# Lung Diseases Classification using X-ray images and machine learning

1st Ahmad Saigol
*dept. name of organization (of Aff.)*
*name of organization (of Aff.)*
Hamburg, Germany

2nd Joel Romero
*dept. name of organization (of Aff.)*
*name of organization (of Aff.)*
Hamburg, Germany
email address or ORCID

*Abstract*—This document is a model and instructions for LaTeX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.

*Index Terms*—component, formatting, style, styling, insert

## I. Introduction

### *A. Background and motivation*

Due to recent worldwide COVID-19 outbreak, diagnosis and appropriate treatment methods are required. First diagnosis and medications are important to prevent the disease from progressing.

### *B. Problem overview*

### *C. Objectives*

Classify X Ray Images of patients who may or not be affected by a respiratory disease. Two types of Classification were required: Binary and Multiclass Classification. Binary Classifications would determine whether a patient had Covid or not. Multiclass Classification would diagnose 4 different conditions: Healthy (Normal) lungs, Covid-19, Pneumonia and Lung Opacity. To tackle with this classification problem, this study was divided into two phases. During Phase 1, the use of feature extractors and supervised machine learning models was demanded for Classification. During Phase 2, the use of Deep Learning models was required.

## II. Materials and Methods

### *A. Dataset Description*

To perform this study a dataset of Chest X-Ray images inside different folders was provided [1]. A folder with 16930 X-Ray Images and its labels were provided to train the Machine Learning models. A folder with 4235 images was also provided to test the accuracy of the model. Finally, a folder with noisy dataset 4235 was provided to test the DL models developed in Phase 2.

### *B. Data Splitting*

The images of the training set were split into training and validation set.

### *C. Data Preprocessing*

To effectively process the information, different operations where performed to the images.

Normalization: It is commonly applied so feature values can be comparable. Min-max method takes the pixel values of the image to a range from 0 to 1, which is given by (1) [2].

$$X_{new} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (1)$$

Resizing: Images of the same dataset must be converted to the same size, so machine learning algorithms can handle them in a consistent manner. Neuronal Networks (NN) required a fixed input size which are optimal to the pretrained NN to be used. Making the images smaller may also be used to reduce the computational complexity of the model, and thus reducing the processing time. [3]

Bilateral Filter: Filtering denoises an image and better preserves the details. Bilateral filter performs a pixelwise operation that uses the neighborhood pixels to determine the spatial and intensity distance between them. [4]

### *D. Feature extraction*

Describe all the features used. (Histogram, Harlick)

### *E. Phase 1: ML-based Classification*

Describe all the ML models used (

*F. Phase 2: NN-based Classification*

Overview of NN and CNN used

*F. Metrics*

Overview of metrics used in results

## III. Results and Discussion

Main inconvenient was computational complexity of the models. Deeper analysis would require more time to train models.

## IV. Conclusion

References


[1] ISM, *Chest X-Ray images dataset.* [Online]. Available: http://gofile.me/5AIED/xNqNoKPxu

[2] Brett Lantz, *Machine Learning with R,* 2nd ed. Packt Publishing, 2015.

[3] Emre AVUÇLU, "COVID-19 detection using X-ray images and statistical measurements," 2022.

[4] Sidheswar Routraya, Arun Kumar Raya, Chandrabhanu Mishrab, "Image denoising by preserving geometric components based on weighted bilateral filter and curvelet transform," 2018.