DEEP LEARNING FOR PNEUMONIA DETECTION USING CHEST X-RAYS

Grace Lin
Student# 1009885929
grrace.lin@mail.utoronto.ca

Ethan Wang Jan
Student# 1010123331
ethan.jan@mail.utoronto.ca

Yun Zhan Edwyn Zhou
Student# 1009948933
edwyn.zhou@mail.utoronto.ca

Renzo Ugarte Basurco
Student# 1010331570
renzo.ugarte@mail.utoronto.ca

ABSTRACT

Healthcare providers are interested in an efficient diagnostic method for pneumonia as traditional methods rely on expert radiologists manually analyzing chest X-rays, a process that is not only time-consuming but also prone to interobserver variability. The goal is to leverage machine learning techniques to provide a preliminary screening tool that flags potential cases of pneumonia. We propose a convolutional neural network (CNN) model for pneumonia detection and classification using chest x-ray images. The model will extract spatial and geometric features from the images to be able to classify them into three categories: normal, bacterial pneumonia, and viral pneumonia. To optimize feature extraction, the model will employ convolutional and max pooling layers, followed by passing it through an Artificial Neural Network (ANN) to transform the feature space into a linearly separable problem. By leveraging deep learning, this project aims to improve the efficiency and scalability of medical image analysis, while maintaining the indispensable role of human experts in making the final diagnostic decisions. Through this research, this project aims to demonstrate the potential of AI-driven image analysis in supporting more efficient pneumonia classification.

1 INTRODUCTION

Human health relies on timely and efficient medical diagnosis, particularly for conditions that require rapid intervention. Pneumonia, a severe respiratory infection that targets the lungs, remains a major health concern, particularly in pediatric and immunocompromised patients (Kallander et al. (2016)). Early detection of pneumonia is crucial for effective treatment, yet traditional diagnostic methods- such as expert radiologists analyzing patient chest x-rays- can be time-consuming and subject to interobserver variability, where different healthcare providers interpret the same chest x-ray differently (Melbye & Dale (1992)). With the growing availability of medical imaging data, machine learning techniques such as deep learning offer a promising solution to automating key aspects of the diagnostic process.

This project explores the application of deep learning techniques to classify chest x-ray images into normal, viral pneumonia, and bacterial pneumonia patients. The goal of this model is not to replace expert radiologists in pneumonic diagnosing, but rather, to provide a preliminary screening tool that flags potentially concerning cases, allowing for faster prioritization and review by medical professionals.

To accomplish this, the team will leverage the efficiency of a Convolutional Neural Network (CNN) to extract spatial and geometric features from the chest x-ray images. By passing these processed features to a fully connected Artificial Neural Network (ANN) it will be transformed into a linearly separable problem, facilitating classification. The use of neural networks in this way offers a scalable and efficient approach to medical imaging analysis, without overlooking the indispensable role of human expertise in final diagnosis decisions.

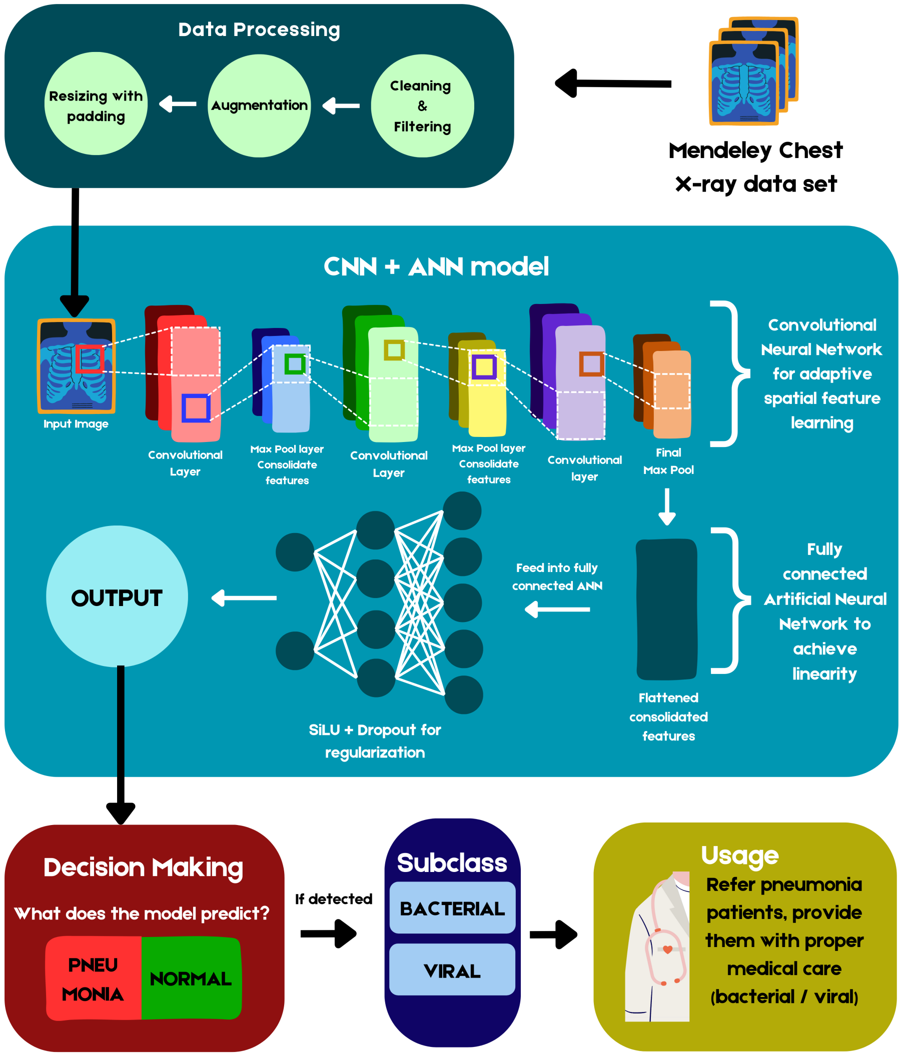

Figure 1: Project Overview

2 Background and Related Work

The use of deep learning techniques for detecting and classifying pneumonia is continuously advancing, driving significant progress in medical imaging analysis. The following section explores key developments in the field, highlighting various machine learning algorithms and neural network models and their unique approaches to enhancing diagnostic efficiency.

2.1 CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Pranav Rajpurkar (2017) at the Stanford Machine Learning Group developed a 121-layer CNN that takes a chest x-ray input and outputs a probability of pneumonia as well as a heatmap that localizes the areas on the x-ray most indicative to pneumonia. The model is trained on ChestX-ray14, a large dataset of chest x-rays from over 30,000 unique patients. They found that CheXNet's performance exceeds the capabilities of the average radiologist in accurately detecting pneumonia in chest x-rays, with a significantly higher F1 score (performance metric combining precision and recall) of 0.435 compared to 0.387.

2.2 Seven classification and regression models to predict pneumonia among schizophrenic patients

Kuang Ming Kuo (2019) predicted pneumonia among schizophrenic patients by applying 11 different features to seven classification and regression models including decision trees, k-nearest neighbors, naïve Bayes, random forest, support vector machines, and logistic regression. They achieved the highest accuracy rate of 94.5%, using a decision tree classifier. However, their research suffers from low generalizability as their models were evaluated on small datasets.

2.3 Simple CNN with Data Augmentation

Stephen O (2019) developed a simple CNN architecture to classify pneumonic chest x-rays. Because they were not able to obtain a large image dataset, they used data augmentation to improve the validation accuracy of the model. To augment the data, they applied various transformations such as rotations and scaling to existing images to increase the diversity of the training dataset. They achieved an accuracy of 93.73%.

2.4 Viral Pneumonia Screening on Chest X-Rays Using Confidence-Aware Anomaly Detection

Zhang J (2021) proposed the Confidence-Aware Anomaly Detection (CAAD) model consisting of a shared feature extractor, an anomaly detection module, and a confidence prediction module to detect anomaly cases classified as viral pneumonia. By using binary classification, they only determined only the anomaly cases. They achieved a 83.61% area under the curve (AUC) score on their dataset, which they deemed to be comparable to human radiologists.

2.5 Computer-aided diagnosis system for automatic pneumonia detection

Rohit Kundu (2021) developed an ensemble of three CNN models: GoogLeNet, ResNet-18, and DenseNet-121. Ensemble learning is a strategy in which the decisions of multiple classifiers are fused to obtain the final prediction for a test sample. Their model was evaluated on two pneumonia x-ray datasets using five-fold cross-validation. Using accuracy as their performance metric, their model produced 87.02%, and 98.8% accuracy for their two datasets.

3 Data Processing

The Kermany (2018) Kaggle dataset consists of images split into "NORMAL" and "PNEUMONIA" folders. "PNEUMONIA" also has subclasses that include either bacteria or viral pneumonia. The dataset is split into 89/1/10 training, validation, and test sets. The training dataset is imbalanced such that there are three times as many pneumonia images as normal images in the training set. The following data processing techniques will be employed when applicable:

3.1 Data Preprocessing and Cleaning

The images will be labelled as either "NORMAL", "PNEUMONIA - BACTERIAL", or "PNEUMONIA - VIRAL". The images will be stored as Numpy arrays and then be stored in Pandas data frames using Python. The images will then be resized to a consistent resolution by padding the images to a specific dimension. In the case of noise, the team will implement noise reduction techniques to clean the data.

3.2 Data Normalization and Standardization

Images will be standardized to provide the neural network with consistent data. The images will then be normalized to prevent the neural network from learning non-existent relations and patterns in the data due to differences in magnitude.

3.3 Data Augmentation

The dataset can be balanced by performing data augmentation to artificially create more normal data points and develop a more robust model. Techniques include rotations, random cropping, and flipping on existing images to increase the diversity of the training dataset.

4 Architecture

Following data processing, a Neural Network model will be used to classify samples into normal, bacterial pneumonia, and viral pneumonia patients. For this project, the dataset consists of images of chest x-rays. The dimensions of 1 sample are big enough to make a standard fully connected network not appealing, due to the computational cost and depth it might require to accurately discern between important features.

Thus, a Convolutional Neural Network (CNN) will take the input, identify important spatial/geometric features and consolidate information alternating between convolutional layers and max pooling layers. After convolution, consolidated features will be flattened and passed to a fully connected neural network (ANN) to convert the feature space into a linearly separable problem. Finally, the model will classify the sample into normal, bacterial or viral pneumonia. See figure 2 for an illustration of the model.

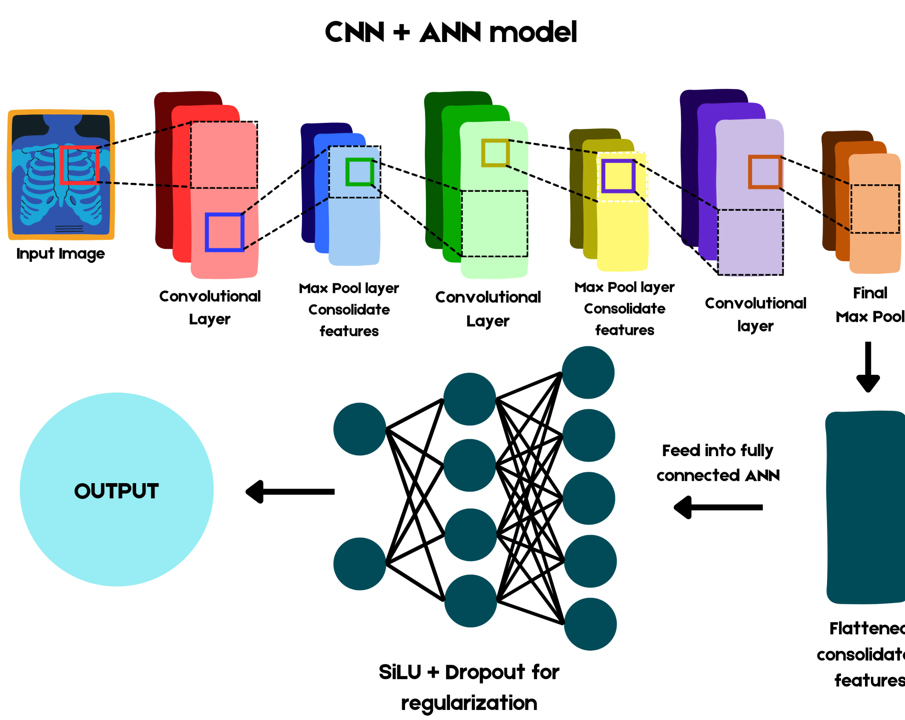

Figure 2: Proposed Model Architecture

5 Baseline Model

A Support Vector Machine (SVM) model will be used as a baseline measure for model performance. The data processing will be consistent with the neural network, and features such as mean, standard deviation, and accuracy will be recorded. With a linear kernel, the SVM model aims to diagnose images into either “NORMAL” or “PNEUMONIA” classes. Another SVM model will be implemented to further classify whether “PNEUMONIA” classes are either “BACTERIAL” or “VIRAL” subclasses. A linear model serves as an efficient and simplified control model to compare performance and computational cost against.

6 Ethical Considerations

Given the proposed pneumonia diagnosis model works with x-ray images of chest scans, there are predominant privacy issues regarding the dataset. The dataset contains sensitive medical information on real patients, and as a result makes it difficult to collect new data to inference on given the privacy considerations for patients.

The Kaggle dataset also states that ”chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children’s Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients’ routine clinical care.”

Given the patients are all one to five years old, data bias is present in the dataset. Specifically, the adult and elderly population is not represented in the data, and will likely result in poor model performance on age ranges outside of pediatric patients. The dataset is also specifically concerning pediatric patients from Guangzhou, which lacks diversity. Considering bone structure differs across race and ethnicity, the model may not be robust when testing across diverse individuals.

As the model aims to diagnose patients, there is an ethical risk in false negatives occurring, as overlooking patients with pneumonia and leaving the condition untreated can result in death and harm to the individual.

The model also has its limitations. As the model is trained to specifically classify normal and pneumonia cases, it may not generalize well when there are co-occurrences of other diseases. Varying changes in pneumonia cases may also cause the model to misclassify, as pneumonia differs in early and late stages.

7 Project Plan

This team will work together by having weekly Saturday meetings to discuss project and task progress, allocate tasks among group members, and work out any difficulties and challenges that have been encountered. For workload management, group members are assigned major tasks and responsible for distributing and creating the necessary subtasks for successful completion by the associated deadline. This process can be seen below in figure 3 where task management was carried out for data processing of the progress report.

Progress Report				
- Data Processing	Edwyn	2/8/2025	3/7/2025	Not Start...
-> Create Script	Renzo	2/8/2025	3/7/2025	Not Start...
-> Clean Data	Edwyn	2/8/2025	3/7/2025	Not Start...
- Baseline Model	Ethan	2/8/2025	3/7/2025	Not Start...
- Primary Model	Grace	2/8/2025	3/7/2025	Not Start...

Figure 3: Assigned Subtasks and Management

Members responsible for the major task are associated with subtask delegation as they have done the most research on the topic, giving them the best perspective of what needs to be completed and how it needs to be completed to fit the project requirements. This structure will be used because throughout the project, focus will be given to each subtask at different points of the timeline. Assigning subtasks is also more flexible, allowing team members to work on tasks that can be completed within their schedule.

All progress and completion will be kept track of using Google Sheets. Each team member has completed and is responsible for the tasks seen in figure 4. Future tasks and deadlines are outlined in figure 3. Softwares and methods for communication and sharing are outlined in Table 1.

Tasks	Assigned To	Start Date	End Date	Status
Team Formation				
- Find Group members	Ethan, Renzo, Grace, Edwyn			Complete
Project Proposal				
- Introduction	Grace	1/17/2025	02/07/25	Complete
- Illustration/Figure	Renzo	1/17/2025	02/07/25	Complete
- Background & Related Work	Grace	1/17/2025	02/07/25	Complete
- Data Processing	Edwyn	1/17/2025	02/07/25	Complete
- Architecture	Renzo	1/17/2025	02/07/25	Complete
- Baseline Model	Edwyn	1/17/2025	02/07/25	In Progress
- Ethical Considerations	Edwyn	1/17/2025	02/07/25	In Progress
- Project Plan	Ethan	1/17/2025	02/07/25	In Progress
- Risk Register	Ethan	1/17/2025	02/07/25	Complete
- Link Github Repo	Edwyn	1/17/2025	02/07/25	Complete
- References	Grace	1/17/2025	02/07/25	In Progress

Figure 4: Task Completion and Schedule

Method	Purpose
Google Drive	Written File Management
Github	Code Management
Google Sheets	Project and Task Scheduling
Discord	Messaging and Communication

Table 1: Software selected for project development

8 Risk Register

Table 2 describes the various types of risk associated with the project and possible steps to solving them.

Risk Factor/Per Type	**Likelihood**	**Solution**
Teammate: Drops the course	10% Is an extra course on top of regular 2nd year coursework, though mandatory for certain degrees.	1) Reallocate work and reconsider internal deadlines. 2) More frequent check-ins to ensure work is completed on time.
Model: Incapable of producing output	50% Likely to happen given computational resource limitations and human error.	1) Consider splitting data further into epochs. 2) Ensure scripts work individually before implementation.
Model: Bias towards specific groups of data	20% Likely to happen during the beginning of training when we use a small amount of data but will happen less as we utilize more data.	1) Utilize multiple datasets. 2) Augment data/create our own data.
Data: Incorrect processing of data	40% Sound and noise can be easily added or data can be incorrectly altered due to human coding error.	1) Process data in small chunks to ensure correct output. 2) Develop a program filter to remove noise and clean data.
Technical: Merge and Pull conflicts leading to lost data	60% Likely to happen as some teammates are not experienced with Git and Push-Pull concepts.	1) Ensure communication during Git actions. 2) Save progress locally.

Table 2: Risk Management Plan

9 Link to Github repository

The following is the link to this project's GitHub repository: `https://github.com/edwynzhou/aps360_project_team_77`

References

Kaggle. Chest x-ray images (pneumonia). URL https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia.

Karin Kallander, Deborah H Burgess, and Shamim A Qazi. Early identification and treatment of pneumonia: a call to action. *The Lancet Global Health*, 4(1):e12–e13, 2016. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5357734/.

Daniel S. et al. Kermany. Identifying medical diagnoses and treatable diseases by image-based deep learning. *CellPress*, 172(5):1122 – 1131.e9, 2018. URL https://www.cell.com/cell/fulltext/S0092-8674(18)30154-5.

Chi Hsien Huang Liang Chih Cheng Kuang Ming Kuo, Paul C Talley. Predicting hospital-acquired pneumonia among schizophrenic patients: a machine learning approach. 2019. URL https://pubmed.ncbi.nlm.nih.gov/30866913/.

H. Melbye and K. Dale. Interobserver variability in the radiographic diagnosis of adult outpatient pneumonia. *National Library of Medicine*, 33(1), 1992. doi: 10.1177/028418519203300117. URL https://pubmed.ncbi.nlm.nih.gov/1731850/.

Kaylie Zhu Brandon Yang Hershel Mehta Tony Duan Daisy Ding Aarti Bagul Curtis Langlotz Katie Shpanskaya Matthew P. Lungren Andrew Y. Ng Pranav Rajpurkar, Jeremy Irvin. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *National Library of Medicine*, 2017. URL https://stanfordmlgroup.github.io/projects/chexnet/.

Zong Woo Geem Gi-Tae Han Ram Sarkar Rohit Kundu, Ritacheta Das. Pneumonia detection in chest x-ray images using an ensemble of deep learning models. 2021. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0256630.

Maduh UJ Jeong DU Stephen O, Sain M. An efficient deep learning approach to pneumonia classification in healthcare. *National Library of Medicine*, 2019. URL https://pubmed.ncbi.nlm.nih.gov/31049186/.

Pang G Liao Z-Verjans J Li W Sun Z He J Li Y Shen C Xia Y Zhang J, Xie Y. Viral pneumonia screening on chest x-rays using confidence-aware anomaly detection. *National Library of Medicine*, 2021. URL https://pubmed.ncbi.nlm.nih.gov/33245693/.