Automated Disease Detection in Medical Images using Deep Learning

Capstone Project

Syed Hussain Ather

Problem Statement and Context

Problem Statement: Early and accurate detection of diseases such as pneumonia and other conditions from chest X-rays is crucial for effective treatment and patient outcomes.

Context: Radiologists face high workloads, which can lead to diagnostic errors and delayed treatment.

Success Metrics: Improve diagnostic accuracy and efficiency using automated systems.

Stakeholders: Hospitals, radiologists, patients.

Constraints: Data privacy, model interpretability, computational resources.

Project Formulation as a Data Science Problem

Goal: Develop a deep learning model to detect and classify diseases from chest X-ray images.

Approach: Use convolutional neural networks (CNNs) to build the model.

Challenges: Large dataset, imbalance in class distribution, ensuring model generalizability.

Dataset Description

Dataset: NIH Chest X-ray Dataset

Data Source: NIH public dataset containing over 100,000 chest X-ray images labeled with 14 different conditions.

Key Features: Image paths, disease labels.

Data Wrangling Steps

Loading Data: Images and labels from the dataset.

Handling Missing Values: Ensure all data entries are complete.

Data Augmentation: Applied to increase dataset size and variability (e.g., rotation, zoom, flip).

Key Takeaways from Data Wrangling

Balanced Dataset: Addressed class imbalance.

Data Normalization: Improved model performance by normalizing pixel values.

Preprocessed Data: Ready for training and evaluation.

Exploratory Data Analysis

Image Distribution: Visualization of the number of images per class.

Pixel Intensity Analysis: Histogram of pixel intensities.

Key Findings from EDA

Class Imbalance: Significant imbalance between different disease categories.

Image Quality: Variations in image quality and resolution.

Visual Insights

Sample Images: Examples of normal and diseased chest X-rays with annotations.

Additional EDA Insights

Correlation Analysis: Correlation between different diseases.

Feature Engineering: Extracted features for model training.

Modeling Approach

Chosen Models: Simple CNN and Deeper CNN.

Why CNN?: CNNs are effective in image classification due to their ability to capture spatial hierarchies.

Model Architecture

Simple CNN: Architecture overview (Conv2D, MaxPooling, Dense).

Deeper CNN: Architecture overview (additional Conv2D and Dense layers).

Model Architecture

Simple CNN: Architecture overview (Conv2D, MaxPooling, Dense).

Deeper CNN: Architecture overview (additional Conv2D and Dense layers).

Model Performance

Accuracy: Comparison of training and validation accuracy for both models.

Loss: Training and validation loss comparison.

Best Model Selection

Final Model: Selected based on highest validation accuracy.

Model Evaluation: Precision, recall, and F1-score for the best model.

Recommendations

Implementation: Deploy the best model in clinical settings for assisting radiologists.

Further Validation: Validate the model on additional datasets to ensure generalizability.

Continuous Improvement: Regular updates and retraining with new data.

Practical Considerations

Ethical Concerns: Ensuring patient data privacy and model fairness.

Technical Challenges: Addressing computational requirements and latency.

User Training: Training radiologists to effectively use the automated system.

Future Work

Enhancements: Explore advanced architectures like ResNet or DenseNet.

Additional Data: Incorporate more diverse datasets for training.

Real-time Processing: Optimize the model for real-time inference.