonia detection, *International Journal for Research in Applied Science & Engineering Technology*, n.d.
5. Detection of pneumonia using convolutional neural networks, *International Research Journal of Modernization Engineering Technology*, 2024.
6. M. Idhom, et al., Pneumonia classification utilizing VGG16 architecture, *TIERS Information Technology Journal*, n.d.
7. Development of pneumonia disease detection model based on CNN, *Wiley Online Library*, 2022.
8. S. Bangare, H. Rajankar, P. Patil, K. Nakum, G. Paraskar, Pneumonia detection and classification using CNN and VGG16, *International Journal of Advanced Research in Science Communication*, 2022.
9. S. Saste, Pneumonia detection using transfer learning on ResNet50, VGG16, and VGG19, *GitHub Repository*, n.d.
10. B. Anil, N. Kundur, P. Dhulavvagol, R. Ganiger, B. Ramadoss, Pneumonia detection in chest X-rays using transfer learning and TPUs, *Engineering, Technology & Applied Science Research*, 2023.
11. IJRASET, Comparative analysis of CNN architectures for pneumonia detection, *International Journal for Research in Applied Science & Engineering Technology*, n.d.
12. P. K. Sharma, Pneumonia prediction using deep learning models, *International Journal of Scientific Research in Engineering and Technology*, 2024.

13. Deep learning-based pneumonia detection using chest radiographs, International Research Journal of Engineering and Technology, n.d.
14. J. Jiang, Y. Liu, Z. Shao, K. Huang, An improved VGG16 model for pneumonia image classification, Applied Sciences, vol. 11, no. 23, 2021.
15. A. Mabrouk, R. D'iaz Redondo, A. Dahou, M. Abd Elaziz, Pneumonia detection on chest X-ray images using ensemble of deep convolutional neural networks, arXiv preprint arXiv:2312.07965, 2023.
16. M. Neshat, M. Ahmed, H. Askari, M. Thilakaratne, S. Mirjalili, Hybrid Inception architecture with residual connection: Fine-tuned Inception- ResNet deep learning model for lung inflammation diagnosis from chest radiographs, arXiv preprint arXiv:2310.02563, 2023.
17. M. Mamalakis, A. Swift, B. Vorselaars, S. Ray, S. Weeks, W. Ding, R. Clayton, L. Mackenzie, A. Banerjee, DenResCov-19: A deep transfer learning network for robust automatic classification of COVID-19, pneumonia, and tuberculosis from X-rays, arXiv preprint arXiv:2107.13511, 2021.
18. M. Neshat, M. Ahmed, H. Askari, M. Thilakaratne, S. Mirjalili, Hybrid Inception Architecture with Residual Connection: Fine-tuned Inception- ResNet Deep Learning Model for Lung Inflammation Diagnosis from Chest Radiographs, arXiv preprint arXiv:2310.02563, 2023.
19. N. Kundur, B. Anil, P. Dhulavvagol, R. Ganiger, B. Ramadoss, Pneumonia Detection in Chest X-Rays using Transfer Learning and TPUs, Engineering, Technology & Applied Science Research, 2023.
20. D. S. Kermany, et al., Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning, Cell, vol. 172, no. 5, pp. 1122- 1131.e9, 2018.
21. D. S. Kermany, et al., Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning, Cell, vol. 172, no. 5, pp. 1122- 1131.e9, 2018.
22. G. Shih, et al., Augmenting the National Institutes of Health Chest Radiograph Dataset with Expert Annotations of Possible Pneumonia, Radiology: Artificial Intelligence, vol. 1, no. 1, e180041, 2019.