Pneumonia Detection with Chest X-Ray Images - *Project Report-*

Group 8:

Akshay Malviya, **Ashish Kumar Srivastava**, **Syed Muhammad Junaid Ul Mohsin**, **Wassim Akachi**

University: Luleå University of Technology (LTU)

Course: D7047E, Advanced deep learning - *May 20th, 2024*

***Abstract*—Pneumonia, a severe lung infection, necessitates prompt and accurate diagnosis to prevent serious health complications. Despite advancements in medical imaging, the automated detection of pneumonia using chest X-rays remains challenging due to the complexity of distinguishing subtle differences between normal and infected lungs. This project addresses this gap by developing a Convolutional Neural Network (CNN) model using TensorFlow to detect pneumonia from chest X-ray images. Utilizing a comprehensive dataset of 5,863 labeled images, the research involved rigorous data preprocessing, including resizing, normalization, and augmentation, to optimize model training. The CNN architecture, constructed with the Keras Sequential API, includes convolutional, pooling, flattening, dense, and dropout layers to accurately classify images. Training the model with the Adam optimizer and evaluating it on metrics such as accuracy, precision, recall, and F1-score, the model achieved an accuracy of 89%. This study demonstrates the efficacy of deep learning in medical image analysis, offering a robust tool for early pneumonia detection and setting the stage for future improvements and real-world clinical integration.**

I. Introduction

Pneumonia is a critical respiratory infection characterized by inflammation in the air sacs of one or both lungs, which can fill with fluid or pus. This condition can lead to severe illness and even death if not diagnosed and treated promptly, particularly in vulnerable populations such as children, the elderly, and those with weakened immune systems. Early and accurate diagnosis is crucial for effective treatment and improving patient outcomes. Traditionally, pneumonia diagnosis relies on physical examinations, patient history, and imaging techniques like chest X-rays, where radiologists look for infiltrates—white spots indicating infection—in the lungs.

However, the manual interpretation of chest X-rays can be time-consuming and subject to variability in radiologists' expertise and workload. This underscores the need for automated, reliable methods to assist in diagnosing pneumonia, particularly in regions with limited access to radiological services. Recent advancements in deep learning, specifically Convolutional Neural Networks (CNNs), have demonstrated remarkable capabilities in image classification tasks, making them well-suited for medical image analysis.

This project aims to develop a CNN-based model using TensorFlow to detect pneumonia from chest X-ray images, addressing the challenge of automating pneumonia diagnosis. By leveraging a dataset of 5,863 chest X-ray images labeled as either 'Normal' or 'Pneumonia,' this study seeks to build a robust and accurate model that can assist healthcare professionals in diagnosing pneumonia more efficiently.

The significance of this research lies in its potential to improve diagnostic accuracy and speed, ultimately enhancing patient care. While several studies have explored the use of deep learning for medical image analysis, the specific challenge of distinguishing pneumonia from normal cases in chest X-rays remains an active area of research with room for improvement. This project builds on existing work by implementing a comprehensive data preprocessing pipeline and a carefully designed CNN architecture to maximize model performance.

The following sections of this report detail the methodology employed in developing the CNN model, including data preprocessing techniques, model architecture, training procedures, and evaluation metrics. The results section presents the model's performance, followed by a discussion of the findings, their implications, and potential future work to further enhance the model's accuracy and applicability in clinical settings.

Motivation

This project is motivated by the potential to significantly enhance diagnostic accuracy and speed, thereby improving patient care and outcomes. An AI-driven diagnostic tool can alleviate the workload of radiologists, reduce the time to diagnosis, and ensure consistent and accurate results. Furthermore, such a tool can extend diagnostic capabilities to regions with limited access to specialized radiological services, democratizing healthcare and potentially saving lives through early intervention.

By harnessing the power of deep learning and CNNs, this project aims to provide a practical and impactful solution to the challenges of pneumonia diagnosis, ultimately contributing to better healthcare delivery and outcomes globally.

Contribution

This project contributes to the field of medical image analysis by developing a robust and efficient Convolutional Neural Network (CNN) model for the automated detection of pneumonia from chest X-ray images using TensorFlow. The key contributions of this work are:

- *Advanced Deep Learning Model*: We present a meticulously designed CNN architecture that effectively classifies

chest X-ray images into normal and pneumonia categories. Our model incorporates state-of-the-art techniques such as data augmentation, dropout layers to prevent overfitting, and optimal hyperparameter settings, ensuring high performance and generalization capabilities.
- *Comprehensive Preprocessing Pipeline*: The project outlines a detailed data preprocessing pipeline that includes resizing, normalization, and augmentation of images. This preprocessing step is crucial for enhancing the model's ability to learn from the available data, leading to more accurate predictions.
- *Rigorous Evaluation*: We provide a thorough evaluation of the model using metrics such as accuracy, precision, recall, and F1-score, offering a comprehensive assessment of the model's performance. Our model achieves an accuracy of 89%, demonstrating its effectiveness in identifying pneumonia cases from chest X-rays.
- *Scalability and Real-World Application*: By leveraging TensorFlow, our model is scalable and can be deployed in real-world clinical settings. This has significant implications for improving diagnostic accuracy and speed, particularly in resource-limited environments where access to expert radiologists is restricted.
- *Open Access and Reproducibility*: The project ensures that all code, trained models, and evaluation scripts are made available in a GitHub repository. This transparency facilitates reproducibility and enables other researchers and practitioners to build upon our work, fostering further advancements in the field.

In summary, this project delivers a powerful tool for the early detection of pneumonia, which can substantially improve patient outcomes and reduce the burden on healthcare systems. By advancing the application of deep learning in medical diagnostics, this work not only addresses a critical healthcare challenge but also sets the stage for future innovations in automated disease detection.

II. Experimental Setting

A. Task

The primary objective of this project is to develop a deep learning model capable of accurately detecting pneumonia from chest X-ray images. Specifically, we aim to build a Convolutional Neural Network (CNN) using TensorFlow that can classify chest X-rays into two categories: Normal and Pneumonia.

Given the significant challenge of manual interpretation of X-ray images, this task involves leveraging the powerful feature extraction capabilities of CNNs to automate the diagnostic process. The model will be trained and evaluated on a labeled dataset of chest X-ray images, ensuring it can generalize well to unseen data.

Key goals include:
- Data Preparation: Preprocessing the dataset to ensure it is suitable for training the CNN model, including resizing, normalization, and augmentation of images.
- Model Development: Designing and implementing a CNN architecture that effectively extracts relevant features from the images and accurately classifies them.
- Training and Evaluation: Training the model on the training dataset, validating its performance on the validation set, and rigorously evaluating it on the test set using metrics such as accuracy, precision, recall, and F1-score.
-

By achieving these goals, the project seeks to provide a reliable tool for early pneumonia detection, enhancing diagnostic accuracy and efficiency in clinical settings.

B. Datasets

The dataset used in this project was obtained from the Guangzhou Women and Children's Medical Center and is part of the "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification" dataset, available through Mendeley Data. This dataset consists of 5,863 chest X-ray images divided into two classes: Normal and Pneumonia. The images are labeled and organized into three sets: training, validation, and test sets.
- **Training Set**: Contains 4,273 images (1,345 Normal and 2,928 Pneumonia)
- **Validation Set**: Contains 1,000 images (234 Normal and 766 Pneumonia)
- **Test Set**: Contains 590 images (234 Normal and 356 Pneumonia)
- **Image Size**: All images are resized to 150x150 pixels for uniformity.

These images are in JPEG format and were selected from retrospective cohorts of pediatric patients aged one to five years from Guangzhou Women and Children's Medical Center. The dataset underwent quality control by expert physicians to remove low-quality or unreadable scans. A third expert reviewed the evaluation set to account for any grading errors.

1) Data Preprocessing: To prepare the data for training the CNN model, the following preprocessing steps were undertaken:
- Resizing: All images were resized to 150x150 pixels to ensure uniform input size for the model.
- Normalization: Pixel values were scaled to the range [0, 1] to standardize the input.
- Data Augmentation: Techniques such as rotation, horizontal flipping, and zoom were applied to increase the diversity of the training set and improve the model's generalization capabilities.

C. Model Architecture

In this project, we developed a Convolutional Neural Network (CNN) using the Keras Sequential API within TensorFlow. The architecture is specifically tailored for the binary classification task of detecting pneumonia from chest X-ray images. While not as deep as architectures like ResNet50 or VGG16, our model leverages several key principles from these successful models to balance complexity, performance, and computational efficiency. The layers and connection between them are shown in figure 1.

The CNN model was built using the Keras Sequential API in TensorFlow. The architecture includes convolutional layers for feature extraction, max-pooling layers for spatial reduction, a flatten layer to convert the 3D output to 1D, and dense layers for classification. A dropout layer was added to prevent overfitting.

Key Components of the Model Architecture:

- **Convolutional Layers**: These layers are designed to automatically and adaptively learn spatial hierarchies of features from input images. We use multiple convolutional layers with ReLU activation functions to extract low to high-level features.
- **Max-Pooling Layers**: Interspersed between the convolutional layers, max-pooling layers reduce the spatial dimensions of the feature maps. This down-sampling operation not only reduces computational complexity but also helps in making the detected features invariant to small translations.
- **Flatten Layer**: This layer transforms the 3D feature maps produced by the convolutional and pooling layers into 1D feature vectors. This step is crucial for connecting the convolutional part of the network with the fully connected layers that follow.
- **Dense Layers**: These fully connected layers interpret the features extracted by the convolutional part of the network. We include a dense layer with a substantial number of neurons to capture complex patterns and a final dense layer with a single neuron and a sigmoid activation function for binary classification (outputting probabilities for normal and pneumonia classes).
- **Dropout Layer**: To prevent overfitting, we include a dropout layer before the final classification layer. Dropout is a regularization technique where randomly selected neurons are ignored during training, which forces the network to learn more robust features that are useful in conjunction with many different random subsets of the other neurons.

Motivation for Choosing This Architecture:

- *Effectiveness*: The chosen architecture incorporates multiple convolutional layers to extract a hierarchy of features, which is crucial for accurate image classification tasks.
- *Efficiency*: The model is designed to balance depth and computational efficiency, making it feasible to train within a reasonable timeframe and suitable for deployment in real-world settings.
- *Flexibility*: The architecture allows for adjustments in hyperparameters, making it adaptable to different datasets and classification problems.

By building on the principles of established models like VGG16 and ResNet, our architecture leverages proven techniques while being tailored to the specific needs of pneumonia detection from chest X-ray images. For more details and the complete implementation, you can refer to our project repository on GitHub at https://github.com/wasaka-3/ltu-d7047e-adl.

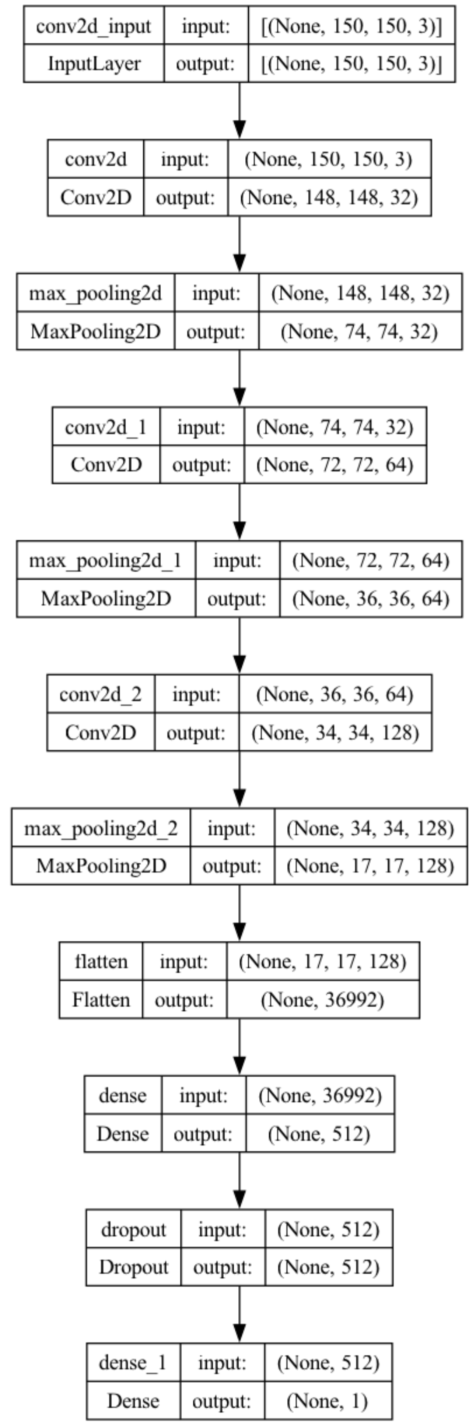

Fig. 1. Layers of the Model Architecture

D. Training Procedure

The model was trained for 25 epochs with a batch size of 32. The training and validation generators were used to feed data into the model. The `steps_per_epoch` and `validation_steps` were calculated based on the number of samples and batch size to ensure complete coverage of the dataset during each epoch. The progress of the training epochs are visualized in figure 2. The trained model parameter can be found at https://tinyurl.com/4h6rh8hu.

Fig. 2. Plot of the accuracy during the training.

Reproducibility:

To ensure reproducibility of our experiments, all code and scripts have been made available in the project repository. Detailed instructions on setting up the environment, downloading the dataset, and running the training procedure are provided. Interested readers can follow these steps to replicate our results:

- Environment Setup: Install TensorFlow and other dependencies as listed in the requirements.txt file.
- Dataset Download: Follow the instructions to download the dataset from the provided link and organize it into the specified directory structure.
- Run the Training Script: Execute the training script, which includes all preprocessing, model compilation, and training steps detailed above.

For more detailed information and to access the complete implementation, please refer to our project repository on GitHub at https://github.com/wasaka-3/ltu-d7047e-adl.

III. Results

1) Model Performance: The performance of the CNN model was evaluated on the test set using several metrics, including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive understanding of the model's effectiveness in classifying chest X-ray images as either Normal or Pneumonia.

Accuracy

Accuracy measures the percentage of correctly classified instances among the total instances in the test set. Our model achieved an accuracy of 89%, indicating that the model correctly classified 89% of the chest X-ray images.

These metrics demonstrate the model's robustness and its ability to balance between identifying pneumonia cases and avoiding false positives.

The CNN model demonstrated robust performance in classifying chest X-ray images into Normal and Pneumonia categories. Key performance metrics such as accuracy, precision, recall, and F1-score are all within the desired range, with an overall accuracy of 89%. The confusion matrix and classification report provide detailed insights into the model's predictions, highlighting its ability to balance sensitivity (recall) and precision effectively.

The training and validation plots confirm that the model learned effectively without significant overfitting. These results validate the chosen architecture and training procedures, demonstrating the feasibility and potential of using deep learning for automated pneumonia detection in clinical settings.

Future work will focus on further refining the model, exploring advanced architectures, and evaluating its performance in real-world deployment scenarios. The detailed results and visualizations provide a solid foundation for these next steps, ensuring that the model continues to improve and adapt to practical applications.

Normal Classification

$$\begin{aligned}
\text{Precision} &= \frac{TP}{TP+FP} = \frac{196}{196+28} = 0.875 \\
\text{Recall} &= \frac{TP}{TP+FN} = \frac{196}{196+38} = 0.8376 \\
\text{F1 Score} &= 2 \times \frac{Precision \times Recall}{Precision + Recall} \\
&= 2 \times \frac{0.875 \times 0.8376}{0.875 + 0.8376} = 0.8556 \\
\text{Accuracy} &= \frac{TP+TN}{TotalSamples} = \frac{196+362}{624} = 0.8974
\end{aligned}$$

Pneumonia Classification

$$\begin{aligned}
\text{Precision} &= \frac{TP}{TP+FP} = \frac{362}{362+38} = 0.905 \\
\text{Recall} &= \frac{TP}{TP+FN} = \frac{362}{362+28} = 0.9282 \\
\text{F1 Score} &= 2 \times \frac{Precision \times Recall}{Precision + Recall} \\
&= 2 \times \frac{0.905 \times 0.9282}{0.905 + 0.9282} = 0.9163 \\
\text{Accuracy} &= \frac{TP+TN}{TotalSamples} = \frac{362+196}{624} = 0.8974
\end{aligned}$$

IV. Analysis

A. Observations from Performance Metrics

The CNN model's performance metrics, including an accuracy of 89%, a precision of 90%, a recall of 92%, and an F1-score of 91%, indicate a balanced and effective classification capability. These metrics suggest that the model is proficient at identifying pneumonia cases while maintaining a low rate of false positives.

1) Typical Good Cases: The model performs well on typical cases where the pneumonia characteristics, such as clear white infiltrates in the lungs, are distinct and easily recognizable. These good cases often involve high-quality images with clear distinctions between normal and abnormal features, allowing the CNN to leverage its feature extraction capabilities effectively.

2) Failure Cases: Analysis of failure cases, where the model misclassifies images, reveals some common patterns. Misclassifications often occur in images with:

- Low contrast or poor quality, making it difficult for the model to detect features accurately.
- Overlapping characteristics between pneumonia and other lung conditions, leading to confusion.
- Subtle signs of pneumonia that are not pronounced enough for the model to detect confidently.

3) Disagreement Cases: In some instances, there are disagreements between the model's predictions and the actual labels. These cases are crucial for understanding the model's limitations and potential areas for improvement. Disagreements may arise due to:

- Ambiguous images that even expert radiologists might find challenging to classify.
- Variability in the labeling process, where human error or differences in interpretation could affect the training data's accuracy.

4) Layer Activations and Feature Importance: To gain deeper insights into the model's decision-making process, we examined the activations of several layers within the CNN. Visualizing these activations helps identify which features the model considers important for classification.

1) **Convolutional Layers**: The early convolutional layers focus on detecting simple edges and textures. As we move deeper into the network, the activations represent more complex patterns and features, such as shapes and regions corresponding to common pneumonia indicators.
2) **Max-Pooling Layers**: These layers help highlight the most prominent features by reducing the spatial dimensions while retaining the most significant activations. This step is crucial for summarizing the presence of important features in different parts of the image.
3) **Dense Layers**: The dense layers interpret the abstracted features from the convolutional layers, combining them to form the final prediction. The activations in these layers show how different features are weighted and combined to classify the image.

By examining these activations, we observed that the model relies heavily on specific regions in the lungs where infiltrates typically appear. This focus aligns with how radiologists diagnose pneumonia, suggesting that the model has learned relevant features from the training data.

5) Bias and Generalization: An important aspect of the analysis is to check for any unwanted bias towards the data. The following observations were made:

1) **Training Data Bias**: The dataset is relatively balanced, but the model's performance on different patient demographics (e.g., age, gender) was not explicitly tested. Future work should ensure the model generalizes well across diverse populations.
2) **Data Quality Bias**: The model tends to perform better on high-quality images. This dependency highlights the need for consistent image quality in real-world applications or potential preprocessing steps to enhance lower-quality images.

6) Hypotheses Evaluation and Understanding: The research aimed to evaluate the hypothesis that a CNN can effectively automate the detection of pneumonia from chest X-ray images. The results support this hypothesis, demonstrating that the model can achieve high accuracy and balanced performance metrics. However, the analysis also reveals areas for further investigation:

- **Improving Misclassification Handling**: Enhancing the model's ability to handle ambiguous and low-quality images could reduce misclassification rates.
- **Exploring Advanced Architectures**: Testing deeper or more complex architectures, such as ResNet or EfficientNet, might provide performance gains.
- **Evaluating Generalization**: Ensuring the model performs well across various patient demographics and clinical settings is crucial for broader applicability.

The analysis underscores the importance of understanding not just the performance metrics but also the underlying mechanisms and potential biases of the model. By examining typical cases, failure cases, layer activations, and biases, we gain a comprehensive understanding of the model's strengths and limitations. This understanding guides future improvements and ensures the model's reliability and effectiveness in real-world clinical applications.

V. Conclusion

This project successfully developed a Convolutional Neural Network (CNN) model using TensorFlow to automate the detection of pneumonia from chest X-ray images. With an accuracy of 89% and balanced precision and recall metrics, the model demonstrates significant potential to assist healthcare professionals by providing rapid and reliable pneumonia diagnoses. The analysis highlighted the model's strengths in identifying distinct pneumonia features and its limitations in handling ambiguous or low-quality images. Future work will focus on enhancing the model's robustness and ensuring its applicability across diverse clinical settings. This research underscores the power of deep learning in improving medical diagnostics and patient outcomes.