| ISM 2022w |
| --- |

| **Group Number : 07** | | |
| --- | --- | --- |
| Name | Role | Contact Information |
| Ahmad Saigol | Technical Expert/Programmer | [ahmad.saigol@tuhh.de](mailto:john.doe@tuhh.de) |
| Joel Romero | Programmer | [joel.romero.munoz@tuhh.de](mailto:joel.romero.munoz@tuhh.de) |

**Contributions - Phase 1**

* **Literature Review (Author: Everyone):**

Everyone reviewed different papers in literature to understand what kind of feature extractors have been successfully used to solve classification of lung diseases including COVID, as well as other similar problems. Furthermore, it was agreed to spend time on papers that were similar to our project and that could be implemented easily. To name a few, Ahmad studied “IKONOS: an intelligent tool to support diagnosis of COVID-19 by texture analysis of X-ray images” and Joel studied “COVID-19 detection using X-ray images and statistical measurements”.

Also, some of the members did not have much experience with python and had to brush up their skills.

* **Pipeline:**

Ahmad set up a scalable pipeline in Python which allowed the easy integration of new preprocessing methods, feature extractors, classifiers, evaluation metrics and plots by following a template. It also allowed the application of different classifiers at the same time. Moreover, It also saved all the important information for future use such as model, training and validation results, complete pipeline associated with the run and the plots. Furthermore, the code also allowed us to generate predictions (.txt file) with one function call. This pipeline allowed us to train and test different models very quickly.

* **Pre-processing and Feature Extractors:**

Ahmad worked on the splitting of data (that mainly included simple and simple stratified split) along with batch processing of the images and Joel worked on label encoding.

Depending upon which feature extractor we were testing, different kinds of preprocessing had to be done. Furthermore, different combinations of features were tested while playing with the hyperparameters.

Joel along with associated preprocessing (such as resizing, normalizing), worked on the basic extractors such as Contrast, Histogram, Skewness and Kurtosis. In addition, he also worked on an algorithm lung segmentation for better pre-processing. Due to the noise and different shapes of the images it was not successful.

Ahmad worked on data preprocessing methods such as Edge Detection and Gaussian blur (denoising) and on feature extractors such as Haralick and Zernike features.

* **Classifiers:**

We mainly worked on two classifiers: Ahmad worked on SVM and Joel worked on Random Forest Tree. We also played with the values of hyperparameters to study their effect on the results. Furthermore, we also worked on the boosting in an attempt to improve the results.

* **Evaluation metrics:**

Joel worked on the evaluation metrics which included accuracy, and balanced accuracy while Ahmad worked on the plotting of confusion matrix.

**Contributions - Phase 2**

* **Literature Review:**

Everyone reviewed different papers in literature to analyze different deep neural network approaches that have been used for the detection of COVID and other similar lung diseases using X-Rays. To name a few, Ahmad studied “A residual network-based framework for COVID-19 detection from CXR images” and Joel studied “A multi-modal bone suppression, lung segmentation, and classification approach for accurate COVID-19 detection using chest radiographs”.

Furthermore, members did not have experience with Pytorch and Kaggle. So they had brush skills in this regard.

* **Pipeline:**

Ahmad set up a scalable pipeline in Python which allowed the easy integration of new data transformations, models, evaluation metrics and plots by following a template. Moreover, It also saved all the important information for future use such as model, training and validation results, complete pipeline associated with the run and the plots. Furthermore, the code also allowed us to generate predictions (.txt file) with one function call. This pipeline allowed us to train and test different models very quickly.

* **Data pre-processing and Data Loading:**

We reused a couple of preprocessing from phase 1 (such as splitting data and label encoding).

Depending upon which model we were working on, different preprocessing was required and it was an individual's task to complete it.

Overall, it included: Normalization, Resizing, Cropping, Horizontal and Vertical Flip, Rotating, Affine, adjusting sharpness and contrast

Furthermore, Joel studied the noisy data set’s characteristics to tackle with the introduced noise. It was recommended, depending upon time and results, we will tackle it with online and offline data augmentation

* **Transfer Learning/Architecture:**

We tried to train a number of models but because of limited computational resources not all models could be trained/optimized in the way they should be. Further, different experiments were conducted by varying hyperparameters and data preprocessing methods, looking for the most effective results. Few of them included:

Ahmad developed a Neural Network that made use of the Haralick and Zernike features (obtained on Phase no. 1). Furthermore, he also trained a simple CNN network which had raw images as input. Also, He worked on ResNet18 and ResNet50 for transfer learning and trained their last layer.

Joel worked on MobileNet V2. He used it for transfer learning and trained its last layers.

* **Evaluation Metrics:**

. We reused everything from Phase 1.