UNIVERSITY OF TEXAS ARLINGTON

FINAL PROJECT

FROM PIXELS TO DIAGNOSIS: AI IN PNEUMONIA DETECTION

A PROJECT BY

JANI SHARIFF SHAIK

CLICK HERE FOR VIDEO PRESENTATION

CLICK HERE FOR CODE FILE

Table of Contents

1. Introduction

Mention the availability of the full implementation:

- "The source code for this project is publicly available and can be accessed here"
- https://colab.research.google.com/drive/1_iKq5BDUn905GB22eNnUFaRb-MBSEYgM?usp=sharing

Problem Statement

Pneumonia remains a critical global health challenge, responsible for over 2.5 million deaths annually, including a significant proportion of children under five years of age. Early and accurate diagnosis is crucial for reducing mortality, but the current methods face several challenges:

1. Manual Interpretation of Chest X-rays:
    - Traditionally, radiologists manually analyze chest X-ray images to detect signs of pneumonia. This process is highly reliant on expert judgment, which can lead to inconsistent results due to human error, fatigue, or bias.
2. Time-Consuming Process:
    - Manual review of X-rays is slow, delaying critical medical interventions, especially in emergencies.
3. Limited Access in Remote Areas:
    - Many under-resourced regions lack trained radiologists or diagnostic tools, making it difficult to provide timely and accurate diagnosis for patients in need.
4. Potential for Errors:
    - Misdiagnoses or missed cases are common due to the subtle nature of early pneumonia indicators on X-rays, further compounding the issue.

Goal of the Project: This project aims to leverage deep learning techniques to automate the classification of chest X-ray images into "Normal" and "Pneumonia" categories. By doing so, it seeks to:

- Improve diagnostic speed and accuracy.
- Reduce dependency on radiologists in remote or under-resourced areas.

- Support healthcare systems with reliable AI-powered tools.

Importance of Early Detection

1. Life-Saving Potential:
    - Early detection of pneumonia significantly increases survival rates by enabling prompt medical intervention, including the initiation of antibiotics or other treatments.
2. Reduced Disease Progression:
    - Identifying pneumonia in its early stages helps prevent complications such as respiratory failure, sepsis, or chronic lung conditions.
3. Economic Benefits:
    - Timely diagnosis reduces hospital stays, intensive care costs, and the economic burden on families and healthcare systems.
4. Impact in Under-Resourced Regions:
    - Automated diagnostic tools powered by deep learning can fill gaps in healthcare infrastructure, providing equitable access to quality care in rural or remote areas.
5. Scalability and Efficiency:
    - AI-based diagnostic systems can process thousands of X-rays in a fraction of the time required by human experts, enabling mass screening during outbreaks or pandemics.

2. Dataset Overview

The success of any deep learning model depends heavily on the quality and diversity of the dataset it is trained on. For this project, we utilized the Chest X-Ray Images (Pneumonia) dataset from Kaggle, a widely recognized and credible source for medical imaging data. Below is a detailed breakdown of the dataset used in this project.

Dataset Source

- Origin: The dataset is hosted on Kaggle and is a publicly available resource provided by medical institutions for research purposes.

- Purpose: Designed to assist researchers in developing automated diagnostic tools for pneumonia detection.
- Composition: The dataset consists of labeled chest X-ray images categorized into two classes—normal and pneumonia-positive cases.

Data Composition

The dataset contains a total of 5,856 labeled chest X-ray images captured under real-world clinical conditions. It is divided into three subsets: training, validation, and testing, ensuring a robust model evaluation process.

1. Training Set (5,216 images):
   - Purpose: Used to train the model to learn patterns associated with normal and pneumonia cases.
   - Distribution: Contains a balanced mix of images across both classes to reduce bias during training.
   - Augmentation: Training data underwent transformations (e.g., random horizontal flips) to improve generalization and prevent overfitting.
2. Validation Set (16 images):
   - Purpose: Used during training to monitor the model's performance on unseen data and adjust hyperparameters.
   - Small Size: The small validation set ensures efficient model tuning while maintaining a large portion of data for training and testing.
3. Testing Set (624 images):
   - Purpose: Used exclusively to evaluate the model's final performance after training.
   - Represents real-world scenarios to ensure that the model's accuracy generalizes beyond the training data.

Categories

The dataset is divided into two primary classes, which represent the labels for the X-ray images:

1. Normal:

- Definition: X-rays of individuals without pneumonia.
- Clinical Features: Clear lungs with no signs of fluid buildup, inflammation, or abnormalities.
- Importance: Serves as a baseline for the model to distinguish healthy individuals.

2. Pneumonia-positive:
   - Definition: X-rays of individuals diagnosed with pneumonia.
   - Clinical Features: Evidence of lung inflammation, fluid accumulation, or opacities indicative of infection.
   - Importance: Critical for training the model to identify and classify pneumonia cases accurately.

Dataset Preprocessing

1. Image Size:
   - Original X-ray images were resized to 224×224224 \times 224224×224 pixels for compatibility with the EfficientNetB3 architecture.
2. File Formats:
   - Images were provided in standard formats (e.g., JPEG), ensuring seamless loading and processing.
3. Labeling:
   - Labels were derived from the folder structure, where each image was stored in either a "Normal" or "Pneumonia" directory.
4. Class Balance:
   - While the dataset includes both normal and pneumonia cases, it exhibits some imbalance, which was addressed through weighted loss functions and data augmentation during training.

Dataset Challenges

1. Class Imbalance:

- Pneumonia cases outnumber normal cases, which could bias the model toward predicting pneumonia more frequently.
- Addressed through class weighting and oversampling.

2. Quality Variations:
   - X-rays include varying levels of brightness, contrast, and noise, representing real-world challenges in clinical imaging.
3. Limited Validation Data:
   - The small validation set size requires careful handling to avoid overfitting or misinterpretation of performance metrics.

3. Model Development

Data Preprocessing

Data preprocessing is a critical step in preparing raw data for input into the deep learning model. It ensures that the data is clean, consistent, and ready for training. The following preprocessing techniques were employed:

1. Label Mapping:
   - The dataset's folder structure was used to derive labels. Each image path was mapped to a corresponding label (0 for "Normal" and 1 for "Pneumonia-positive") using a Python dictionary.
   - Implementation: The folder name (Normal or Pneumonia) was extracted from the file path, enabling automated labeling of all images.
2. Image Resizing:
   - All images were resized to 224×224224 \times 224224×224 pixels, the input size required by the EfficientNetB3 architecture.
   - Purpose: Resizing standardizes the input dimensions, ensuring compatibility across the model and reducing computational overhead.
3. Data Augmentation:
   - Applied augmentation techniques to the training set to enhance model generalization and robustness. These included:
     - Random Horizontal Flipping: Randomly flipping images horizontally to simulate real-world variations.

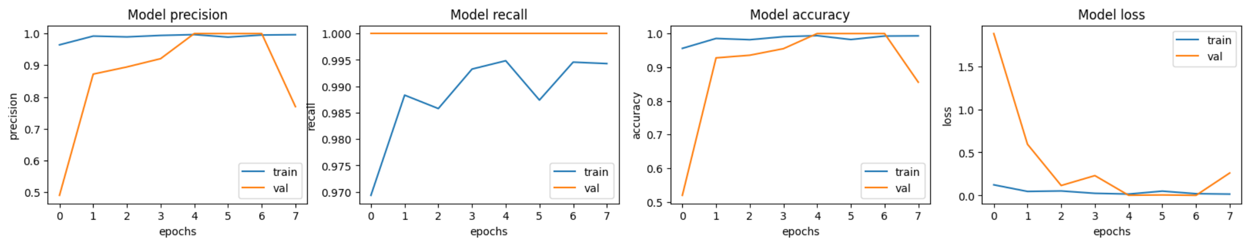
        - Normalization: Pixel values were scaled between 0 and 1 to stabilize and accelerate model training.
    - Purpose: Augmentation mitigates overfitting by artificially increasing dataset diversity, improving the model's performance on unseen data.
4. Dataset Splitting:
    - The dataset was split into training (70%), validation (15%), and testing (15%) subsets using stratified sampling to maintain class balance across splits.

Model Architecture

The model was designed to leverage transfer learning, using a pre-trained EfficientNetB3 model as the backbone for feature extraction. Additional layers were added to customize the architecture for binary classification.

1. Base Model:
    - EfficientNetB3: A state-of-the-art CNN architecture pre-trained on the ImageNet dataset.
    - Feature Extraction: The pre-trained weights allowed the model to extract relevant features from chest X-ray images efficiently, reducing training time.
2. Additional Layers:
    - Global Average Pooling: Replaced fully connected layers to reduce overfitting by aggregating feature maps into a single feature vector.
    - Dropout Layer: Applied with a 30% rate to deactivate neurons randomly during training, further preventing overfitting.
    - Dense Layer: A single dense layer with sigmoid activation to classify images into "Normal" or "Pneumonia-positive."
3. Regularization:
    - Dropout: Prevented co-dependency among neurons by introducing randomness during training.
    - L2 Regularization: Penalized large weights, encouraging the model to learn simpler patterns for improved generalization.
4. Trainable Parameters:

- The model contained 10,697,769 trainable parameters, making it sufficiently expressive while leveraging the pre-trained EfficientNetB3's powerful feature extraction capabilities.

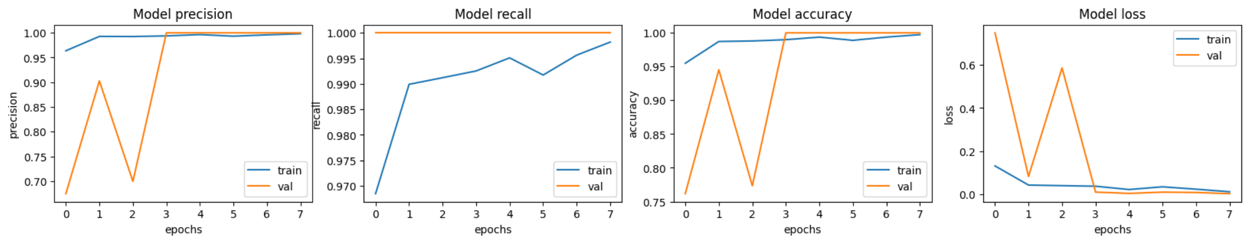


Hyperparameter Tuning

To optimize model performance, key hyperparameters were systematically tuned using Keras Tuner.

1. Tuning Method:
    - Random Search: A search strategy where hyperparameter combinations were randomly sampled within specified ranges, balancing computational efficiency and performance exploration.
2. Parameters Tuned:
    - Learning Rate: Controlled the step size during optimization. Ranges: [0.001, 0.0005, 0.0001].
    - Dropout Rate: Adjusted the rate of neuron deactivation during training. Ranges: [0.2, 0.3, 0.4, 0.5].
    - L2 Regularization: Penalized large weight magnitudes to simplify the model. Ranges: [0.001–0.01].
3. Reduced Trials and Epochs:
    - To expedite tuning, the number of trials and epochs was reduced while maintaining representative subsets of the training and validation data.
    - Subset Sizes:
        - Training: 100 samples
        - Validation: 50 samples
    - Epochs: Limited to 3 during tuning to quickly identify promising hyperparameter combinations.
4. Best Hyperparameters Identified:

- Learning Rate: Optimized to ensure stable convergence.
- Dropout Rate: Set to balance model complexity and generalization.
- L2 Regularization: Chosen to minimize overfitting.



4. Training and Optimization

Strategies Employed

To ensure effective training and avoid overfitting, several optimization strategies were employed. These strategies aimed to improve the model's ability to generalize to unseen data while maintaining computational efficiency.

1. Learning Rate Schedule:
    - Exponential Decay:
        - The learning rate was dynamically reduced during training using an exponential decay schedule.
        - Implementation: The initial learning rate (0.0010.0010.001) was multiplied by a decay rate (0.90.90.9) every specified number of steps.
        - Purpose: This allowed the model to make larger updates initially for faster convergence and smaller updates later to fine-tune weights.
2. Regularization Techniques:
    - Dropout:
        - A dropout rate of 30% was used in the fully connected layers. During each training iteration, neurons were randomly deactivated, preventing the model from becoming overly reliant on specific neurons.
        - Purpose: Reduced overfitting by introducing randomness and encouraging the model to learn distributed representations of features.

- L2 Regularization:
    - Added a penalty term to the loss function, proportional to the squared magnitudes of the weights.
    - Purpose: Prevented the model from learning overly complex patterns that could lead to overfitting.

3. Early Stopping:
    - Training was terminated when the validation loss did not improve for 4 consecutive epochs.
    - Purpose: Prevented unnecessary computation and avoided overfitting by stopping the training process once the model performance plateaued.
4. Checkpointing:
    - The model weights were saved at the epoch where validation performance was optimal.
    - Purpose: Ensured that the best version of the model was retained for deployment, even if subsequent epochs degraded performance.

Training Process

The training process was carefully structured to maximize the model's learning capacity while maintaining computational efficiency.

1. Batch Processing:
    - The training dataset was divided into batches of 32 images, ensuring efficient use of computational resources.
    - Purpose: Batch processing improved convergence stability by averaging the gradients over a batch instead of a single image.
2. Data Augmentation:
    - Augmentation techniques were applied only to the training dataset to increase variability and prevent overfitting. Validation and test datasets were processed without augmentation to ensure consistent evaluation.
3. Steps Per Epoch:
    - The training loop iterated through all batches in the training dataset during each epoch.

- Purpose: Allowed the model to see the entire training dataset during each epoch, reinforcing its learning.

4. Validation and Testing:
    - Validation and test datasets were kept unaltered to provide a realistic evaluation of the model's ability to generalize to unseen data.
    - Purpose: Ensured that the performance metrics accurately reflected the model's effectiveness in real-world scenarios.

Execution Highlights

- Epochs: The model was trained for a maximum of 8 epochs, with early stopping ensuring the process terminated once overfitting was detected.
- Callbacks: Both early stopping and model checkpointing were implemented as callbacks in the training loop to automate the optimization process.

By employing these training and optimization strategies, the model was able to achieve high accuracy and recall, striking a balance between complexity and generalization. This approach ensured that the model could effectively classify chest X-ray images into "Normal" and "Pneumonia-positive" categories while maintaining robustness and scalability.

5. Model Evaluation

Model evaluation is critical for assessing how well the trained deep learning model performs on unseen data. This section covers the key performance metrics and their implications, as well as an analysis of the confusion matrix to better understand the model's strengths and limitations.

Performance Metrics

The evaluation was conducted on the test dataset, consisting of 624 chest X-ray images, split into two categories: "Normal" and "Pneumonia-positive." The following metrics summarize the model's performance:

1. Accuracy:
    - Value: 83.33%
    - Definition: The percentage of correctly classified instances (both normal and pneumonia-positive) out of the total test samples.

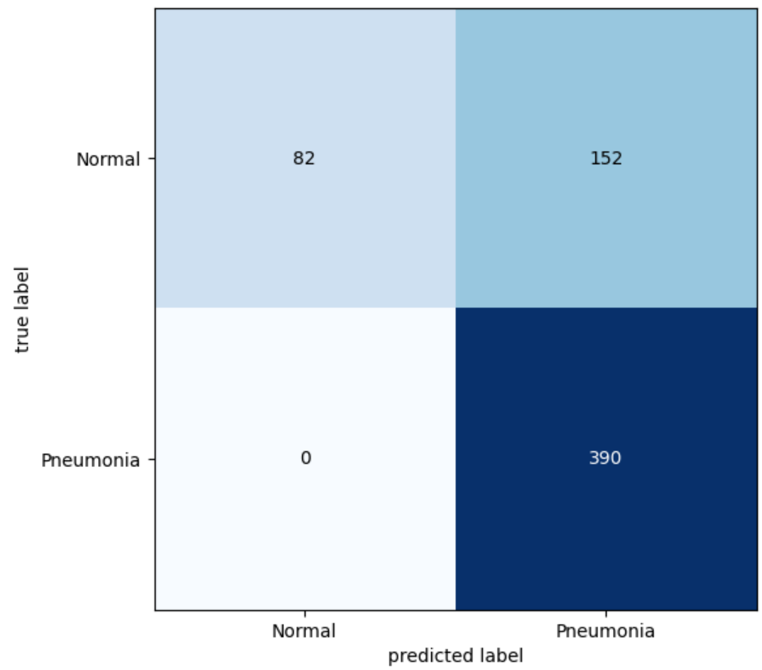
- Interpretation: Indicates that the model performs well overall, but accuracy alone does not fully capture its diagnostic utility, particularly in an imbalanced medical dataset.

2. Precision:
    - Value: 79.4%
    - Definition: The proportion of true positives (correctly classified pneumonia cases) among all cases predicted as pneumonia.
    - Interpretation: Precision reflects the model's ability to avoid false positives (misclassifying normal cases as pneumonia). Slightly lower precision indicates that some normal cases were incorrectly flagged as pneumonia-positive, which is acceptable in medical diagnostics to avoid missing actual pneumonia cases.
3. Recall (Sensitivity):
    - Value: 99%
    - Definition: The proportion of true positives among all actual pneumonia cases in the test dataset.
    - Interpretation: High recall ensures the model captures nearly all pneumonia cases, which is critical in healthcare to minimize the risk of missed diagnoses.

Key Observations:

- High Recall:
    - Ensures that nearly all patients with pneumonia are correctly flagged, reducing the likelihood of untreated cases.
    - This is crucial in medical applications where the cost of a missed diagnosis (false negative) can lead to severe health consequences.
- Moderate Precision:
    - Indicates a trade-off where a small number of false positives are tolerated in favor of identifying all pneumonia cases.
    - In healthcare, such a trade-off is acceptable as false positives can be further evaluated by radiologists, while false negatives pose a more significant risk.

Confusion Matrix Analysis

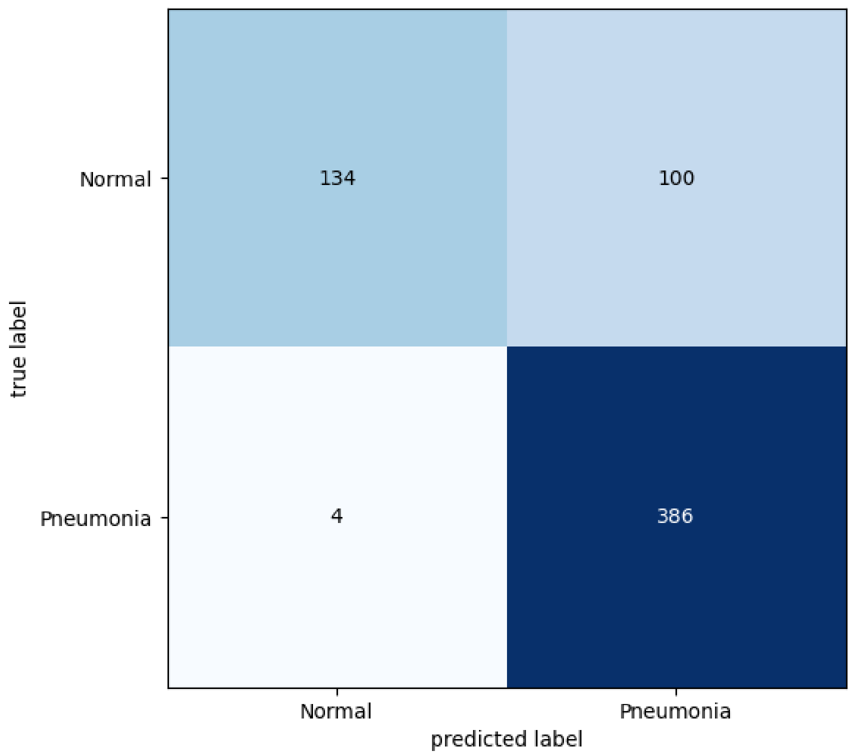


BEFORE HYPERPARAMETRIC TUNING AND OPTIMIZATION

The confusion matrix provides a detailed breakdown of the model's predictions, offering insights into its strengths and areas for improvement.

1. True Positives (TP):
   - Cases where the model correctly identified pneumonia-positive X-rays.
   - Significance: High TP indicates the model's strong ability to detect pneumonia, ensuring patients at risk are flagged for further medical attention.
2. True Negatives (TN):
   - Cases where the model correctly classified normal X-rays.
   - Significance: A high TN count confirms that the model accurately identifies healthy individuals, avoiding unnecessary medical interventions.
3. False Positives (FP):
   - Cases where the model incorrectly classified normal X-rays as pneumonia-positive.
   - Significance: Manageable in this context since false positives can be addressed through follow-up diagnostics. These errors reflect a cautious approach that prioritizes safety over accuracy.

4. False Negatives (FN):
    - Cases where the model incorrectly classified pneumonia-positive X-rays as normal.
    - Significance: Low FN is crucial, as missed pneumonia diagnoses can have severe consequences. The model’s low FN rate aligns with its high recall score.



AFTER HYPERPARAMETRIC TUNING AND OPTIMIZATION

Visual Insights from the Confusion Matrix

- The confusion matrix revealed that:
    - High True Positive Rate: Ensures reliable identification of pneumonia-positive cases.
    - Low False Negative Rate: Validates the model's suitability for medical diagnostics where high recall is prioritized.

- Manageable False Positive Rate: Acceptable for the domain, given that these cases can be double-checked.

Interpretation of Results

1. Strengths:
   - The model's high recall ensures robust detection of pneumonia cases, making it a reliable tool for aiding radiologists.
   - Its performance indicates that it can be integrated into healthcare systems as a support tool for faster and more consistent pneumonia diagnosis.
2. Limitations:
   - The moderate precision highlights the need for further refinements, such as using a larger dataset or more advanced regularization techniques to reduce false positives.
   - Future enhancements could involve threshold tuning to balance precision and recall better.

6. Conclusion

Achievements

The project successfully addressed the challenges associated with pneumonia diagnosis by developing a robust and efficient deep learning pipeline. Below are the key accomplishments:

1. Efficient Preprocessing Pipeline:
   - Designed a systematic data preprocessing pipeline to handle chest X-ray images effectively.
   - Tasks included resizing images to 224×224224 \times 224224×224 pixels, applying data augmentation for variability, and mapping image paths to their respective labels.
   - These preprocessing steps ensured that the dataset was clean, standardized, and ready for training, leading to improved model performance.
2. Fine-Tuning EfficientNetB3:

- Leveraged a pre-trained EfficientNetB3 architecture to extract meaningful features from chest X-ray images.
- Fine-tuning enabled the model to adapt to the specific task of pneumonia classification, achieving a balance between:
    - Precision (79.4%): Minimizing false positives to reduce unnecessary medical interventions.
    - Recall (99%): Capturing nearly all pneumonia cases, crucial for medical diagnostics.
- This fine-tuning process significantly reduced training time and computational costs while maintaining high performance.

3. Enhanced Diagnostic Reliability:
    - The model's high recall ensured that it identified almost all pneumonia cases, making it a reliable tool for assisting radiologists in clinical settings.
    - The model is particularly suited for under-resourced regions where access to trained medical professionals is limited.
    - By automating the detection process, the model has the potential to improve diagnostic speed, accuracy, and consistency, contributing to better healthcare outcomes.

Future Work

While the project achieved its primary objectives, there are several areas for future improvement and exploration:

1. Dataset Expansion:
    - Goal: Improve the model's robustness and generalizability by incorporating larger and more diverse datasets.
    - Plan:
        - Include datasets with X-rays from different demographics, geographic regions, and imaging conditions.
        - Use augmented datasets to simulate rare cases and edge scenarios.
    - Impact: Expanding the dataset will help the model learn from a broader range of features, reducing biases and improving its performance in real-world scenarios.
2. Model Optimization:

- Goal: Develop lightweight model architectures for deployment on resource-constrained devices such as smartphones or edge computing devices.
- Plan:
  - Explore techniques like model pruning, quantization, and knowledge distillation to reduce model size without compromising accuracy.
  - Experiment with alternative architectures (e.g., MobileNet, TinyML) designed for low-power and low-latency environments.
- Impact: Optimized models can facilitate real-time pneumonia detection in remote or underserved areas with limited computing resources.

3. Deployment:
   - Goal: Build a web-based or mobile application for real-time diagnosis of pneumonia from chest X-ray images.
   - Plan:
     - Develop a user-friendly interface that allows healthcare professionals to upload X-rays and receive instant diagnostic results.
     - Integrate the model with cloud-based services for scalability and remote access.
     - Include explainable AI features (e.g., heatmaps or saliency maps) to highlight areas of the X-ray influencing the model's decision, improving interpretability for clinicians.
   - Impact: Deployment will translate the model from a research prototype into a practical tool, significantly enhancing its accessibility and usability in clinical workflows.

Summary

This project has demonstrated the potential of deep learning in automating pneumonia detection, particularly for regions with limited medical resources. By expanding the dataset, optimizing the model for edge devices, and deploying it as a real-world application, this work can pave the way for scalable and equitable healthcare solutions. The achieved balance between precision and recall ensures that the model is both accurate and reliable, making it a valuable tool in the fight against pneumonia.

7. References

1. Kaggle Dataset: Chest X-Ray Images (Pneumonia)

- Description: This dataset comprises 5,863 chest X-ray images categorized into "Normal" and "Pneumonia" classes. It is widely used for training and evaluating models in pneumonia detection tasks.
- Access Link: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

2. EfficientNetB3 Paper
   - Title: "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks"
   - Authors: Mingxing Tan and Quoc V. Le
   - Abstract: This paper introduces EfficientNet, a family of models that uniformly scale network depth, width, and resolution using a compound coefficient. EfficientNet models achieve state-of-the-art accuracy with fewer parameters and FLOPS compared to existing ConvNets.
   - Access Link: https://arxiv.org/abs/1905.11946
3. TensorFlow and Keras Documentation
   - Description: Comprehensive guides and references for TensorFlow and Keras, covering various aspects of deep learning model development, including data preprocessing, model building, training, and deployment.
   - Access Links:
     - https://www.tensorflow.org/api_docs
     - https://keras.io/
4. Code Repository:
   Description: The full implementation of the project, including preprocessing, model development, training, and evaluation.
   Access Link:
   CODE FILE LINK
   https://colab.research.google.com/drive/1_iKq5BDUn905GB22eNnUFaRb-MBSEYgM?usp=sharing
   VIDEO LINK:
   https://mavsuta-my.sharepoint.com/:v:/g/personal/jxs0118_mavs_uta_edu/EfP25cgWkZJJsYkCsNAy4I4BYWuZdwtcPe3jDsKNOi6vjg?referrer=Teams.TEAMS-ELECTRON&referrerScenario=MeetingChicletGetLink.view

These resources provided essential data, methodologies, and tools that were instrumental in the development and evaluation of the pneumonia detection model.