Computational Methods for Biomedical Image Analysis

Assignment 2

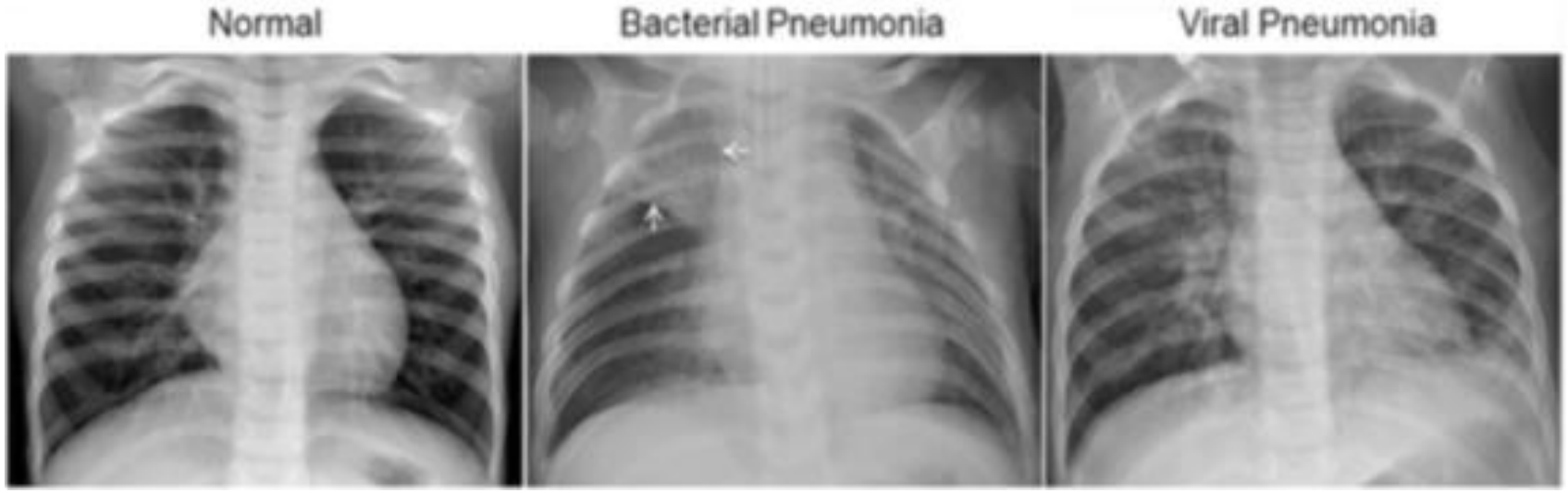
Pneumonia Classification with Chest X-Ray Images

Introduction

- Machine learning can help to detect pneumonia cases from chest X-ray images. The task is a simple classification problem where given an input chest X-ray image, the machine learning-based model must detect whether the subject of study has been infected or not.
- In this assignment, we will learn
  - Image preprocess
  - Classification methods
  - Evaluation metrics

Dataset

- The chest X-Ray dataset contains normal, bacterial pneumonia, and viral pneumonia cases, with a total number of 5,863 X-Ray images.
- For simplicity, we only have to classify it into normal class or pneumonia class.



Dataset (cont.)

- The dataset is organized into 3 folders (train, val, test). Train and val contain subfolders for each image category (pneumonia/normal). Test folder contain images without label.

- Please download the dataset from the following link: https://drive.google.com/file/d/14WBR3W0VIA8xaO57zhCVSoStQiV18uE5/view?usp=sharing

Part 1 (10%)

- Read all training data and apply preprocessing such as normalize, resize, or augment. Print out the image size, min, max and mean of both original image and the pre-processed image (You will only need to show the value of one example image). (10%)

Part 2 (40%)

- Please try at least one classifier to classify whether the image is normal or pneumonia.
- Please use the validation set to evaluate the performance.
- Classification method:
  - K-means Clustering
  - Support Vector Machine
  - Random Forest Classifier
  - Convolution Neural Network
  - Any other existing methods

Part 2 (Cont.)

- Classification method: (40%)
  - Please list the detail settings of your method. (20%)
  - Please evaluate the classification accuracy of the validation data through your method. Please report the following four metrics. (5% for each)

  - Evaluation metrics:
    - Accuracy = (TP+TN) / (TP+FN+TN+FP)
    - Sensitivity (Recall) = TP / P
    - Precision = TP / (TP+FP)
    - ROC curve
    - https://en.wikipedia.org/wiki/Precision_and_recall

Part 3 (40%)

- Output the predict result of test data to csv file (5%).
- Please refer to the example csv and follow the format: https://drive.google.com/file/d/1oNnPAh8I9YBn4hV71b3lZz0nn7_tPMOf/view?usp=sharing

Part 3 (Cont.)

- Accuracy scoring: (35%)
  - Baseline: If you achieve 0.6 accuracy (15%)
  - Ranking:
    - Your test accuracy with be compared with all students in class.
      - If your accuracy is in top one-third (20%)
      - If your accuracy is between top one-third and last one (15%)
      - If your accuracy is in last one-third (10%)

Report (10%)

- Please write your report with the given latex template in English.
- Show your print out result inside the report.
- Try to give more explanation and discussion for your assignment.
- List out what package you had used.
- Please give a summary about what you have learned.

Hand in Rules

- Please zip your code (.py) and report (.pdf) together and name as {Student_id}_{Name}.zip
- The test result in Part 3 should be name as {Student_id}_{Name}.csv
- Deadline: Upload to iLMS before 2021/5/3 23:59 (UTC+8).
- Late submission: 20% penalty.
- Wrong hand in format: 10% penalty.
- This is an individual assignment, no cheating.

Hint

- You can write your report with Overleaf, an online latex editor.
- You can write your code with jupyter notebook but convert to (.py) when you hand in.