Comparative Analysis of Deep Learning Models for Pneumonia Detection

Keshav Lodha, Oscar Wang, Jason Nyachhayon, Liew Saefong

Fall 2024

1 Introduction, Motivation, & Problem

The manual analysis of medical images, such as chest X-rays, is a time-consuming and error-prone process that can delay diagnoses and lead to suboptimal patient outcomes. With advancements in technology and computer vision, we aim to address these challenges by developing and evaluating models capable of accurately detecting pneumonia in chest X-ray images. These models can assist healthcare professionals in making timely and precise diagnoses, especially in areas with limited access to specialized expertise. Automating medical image analysis offers significant benefits. It reduces diagnostic time, allows healthcare providers to focus on critical tasks, and enhances diagnostic accuracy by identifying subtle patterns that may elude human experts. Early detection of pneumonia is particularly vital for initiating prompt treatment and improving patient outcomes.

Our project not only seeks to improve diagnostic efficiency but also aims to alleviate the burden on healthcare systems, reduce costs, and improve patient satisfaction. We will analyze and compare the performance of ResNet50, MobileNet V2, and ConvNeXT models to identify the most effective approach. By contributing to the development of reliable and accessible diagnostic tools, this work supports broader societal goals and lays the groundwork for future advancements in medical image analysis.

2 Dataset

The data set used in this study is described in the paper 'Deep Learning-Based Classification and Referral of Treatable Human Diseases.' We are using Chest X-ray images, which is divided into training and testing sets, with each set consisting of images categorized as NORMAL or PNEUMONIA. There are approximately 1500 NORMAL images and 3500 PNEUMONIA images for each set. The NORMAL images have a resolution of 1857×1317 pixels while the PNEUMONIA images have a resolution of 1304×968 pixels. We loaded the images into the code and split the sets into training, validation, and testing sets. Then, we resized the images to 224×224 pixels and applied a center crop to focus on the most important part of the image. Finally, the pixel values are normalized with a mean of 0.5 and standard deviation of 0.5. These sequence of transformations ensures that the images are consistent with each testing and training of the models.

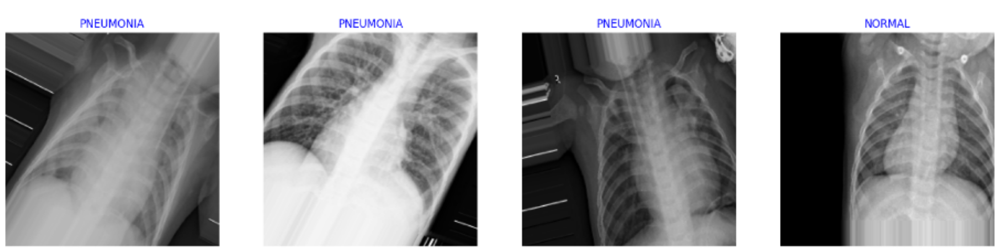


Figure 1: Example images from our dataset

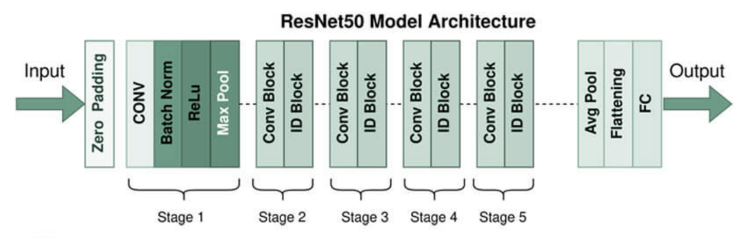
3 Methodology

3.1 Models

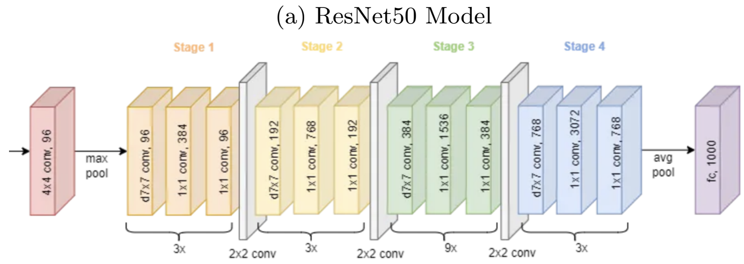
In this project, we analyze the performance of three computer vision models—ResNet50, MobileNet V2, and ConvNeXt—for detecting pneumonia in chest X-ray images.

- **ResNet50**: ResNet50 (Residual Network) introduces the concept of residual learning to mitigate the vanishing gradient problem in deep neural networks. It consists of 50 layers, including convolutional, pooling, and fully connected layers, and uses skip connections to allow gradients to flow directly through the network. This design enables the model to train deeper networks without performance degradation. ResNet50 is an excellent choice for a classification project due to its proven ability to learn complex features in images while maintaining high accuracy across various datasets. **We will be using the performance of the ResNet50 model as the baseline to compare all other model performances.**

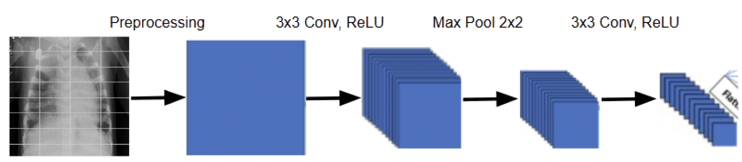
- **MobileNet V2**: MobileNet V2 is designed for resource-constrained environments, prioritizing efficiency and speed. It employs depthwise separable convolutions to reduce the number of parameters and computational complexity. Additionally, it introduces inverted residuals and linear bottleneck layers, which enhance computational efficiency while maintaining accuracy. MobileNet V2 is ideal for a classification project because it balances performance and computational efficiency, making it suitable for deployment in real-world medical applications, including mobile or edge devices.
- **ConvNeXT**: ConvNeXT is a modern adaptation of convolutional networks inspired by transformer architectures. It incorporates design elements like depthwise convolutions, layer normalization, and carefully tuned architectural hyperparameters to achieve state-of-the-art performance. ConvNeXt combines the simplicity of convolutional networks with the efficiency of transformers, making it highly effective for image classification. Its ability to capture fine-grained features and handle complex image data makes it a strong choice for accurately classifying pneumonia in medical images.



(a) ResNet50 Model



(b) ConvNeXT Model



(c) MobileNet V2 Model

Figure 2: Architectures for the different models trained

3.2 Data Augmentation & Fine Tuning

Since each model was initially trained with a large dataset, we needed to fine tune the models to recognize the complex patterns in our chest X-ray images. To diversify our dataset, we applied data augmentation before testing and training the model. These included random cropping, flipping, and adjustments to brightness, contrast, and saturation to simulate real-world variations in data.

Several fine-tuning techniques were applied to improve the performance of each model. First, the later layers of each pre-trained model were unfrozen to allow for more targeted adaptation to our dataset. We froze earlier layers to preserve the low-level features gained from initial training with ImageNet. In the classifier layer, we added dropout for regularization and batch normalization to stabilize training. We experimented with several learning rates and finalized on 0.001.

3.3 Our Approach

We began our experiments with the ResNet50 model, using a set of predefined transformations to preprocess the dataset. The training process resulted in extremely high training accuracy, indicating that the model was able to learn the patterns in the training data effectively. However, as shown in Figure 3, the test accuracy fluctuated significantly across epochs. This suggests overfitting, where the model performed well on the training set but struggled to generalize to unseen data. Ideally, we would expect both the training and test accuracies to steadily increase with minimal divergence between the two, indicating a well-generalized model.

To address the overfitting observed in the initial training of ResNet50, we applied fine-tuning techniques. Fine-tuning involved unfreezing specific layers and adjusting hyperparameters such as learning rates. As shown in Figure 4, this approach reduced the overfitting effect, as evidenced by a more stable test accuracy. The training accuracy remained high, while the test accuracy showed a trend of improvement, suggesting better generalization. Ideally, we expect fine-tuning to mitigate overfitting by striking a balance between training and testing performance.

We next trained the ConvNeXt model using the same set of transformations. As depicted in Figure 5, this model demonstrated strong performance, with a steadily decreasing training loss and a consistently increasing test accuracy across epochs. This behavior aligns closely with our ideal expectations: the model efficiently learns the training data while generalizing well to the test set. Given these promising results, we deemed additional fine-tuning unnecessary, as the model already exhibited excellent generalization capabilities.

For MobileNetV2, the initial training process displayed patterns similar to ResNet50, as shown in Figure 6. The model achieved high training accuracy, but the test accuracy fluctuated significantly, again pointing to overfitting. Ideally, we would expect a stable and increasing test accuracy alongside high training accuracy, indicating good generalization. However, the divergence between the two metrics suggested that the model required additional fine-tuning.

To improve the MobileNetV2 model, we applied fine-tuning techniques, similar to ResNet50. As seen in Figure 7, fine-tuning helped stabilize the test accuracy and reduced overfitting, with the model showing improved generalization. The training accuracy remained high, while the test accuracy exhibited a more consistent upward trend. This behavior is closer to the ideal, where fine-tuning balances both training and test performance, allowing the model to generalize better.

4 Results

4.1 Model Performances

In this section, we present the results of our experiments and compare the performance of ResNet50, ConvNeXt, and MobileNetV2 models. By analyzing metrics such as loss, training accuracy, and test accuracy, we aim to determine the most effective model for detecting pneumonia in chest X-ray images.

Model	Loss	Test Accuracy (%)
ResNet50 (fine-tuned)	0.2799	91.03
ConvNeXt	0.222	92.15
MobileNetV2 (fine-tuned)	0.3063	83.97

Table 1: Comparison of model performance statistics.

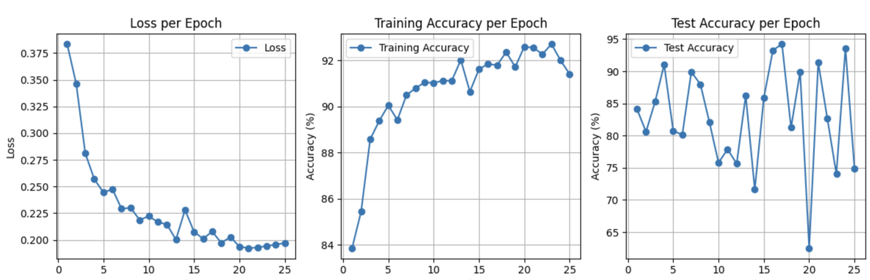


Figure 3: ResNet50 without fine-tuning

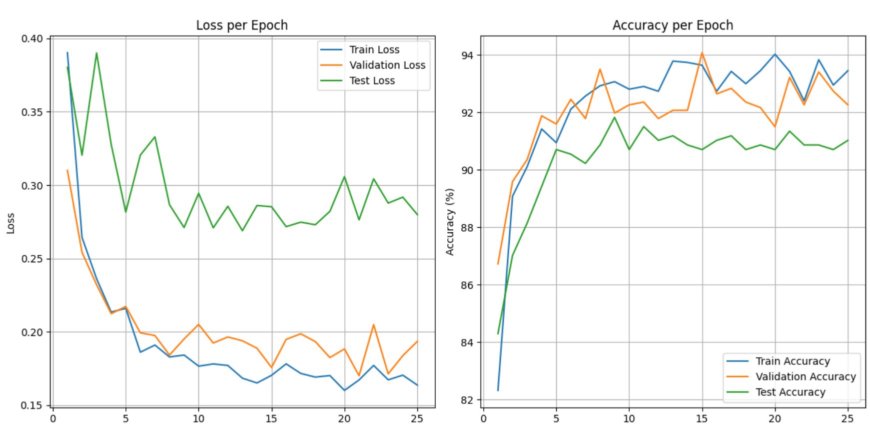


Figure 4: ResNet50 with fine-tuning

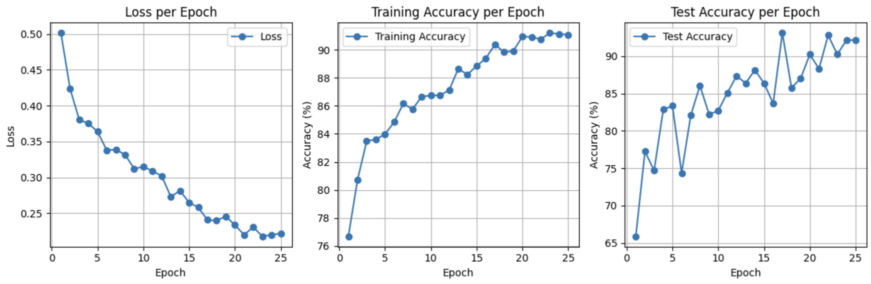


Figure 5: ConvNeXT with transformations

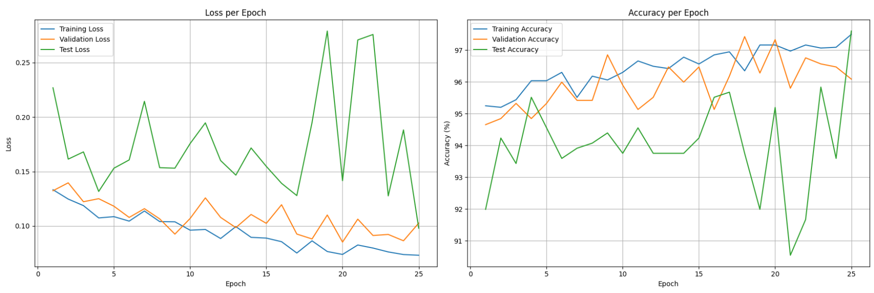


Figure 6: MobileNet V2 without fine-tuning

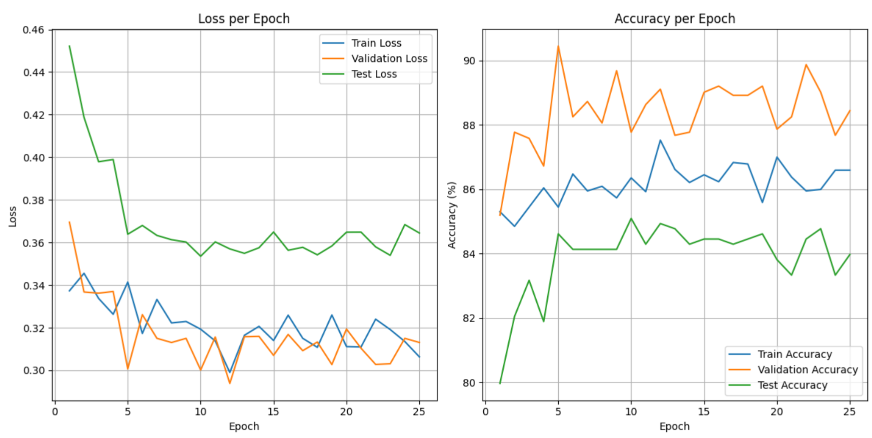


Figure 7: MobileNet V2 with fine-tuning

4.2 Final Thoughts

In this project, we evaluated the performance of ResNet50, ConvNeXt, and MobileNetV2 for pneumonia detection from chest X-ray images. Using data augmentation and fine-tuning techniques, we sought to improve the generalization ability of these pre-trained models. Among the three, ConvNeXt emerged as the best performer, achieving the highest test accuracy of 92.15% with a steadily decreasing loss and consistent performance across epochs. ResNet50, while demonstrating competitive performance with a test accuracy of 91.03% after fine-tuning, exhibited initial signs of overfitting that required additional adjustments. MobileNetV2, designed for computational efficiency, achieved a test accuracy of 83.97% but struggled to capture the intricate patterns in X-ray images, reflecting its trade-off between efficiency and feature extraction. Overall, ConvNeXt's robustness and generalization make it the most suitable model for this task, offering a reliable approach to automating pneumonia detection in medical imaging. Future work could explore larger, domain-specific datasets and additional fine-tuning techniques to further improve performance and scalability.

4.3 Future Improvements

A potential improvement would be to train the models with a larger and more diverse dataset, which could help them generalize better across a wider range of images. Furthermore, fine-tuning more layers and exploring other model architectures could further improve accuracy. Although all three models were initially trained on the large ImageNet dataset, training and testing on a specific data set for medical images would likely improve their performance in the medical imaging domain.

5 Team Member Contributions

- Keshav Lodha: Model training, report writing, slide preparation, final presentation
- Oscar Wang: Model training, report writing, slide preparation, final presentation
- Jason Nyachhayon: Model training, report writing, slide preparation, final presentation
- Liew Saefong: Model training, report writing, slide preparation, final presentation

6 References

1. Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification", Mendeley Data, V2, doi: 10.17632/rscbjbr9sj.2

2. D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan and A. Mittal, "Pneumonia Detection Using CNN based Feature Extraction," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-7, doi: 10.1109/ICECCT.2019.8869364.

3. Alapat DJ, Menon MV, Ashok S. A Review on Detection of Pneumonia in Chest X-ray Images Using Neural Networks. J Biomed Phys Eng. 2022 Dec 1;12(6):551-558. doi: 10.31661/jbpe.v0i0.2202-1461. PMID: 36569568; PMCID: PMC9759647.

4. Liu, Z. et al. (2022) A convnet for the 2020s, arXiv.org. Available at: https://arxiv.org/abs/2201.03545

5. Sandler, M. et al. (2019) MobileNetV2: Inverted residuals and linear bottlenecks, arXiv.org. Available at: https://arxiv.org/abs/1801.04381