Medical Image Classification Using Deep Learning

Chest X-Ray Pneumonia Detection

Tasneem Mohamed Mohamed

1. Project Overview

This project focuses on developing a **deep learning–based medical image classification system** capable of analyzing chest X-ray images to detect **Pneumonia**. The primary objective is to design a robust and reliable **Convolutional Neural Network (CNN)** model that can assist in medical decision-making by accurately identifying abnormal lung conditions.

Given the critical nature of medical diagnosis, the project emphasizes **clinical reliability**, prioritizing sensitivity and recall over raw accuracy to reduce the risk of undetected disease cases.

---

2. Dataset Description

The project utilizes the **Chest X-Ray Images dataset** obtained from Kaggle. This dataset consists of labeled chest radiographs categorized into two classes:

- **Normal**
- **Pneumonia**

Dataset Characteristics:

- Image format: JPEG
- Varying image resolutions
- Binary classification problem
- Dataset split into training, validation, and testing subsets

A notable challenge within the dataset is **class imbalance**, where pneumonia cases are more prevalent than normal images. This imbalance reflects real-world clinical data and required careful handling during training.

---

3. Exploratory Data Analysis (EDA)

An exploratory data analysis phase was conducted to understand the dataset structure and characteristics:

- Class distribution analysis
- Image quality inspection
- Verification of labeling consistency
- Detection of corrupted or duplicate images
- Confirmation of strict separation between training and testing data

Key Observation:
The dataset is suitable for deep learning training but requires augmentation and class-balancing strategies to ensure fair learning and generalization.

---

4. Data Preprocessing

4.1 Image Preparation

To standardize the input data:

- All images were resized to **224 × 224 pixels**
- Pixel values were normalized to the range **[0, 1]**
- Color channels were retained to match pre-trained model requirements

4.2 Data Augmentation

To improve model generalization and reduce overfitting, several augmentation techniques were applied during training:

- Random rotation
- Horizontal flipping
- Zoom transformations
- Width and height shifting

4.3 Handling Class Imbalance

To mitigate the effects of class imbalance, **class weights** were introduced during training. This ensured that misclassification of minority or clinically important cases carried a higher penalty.

---

5. Model Architecture

5.1 Baseline CNN

A baseline CNN model was initially implemented to establish a performance benchmark. While this model demonstrated learning capability, its performance was insufficient for medical-grade classification, leading to the adoption of transfer learning.

5.2 Transfer Learning with EfficientNet

The final architecture is based on **EfficientNetB0**, selected for its optimal balance between performance and computational efficiency.

Architecture Components:

- Pre-trained EfficientNetB0 backbone (initially frozen)
- Global Average Pooling layer
- Batch Normalization
- Fully connected dense layer
- Dropout regularization
- Sigmoid activation for binary classification

This architecture significantly improved feature extraction while minimizing overfitting.

---

6. Training Strategy

6.1 Optimization Configuration

- Optimizer: Adam
- Initial learning rate: 1e-4
- Fine-tuning learning rates: 1e-5 to 3e-6
- Loss function: Binary Cross-Entropy

6.2 Callbacks

To ensure controlled and efficient training, the following callbacks were implemented:

- Early stopping to prevent overfitting
- Learning rate reduction upon validation plateau
- Model checkpointing to preserve the best performing weights

6.3 Fine-Tuning

Once the classifier head converged:

- Selected upper layers of EfficientNet were unfrozen
- Training continued with a reduced learning rate
- This allowed domain-specific feature refinement without destabilizing learned representations

---

7. Model Evaluation

7.1 Evaluation Metrics

Due to the medical context, evaluation focused on clinically meaningful metrics:

- Accuracy
- Precision
- Recall (Sensitivity)
- F1-Score
- ROC-AUC
- Confusion Matrix analysis

Accuracy alone was not considered sufficient, as **false negatives represent a serious clinical risk**.

7.2 Performance Results

- Validation accuracy exceeded **95%**
- High recall for pneumonia cases
- ROC-AUC values consistently above **0.9**
- Balanced precision-recall trade-off suitable for medical screening

---

8. Confusion Matrix Analysis

The confusion matrix was analyzed to assess real diagnostic behavior:

- True positives indicated correctly detected pneumonia cases
- False negatives were minimized to reduce missed diagnoses
- False positives were accepted within reasonable bounds to favor patient safety

This analysis provided deeper insight beyond numerical metrics.

---

9. Decision Threshold Optimization

Instead of relying on the default probability threshold (0.5), alternative thresholds were evaluated to improve sensitivity:

- Lower thresholds increased recall
- Slight reduction in precision was considered acceptable in a medical context
- The final threshold was selected based on clinical safety considerations

---

10. Medical Interpretation of Results

The trained model demonstrates a strong ability to distinguish between normal and pneumonia-affected lungs. Emphasis on recall ensures that potentially dangerous cases are less likely to be overlooked, making the system suitable as a **clinical decision support tool** rather than a replacement for medical professionals.

---

11. Limitations and Future Work

Current Limitations:

- Single dataset dependency
- Lack of patient metadata integration
- Binary classification scope

Future Enhancements:

- Use of larger EfficientNet variants
- Ensemble modeling approaches
- Integration of explainability tools such as Grad-CAM
- Incorporation of clinical patient data

---

12. Conclusion

This project successfully demonstrates the application of deep learning techniques to medical image classification. By combining transfer learning, careful preprocessing, and clinically relevant evaluation strategies, the resulting model achieves high diagnostic reliability. The approach aligns with real-world medical constraints and serves as a strong foundation for future medical AI applications.