

Machine Learning

Course Project Proposal

Group - 43 Jaskirath Singh (2018150), Sarandeep Singh (2018364), Simarjeet Singh (2018103)

1. Motivation and precise problem statement

- **Motivation** - According to WHO, there are about 4 million cases of premature deaths annually due to air pollution-related diseases like Pneumonia. Detection of Pneumonia in the early stage can help patients in timely medical treatment. This has the potential for generalized high-impact application in biomedical imaging by rapidly reviewing immense amounts of images.
- **Problem Statement** - Identify if a person is suffering from pneumonia using an image of his/her chest x-ray.
- **Learning Task** - Binary classification

2. Dataset

- [Kermany, Daniel; Zhang, Kang; Goldbaum, Michael \(2018\). "Labeled Optical Coherence Tomography \(OCT\) and Chest X-Ray Images for Classification". Mendeley Data. V2. doi: 10.17632/rscbjbr9sj.2](#)

3. Preprocessing Techniques

- Data Refining and Image Resizing
- Morphological Transformations and Augmentation
- Feature extraction using Convolutional Neural Network (CNN)

4. Learning Techniques

- Logistic regression (Baseline)
- K-Nearest Neighbors
- XGBoost
- Neural Network

5. Strategy for model selection and hyperparameter tuning

- Hyperparameter Tuning: GridSearchCV
- Model Selection: Cross Validation Score

6. Training Approaches

- Newton Based Method
- Stochastic Gradient Descent

7. Ensemble Approaches

- Boosting

8. Evaluation Metrics and Error analysis

- Evaluation Metrics: Recall, Accuracy, and Area under ROC curve (AUC)
- Error Analysis: Model Bias, Variance and Ablation Analysis (in case of NN)

9. Deliverables - (Every member will analyze errors in their own models)

- Jaskirath Singh - Exploratory Data Analysis (EDA), Data Preprocessing and K-Nearest Neighbors
- Sarandeep Singh - EDA, Data Preprocessing, Logistic Regression, and XGBoost
- Simarjeet Singh - Feature Extraction and Selection, Neural Network, and Ablation Analysis