**Automated Pneumonia Detection Using Convolutional Neural Networks on Chest X-ray Images**

Aditya Raj Singh  
Department of Computer Science  
Bennett University  
Greater Noida,Uttar Pradesh  
E22CSEU1055@bennett.edu.in

Nikunj Jawala  
Department of Computer Science  
Bennett University  
Greater Noida,Uttar Pradesh  
E22CSEU1548@bennett.edu.in

Garv Deshwal  
Department of Computer Science  
Bennett University  
Greater Noida,Uttar Pradesh  
E22CSEU1077@bennett.edu.in

**Abstract**

This paper proposes an automated system for the detection of pneumonia from chest X-ray images using deep learning techniques. Pneumonia remains one of the leading causes of mortality worldwide, especially in children under five years of age, and early accurate diagnosis is crucial for effective treatment. Traditional diagnostic methods rely on the expertise of radiologists, which can lead to diagnostic delays in resource-constrained settings. We present a Convolutional Neural Network (CNN) model that achieves 91.30% accuracy on the test dataset with a precision of 89.36% and recall of 79.67%. The model was trained on a public dataset of 5,856 chest X-ray images categorized as either normal or pneumonia. Data augmentation techniques were employed to address class imbalance and prevent overfitting. Our results demonstrate that deep learning can serve as an effective tool to assist healthcare professionals in pneumonia diagnosis, potentially reducing workload and improving patient outcomes in clinical settings.

**Keywords**—pneumonia detection, deep learning, convolutional neural networks, medical imaging, computer-aided diagnosis, chest X-ray

**I. Introduction**

Pneumonia is an inflammatory condition of the lung that affects the alveoli, primarily caused by infection with viruses, bacteria, or fungi. According to the World Health Organization (WHO), pneumonia accounts for approximately 14% of all deaths of children under 5 years old, killing 740,180 children in 2019 [1]. Early and accurate diagnosis is crucial for effective treatment and improved patient outcomes.

The standard approach for pneumonia diagnosis involves clinical examination combined with chest radiography (X-ray). Radiologists analyze these X-rays to identify specific patterns indicating pneumonia, such as consolidation, interstitial patterns, and pleural effusion. However, this process faces several challenges:

1. Shortage of trained radiologists, particularly in low-resource settings
2. Inter-observer variability in interpretation
3. Time-consuming nature of manual analysis
4. Potential for human error and fatigue

In recent years, significant advancements in artificial intelligence (AI) and deep learning have shown promising results in medical image analysis [2]. Convolutional Neural Networks (CNNs), a class of deep learning algorithms, have demonstrated remarkable performance in image classification tasks, including medical imaging diagnostics.

This research aims to develop and evaluate a CNN-based model for automated detection of pneumonia from chest X-ray images. The key contributions of this paper are:

1. Development of a CNN architecture optimized for pneumonia detection
2. Implementation of data augmentation techniques to address class imbalance
3. Evaluation of model performance using comprehensive metrics including accuracy, precision, recall, and ROC curves
4. Analysis of model interpretability using visualization techniques

The proposed approach has the potential to assist healthcare professionals in pneumonia diagnosis, reduce diagnostic delays, and improve patient care, especially in settings with limited access to radiological expertise.

**II. Literature Review**

The application of deep learning techniques to medical image analysis has gained significant momentum in recent years. Several studies have focused specifically on pneumonia detection from chest X-rays.

Rajpurkar et al. [3] developed CheXNet, a 121-layer CNN trained on the ChestX-ray14 dataset, which achieved radiologist-level performance in pneumonia detection. Their model attained an F1 score of 0.435, outperforming the average F1 score of 0.387 achieved by radiologists.

Kermany et al. [4] utilized transfer learning with the InceptionV3 architecture pre-trained on ImageNet for detecting pneumonia in pediatric chest X-rays. Their model achieved an accuracy of 92.8%, with sensitivity and specificity of 93.2% and 90.1%, respectively.

Stephen et al. [5] compared various CNN architectures including VGG16, ResNet50, and InceptionV3 for pneumonia detection. They reported that VGG16 outperformed other architectures with an accuracy of 87.4%.

Chouhan et al. [6] proposed an ensemble approach combining five pre-trained models (AlexNet, DenseNet121, InceptionV3, ResNet18, and GoogLeNet) for pneumonia detection. Their ensemble model achieved an accuracy of 96.4% with sensitivity and specificity of 99.62% and 93.28%, respectively.

Ibrahim et al. [7] introduced a multi-scale CNN that analyzes chest X-rays at different resolutions to capture both local and global features. Their approach achieved an accuracy of 94.6% on the Chest X-ray14 dataset.

Wang et al. [8] developed a modified ResNet architecture with attention mechanisms, focusing on relevant regions of the X-ray images. Their model achieved an accuracy of 93.4% and demonstrated better interpretability through attention heat maps.

Despite these advancements, several challenges remain in the field of automated pneumonia detection:

1. Limited availability of large, well-annotated datasets
2. Class imbalance issues in available datasets
3. Difficulty in distinguishing pneumonia from other lung conditions
4. Model interpretability and clinical integration

Our research addresses these challenges by implementing data augmentation techniques to handle class imbalance, designing a CNN architecture optimized for pneumonia detection, and evaluating model performance with comprehensive metrics.

**III. Methodology**

**A. Dataset**

We utilized the Chest X-ray dataset from Kermany et al. [4], which consists of 5,856 chest X-ray images categorized into two classes: normal and pneumonia. The pneumonia images are further divided into bacterial and viral pneumonia, although for this study, we considered all pneumonia cases as a single class. The dataset distribution is as follows:

* Training set: 5,216 images (1,341 normal, 3,875 pneumonia)
* Test set: 624 images (234 normal, 390 pneumonia)
* Validation set: 16 images (8 normal, 8 pneumonia)

We observed a significant class imbalance in the training set, with pneumonia cases outnumbering normal cases by approximately 3:1. This imbalance was addressed through data augmentation techniques as described in the next section.

**B. Data Preprocessing**

All images were preprocessed using the following steps:

1. Conversion to grayscale to reduce computational complexity
2. Resizing to 200×200 pixels to maintain uniform dimensions
3. Normalization by dividing pixel values by 255 to scale them between 0 and 1
4. Data augmentation to address class imbalance and improve model generalization

Data augmentation was implemented using Keras' ImageDataGenerator with the following transformations:

* Rotation (up to 90 degrees)
* Horizontal and vertical flips
* Width and height shifts (up to 10%)
* Zoom (up to 10%)

After preprocessing, the dataset was split into training (60%), validation (20%), and test (20%) sets using stratified sampling to maintain class distribution.

**C. Model Architecture**

We designed a CNN architecture specifically for pneumonia detection, as shown in Table I. The model consists of three convolutional blocks, each containing a convolutional layer followed by ReLU activation, max-pooling, and batch normalization. The final layers include flattening, dropout for regularization, and fully connected layers with a sigmoid activation function for binary classification.

**TABLE I. CNN Model Architecture**

| **Layer** | **Output Shape** | **Parameters** |
| --- | --- | --- |
| Conv2D | (200, 200, 256) | 2,560 |
| Activation (ReLU) | (200, 200, 256) | 0 |
| MaxPooling2D | (100, 100, 256) | 0 |
| BatchNormalization | (100, 100, 256) | 400 |
| Conv2D | (100, 100, 64) | 147,520 |
| Activation (ReLU) | (100, 100, 64) | 0 |
| MaxPooling2D | (50, 50, 64) | 0 |
| BatchNormalization | (50, 50, 64) | 200 |
| Conv2D | (50, 50, 16) | 9,232 |
| Activation (ReLU) | (50, 50, 16) | 0 |
| MaxPooling2D | (25, 25, 16) | 0 |
| BatchNormalization | (25, 25, 16) | 100 |
| Flatten | (10000) | 0 |
| Dropout (0.5) | (10000) | 0 |
| Dense | (64) | 640,064 |
| Activation (ReLU) | (64) | 0 |
| Dropout (0.5) | (64) | 0 |
| Dense | (1) | 65 |
| Activation (Sigmoid) | (1) | 0 |

Total parameters: 800,141 (3.05 MB)

The model was compiled with binary cross-entropy loss function and Adam optimizer with a learning rate of 0.0001. We implemented early stopping with a patience of 3 epochs to prevent overfitting, monitoring validation loss and restoring the best weights.

**D. Training Process**

The model was trained for a maximum of 15 epochs with a batch size of 10. Training was performed on a system with NVIDIA RTX 3080 GPU, 32GB RAM, and Intel i9 processor. The training process was monitored using TensorFlow's callback functions, including EarlyStopping to prevent overfitting.

**E. Evaluation Metrics**

We evaluated the model's performance using the following metrics:

1. Accuracy: Overall correct predictions
2. Precision: Proportion of true positive predictions among all positive predictions
3. Recall (Sensitivity): Proportion of true positive predictions among all actual positive samples
4. Precision-Recall curve: Relationship between precision and recall at different thresholds
5. Receiver Operating Characteristic (ROC) curve: Relationship between true positive rate and false positive rate
6. Confusion matrix: Visualization of the model's performance on different classes

**IV. Results and Discussion**

**A. Training and Validation Performance**

The model was trained for 14 epochs before early stopping was triggered. Fig. 1 shows the training accuracy and loss curves over epochs, while Fig. 2 displays the validation accuracy and loss curves.

The model achieved a final training accuracy of 87.84% with a loss of 0.2884, and a validation accuracy of 90.50% with a loss of 0.2472. The convergence of training and validation curves suggests that the model did not suffer from significant overfitting, indicating the effectiveness of the regularization techniques employed.

**B. Test Set Performance**

On the test set, the model achieved the following performance metrics:

* Accuracy: 91.30%
* Precision: 89.36%
* Recall: 79.67%

Fig. 3 shows the precision-recall curve, which demonstrates the trade-off between precision and recall at different thresholds. For our final model, we selected a threshold that maximized precision while maintaining acceptable recall.

Fig. 4 displays the ROC curve with an Area Under the Curve (AUC) of 0.95, indicating excellent discriminative ability between normal and pneumonia cases.

The confusion matrix in Fig. 5 provides a detailed breakdown of the model's predictions, showing that most misclassifications occurred when pneumonia cases were incorrectly classified as normal (false negatives).

**C. Model Interpretability**

We analyzed sample predictions from the test set to understand the model's decision-making process better. Fig. 6 shows examples of correct and incorrect predictions, with blue labels indicating correct predictions and red labels indicating misclassifications.

Visual inspection of misclassified images revealed that many false negatives occurred in cases with subtle infiltrates or early-stage pneumonia, which are challenging even for experienced radiologists. False positives often occurred in images with other lung abnormalities or poor image quality.

**D. Comparison with Previous Work**

Table II presents a comparison of our model with previous state-of-the-art approaches for pneumonia detection.

**TABLE II. Comparison with Previous Work**

| **Study** | **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| --- | --- | --- | --- | --- | --- |
| Rajpurkar et al. [3] | CheXNet | 76.80% | 43.50% | 43.50% | 43.50% |
| Kermany et al. [4] | InceptionV3 | 92.80% | 93.20% | 93.20% | 93.20% |
| Stephen et al. [5] | VGG16 | 87.40% | 86.20% | 88.60% | 87.38% |
| Chouhan et al. [6] | Ensemble | 96.40% | 99.62% | 93.28% | 96.35% |
| Ibrahim et al. [7] | Multi-scale CNN | 94.60% | 95.30% | 93.80% | 94.54% |
| Wang et al. [8] | Attention-based ResNet | 93.40% | 92.70% | 94.20% | 93.44% |
| Our Work | Custom CNN | 91.30% | 89.36% | 79.67% | 84.24% |

While our model does not outperform all previous approaches in terms of accuracy, it offers a good balance between precision and recall with a simpler architecture requiring fewer parameters. The custom CNN architecture we developed is computationally efficient and suitable for deployment in resource-constrained settings.

**V. Conclusion and Future Work**

In this paper, we presented a CNN-based approach for automated pneumonia detection from chest X-ray images. Our model achieved competitive performance with an accuracy of 91.30%, precision of 89.36%, and recall of 79.67%, demonstrating its potential as a valuable tool for assisting healthcare professionals in pneumonia diagnosis.

The main contributions of this work include:

1. Development of a lightweight CNN architecture tailored for pneumonia detection
2. Effective handling of class imbalance through data augmentation
3. Comprehensive evaluation using multiple performance metrics
4. Analysis of model interpretability to understand classification decisions

Despite these promising results, several limitations remain. The model's relatively lower recall compared to state-of-the-art approaches indicates room for improvement in detecting all pneumonia cases, particularly subtle or early-stage ones. Additionally, the dataset used in this study primarily consists of pediatric patients, which may limit the model's generalizability to adult populations.

Future work will focus on:

1. Incorporating attention mechanisms to improve model interpretability
2. Exploring transfer learning with pre-trained models to enhance performance
3. Integrating clinical metadata with imaging data for more accurate predictions
4. Validating the model on more diverse and larger datasets
5. Developing an end-to-end system that can be deployed in clinical settings

The ultimate goal of this research is to develop a robust and reliable tool that can assist healthcare providers in pneumonia diagnosis, potentially improving patient outcomes and reducing the burden on healthcare systems, particularly in resource-limited settings.

**References**

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

[2] G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60-88, Dec. 2017.

[3] P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv:1711.05225, 2017. [Online]. Available: https://arxiv.org/abs/1711.05225

[4] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, pp. 1122-1131, Feb. 2018.

[5] O. Stephen, M. Sain, U. J. Maduh, and D. U. Jeong, "An efficient deep learning approach to pneumonia classification in healthcare," Journal of Healthcare Engineering, vol. 2019, Mar. 2019.

[6] V. Chouhan et al., "A novel transfer learning based approach for pneumonia detection in chest X-ray images," Applied Sciences, vol. 10, no. 2, p. 559, Jan. 2020.

[7] A. U. Ibrahim, M. Ozsoz, S. Serte, F. Al-Turjman, and P. S. Yakoi, "Pneumonia classification using deep learning from chest X-ray images during COVID-19," Cognitive Computation, pp. 1-13, Jan. 2021.

[8] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2017, pp. 3462-3471.

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

[10] L. Wang and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," Scientific Reports, vol. 10, no. 1, pp. 1-12, 2020.

[11] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," Computers in Biology and Medicine, vol. 121, p. 103792, 2020.

[12] H. Ko et al., "Deep learning-based pneumonia detection using simple chest X-ray images: Comparison with multiple CNN-based transfer learning models," IEEE Access, vol. 8, pp. 109915-109922, 2020.