Enhancing Clinical Predictability in Lung Disease Diagnosis Using Deep Learning on Chest X-Rays

Muhammad Qasim Shah[1], Nafey Ahmed[1], Rabia Tehseen[2*], Uzma Omer[3]
[1]Department of Data Science, University of Central Punjab, Lahore, Pakistan
[2]Department of Computer Science, University of Central Punjab, Lahore, Pakistan
[3]Department of Information Sciences, University of Education, Lahore, Pakistan
*Correspondence: rabia.tehseen@ucp.edu.pk

Citation | Shah. M. Q, Ahmed. N, Tehseen. R, Omer. U, "Enhancing Clinical Predictability in Lung Disease Diagnosis Using Deep Learning on Chest X-Rays", IJIST Vol. 08 Issue. 01 pp 296-306, February 2026

Received| December 30, 2025 **Revised|** January 31, 2026 **Accepted|** February 04, 2026 **Published|** February 07, 2026.

TO increase diagnostic efficiency and accuracy, automated disease classification using medical images has become more important. To classify diseases based on labeled image data, this study investigates the use of several deep learning architectures, such as adapted convolutional neural networks (CNN), VGG16, ResNet50, and EfficientNetB0. The models were evaluated on accuracy, loss, and specific performance metrics using a rigorous training protocol and transfer learning. According to experimental results, VGG16 outperforms other models with the highest validation accuracy of approximately 97.6%. Simple CNNs also achieved competitive performance. Under the current training conditions, more complex models such as ResNet50 and EfficientNetB0 perform worse, indicating the need for further tuning or larger data sets. In addition to highlighting the effectiveness of pre-trained models in medical image classification tasks, this work provides a framework for comparative analysis. To improve clinical applicability, future directions include the integration of interpretability, advanced refinements, and dataset expansion.

Keywords: Deep Learning, Medical Image Classification, Transfer Learning, VGG16, CNN, Disease Diagnosis

Introduction:

As one of the leading causes of morbidity and mortality, lung conditions such as pneumonia, tuberculosis, and chronic obstructive pulmonary disease (COPD) continue to pose significant global health challenges. Early and accurate diagnosis is important to reduce complications, promote successful treatment, and increase survival rates. Because of its rapid acquisition, minimal radiation exposure, and wide availability in both urban and rural health care settings, the chest x-ray (CXR) remains the most popular and inexpensive diagnostic modality for visualizing lung abnormalities. However, overlapping anatomic structures, subtle pathologic variations, poor image quality, and variations in the clinical experience of radiologists often complicate the interpretation of chest radiographs. Prior studies state this variation from 15-30% depending on x-ray image complexity and reader expertise [1][2][3][4]. These difficulties can lead to inconsistent interpretations and delayed or incorrect diagnoses, especially in areas with limited access to professional radiologists [5][6].

Deep learning (DL) and artificial intelligence (AI) methods have shown promise in medical image analysis in recent years. Convolutional Neural Networks (CNNs) have the ability to extract discriminative features from chest radiographs for disease identification.

CNNs have been applied to detect many lung diseases, such as tuberculosis, lung nodules, and pneumonia. According to recent research, deep learning models, when properly trained and validated, can achieve diagnostic accuracy equal to, and sometimes even surpass, that of human experts [7]. In addition, the development of hybrid CNN- Transformer architectures and Transformer-based models has improved the ability of AI systems to capture both local and global context, enabling more accurate identification of complex patterns in chest X-rays [8].

Increasing model transparency and clinical reliability is the main goal of current research. Despite the high accuracy reported by many previous studies, their acceptability in actual clinical practice is limited due to their lack of interpretability and generalizability. To overcome this limitation, researchers have developed interpretable artificial intelligence (XAI) technology that allows doctors to see areas of interest that influence the model's decisions. These advances reduce the "black-box" aspect of deep learning solutions, promote clinical reasoning, and increase trust [9]. Furthermore, recent research highlights the importance of multi-label and multi-task learning frameworks, which enable AI systems to recognize multiple coexisting lung conditions in a single radiograph, more accurately representing real-world clinical conditions [10].

This research builds on these advances and proposes an intelligent system to improve clinical prediction in the diagnosis of lung disease from chest radiographs. The new architecture is not only concerned with experimental calculations for classification accuracy, but places special emphasis on the clinical interpretability and robustness of the system. A healthcare professional can be assured of more effective detection and discrimination of lung conditions with overlapping patterns using the described system, which uses multi-function fusion, deep learning, based pattern recognition, and interpretation modules. Reducing diagnostic uncertainty, enhancing efficiency in the clinical workflow, and ultimately improving patient outcomes are the goals of this new strategy in the healthcare sector in various clinical settings. The novelty of the study is a comprehensive comparison of multiple transfer learning architectures under unique preprocessing conditions and the evaluation of multi-disease classification rather than binary classification.

Literature Review:

The use of artificial intelligence in medical image analysis has been a gradual change, with research initially focused on classical machine learning techniques to detect lung diseases through chest X-ray analysis. Researchers usually combine traditional models such as Support Vector Machine (SVM), k-nearest neighbors (k-NN), decision trees, and random forests with hand-crafted feature extraction methods. The logic of these models was to identify the size, shape, and intensity of anatomical features from radiographic images, which were then used as input for the classification model. Although the methods were moderately successful in detecting lung abnormalities, the results were highly dependent on the quality of the extracted features. Furthermore, manual feature extraction techniques was very demanding in terms of domain expertise and had difficulty capturing the complex visual patterns of X-rays, thus limiting diagnostic accuracy [11][12].

Deep learning techniques began to replace traditional machine learning methods as computational power and large datasets became available. The dominant method was Convolutional Neural Networks (CNN) due to its feature extraction capability for hierarchical features directly from raw image data. Many well-known CNN architectures, such as VGGNet, ResNet, DenseNet, InceptionNet, and EfficientNet, were used to identify lung diseases such as pneumonia, tuberculosis, pulmonary nodules, chronic obstructive pulmonary disease (COPD), and COVID-19-related conditions. These architectures outperformed classical methods primarily because of their ability to detect subtle and complex spatial patterns. The use of transfer learning made these models even more powerful as it allowed them to use knowledge learned from large image repositories, resulting in significant improvements in accuracy and convergence speed when applied to medical datasets of limited size [13][14].

Recent studies have emphasized the importance of multi-label classification frameworks compared to binary or single-label classification frameworks. Such systems can identify multiple diseases in a single chest X-ray, thereby simulating clinical scenarios where overlapping abnormalities often occur. With the ability to model correlations between different diseases, multi-label deep learning models not only improve the analysis of patient conditions but also better adapt clinical practice. As a result, this success has significantly increased the confidence and deployment of automated diagnostic systems in healthcare [15][16].

Along with CNN-based architectures, transformer-based and hybrid deep learning models have attracted much interest in recent years. Vision Transformers (ViTs) use their own attention mechanism that equips the model with the ability to understand long-range dependencies in the image. This function finds relevant applications in chest X-ray analysis, where pathology can be detected anywhere in different lung fields. The hybrid CNN Transformer architecture combines the advantages of both strategies, using CNNs for local feature extraction and Transformers for global contextual relations. These sophisticated architectures have reported better performance in terms of generalization, robustness, and sensitivity compared to analog CNN models, and are therefore potential candidates for complex medical imaging tasks [17][18][19].

Deep learning models have achieved impressive results; however, their opacity has been a major factor preventing their large-scale deployment in clinical settings. Most of these models are considered black boxes, making it difficult for doctors to understand how the model arrives at its predictions. This problem gave rise to interpretable artificial intelligence (XAI) methods. In general, these techniques (eg, GRADE, CAM, saliency maps, and attention-based visualization) visually pinpoint specific areas of the chest x-ray that have the greatest impact on model decisions. As a result, these tools increase visibility and enable medical personnel to examine and explain the output of automated systems, thereby increasing the level of trust and reliability in AI-based diagnostic technology [20].

Although significant progress has been made, major problems of limited generalizability, class imbalance, variable image quality, and insufficient validation in heterogeneous populations still exist. A large part of the current research effort is mainly concerned with improving the predictive performance while recognizing clinical interpretation, scalability, and practical applicability aspects. Additionally, models trained on data from one source may produce different results when tested on data collected from other hospitals or imaging units. These shortcomings require the development of more flexible, generalizable, and clinically integrated solutions.

This study addresses the highlighted research gap by establishing an intelligent system to increase the clinical predictability of lung disease diagnosis from chest radiographs. The proposed solution is conceptualized as an end-to-end deep learning framework that can automatically load images from structured folders and assign labels based on folder names, thus eliminating the need for manual annotation. In the same experimental environment, four deep learning models are implemented and evaluated: an adapted simple CNN and three powerful transfer learning models, i.e., VGG16, ResNet50, and EfficientNetB0. The system, during training, keeps records and displays accuracy and loss values for the training and validation sets, thus learning behavior and model stability can be analyzed in detail. Then, a detailed comparison between the four models is made to determine the most reliable and efficient architecture for lung disease classification. By integrating automation, state-of-the-art deep learning techniques, performance visualization, and model comparison, the proposed system attempts to provide a practical, interpretable, and high-performance tool that enables physicians to make accurate and timely diagnostic decisions [6].

Methodology:

This study presents a deep learning-based framework aimed at automating disease classification through medical imaging. The aforementioned method proceeds systematically through various steps such as compilation of data sets, pre-processing, model building, training, and detailed performance evaluation. Many deep learning architectures are used to analyze classification performance and perform comparative evaluation.

Table 1. Dataset Description

Attribute	Description
Dataset Type	Medical image dataset
Imaging Modality	Chest X-ray (CXR) images
Application Domain	Automated lung disease classification
Total Number of Images	11,007 chest X-ray images
Classification Task	Multi-class image classification
Number of Classes	5
Class Distribution	Bacterial Pneumonia (3,617), COVID-19 (2,016), Normal (3,620), Tuberculosis (700), Viral Pneumonia (3,054)
Image Categories	Bacterial Pneumonia, COVID-19, Normal, Tuberculosis, Viral Pneumonia
Image Format	Digital chest X-ray images (PNG/JPEG)
Image Color Mode	Grayscale
Image Preprocessing	Resizing, normalization, and augmentation
Data Source	Publicly available medical X-ray repositories
Data Labeling	Expert-annotated clinical labels
Missing or Corrupted Images	None

Dataset Organization:

The dataset contains medical images classified by diseases. Each disease category is represented by a folder containing related images. Since the data is organized in a directory-based structure, labels can be generated automatically during data loading, making external annotation files unnecessary. Although image filenames may include disease names and serial numbers, labels are assigned based solely on folder names. If there is no separate test data set, the data is split internally into training and validation sets. Dataset labels were cross-verified using the dataset documentation. In addition, random manual inspection was performed to confirm label consistency.

Data Preprocessing:

All images are uniformly resized to 224 x 224 pixels to maintain model compatibility for all deep learning architectures. Normalizing pixel values to the range [0, 1] provides better numerical stability and thus increases convergence during training. TensorFlow's catalog, a built-in image loader that supports batching, shuffling, and label encoding, is used.

Deep Learning Models:

This work utilizes four models: an adapted convolutional neural network (CNN), VGG16, ResNet50, and EfficientNetB0. While the CNN is a baseline model, the other three models are fine-tuned with ImageNet pretrained weights via transfer learning. Classification layers are trained only on the target data set. We did not include DenseNet and Vision Transformers in this study because of limited computational resources and the relatively small size of the dataset. These models typically require larger training datasets and higher computational resources. Their use was beyond the scope of this project.

Model Training:

The models use the Adam optimizer and adopt hierarchical cross-entropy as the loss function. The models are trained for a fixed number of epochs, with performance assessed using accuracy and loss metrics.

Performance Evaluation:

The evaluation includes precision and loss curves, confusion matrix analysis, precision, recall, F1, score, ROC curve, precision recall curve, as well as TP, FP, FN, and TN analysis. Results are clearly presented using visualizations such as bar and pie charts.

Comparative Analysis:

All models are compared based on validation metrics to determine the most effective architecture. The best models are selected and saved in native Keras format for future use.

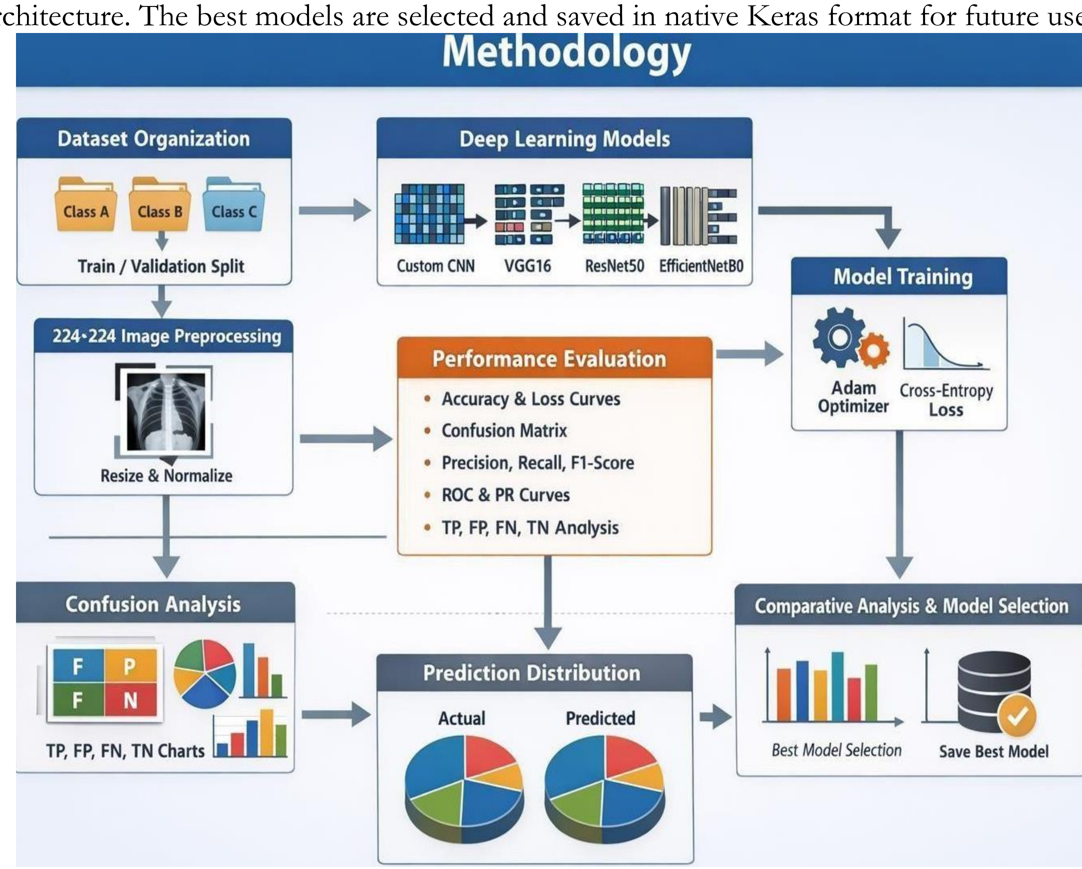

Figure 1. Methodology

Results and Analysis:

This section describes the evaluation of four deep learning models, namely Simple CNN, VGG16, ResNet50, and EfficientNetB0, for the disease classification task. The performance calculations for training and validation are summarized in Table 2.

Training and Validation Performance:

Both the Simple CNN and VGG16 models achieved high training accuracies of 98.9% and 99.8%, respectively. Their corresponding validation accuracy was 93.2% and 97.6%. These results demonstrate that these architectures exhibit strong learning capacity and good generalization performance.

On the other hand, ResNet50 and EfficientNetB0 showed significantly lower training and validation accuracies, with EfficientNetB0 being the hardest case (training accuracy ~27.3%, validation accuracy ~28.3%). The increased losses of these models indicate that they may have had difficulty in convergence. This can be attributed to factors such as the size of the data set, the settings of the hyperparameters or the complexity of the model.

Comparative Accuracy and Loss:

Table 2. Summary of validation performance and model complexity.

Model	Train Accuracy	Validation Accuracy	Train Loss	Validation Loss
Simple CNN	0.989	0.932	0.033	0.355
VGG16	0.998	0.976	0.010	0.082
ResNet50	0.679	0.665	0.836	0.875
EfficientNetB0	0.273	0.283	1.502	1.492

CNN Model Results (Per Class):

Table 3. CNN Per Class Results

Class	Precision	Recall	F1-Score	Accuracy	Support
Bacterial_Pneumonia	0.92	0.95	0.94	0.95	706
COVID	0.95	0.91	0.93	0.91	394
Normal	0.93	0.95	0.94	0.95	736
Tuberculosis	0.96	0.89	0.92	0.89	131
Viral_Pneumonia	0.99	0.97	0.98	0.97	634

VGG16 Model Results (Per Class):

Table 4. VGG16 Per Class Results

Class	Precision	Recall	F1-Score	Accuracy	Support
Bacterial_Pneumonia	0.99	0.91	0.95	0.91	706
COVID	1.00	0.92	0.96	0.92	394
Normal	0.93	0.96	0.95	0.96	736
Tuberculosis	0.99	0.95	0.97	0.95	131
Viral_Pneumonia	0.90	1.00	0.95	0.99	634

RESTNET50 Model Results (Per Class):

Table 5. RESTNET50 Per Class Results

Class	Precision	Recall	F1-Score	Accuracy	Support
Bacterial_Pneumonia	0.66	0.72	0.69	0.72	706
COVID	0.60	0.54	0.57	0.54	394
Normal	0.67	0.81	0.73	0.81	736
Tuberculosis	0.47	0.33	0.39	0.39	131
Viral_Pneumonia	0.92	0.73	0.81	0.73	634

EFFICIENTNETB0 Model Results (Per Class):

Table 6. EFFICIENTNETB0 Per Class Results

Class	Precision	Recall	F1-Score	Accuracy	Support
Bacterial_Pneumonia	0.42	0.99	0.59	0.98	706

COVID	0.38	0.01	0.01	0.01	394
Normal	0.14	0.00	0.00	0.00	736
Tuberculosis	0.00	0.00	0.00	0.00	131
Viral_Pneumonia	0.52	0.75	0.62	0.75	634

Simple CNN and VGG16 are clearly more efficient than ResNet50 and EfficientNetB0, both in terms of accuracy and loss. Indeed, VGG16 achieves the highest validation accuracy overall across all classes, demonstrating its suitability for this dataset and classification task.

Interpretation of Results:

Simple CNN serves as a robust baseline, achieving high accuracy with low loss, confirming that a custom-designed architecture tailored for this problem is highly effective.

VGG16 achieves the highest validation accuracy, leveraging pre-trained ImageNet weights and the deep VGG16 architecture.

ResNet50 shows average performance, but it is below the peak, and this may be a result of overfitting or a lack of sufficient fine-tuning due to the nature of the dataset.

EfficientNetB0 performs very poorly, and this performance may have been affected by a mismatch between the model's expected input resolution and the size of the images used. In addition, only a limited number of layers were fine-tuned, which could have restricted the model's ability to fully adapt to the dataset.

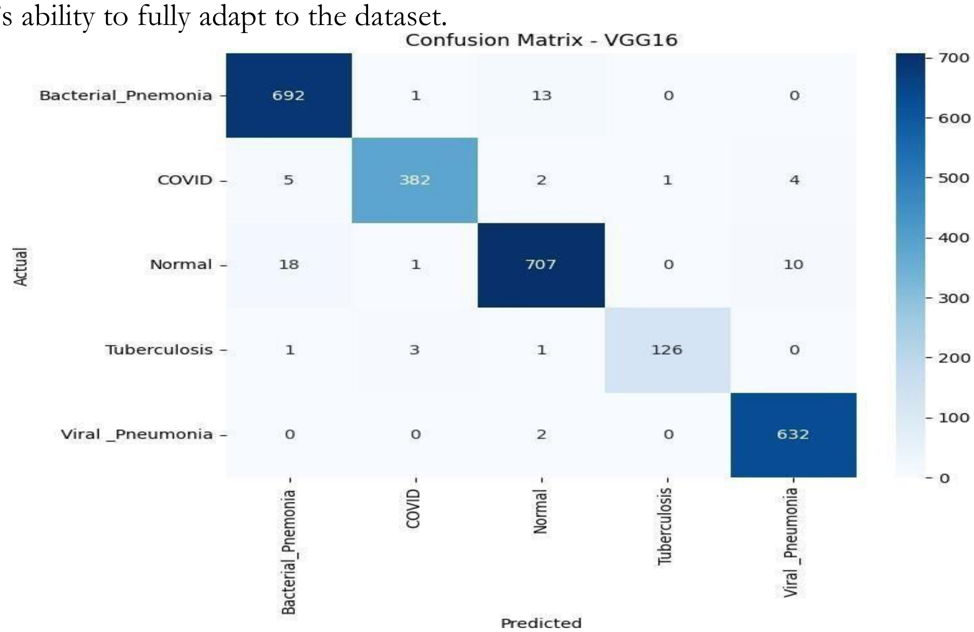

Figure 2. Confusion Matrix - VGG16

Figure 2 illustrates the confusion matrix of the VGG16 model for multi-class chest X-ray image classification. The confusion matrix reveals a high degree of diagonal dominance, indicating high correct classification rates across all five classes. The model was able to correctly classify 692 cases of Bacterial Pneumonia, 382 cases of COVID, 707 cases of Normal, 126 cases of Tuberculosis, and 632 cases of Viral Pneumonia. The high number of true positives indicates the excellent discriminative power of the model for across both dominant and less-represented classes.

There were very few misclassifications, and these were mostly between classes that are similar in nature. For example, there were a few cases of Normal being classified as Bacterial Pneumonia (18 cases) and Viral Pneumonia (10 cases). Similarly, minimal confusion occurred between COVID and Bacterial Pneumonia. Tuberculosis showed excellent classification consistency with very few false positives. The confusion matrix thus confirms the high validation accuracy (97.6%) and also indicates that the VGG16 model can make consistent class-wise predictions with very few instances of overlap between classes.

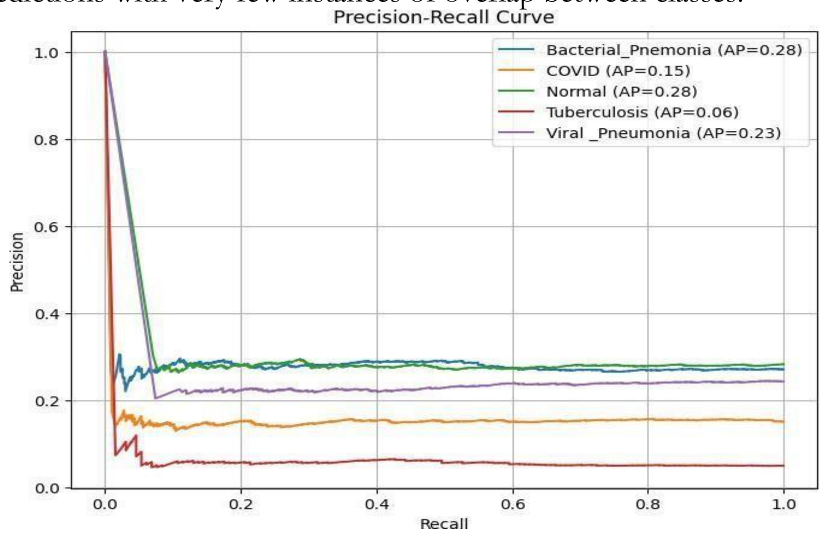

Figure 3. Precision -Recall Curve

In Figure 3 shows the precision-recall (PR) curves of the VGG16 model for each of the five classes of diseases. The PR curves represent the relationship between precision and recall for varying thresholds of classification. The Average Precision (AP) scores show that Bacterial Pneumonia (AP = 0.28) and Normal (AP = 0.28) have relatively stronger performance, followed by Viral Pneumonia (AP = 0.23). COVID-19 has a moderate AP score of 0.15, whereas Tuberculosis, a minority class, has the lowest AP score of 0.06, reflecting its greater classification difficulty

The PR curves in Figure 3 show that precision decreases as recall increases, which is typical in multi-class classification with imbalanced datasets. The stable curves for Bacterial Pneumonia and Normal demonstrate that the model maintains consistent predictive confidence across all thresholds. On the other hand, the lower and less steep curve for Tuberculosis indicates lower discriminative power across different recall levels. The results from the PR analysis are consistent with the observation that VGG16 is a reliable model for the majority classes but relatively less sensitive for the minority classes.

Figure 4 depicts the class distribution predicted by the VGG16 model for the five categories of diseases. The class distribution indicates that Normal (27.9%) and Bacterial Pneumonia (27.5%) have the highest proportion of predictions, followed by Viral Pneumonia (24.8%). COVID-19 has a proportion of 14.9% in the total predictions, and Tuberculosis has the lowest proportion of 4.9%.

The fact that the proportions of Normal and Bacterial Pneumonia are quite close indicates that the VGG16 model has equal capability to distinguish between healthy and diseased samples. The lower proportion of Tuberculosis, the least represented class in the dataset, indicates that the model does not overpredict minority classes. In addition, the absence of extreme class proportions indicates that the VGG16 model does not favor any particular class.

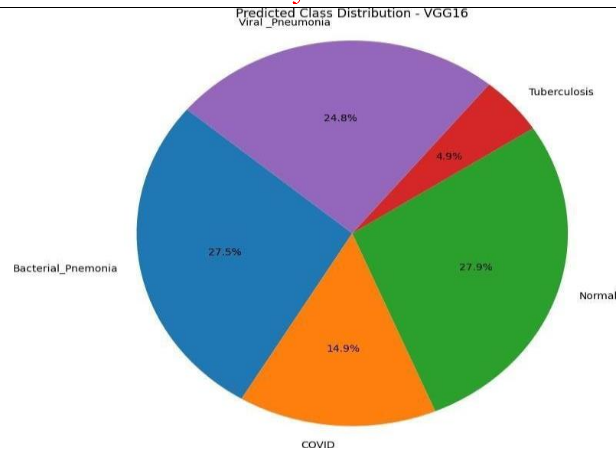

Figure 4. Predicted Class Distribution - VGG16

Discussion:

The results show that VGG16 achieved the highest validation accuracy (97.6%) and consistently improved F1 scores per class across all five disease categories. These findings are consistent with comparative studies such as Yao and Al-Sheikh, who reported that transfer learning with a pre-trained CNN architecture could outperform deep residual models in multi-class chest X-ray classification. Stable performance in minority classes, especially Tuberculosis and COVID-19, suggests efficient feature extraction and strong generalizability.

The clinical relevance of these findings is reinforced by the documented variation between readers in the interpretation of radiographs. Van Essen and Nair's studies showed variation among radiologists in COVID-19 assessment. Similarly, Lee highlighted inconsistencies in diagnostic labeling. The high and consistent performance of VGG16 in this study indicates the ability of deep learning models to provide standardized and reproducible classification support.

In contrast, ResNet50 and EfficientNetB0 showed relatively weak results, indicating that deeper or more complex architectures do not necessarily guarantee superior performance. While some studies, such as Nawaz's, have reported robust results for residual networks, model efficiency appears to be highly dependent on dataset properties and optimization strategies. Although newer transformer-based approaches (e.g., LION) are promising, our findings indicate that well-tuned CNN-based transfer learning models remain highly competitive for multi-class lung disease detection.

Overall, the study emphasizes that architecture choice and task-specific fine-tuning are important to achieve strong performance. Future work should include explanatory frameworks and cross-dataset validation to further improve clinical applicability and reliability.

Conclusion:

This research compares the effectiveness of different deep learning models to classify diseases based on medical images. In summary, transfer learning models such as VGG16 were found to be superior to a custom-built CNN and even complex architectures such as ResNet50 and EfficientNetB0 with respect to the current dataset and training configuration. The high validation accuracy of VGG16 demonstrates the effectiveness of pre-trained networks for medical image analysis, particularly when data and computational resources are limited.

On the other hand, the relatively low performance of ResNet50 and EfficientNetB0 implies that these models require very careful hyperparameter tuning and perhaps a larger data set to realize their full potential. Thus, this study provides a comparative framework to guide the selection of models for disease classification tasks. Furthermore, it emphasizes the importance of choosing a model based on the nature of the data. Subsequent studies will be dedicated to improving the generalizability of the model through data augmentation, interpretability to achieve clinical confidence, using capacity methods, and improving the dataset to increase robustness. These results are based only on publicly available datasets. To confirm how well the model performs in real-world settings, it should be tested in the future on data collected from multiple institutions.

Future Directions:

Future work should include expanding the data set, incorporating advanced data enhancements, and integrating interpretation techniques such as Grad-CAM (XAI) to improve clinical applicability. Dense Net and Vision Transformers can be used with better resources and a large dataset for more accurate results. Future work should also address potential dataset noise to enhance model robustness.

References:

[1] Marly van Assen, Mohammadreza Zandehshahvar, Hossein Maleki, Yashar Kiarashi, Timothy Arleo, "COVID-19 pneumonia chest radiographic severity score: variability assessment among experienced and in-training radiologists and creation of a multireader composite score database for artificial intelligence algorithm development," *Br J Radiol*, vol. 95, 2022, doi: 10.1259/bjr.20211028.

[2] Arjun Nair, Alexander Procter, Steve Halligan, Thomas Parry, "Chest radiograph classification and severity of suspected COVID-19 by different radiologist groups and attending clinicians: multi-reader, multi-case study," *Eur Radiol*, vol. 33, no. 3, pp. 2096–2104, 2023, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/36282308/

[3] Dana Li, Lea Marie Pehrson, Lea Tøttrup, Marco Fraccaro, "Inter- and Intra-Observer Agreement When Using a Diagnostic Labeling Scheme for Annotating Findings on Chest X-rays-An Early Step in the Development of a Deep Learning-Based Decision Support System," *Diagnostics*, vol. 12, no. 12, 2022, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/36553118/

[4] Lea Marie Pehrson, Dana Li, "Clinicians' Agreement on Extrapulmonary Radiographic Findings in Chest X-Rays Using a Diagnostic Labelling Scheme," *Diagnostics*, vol. 15, no. 7, p. 902, 2025, doi: https://doi.org/10.3390/diagnostics15070902.

[5] G. R. John T. Murchison, "Validation of a deep learning computer aided system for CT based lung nodule detection, classification, and growth rate estimation in a routine clinical population," *PLoS One*, vol. 17, no. 5, 2022, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/35511758/

[6] S. J. Raheel Siddiqi, "Deep Learning for Pneumonia Detection in Chest X-ray Images: A Comprehensive Survey," *J. Imaging*, vol. 10, no. 8, p. 176, 2024, [Online]. Available: https://www.mdpi.com/2313-433X/10/8/176

[7] M. B. Jakub Kufel, "Multi-Label Classification of Chest X-ray Abnormalities Using Transfer Learning Techniques," *J. Pers. Med.*, vol. 13, no. 10, 2023, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/37888037/

[8] K. A. & S. S. Sukhendra Singh, Manoj Kumar, Abhay Kumar, Birendra Kumar Verma, "Efficient pneumonia detection using Vision Transformers on chest X-rays," *Sci. Rep.*, vol. 14, no. 2487, 2024, doi: https://doi.org/10.1038/s41598-024-52703-2.

[9] M. J. I. Muhammad Aasem, "Toward explainable AI in radiology: Ensemble-CAM for effective thoracic disease localization in chest X-ray images using weak supervised learning," *Front. big Data*, 2024, [Online]. Available: https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2024.1366415/full

[10] B. B. Izegbua E. Ihongbe ,Shereen Fouad ,Taha F. Mahmoud,Arvind Rajasekaran, "Evaluating Explainable Artificial Intelligence (XAI) techniques in chest radiology

imaging through a human-centered Lens," *PLoS One*, 2024, doi: https://doi.org/10.1371/journal.pone.0308758.

[11] H. A. J. Mona Hmoud Al-Sheikh, Omran Al Dandan, Ahmad Sami Al-Shamayleh, "Multi-class deep learning architecture for classifying lung diseases from chest X-Ray and CT images," *Sci. Rep.*, vol. 13, no. 19373, 2023, [Online]. Available: https://www.nature.com/articles/s41598-023-46147-3

[12] M. A. Mehdi Neshat, "Hybrid Inception Architecture with Residual Connection: Fine-tuned Inception-ResNet Deep Learning Model for Lung Inflammation Diagnosis from Chest Radiographs," *Procedia Comput. Sci.*, vol. 235, pp. 1841–1850, 2024, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050924008512

[13] M. M. Muhammad Ahmad, Sardar Usman, Ildar Batyrshin, "Automated diagnosis of lung diseases using vision transformer: a comparative study on chest x-ray classification," *arXiv:2503.18973*, 2025, [Online]. Available: https://arxiv.org/abs/2503.18973

[14] S. G. Burcu Oltu, "Automated classification of chest X-rays: a deep learning approach with attention mechanisms," *BMC Med. Imaging*, vol. 25, no. 71, 2025, [Online]. Available: https://link.springer.com/article/10.1186/s12880-025-01604-5

[15] Iyad Sultan, Hasan Gharaibeh, "LungVisionNet: A Hybrid Deep Learning Model for Chest X-Ray Classification—A Case Study at King Hussein Cancer Center (KHCC)," *Technologies*, 2025, [Online]. Available: https://www.semanticscholar.org/paper/LungVisionNet%3A-A-Hybrid-Deep-Learning-Model-for-at-Sultan-Gharaibeh/115225758f1887b002b801860ce9346296e62fe2

[16] Yilin Yao, Yinghan Li, Shirong Zheng, Taoyu Zhu, "Deep Learning Models for Multi-Class Pneumonia Detection in Chest X-Rays: A Comparative Study of VGG16, MobileNet, and ResNet152," *Comput. Simul. Appl.*, 2025, [Online]. Available: https://journal.whioce.com/index.php/csa/article/view/916

[17] C. T. P. Rahul Kumar, "Enhanced Multi-Model Deep Learning for Rapid and Precise Diagnosis of Pulmonary Diseases Using Chest X-Ray Imaging," *Diagnostics*, vol. 15, no. 3, p. 248, 2025, [Online]. Available: https://www.mdpi.com/2075-4418/15/3/248

[18] H. M. Bouthaina Slika, Fadi Dornaika, "Lung pneumonia severity scoring in chest X-ray images using transformers," *Med. Biol. Eng. Comput.*, vol. 62, pp. 2389–2407, 2024, [Online]. Available: https://link.springer.com/article/10.1007/s11517-024-03066-3

[19] A. T. Xiaoyang Fu, Rongbin Lin, Wei Du, "Explainable hybrid transformer for multi-classification of lung disease using chest X-rays," *Sci. Rep.*, vol. 15, 2025, [Online]. Available: https://www.nature.com/articles/s41598-025-90607-x

[20] S. R. Sobia Nawaz, "Deep Learning ResNet101 Deep Features of Portable Chest X-Ray Accurately Classify COVID-19 Lung Infection," *Comput. Mater. Contin.*, vol. 75, no. 3, 2023, [Online]. Available: https://www.techscience.com/cmc/v75n3/52600/html