CS 583 — Deep Learning April 12, 2024

CS583 Final Project Report

Lecturer: Zhaozhuo Xu *By: Hemalatha Katta*

1 What problem do we want to solve?

The goal of this project is to develop an automated deep learning system for detecting pneumonia from chest X-ray images. Pneumonia diagnosis using radiographic images is often time-consuming and requires expert interpretation. By training deep learning models to classify X-rays as either *NORMAL* or *PNEUMONIA*, we aim to build a fast, accurate, and scalable tool that can support clinicians in early detection and reduce diagnostic burden in high-volume medical settings.

2 What datasets did you use?

We used the publicly available **Chest X-Ray Pneumonia Dataset** from Kaggle, containing pediatric chest X-ray images. The dataset includes two classes: *NORMAL* and *PNEUMONIA*. The data is organized into training, validation, and test directories.

- **Training set:** 5216 images (1341 Normal, 3875 Pneumonia)
- **Validation set (created via 15% split):** 782 images
- **Test set:** 624 images (234 Normal, 390 Pneumonia)

Preprocessing steps included resizing images to 224×224, normalizing pixel intensities, and applying data augmentation (rotation, shift, shear, zoom, and horizontal flip). Class weights were computed to address class imbalance.

3 What models have you tried?

Two models were implemented and evaluated:

(a) CNN (Baseline Model)

A custom convolutional neural network built with:

- Four convolutional blocks (32, 64, 128, 256 filters)
- Batch Normalization and ReLU activations
- MaxPooling layers for downsampling
- Dropout (0.25–0.5) for regularization
- GlobalAveragePooling2D
- Fully connected layer with 256 units

This model has approximately 457k parameters.

(b) Transfer Learning Model: DenseNet121

A significantly more advanced model leveraging ImageNet-pretrained DenseNet121:

- Base model frozen to retain learned features
- Added classification head:
  - GlobalAveragePooling2D
  - Batch Normalization
  - Dropout layers
  - Dense layer with 256 ReLU units
  - Sigmoid output layer
- Approximately 265k trainable parameters

DenseNet121 was selected due to its strength in medical image feature extraction and ability to reuse information through dense connections.

4 How to evaluate the performance of the model on your dataset?

Model performance was evaluated using both quantitative metrics and visualization tools. The following evaluation methods were implemented:

- **Accuracy and Loss** on training, validation, and test sets
- **Precision, Recall, and F1-score** for each class
- **Confusion Matrix** to analyze prediction errors
- **Sensitivity (Recall for Pneumonia)** and **Specificity (Recall for Normal)**
- **ROC Curve and AUC** to measure discriminative ability
- **Training curves** (accuracy and loss) for overfitting detection
- **Class weights** to mitigate class imbalance
- **Early Stopping** and **Learning Rate Scheduler** to optimize training

These evaluation techniques provide a comprehensive understanding of model behavior and reliability.

5 How does your model perform?

Two models were trained and tested. Their final performance on the test dataset is:

CNN (Baseline)

- **Test Accuracy:** 62.50%
- **Test Loss:** 1.3582

The baseline CNN exhibited signs of underfitting and did not generalize well to unseen data.

DenseNet121 (Transfer Learning)

- **Test Accuracy:** 89.10%
- **Test Loss:** 0.2564
- **Sensitivity (Pneumonia Recall):** 90.3%
- **Specificity (Normal Recall):** 87.2%
- **AUC:** 0.960 (Excellent classifier)

Confusion Matrix Summary:

- True Negatives: 204
- False Positives: 30
- False Negatives: 38
- True Positives: 352

DenseNet121 substantially outperforms the CNN model with a performance gain of more than 26% in accuracy. The high AUC indicates a strong ability to distinguish between pneumonia and normal chest X-rays.

6 How does your AI teammate perform?

The AI teammate, represented by the DenseNet121-based classifier, performs as a reliable and efficient assistant for medical professionals.

- It provides fast and consistent predictions across large datasets.
- High sensitivity ensures that pneumonia cases are detected early.
- Strong specificity reduces false alarms and unnecessary follow-up testing.
- The model reduces radiologist workload by pre-screening X-rays.
- It enhances clinical decision-making by acting as a second reader.
- Visualization tools such as ROC curves, confusion matrices, and sample predictions make the system transparent and interpretable.

Overall, the AI teammate acts as a supportive diagnostic partner that improves accuracy, reduces human fatigue, and helps prioritize patients needing urgent care.