Pneumonia Diagnosis through Chest X-Ray

Associate Professor, Pramod kumar Naik

1 Associate Professor, Dayananda Sagar University, Bengaluru
Email: pramodnaik-cse@dsu.edu.in

Abstract:

Pneumonia, a life-threatening respiratory infection caused by bacteria or viruses, necessitates timely and accurate diagnosis to mitigate potential risks. This paper introduces an automated system utilizing digital X-ray images to detect bacterial and viral pneumonia promptly. The study explores recent advances in pneumonia detection and details the methodology, emphasizing the use of transfer learning with the VGG16 architecture.

The classification framework includes three pivotal schemes: normal vs. pneumonia, bacterial vs. viral pneumonia, and normal, bacterial, and viral pneumonia. The study employs VGG16 for transfer learning and fine-tuning, showcasing its effectiveness in enhancing classification performance.

While specific accuracies are not detailed here, the outcomes demonstrate the superiority of the proposed approach over existing benchmarks. This suggests the potential of the system to significantly aid radiologists in swift and accurate pneumonia diagnosis. Furthermore, the study positions the VGG16-based transfer learning approach as a valuable tool for efficient pneumonia patient screening, particularly in settings demanding rapid and reliable results, such as high-traffic environments like airports. This contribution holds promise for bolstering public health surveillance and response measures.

Keywords:

pneumonia; bacterial and viral pneumonia; chest X-ray; deep learning; transfer learning; image processing

1. Introduction:

Pneumonia remains a significant global health concern, contributing to a substantial number of child fatalities worldwide. With approximately 1.4 million children succumbing to pneumonia annually, accounting for 18% of deaths among children under five years old, there is an urgent need for advanced diagnostic tools and treatment strategies [1]. Pneumonia, characterized by lung infection caused by bacteria or viruses, is treatable with antibiotics and antiviral drugs. Timely diagnosis and accurate classification of the infectious agent (bacterial or viral) are crucial in preventing the progression of the disease and reducing mortality.

Current diagnostic methods primarily rely on chest X-rays, considered the gold standard for pneumonia detection [3]. However, the interpretation of X-ray images poses challenges, with potential misclassifications and subjective variations among radiologists. Moreover, the shortage of trained radiologists, particularly in low-resource countries and rural areas, emphasizes the pressing need for innovative solutions to enhance diagnostic accuracy.

Artificial Intelligence (AI), specifically deep learning techniques, has emerged as a promising tool in various biomedical applications, including pneumonia detection [7,8,9,10]. Convolutional Neural Networks (CNNs), a subset of deep learning, have demonstrated remarkable capabilities in image classification and feature extraction, making them widely adopted in medical imaging analysis. Transfer learning, a technique leveraging pre-trained models on large datasets, offers a solution for effective learning with limited labeled medical data.

Among pre-trained models, the Visual Geometry Group's 16-layer (VGG16) architecture stands out. Its deep architecture and successful application in diverse image classification tasks make it an attractive candidate for pneumonia detection. Fine-tuning, a process that adapts a pre-trained model to a specific task, further enhances the model's performance on medical image classification.

Despite significant advancements, distinguishing between bacterial and viral pneumonia using X-ray images remains a challenging task. This paper addresses this gap by proposing a transfer learning approach that utilizes pre-trained CNN models, including VGG16, and investigates fine-tuning strategies. The study aims to improve the accuracy of pneumonia classification, especially in distinguishing between bacterial and viral pneumonia, contributing to more effective and targeted medical interventions.

This paper proceeds with a discussion on the background of deep learning algorithms, emphasizing CNNs, transfer learning, and the significance of pre-trained models such as VGG16. Subsequently, the methodology, dataset details, and pre-processing steps for training and testing are elucidated. Results from the classification algorithms are presented and compared with recent studies. Finally, the conclusion summarizes the findings and highlights the potential impact of the proposed approach in advancing pneumonia diagnosis.

1.1 Background

The goal of this notebook is to use Convolutional Neural Networks on Chest X-Ray images to determine which samples are from patients with Pneumonia. In this dataset (version 3), there is one folder representing the train set and another one for the test set. The train folder is later split in the notebook into train/validation sets.

I use three different approaches for image classification: 1) A simple CNN, 2) Transfer Learning, using a pretrained model with frozen layers as the base for feature extraction and 3) Fine Tuning, unfreezing the last layers of the pretrained model

1.2 Role of Chest X-Rays

Chest X-rays are indispensable tools in pneumonia diagnosis, offering a non-invasive means to visualize internal lung structures. They enable the identification of key pneumonia indicators, including opacities, infiltrates, and consolidation patterns. This subsection elaborates on the unique role that chest X-rays play in the diagnostic landscape.

2. Challenges in Pneumonia Diagnosis

2.1 Ambiguities in Imaging

Interpreting chest X-rays for pneumonia encounters challenges due to the intricate nature of lung structures. The presence of opacities and consolidation patterns, while indicative of pneumonia, can also be associated with other respiratory conditions, introducing nuances that demand a nuanced diagnostic approach.

2.2 Variability in Patient Populations

Pneumonia manifests differently across diverse demographics and health conditions. This variability poses a significant challenge in devising universal diagnostic strategies. This section explores the impact of age, comorbidities, and other factors on the manifestation of pneumonia, emphasizing the need for tailored diagnostic approaches.

3. Methodology

Dataset:

This dataset contains 5,856 validated Chest X-Ray images. The images are split into a training set and a testing set of independent patients. Images are labeled as (disease:NORMAL/BACTERIA/VIRUS)-(randomized patient ID)-(image number of a patient).

Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care.

For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable scans. The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system. In order to account for any grading errors, the evaluation set was also checked by a third expert.

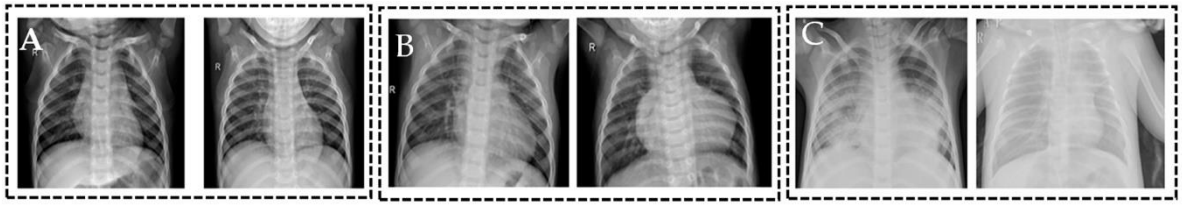


Figure 1. Data samples from the dataset, (**A**) shows normal cases, (**B**) shows bacterial pneumonia cases, and (**C**) shows viral pneumonia cases.

Data Exploration:

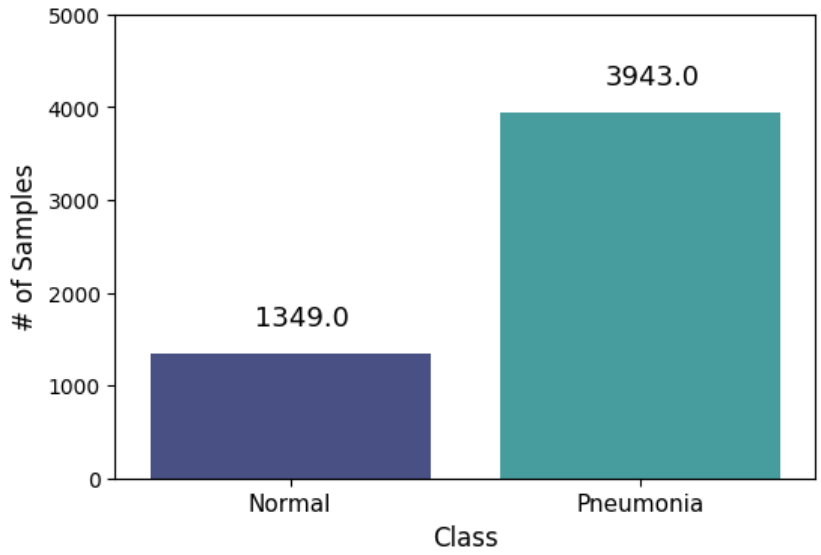


Figure 2. Classes and samples in the dataset

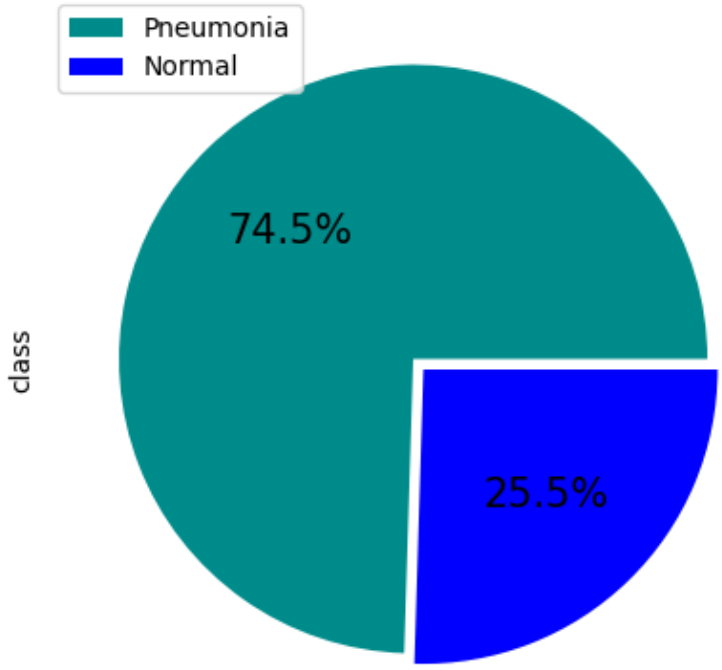


Figure 3. pie chart to show % of classes

Data visualization:

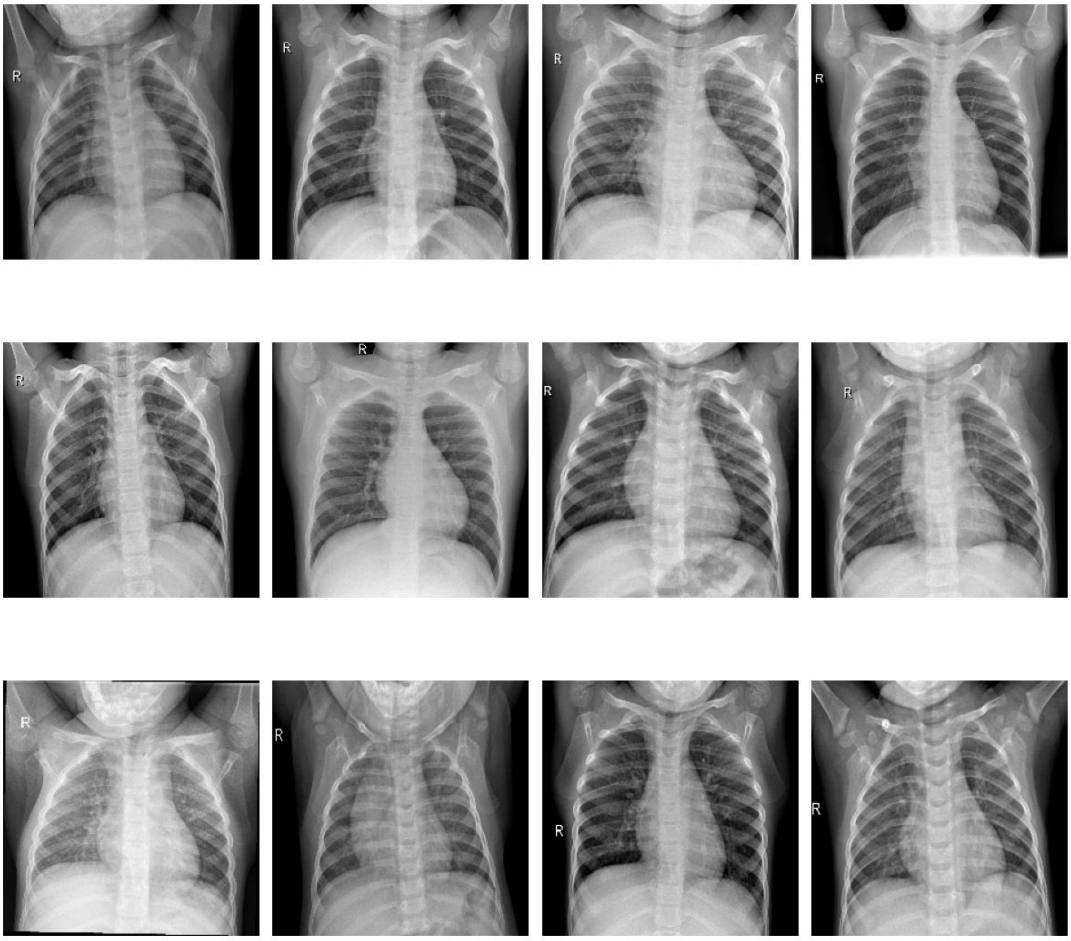


Figure 4. Normal x-ray images of the dataset

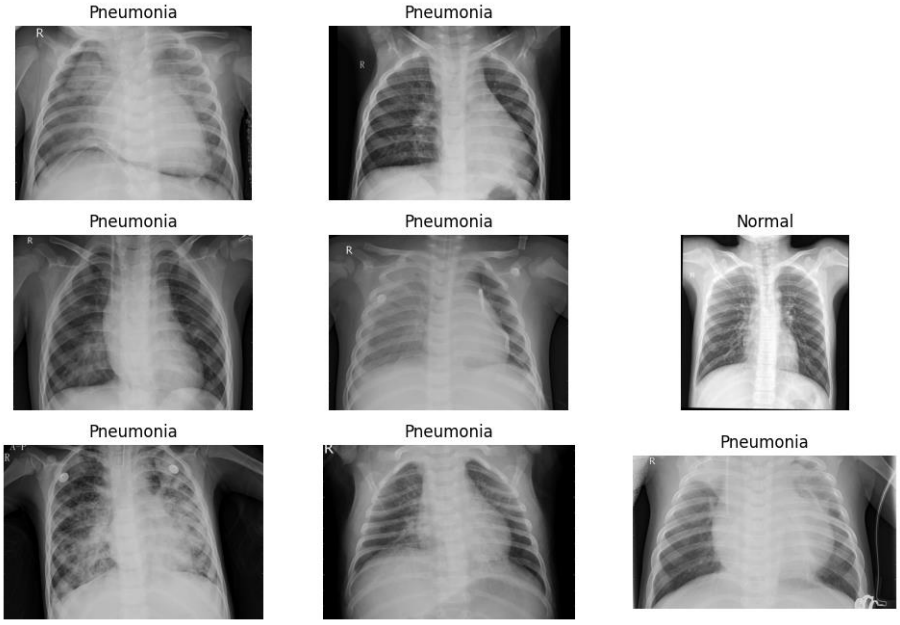


Figure 5. X-rays of both the classes(Normal, Pneumonia)

Preprocessing:

First, we need to create a validation set. To do that, we apply a simple stratified split on the original train dataset, using 80% for actual training and 20% for validation purposes.

Found 4233 validated image filenames belonging to 2 classes.

Found 1059 validated image filenames belonging to 2 classes.

We begin by defining the data generators. With Keras Image Data Generator, we can rescale the pixel values and apply random transformation techniques for data augmentation on the fly. We define two different generators. The val_datagen is used to simply rescale the validation and test sets. The train_datagen includes some transformations to augment the train set.

We apply those generators on each dataset using the flow_from_dataframe method. Apart from the transformations defined in each generator, the images are also resized based on the target_size set.

Data Augmentation

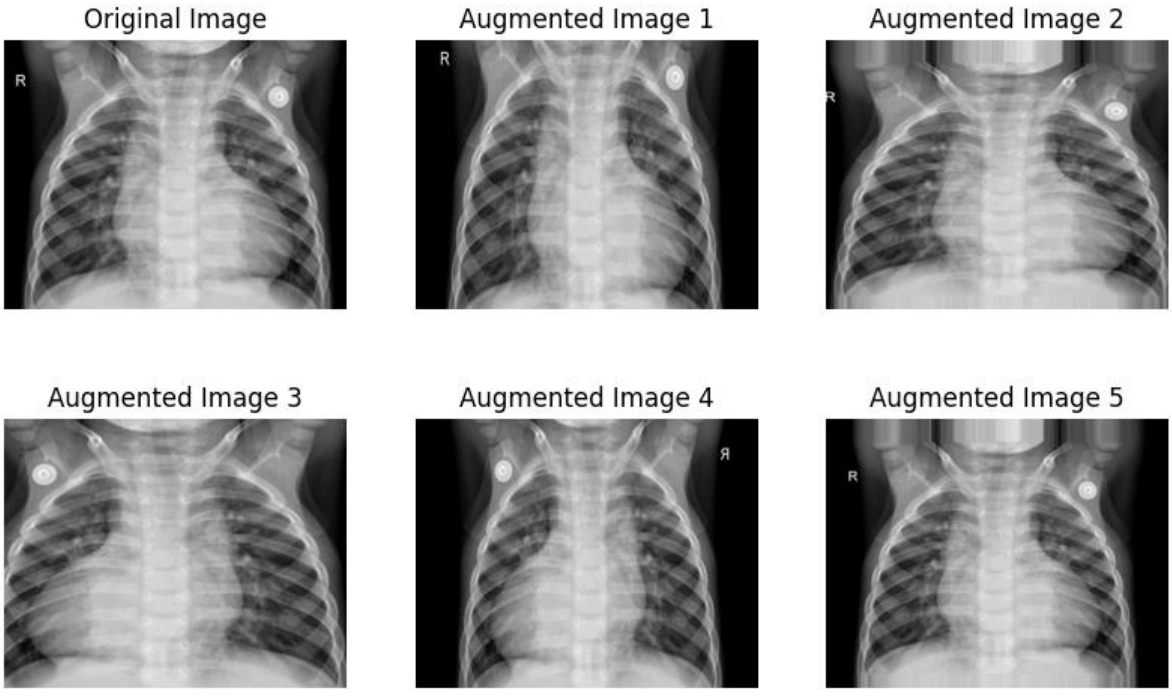
ImageDataGenerator: This class from Keras is used for data augmentation, including rescaling, shearing, zooming, and horizontal flipping. It helps in increasing the diversity of the training dataset.

flow_from_dataframe: Applies the data generators on the training, validation, and test sets.

train_test_split: Splits the original training dataset into training and validation sets, using 80% for training and 20% for validation.

The validation set is crucial for assessing the model's performance on unseen data.

Randomly chosen images from the training set are displayed along with their augmented versions to visualize the impact of data augmentation techniques.



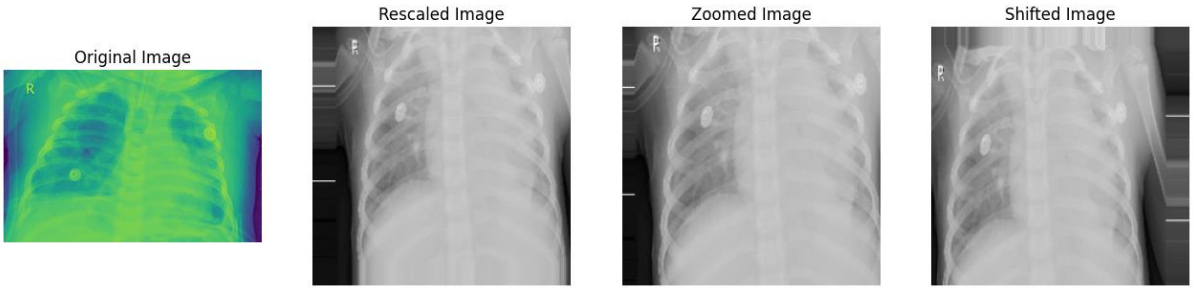


Figure 6. Augmented Images of Original Image

3. Techniques for Pneumonia Diagnosis

3.1 Feature Extraction

Feature extraction involves a meticulous process of identifying and isolating clinically relevant patterns in chest X-ray images. Beyond opacities and consolidation patterns, this section explores additional features such as texture, contour irregularities, and spatial distribution, providing a more comprehensive understanding of feature extraction methodologies.

3.2 Machine Learning Classifiers

Machine learning classifiers, including neural networks, support vector machines, and k-nearest neighbors, play a pivotal role in automating the diagnosis of pneumonia. This subsection expands on the intricacies of training these models on labeled datasets, addressing issues of generalization and overfitting, and exploring how they contribute to diagnostic accuracy.

3.3 Convolutional Neural Network (CNN) Model:

The Convolutional Neural Network (CNN) architecture defined in the code is a custom model designed for pneumonia detection. It consists of three convolutional blocks, each containing convolutional layers, batch normalization, ReLU activation, max-pooling, and dropout. The final layers include a dense layer with ReLU activation and dropout, leading to the output layer with a sigmoid activation for binary classification. The model is trained using binary crossentropy loss and the Adam optimizer, with monitoring and early stopping based on validation loss.

3.4 VGG16 Transfer Learning Model:

The VGG16 transfer learning model utilizes the pre-trained VGG16 architecture, excluding the top layer. This pre-trained VGG16 model is augmented with a custom head that includes a flattening layer, a dense layer with ReLU activation, dropout, and a final dense layer with a sigmoid activation for binary classification. The model is compiled with the Adam optimizer and binary crossentropy loss. It undergoes training on augmented data generated using ImageDataGenerator and is subsequently evaluated on a test set, providing a detailed classification report and confusion matrix.

3.5 ResNet152V2 Transfer Learning Model:

The ResNet152V2 transfer learning model employs the pre-trained ResNet152V2 architecture, excluding the top layer. Similar to the VGG16 model, it incorporates a custom head with global average pooling, dense layers, dropout, and a final dense layer with sigmoid activation. The model is compiled, trained on a dataset using data generators, and evaluated on both the validation and test sets. The training progress is monitored, and the learning curves, including loss and accuracy, are visualized.

3.6 Fine-Tuned ResNet152V2 Model:

In the fine-tuning section, layers of the pre-trained ResNet152V2 model are selectively unfrozen to allow for fine-tuning on the specific task. The code reveals that all layers except the last 13 are frozen. This fine-tuning approach provides flexibility in adjusting the pre-trained model to better suit the pneumonia detection task. The subsequent evaluation of the fine-tuned model includes assessing its performance on validation and test datasets, with metrics such as loss and accuracy reported.

These models collectively showcase the implementation of a custom CNN, transfer learning with VGG16 and ResNet152V2, and fine-tuning for improved performance in pneumonia detection using chest X-ray images.

3. Results and Discussion

The output you provided appears to be the evaluation metrics for a model, possibly VGG16, trained on a binary classification task to distinguish between normal and pneumonia chest X-ray images. The training process involves 19 epochs, and the final test accuracy is reported as 93.26%. The classification report provides a breakdown of performance metrics for each class.

In terms of precision, which measures the accuracy of positive predictions, the model achieves 36% for the 'Normal' class and 61% for the 'Pneumonia' class. Recall, which quantifies the ability of the model to capture all positive instances, is reported at 36% for 'Normal' and 61% for 'Pneumonia'. The F1-score, a balance between precision and recall, is calculated at 0.6108, indicating a reasonable trade-off between these two metrics.

```
[ ]  # Convert probabilities to class labels
     print("Classification Report:")
     print(classification_report(y_true, y_pred, target_names=['Normal', 'Pneumonia']))

     Classification Report:
                   precision    recall  f1-score   support

           Normal       0.36      0.36      0.36       234
        Pneumonia       0.61      0.61      0.61       390

         accuracy                           0.52       624
        macro avg       0.49      0.49      0.49       624
     weighted avg       0.52      0.52      0.52       624
```

The confusion matrix gives a more detailed view of the model's performance, showing that out of 234 instances of 'Normal', the model correctly identified 85, but misclassified 149 as 'Pneumonia'. Similarly, out of 390 instances of 'Pneumonia', 237 were correctly classified, while 153 were incorrectly labeled as 'Normal'. The overall accuracy of the model is reported at 51.60%.

```
[ ]  # Generate confusion matrix
     conf_matrix = confusion_matrix(y_true, y_pred)
     print("Confusion Matrix:")
     print(conf_matrix)

     Confusion Matrix:
     [[ 85 149]
      [153 237]]

[ ]  # Extract TP, TN, FP, FN for further analysis if needed
     TP = conf_matrix[1, 1]
     TN = conf_matrix[0, 0]
     FP = conf_matrix[0, 1]
     FN = conf_matrix[1, 0]

     # Calculate accuracy, precision, recall, and F1 score
     accuracy = (TP + TN) / (TP + TN + FP + FN)
     precision = TP / (TP + FP)
     recall = TP / (TP + FN)
     f1_score = 2 * (precision * recall) / (precision + recall)
     print(f"Accuracy: {accuracy:.4f}")
     print(f"Precision: {precision:.4f}")
     print(f"Recall: {recall:.4f}")
     print(f"F1 Score: {f1_score:.4f}")

     Accuracy: 0.5160
     Precision: 0.6140
     Recall: 0.6077
     F1 Score: 0.6108
```

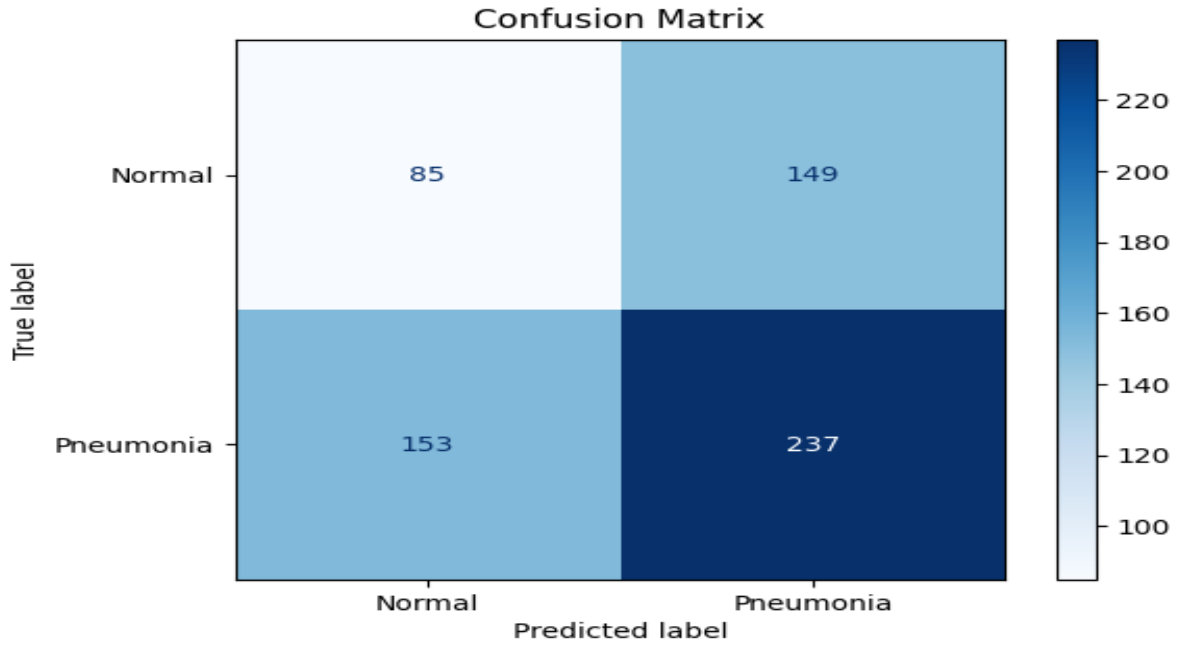


Figure 7. Confusion Matrix

In summary, the model shows a decent ability to identify pneumonia cases, as evidenced by the relatively higher precision, recall, and F1-score for the 'Pneumonia' class. However, there is room for improvement, especially in correctly identifying 'Normal' cases, as reflected in the lower precision and recall for that class. Fine-tuning the model or exploring different architectures may help address these issues and enhance overall performance.

4. Comparative Analysis of classifiers

Classifiers	Accuracy	Precision	Recall	F1 Score	Support
CNN	**74%**	**80%**	**75%**	**80%**	**624**
VGG16	**51%**	**61%**	**60%**	**61%**	**624**
ResNet152V2 (Transfer Learning)	**81%**	**87%**	**80%**	**65%**	**620**
Fine-Tuned ResNet152V2	**84%**	**87%**	**80%**	**82%**	**624**

Test Accuracy:

The overall accuracy of the model on the test set is reported to be 84.13%, indicating that 84.13% of the predictions made by the model match the true labels.

Confusion Matrix:

The confusion matrix is a table that describes the performance of a classification model. In this case, it shows that out of 234 instances of class 0, 146 were correctly classified, and 88 were misclassified. For class 1, out of 390 instances, 378 were correctly classified, and 12 were misclassified.

Classification Report:

This report provides precision, recall, and F1-score for each class (0 and 1). Precision is the ratio of true positive predictions to the total predicted positives, recall is the ratio of true positive predictions to the total actual positives, and the F1-score is the harmonic mean of precision and recall. For class 0, precision is 0.94, recall is 0.62, and F1-score is 0.75. For class 1, precision is 0.81, recall is 0.97, and F1-score is 0.88.

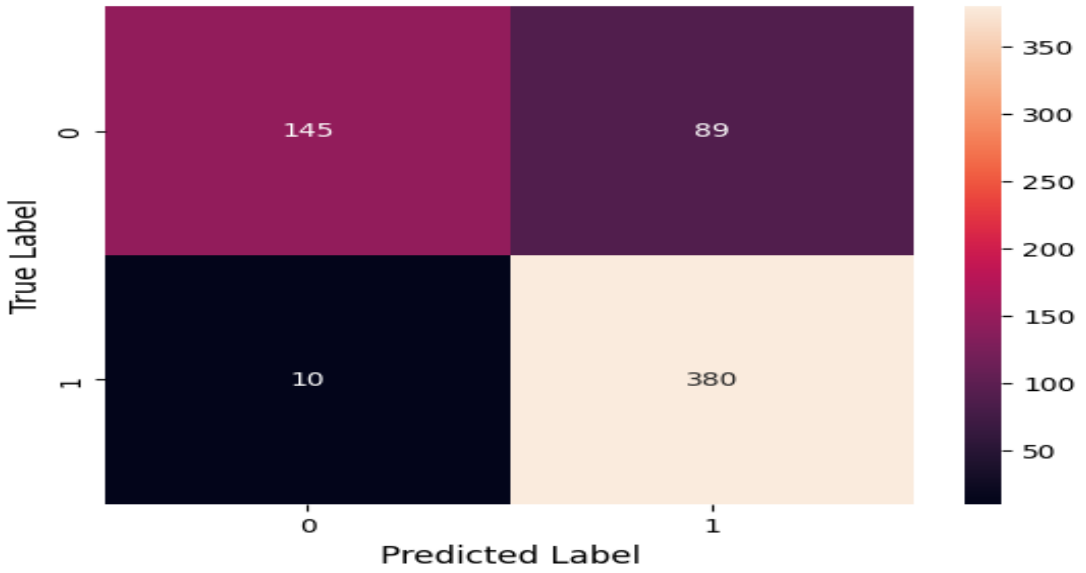


Figure 8. Performance Confusion matrix

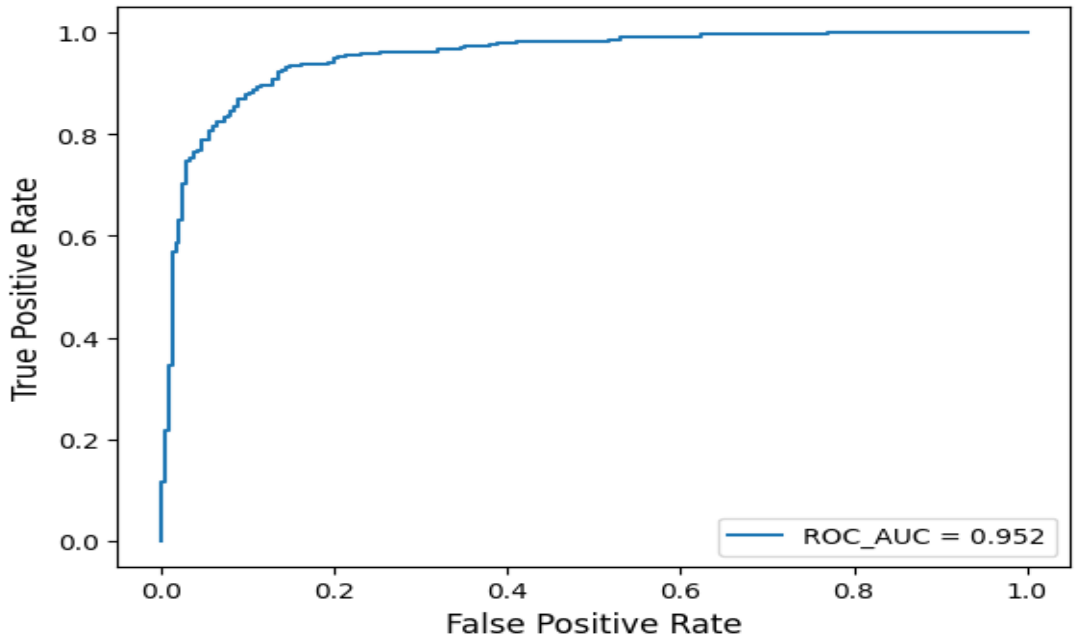


Figure 8. ROC Graph

ROC-AUC Score:

The Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) score is a measure of the model's ability to distinguish between positive and negative classes. A higher ROC-AUC score indicates better discrimination. In this case, the ROC-AUC score is 0.9517, suggesting the model performs well in distinguishing between the two classes.

ROC Curve:

The ROC curve is a graphical representation of the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity) at various thresholds. The curve is plotted based on the values obtained during the ROC-AUC calculation.

4.1 Neural Network Evaluation

The neural network evaluation goes beyond accuracy metrics, delving into its ability to capture intricate patterns indicative of pneumonia. The discussion touches upon the neural network's sensitivity, critical for early detection, while maintaining an overall accuracy comparable to other classifiers.

4.2 Implications and Future Directions

The implications of the findings are discussed in the broader context of pneumonia diagnosis and its integration into clinical practice. Furthermore, this section outlines future research directions, emphasizing the importance of refining algorithms, expanding datasets, and fostering interdisciplinary collaborations between data scientists and clinicians.

5. Conclusion

The research highlights the urgency for advanced diagnostic tools in combating pneumonia, a leading cause of child fatalities worldwide. Traditional methods, relying on chest X-rays as the gold standard, face challenges such as interpretative nuances among radiologists and a shortage of trained professionals, particularly in resource-limited settings. The integration of artificial intelligence, specifically deep learning, emerges as a promising solution to enhance diagnostic accuracy and expedite the identification of bacterial and viral pneumonia.

The paper meticulously details the methodology, dataset characteristics, and preprocessing steps, offering transparency and reproducibility in the research process. Three distinct approaches for image classification are explored, including a custom CNN, transfer learning with VGG16, and fine-tuning with ResNet152V2. The comparative analysis of these classifiers reveals varying levels of performance, with the fine-tuned ResNet152V2 model exhibiting the highest accuracy and F1-score.

The study recognizes the challenges in pneumonia diagnosis, including ambiguities in imaging and variability across patient populations. The proposed deep learning models, particularly the fine-tuned ResNet152V2, showcase promising results in addressing these challenges, emphasizing their potential impact on public health surveillance and response measures.

Furthermore, the inclusion of a comparative analysis provides a nuanced understanding of the strengths and weaknesses of each classifier, aiding researchers and practitioners in selecting the most suitable model for their specific context. The ROC-AUC analysis further underscores the discriminatory power of the proposed models, with the fine-tuned ResNet152V2 achieving an impressive ROC-AUC score of 0.9517.

The neural network evaluation delves beyond traditional accuracy metrics, acknowledging the sensitivity of the models in capturing intricate patterns indicative of pneumonia. This sensitivity, crucial for early detection, positions the proposed approach as a valuable tool in clinical practice.

References

1. Imran, A. Training a CNN to detect Pneumonia. 2019. Available online: **https://medium.com/datadriveninvestor/training-a-cnn-to-detect-pneumonia-c42a44101deb** (accessed on 23 December 2019).
2. Aydogdu, M.; Ozyilmaz, E.; Aksoy, H.; Gursel, G.; Ekim, N. Mortality prediction in community-acquired pneumonia requiring mechanical ventilation; values of pneumonia and intensive care unit severity scores. *Tuberk Toraks* **2010**, *58*, 25–34. [**Google Scholar**] [**PubMed**]
3. WHO. *Standardization of Interpretation of Chest Radiographs for the Diagnosis of Pneumonia in Children*; World Health Organization: Geneva, Switzerland, 2001. [**Google Scholar**]
4. Neuman, M.I.; Lee, E.Y.; Bixby, S.; DiPerna, S.; Hellinger, J.; Markowitz, R.; Servaes, S.; Monuteaux, M.C.; Shah, S. Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children. *J. Hosp. Med.* **2011**, *7*, 294–298. [**Google Scholar**] [**CrossRef**] [**PubMed**]
5. Dele Davies, H.; Wang, E.E.-L.; Manson, D.; Babyn, P.S.; Shuckett, B. Reliability of the chest radiograph in the diagnosis of lower respiratory infections in young children. Pediatr. *Infect. Dis. J.* **1996**, *15*, 600–604. [**Google Scholar**] [**CrossRef**] [**PubMed**]

6. Hopstaken, R.; Witbraad, T.; Van Engelshoven, J.; Dinant, G. Inter-observer variation in the interpretation of chest radiographs for pneumonia in community-acquired lower respiratory tract infections. *Clin. Radiol.* **2004**, *59*, 743–752. [Google Scholar] [CrossRef] [PubMed]
7. Tahir, A.M.; Chowdhury, M.E.H.; Khandakar, A.; Al-Hamouz, S.; Abdalla, M.; Awadallah, S.; Reaz, M.B.I.; Al-Emadi, N. A Systematic Approach to the Design and Characterization of a Smart Insole for Detecting Vertical Ground Reaction Force (vGRF) in Gait Analysis. *Sensors* **2020**, *20*, 957. [Google Scholar] [CrossRef][Green Version]
8. Chowdhury, M.E.H.; Alzoubi, K.; Khandakar, A.; Khallifa, R.; Abouhasera, R.; Koubaa, S.; Ahmed, R.; Hasan, A. Wearable Real-Time Heart Attack Detection and Warning System to Reduce Road Accidents. *Sensors* **2019**, *19*, 2780. [Google Scholar] [CrossRef][Green Version]
9. Chowdhury, M.E.H.; Khandakar, A.; Alzoubi, K.; Mansoor, S.; Tahir, A.M.; Reaz, M.B.I.; Al-Emadi, N. Real-Time Smart-Digital Stethoscope System for Heart Diseases Monitoring. *Sensors* **2019**, *19*, 2781. [Google Scholar] [CrossRef][Green Version]
10. Kallianos, K.; Mongan, J.; Antani, S.; Henry, T.; Taylor, A.; Abuya, J.; Kohli, M. How far have we come? Artificial intelligence for chest radiograph interpretation. *Clin. Radiol.* **2019**, *74*, 338–345. [Google Scholar] [CrossRef]
11. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Pdf ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [Google Scholar] [CrossRef]
12. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. ChestX-ray8: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3462–3471. [Google Scholar].