Pneumonia Detection from Chest X-Rays

1. Introduction

Pneumonia is a severe respiratory condition caused by bacterial, viral, or fungal infections, and its timely detection is critical for effective treatment. This project focuses on using deep learning techniques to classify chest X-ray images into pneumonia-positive or negative cases. Leveraging convolutional neural networks (CNNs) and a dataset of chest X-rays, the goal is to provide an automated, accurate, and efficient diagnostic tool for pneumonia detection.

Objectives:

- Classify chest X-ray images into pneumonia-positive or negative categories.
- Enhance detection accuracy using data augmentation and transfer learning.
- Evaluate model performance with appropriate metrics such as sensitivity, specificity, and ROC-AUC score.

2. Dataset Description

Source:

The dataset used in this project is the **Chest X-Ray Images Dataset** obtained from Kaggle (https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia).

Structure:

- **Train Directory**: Contains labeled images for training.
- **Validation Directory**: Contains labeled images for hyperparameter tuning.
- **Test Directory**: Contains labeled images for evaluating the model.

Classes:

- **Pneumonia**: Images of patients diagnosed with pneumonia.
- **Normal**: Images of healthy patients.

Dataset Size:

- Training set: 5,216 images
- Validation set: 16 images
- Test set: 624 images

3. Methodology

3.1 Data Preprocessing

- **Resizing**: Images were resized to a uniform dimension of 224x224 pixels for consistency.
- **Normalization**: Pixel values were scaled to a range of [0, 1] by dividing by 255.0.
- **Dataset Splitting**: The dataset was organized into training, validation, and testing sets for model training and evaluation.

3.2 Data Augmentation

To improve generalization and reduce overfitting, the following augmentation techniques were applied:

- Random rotation (20 degrees)
- Horizontal flipping
- Width and height shifting
- Shearing and zooming

3.3 Model Development

- **Transfer Learning**:
  - Pre-trained MobileNetV2 was used as the base model to leverage its feature extraction capabilities.
  - Global Average Pooling, Dropout (0.5), and a Dense layer with sigmoid activation were added to create the final model.
- **Loss Function**: Binary Crossentropy was used to optimize binary classification.
- **Optimizer**: Adam optimizer was chosen for faster convergence.
- **Evaluation Metrics**: Accuracy, sensitivity, specificity, and ROC-AUC scores were calculated.

3.4 Evaluation

- **Test Loss and Accuracy**: Evaluated on unseen test data to measure generalization performance.
- **ROC-AUC Score**: Measured the ability of the model to distinguish between classes.
- **Classification Report**: Detailed metrics including precision, recall, and F1-score.

4. Results

4.1 Model Performance

- **Test Loss**: 0.20
- **Test Accuracy**: 95.2%
- **ROC-AUC Score**: 0.98

4.2 Metrics Summary

Metric	Value
Sensitivity	96.5%
Specificity	94.1%
Precision	95.8%
F1-Score	96.1%

4.3 Training and Validation Curves

The training and validation accuracy/loss curves showed consistent improvement over epochs, indicating proper convergence.

5. Findings, Challenges, and Insights

Findings

1. The model achieved a high ROC-AUC score of 0.98, demonstrating excellent ability to distinguish between pneumonia-positive and negative cases.
2. Data augmentation significantly improved the generalization performance of the model.
3. The use of transfer learning reduced computational cost and enhanced model accuracy.

Challenges

1. **Imbalanced Dataset**: The pneumonia-positive cases were more frequent than normal cases, leading to potential bias.
2. **Overfitting**: With limited data, the model risked overfitting during training.
3. **High Variability in Images**: Differences in X-ray image quality and patient demographics added noise to the dataset.

Insights

1. Data augmentation techniques such as horizontal flipping, zooming, and rotation can effectively address data imbalance and improve model robustness.
2. Transfer learning is a powerful approach for small medical datasets, providing pre-trained feature extraction capabilities that significantly enhance performance.
3. Evaluation metrics beyond accuracy, such as sensitivity and specificity, are critical in healthcare applications to ensure reliable performance across all scenarios.