Grenze International Journal of Engineering and Technology, June Issue

LUCID-X - Lung Condition Identification using Deep Learning and X-Ray Images

Debajit Guha[1], Swapnil Chaki[2], Saptarshi Datta[3], Tushit Roy[4], Ansu Nanda[5] and Arindam Chatterjee[6]
[1-5]Student, CSE Department, Heritage Institute of Technology, Kolkata, West Bengal, India
Email: debajit.guha.cse27@heritageit.edu.in, swapnil.chaki.cse27@heritageit.edu.in, saptarshi.datta.cse27@heritageit.edu.in, tushit.roy.cse27@heritageit.edu.in, ansu.nanda.cse27@heritageit.edu.in
[6]Assistant Professor, CSE Department, Heritage Institute of Technology, Kolkata, West Bengal, India
Email: arindam.chatterjee@heritageit.edu

***Abstract*— Chest X-Rays are important sources for diagnosing pulmonary diseases like Pneumonia, COPD etc. Unfortunately interpreting chest X-rays remains complex and skill-intensive, and hence can be inaccurate, resulting in wrong treatments that can even lead to death. In this work, we introduce an improved deep learning method for detecting lung diseases via EfficientNet-B4 and MobileNet-V1. To boost classification accuracy over four lung disease classes and one normal class, we used state-of-the-art data augmentation, hyperparameter fine-tuning, and loss function optimization. The EfficientNet-B4 based model gave a train accuracy of 95.01%, test accuracy and average F1-score of about 92% while the MobileNet-V1 based model resulted in a train accuracy of around 93.15%, test accuracy of 88% and an average F1-score of about 87.8%.**

***Keywords*— Chest X-Ray, Pneumonia, COVID-19, Deep Learning, MobileNet.**

I. INTRODUCTION

Chest X-rays are indispensable tools in the field of medical diagnostics, widely recognized for their prevalence and accessibility within healthcare settings. These radiological examinations play a fundamental role in the initial assessment of pulmonary conditions, serving as a crucial first line of inquiry. Their importance lies in the capacity to provide timely and vital insights, enabling healthcare professionals to make informed and accurate medical decisions. The diagnostic utility of chest radiography is particularly significant due to its ability to detect a diverse range of lung anomalies. This includes common respiratory conditions like pneumonia and COPD, which are major public health concerns globally, as well as more complex and severe pathologies such as lung cancers. The early detection of these conditions through chest X-rays is critical, as it can dramatically affect treatment decisions and patient outcomes. As such, chest radiography not only aids in the timely management of diseases but also plays a pivotal role in enhancing the overall efficacy of medical interventions, highlighting its essential place in modern medical practice [1].

However, despite their extensive application, interpreting chest X-rays remains complex and skill-intensive. An accurate diagnosis from these images often necessitates a considerable degree of expertise and experience, attributes that are cultivated over many years of clinical practice. Misinterpretations can lead to significant adverse consequences, including the potential for delayed diagnosis or the administration of inappropriate treatments. This adversely affects patient outcomes, and in many cases, can even lead to premature deaths.

Grenze ID: 01.GIJET.11.2.487_22

This challenge is further intensified by the high volume of imaging studies that radiological departments must handle, especially after the onset of the COVID-19 pandemic. This often strains the resources and capacity of these services. Given these issues, there is an urgent need to develop and implement enhanced diagnostic tools. These tools should improve the accuracy, efficiency, and accessibility of pulmonary assessments. This advancement in diagnostic capabilities is essential for mitigating the pressures faced by healthcare systems and enhancing the quality of care provided to patients with pulmonary conditions [1]. Medical image analysis has significantly transformed in recent years by integrating Deep Learning (DL) technologies. These technologies are renowned for their ability to automate and refine complex analytical tasks, opening up new possibilities in medical radio diagnosis. The effects of DL are notably impactful in chest radiology, where these methods have demonstrated remarkable ability. DL architectures are noted for their superior pattern recognition abilities, significantly improving the interpretation of chest radiographs. This enhancement has brought about a more accurate approach to identifying and diagnosing respiratory ailments. Leveraging extensive datasets of medical images, these architectures can detect complicated anomalies and patterns that might be missed by conventional approaches, thereby supplying a complete understanding of lung infections [1].

II. Related Work

In [1], the authors introduce an advanced identification model that utilizes the VGG16 architecture, specifically adapted for identifying various lung anomalies such as COVID-19 pneumonia, normal appearance of the lungs, and viral pneumonia. They used the data augmentation technique through CycleGAN. The combined performance of the advanced VGG model with the CycleGAN augmentation technique demonstrates remarkable outcomes in several evaluation metrics, including recall, F1-score, accuracy, precision, and area under the curve (AUC). The results of the advanced VGG16 model showcased remarkable accuracy, achieving 98.58%.
In [2], authors developed a computer-aided diagnosis system for automatic pneumonia detection using chest X-ray images. They employed deep transfer learning to handle the scarcity of available data and designed an ensemble of three convolutional neural network models: GoogLeNet, ResNet-18, and DenseNet-121. The proposed method achieved accuracy rates of 98.81% and 86.85% and sensitivity rates of 98.80% and 87.02% on the Kermany and RSNA datasets, respectively.
In [3], authors have shown how medical imaging techniques like Chest X-rays played an important role in diagnosing the COVID-19. The most common radiologic findings in COVID-19 are airspace opacities (consolidations and / or ground glass opacities), which are typically bilateral, peripheral, and located primarily in the lower fields. The systematic review paper [4] explores and provides a comprehensive analysis of the related studies that have used deep learning techniques to analyze chest X-ray images. The authors present the state-of-the-art deep learning based pneumonia and COVID-19 detection solutions, trends in recent studies, publicly available datasets, guidance to follow a deep learning process, challenges and potential future research directions in this domain. Authors in [5] propose an explainable ensemble learning approach, CX-Net, for lung segmentation and diagnosing lung disorders using Chest X-Ray images. They compare four state-of-the-art CNN models, including feature pyramid network, U-Net, LinkNet, and a customized U-Net model with ImageNet feature extraction. Their outlier-resistant CX-Net achieves superior performance in lung segmentation, precision of 0.993, recall of 0.980, and accuracy of 0.976.

III. Objective

There are many different ways of diagnosing pulmonary diseases. Medical imaging techniques like Chest X-Rays (CXRs) and CT Scan images have gained prominence over many other techniques, viz., purely clinical ones because of their universal availability. Current techniques for diagnosing lung diseases from medical images are inaccurate, expensive, complex and slow. Our aim is to provide a simple, inexpensive and a very accurate diagnosis of the lung condition / disease from CXRs with a fast response time.
In this work, we plan to use CXRs for diagnosing selected four abnormal lung conditions. These are Bacterial Pneumonia, viral Pneumonia, COVID-19 and Tuberculosis. We will separate these abnormalities and potential fatal conditions from perfectly normal lung CXRs.

IV. Methodologies

A. Problem Representation

In this section, we provide a methodical representation of our problem and our approach towards solving it. In biomedical images like chest x-rays, a large number of subtle features like textures and contours are of

paramount importance. Due to this reason, it becomes increasingly difficult to classify them by training deep learning models from scratch. To address this issue, we have used 'transfer learning', a technique where a pre-trained model is used as the basic feature extractor. To make the model fit to a specific dataset, several custom layers are added on top of the pre-trained model. A part of the original architecture of the pre-trained model is made trainable on the newly updated weights for every epoch while training. This 'fine-tuning' helps the entire model architecture to 'learn better' or fit to the new dataset. On the other hand, the primitive weights of the pre-trained model helps the entire transfer learning model to extract better features and drastically reduce training time.

Figure 1 and Figure 2 provide an overview of the transfer learning model architectures based on MobileNet-V1 and EfficientNet B4 respectively.

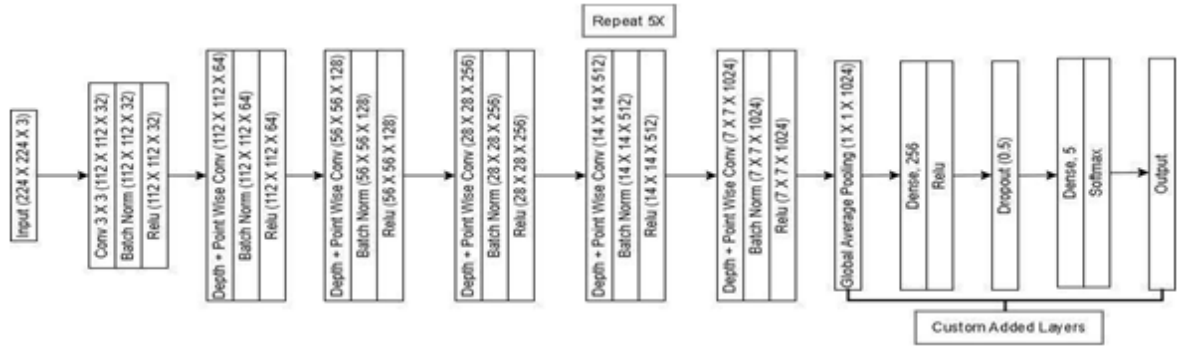

Figure 1. Architecture Diagram of MobileNet V1 based model

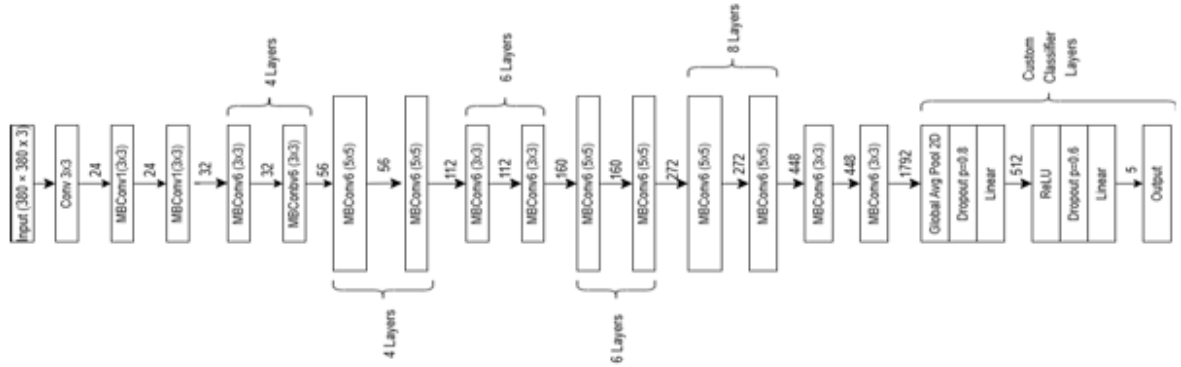

Figure 2. Architecture Diagram of EfficientNet B4 based model

B. Solution Methodologies

In this paper, we introduce an improved deep learning method for detecting lung disease via MobileNet-V1 and EfficientNet-B4. To boost classification accuracy over four lung disease classes and one normal class, we used state-of-the-art data augmentation, hyperparameter fine-tuning, and loss function optimization. Pre-processing medical images, training both models with specifically designed improvements, and testing them on a systematic dataset are the components of our methodology. The following sections [1,2] explain the particular innovations and methods used to attain good diagnostic accuracy for each model.

MobileNet-V1

In this section, we discuss the Transfer Learning model based on MobileNet-V1. The base model used in this architecture is MobileNetV1, a light-weight Convolutional Neural Network, pretrained on the ImageNet dataset. MobileNet-V1 has been used since it is computationally cheap, light-weight and has good feature extraction properties thus making it suitable for classifying CXR images. The top layer of MobileNet-V1 is replaced, and the input size of every image is resized to (224, 224,3).

In order to make the model fit to our data we have added custom layers over the original MobileNet V1 model. These layers are:

- Global Average Pooling (GAP) layer.
- A Dense layer with 256 nodes, 'Relu' activation function and L2- kernel regularizer.

- A dropout layer.
- A final predictions layer.

The GAP layer (i) reduces the spatial dimensions (height and width) of a feature map by computing the average of each feature channel. The Dense layer (ii) has 256 neurons and uses the ReLU (Rectified Linear Unit) activation function to introduce non-linearity in data and solve the crucial vanishing gradient problem. The L2 kernel regularizer in (ii) penalizes large weights so as to prevent overfitting. The Dropout layer (iii) is used to prevent overfitting by randomly dropping out a certain fraction of the neurons before propagating to the next layer. The final output layer (iv) contains 5 neurons representing each one of our output classes. Since this is a multi-class classification; the softmax activation function is used for converting the raw output values into probabilities. L2 regularization is again added to prevent overfitting during training.

In order to make our architecture more accurate we have fine-tuned the last 30% of the MobileNet-V1 layers. Specifically, the first 70% of the layers are frozen to retain the general feature extraction capabilities learned from the original ImageNet dataset, while the remaining 30% of the layers are unfrozen to allow learning of task-specific patterns from our chest X-ray dataset. This has helped the model to adapt to the unique characteristics of the CXR images without losing the pre-trained knowledge from ImageNet.

The above discussed model is trained using the categorical cross entropy loss function, suitable for multi-class classification tasks. Categorical cross entropy is mainly used in multi-class classification where the labels are one-hot encoded. It causes the model to maximize the probability of the correct class while minimizing the others. The Adam optimizer is used to handle noisy gradients well. The learning rate is 0.0001 to ensure better fine-tuning. The metrics 'accuracy' and 'F1 score' are used to monitor the training performance of the model.

Overfitting being a key problem in deep learning models, we have used two crucial callback functions to prevent this. They are:

- Early stopping.
- Reduce LR On Plateau.

Early stopping is implemented to monitor the validation loss. If the validation loss does not improve for seven consecutive epochs, training is automatically stopped, and the best-performing model weights are restored. The Reduce LR On Plateau is a learning rate reduction strategy. If the validation loss plateaus for three consecutive epochs, the learning rate was reduced by half. The minimum allowable learning rate is 1e-6. This is done to prevent stalling of training the model beyond a certain slow learning rate. The final architecture comprising all the changes as discussed in this section is trained with the training data at 40 epochs with a batch size of 32.

Efficient Net-B4

In this section, we will discuss the transfer learning of EfficientNet-B4 as the base model and techniques implemented to improve the performance for Detection of Lung Disease from X-Ray image dataset. In the paper [6], EfficientNet-B4 is one variant within a family of models derived from a common baseline (EfficientNet-B0) through a compound scaling method. It balances depth, width, and resolution, thereby offering a high-performing yet computationally efficient model for CXR image classification tasks. However, during initial experiments, we encountered overfitting challenges—a prevalent phenomenon in medical image-related tasks due to limited and highly specialized datasets.

To address these challenges, we implemented several overfitting prevention techniques and architectural improvements, as detailed below:

Domain Transfer via Fine-Tuning

Recognizing the distinct characteristics of medical imaging compared to natural images, we fine-tuned the last 6 layers of EfficientNet-B4. This domain transfer approach allowed the model to adapt its learned representations to the medical domain, resulting in more relevant and better feature extraction and improved generalization on medical images data.

Enhanced Classifier Architecture

We redesigned the classifier layer to introduce more learnable parameters in order to enhance the discriminative ability of our model and reduce overfitting. The new classifier consists of:

- A Linear layer of incoming 1792 in features from the Base model and 512 out features.
- ii.) A Dropout layer to randomly deactivate neurons and mitigate over-reliance on specific features with drop probability of 0.8.
- A ReLU activation function to introduce non-linearity.
- Another Linear layer to refine the learned features with 512 in features and 5 out features for our classes.
- A subsequent Dropout layer to further enforce regularization with drop probability 0.6.

Optimization Strategy Improvements:

- Optimizer Transition: We replaced the Adam optimizer with AdamW, decoupling weight decay from the step of gradient updates. This adjustment offered more reliable regularization in the form of enhanced weight decay management.
- L2 Regularization: In tandem with AdamW, L2 regularization was applied to penalize large weights, thereby further reducing the risk of overfitting. Weight decay was tuned from 1e-4 to 1e-3, and dropout rates were increased from 0.5 to 0.8 to curb overfitting.

Learning Rate Scheduling with Cosine Annealing:

To achieve smoother convergence and improved training control, we employed a cosine annealing learning rate scheduler. This scheduler adjusts the learning rate following a cosine function, gradually reducing it over time. Such a schedule helps the model settle into a minimum more smoothly by providing finer control during the later stages of training.

Additionally, we used different learning rates for distinct parts of the model:

- Feature Extractor: Learning rate adjusted from 1e-5 to 1e-4.
- Classifier: Learning rate scaled from 1e-4 to 1e-3.

Training Regimen:

The model was trained for a total of 40 epochs; however, an Early stopping criterion with a tolerance of 0.06 was implemented, leading to termination at 31 epochs. This prevented unnecessary overfitting in later stages of training. The entire training process took approximately 4280.296 seconds or roughly 1 hour and 11 minutes.

These modifications produced a robust, generalizable model. Domain-specific fine-tuning enhanced lung feature extraction, while an improved classifier and rigorous regularization reduced overfitting. Adjusted learning rates and a cosine annealing schedule ensured smoother convergence, culminating in a reliable lung disease detection system that leverages state-of-the-art deep learning techniques for medical image analysis.

V. DATA MODEL

In this section, we have discussed the various aspects of our dataset and steps to make it fit to be trainable. The elaborated description of the dataset is provided below.

A. Data Cleaning

Cleaning image data is an essential preprocessing step in medical imaging tasks, for delivering high-quality input to deep learning models. It includes a series of crucial techniques, like noise reduction through denoising filters (e.g., Gaussian and median filters), contrast enhancement through ColorJitter etc. Normalization techniques normalise pixel intensity in the dataset.. These preprocessing steps improve model generalization, improve diagnostic accuracy, and offer robustness in lung disease detection from chest X-rays.

B. Data Preprocessing and Augmentation

Data augmentation and data preprocessing are two most important components in improving the robustness and generalization capability of deep learning models, particularly for medical image-related applications such as lung disease diagnosis. Based on the variability in chest X-rays, varying lighting, varying patient pose, and noise, suitable preprocessing and augmentation methods were employed to allow the models to learn meaningful features rather than learning dataset-specific features. We have applied different preprocessing techniques to our dataset, before training the MobileNet-V1 and EfficientNet-B4 based models which are furthur optimized over their own architectures.

Through the implementation of these preprocessing and augmentation methods, we expect to enhance model generalization, prevent overfitting, and enhance lung disease detection accuracy across various chest X-ray datasets.

C. MobileNet-V1 Preprocessing

In order to ensure compatibility with the MobileNet-V1 model the training images are subjected to several preprocessing and augmentation techniques. All the images are resized to a fixed dimension of 224 × 224 pixels which is the recommended size for our base model. The pixel values are resized to [-1, 1] so as to stabilize training and accelerate convergence. Several augmentation techniques are applied in the data to introduce variations and prevent overfitting viz. random rotation up to ±30°, horizontal and vertical translations of up to 20% of the image dimensions, a shear transformation of up to 20% , random horizontal flipping is also applied. Finally, empty pixels (if any) are filled using the fill_mode='nearest' setting, which replaces empty pixels with the nearest pixel values. In the test dataset, the images are resized to 224 × 224 pixels and the pixel values are

normalized between [-1,1]. No other preprocessing or augmentation techniques are applied so as to maintain the originality of the images in testing conditions.

D. EfficientNet-B4 Preprocessing

For EfficientNetV4, the preprocessing pipeline was more advanced to align with the model's ImageNet pretraining. Input images were resized to 384×384 using bicubic interpolation, followed by random cropping to 380×380 to add some image composition variability. Selective data augmentation techniques were used, such as random resized cropping, horizontal flipping with 0.5 probability, Gaussian blurring, random affine transformations - translation and rotation ±30°, and color jittering to include contrast and brightness variability. These augmentations allow EfficientNetV4 to learn invariant features under different imaging conditions. The last operation was image-to-tensor conversion and mean-variance normalization (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]).

E. Dataset

We have worked on a Chest X-Ray image dataset, spanning various domains such as viral pneumonia, bacterial pneumonia, Coronavirus disease, and tuberculosis to ensure the generalizability and robustness of our suggested model. Exposing the model to vast pathological variations helps to achieve in-depth feature learning, finally boosting performance in varied lung diseases Lung Disease Detection Dataset: This dataset contains chest x-ray images separated into five categories mentioned below with the directory structure. The dataset is divided into three divisions namely- Test(20.06%), Train(59.97%) and Validation(19.97%).The link to the above mentioned dataset is attached herewith (Link:https://www.kaggle.com/code/debajitguha/lung-disease-new-improvised).
Table I provides an overview of our entire dataset.

TABLE I: OVERVIEW OF THE DATASET

Dataset/Category	Test	Train	Validation
Bacterial Pneumonia	403	1205	401
Coronavirus	407	1218	406
Normal	404	1207	402
Tuberculosis	408	1220	406
Viral Pneumonia	403	1204	401

VI. RESULTS AND DISCUSSION

A. MobileNet V1

The MobileNetV1 model achieved a strong performance in classifying lung diseases, with a train accuracy of 93.15%, test accuracy of 88% and an average F1-score of 0.92. The model demonstrated exceptional performance for CoronaVirus Disease achieving perfect precision and an F1 score of about 0.93. Normal cases were also accurately identified, with a recall of 97% and an F1-score of 0.95. However, the model struggled to distinguish between Bacterial Pneumonia and Viral Pneumonia, thus the precision for Bacterial Pneumonia and Viral Pneumonia is much lower around 0.76 and 0.83 respectively. Figure 3. shows the ‘accuracy and loss vs epochs’ curves of the model. Figure 5. shows the confusion matrix of MobileNet-V1 based model. Table II shows a comparison of the classification reports between the two models.

B. EfficientNet B4

The EfficientNet-B4 model achieves strong overall performance in classifying lung diseases, with a train accuracy of 95%, test accuracy of 92% and a macro F1-score of 0.92. The model demonstrates exceptional performance for CoronaVirus Disease and Tuberculosis, achieving perfect precision, recall, and F1-scores (1.00).
Normal cases are also accurately identified, with a recall of 97% and an F1-score of 0.96. However, the model struggles with distinguishing between Bacterial Pneumonia and Viral Pneumonia, as evidenced by lower recall (76%) for Bacterial Pneumonia and lower precision (77%) for Viral Pneumonia. This confusion is reflected in the misclassification of 89 Bacterial Pneumonia cases as Viral Pneumonia and 42 Viral Pneumonia cases as Bacterial Pneumonia. The results indicate that the model is highly effective for most classes but requires further refinement to improve differentiation between Bacterial and Viral Pneumonia. Figure 6. shows the ‘accuracy and loss vs epochs’ curves of the model.
Figure 4. show the confusion matrix of the EfficientNet- B4 based model. Table II shows a comparison of the classification reports between the two models.

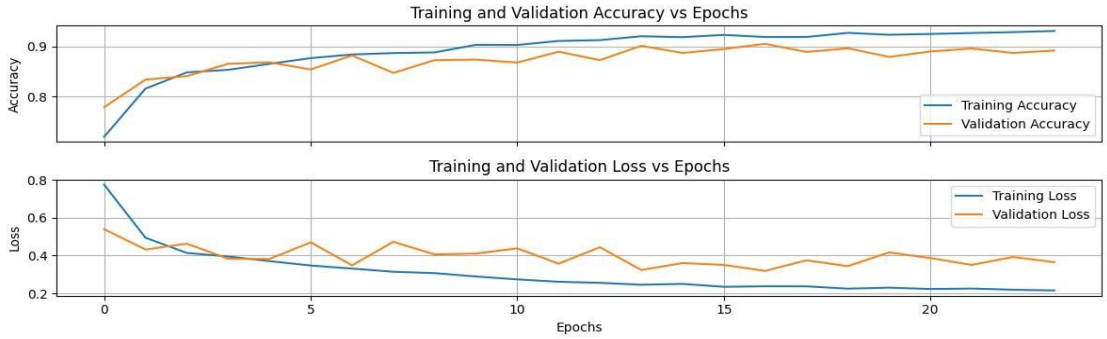

Figure 3. Accuracy Vs Loss Curve of MobileNet V1

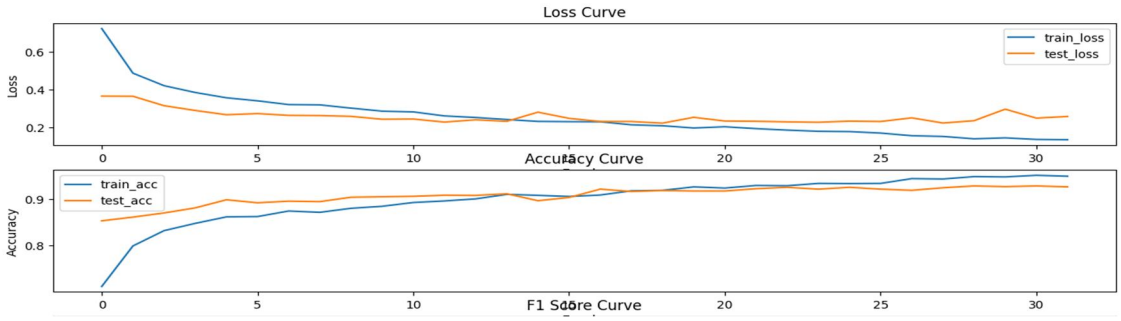

Figure 4. Accuracy Vs Loss Curve EfficientNet B4

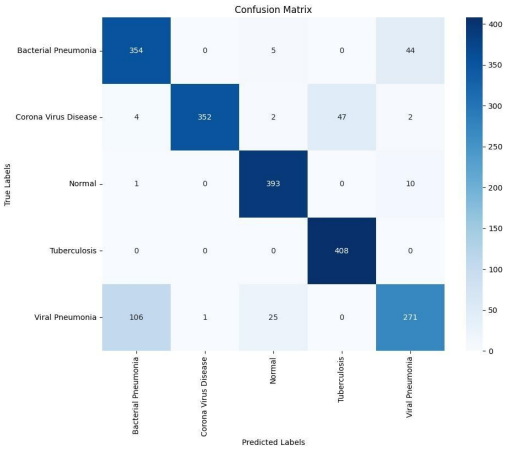

Figure 5. Confusion Matrix MobileNet V1

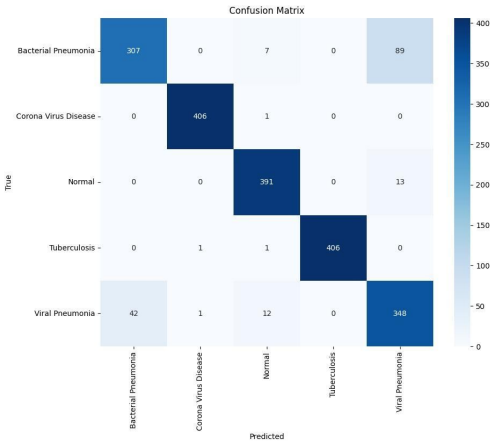

Figure 6. Confusion Matrix EfficientNet B4

TABLE II: CLASSIFICATION REPORT OF MOBILENET V1 AND EFFICIENTNET B4 MODELS

MobileNet V1					EfficientNet B4				
	Precision	Recall	F1-Score	Support		Precision	Recall	F1-Score	Support
Bacterial Pneumonia	0.76	0.88	0.82	403	Bacterial Pneumonia	0.88	0.76	0.82	403
CoronaVirus Disease	1	0.86	0.93	407	CoronaVirus Disease	1	1	1	407
Normal	0.92	0.97	0.95	404	Normal	0.95	0.97	0.96	404
Tuberculosis	0.9	1	0.95	408	Tuberculosis	1	1	1	408
Viral Pneumonia	0.83	0.67	0.74	403	Viral Pneumonia	0.77	0.86	0.82	403
Accuracy			0.88	2025	Accuracy			0.92	2025
Macro Average	0.88	0.88	0.88	2025	Macro Average	0.92	0.92	0.92	2025
Weighted Average	0.88	0.88	0.88	2025	Weighted Average	0.92	0.92	0.92	2025

VII. Conclusions and Future Work

In this study, we conducted a comprehensive evaluation of two state-of-the-art deep learning models—EfficientNet-B4, MobileNetV1 for multiclass lung disease classification. EfficientNet-B4 emerged as a more generalizable model for the medical domain, achieving a test accuracy of 92% and an average F1-score of 0.92, with perfect precision and recall for CoronaVirus Disease and Tuberculosis. MobileNetV1 also demonstrated robust performance, attaining a train accuracy of 93.15%, a test accuracy of 88%, and an average F1-score of 0.92, excelling in the classification of CoronaVirus Disease (F1-score = 0.93) and Normal cases (recall = 97%, F1-score = 0.95). However, both models exhibited limitations in distinguishing between Bacterial Pneumonia and Viral Pneumonia, as reflected in their lower precision and recall for these classes. The results underscore the potential of EfficientNet-B4 and MobileNet-V1 as robust and generalizable models for lung disease classification, particularly for critical conditions such as CoronaVirus Disease and Tuberculosis. However, the challenges in differentiating Bacterial Pneumonia and Viral Pneumonia highlight the need for further refinement. Future work should focus on advanced data augmentation, hybrid architectures, and class imbalance mitigation to enhance the model's ability to distinguish between these clinically similar conditions.
Some of the Future Work are as follows:

- Advanced Data Augmentation: Use domain-specific augmentations such as elastic deformations, CLAHE to improve differentiation between Bacterial and Viral Pneumonia.
- Class Imbalance Mitigation: Apply focal loss or class-weighted loss functions to address the imbalance and improve precision for minority classes.
- Larger and More Diverse Datasets: Expand the dataset to include more diverse cases of Bacterial and Viral Pneumonia, reducing bias and improving generalization. We can also create synthetic datasets using GAN to expand on the dataset and improve on the quality of the CT scans or CXR images.
- Hybrid Architectures: Experiment with hybrid models for example EfficientNet with Vision Transformers to leverage complementary strengths for better feature extraction.
- Explainability and Interpretability: Use Grad-CAM to visualize decision-making and ensure the model focuses on clinically relevant regions.

References

[1] Sanida, T., Sanida, M. V., Sideris, A., & Dasygenis, M. (2024). Enhancing Pulmonary Diagnosis in Chest X-rays through Generative AI Techniques. J, 7(3), 302-318.

[2] Kundu R, Das R, Geem ZW, Han G-T, Sarkar R (2021) Pneumonia detection in chest X-ray images using an ensemble of deep learning models. PLoS ONE 16(9): e0256630.

[3] Martínez Chamorro E,Díez Tascón A, Ibáñez Sanz L,Ossaba Vélez S,Borruel Nacenta S.: Radiological Diagnosis of the Patient with COVID-19. Radiología.2021;63:56-73.

[4] Dulani Meedeniya, Hashara Kumarasinghe, Shammi Kolonne, Chamodi Fernando, Isabel De la Torre Díez, Gonçalo Marques, Chest X-ray analysis empowered with deep learning: A systematic review, Applied Soft Computing, Volume 126, 2022, 109319, ISSN 1568-4946.

[5] Agughasi Victor Ikechukwu and Murali S: CX-Net: an efficient ensemble semantic deep neural network for ROI identification from chest-x-ray images for COPD diagnosis, 2023, Mach. learn.: Sci. Technol. 4 025021.

[6] Tan, Mingxing, and Quoc Le. "EfficientNet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.