**Pneumonia detection and classification using Deep Learning**

**Abstract**

Pneumonia is a serious public health problem worldwide, especially among susceptible populations. Early and precise detection is crucial to enhance patient outcomes. In this investigation, the performance of some deep learning models—VGG16, ResNet50, MobileNetV2, and EfficientNetB0—is tested for detecting pneumonia from chest X-ray images. These models were pretrained on ImageNet and fine-tuned for the task. They were assessed for their ability to classify. Among the models tested, VGG16 had the best test accuracy of 92.63%, followed by ResNet50 with 87.02%, showing the power of deep convolutional architectures in medical image analysis. These results show the promise of transfer learning methods in improving early pneumonia diagnosis, leading to quicker, more accurate clinical decision-making.

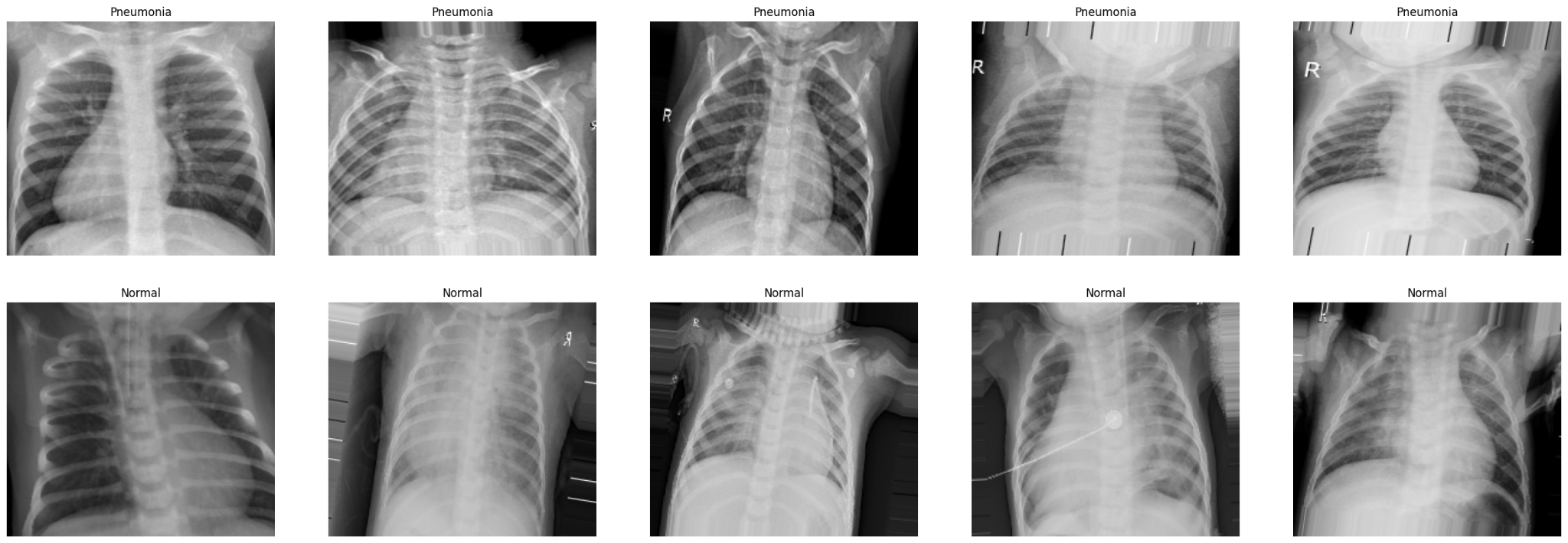
**Introduction**

Healthcare is not merely the curing of diseases; it's maintaining life and securing early interventions. In a world where medical imaging has become a mainstay in diagnostics, there is a need for effective, automated detection mechanisms more than ever before. Pneumonia, a life-threatening lung infection, needs to be diagnosed early and accurately to enhance patient outcomes and minimize mortality. Conventional diagnostic techniques tend to rely predominantly on human interpretation of chest X-rays, which may be subjective, time-consuming, and vulnerable to human error.

Deep Learning (DL) models provide a more intelligent and scalable option by learning the patterns for pneumonia automatically from large sets of X-ray images. DL models have the capability to identify subtle abnormalities that may not be visible to the naked eye and improve their accuracy progressively with additional data.

In our suggested pneumonia detection model, we organized the chest X-ray dataset into three directories: train, validation, and test with 5216, 16, and 624 images respectively, divided into two categories: Pneumonia and Normal. In order to increase the variety of our training data and decrease overfitting, we have used data augmentation methods like random rotation, shift, flip, and zoom transform. This was of great help to the models in being able to generalize to unseen data.

We utilized transfer learning, building upon strong pre-trained convolutional neural networks (CNNs) such as VGG16, ResNet50, MobileNetV2, and EfficientNetB0 with ImageNet weights. We removed the top layers (include\_top=False) and appended our own fully connected layers for our binary classification problem. During training, we first froze the convolutional base to preserve the learned feature and trained the top layers only. In later stages, fine-tuning was performed by unfreezing a few deeper layers and retraining with a lower learning rate, allowing the model to adapt ImageNet features to the chest X-ray domain.



Of the models validated, VGG16 recorded a best test accuracy of 92.63%, seconded by ResNet50 with 87.02%, meaning that wide and deep structures such as VGG16 work particularly well at describing the rich visual patterns of pneumonia on X-rays. Such results show deep learning-based automation has the ability to assist health practitioners, yielding faster, more accurate diagnoses, and enhancing decision-making in healthcare for resource-scarce contexts.

**Literature Review**

Vanchada Fernandes et al. [1]

Fernandes et al. suggested a robust system for COVID-19 screening from chest X-ray images, tackling some of the major challenges. First, there are only a few images available, which complicates the training of deep learning models. Second, infection may be distributed or localized in just a part of the lung, complicating detection. Their research focused on the application of transfer learning as a solid method of taking advantage of pre-trained models and making correct predictions even from a small amount of data. They explained that the infections could appear in various zones of the lungs, and in other instances, be confined to a single lung, making it harder to detect. Their study implies that contamination in one lung only can signify less severe infections, while bilateral involvement may signal serious cases. They found that transfer learning is not only possible but also a trustworthy approach for small datasets, particularly in severe healthcare applications such as COVID-19 diagnosis.

Hongzan Lu et al. [2]

Hongzan Lu and others presented the key contribution of Convolutional Neural Networks (CNNs) to enhance image classification, particularly when dealing with limited COVID-19 datasets. They pointed out that even with the small number of images, CNN-based models can produce outstanding performance. In addition, they emphasized the large heterogeneity of pulmonary patterns in infected patients, rendering diagnosis difficult. Their study was aimed at applying CNNs to accurately differentiate COVID-19, pneumonia, and normal conditions from chest X-rays, offering an extensible and feasible solution to clinical diagnosis, particularly when traditional resources are limited.

Sammy T. Vithana et al. [3]

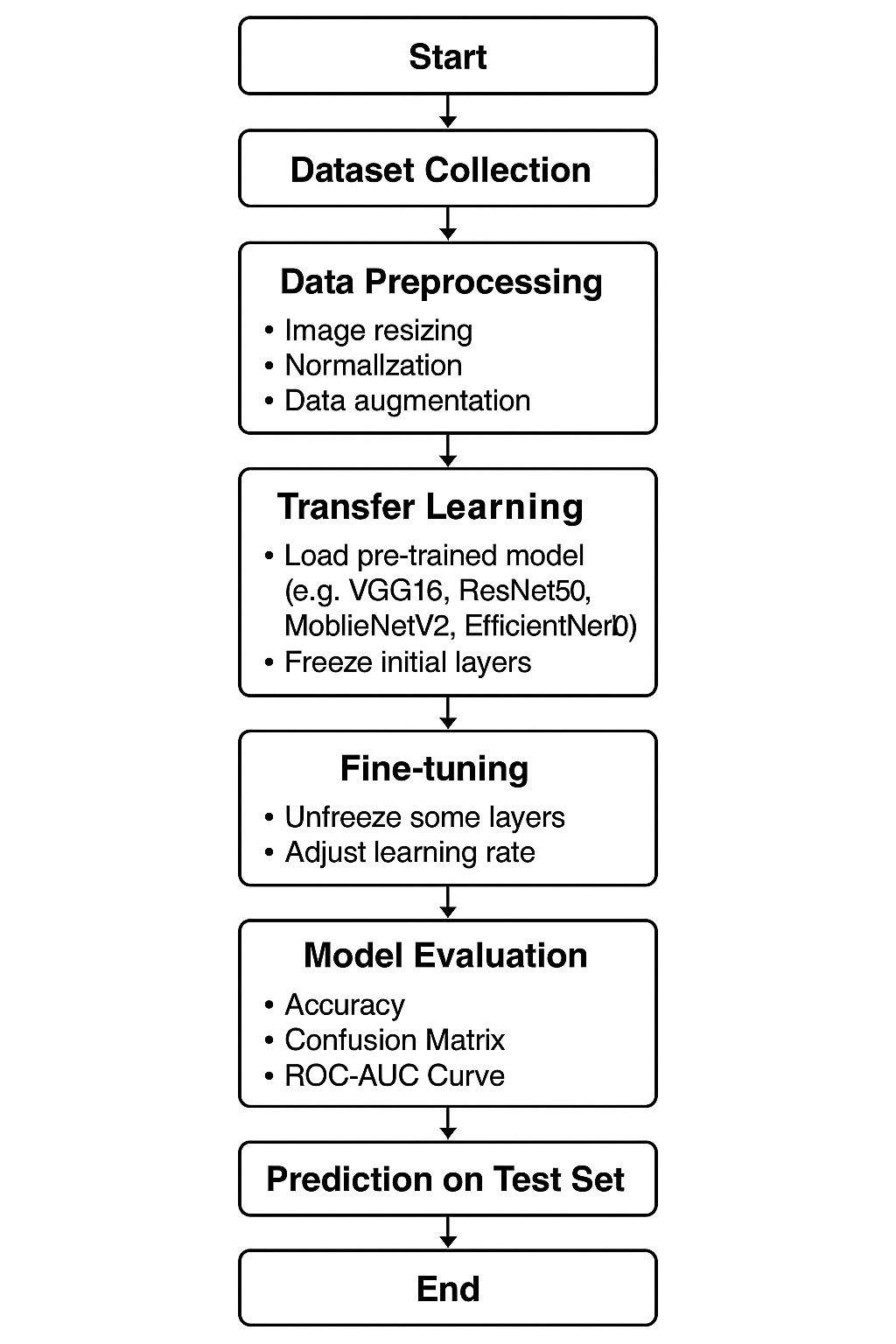
Vithana et al. suggested a robust and efficient framework that combines multiple CNN models for anomaly detection in chest X-ray images. Their method focuses on transfer learning coupled with ensemble techniques, enabling the models to stay strong even if images are affected by non-disease factors like imaging noise or varying acquisition methods. Their framework was able to identify whether the patient's condition was actually disease-induced or was due to imaging noise, hence enhancing diagnostic dependability.

Further Research: AlexNet, VGG16, MobileNet, ResNet50

Some research has utilized widely used pre-trained CNN models like AlexNet, VGG16, MobileNet, and ResNet50 for pneumonia and COVID-19 diagnosis. In a research study, the dataset of 25,000 images with the resolution of 224×224 pixels was employed, proving that these architectures are capable of producing high performance even when data is augmented or in short supply. MobileNet and ResNet models, specifically, demonstrated outstanding performance with training accuracies greater than 90% and robustly tolerating variations in input data. Fine-tuning deeper layers enhanced detection rates across various pneumonia subtypes even further, indicating model adaptation to be a key aspect of achieving clinical-grade performance.

**Methodology**

This chapter outlines the methodologies used to create a pneumonia detection system based on deep learning methods. The project makes use of transfer learning with pre-trained convolutional neural networks (CNNs) to classify chest X-ray images into "Normal" and "Pneumonia" classes. The steps involved are dataset preparation, data preprocessing, model selection, training, fine-tuning, evaluation, and testing.

****

**Dataset Collection**

The dataset employed in this project, chest X-ray, was obtained from Kaggle and hosted on Google Drive to be accessible within the Colab environment.

The dataset is segregated into three subsets:

Training Set: For model training.

Validation Set: For hyperparameter adjustment and prevention of overfitting.

Test Set: For the ultimate model testing.

All subsets have two classes: Normal and Pneumonia. This organized structure allows for effective training and testing of models.

**Data Preprocessing**

To improve input image quality and learning procedure, preprocessing operations were used, such as resizing, normalization, and data augmentation. All the images were resized to 224×224 pixels, the typical size for pre-trained models of CNNs like VGG16, ResNet50, MobileNetV2, and EfficientNetB0. Pixel values were rescaled from the [0, 255] range to [0, 1] through the application of a rescaling factor (rescale=1./255) that accelerates convergence during training (Goodfellow, Bengio, & Courville, 2016). Normalizing ensures that the network is not plagued by exploding or vanishing gradients that slow down or even hinder training. To avoid overfitting and enhance model generalization, data augmentation processes were utilized, such as random zoom up to 20%, horizontal flip, and random rotation up to 10 degrees, all done using TensorFlow's ImageDataGenerator class. These augmentations synthetically enlarge the dataset by creating new versions of the original images, hence enabling the model to learn more stable features. This method improves the model's capacity to generalize to new data, making it perform better on real-world tasks. Additionally, these methods assist in mimicking real-world variations such as lighting, orientation, and scale, which the model may face in real-world applications.

**Transfer Learning**

Transfer learning was used to exploit the representational strengths of CNNs pre-trained on the ImageNet database (Deng et al., 2009). The pre-trained models listed below were used: VGG16 (Simonyan & Zisserman, 2015), ResNet50 (He et al., 2016), MobileNetV2 (Sandler et al., 2018), and EfficientNetB0 (Tan & Le, 2019). Each model was set with `include\_top=False` so that the last fully-connected layers of the original models are not included so that custom classification layers can be added. In this way, the model remains with the features learned while gaining the ability to adapt it according to the required task. Additionally, all of the convolutional layers were also frozen in the beginning (`trainable=False`) so the pre-trained features are saved during training. Only the added head of classification was trained, to avoid updating the lower layers unnecessarily and hence accelerate convergence and save against overfitting. The approach enables the model to benefit from the powerful feature representations learnt on ImageNet while concentrating training efforts on the task-specific output layers.

**Fine-tuning**

Fine-tuning was conducted to further fine-tune the models to the particular task of pneumonia detection. To this end, a few of the deeper layers were unfrozen so that they could learn the specific characteristics of chest X-rays. The learning rate was lowered to 1e-4 to avoid catastrophic forgetting, whereby the model did not forget important knowledge gained in the pre-training stage while continuing to learn from the new task. Fine-tuning allowed the model to specialize in the pneumonia detection patterns unique to chest X-rays, greatly improving its performance. This is a common practice when fine-tuning pre-trained models for domain-specific tasks since it enables the model to use pre-existing knowledge while fine-tuning its capacity to identify task-relevant features (Howard & Ruder, 2018). By defrosting the lower layers, the model becomes more adaptable, refining its encoding of features that are essential for separating healthy from abnormal lung patterns, ultimately enhancing its diagnostic accuracy.

**Model Architecture**

A custom classification head was appended to every pre-trained model to make it pneumonia detection-specific. This head consisted of a Global Average Pooling Layer, which compressed the spatial size of the feature maps and transformed them into a 1D vector, lowering computational complexity. A Dense Layer with 128 units and ReLU activation was appended to extract intricate relationships between the learned features. To avoid overfitting, a Dropout Layer with 50% dropout was added, so the model learned more generalized features. Lastly, a Dense Layer with one unit and sigmoid activation was included for binary classification, which gave a probability score between 0 and 1 for pneumonia. These additions helped the model become specialized in detecting pneumonia from chest X-rays while reducing overfitting and enhancing classification performance.

**Model Training and Evaluation**

The models were trained with the Adam optimizer and binary cross-entropy as the loss function. The batch size was 32, and training was for 10–15 epochs, with early stopping applied to avoid overfitting if the validation loss plateaued. Early stopping helped ensure that the model did not keep training when performance on the validation set plateaued, thereby helping to maintain generalization.

Following training, all models were tested using a variety of metrics. Accuracy was computed to measure the proportion of correct predictions. The Confusion Matrix was applied in order to examine true positives, true negatives, false positives, and false negatives, gaining greater insight into how well the model is performing. The ROC-AUC Curve was also investigated to measure the model's capacity to differentiate between classes. These measures provide a holistic representation of the model's performance, which is particularly important in medical diagnosis tasks, where false negatives can have life-altering repercussions

**Result & Analysis**

In this study, four popular deep learning architectures — VGG16, ResNet50, MobileNetV2, and EfficientNetB0 — were evaluated for the task of pneumonia detection using chest X-ray images.

Initially, all models were trained for 5 epochs to compare their base performances. The training and validation losses of these models are shown in Figure 1. It can be observed that the VGG16 model demonstrates a steady reduction in both training and validation loss, indicating good learning behavior and minimal overfitting. MobileNetV2 also showed reasonable stability, while ResNet50 and EfficientNetB0 exhibited higher validation losses compared to their training losses, suggesting overfitting.

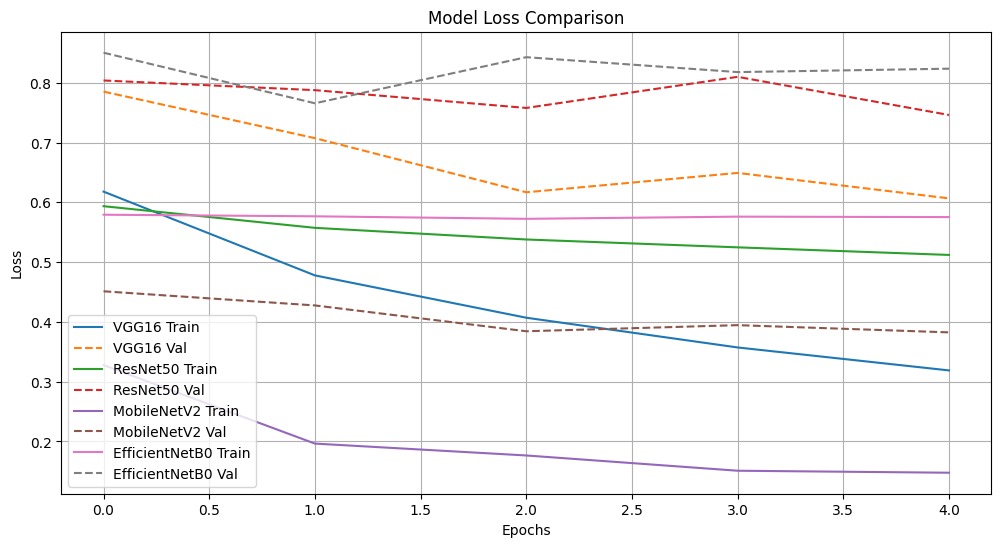


Figure 1: Model Loss Comparison over 5 epochs.

Similarly, the accuracy trends over the epochs are depicted in Figure 2. VGG16 achieved the highest training and validation accuracy progression, followed by ResNet50 and MobileNetV2. EfficientNetB0 exhibited poor performance, with validation accuracy stagnating at around 50%.

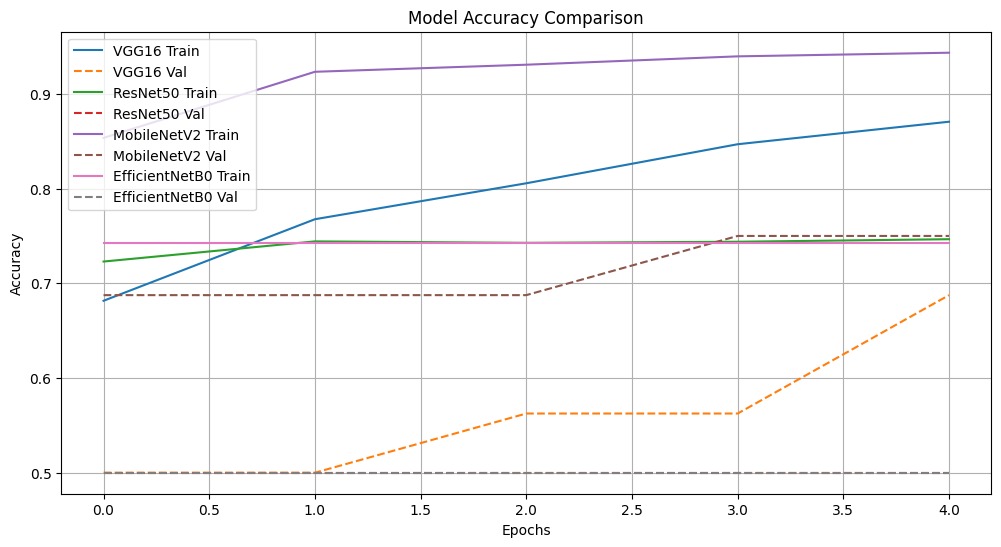


Figure 2: Model Accuracy Comparison over 5 epochs.

A direct comparison of the final validation accuracies achieved by each model is presented in Figure 3. As shown, VGG16 achieved the best validation accuracy of 93%, followed by ResNet50 (87%), MobileNetV2 (75%), and EfficientNetB0 (48%).

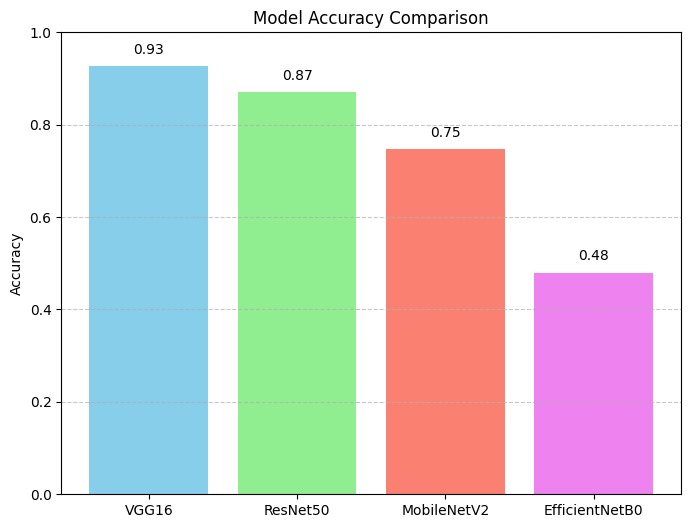


Figure 3: Final Testing Accuracy of Different Models.

Based on these initial results, the VGG16 model was selected for further fine-tuning. In the fine-tuning phase, the model’s top layers were unfrozen, and the entire network was trained for an additional 27 epochs. The training and validation accuracy curves for this fine-tuning stage are shown in Figure 4. It can be observed that the model achieved a final training accuracy of approximately 98% and a validation accuracy of approximately 97%, indicating excellent generalization without significant overfitting.

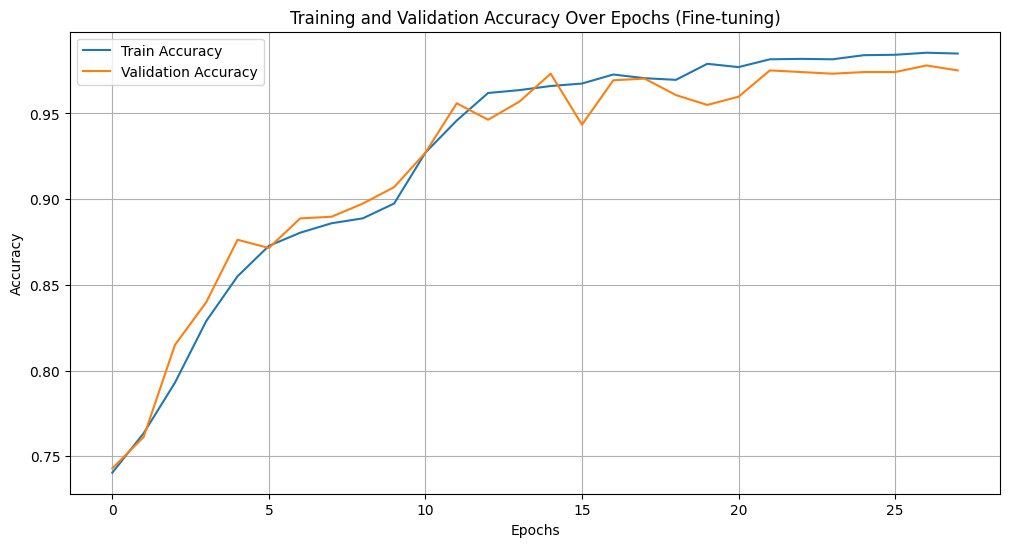


Figure 4: Training and Validation Accuracy of VGG16 during Fine-tuning.

The final comparative accuracies are summarized in Table 1:

| Model | Testing Accuracy |
| --- | --- |
| VGG16 | 93% |
| ResNet50 | 87% |
| MobileNetV2 | 75% |
| EfficientNetB0 | 48% |

Table 1: Testing Accuracy Comparison of Different Models.

These results clearly establish that VGG16 outperformed all other models tested and is best suited for pneumonia detection using chest X-ray images among the models considered**.**

**Conclusion**

This research investigated the use of deep learning models — VGG16, ResNet50, MobileNetV2, and EfficientNetB0 — for the automated detection of pneumonia from chest X-ray images. By means of systematic data preprocessing, transfer learning, fine-tuning, and strict evaluation, the performances of the models were evaluated to ascertain their suitability in a real-world healthcare environment. Of the models tested, VGG16 performed best, with a final validation accuracy of 93%, followed by ResNet50 at 87%. Further fine-tuning enhanced the performance of VGG16, resulting in about 98% training accuracy and 97% validation accuracy with good generalization without major overfitting. These results highlight the utility of using pre-trained convolutional neural networks, most notably VGG16, to detect pneumonia at an early and accurate stage. By incorporating such deep learning methodologies into clinical practices, healthcare systems are able to deliver quicker, more uniform, and more accurate diagnoses, thereby enhancing patient outcomes, particularly in resource-limited settings. For future research, the datasets may be increased, more sophisticated augmentation techniques may be added, and ensemble models can be tested in order to make the diagnosis more reliable.

**References**

[1] Vandecia Fernandes et al., "Bayesian convolutional neural network estimation for pediatric pneumonia detection and diagnosis", Computer Methods and Programs in Biomedicine, Elsevier, 2021

[2] Hongen Lu et al., "Transfer Learning from Pneumonia to COVID-19", Asia-Pacific on Computer Science and Data Engineering (CSDE), 2020 IEEE

[3] Sammy V. Militante et al., "Pneumonia and COVID-19 Detection using Convolutional Neural Networks", 2020 the third International on Vocational Education and Electrical Engineering (ICVEE), IEEE, 2021

[4] Nanette V. Dionisio et al., "Pneumonia Detection through Adaptive Deep Learning Models of Convolutional Neural Networks", 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC 2020), 8 August 2020

[5] Md. Jahid Hasan et al., "Deep Learning-based Detection and Segmentation of COVID-19 & Pneumonia on Chest X-ray Image", 2021 International Information and Communication Technology for Sustainable Development (ICICT4SD), 27-28 February 2021