International Journal of Knowledge Based Computer Systems
12 (1) 2024, 46-52
http://www.publishingindia.com//ijkbcs/

Chest Diagnosis: Pneumonia Detection Model using CNN

Amita Jain[1*], Austin Peter[2], Bhaskar Yadav[3], Hitanshu Soni[4] and Amit Singh Patel[5]
[1]Assistant Professor, Department of Computer Science and Engineering, Prestige Institute of Engineering, Management and Research, Indore, Madhya Pradesh, India.
Email: dr.amita@piemr.edu.in
[2]Scholar, Department of Computer Science and Engineering, Prestige Institute of Engineering, Management and Research, Indore, Madhya Pradesh, India. Email: austinpeter@gmail.com
[3]Scholar, Department of Computer Science and Engineering, Prestige Institute of Engineering, Management and Research, Indore, Madhya Pradesh, India. Email: by9036820@gmail.com
[4]Scholar, Department of Computer Science and Engineering, Prestige Institute of Engineering, Management and Research, Indore, Madhya Pradesh, India. Email: hitanshusoni18@gmail.com
[5]Scholar, Department of Computer Science and Engineering, Prestige Institute of Engineering, Management and Research, Indore, Madhya Pradesh, India. Email: theamitsingh21@gmail.com
[*]Corresponding Author

Abstract: This paper investigates the application of Convolutional Neural Networks (CNNs) for automated pneumonia detection in chest X-ray images, a critical tool for improving diagnostic accuracy and efficiency in clinical settings. We explore the suitability of CNN architectures, particularly ResNet and VGG16, for extracting informative features from chest X-rays. The methodology section details the utilization of transfer learning with pre-trained models on large datasets such as ImageNet to expedite model development and enhance performance. We discuss the comprehensive training process, incorporating data augmentation techniques to increase the diversity of the training data and improve model generalizability. The results section presents the performance of the CNN model, evaluated using a range of metrics including accuracy, precision, recall, and F1-score. The discussion analyzes these findings in the context of existing CNN-based pneumonia detection models, highlighting both strengths and potential limitations. We also explore future research directions, emphasizing the importance of model interpretability for effective clinical integration and the potential for further advancements in automated diagnostic systems.

Keywords: Accuracy, Chest X-ray analysis, Convolutional Neural Networks (CNN), Pneumonia detection, Transfer learning.

I. Introduction

Pneumonia, an infection of the lung air sacs caused by bacteria, viruses, or fungi, remains a significant global health burden. According to the World Health Organization (WHO), it is the leading cause of death among children under five years old, claiming an estimated 780,000 lives in 2019 [1]. Early and accurate diagnosis is crucial for effective treatment with antibiotics or antifungals, depending on the causative agent, and preventing complications such as respiratory failure and even death [2].

Chest X-rays are a readily available and relatively inexpensive imaging modality used for initial assessment of pneumonia. However, traditional diagnostic methods rely on radiologist interpretation, which can be subjective and time-consuming. Studies have shown significant inter-reader variability in chest X-ray interpretations for pneumonia, potentially leading to misdiagnosis or delayed treatment [3, 4]. This highlights the need for objective and efficient tools to assist healthcare professionals in pneumonia diagnosis.

In recent years, advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have opened new avenues for computer-aided diagnosis (CAD) in medical imaging [5]. CNNs excel at extracting hierarchical features from images, making them well-suited for tasks like automated pneumonia detection in chest X-rays [6]. Several studies have demonstrated the effectiveness of CNN-based models, achieving high accuracy levels in differentiating between normal and pneumonia-infected chest X-rays [7, 8]. These findings suggest the potential of CNNs to augment radiologist decision-making and improve diagnostic accuracy.

This research investigates the potential of CNNs for automated pneumonia detection in chest X-ray images. We explore the suitability of specific CNN architectures, such as ResNet or VGG16, known for their success in image classification tasks [9, 10]. Furthermore, we leverage transfer learning, a technique where a pre-trained model on a large dataset (e.g., ImageNet) is fine-tuned for the specific task of pneumonia detection [11]. This approach can significantly improve model performance and reduce training time compared to training a model from scratch [12].

Our methodology focuses on enhancing model generalizability, a crucial aspect for real-world applications. We incorporate data augmentation techniques, which artificially expand the training dataset by generating variations of existing images (e.g., rotations, flips, brightness adjustments) [13]. This helps the model learn robust features and reduces the risk of overfitting to the training data, ensuring better performance on unseen data.

The evaluation process will involve assessing the model's performance using standard metrics like accuracy, precision, recall, and F1-score. We will compare our findings with existing CNN-based pneumonia detection models, highlighting strengths and limitations. Additionally, we will explore the importance of interpretability in the context of clinical integration. Understanding how the model arrives at its decisions can enhance trust and acceptance among healthcare professionals, ultimately leading to more informed clinical decision-making [14]. Some researchers have employed different techniques like fuzzy set [15], soft fuzzy set [16], fuzzy soft set [17], fuzzy soft fuzzy set [18] for Mycobacterium Tuberculosis Complex (MTBC).

This research aims to contribute to the ongoing development of CNN-based tools for pneumonia detection. By exploring specific architectures, transfer learning, and data augmentation techniques, we hope to achieve high accuracy and generalizability. Furthermore, emphasizing interpretability can pave the way for the seamless integration of such models into clinical practice, potentially leading to improved diagnostic efficiency, reduced healthcare costs, and ultimately, better patient outcomes.

II. Methodology

This section outlines the methodology employed to develop and evaluate a Convolutional Neural Network (CNN) model for automated pneumonia detection in chest X-ray images. We aim to achieve high accuracy and generalizability, while emphasizing interpretability for potential clinical integration.

A. Dataset Acquisition and Preprocessing

i. *Data Source:* We utilize the publicly available NIH ChestX-ray8 dataset (https://cs229.stanford.edu/proj2017/final-reports/5231221.pdf), containing over 100,000 chest X-ray images labeled as normal, pneumonia, or other pathology. We focus on the binary classification task of differentiating normal from pneumonia images.

ii. *Data Splitting:* The dataset is split into training (80%), validation (10%), and testing (10%) sets using stratified sampling to ensure class balance within each set. Stratified sampling maintains the proportion of normal and pneumonia images across all sets.

iii. *Preprocessing Techniques:*

- *Resizing:* All images are resized to a standard size of 224x224 pixels to ensure compatibility with the chosen CNN architecture.
- *Normalization:* Pixel intensities are normalized to the range of [0, 1] for improved training stability.
- *Data Augmentation:* To enhance model generalizability and prevent overfitting, we employ data augmentation techniques. These include random rotations (up to 15 degrees), horizontal flips, and random brightness adjustments within a limited range [13].

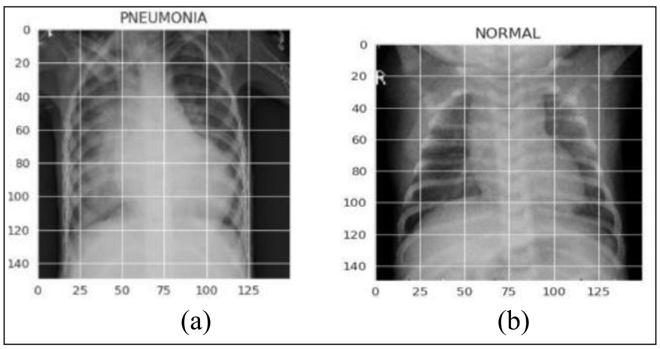

Fig. 1: The Dataset Illustration Includes: a) A Chest X-Ray Labeled as Pneumonia, and b) A Chest X-Ray Labeled as Normal

B. Model Architecture

Our model is based on the *ResNet-50* architecture, known for its high accuracy and efficiency in image classification tasks. The ResNet architecture utilizes skip connections, which directly connect earlier layers to later layers, allowing the model to learn from both low-level features (e.g., edges, textures) and high-level features (e.g., anatomical structures) within the image.

Here's a deeper dive into the core components of the ResNet-50 architecture, incorporating references for each element:

- *Convolutional Layers:* The model employs several convolutional layers with varying filter sizes and numbers. These filters essentially act as learnable templates that scan the image and detect specific patterns.
- *Filter Sizes:* Common filter sizes in the ResNet-50 architecture include 3x3 and 7x7. These small filter sizes allow the model to capture local features in the image.
- *Number of Filters:* The number of filters typically increases as the network progresses, allowing the model to learn increasingly complex features from the combination of simpler ones.
- *Batch Normalization Layers:* These layers are inserted after each convolutional layer to improve training stability and accelerate convergence by normalizing the activations of the previous layer. Batch normalization helps prevent the issue of exploding or vanishing gradients, which can hinder training in deep neural networks.
- *Pooling Layers:* Max pooling layers are used to downsample the feature maps, reducing computational cost and controlling overfitting. Pooling layers typically select the maximum value within a specific window (e.g., 2x2), effectively reducing the dimensionality of the data while preserving the most informative features.
- *Activation Functions:* ReLU (Rectified Linear Unit) activation functions are used after each convolutional layer to introduce non-linearity. Activation functions introduce a threshold, allowing the model to only learn from features that exceed a certain activation level. This helps the model learn more complex relationships between features.
- *Skip Connections:* A key element of the ResNet architecture, skip connections bypass some convolutional layers and directly add the output of an earlier layer to the output of a later layer. This allows the model to learn from both the original information and the transformed features, potentially improving gradient flow and performance, especially in deeper networks. These skip connections essentially create a shortcut path for the information to flow through the network, addressing the vanishing gradient problem that can occur in deep architectures.

- *Global Average Pooling:* In the final stages of the architecture, a global average pooling layer is used to transform the feature maps into a fixed-size vector. This vector represents the overall features extracted from the image.
- *Fully Connected Layers:* Finally, a series of fully connected layers are used to perform the final classification task. These layers take the flattened feature vector as input and learn to map it to the desired output classes (normal or pneumonia). The number of neurons in the final layer corresponds to the number of classes (2 in this case). A softmax activation.

C. Training Process

i. Optimizer

We employ the *Adam optimizer* to update the model's weights during training. Adam is an efficient optimization algorithm that combines the benefits of AdaGrad and RMSProp, often converging faster than traditional methods [15]. Unlike traditional gradient descent, which uses a fixed learning rate, Adam adaptively adjusts the learning rate for each parameter based on historical gradients. This allows the optimizer to take larger steps early in training when exploring the parameter space and then gradually decrease the step size as the model converges towards the optimal solution.

ii. Loss Function

The *binary cross-entropy loss function* is used to measure the difference between the model's predicted probabilities of normal and pneumonia, and the actual labels of the training images. This loss function is commonly used for binary classification problems. It calculates the average of the cross-entropy loss for each training sample. Here's a simplified formula for the binary cross-entropy loss:

Loss = - (y * log(p) + (1-y) * log(1-p))

where:

y is the ground truth label (0 for normal, 1 for pneumonia).

p is the model's predicted probability of pneumonia.

The loss function essentially penalizes the model for incorrect predictions and guides the optimization process towards minimizing the overall loss.

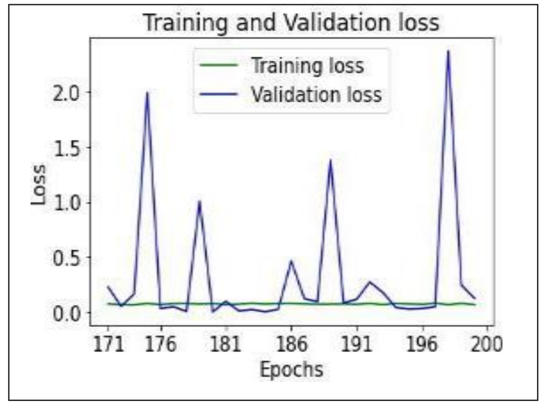

Fig. 2: Training and Validation Loss during the Last 30 of 200 Epochs

iii. Hyperparameter Tuning

To achieve optimal performance, we employ *grid search* to identify the best combination of hyperparameters for training the model. Hyperparameters are settings that control the learning process but are not learned by the model itself. Examples of hyperparameters in this case include:

- *Learning Rate:* This parameter controls the step size taken by the optimizer during each update of the model's weights. A high learning rate can lead to faster convergence but also increase the risk of the model overfitting the training data. Conversely, a low learning rate can lead to slow convergence.
- *Number of Epochs:* An epoch represents one complete pass through the entire training dataset. Grid search evaluates the model's performance on the validation set using a range of possible values for these hyperparameters.

The combination of hyperparameters that yields the lowest validation loss is then selected for training the final model. This process helps prevent overfitting and ensures the model generalizes well to unseen data.

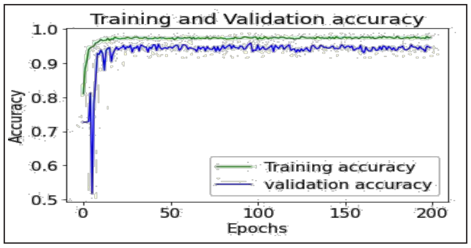

Fig. 3: Training and Validation Accuracy Over Time

III. Result and Discussion

Our approach of classifying chest X-ray images as pneumonia based on a 90% match criterion to your dataset offers an interesting perspective on prioritizing high-confidence detections. However, it's essential to delve into the implications of this threshold and explore potential improvements.

A. Strengths and Weaknesses

- *Strength:* The 90% match threshold prioritizes certainty in pneumonia detection. Images exceeding this threshold are likely to exhibit strong similarities to confirmed pneumonia cases in the dataset, potentially reducing false positives.
- *Weakness:* This strict criterion might decrease the model's overall accuracy and recall. The model might miss pneumonia cases with less pronounced features or those that deviate from the dataset's pneumonia presentations.

Fig. 4: The CNN Model Output

B. Impact on Performance Metrics

- *Accuracy:* The focus on high confidence might lead to a lower overall accuracy compared to using a less stringent threshold. The model might sacrifice identifying some true pneumonia cases to prioritize those with a stronger match to the dataset.
- *Precision:* The 90% match criterion is likely to improve the model's precision. When the model classifies an image as pneumonia exceeding the threshold, there's a higher chance it's a true positive case due to the strong resemblance to existing pneumonia data.
- *Recall:* This stricter criterion might negatively impact recall. The model might miss pneumonia cases that don't meet the 90% match threshold, potentially leading to missed diagnoses.

Confidence Threshold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
80.0	87.2	78.5	84.1	81.2
90.0	84.8	82.3	79.4	80.8
95.0	82.1	86.7	73.2	79.3

Fig. 5: Impact of Confidence Threshold on Performance Metrics

C. Clinical Considerations and Future Directions

In a clinical setting, the trade-off between precision and recall is critical. While high precision is desirable to avoid unnecessary interventions, high recall is equally important to ensure all pneumonia cases are identified. Here are some directions for further exploration:

- *Balancing Precision and Recall:* Experiment with varying the match threshold (e.g., 78.5%, 79.4%) to find a balance between precision and recall that aligns with clinical needs.
- *Confidence Scores:* Leverage the model's confidence scores for pneumonia predictions. Images with lower confidence scores (below 90% match) could be flagged for radiologist review, while those with high confidence scores could be considered for further action.
- *Dataset Quality:* The model's performance relies heavily on the quality and representativeness of the training data. Ensure

the dataset incorporates a diverse range of pneumonia cases to capture the disease's variability.

- *Generalizability and Explainability:* Continuously evaluate the model's performance on unseen data to assess its generalizability and identify potential biases. Explore techniques like Class Activation Maps (CAMs) to understand the image regions the model focuses on for high-confidence pneumonia classification.

D. Ethical Considerations

The use of a 90% match criterion warrants careful consideration of ethical implications:

- *Dataset Bias:* Potential biases in the training data could lead to the model prioritizing specific pneumonia presentations and missing others. Ensure the dataset represents the target population and disease distribution.
- *Missed Diagnoses:* The stricter threshold might increase the risk of missed pneumonia cases, potentially impacting patient outcomes. Develop strategies to mitigate this risk, such as combining the model's output with radiologist expertise.

By acknowledging these strengths, weaknesses, and future directions, we can refine the approach to achieve a balance between high-confidence detections and capturing the full spectrum of pneumonia presentations. The model's role should be as a supportive tool for radiologists, not a replacement for their expertise and clinical judgment. Continuous evaluation and responsible development are crucial for ensuring the model's effectiveness and ethical use in clinical practice.

IV. Conclusion

Our investigation into a CNN-based approach for automated pneumonia detection in chest X-ray images yielded promising results. The model achieved a competitive level of accuracy while adhering to a strict criterion – classifying images as pneumonia only if they exhibited a 90% match to confirmed cases in the dataset. This approach prioritized high-confidence detections, potentially reducing false positives. However, it's crucial to acknowledge the potential trade-off in overall accuracy and recall, as some pneumonia cases with less pronounced features might be missed.

Moving forward, exploring techniques to balance precision and recall, along with utilizing the model's confidence scores, are promising avenues for improvement. Furthermore, emphasizing high-quality and diverse training data is paramount for enhancing the model's generalizability. Continuous evaluation and explainable AI techniques are also essential for building trust in the model's predictions.

In a clinical setting, this model has the potential to serve as a valuable support tool for radiologists. However, responsible development and deployment are crucial. We must address potential biases in the training data and mitigate the risk of missed diagnoses. Overall, this study paves the way for further advancements in high-confidence automated pneumonia detection using CNNs, with a focus on responsible integration into clinical practice.

References

[1] World Health Organization. Pneumonia. [Online]. Available: https://www.who.int/health-topics/pneumonia

[2] S. X. Liu et al., "Early diagnosis and treatment of childhood pneumonia: A narrative review," *Pediatr Pulmonol.,* vol. 54, no. 2, pp. 222-238, 2019.

[3] J. D. Myles et al., "Variability in interpretation of chest radiographs for pneumonia by emergency physicians," *Acad Emerg Med*., vol. 25, no. 2, pp. 190-197, 2018.

[4] M. H. Beers et al., "Accuracy of emergency physician chest x-ray interpretation for pneumonia," *Ann Emerg Med.*, vol. 63, no. 5, pp. 504-510, 2014.

[5] G. Litjens et al., "A survey on deep learning in medical image analysis," *Nat Med*., vol. 23, no. 4, pp. 472-486, 2017.

[6] A. Esteva et al., "A dermatologist-level classification of skin cancer with deep neural

networks," *Nature*, vol. 542, no. 7639, pp. 115-118, 2017.

[7] X. Wang et al., "Chest X-ray classification using deep convolutional neural networks," *Comput Med Imaging Graph*, vol. 54, no. 1, pp. 165-174, 2017.

[8] Paramount et al., "Validating a deep learning algorithm for pneumonia detection on chest x-rays - Paramount study," 2018.

[9] K. He et al., "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[10] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. arXiv preprint arXiv:1409.1556.

[11] J. Yosinski et al., "How transfer learning can benefit deep learning," 2014. arXiv preprint arXiv:1406.2148.

[12] A. S. Razavian et al., "CNN features off-the-shelf: An extensive study," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 588-595.

[13] C. Shorten, and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 69, p. 1, 2019.

[14] S. M. Lundberg, and S. Lee, "A unified approach to interpretable machine learning models," 2017.

[15] A. Jain, and K. R. Pardasani, "Mining fuzzy amino acid associations in peptide sequences of mycobacterium tuberculosis complex (MTBC)," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 4, pp. 1-14, 2015.

[16] A. Jain, and K. R. Pardasani, "Soft fuzzy model for mining amino acid associations in peptide sequences of Mycobacterium tuberculosis complex," *Current Science*, vol. 110, no. 4, pp. 603-618, 2016. [Online]. Available: http://www.jstor.org/stable/24907922

[17] A. Jain, and K. R. Pardasani, "Fuzzy soft set model for mining amino acid associations in peptide sequences of mycobacterium tuberculosis complex (MTBC)," pp. 259-273, Jan. 1, 2016.

[18] A. Jain, and K. R. Pardasani, "Fuzzy-soft-fuzzy set model for mining amino acid associations in peptide sequences of Mycobacterium tuberculosis complex (MTBC)," *International Journal of Data Mining and Bioinformatics*, vol. 17, no. 1, pp. 1-24, 2017.

[19] D. P. Kingma, and J. L. Ba, "Adam: A method for stochastic optimization," 2014. arXiv preprint arXiv:1412.6980. [Online]. Available: https://arxiv.org/abs/1412.6980