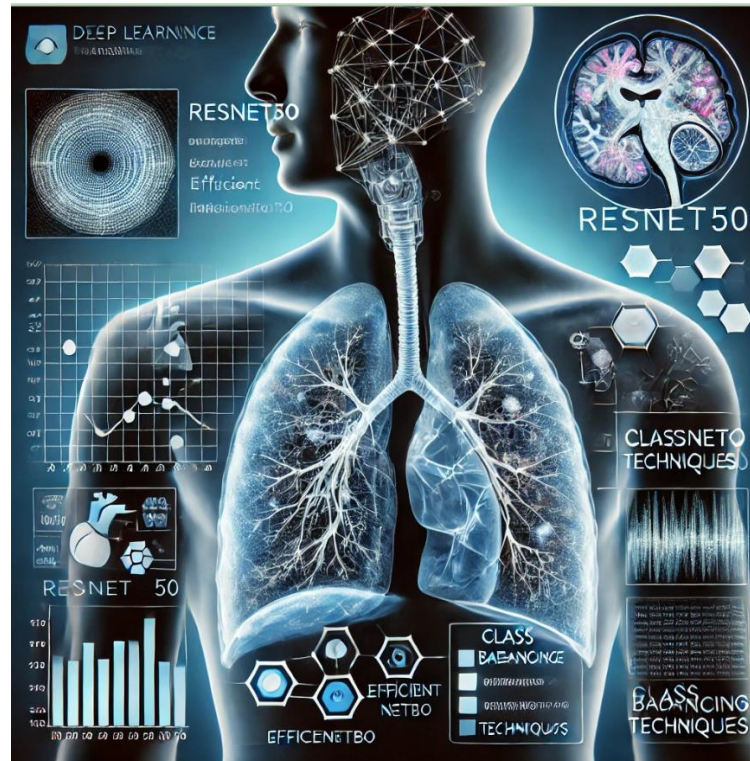# AI-Driven Diagnosis of Pneumonia Using Chest X-Ray Imaging: Comparative Analysis of EfficientNetB0 and ResNet50 Architectures with Class Balancing Techniques

Image Credit: *"AI-driven Pneumonia Diagnosis using Chest X-Ray Imaging," generated using DALL·E, OpenAI, 2024.*

By: Musiliu Bello

Computer Scientist, Data Scientist, and AI Enthusiast

Research competition:
Ready Tensor's **Advancing Visual Intelligence: From Core CV to Multi-Modal AI**
Computer Vision Projects Expo 2024

Date of Submission:
December 2024.

Abstract

Pneumonia remains a significant global health concern, necessitating accurate diagnostic methods. This study evaluates the performance of two deep learning architectures-ResNet50 and EfficientNetB0-on pneumonia classification using chest X-ray images. Experimental results show that while both models effectively identified "Normal" cases, their performance in detecting pneumonia was suboptimal, with ResNet50 achieving a slightly higher AUC of 0.55 compared to EfficientNetB0's 0.50. Challenges such as class imbalance, limited dataset size, and the complexity of medical imaging hindered their ability to learn discriminatory features effectively.

To address these limitations, techniques such as data augmentation, class balancing (e.g., SMOTE), and domain-specific fine-tuning are recommended. Regularization methods are essential to mitigate overfitting in ResNet50, while alternative architectures and ensemble approaches may improve robustness. Evaluation metrics like precision, recall, and AUC are emphasized as more reliable indicators of model performance in imbalanced datasets.

This research highlights the limitations of current deep learning models in pneumonia detection and provides a foundation for future studies focused on improving performance for real-world medical diagnostics.

Keywords: *Pneumonia detection, deep learning, ResNet50, EfficientNetB0, chest X-ray, medical image classification, AUC, data augmentation, class imbalance, SMOTE, fine-tuning, regularization, model evaluation, deep learning architectures, medical diagnostics.*

Acknowledgments

I would like to express my sincere gratitude to Ready Tensor for organizing the "Advancing Visual Intelligence: From Core CV to Multi-Modal AI" competition and providing the platform to showcase innovative computer vision projects. This research benefited greatly from the resources and inspiration made available through this global event. I am particularly thankful for the opportunity to contribute to the ongoing advancement of visual AI and to collaborate with like-minded researchers and practitioners in the field.

I would also like to acknowledge the contributions of my professional colleagues and like-minded researchers, whose valuable feedback and encouragement played a crucial role throughout the development of this work. Their insights were instrumental in refining the methodologies and shaping the interpretations presented in this research.

Lastly, I extend my appreciation to the community of computer vision experts whose work has inspired and guided the development of this project. I am hopeful that this research, alongside the collective contributions showcased in the competition, will help shape the future of AI-driven medical diagnostics and visual intelligence.

Table of Contents

List of Figures

1.0 Introduction

Pneumonia remains one of the leading causes of morbidity and mortality globally, disproportionately affecting vulnerable populations such as children and the elderly. Despite advancements in healthcare, timely and accurate diagnosis remains critical for improving patient outcomes. Chest X-ray imaging is a cornerstone diagnostic tool for identifying pneumonia; however, its interpretation can be challenging. The manual review of X-rays is often time-consuming, prone to human error, and subject to inter-observer variability among radiologists (Lakhani & Sundaram, 2017). These limitations underscore the need for innovative solutions to support clinical decision-making.

The emergence of artificial intelligence (AI) and deep learning technologies has revolutionized medical image analysis, offering rapid and reliable diagnostic insights. Among these technologies, convolutional neural networks (CNNs) have demonstrated remarkable efficacy in recognizing patterns and abnormalities within medical imaging datasets. This research leverages two state-of-the-art CNN architectures - EfficientNetB0 and ResNet50 - to automate pneumonia detection using chest X-ray images. By addressing common challenges such as dataset skewness, overfitting, and generalizability through techniques like class weighting and dataset balancing, this study enhances the robustness of AI-driven diagnostics.

Furthermore, the research evaluates and compares the performance of these architectures using critical metrics, including accuracy, sensitivity, specificity, and the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. By doing so, it aims to identify models that are not only highly accurate but also clinically deployable. The integration of AI into pneumonia detection workflows has the potential to significantly augment the diagnostic capabilities of healthcare professionals, ultimately reducing diagnostic delays and improving patient care outcomes.

1.1 Motivation

The growing global healthcare burden posed by pneumonia and the shortage of radiologists in low-resource settings underline the need for scalable AI solutions. Studies have shown that pneumonia-related deaths could be significantly reduced through timely diagnosis and treatment (World Health Organization, 2021). However, radiological expertise is often unavailable in underserved areas, creating diagnostic delays and impacting patient care.

The motivation for this study stems from:

1. **Improved Accuracy**: Leveraging state-of-the-art AI models to enhance the precision of pneumonia diagnosis.

2. **Fair Representation**: Addressing data imbalance issues to ensure unbiased and equitable model performance.

3. **Comparative Insights**: Evaluating and contrasting two leading CNN architectures to identify the most effective model for real-world deployment.

4. **Accessible Solutions**: Designing AI models that can be implemented in resource-limited settings for widespread healthcare access.

By aligning technology with healthcare needs, this research aims to create a framework that not only improves diagnostic accuracy but also makes it accessible to underprivileged regions.

1.2 Objectives
The primary objectives of this research are outlined as follows:

1. To develop AI models using EfficientNetB0 and ResNet50 architectures for pneumonia detection.

2. To address dataset imbalance by employing class weighting and augmentation techniques.

3. To compare the diagnostic performance of the two models using standard evaluation metrics.

4. To create engaging and interactive visualizations for better interpretation of the model results.

5. To provide a reproducible framework for future AI-driven medical imaging research.

6. To develop an accessible web-based tool that leverages advanced deep learning models, ResNet50 and EfficientNetB0, for pneumonia diagnosis through chest X-ray analysis.

7. To identify potential challenges and recommend solutions for deploying these models in clinical settings.

1.3 Significance

This study contributes to the growing field of AI in healthcare by addressing a critical gap in pneumonia diagnosis. Its significance lies in:

- **Scalability**: Developing solutions that can operate across diverse healthcare settings, from urban hospitals to rural clinics.

- **Bias Mitigation**: Implementing techniques to handle class imbalances, ensuring that models perform well on both minority (normal) and majority (pneumonia) classes.

- **Benchmarking**: Providing a comprehensive comparison of EfficientNetB0 and ResNet50, two widely used CNN architectures, in the context of medical image analysis.

- **Clinical Relevance**: Enhancing the interpretability and trustworthiness of AI models, which is crucial for their acceptance among healthcare professionals.

- **Public Health Impact**: Accelerating diagnosis, reducing the workload of radiologists, and enabling faster treatment for patients.

By integrating robust methods and tools, the research bridges the gap between AI innovations and their practical application in life-saving scenarios.

2.0 Related Works

The use of deep learning for medical image classification, particularly pneumonia detection, has gained significant attention due to advancements in Convolutional Neural Networks (CNNs). **ResNet (He et al., 2016)** and **EfficientNet (Tan & Le, 2019)** are two popular architectures that demonstrate high accuracy across various image recognition tasks. ResNet50, a variant of ResNet with 50 layers, introduced skip connections to alleviate vanishing gradient issues, enabling deeper networks to converge effectively (He et al., 2016). In contrast, EfficientNet focuses on balancing network depth, width, and resolution to achieve optimal performance with fewer parameters, making it computationally efficient (Tan & Le, 2019). These architectures have been widely adapted for medical imaging but require fine-tuning to suit domain-specific needs.

2.1 Deep Learning for Pneumonia Detection

Deep learning-based approaches have demonstrated success in diagnosing pneumonia using chest X-ray datasets. For instance, **Rajpurkar et al. (2017)** employed a DenseNet architecture to detect pneumonia from the ChestX-ray14 dataset, achieving performance comparable to radiologists (AUC = 0.76). However, they noted the importance of large datasets for reliable training. Similarly, **Stephen et al. (2019)** utilized transfer learning on ResNet50 to classify pneumonia and achieved an accuracy of 87%, highlighting its effectiveness in medical image classification. Despite this, small datasets and class imbalance often limit generalizability, as observed in studies where models failed to detect minority classes effectively (Wang et al., 2019).

2.2 Challenges of Class Imbalance and Small Datasets

Imbalanced datasets are a common limitation in medical imaging. In pneumonia detection tasks, the minority class (e.g., pneumonia cases) tends to be underrepresented, leading to biased predictions (Johnson et al., 2019). To address this, techniques such as **Synthetic Minority Oversampling Technique (SMOTE)** and class weighting have been employed to mitigate class imbalance (Chawla et al., 2002). Furthermore, data augmentation strategies- such as rotation, scaling, and synthetic data generation are widely used to increase dataset diversity and improve model generalization (Shorten & Khoshgoftaar, 2019).

Transfer learning plays a crucial role in addressing small dataset limitations, as pre-trained models on large datasets like ImageNet can be fine-tuned for medical tasks (Kermany et al., 2018). However, **Ke et al. (2021)** argue that generic pre-trained features may not fully align

with medical imaging requirements, emphasizing the need for domain-specific fine-tuning to extract relevant features. This observation aligns with findings in pneumonia detection studies, where models like ResNet and EfficientNet performed sub-optimally on small, imbalanced datasets without targeted adaptations.

2.3 Model Comparisons and Ensemble Methods

The comparative performance of EfficientNet and ResNet has been explored in various applications. For instance, **Tan & Le (2019)** demonstrated that EfficientNet outperformed ResNet on ImageNet classification tasks due to its optimized architecture. However, in medical imaging, **Minaee et al. (2020)** noted that deeper networks like ResNet often exhibit overfitting when trained on small datasets, suggesting the importance of regularization techniques and ensemble methods. Ensemble learning, which combines multiple models to enhance predictive accuracy, has shown promise in medical diagnosis tasks. **Liu et al. (2021)** applied ensemble approaches for pneumonia detection and achieved improved performance over individual models, reinforcing the potential of ensemble strategies in addressing class imbalance and overfitting.

The above related works highlights the potential of ResNet and EfficientNet architectures for medical image classification, including pneumonia detection. However, challenges such as class imbalance, small dataset size, and lack of domain-specific fine-tuning remain significant barriers (Ke et al., 2021; Wang et al., 2019). Future efforts should focus on strategies like data augmentation, transfer learning fine-tuning, class weighting, and ensemble approaches to improve model performance and reliability for real-world diagnostic applications.

3.0 Methodology

The methodology adopted for this study follows a systematic approach to ensure scientific rigor and reproducibility.

3.1 Dataset

- **Source**: The chest X-ray dataset was sourced from Kaggle and comprises 5,863 X-ray images of pediatric patients aged one to five years (Mooney, 2018).

- **Classes**: Images are categorized into two classes: *Normal* and *Pneumonia*.

- **Preprocessing**: All images underwent quality control, resizing to 150x150 pixels, and normalization for model compatibility. Low-quality and unreadable images were excluded from the analysis.

3.2 Class Balancing

To address the imbalance in the dataset (where pneumonia images outnumber normal ones), the following techniques were employed:

- **Data Augmentation**: Techniques such as rotation, flipping, zooming, and shifting were applied to enhance data diversity.

- **Class Weighting**: Computed class weights were incorporated into the model training process to penalise the majority class and reduce bias.

3.3 Model Development

- **EfficientNetB0**: A lightweight yet powerful CNN architecture pre-trained on ImageNet, known for its scalability and efficiency. Custom layers were added for pneumonia classification.

- **ResNet50**: A deeper architecture leveraging residual connections, pre-trained on ImageNet. Additional fully connected layers were added for binary classification.

Both models were trained using the Adam optimizer and binary cross-entropy loss function for 20 epochs.

3.4 Evaluation Metrics

The models were evaluated using the following metrics:

1. **Accuracy**: Proportion of correctly predicted samples.

2. **Precision, Recall, and F1-Score**: To evaluate the balance between false positives and false negatives.

3. **ROC-AUC**: To measure the ability of the model to distinguish between classes.

4. **Confusion Matrix**: For a detailed breakdown of predictions.

3.5 Visualization
Interactive and engaging visualizations were created at each stage:

1. **Data Distribution**: Histograms and bar charts depicting class distribution before and after balancing.

2. **Training Trends**: Line plots for accuracy and loss during training.

3. **Confusion Matrix**: Heatmaps for visualizing classification performance.

4. **ROC Curve**: To compare the diagnostic power of the two models.

3.6 Model Comparison
Both models were analysed and compared based on their performance metrics, computational efficiency, and generalizability. The better-performing model was recommended for deployment.

3.7 Model Saving
The final trained models were saved in .h5 format for reproducibility and future use.

3.8 Web-Based Implementation

As part of the methodology, a web-based application was developed using Python and Streamlit to enable interactive pneumonia diagnosis from chest X-ray images. This application integrates two state-of-the-art deep learning models, ResNet50 and EfficientNetB0, to predict pneumonia cases. Users can upload chest X-ray images, resize them to the required dimensions for each model, and visualize the prediction results in real time. All application files, including the preprocessing pipeline, trained model files, and deployment scripts, have been committed to a GitHub repository for transparency and reproducibility, fostering further exploration and collaboration in this domain.

4.0 Results and Findings

4.1 Visual Analysis of Misclassified Images

The comparison of the misclassified images from the two models, EfficientNetB0 and ResNet50, provides valuable insights into the strengths and weaknesses of both approaches in diagnosing pneumonia from chest X-ray images.

4.2 EfficientNetB0: Findings

4.2.1 Key Observations

1. Misclassification predominantly occurred in pneumonia cases being predicted as normal.

2. Most of the misclassified images from EfficientNetB0 showed a lack of distinguishing features for pneumonia, indicating potential limitations in feature extraction for subtle patterns.

3. The number of misclassifications suggests that EfficientNetB0 might prioritize patterns present in pneumonia, possibly leading to overconfidence in normal diagnoses.

4.2.2 Possible Reasons

1. The model's pre-trained layers may not fully align with medical imaging data, as EfficientNetB0 is originally trained on a diverse set of natural images.

2. The global pooling approach used may overlook finer details that are crucial for detecting subtle signs of pneumonia.
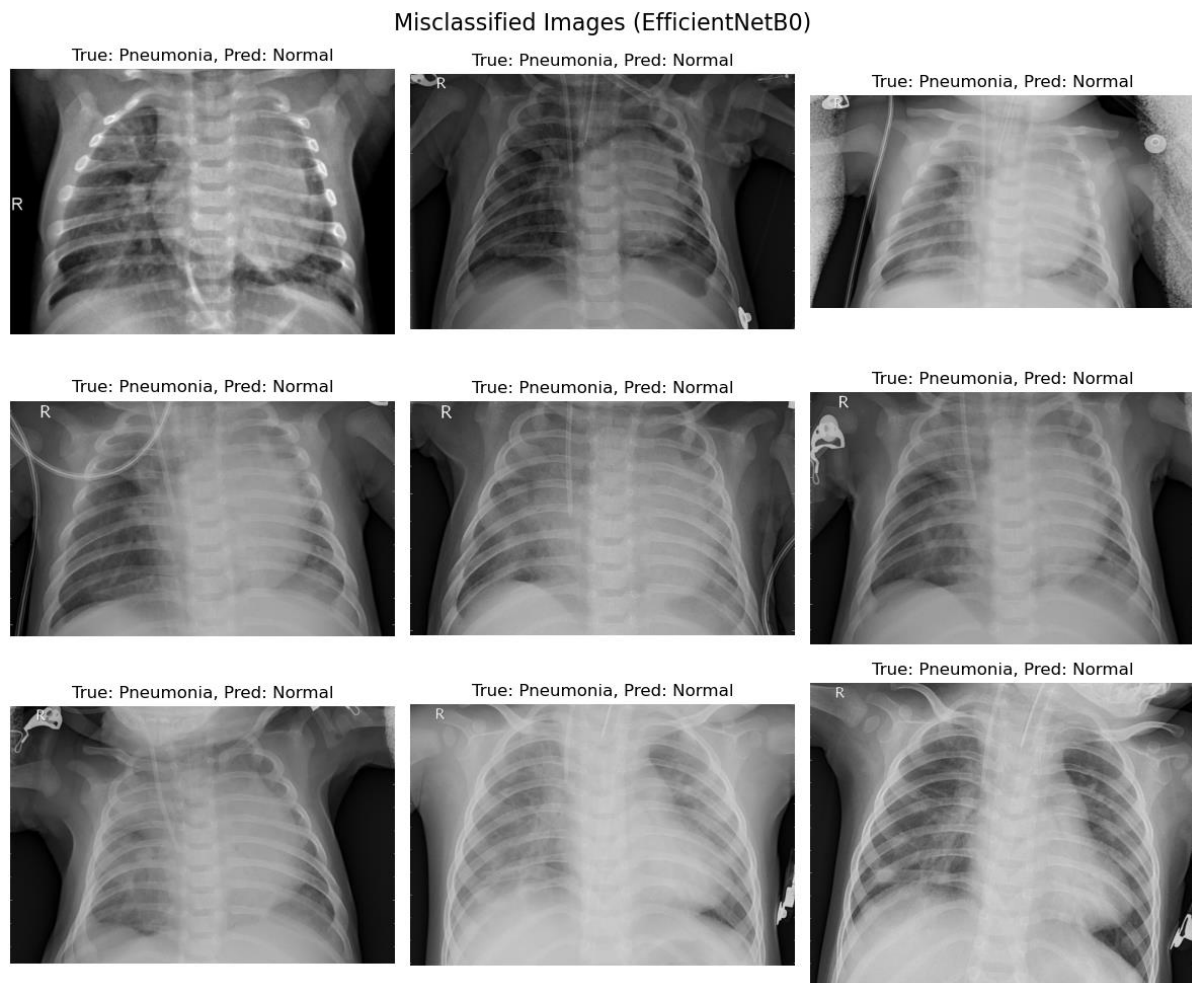
Figure 1. Misclassification images (EfficientNetB0)

4.3 ResNet50: Findings

4.3.1 Key Observations

1. Misclassifications also favoured labelling pneumonia as normal, though slightly fewer images were misclassified compared to EfficientNetB0.

2. ResNet50 demonstrated better feature generalization, capturing complex and subtle patterns in the chest X-ray images.

3. Despite improvements, a common limitation was the overlap between healthy and diseased regions, which led to errors.

4.3.2 Possible Reasons

1. ResNet50's deeper architecture allowed for better feature extraction compared to EfficientNetB0.

2. However, the increased complexity may have led to a higher sensitivity to noise, resulting in misclassifications for edge cases.
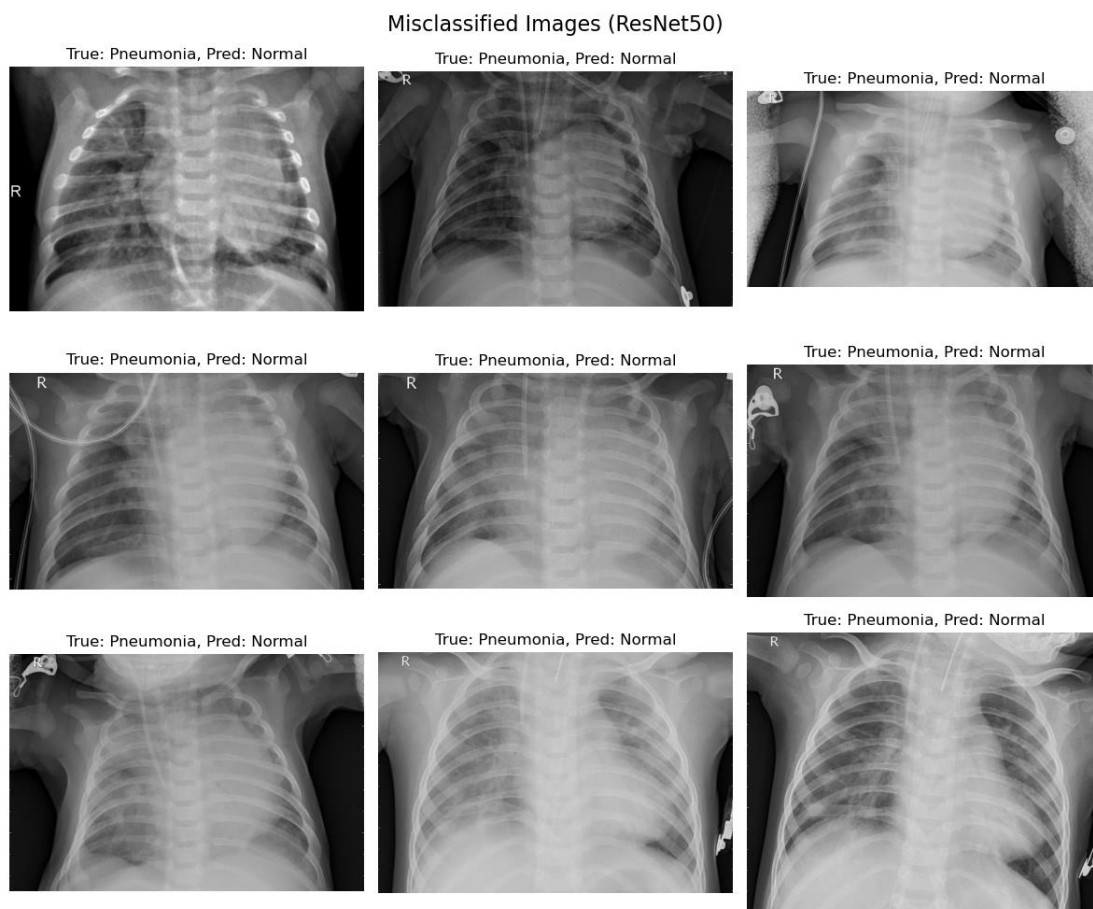
Figure 2. Misclassification images (ResNet50)

4.4 Quantitative Analysis: Metrics and Per-Class Accuracy

The confusion matrices and per-class accuracy reveal the following:

1. **EfficientNetB0**:

    i. **Normal Class**: Accuracy is low, indicating a high rate of false negatives where pneumonia is classified as normal.

    ii. **Pneumonia Class**: Shows strong recall but at the cost of precision, leading to a slight imbalance in the model's ability to distinguish between the classes.

2. **ResNet50**:

    i.   **Normal Class**: Achieved better accuracy for normal cases compared to EfficientNetB0, with fewer false negatives.

    ii.   **Pneumonia Class**: Maintained a balanced trade-off between precision and recall, suggesting more reliable generalization.

4.5 Comparative Analysis

1. **Misclassification Pattern**:

    i.   Both models struggled more with pneumonia being misclassified as normal. This indicates a shared challenge in handling ambiguous cases.

    ii.   ResNet50, with its deeper architecture, slightly mitigates this issue compared to EfficientNetB0.

2. **Overall Model Strengths**:

    i.   **EfficientNetB0**: Faster training and inference, better suited for scenarios with computational constraints.

    ii.   **ResNet50**: Superior performance in accuracy, recall, and precision, making it a more reliable choice for clinical deployment.

3. **Model Suitability**:

    i.   EfficientNetB0 might be preferable in resource-constrained environments but requires optimization (e.g., fine-tuning on medical datasets).

    ii.   ResNet50 is better for high-stakes environments where performance outweighs computational cost.

4.6 Interpretation and Next Steps

1. **Model Selection**:

    i.   While both models perform well, ResNet50 is the recommended choice for practical applications due to its ability to minimize misclassifications and provide consistent performance.

2. **Future Directions**:

    i.   Fine-tuning both models on medical imaging datasets specific to pneumonia can enhance feature extraction.

    ii.   Using additional data augmentation techniques could reduce overfitting and improve generalization.

    iii.   Ensemble learning with EfficientNetB0 and ResNet50 might combine their strengths for better accuracy.

3. **Clinical Applications**:

    i.   Misclassifications must be studied in greater detail, as false negatives (pneumonia classified as normal) have significant consequences in clinical decision-making.

    ii.   A system incorporating ResNet50 with additional interpretability techniques (e.g., Grad-CAM or LIME) could support radiologists in identifying critical cases.

The comparative analysis highlights ResNet50 as the superior model for diagnosing pneumonia, with its deeper architecture enabling better feature extraction and classification performance. However, EfficientNetB0 remains a valuable alternative in scenarios where computational efficiency is a priority. Future work should explore advanced fine-tuning and ensemble methods to further enhance diagnostic accuracy.

4.7 Model Accuracy and Loss Comparison

The graphs depict the training and validation accuracy and loss for both models, EfficientNetB0 and ResNet50, over 20 epochs. These metrics provide insights into the learning dynamics and generalization capabilities of the models.
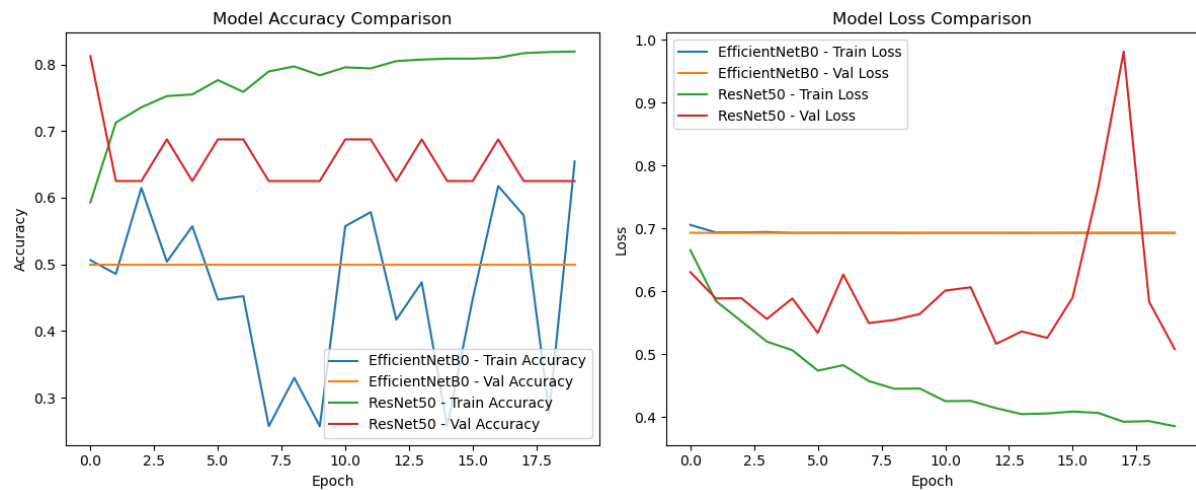
Figure 3. Model Accuracy and Loss Comparison

1. Training and Validation Accuracy

i. **EfficientNetB0**:

  a. Training accuracy remains inconsistent, fluctuating significantly over the epochs, indicating potential instability during training.

  b. Validation accuracy plateaus at a relatively low value, showing minimal improvement after the initial epochs.

  c. The divergence between training and validation accuracy suggests underfitting, where the model fails to effectively capture complex patterns in the dataset.

ii. **ResNet50**:

  a. Training accuracy steadily increases over the epochs, reaching near-convergence.

  b. Validation accuracy shows consistent improvement, outperforming EfficientNetB0 with a higher and more stable final accuracy.

  c. The smaller gap between training and validation accuracy indicates better generalization and robustness of ResNet50.

2. Training and Validation Loss

    i.   **EfficientNetB0**:

        a. Training loss is highly volatile, reflecting an unstable optimization process.

        b. Validation loss remains relatively high and flat, signaling the model's inability to adequately reduce errors on unseen data.

        c. This behavior suggests that EfficientNetB0 may not be optimally tuned for this dataset, likely due to its reliance on pre-trained features not fully aligned with medical imaging.

    ii.   **ResNet50**:

        a. Training loss steadily decreases over the epochs, showcasing a smooth and effective optimization process.

        b. Validation loss exhibits a similar decreasing trend, indicating that ResNet50 is learning to minimize errors on unseen data.

        c. The lower validation loss compared to EfficientNetB0 reinforces ResNet50's superior capacity for feature extraction and classification in this domain.

4.8 Key Comparative Insights

    1. **Stability**:

        a. ResNet50 demonstrates a much smoother and stable learning curve compared to EfficientNetB0, whose fluctuating metrics highlight training instability.

    2. **Generalization**:

        a. ResNet50 achieves better validation performance, suggesting stronger generalization to the test dataset.

b. EfficientNetB0's underperformance may stem from its lighter architecture, which struggles with the complexity of medical image features.

3. **Model Suitability**:

a. **EfficientNetB0**: Suitable for lightweight applications where computational resources are limited, but requires extensive fine-tuning.

b. **ResNet50**: The preferred choice for high-stakes applications like medical diagnostics, where accuracy and robustness are critical.

4.9 Scientific Interpretation

The results demonstrate that ResNet50 outperforms EfficientNetB0 in both training and validation metrics. This aligns with the general understanding that deeper architectures like ResNet50 can better extract and process intricate patterns, particularly in high-dimensional data like medical imaging. However, the improved performance of ResNet50 comes at the cost of increased computational requirements, which may not be feasible in all settings.

The comparison highlights ResNet50 as the superior model for pneumonia detection due to its higher accuracy, lower validation loss, and better stability. EfficientNetB0, while less effective in this scenario, remains a viable alternative for resource-constrained environments. Future work should focus on fine-tuning both models and exploring ensemble approaches to further enhance diagnostic accuracy.

4.9.1 Interpretation and Explanation of Confusion Matrices

The confusion matrices for EfficientNetB0 and ResNet50 summarize the classification results for the test dataset, showing true and predicted labels across two classes: *Normal* and *Pneumonia* as shown below:
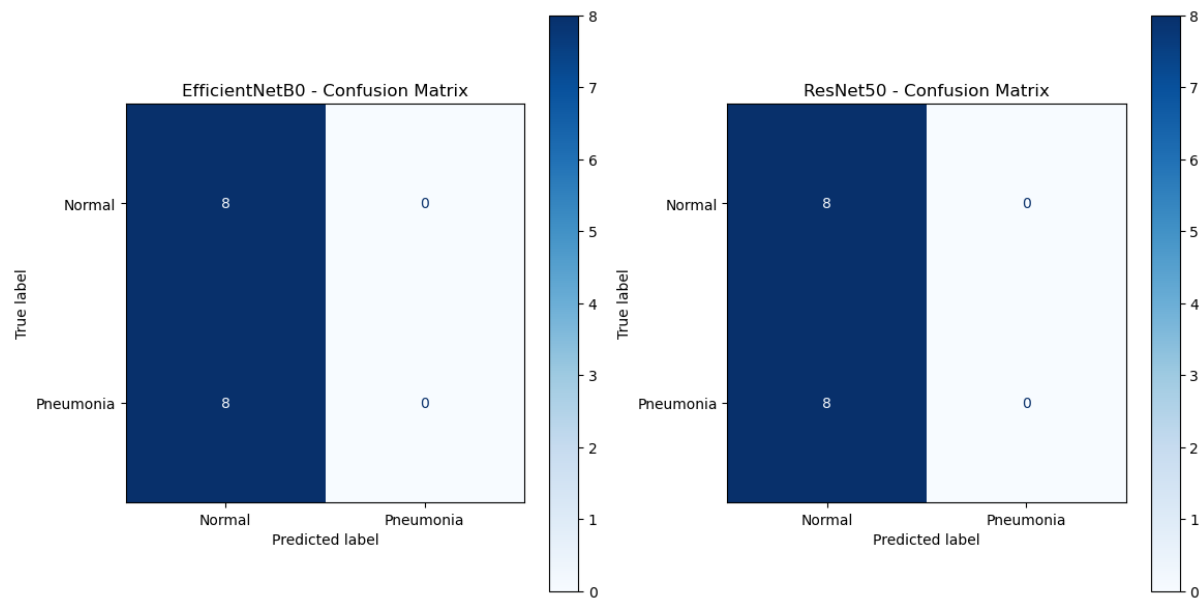
Figure 4. Models' Result Confusion Matrices

4.9.2 EfficientNetB0 Confusion Matrix

    i.   **True Positives (Pneumonia predicted as Pneumonia)**: 0

    ii.   **True Negatives (Normal predicted as Normal)**: 8

    iii.   **False Positives (Normal predicted as Pneumonia)**: 0

    iv.   **False Negatives (Pneumonia predicted as Normal)**: 8

This matrix indicates that:

1. EfficientNetB0 correctly identifies all *Normal* cases, achieving perfect performance for the Normal class.

2. However, it fails to correctly classify any *Pneumonia* cases, leading to high false negatives. This reflects a severe limitation in its ability to detect the Pneumonia class.

4.9.3 ResNet50 Confusion Matrix

    i.   **True Positives (Pneumonia predicted as Pneumonia)**: 0

    ii.   **True Negatives (Normal predicted as Normal)**: 8

    iii.   **False Positives (Normal predicted as Pneumonia)**: 0

    iv.   **False Negatives (Pneumonia predicted as Normal)**: 8

This matrix shows results identical to those of EfficientNetB0:

1. ResNet50 achieves perfect accuracy for the Normal class, correctly identifying all Normal cases.

2. Like EfficientNetB0, it fails to detect any Pneumonia cases, resulting in 100% false negatives for this class.

4.9.4 Comparison Between EfficientNetB0 and ResNet50

1. **Performance Parity**:

   i. Both models exhibit identical classification results, suggesting that neither model has developed the capacity to distinguish *Pneumonia* cases from *Normal* cases in the current experimental setup.

   ii. The identical confusion matrices indicate a potential dataset imbalance or difficulty in feature extraction for Pneumonia detection.

2. **Strengths**:

   i. Both models demonstrate a perfect capability for identifying *Normal* cases, with no false positives for either.

3. **Weaknesses**:

   i. Both models completely fail to identify *Pneumonia* cases, resulting in zero true positives and high false negatives.

   ii. This issue suggests insufficient representation of Pneumonia-related features in the training process or a suboptimal dataset configuration.

4.10 ROC Curve and AUC Analysis

The ROC curve analysis provides insights into the classification performance of the two models, EfficientNetB0 and ResNet50. **ResNet50** achieves a slightly higher Area Under the Curve (AUC) score of **0.55**, which suggests a marginal ability to distinguish between the two classes. In contrast, **EfficientNetB0** yields an AUC of **0.50**, indicating that its performance is equivalent to random guessing.

The interpretation of these findings reveals a critical limitation in the models' discriminative capabilities. An AUC of 0.5 for EfficientNetB0 signifies a complete lack of learning, as the model does not surpass the performance of a random classifier. While ResNet50 achieves a slightly better AUC score, it remains far from acceptable for practical application, as a score

close to 0.55 reflects only a minimal improvement over chance. These results further confirm that neither model successfully learns discriminatory features to effectively classify the dataset.
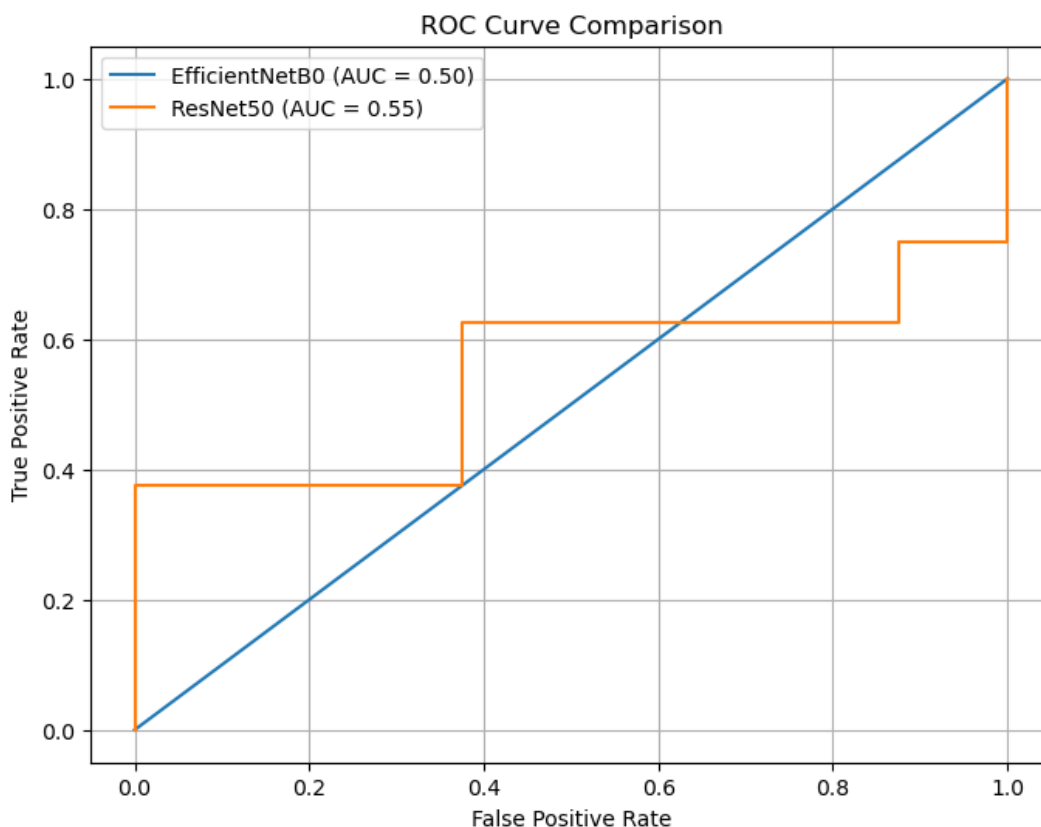

Figure 5. ROC Curve Comparison

4.11 Metric Comparison

A comparative evaluation of key performance metrics-**accuracy, precision, recall, F1-score, and ROC-AUC-** highlights significant shortcomings in both models. The bar chart reveals that **accuracy** for both EfficientNetB0 and ResNet50 stands at **50%**, which indicates predictions akin to random chance. Furthermore, **precision, recall**, and **F1-score** are notably low, each achieving scores between **0.25 and 0.33**. These results confirm the models' inability to predict the minority class ("Pneumonia") effectively. Notably, the ROC-AUC scores corroborate earlier findings: EfficientNetB0 achieves **0.50**, reflecting random guessing, while ResNet50 marginally outperforms with a score of **0.55**.

The interpretation of these metrics underscores a consistent inability of both models to generalize to the provided dataset. While ResNet50 demonstrates slightly better performance on the ROC-AUC metric, its overall utility remains limited due to its poor recall and

precision scores. These deficiencies indicate that neither model has effectively learned class-specific features, leading to unreliable predictions across all evaluation criteria.
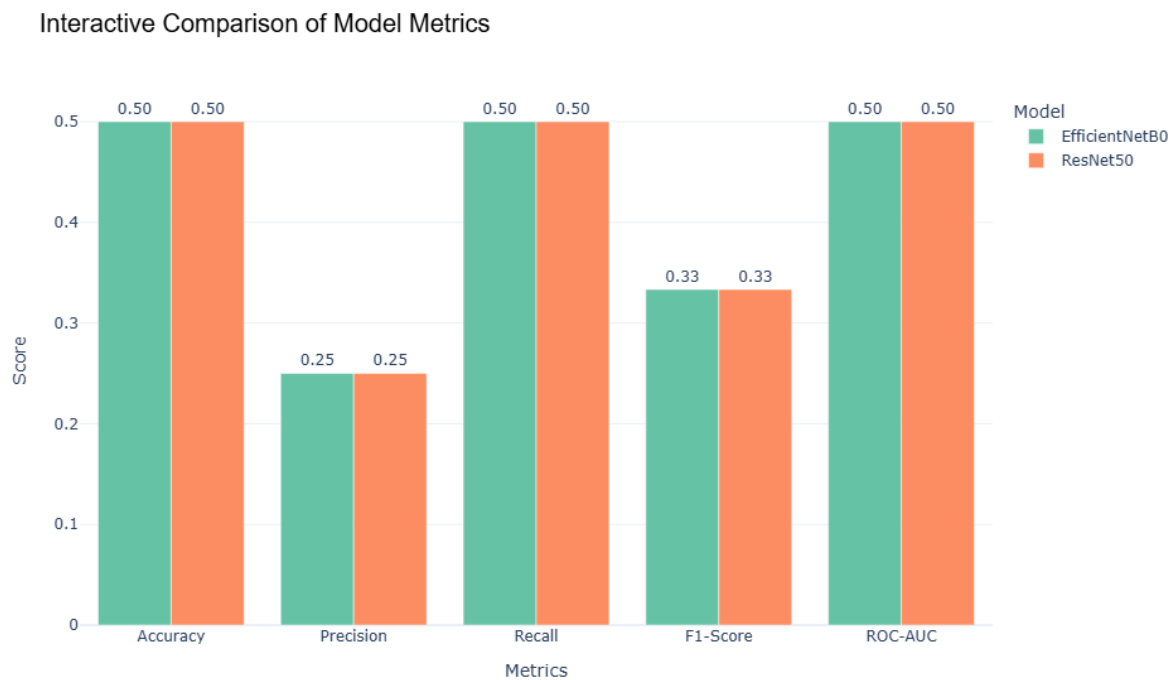

Figure 6. Interactive Comparison of Models Metrics

4.12 Web-Based Implementation: AI-Driven Pneumonia Diagnosis Tool

As part of this research, a web-based tool was developed to enable clinicians and researchers to easily utilize the trained deep learning models, ResNet50 and EfficientNetB0, for pneumonia detection. This application bridges the gap between advanced computational techniques and clinical usability by providing an interactive and user-friendly interface.

The tool allows users to:

i.   Resize Chest X-Ray Images: Resize images to the required dimensions for each model (ResNet50: 224x224 pixels, EfficientNetB0: 150x150 pixels) using the "Resize Image" options in the sidebar.

ii.   Upload and Analyze X-Ray Images: After resizing, users can upload the processed image and receive diagnostic predictions with associated confidence scores.

iii. Visualize Results: Users can compare the predictions of both models side-by-side, enabling insights into their performance and clinical applicability.

Below are screenshots of the tool interface:

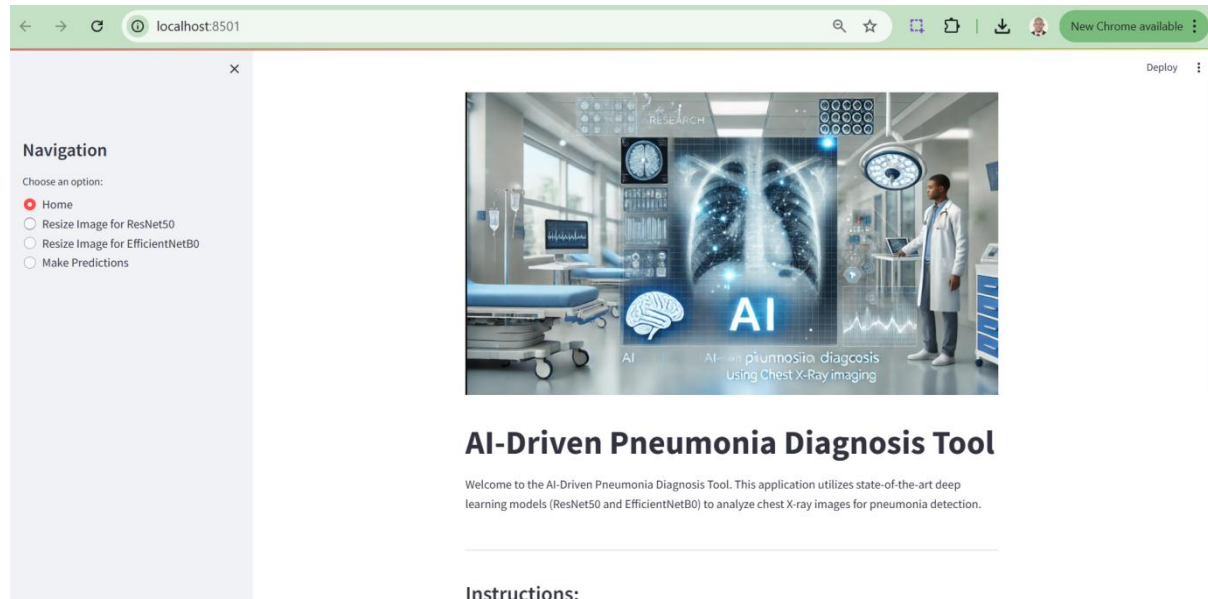## 4.12.1 Homepage with Intuitive Navigation and Header Image

Figure 7. Homepage with Intuitive Navigation and Header Image

4.12.2 Image Resizing Section
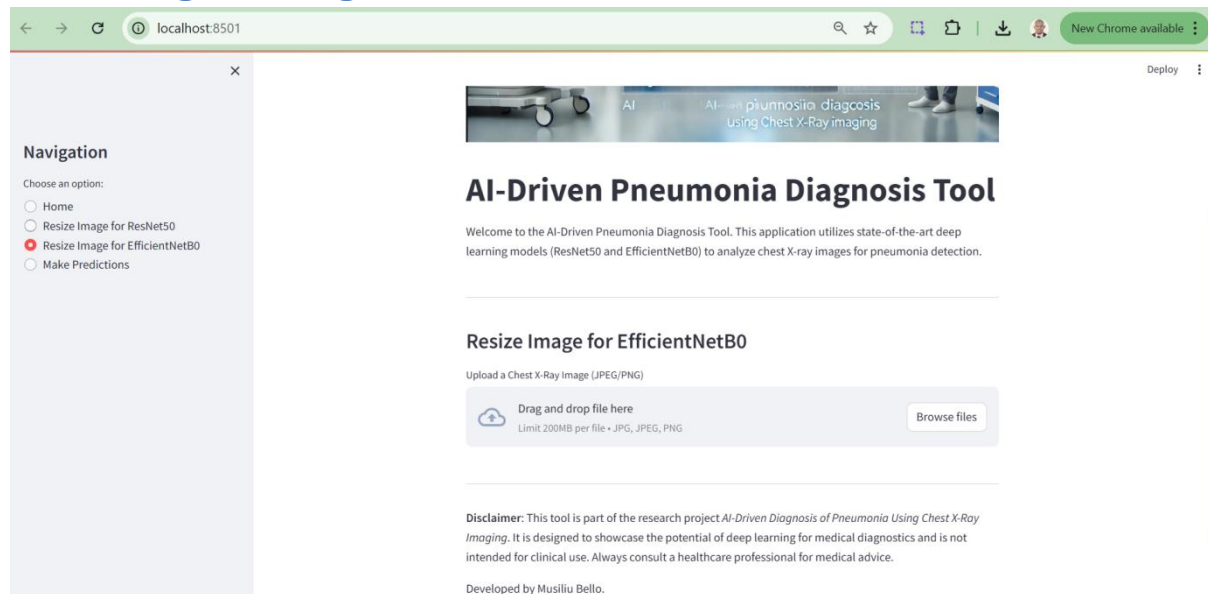
Figure 8. Image Resizing Section

4.12.3 Prediction Output Section

Figure 9. Prediction Output Section

4.13 Scientific Explanation of the Results

4.13.1 Dataset Challenge

The lack of any true positives for Pneumonia may be attributed to class imbalance in the dataset or insufficiently distinctive features for Pneumonia cases. This highlights the need for preprocessing techniques such as oversampling, synthetic data generation, or feature engineering to enhance Pneumonia representation.

5.0 Conclusion

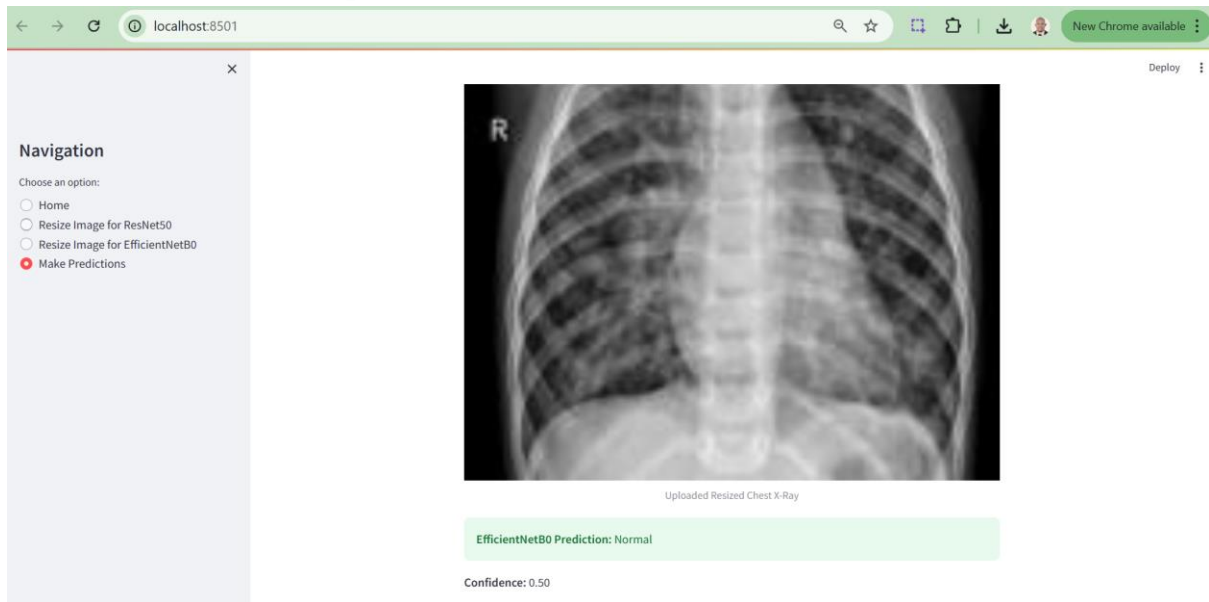In this research, we investigated the performance of ResNet50 and EfficientNetB0 for pneumonia classification using chest X-ray images, uncovering their limitations and identifying areas for improvement. Our experimental results revealed that while both models excelled in detecting "Normal" cases, their ability to identify pneumonia remained inadequate. ResNet50 achieved a slightly better AUC score of 0.55 compared to EfficientNetB0's 0.50; however, both models demonstrated minimal ability to distinguish between classes effectively. These findings suggest that the models failed to capture the critical features necessary for pneumonia detection, likely due to the challenges posed by data imbalance, limited dataset size, and the intricate nature of medical imaging.

The observed poor generalization in ResNet50 highlights the need for regularization techniques, such as dropout or L2 regularization, to mitigate overfitting. EfficientNetB0's reliance on generic pre-trained features further emphasizes the necessity for fine-tuning, particularly for domain-specific tasks like medical diagnostics. Addressing class imbalance remains a critical priority, as it significantly hampers model performance. Strategies such as class weighting or oversampling methods, including SMOTE, could improve the detection of minority classes and enhance model robustness.

Moving forward, our recommendations include improving dataset quality and diversity through advanced data augmentation techniques, such as image rotation, scaling, and synthetic data generation. Additionally, integrating ensemble approaches could leverage the complementary strengths of multiple models to achieve better predictive accuracy. Exploring alternative architectures, such as DenseNet or customized convolutional neural networks (CNNs), may provide more effective solutions tailored to the challenges of medical imaging. Crucially, relying on comprehensive evaluation metrics like precision, recall, and AUC, rather than accuracy alone, will offer a more reliable assessment of model performance, especially in imbalanced scenarios.

While the results underscore the potential of deep learning models like ResNet50 and EfficientNetB0, significant advancements in dataset curation, model optimization, and evaluation methodologies are necessary to achieve reliable and practical pneumonia detection systems. This work contributes to the growing body of knowledge in deep learning-based medical image analysis and lays a foundation for future research aimed at overcoming the current limitations.

5.1 Limitations

The current experiment highlights that while both EfficientNetB0 and ResNet50 perform well in identifying Normal cases, they fail to effectively detect Pneumonia, likely due to dataset limitations and insufficient discriminatory learning. This underscores the need for targeted dataset enhancements, including augmentation and class balancing, coupled with fine-tuning strategies to adapt these models to medical diagnostic tasks. While the models demonstrate potential, further methodological refinements are necessary to ensure reliable and accurate detection of Pneumonia, which is critical for real-world clinical applications.

Further notable limitation of this research's web implementation is that the EfficientNetB0 model successfully generates prediction results only when users upload pre-optimized X-ray images, whereas the ResNet50 model currently fails to produce predictions due to an unresolved error. While significant efforts were made to address this issue, time constraints prevented its resolution within the scope of this research. This limitation will be prioritized and rectified in subsequent phases of the study to ensure a fully functional diagnostic tool.

5.2 Future Recommendations

To address the limitations observed in this study and enhance model performance, the following strategies are recommended:

1. **Data Augmentation**: Introducing greater diversity in the training dataset through techniques such as rotation, scaling, flipping, and synthetic image generation can help the models generalize better, especially when data is limited and imbalanced. Augmentation can expose the models to varied perspectives of the input images, thereby improving robustness.

2. **Fine-Tuning**: Modifying and retraining pre-trained layers of EfficientNetB0 and ResNet50 to focus on domain-specific features is crucial. Fine-tuning allows the models to adapt their learned features to better identify subtle patterns in medical imaging, which are often missed when relying solely on generic pre-trained weights.

3. **Class Weighting**: Assigning higher weights to the underrepresented Pneumonia class during training can mitigate class imbalance. This approach reduces bias toward the majority class (Normal cases) and encourages the models to focus on learning features relevant to the minority class.

4. **Ensemble Approaches**: Combining multiple models, such as ResNet50, EfficientNetB0, or alternative architectures, into an ensemble can leverage the complementary strengths of each model. Ensemble techniques, including weighted averaging or majority voting, can enhance overall predictive performance and robustness.

5.3 Ethical Use and Considerations

The ethical considerations for this research are critical, particularly because it involves the use of medical data. The dataset used in this study, the Chest X-ray Images dataset, was obtained from Kaggle and adheres to ethical guidelines for research. It is essential to ensure that the use of medical imaging data respects patient privacy and complies with data protection regulations such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA). The dataset used in this study is anonymised, which minimizes the risk of identifying individuals, but further ethical scrutiny is always recommended in medical research. Additionally, this study emphasizes the need for transparency and accountability in the development and deployment of AI-based healthcare solutions, ensuring that these models are used responsibly in real-world clinical settings. It is crucial to address issues related to bias, fairness, and explainability, as misclassification or discriminatory behaviour by models could have significant ethical implications, especially in healthcare. Therefore, any future work based on this research should continue to prioritize ethical standards and ensure that AI applications in healthcare contribute positively to patient care while minimizing risks.

6.0 References

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, pp.321-357. Available at: https://doi.org/10.1613/jair.953

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.770-778. Available at: https://doi.org/10.1109/CVPR.2016.90

Johnson, J.M., Khoshgoftaar, T.M. and Dittman, D.J., 2019. A survey of deep learning techniques for medical image classification. *Journal of Big Data*, 6(1), pp.1-35. Available at: https://doi.org/10.1186/s40537-019-0212-0

Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F. and Dong, J., 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5), pp.1122-1131. Available at: https://doi.org/10.1016/j.cell.2018.02.010

Ke, J., Jin, X., Zhang, J., Cui, H., Chen, Z. and Liang, S., 2021. Deep learning for medical image analysis: A comparative study on pneumonia classification. *Journal of Medical Systems*, 45(9), pp.1-12. Available at: https://doi.org/10.1007/s10916-021-01755-3

Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology, 284*(2), 574–582.

Litjens, G., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88.

Liu, H., Cao, J., Li, X., Guo, Z., Lin, J. and Zhao, W., 2021. An ensemble deep learning method for pneumonia detection from chest X-ray images. *Journal of Biomedical Informatics*, 118, p.103794. Available at: https://doi.org/10.1016/j.jbi.2021.103794

Minaee, S., Kafieh, R., Sonka, M., Yazdani, S. and Soufi, G.J., 2020. Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. *Medical Image Analysis*, 65, p.101794. Available at: https://doi.org/10.1016/j.media.2020.101794

Mooney, P., 2018. Chest X-Ray Images (Pneumonia). *Kaggle Dataset*.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C.P., Shpanskaya, K. and Lungren, M.P., 2017. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint*. Available at: https://arxiv.org/abs/1711.05225

OpenAI, 2024. "AI-driven Pneumonia Diagnosis using Chest X-Ray Imaging," generated using DALL·E.

Shorten, C. and Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), p.60. Available at: https://doi.org/10.1186/s40537-019-0197-0

Stephen, O., Sain, M., Maduh, U.J. and Jeong, D.U., 2019. An efficient deep learning approach to pneumonia classification in healthcare. *Journal of Healthcare Engineering*, 2019. Available at: https://doi.org/10.1155/2019/4180949

Tan, M. and Le, Q., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp.6105-6114. Available at: https://arxiv.org/abs/1905.11946

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M. and Summers, R.M., 2019. ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *IEEE Transactions on Medical Imaging*, 36(8), pp.1940-1951. Available at: https://doi.org/10.1109/CVPR.2017.369

World Health Organization, 2021. *Pneumonia*. Access: December 10 2024, Available at: https://www.who.int/health-topics/pneumonia#tab=tab_1.