PneumoCatcher: Automated Pneumonia Detection from Chest X-Rays Using Deep Learning

Saifullah Shoaib
Department of Computer Science
Missouri University of Science & Technology
Rolla, MO, 65401, USA
sbsr4c@umsystem.edu

***Abstract*—Pneumonia is a significant global health concern, especially for vulnerable populations like children and the elderly. This study investigates the potential of deep learning for rapid and reliable pneumonia detection from chest X-ray images. A custom convolutional neural network (CNN) model was designed and trained to classify images as either pneumonia or healthy/normal. Additionally, another two models based on the ResNet-50 architecture were fine-tuned. Using a publicly available dataset, the models' performances were evaluated and compared. Critical regions in the X-rays were visualized through Grad-CAM (Gradient-weighted Class Activation Mapping) heatmaps to enhance interpretability. The findings suggest that the custom CNN model outperforms the two fine-tuned ResNet-50 models in accuracy and diagnostic consistency. Additionally, the Grad-CAM heatmap results suggest potential overfitting and the need for further model refinement. This study highlights deep learning's promise as a diagnostic tool, particularly in areas with limited medical resources.**

Keywords—Pneumonia detection, Convolutional Neural Networks (CNNs), ResNet-50, Grad-CAM, Medical Imaging, Explainable AI, Deep Learning, Chest X-ray

I. Introduction

Pneumonia, an inflammatory infection of the lungs, is a major cause of illness and death worldwide, especially affecting children under five and the elderly. According to the World Health Organization (WHO), pneumonia is responsible for a substantial portion of childhood mortality, particularly in low-resource settings where healthcare access is limited [1]. Early and accurate detection is critical, as prompt treatment can significantly improve patient outcomes [2]. Currently, chest X-rays are among the most common diagnostic tools for pneumonia; however, interpreting these images requires expertise and is often time-consuming, especially in resource-limited settings where access to skilled radiologists is restricted [3]. This presents a compelling need for automated, reliable methods to assist or even perform pneumonia diagnosis, thereby enhancing diagnostic efficiency and accessibility.

Deep learning, specifically convolutional neural networks (CNNs), has shown remarkable success in image classification tasks [4] and holds promise for medical imaging applications, including pneumonia detection [5]. By training on large image datasets, CNNs can potentially learn to recognize complex patterns in X-ray images that signify pneumonia, providing consistent and objective diagnoses. Moreover, advancements in transfer learning allow models like ResNet-50—an established deep learning architecture pre-trained on extensive datasets [6]—to be fine-tuned for specialized medical tasks with relatively small amounts of medical data [7].

This study aims to leverage these capabilities by building and evaluating two deep learning approaches: a custom CNN model designed specifically for the nuances of X-ray images, and fine-tuning the ResNet-50 model (two fine-tuned models were evaluated). The performance of these models is analyzed in terms of their classification accuracy, precision, recall, and F1 scores on a publicly available chest X-ray dataset from Kaggle, consisting of normal and pneumonia cases. In addition, Grad-CAM (Gradient-weighted Class Activation Mapping) is applied to generate heatmaps highlighting regions of the X-ray images that contribute most to each classification, thereby offering interpretability and insight into the models' decision-making processes [8]. This interpretability is essential for clinical settings, where understanding the model's focus can build trust among healthcare providers.

By comparing the custom CNN with the two fine-tuned ResNet-50 models, our study seeks to provide insights into the efficacy of different approaches in pneumonia detection and the potential of deep learning as a diagnostic aid. The outcomes of this research could have significant implications for enhancing diagnostic capacity in under-resourced healthcare environments, where the burden of pneumonia is often the greatest.

II. Project Approach

A. Dataset Selection

This project utilized the publicly available *Kaggle Pneumonia Dataset* [9], which contains 5,856 labeled chest X-ray images divided into three categories: normal, bacterial pneumonia, and viral pneumonia. For the purpose of this study, the dataset was simplified into a binary classification problem—

normal versus pneumonia. Pneumonia cases were aggregated into a single category, irrespective of bacterial or viral origin. More specifically, the dataset features anterior-posterior (AP) chest X-rays of pediatric patients aged one to five years, collected retrospectively from Guangzhou Women and Children's Medical Center in Guangzhou, China. To ensure data quality and reliability, low-quality and unreadable scans were excluded from the dataset. Additionally, diagnoses were validated by two expert physicians, with a third physician conducting a secondary review of the evaluation set to address any grading errors.

Notably, the dataset offers clinically significant patterns as depicted in Fig. 1 below:

- Normal Chest X-rays: Depict clear lungs without abnormal opacification.
- Bacterial Pneumonia: Typically presents as a localized area of lung inflammation, often appearing as a bright, dense region in the X-ray (referred to as "focal lobar consolidation"), such as in the center X-ray in Fig. 1.
- Viral Pneumonia: Characterized by a widespread, defused patterns of inflammation throughout both lungs, often referred to as an "interstitial pattern," which appears as a hazy or streaky texture in the X-ray.

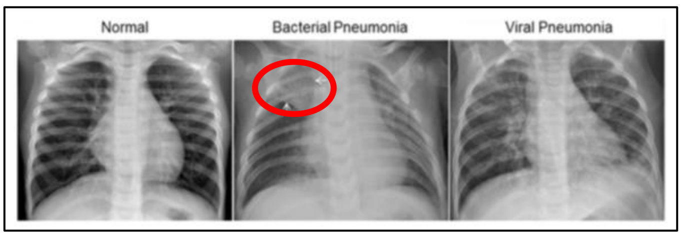

Fig. 1. Typical Normal, Bacterial Pneumonia, and Viral Pneumonia X-ray images.

These clinically significant patterns contained in the X-ray data provide a foundation for training machine learning models and evaluating their ability to distinguish normal from abnormal cases.

B. *Preprocessing*

As mentioned previously, the dataset contains 5,856 labeled chest X-ray images. This dataset was originally in JPEG format with varying resolutions. Preprocessing was completed to prepare the data to be in a more usable format for model training and development. These steps were crucial to standardizing the input data and enhancing model performance. More specifically, the following preprocessing steps were implemented:

1. Gray Scaling: The data was originally in JPEG format with RGB color channels, and was converted to a single, grayscale color channel to improve model speed and decrease latency.
2. Resizing: All images were resized to 180x180 pixels to ensure uniformity.
3. NumPy Formatting: The data was preprocessed to be NumPy arrays for model usability from the original JPEG format.
4. Normalization: Pixel values were normalized to a range of [0, 1] to expedite convergence during training.
5. Training/Test Split: The data was split into training and test sets in an 80/20 ratio to ensure robust evaluation while mitigating overfitting.

C. *Model Development*

Two different approaches to model development were utilized in this paper. The first approach was building and training a custom Convolutional Neural Network (CNN) model. The second approach was to fine-tune the well-known ResNet-50 model to our specific task. These two approaches are discussed in more detail as follows.

Custom CNN Model: The first approach in this paper was to develop a custom CNN model (architecture shown in Fig. 2) for the task of X-ray classification into healthy and pneumonia categories (Model 1). This model was trained specifically for this task and includes the following layer types:

- Convolutional Layers (x 3): Extract features such as edges and textures specific to chest X-rays.
- Pooling Layers (x 3): Reduce spatial dimensions while preserving essential information.
- Dense Layers (x 2): Enable classification by combining extracted features. The architecture was intentionally lightweight to ensure computational efficiency,

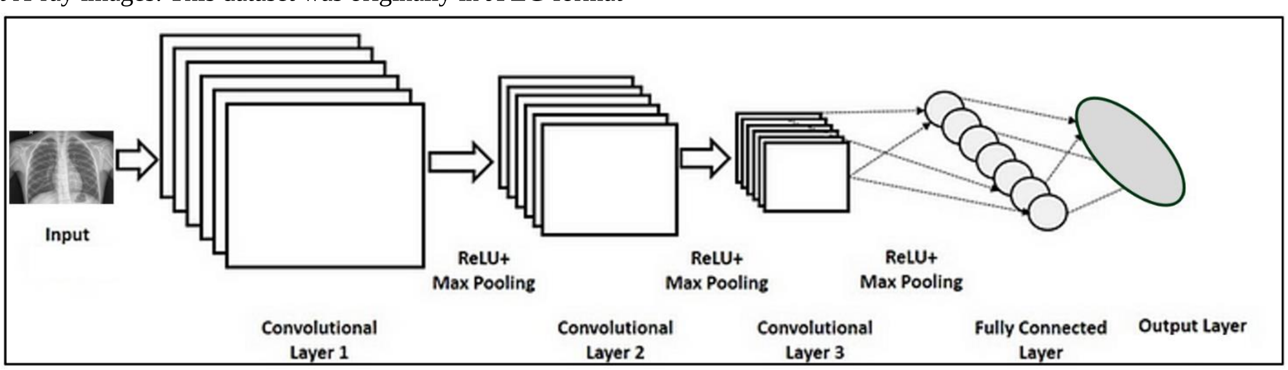

Fig. 2. Custom CNN Model Architecure (Model 1)

making it suitable for deployment in resource-limited settings.

- Dropout Layer (x 1): To prevent overfitting and improve model generalization, a dropout layer was included.

Fine-Tuned ResNet-50 Models: In addition to the custom CNN model, the pre-trained ResNet-50 architecture was fine-tuned for pneumonia detection. As mentioned previously, ResNet-50 is an established deep learning architecture pre-trained on extensive datasets [6]. Its deep residual learning framework allows it to mitigate vanishing gradient issues while maintaining high performance on image recognition tasks. By transferring knowledge from ImageNet pretraining, the model required only minimal retraining on the pneumonia dataset, making it highly effective even with limited data.

Fine-tuning involved replacing the top layer of ResNet-50 with a single pooling layer and two fully connected layers tailored for binary classification. The model was then retrained on the pneumonia dataset with a low learning rate. This approach leverages ResNet-50's robustness and capacity for hierarchical feature learning, enabling it to detect subtle patterns associated with pneumonia. Additionally, two model instances of this approach were evaluated (Models 2 and 3). The first model (Model 2) limited retraining to the newly added layers, while the second model (Model 3) allowed retraining to extend to the later layers of the base ResNet-50 model as well. This was done by applying a second phase of fine-tuning in which we unfroze the later layers of the base model for retraining.

D. *Model Development Justification*

The custom CNN model was developed to provide a simple, interpretable baseline tailored to the specific features of chest X-ray images, enabling a clearer comparison against more complex architectures. This model serves as a proof of concept for lightweight, deployable solutions in real-world scenarios.

Conversely, the base ResNet-50 model was chosen for its demonstrated success in diverse image classification tasks and its suitability for transfer learning. Fine-tuning ResNet-50 leverages its capacity for deep feature extraction, allowing it to identify intricate patterns that may escape a custom-built model. Additionally, two model instances of the fine-tuning approach were evaluated in order to study the impact of greater fine-tuning on model performance.

By comparing these two approaches, this study aims to assess the trade-offs between computational efficiency and diagnostic accuracy. Moreover, exploring multiple alternative approaches allows us to more thoroughly investigate the solution space for the task at hand, pneumonia detection, and observe which approaches yield better performance.

This comprehensive methodology ensures robust evaluation of the models while addressing the challenges of medical image analysis, including limited labeled data and the need for interpretability.

III. Results & Discussion

A. *Results*

The performance of the custom CNN model and the two fine-tuned ResNet-50 models was evaluated using accuracy, precision, recall, and F1 score as primary metrics. The results are summarized in Table 1 and Fig. 3.

Table 1. Model Performance Comparison.

Model	Accuracy	Precision	Recall	F1 Score
Model 1	95.6%	95.8%	98.2%	97.0%
Model 2	90.1%	91.7%	94.6%	93.1%
Model 3	93.3%	92.3%	99.3%	95.7%

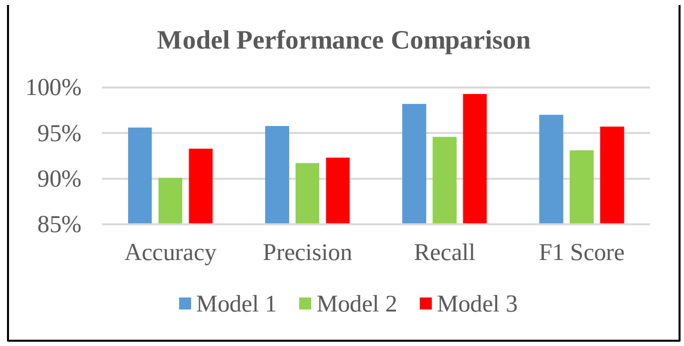

Fig. 3. Model Performance Comparison.

The three models are as follows:

- Model 1: Custom CNN
- Model 2: ResNet-50 Model with one phase of Fine-Tuning
- Model 3: ResNet-50 Model with two phases of Fine-Tuning

The custom CNN model (Model 1) achieved the highest overall accuracy (95.6%) and F1 score (97.0%), demonstrating its ability to effectively differentiate between normal and pneumonia X-rays. Among the fine-tuned ResNet-50 models, Model 3—where additional fine-tuning was applied to the later layers—outperformed Model 2 in all metrics. Moreover, Model 3 performance was only slightly less than that of the custom CNN model (93.3% as compared to 95.6% for accuracy and 95.7% as compared to 97.0% for F1 score).

Additionally, visualizations of the confusion matrices for all three models are shown in Fig. 4 to Fig. 6 below. We see that the number of false positive and false negative cases is low, indicative of strong models.

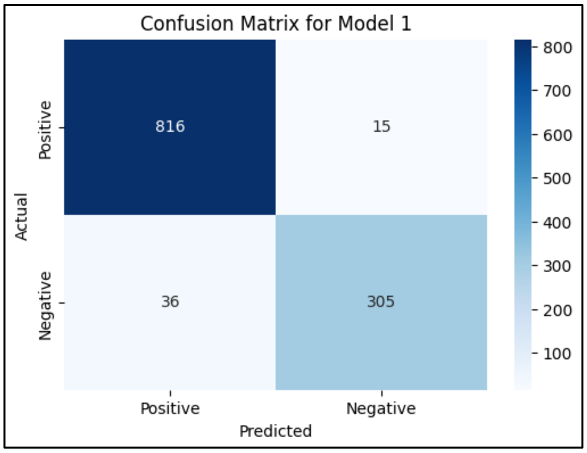

Fig. 4. Confusion Matrix for Model 1.

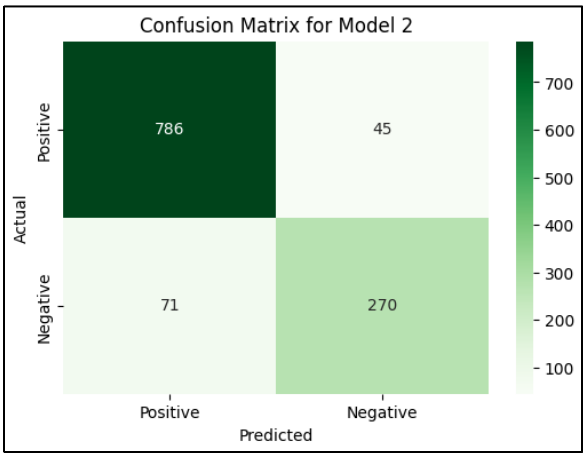

Fig. 5. Confusion Matrix for Model 2.

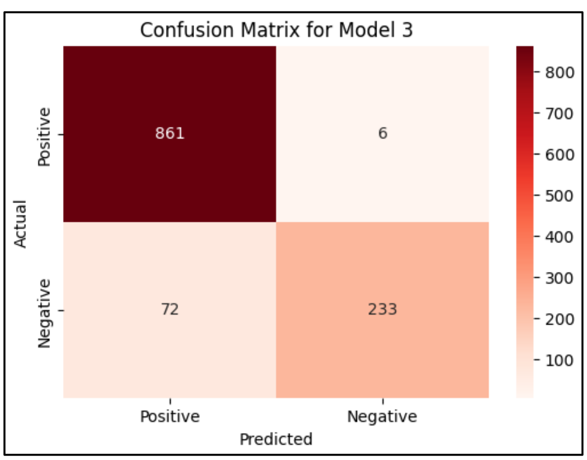

Fig. 6. Confusion Matrix for Model 3.

The Grad-CAM visualizations provided further insights into model decision-making processes. Heatmaps generated from correctly classified pneumonia cases revealed that the models often concentrated on regions of the lungs consistent with clinically significant patterns, such as areas of focal consolidation or interstitial patterns. However, very frequently the Grad-CAM heatmaps also highlighted irrelevant regions of the X-rays, such as regions outside the chest area, especially artifacts. Fig. 7 below shows a reasonable Grad-CAM heatmap while Fig. 8 and Fig. 9 show Grad-CAM heatmaps that are obviously erroneous.

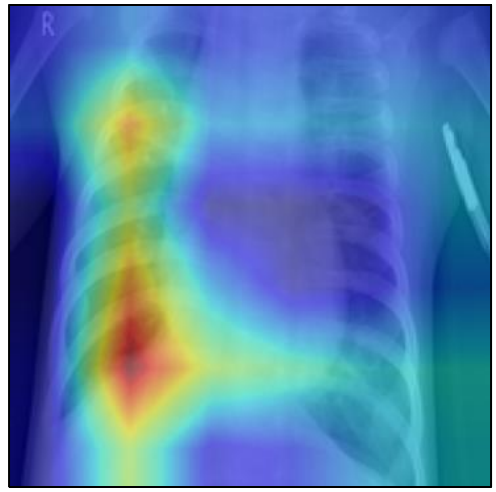
Fig. 7. Reasonable Grad-CAM Heatmap.

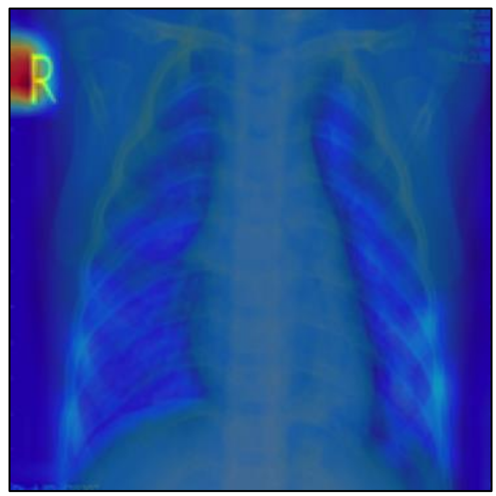
Fig. 8. Erroneous Grad-CAM Heatmap: Artifact.

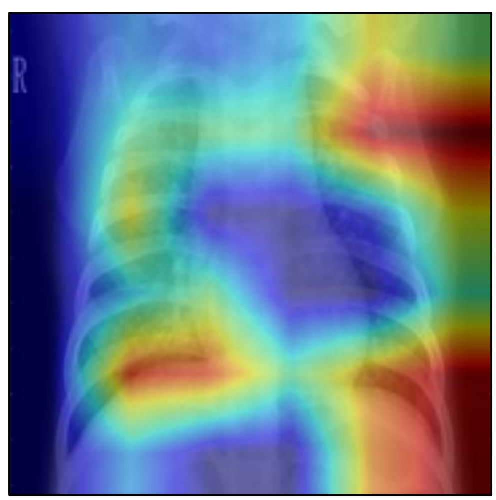
Fig. 9. Erroneous Grad-CAM Heatmap: Highlighting region outside of chest area.

B. Discussion

The results highlight the effectiveness of deep learning models in pneumonia detection from chest X-rays, with the custom CNN model demonstrating better overall performance than the fine-tuned ResNet-50 models. This suggests that a task-specific, lightweight architecture can compete with more complex pre-trained models when designed appropriately for the dataset.

The fine-tuned ResNet-50 models, while slightly behind the custom CNN in metrics, still exhibited robust performance. Model 3's improvement over Model 2 confirms that extending fine-tuning to deeper layers of the ResNet-50 base architecture enables the model to capture more domain-specific features.

Interpretability remains a crucial factor in deploying AI models for medical diagnostics. The Grad-CAM visualizations with their erroneous highlighting underscore the need for further model refinement in order to gain confidence in the models for clinical use. The incorrect heatmaps are indicative of potential overfitting that must be addressed. Interpretability is particularly important for gaining acceptance in clinical settings, where understanding the reasoning behind a model's prediction is critical.

IV. Conclusion & Future Work

A. Conclusion

This study demonstrates the potential of deep learning models, including custom CNNs and fine-tuned ResNet-50 architectures, for automated pneumonia detection using chest X-ray images. The custom CNN model achieved the best overall performance, while the fine-tuned ResNet-50 models offered comparable results with the added advantage of transfer learning.

The results underscore the promise of deep learning in enhancing diagnostic capacity, particularly in resource-limited settings where access to radiologists is scarce. Additionally, the use of Grad-CAM for explainability highlights the importance of model transparency in clinical applications, fostering trust and reliability in AI-assisted diagnostics. However, further refinement of the model is necessary to obtain satisfactory results of the heatmaps.

B. Future Work

To further enhance the reliability and applicability of this study's findings, several avenues for future research are proposed:

- Enhanced Preprocessing Methods: To prevent overfitting of the model and learning from artifacts, we can use more enhanced and thorough preprocessing methods, such as removing artifacts from the dataset. This would help improve the results of the Grad-CAM heatmaps as well.
- Dataset Expansion: Incorporating additional datasets, including those with adult X-rays and a greater diversity of pneumonia cases, to improve the generalizability of the models.
- Data Augmentation: Related to the previous point, adding data augmentation techniques can greatly increase the dataset size without the overhead of added data collection costs. In this way, the model can be enhanced without much additional investment. This would also help improve the results of the Grad-CAM heatmaps.
- Model Architecture Comparison: In this report, the ResNet-50 model was fine-tuned in addition to a custom CNN model. Other model architectures such as the VGG-16 model can also be evaluated for their effectiveness in addressing the task of automated pneumonia detection.
- Multi-Class Classification: Extending the model to differentiate between bacterial and viral pneumonia, which could provide additional diagnostic utility.
- Real-Time Deployment: Developing lightweight implementations for real-time pneumonia detection on smaller devices, such as mobile phones or portable X-ray machines.
- Collaboration with Clinicians: Engaging healthcare professionals in model evaluation to assess clinical usability and gather feedback for improvement.

By addressing these areas, future studies can build upon the results presented here to create even more robust, accessible, and clinically relevant AI tools for pneumonia detection.

Acknowledgment

The author would like to thank all those who contributed to the gathering of the X-rays and preparation of the dataset. The author would also like to thank the faculty at Missouri University of Science & Technology (Rolla, MO) for their dedication to science and knowledge. In particular, Dr. Cihan Dagli deserves acknowledgement for the guidance and mentorship given in the course of this project.

References

[1] World Health Organization, "Pneumonia," *World Health Organization Fact Sheets*, Nov. 2019. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia

[2] D. G. Musher and A. Thorner, "Community-acquired pneumonia," *The New England Journal of Medicine*, vol. 371, no. 17, pp. 1619-1628, 2014.

[3] P. Rajpurkar et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," *arXiv preprint arXiv:1711.05225*, 2017.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

[5] T. B. Moen, J. E. Li, K. C. Rogers, and M. M. Doshi-Velez, "An Automated Approach to Pneumonia Screening on Chest Radiographs Using Convolutional Neural Networks," *Proceedings of Machine Learning Research (MLHC)*, vol. 85, pp. 332-349, 2018.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[7] A. Tajbakhsh et al., "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299-1312, 2016.

[8] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.

[9] “Chest X-Ray Images (Pneumonia),” Kaggle, [Online]. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia. [Accessed: Nov. 16, 2024].