Rhea Samuel (rss3488)
Nicole Olvera (no4342)
Ira Narang (in2933)

Machine Learning Initial Proposal

Topic

Early and accurate prediction of pneumonia is vital for effective patient care and to prevent long-term implications of this infection. In this project, we aim to develop an image classification model to detect pneumonia from X-ray images of pediatric patients. To do this, we will optimize neural network architectures to maximize image classification accuracy on a publicly available dataset containing images of patients with normal lungs and patients diagnosed with pneumonia. A well-tuned model can assist medical experts in diagnosing patients and potentially automate the diagnosis process. In medical settings where access to expert interpretation is limited, this model could help diagnose cases that require immediate attention.

Problem Statement

In order to find the best model configuration, we will employ hyperparameter tuning approaches and neural network architecture search to classify X-ray pictures into normal and pneumonia categories. Our goal for this project is to maximize accuracy while also maintaining model efficiency and interpretability.

Dataset

Our dataset, Pneumonia Dataset, contains 5,863 images with 2 categories: normal and pneumonia. The image below is an example of those in the dataset.

The image on the left depicts clear lungs without any abnormalities, the image in the middle (pneumonia) exhibits a focal global consolidation (pointed by white arrows), and the image to the right also depicts pneumonia with a more diffuse interstitial pattern in both lungs.

High-Level Methods

We will be using the following technologies and techniques for our project. We will be testing which parameters and search techniques provide the optimal performance on our dataset.

- Framework: Keras
- Image Preprocessing: Rescaling, normalization
- Model Building: Keras CNN
- Search Techniques: Random Search, Hyperband, and Bayesian Optimization
- Evaluation Metrics: Accuracy, precision, recall, F1-score
- Hyperparameter Tuning: We will be using Keras Tuner to explore:
    - Number of layers
    - Activation functions
    - Optimizers

We will also be deploying our project within Docker and will use a Flask API, including the following curl commands.

- GET /summary: Returns a summary of the metadata from our model
- GET /hyperparameters: Returns the optimal number of hyperparameters for our model
- POST /inference: Returns whether the image is pneumonia or normal.

Deliverables

Finally, the primary deliverables for this project will be the following,

- Jupyter notebook: different variations on our model based on our tools and techniques
- Dockerfile and Docker Compose file: deployment and containerization
- Flask API file, api.py: the different curl commands
- Pneumonia dataset files: train, test, images to test
- Final report: our optimal model variation and other results
- README: information to deploy model and test different endpoints