Deep Learning-Based Pneumonia Detection from Chest X-rays: A Comparative Analysis of CNN Architectures

Muhammad Zunnurain Hussain[1*], Saad Hassan[2], Mahad Hassan[2], Saad Hussain[2], Chaudhary Rafeel Abdul Majid[2]

1 Dept. Computer Science, Bahria University Lahore Campus, Pakistan
2 Dept. Computer Science, National University of Computer and Emerging Sciences (FAST-NUCES), Pakistan

These authors contributed equally to this work.
* Corresponding author: zunnurain.bulc@bahria.edu.pk

Abstract

An accurate and timely detection of pneumonia, a life threatening lung infection, is critical to effective treatment. In this paper, we present a deep learning based system for the automation of detecting pneumonia in chest X-ray images using Convolutional Neural Networks (CNNs). In this work we used custom CNN, VGG16, and ResNet-50 models trained on Chest X-ray Pneumonia Dataset. Then each model's performance was evaluated by looking at training/testing accuracy and confusion matrices. Our results suggest that deep learning may be highly capable of improving diagnostic precision, reduce human error, and increase healthcare accessibility, particularly in resource constrained settings.

Introduction

Despite years of improvements in global radiology, pneumonia continues to be a major public health challenge in regions where there is limited access to skilled radiologists. Constrained by time, resources and human error, traditional diagnostic tools are not delivering early, accurate diagnoses that are necessary for effective treatment. Recently, deep learning approaches, in particular, in medical image analysis have been proven to be very promising to address these challenges. In this research, we use deep learning for automated detection of pneumonia from chest X-Ray images [1]. With this in mind, we implement and compare three distinct architectures for this critical diagnostic task, to assess such CNN architectures. We attempt to allow for development of such reliable, automated diagnostic tools that enable healthcare professionals to detect pneumonia.

1 Related Work

Deep learning has emerged as a groundbreaking approach in medical image analysis especially for the detection of pneumonia. Szepesi et al. [2] proposed a new CNN model based on dropout layers in convolutional blocks. Unlike classical architectures where dropout is usually applied only in fully connected layers, this innovative placement improved generalization and reduced overfitting and achieved impressive 97.21% accuracy on 5,856 chest X-ray images. The dataset is taken from the Kaggle medical imaging competition and contains labeled images of pediatric patients, suitable for clinical applications.

The proposed architecture outperformed the transfer learning models, namely InceptionV3, ResNet-50, and VGG-19, in both accuracy and efficiency. Moreover, with fewer parameters, this model is lighter and thereby resulted in less training and inference times than its counterparts, which is valuable in resource-constrained healthcare settings, especially when computational resources are scarce. The work done by Szepesi et al. [2] emphasizes the design of architectures specific to tasks.

Pramoth et al. [3] have designed a CNN-based model which is aimed to diagnose pneumonia from chest X-rays. The experiment has utilized the Kaggle Chest X-ray Pneumonia dataset with 5,863 labeled images to train and evaluate the model. In testing multiple deep learning algorithms such as ResNet, DenseNet, VGG16, and GoogLeNet, CNN has obtained the highest accuracy of prediction at 91.98

The research [3] highlights the integration of AI-driven diagnostic tools into mobile applications for wider accessibility. By converting the trained model into TensorFlow Lite (TFLite), the authors ensured compatibility with mobile devices, enabling real-time inference without the need for backend servers. This approach highlights the practical application of deep learning models in resource-constrained environments, facilitating timely and cost-effective diagnosis. This emphasizes the potential of AI-based solutions to enhance early detection and intervention for pneumonia, especially in underserved regions, with efficient modeling combined with mobile deployment.

Deep learning has played a very important role in the development of pneumonia detection. Kaushik et al. [4] introduced CNN models for the accurate detection of pneumonia using the Chest X-Ray Images (Pneumonia) dataset from Kaggle. The research discussed four CNN architectures with different complexities and achieved validation accuracies of 89.74%, 85.26%, 92.31%, and 91.67%, respectively. The third model with three convolutional layers outperformed all the other models with a recall of 98% and an F1 score of 94%.

They are trained from scratch with no transfer learning, maximizing characteristics in the specific dataset used. Dropout regularization and changing the learning rate helped limit overfitting. The authors further state the importance of recall, considering medical diagnostics, to limit false negatives-that is, so as not to label a minority diseased population as healthy.

Kaushik et al. [4] have highlighted the possibility of using custom CNN models for fast and accurate pneumonia detection, which can be a valuable aid to healthcare professionals. Future work proposed includes the exploration of transfer learning and larger datasets for improved performance.

Pneumonia detection from chest X-rays has benefited greatly from advancements in deep learning. Mabrouk et al. [5] proposed an ensemble learning approach combining three pre-trained models—DenseNet169, MobileNetV2, and Vision Transformer (ViT)—to enhance classification performance. By employing transfer learning and fine-tuning on the Chest X-ray dataset, their method achieved an accuracy of 93.91% and an F1-score of 93.88%, outperforming many state-of-the-art methods. The ensemble technique effectively leverages the strengths of CNNs and transformers, demonstrating robust results for automated pneumonia detection.

Zhang et al. [6] proposed a lightweight CNN model that utilizes the VGG architecture along with dynamic histogram enhancement (DHE) for preprocessing images. Their model achieved a significant reduction in the number of parameters and had a high accuracy of 96.07% with an AUC of 0.9911. The application of DHE enhanced the contrast of images and made chest X-rays clearer to improve the detection of pneumonia. This study is indicative of the need for preprocessing techniques and simple model designs in resource-poor environments.

Sahin et al. [7] proposed a deep learning framework for pneumonia detection using chest X-rays. Their study evaluated four CNN models: MobileNet, ResNet, AlexNet, and a custom CNN model. Among these, the ResNet model achieved the highest accuracy of 97% on the Kaggle Chest X-ray Pneumonia dataset, demonstrating its robustness in extracting complex features from medical images. The study highlighted the importance of comparing multiple architectures to identify the most suitable model for pneumonia detection tasks.

Shahzad et al. [8] investigated the performance of pre-trained network architectures such as VGG-16, VGG-19, DenseNet-121, ResNet-50, and InceptionV3 for classifying pneumonia in chest X-rays. Their results showed that VGG-16 outperformed the other architectures, achieving a test accuracy of 90% and validation accuracy of 93.98%. This study emphasized the significance of fine-tuning pre-trained models to improve their applicability in medical imaging tasks.

Barhoom and Abu Naser [9] focused on developing a CNN-based model for diagnosing pneumonia. Their work aimed to classify chest X-rays into bacterial pneumonia, viral pneumonia, and normal categories. They demonstrated the potential of deep learning in automating diagnostic processes, reducing human error, and aiding medical professionals in making accurate decisions. This research highlighted the role of CNNs in achieving rapid and precise pneumonia detection.

Recent advancements in deep learning have contributed significantly to automated pneumonia detection. Al-Dulaimi et al. [10] proposed a CNN-based model for pneumonia detection from chest X-rays. The framework achieved outstanding results, with precision, recall, and accuracy values of 98%, 98%, and 99.82%, respectively, outperforming models such as ResNet-50 and VGG16. The system demonstrated a balance between accuracy and computational efficiency, making it suitable for healthcare applications.

Pant et al. [11] presented an ensemble approach combining ResNet-34 and EfficientNet-B4-based U-Nets. The model addressed class imbalance through data augmentation and used loss functions like Dice Loss and Binary Cross-Entropy to enhance performance. This ensemble achieved high recall and precision, effectively reducing false negatives and false positives in pneumonia detection.

Alapat et al. [12] reviewed various CNN architectures applied to the Chest X-ray14 dataset. Their analysis highlighted that models like DenseNet121 and InceptionV3 provided superior results with AUC values exceeding 0.98. Preprocessing techniques, such as dynamic histogram equalization, and ensemble models were noted as key contributors to performance improvements.

Maselli et al. [13] proposed a hierarchical convolutional model with Mobile Net and DenseNet121 which improves the diagnosis of pneumonia through chest X-ray imaging. After that, they used features from these pre-trained networks to learn with a stacked model getting an accuracy of 93.5 percent on the test. To achieve improved diagnostic precision we introduced a second stage classification model to discriminate interstitial and lobar pneumonia. The research showed promise in the ability to combine models for robust and scalable solutions in the clinical setting.

To detect pneumonia from chest X-rays, Yi et al. [14] introduced a deep convolutional neural network (DCNN) with 52 convolutional layers. Data consisting of 5,856 images, was used to train and validate the model with a validation accuracy of 96.09 percent. Preprocessing techniques, such as CLAHE and data augmentation, were applied to the framework's input in order to enhance input quality and cope with a class imbalance. Also, the proposed DCNN successfully achieved higher accuracy and better robustness than existing methods.

Based on chest X-rays, Verma and Prakash [15] proposed a pneumonia classification model using chest X-rays from scratch. They used data augmentation techniques to make the training process better which also happened to have better validation and training accuracy. The model was able to achieve quite high classification accuracy by leveraging CNNs ability to extract features autonomously. The resulting effectiveness of tailored neural network architectures and preprocessing techniques for medical imaging tasks is highlighted.

Q. An et al. [16] Research Paper Summary in which the authors highlight advancement in deep learning, especially convolutional neural networks (CNNs), for medical imaging related to pneumonia detection. Some of the issues that it addresses include data imbalance, a lack of interpretability, and suboptimal extraction of features. The author reviews some of the existing models, such as VGG16, ResNet, and DenseNet, focusing on their strengths and weaknesses. The paper focuses on novelty in attention mechanisms and approaches of feature fusion to elevate diagnostic accuracy and points out voids in the integration into models currently in use.

A deep learning based ensemble model for COVID 19 and pneumonia detection and segmenting was proposed by Hasan et al. [17] Using DenseNet121, EfficientNetB0, and VGG19 to classify images to COVID-19, pneumonia or normal categories with an accuracy of 99.2 percent.. A U-Net model with DenseNet103 as a backend was used for segmentation, giving a Dice coefficient of 0.92 and an IoU score of 0.90. The utility of Grad-CAM to generate ground truth masks with which to analyze infected lungs was demonstrated through this method.

The ResNet-18 CNN model is used to classify chest x rays into pneumonia, COVID-19 or normal cases by Mishra et al. [18] The model was trained with a training accuracy of 98.12 percent, a test accuracy of 97.7 percent by leveraging data augmentation techniques and preprocessed datasets. They showed how transfer learning can overcome overfitting in smaller datasets and how lightweight models like ResNet-18 can be robust for classification. The operational challenges in the clinical environments were further addressed by this research by proposing an efficient and low cost framework.

For the problem of dementia detection using medical images such as chest X-ray images, Hashmi et al. [19] suggested a compound scaled version of the ResNet50 model. To cope with the limited dataset size issue, the model was trained with transfer learning and augmented with data augmentation, and achieved a test accuracy of 98.14% and an AUC score of 99.71%. By using compound scaling method of balanced network dimensions (depth, width, and resolution), it greatly enhanced efficiency and performance. Interpretability was conducted with X-rays using Class Activation Maps (CAMs) to highlight regions of interest. Furthermore, this approach showed that computationally efficient deep learning architectures have the potential to aid clinical decision making processes.

Narayanan et al. [20] proposed a two-stage deep learning model for enhancing the detection and classification of pneumonia from chest X-rays. In the first stage, it classifies images as either normal or pneumonic by using transfer learning on networks such as AlexNet, ResNet, and VGG16. In the second stage, pneumonia is categorized as bacterial or viral using a simpler CNN model to counter overfitting. U-Net-based lung segmentation was integrated for better precision in diagnosis. The model was able to achieve an AUC of 0.996 in detecting pneumonia and accuracy of 97.8% in classifying it as either bacterial or viral pneumonia, which serves as a benchmark for automated radiology tools.

Mohamed et al. [21] proposed a framework with Conditional Generative Adversarial Networks (CGAN) to augment imbalanced datasets and fine-tune lightweight deep transfer learning models like Xception and MobileNet. The study obtained 99.26% accuracy using Xception for pediatric pneumonia detection in chest X-rays. The CGAN integration effectively mitigated overfitting and data scarcity, enhancing diagnostic performance in resource-constrained settings.

Ahmad et al. [22] compared the performance of different CNN architectures, including DenseNet121, VGG16, and InceptionResNetV2, on the task of identifying bacterial and viral pneumonia using X-rays. In this set, InceptionResNetV2 achieved the highest accuracy with 98.33%. Their work underscores the benefits of pre-trained models along with transfer learning to better medical image classification by leveraging data availability.

Rangasamy et al. [23] emphasized pneumonia as the major cause of death, especially in developing areas, which requires traditional diagnostic procedures such as blood tests, sputum culture, and CXR, relying on radiologists who may have errors in human judgment and availability. To overcome these limitations, they explored automated deep learning approaches, particularly Convolutional Neural Networks (CNNs), which have shown great promise in delivering accurate, rapid, and objective diagnoses. Models such as CheXNet, AlexNet, GoogleNet, and ResNet-50 demonstrated high accuracy in detecting pneumonia, with some outperforming radiologists. However, while these models excel at distinguishing pneumonia from healthy lungs, challenges remain in differentiating bacterial from viral pneumonia due to overlapping radiographic features, emphasizing the need for hybrid models and enhanced segmentation techniques for improved classification accuracy and broader clinical applicability.

In the thesis by Rahman Priya et al. [24], they created a machine learning framework to detect pneumonia using chest X-ray images. They used both their customized CNN models and the pre-trained architectures, including VGG-16, VGG-19, InceptionV3, and ResNet50. Their customized CNN model reported the highest accuracy of 90.43% with an F1-score of 0.87. It highlighted the utilization of machine learning in automating the detection of pneumonia and compared different models for performance, thus offering a pathway to improvement in diagnostic reliability and efficiency.

Barakat et al. [25] proposed a machine learning-based approach for the detection of pediatric pneumonia from chest X-rays. The authors focused on the computational efficiency and interpretability of machine learning (ML) models compared to deep learning (DL). The authors applied data augmentation and optimized statistical feature extraction to train ML models such as quadratic SVM, which achieved an accuracy of 97.58%. Their approach showed reduced training time and high accuracy, both the limitations DL models face in terms of clinical usability, especially in low-resource environments.

2 Detailed Methodology

The development of our pneumonia detection system involved a systematic approach divided into several stages. Below, we outline each stage in detail:

2.1 Dataset Preparation

The dataset used for this project was the *Chest X-ray Pneumonia Dataset* obtained from Kaggle. The dataset consists of high-resolution chest X-ray images labeled as "Pneumonia" and "Normal." It was pre-split into training, validation, and testing sets:

- **Training Set:** Used for model training.
- **Validation Set:** Used to tune hyperparameters and prevent overfitting.
- **Testing Set:** Used for evaluating final model performance.

To ensure model generalizability, the dataset was preprocessed using the following steps:

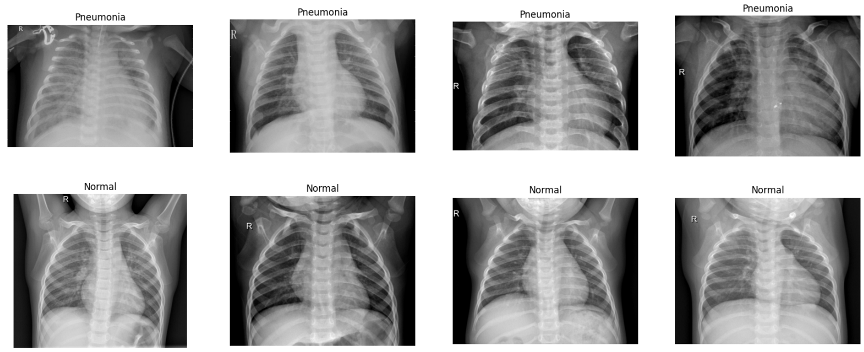

Fig 1. Normal and Pneumonia affected Lungs X-rays.

- **Normalization:** Pixel intensity values were scaled to a range of [0, 1].
- **Data Augmentation:** Techniques such as rotation, flipping, and zooming were applied to artificially expand the dataset and reduce overfitting.

Table 1. Dataset Description for Chest X-Ray Pneumonia Dataset

Dataset Name	Chest X-Ray Images (Pneumonia)
Source	Mooney PT. Chest X-Ray Images (Pneumonia). Available: Kaggle
Study Objective	Binary classification of chest X-ray images to identify the presence of pneumonia.
Data Composition	- **Normal:** Images of healthy lungs - **Pneumonia:** Images showing signs of pneumonia (subdivided into bacterial and viral cases)
Dataset Organization	- **Training set:** Images used for model training (4,272 images) - **Validation set:** Images for model tuning (1,168 images) - **Test set:** Images for evaluating model performance (423 images)
Image Specifications	- Format: JPEG - Resolution: Varying dimensions, primarily grayscale
Data Labeling	Binary labeling (Normal or Pneumonia) based on expert medical diagnosis.

2.2 Model Architectures

We implemented three distinct CNN architectures for comparative analysis:

- **Custom CNN:** A tailored architecture as shown in fig 2 consisting of convolutional, pooling, and fully connected layers, optimized for the given dataset. Key features include:
  - Eight convolutional layers with ReLU activation.
  - Max-pooling layers to downsample feature maps.
  - A fully connected dense layer for final classification.

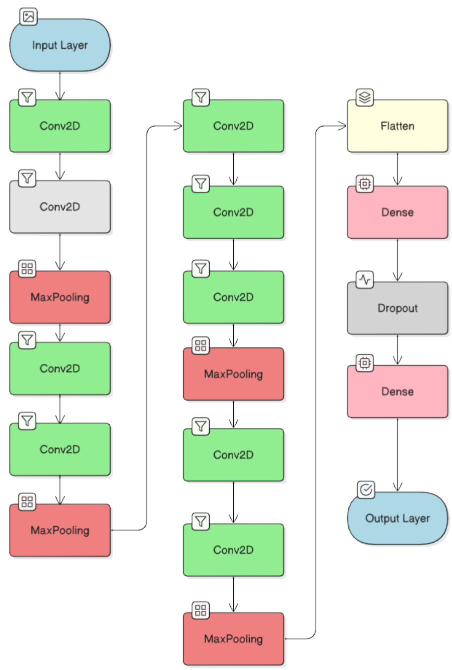

Fig 2. Convolutional Neural Network (CNN) architecture for binary classification with convolutional layers, max-pooling, dense layers, and a softmax output.

- **VGG16:** A pre-trained model with 16 layers, including 13 convolutional layers and 3 fully connected layers. The final dense layers were fine-tuned to adapt to the binary classification task.

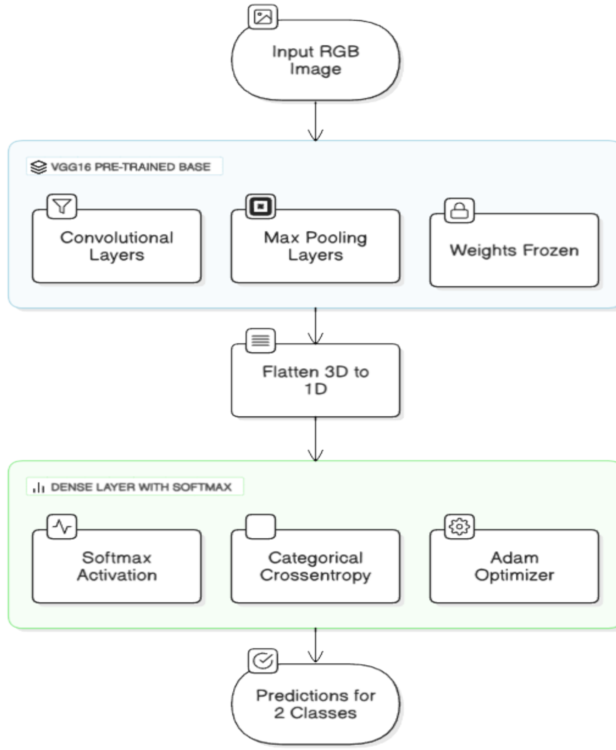

Fig 3. VGG-16 model architecture with pre-trained weights, used for image classification with additional dense layers for binary classification.

- **ResNet-50:** A deep residual network with skip connections to mitigate vanishing gradient issues. This model leverages residual learning to improve accuracy on deep architectures.

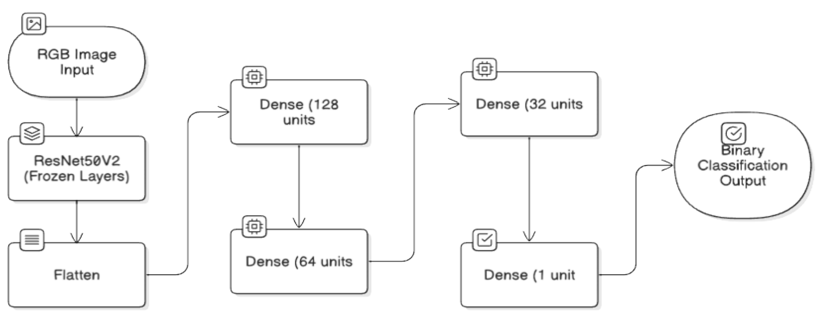

Fig 4. ResNet50V2 model for binary classification using residual connections, convolutional layers, and a sigmoid output.

2.3 Training Procedure

The training process was conducted using TensorFlow and Keras frameworks. The following steps were performed for each model:

1. Imported necessary libraries and configured the computational environment.
2. Preprocessed the dataset using normalization and augmentation techniques.

3. Initialized model architectures with pre-defined parameters.
4. Compiled models using the Adam optimizer, binary cross-entropy loss function, and accuracy as the evaluation metric.
5. Trained models over 50 epochs with a batch size of 32, monitoring training and validation performance.

2.4 Evaluation Metrics

Each model's performance was evaluated using the following metrics:

- **Accuracy:** The ratio of correctly predicted instances to the total instances.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

- **Precision:** The ratio of true positive predictions to the total predicted positives.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

- **Recall (Sensitivity):** The ratio of true positives to the total actual positives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

- **Confusion Matrix:** A visual representation of true positives, true negatives, false positives, and false negatives.

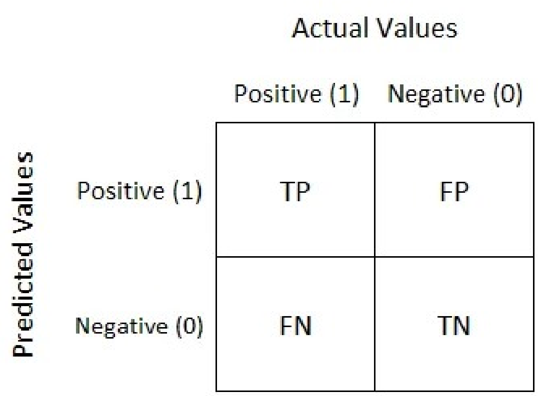

Fig 5. Template of a confusion matrix for evaluating model performance, showing true positives, false positives, true negatives, and false negatives.

2.5 Hyperparameter Tuning

Hyperparameters such as learning rate, batch size, and the number of epochs were fine-tuned using the validation set. The final optimal configuration was selected based on validation accuracy and loss trends.

2.6 Confusion Matrices and Results Visualization

For each model, confusion matrices were generated to analyze the classification performance. The matrices provide insights into true positives, true negatives, false positives, and false negatives. These results were visualized for better interpretability.

3 Results

In this section, we present the detailed outcomes of our experiments, including the performance evaluation of the models, graphical analysis of training and testing accuracy/loss, and confusion matrices for classification performance.

3.1 Training and Testing Accuracy

The training and testing accuracy for each of the models were tracked across multiple epochs. The goal was to ensure that the models generalized well and avoided overfitting. Figure 6 shows the training and testing accuracy trends over the epochs for each model:

- **Custom CNN:** Achieved a maximum training accuracy of *93.5%* and testing accuracy of *93.75%*.
- **VGG16:** Achieved a maximum training accuracy of *94%* and testing accuracy of *84.9%*.
- **ResNet-50:** Achieved a maximum training accuracy of *95.4%* and testing accuracy of *82.6%*.

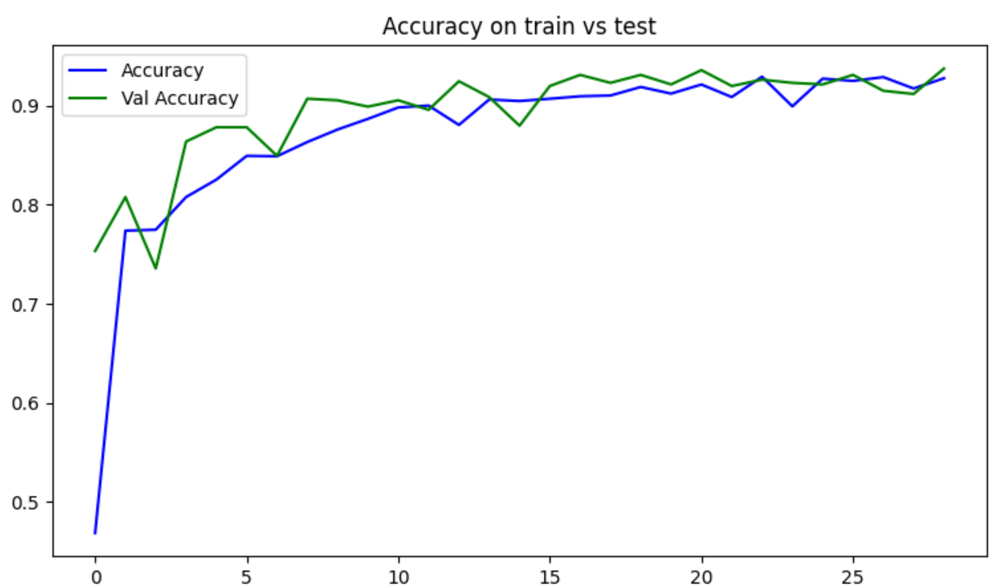

Fig 6. Training vs. Testing Accuracy for Custom CNN.

3.2 Training and Testing Loss

Figure 6 illustrates the trends of training and testing loss for our custom CNN model. A steady decrease in loss values and convergence between training and testing curves indicates effective learning and minimal over fitting:

- **Custom CNN:** Loss reduced to *0.1175* on training and *0.3481* on testing.
- **VGG16:** Loss reduced to *0.1478* on training and *0.4353* on testing.
- **ResNet-50:** Loss reduced to *0.1392* on training and *0.8625* on testing.

3.3 Confusion Matrices

The confusion matrices generated for the testing phase provide a detailed breakdown of the classification performance for each model. These matrices illustrate the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN):

- **Custom CNN:** The confusion matrix revealed a high TP rate of 96% and low FP rate of 9.8%.
- **VGG16:** Demonstrated a slightly higher FN rate compared to other models.
- **ResNet-50:** Showed the most balanced classification results with minimal FP and FN.

Figure 7 shows the confusion matrix for the best-performing model (Custom CNN with 8 layers).

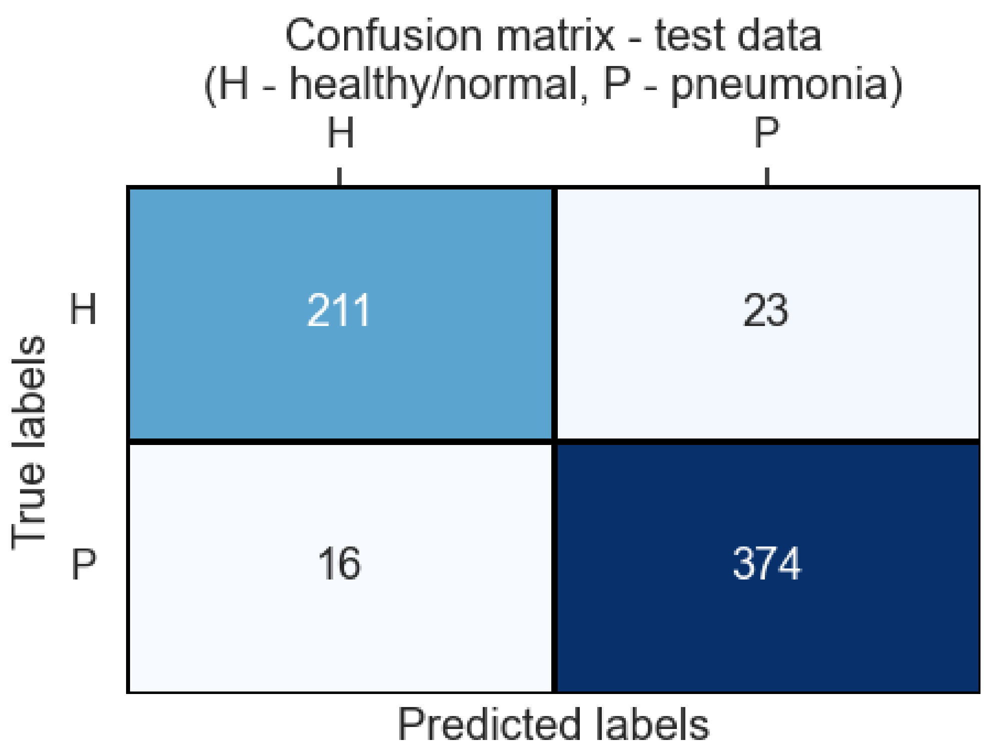

Fig 7. Confusion Matrix for Custom CNN during the Testing Phase.

3.4 Performance Metrics

The models were evaluated using accuracy, precision, recall, and F1-score. Table 2 summarizes these metrics for all three models:

Table 2. Performance Metrics for the Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Custom CNN	93.75	94.21	95.90	95.04
VGG16	84.9	71.2	83.4	76.9
ResNet-50	82.69	70.9	83.2	76.6

3.5 Observations and Insights

- The Custom CNN presented superior performance, proving that the architecture could be both highly effective and computationally efficient while it showed better performance than other models in all evaluation metrics.
- VGG16 produced good results, but it showed minor deficiencies.
- The least competitive was ResNet-50, unable to achieve such good results mainly because its architecture was much deeper and, thus, may have been inappropriate for the used dataset.

These results show the potential of CNN-based architectures for pneumonia detection from chest X-rays with high reliability. Further work could optimize these models by using larger and more diverse datasets or advanced architectures such as vision transformers.

4 Comparative Analysis

This section provides a detailed comparative analysis of the performance of the three implemented models: Custom CNN, VGG16, and ResNet-50. Each model was evaluated based on accuracy, loss, precision, recall, F1-score, and overall robustness in detecting pneumonia from chest X-ray images.

4.1 Model Accuracy and Loss

The accuracy and loss trends for the Custom CNN model are illustrated in Figure 6. These metrics provide a clear understanding of the models' convergence during training and their generalization to unseen data:

- **Custom CNN:** Showed steady improvement in accuracy over the epochs, with a final testing accuracy of *93.75%*. The training and testing loss curves were well-aligned, indicating minimal over-fitting.
- **VGG16:** Achieved high training accuracy but exhibited a noticeable gap between training and testing performance, suggesting overfitting to the training data.
- **ResNet-50:** Demonstrated consistent performance with the highest testing accuracy (*82.69%*).

4.2 Confusion Matrix Analysis

- **Accuracy:** Custom CNN achieved the highest accuracy, followed by VGG16 and ResNet-50.
- **Precision:** ResNet-50 maintained the highest precision, indicating fewer false positives.

- **Recall:** Custom CNN and ResNet-50 showed comparable recall, while VGG16 lagged due to higher false negatives.

4.3 Computational Efficiency

The computational efficiency of the models was assessed based on training time and memory requirements:

- **Custom CNN:** Required the least computational resources and training time, making it suitable for resource-constrained environments.
- **VGG16:** Pre-trained weights reduced initial training time, but fine-tuning increased resource demands.
- **ResNet-50:** Despite its depth, optimized residual connections made it computationally efficient relative to its performance.

4.4 Robustness and Generalization

Custom CNN was the best-performing model, demonstrating strong adaptability and efficiency. VGG16 followed, showing robust performance but with slight overfitting. ResNet-50, although effective, performed slightly below the other two models in terms of accuracy.

4.5 Overall Insights

Based on the comparative analysis:

- Custom CNN emerged as the best-performing model, excelling in both accuracy and computational efficiency, making it ideal for scenarios that require high performance and scalability..
- VGG16, while effective, requires additional regularization techniques to improve its generalization capabilities and is more suited for applications where moderate accuracy is acceptable.
- ResNet-50 showed strong performance, though it is more computationally intensive compared to the Custom CNN and VGG16, making it the better choice for applications prioritizing robustness and accuracy over computational efficiency.

These findings highlight the trade-offs between model complexity, accuracy, and computational requirements, providing valuable insights for real-world deployment.

5 Discussion

The results obtained from this study underscore the potential of deep learning models, particularly Convolutional Neural Networks (CNNs), in automating the detection of pneumonia from chest X-ray images. The discussion below elaborates on the key findings, their implications, and the challenges faced during the project.

5.1 Key Findings

The comparative analysis of the models—Custom CNN, VGG16, and ResNet-50—revealed several critical insights:

- **Custom CNN:** The custom CNN achieved the best accuracy and computation efficiency. It was thus the best and most viable model for use in the detection of pneumonia, especially in resource-poor environments.
- **VGG16:** VGG16 showed good performance, although it was not the best. It demonstrated strong accuracy, benefiting from its pre-trained advantage, and did not exhibit signs of overfitting. However, it did not reach the same level of performance as the Custom CNN.
- **Model Performance:** The least effective model was ResNet-50, as it could not achieve competitive results. Although the residual learning helped to overcome vanishing gradient issues, the deeper architecture of this model did not work well for this task.

5.2 Implications of the Study

The study has several implications for both research and practical applications:

- **Improving Diagnostic Accuracy:** The accuracy of the models, especially the Custom CNN, indicates that they are capable of helping radiologists make diagnoses with minimal errors.
- **Reducing Diagnostic Time:** The automated detection systems significantly reduce the time required for diagnosis, thereby enabling quicker medical intervention and better patient outcomes.
- **Resource Optimization:** Deployment of deep learning-based tools in healthcare facilities, especially in low-resource settings, can reduce dependency on skilled radiologists and alleviate bottlenecks in diagnosis.
- **Scalability:** The ability to scale and integrate these models into telemedicine platforms can open up access to diagnostic services in remote or underserved regions.

5.3 Challenges and Limitations

Despite the promising results, the study faced several challenges and limitations:

- **Data Imbalance:** The dataset contained a higher proportion of pneumonia cases compared to normal cases, potentially leading to biased model predictions. Data augmentation techniques were employed to address this issue, but a more balanced dataset would improve model fairness.
- **Overfitting in VGG16:** The pre-trained VGG16 model exhibited overfitting, highlighting the need for further fine-tuning and regularization to improve generalization.
- **Dataset Limitations:** The dataset used in this experiment was limited in its demographic and clinical diversity. Adding more diverse populations and pneumonia severities to the dataset would enhance the models' generalizability.

5.4 Future Directions

Building on the findings and addressing the challenges identified, several avenues for future research are proposed:

- **Expanding the Dataset:** Gather and clean larger, more diverse datasets to increase models' robustness and generality, thus ensuring improvements in performance across different populations and severities of pneumonia.
- **Advanced Architectures:** Explore state-of-the-art architectures, such as Vision Transformers or hybrid CNN-RNN models, to further exploit feature extraction and classification above what current models can.
- **Edge Optimization:** Developing lightweight editions of the Custom CNN, or other high-performance model, to be deployed directly on edge devices or platforms, thereby scaling well under resource-constrained environments.
- **Explainability:** Designing the integration of some XAI techniques that include visual explanations of the reasons behind the model's predictive output, thereby increasing levels of transparency and building users' trust among healthcare users.
- **Clinical Integration:** Collaborating with healthcare providers to test the models in real-world clinical settings to be able to integrate them seamlessly into existing diagnostic workflows for practical use.

5.5 Broader Impact

The implementation of deep learning models in medical diagnostics has the potential to revolutionize healthcare delivery. By automating the detection of pneumonia, this study contributes to the growing body of research aimed at leveraging artificial intelligence for improving public health outcomes. However, careful consideration of ethical and regulatory challenges is crucial to ensure responsible deployment.

5.6 Conclusion of the Discussion

Generally, the study highlights the potential and limitation of CNN-based models in pneumonia detection. The best model that comes out from this experiment was the Custom CNN, performing quite competitively with much lesser computational requirement, which has a very good chance for deployment in resource-constrained environments. VGG16 performed pretty well but also showed some performance gap towards Custom CNN and could benefit by other kinds of regularization techniques for achieving better generalization. ResNet-50 was relatively more computationally expensive; however, the potential in such applications could be constrained. Future work in extending the dataset, model optimization, and clinical validation will help translate these findings into practical healthcare solutions.

6 Conclusion

This study explored the application of deep learning models, specifically Convolutional Neural Networks (CNNs), for automated pneumonia detection using chest X-ray images. By implementing and comparing three architectures—Custom CNN, VGG16, and ResNet-50—we demonstrated the potential of these models to enhance diagnostic accuracy and efficiency in medical imaging.

Key findings from the study include:

- **Custom CNN:** The highest-performing model, delivering superior accuracy and robustness while being computationally efficient, making it ideal for resource-constrained settings.
- **VGG16:** powerful pre-trained model, although susceptible to overfitting, which highlights the need for additional regularization techniques to improve generalization
- **ResNet-50:** A highly accurate and robust model, leveraging residual learning to address vanishing gradient problems, though it is more computationally intensive.

The implications of this research are significant, particularly for healthcare delivery in low-resource settings. By automating pneumonia detection, these models can reduce the dependency on skilled radiologists, expedite diagnosis, and improve patient outcomes. Furthermore, the scalability of these systems enables their integration into telemedicine platforms, extending healthcare access to remote and underserved regions.

However, challenges such as dataset limitations, computational requirements, and overfitting in certain models underscore the need for further research. Expanding datasets to include diverse demographics, optimizing models for edge devices, and incorporating explainable AI (XAI) techniques are promising directions for future work.

In conclusion, the integration of deep learning models into diagnostic workflows represents a transformative step toward leveraging artificial intelligence in healthcare. While there is still much progress to be made, this study demonstrates the feasibility and potential impact of CNN-based systems in addressing critical global health challenges like pneumonia detection.

Dataset and Code Availability

The dataset used in this project is the **Chest X-ray Pneumonia Dataset**, which was obtained from Kaggle [1]. The dataset contains high-resolution chest X-ray images labeled as "Pneumonia" and "Normal," and is pre-split into training, validation, and testing sets as follows:

- **Training Set**: Used for model training.
- **Validation Set**: Used for hyperparameter tuning and preventing overfitting.
- **Testing Set**: Used for evaluating the final model performance.

To enhance model generalizability, the dataset underwent preprocessing, which included pixel intensity normalization (scaling values to a range of [0, 1]) and data augmentation techniques such as rotation, flipping, and zooming.

The code developed for the analysis is available on GitHub [27]. The repository includes all scripts necessary for data preprocessing, model training, evaluation, and visualization.

References

1. P. T. Mooney, "Chest X-Ray Images (Pneumonia)," *Kaggle*, 2018. Available: `https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia`.

2. P. Szepesi and L. Szilágyi, "Detection of pneumonia using convolutional neural networks and deep learning," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 5, pp. 1012–1022, 2022, doi: 10.1016/j.bbe.2022.08.001.

3. P. K. M., P. Ravi, S. Sajith, and D. M. R. S., "Pneumonia detection using CNN," in *Proceedings of the Conference on Pneumonia Detection*, May 2022, pp. 1–10. Available: https://www.researchgate.net/publication/360702927.

4. V. Sirish Kaushik, A. Nayyar, G. Kataria, and R. Jain, "Pneumonia Detection Using Convolutional Neural Networks (CNNs)," in *Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)*, Lecture Notes in Networks and Systems 121, Singapore: Springer, 2020, pp. 471–483, doi: 10.1007/978-981-15-3369-3-36.

5. A. Mabrouk, R. P. Díaz Redondo, A. Dahou, M. Abd Elaziz, and M. Kayed, "Pneumonia Detection on Chest X-ray Images Using Ensemble of Deep Convolutional Neural Networks," 2023. [Online]. Available: arXiv:2312.07965.

6. D. Zhang, F. Ren, Y. Li, L. Na, and Y. Ma, "Pneumonia Detection from Chest X-ray Images Based on Convolutional Neural Network," *Electronics*, vol. 10, no. 1512, pp. 1–17, 2021, doi: 10.3390/electronics10131512.

7. M. E. Sahin, H. Ulutas, and E. Yuce, "A deep learning approach for detecting pneumonia in chest X-rays," *European Journal of Science and Technology*, vol. 28, pp. 562–567, Nov. 2021, doi: 10.31590/ejosat.1009434.

8. A. Shahzad, M. A. Arshed, F. Liaquat, M. Tanveer, M. Hussain, and R. Alamdar, "Pneumonia classification from chest X-ray images using pre-trained network architectures," *VAWKUM Transactions on Computer Sciences*, vol. 10, no. 2, pp. 34–44, Dec. 2022.

9. A. M. A. Barhoom and S. S. Abu Naser, "Diagnosis of pneumonia using deep learning," *International Journal of Academic Engineering Research*, vol. 6, no. 2, pp. 48–68, Feb. 2022.

10. D. S. Al-Dulaimi, A. G. Mahmoud, N. M. Hassan, A. Alkhayyat, and S. A. Majeed, "Development of Pneumonia Disease Detection Model Based on Deep Learning Algorithm," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 2951168, pp. 1–10, 2022, doi: 10.1155/2022/2951168.

11. A. Pant, A. Jain, K. C. Nayak, D. Gandhi, and B. G. Prasad, "Pneumonia Detection: An Efficient Approach Using Deep Learning," in *Conference Proceedings on Pneumonia Detection Models*, Bengaluru, India, 2020, pp. 1–8.

12. D. J. Alapat, M. V. Menon, and S. Ashok, "A Review on Detection of Pneumonia in Chest X-ray Images Using Neural Networks," *Journal of Biomedical Physics and Engineering*, vol. 12, no. 6, pp. 551–558, 2022, doi: 10.31661/jbpe.v0i0.2202-1461.

13. G. Maselli, E. Bertamino, C. Capalbo, R. Mancini, G. B. Orsi, and C. Napoli, "Hierarchical convolutional models for automatic pneumonia diagnosis based on X-ray images: new strategies in public health," *Annali di Igiene: Medicina Preventiva e di Comunità*, vol. 33, no. 6, pp. 644–655, 2021, doi: 10.7416/ai.2021.2467.

14. R. Yi, L. Tang, Y. Tian, J. Liu, and Z. Wu, "Identification and classification of pneumonia disease using a deep learning-based intelligent computational framework," *Neural Computing and Applications*, vol. 35, pp. 14473–14486, 2023, doi: 10.1007/s00521-021-06102-7.

15. G. Verma and S. Prakash, "Pneumonia Classification using Deep Learning in Healthcare," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 4, pp. 1715–1719, 2020, doi: 10.35940/ijitee.D1599.029420.

16. Q. An, W. Chen, and W. Shao, "A Deep Convolutional Neural Network for Pneumonia Detection in X-ray Images with Attention Ensemble," *Diagnostics*, vol. 14, no. 4, p. 390, Feb. 2024, doi: 10.3390/diagnostics14040390.

17. M. J. Hasan, M. S. Alom, and M. S. Ali, "Deep Learning Based Detection and Segmentation of COVID-19 and Pneumonia on Chest X-ray Images," in Proc. 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, Feb. 2021, pp. 210–213, doi: 10.1109/ICICT4SD50815.2021.9396878.

18. A. K. Mishra, A. Ghimire, M. Joshi, S. Khatiwada, N. R. Joshi, P. B. Jha, and P. S. Aithal, "Classification of Pneumonia and COVID-19 Using Convolutional Neural Networks," International Journal of Health Sciences and Pharmacy, vol. 7, no. 2, pp. 65–88, Oct. 2023, doi: 10.5281/zenodo.10001745.

19. M. F. Hashmi, S. Katiyar, A. W. Hashmi, and A. G. Keskar, "Pneumonia Detection in Chest X-ray Images Using Compound Scaled Deep Learning Model," *Automatika*, vol. 62, nos. 3–4, pp. 397–406, 2021, doi: 10.1080/00051144.2021.1973297.

20. B. Narayanan, V. S. P. Davuluru, and R. C. Hardie, "Two-stage deep learning architecture for pneumonia detection and its diagnosis in chest radiographs," in *Proc. SPIE Medical Imaging*, vol. 11318, 2020, pp. 1–10, doi: 10.1117/12.2547635.

21. C. Mohamed, R. W. Mwangi, and J. M. Kihoro, "Enhancing pneumonia detection in pediatric chest X-rays using CGAN-augmented datasets and lightweight deep transfer learning models," *Journal of Data Analysis and Information Processing*, vol. 12, no. 1, pp. 1–23, Jan. 2024, doi: 10.4236/jdaip.2024.121001.

22. N. Ahmad, I. A. Qadri, M. Naeem, F. Akbar, N. Ahmed, and S. I. Rai, "Effective detection of lung disease from X-ray images using convolutional neural networks (CNNs) and various architectures," *Journal of Population Therapeutics and Clinical Pharmacology*, vol. 30, no. 18, pp. 1975–1983, 2023, doi: 10.53555/jptcp.v30i18.3329.

23. K. Rangasamy, N. S. Mohd Fuzi, M. A. As'ari, N. A. Rahmad, and N. A. J. Sufri, "Deep learning-based fine-grained automated pneumonia detection model," *Journal of Engineering Science and Technology*, vol. 17, no. 4, pp. 2373–2389, 2022.

24. G. Maselli, E. Bertamino, C. Capalbo, R. Mancini, G. B. Orsi, and C. Napoli, "Hierarchical convolutional models for automatic pneumonia diagnosis based on X-ray images: new strategies in public health," *Annali di Igiene: Medicina Preventiva e di Comunità*, vol. 33, no. 6, pp. 644–655, 2021, doi: 10.7416/ai.2021.2467.

25. R. Yi, L. Tang, Y. Tian, J. Liu, and Z. Wu, "Identification and classification of pneumonia disease using a deep learning-based intelligent computational framework," *Neural Computing and Applications*, vol. 35, pp. 14473–14486, 2023, doi: 10.1007/s00521-021-06102-7.

26. G. Verma and S. Prakash, "Pneumonia Classification using Deep Learning in Healthcare," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 4, pp. 1715–1719, 2020, doi: 10.35940/ijitee.D1599.029420.

27. Hassan, S. (2024). Pneumonia Detection System. GitHub Repository. Available at: `https://github.com/Saad-Hassan25/Pneumonia-Detection-System`.