Report – Foundations of AI – Final Project (CS5100)

Reena Sajad Hyder

Pneumonia Detection on X-Ray Images of Chest

I. Introduction:

Lung diseases, including pneumonia, significantly impact breathing and overall health. Pneumonia alone accounts for over 15% of deaths in children under five worldwide and remains a leading cause of death in the U.S., with over 50,000 fatalities in 2015. Accurate diagnosis often requires expert review of chest radiographs (CXRs), which can be complicated by overlapping conditions and imaging variability. This project leverages machine learning to detect pneumonia from CXRs, aiming to serve as a prescreening tool for healthcare providers in under-resourced areas, enabling faster detection and improved recovery rates.

The goal of this project is to develop a deep learning-based solution using CNN (Convolutional Neural Network) to detect a visual signal for pneumonia in medical images. Specifically, it needs to automatically locate lung opacities on CXRs (given in DICOM format) and use that to detect pneumonia. The dataset is sourced from the Radiological Society of North America who collaborated with the US National Institutes of Health, The Society of Thoracic radiology, and MD.ai to develop this dataset. It is a binary classification task where the model predicts whether or not the patient has pneumonia, based on the CXRs and the associated bounding box data for the regions impacted by pneumonia. CXRs are the main diagnostic tool for detecting pneumonia but manual diagnosis is often time consuming and also prone to errors. Automating this process with the help of a deep learning-based solution can potentially assist radiologists by highlighting the pneumonia affected regions, thus increasing diagnostic accuracy and speed.

This project addresses the limitations of manual diagnosis, such as time constraints and human error, by automating the classification process through a deep learning pipeline. By preprocessing medical images to normalize and extract key visual features, the system not only identifies pneumonia but also has scope of highlighting the affected regions, providing an explainable AI tool for radiologists. The integration of bounding box data enhances detection accuracy, ensuring the model is capable of isolating key areas of interest in complex CXR images. Essentially, it seeks to bridge the gap between advanced diagnostic tools and the accessibility challenges faced by healthcare systems worldwide, particularly in low-resource settings.

II. Methods and Approach:

To develop an efficient solution for pneumonia detection from CXRs, this project follows the following structure: data preparation, model development and evaluation. It is based on using deep learning techniques and leverages a comprehensive dataset.

1. Data Preparation

The dataset consists of chest X-rays in DICOM format, along with associated metadata and bounding box annotations for regions affected by pneumonia. The dataset was large (consisted of 30000 images) and hence the compressed zip file was used. The data in the compressed ZIP format was extracted and organized into directories for training and testing. The raw data consists of DICOM images and associated metadata in CSV files, containing bounding box annotations and diagnostic labels. The main steps in this stage were splitting the dataset and directory creation. The dataset was split into 80% training and 20% testing sets. Images were copied into separate directories for training and testing to facilitate efficient data loading

2. Preprocessing

The main goal in this stage was ensuring consistency in image dimensions and scaling. A custom function was implemented to read DICOM files, resize them to 224×224 pixels, and normalize pixel values to a range of [0,1]. Metadata was aggregated at the patient level, combining bounding box annotations for cases with multiple pneumonia regions. Binary labels were assigned to indicate the presence or absence of pneumonia. ‘has_pneumonia’ being 0 meant no pneumonia and 1 meant pneumonia.

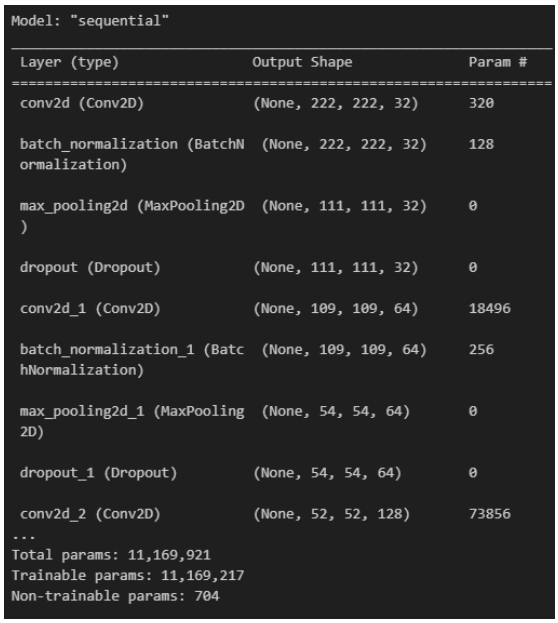
3. Data Augmentation and Generator Design

A custom data generator was developed using the keras.utils.Sequence library. This generator dynamically loaded and preprocessed images in batches during training and validation. The goal of this approach was to optimize memory usage by loading the images on demand. It applied preprocessing transformations uniformly across the dataset.

4. CNN Model Architecture:

The main goal of this project was to build a convolution neural network designed to classify CXRs as having or not having pneumonia. The CNN model developed had the following enhancements. L2 Regularization and dropout layers were added to reduce overfitting. Batch normalization was implemented after convolutional layers to stabilize training and improve generalization. ReLu activation function was used for hidden layers, a sigmoid activation function was used in the output layer for binary

classification. Adaptive learning was done through the Adam optimizer and binary cross-entropy was used as the loss function.



```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 222, 222, 32)      320

 batch_normalization (BatchN  (None, 222, 222, 32)     128
 ormalization)

 max_pooling2d (MaxPooling2D  (None, 111, 111, 32)     0
 )

 dropout (Dropout)           (None, 111, 111, 32)      0

 conv2d_1 (Conv2D)           (None, 109, 109, 64)      18496

 batch_normalization_1 (Batc  (None, 109, 109, 64)     256
 hNormalization)

 max_pooling2d_1 (MaxPooling  (None, 54, 54, 64)       0
 2D)

 dropout_1 (Dropout)         (None, 54, 54, 64)        0

 conv2d_2 (Conv2D)           (None, 52, 52, 128)       73856
...
Total params: 11,169,921
Trainable params: 11,169,217
Non-trainable params: 704
```

Figure 1: CNN Model

5. Training and Validation

The model was trained using the prepared training set and evaluated on a validation split. The following callbacks were employed. ModelCheckPoint was used to save the best performing model based on validation loss. EarlyStopping was used to halt training if validation performance stopped improving thus preventing overfitting. ReduceLrOnPlateau was used to reduce the learning rate when validation loss plateaued.

6. Testing and Evaluation

The best model was saved and evaluated on the test dataset. Threshhold Optimization was carried out by analyzing predictions at various thresholds to identify an optimal balance between sensitivity and specificity. The performance of the model was evaluated with the help of metrics such as the confusion matrix, precision, recall, F1 score, and AUC-ROC. Confusion matrix heatmap and ROC curve visualization was done for further insights.

7. Visualization and Analysis

Training and validation curves were plotted to analyze the convergence and detect potential overfitting. The testing results were visualized with the help of accuracy vs threshold plots, confusion matrix heatmaps and ROC curves, AUC serving as the key performance indicator.

This comprehensive approach was useful in ensuring that the model was robust, generalizable and suited well for deployment in clinical settings.

III. Results and Analysis:

The CNN Model demonstrated strong performance in classifying pneumonia in CXRs. The model achieved a final **training accuracy** of **80%.** The validation accuracy of the model stabilized at 81.15%, showing that the model had good generalization without overfitting, The training and validation loss curves highlight initial stability but converge well, indicating effective optimization.

The model was evaluated on 5337 test samples, achieving an accuracy of 80% and the precision, recall and F1-Scores were also good. The ROC-AUC score was 0.81, showing good discrimination between the classes. The results indicate that the model is highly accurate in detecting cases without pneumonia but still struggles slightly with sensitivity (recall) for pneumonia cases. This could be due to overlap in visual features between classes.

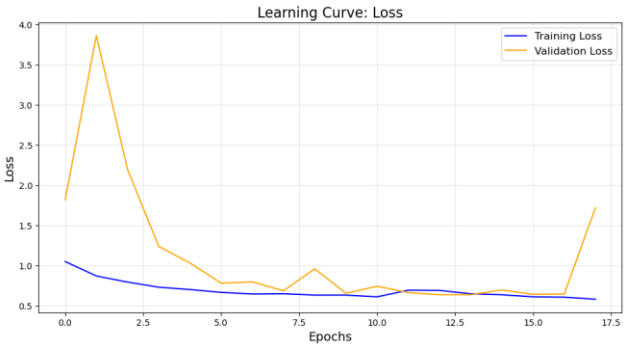


Figure 2: Learning Curve: Loss

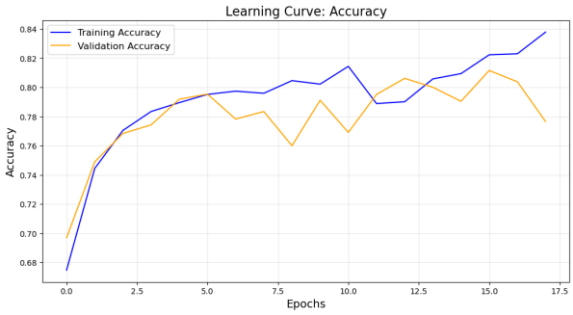


Figure 3: Learning Curve: Accuracy

The visualizations above have provided the following insights. The loss learning curve shows that training and validation loss decreased significantly in the initial epochs. However, slight overfitting in later epochs was addressed by early stopping. The accuracy learning curve shows that both training and validation accuracy steadily improved, with a narrow gap, reinforcing the model's stability.

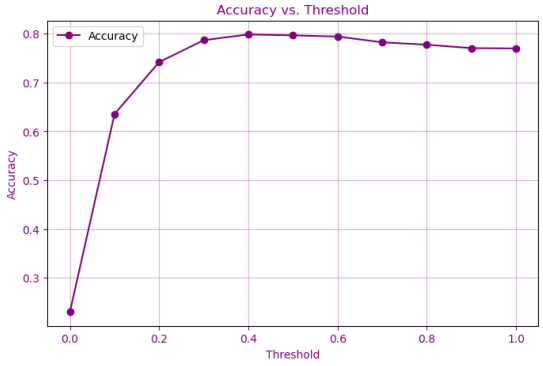


Figure 4: Accuracy vs Threshold

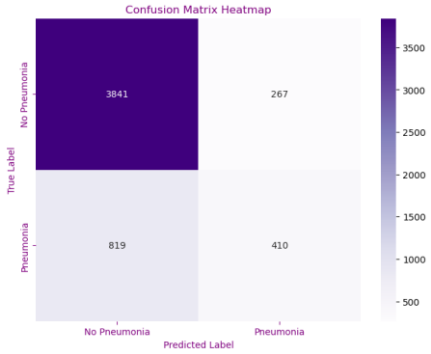


Figure 5: Confusion Matrix Heatmap

The optimal classification threshold of 0.5 yielded an accuracy of 80%. Adjusting the threshold beyond this did not improve model performance, showing that the model is robust. In the confusion matrix, there were 410 True Positives, 819 False Negatives, 3841 True Negatives and 267 False Positives. These figures highlight the model's strength in identifying the non-pneumonia cases (high specificity) and also reveals its weakness in identifying all pneumonia cases (low sensitivity).

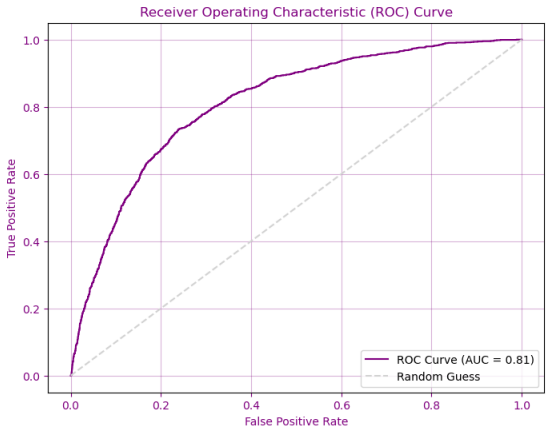


Figure 6: ROC Curve

The ROC curve illustrates the strong separation between the classes, with an AUC score of 0.81. This demonstrates the model's ability to distinguish between the different cases under varying thresholds.

IV. Conclusion and Reflection:

This project successfully developed a deep learning-based solution using CNN for detecting pneumonia from chest X-ray images, achieving a test accuracy of over 80% and AUC-ROC score of 0.81. The results highlight the model's capability to accurately identify non-pneumonia cases (high specificity) while underscoring challenges in achieving higher sensitivity for pneumonia detection. The implementation of a robust CNN architecture with regularization, dropout, and batch normalization ensured stable training and effective generalization.

Despite the promising results, the project faced challenges such as imbalance in class distribution, which affected recall for pneumonia cases. This limitation suggests the need for further exploration of techniques like data augmentation, class weighting, and incorporation of bounding box annotations to enhance the model's sensitivity. Additionally, integrating explainability tools like Grad-CAM could make the solution more interpretable for clinical use, thereby increasing trust and adoption among healthcare providers.

Looking forward, this project demonstrates the potential of machine learning in medical imaging to assist in timely and accurate diagnoses, particularly in resource-limited settings. Future directions include fine-tuning the model for higher recall, expanding the dataset to include diverse cases, and adapting the solution to highlight regions of interest in pneumonia-affected lungs. With these advancements, the model

could serve as a valuable tool for radiologists, enhancing diagnostic accuracy and efficiency, and ultimately contributing to improved patient outcomes in the fight against pneumonia.

References:

[1] Radiological Society of North America (RSNA). (2018). *RSNA Pneumonia Detection Challenge Dataset.* Available at: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge

[2] Litjens, G., Kooi, T., Bejnordi, B. E., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42,* 60–88. https://doi.org/10.1016/j.media.2017.07.005

[3] Stephen, O., Sain, M., Maduh, U. J., & Jeong, D. U. (2019). An efficient deep learning approach to pneumonia classification in healthcare. *Journal of Healthcare Engineering, 2019.* https://doi.org/10.1155/2019/4180949

[4] Rajpurkar, P., Irvin, J., Zhu, K., et al. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv preprint arXiv:1711.05225.*

[5] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767.*

[6] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6(1),* 60. https://doi.org/10.1186/s40537-019-0197-0

[7] Selvaraju, R. R., Cogswell, M., Das, A., et al. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision, 128,* 336–359.

[8] He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering, 21(9),* 1263–1284.

[9] Shanmugam, P., Subramaniam, U., & Yang, L. T. (2020). A review on normalization techniques for medical image analysis. *Pattern Recognition Letters, 134,* 184–192. https://doi.org/10.1016/j.patrec.2020.02.005

[10] World Health Organization (WHO). (2019). *Pneumonia Fact Sheet.* Available at: https://www.who.int/news-room/fact-sheets/detail/pneumonia