

## Detection and Classification of Lung Diseases using Machine and Deep Learning Techniques

Syamala KPL\*, Niharika CS, Jenny AM, Pavani P

Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil-Nadu, India

Received Date: January 09, 2023 Accepted Date: February 09, 2023 Published Date: February 12, 2023

\*Corresponding author: Syamala KPL, Dept of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, TamilNadu, India, Email: poojitakkr@gmail.com

Citation: Syamala KPL, Niharika CS, Jenny AM, Pavani P (2023) Detection and Classification of Lung Diseases using Machine and Deep Learning Techniques. J Comput Sci Software Dev 2: 1-10

### Abstract

The change in the environment, pollution, and some unwanted daily habits, such as smoking, drinking, etc., can lead to many lung diseases, which need early detection. As a result of smoking, smokers and surrounding people are infected with lung diseases, especially by breathing problems. This paper proposes a website that takes the symptoms of the patient and determines if any disease is present and gives a grade that indicates how severe or moderate the disease is. In the case that the user has an x-ray image and wants a cross-verification, he or she can upload the image and view the results. Thus, the user input can be a string or an image. We focus on detecting chronic lung disease in an early stage, which, in turn, enhances the chances of recovery and survival. Our paper contains a hypothesis that utilizes deep learning and machine learning to predict diseases such as COVID-19, Tuberculosis, Pneumonia, and COPD. For Covid-19 and COPD, we achieved accuracy of 96.90%, 90.32% respectively using classification algorithms, and for image dataset, we obtained accuracy of 98.58% using EfficientNet B0, a deep learning algorithm.

**Keywords:** X-ray, Dataset, Machine Learning, Classification, Efficient Net B0, Deep Learning

## Introduction

Lung disease refers to a variety of medical conditions that cause the lungs to work inefficiently. The most common lung diseases are Asthma, COPD, Hypertension, Lung cancer, Pneumonia, Tuberculosis, Pulmonary edema. Among these diseases, in this paper we forecasted Pneumonia, Tuberculosis, Covid-19 and COPD using Machine Learning and Deep Learning techniques. Across the globe, lung diseases cause millions of

deaths each year and are straining healthcare systems. A timely diagnosis is imperative for enhancing long-term survival and improving the chance of recovery.

There are several causes of lung diseases including smoking, alcohol consumption, pollution. COPD affects 65 million people worldwide and kills 3 million people each year, making it the third most common cause of death. Almost 15 percent, or roughly one in seven, middle-aged, older, and adults have lung disease.



Figure 1: Causes of Lung Diseases

In Covid times recently, having a chronic lung disease meant you were at high risk for severe illness and complications, also causing deaths. One-fourth of covid-19 cases involve an infection that affects both lungs.

In this paper we present machine learning algorithms to determine the severity of diseases based on their symptoms,

as well as EfficientNet B0, a deep learning algorithm used to detect diseases based on chest X-ray images. We used different algorithms to detect the disease, among them we choose a best model with good accuracy rate. The development of deep learning technology on medical images, such as Chest X-rays, has shown great potential for detecting lung disease.

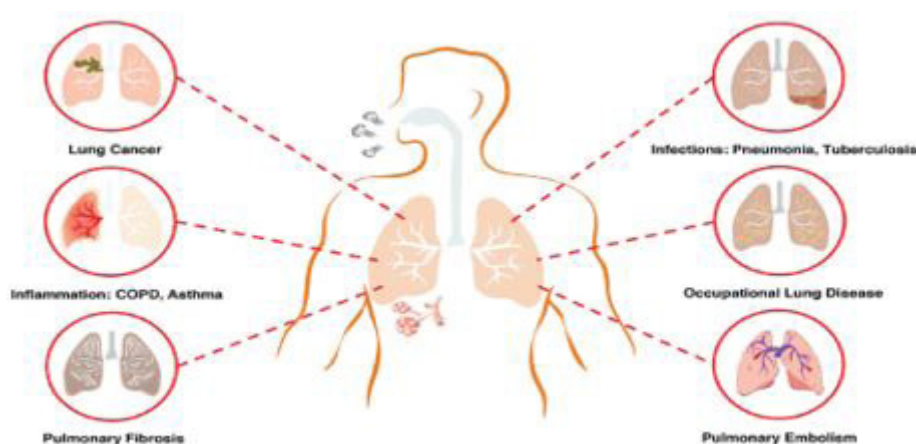


Figure 2: Different Types of Lung Diseases

## Problem Statement

In this project by using both Machine Learning and Deep Learning to take best features by combining the processing of patient information with data from symptoms as information and from X-ray images, using EfficientNetB0 as a well-trained model, to predict patient has a lung disease. As technology increases and world is changing so fast that the pressure on health is rapidly increasing due the changes in environment and climate which increased the risk of disease for people. One of the issues will be focused in this paper i.e., Lung diseases. This application is applied before the treatment of patient in health care systems and in addition patient information can provide better service during the treatment.

## Database

In this paper, we provide the work of experimental analysis of the proposed model on various popular lungdiseases datasets, such as COVID, TUBERCULOSIS, COPD DATASETS and CHEST X-RAY DATASET. This project uses different types of diseases datasets from Kaggle. Before moving into the results, we are going to give a brief overview of our datasets.

## COPD

Our COPD dataset include 101 patients and 24 variables. There is information on their characteristics variables such as AGE, GENDER and Smoking, disease severity, and co-morbidities. It's also having measures of their walking ability, quality of life, and anxiety and depression. The different stages of COPD in the dataset are taken as Gold1 to Gold4 as Mild, Moderate, Severe and VerySevere

## Tuberculosis

The tuberculosis dataset consists 16 columns with symptoms such as fever, coughing blood, chestpain, night sweats, weight loss etc., In Gender column male and female are indicated as 0 refers to women and 1 to men.

## COVID

In COVID dataset has different symptoms as Cough, Fever, Sore Throat, Shortness of Breath, Headache, persons with age above 60 and above, and Gender. The categorical data is pre-processed to convert to numerical data.

	test_date	cough	fever	sore_throat	shortness_of_breath	head_ache	corona_result	age_60_and_above	gender
0	2020-04-30	0	0	0	0	0	negative	None	female
1	2020-04-30	1	0	0	0	0	negative	None	female
2	2020-04-30	0	1	0	0	0	negative	None	male
3	2020-04-30	1	0	0	0	0	negative	None	female
4	2020-04-30	1	0	0	0	0	negative	None	male

**Figure 3:** Depicts the detail information of dataset

## Image dataset

The dataset has a total of 7135 x-ray images are present, which includes four different diseases as subfolders under train, test and Val, the subfolders for each image category as Normal, Pneumonia, covid-19, Tuberculosis. The EfficientNetB0 technique is used to detect and classify the disease from x-ray images

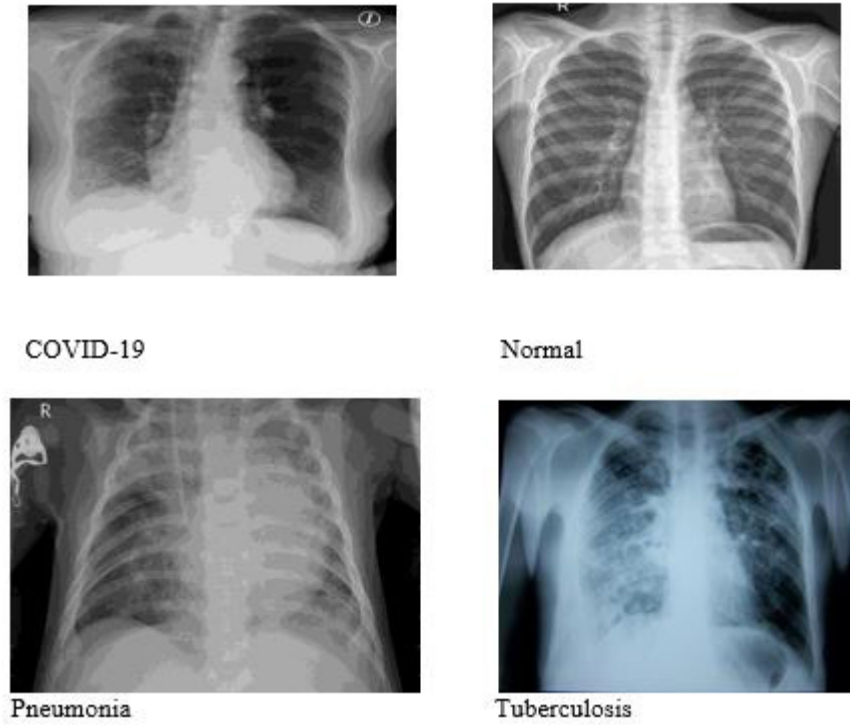


Figure 4: Some Sample pictures of the Image Dataset

### Proposed Methodology

In our model we worked on both csv dataset which contains disease symptoms and image dataset which consists of radiology images, for csv dataset we worked on 3 types of diseases such as Covid, Tuberculosis and COPD. In that for each dataset we use various algorithms such as decision trees for both Covid, kmeans Clustering for Tuberculosis respectively. For Image dataset we use EfficientNetB0 which is pre-defined ImageNet model.

For Image dataset we use activation function as ReLU for input layers and SoftMax for output layer.

In our model we built a website using streamlit a open-source python library used for creating and sharing webapps for Machine Learning projects. This website will be in private and it should be deployed in Heroku to make it as public. In the website the disease is predicted as positive or negative based on the input given by the patients.

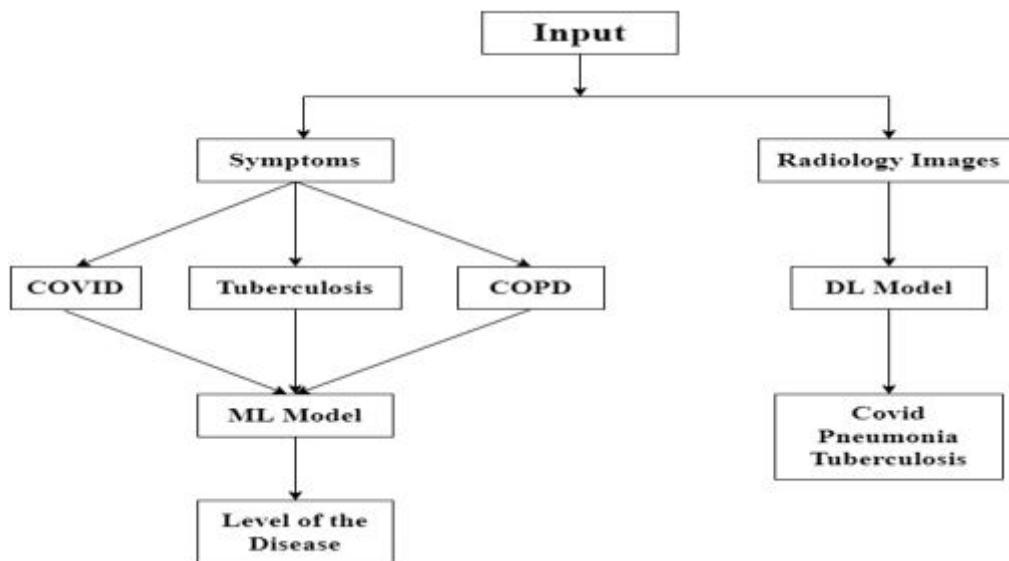


Figure 5: Steps to follow for Building CNN to data

## Implementation

Import all the required libraries for building the model. For numerical calculations import NumPy, for reading csv file import Pandas and machine learning algorithms as KNeighborClassifier, DecisionTreeClassifier for predicting the model, Keras for developing and evolution of deep learning models and we will import the dataset. Here, we also import some layers, some Keras library like dense, Conv2D, Maxpooling2D, Flatten, Dropout & keras applications as EfficientNetB0 Fig 9 shows the steps to follow for building CNN model.

In the dataset there are four types of diseases they are as Covid, Pneumonia, Tuberculosis and without any disease. After importing the images dataset, first step is to preprocess the data by creating a Data Frame with the filepath and the labels of the pictures.

Second step to train the images by labeling the columns and the filepaths and taking the target size.

Third step is to design an neural network model by initializing the Efficient Net B0 model which is used for creating the deep learning models for improving the efficiency and accuracy

In this model Efficient net b0 model is connected to different input layers such as Normalization, ZeroPadding2D, Batch Normalization, Conv2D all these layers are used to normalize the output of the previous layers. In the output layer, the SoftMax function serves as an activation function. The loss function used is categorical cross entropy for multi-class classification for giving two or more output labels. The optimizer used is the Adam which is a stochastic gradient descent method for training the model.

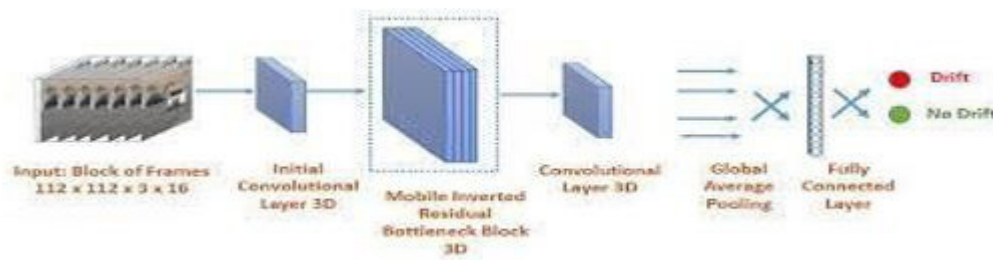


Figure 6: The concept of Convolutional Neural Network (CNN)

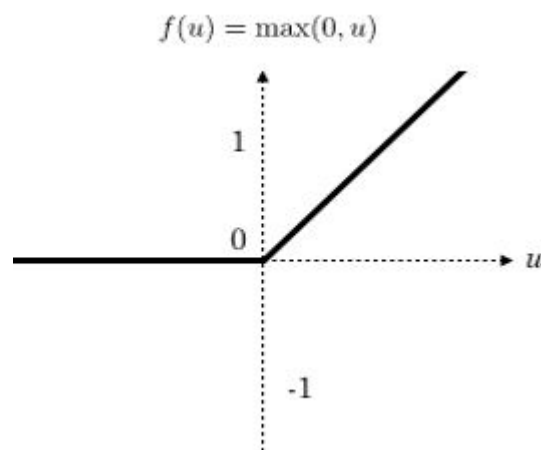


Figure 7: ReLu Activation Function

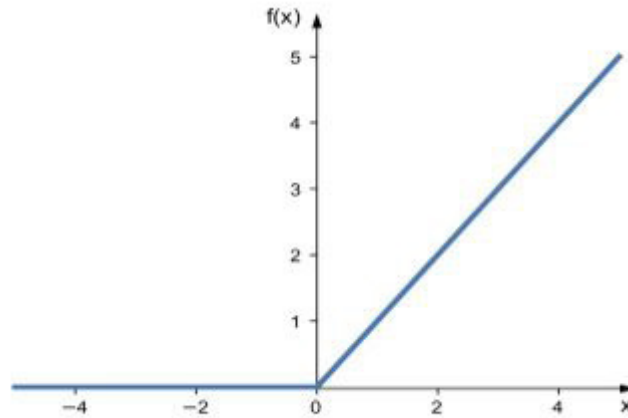


Figure 8: ReLu Activation Function Derivative

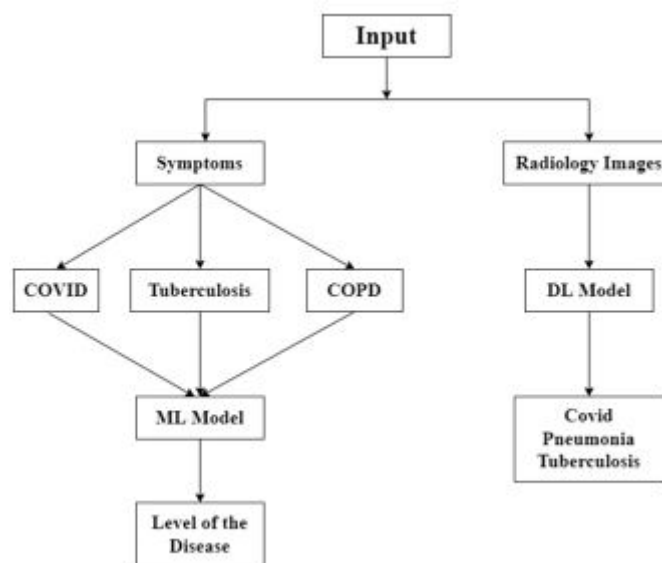


Figure 9: Steps to follow for Building CNN to data

The following are the layers of CNN:

**Convolutional Layer:** Layers of convolution are used to retrieve images features, which are edges, intersection points, giving rich information. The number of layers matters here.

We can change the architecture by using different activation functions with different numbers of features.

The network will include the following components:

**Activation functions:** Our model includes the Relu and Softmax activation functions that are applied to all the output layers.

**Pooling Layers:** A convolution layer is added after it, performing continuous dimensionality reduction i.e., reducing

the number of parameters and computations, thereby shortening the training time and reducing overfitting.

**Dense layers:** CNN's bottom layer is the convolution layer, the layer which takes all the feature data produced by the convolution layers and analyzes it.

**Dropout Layers:** For preventing overfitting of irregular turns of neurons in deep learning networks, we will have many weight parameters and bias parameters.

In the dropout layer, we select specific features from the input layer and a specific set of neurons from the hidden layer, according to the p value. Some neurons and features are deactivated and others are activated.

Dropout ratio:  $0 \leq p \leq 1$

**Batch Normalization:** The method is used for training neural network models. As the batch normalization increases, the epochs required to train the deep neural network model decrease. Each layer of the neural network can learn independently, enabling faster training.

To make the model, we use Adam as the optimizer, loss as categorical cross-entropy, and metrics as accuracy. After building and compiling the model, the data is split into training data and validation data. Our model takes the batch size as 32 with 15 epochs.

After completion of training, we evaluate the model and calculate the loss and accuracy.

To predict the disease for symptoms by using we use the machine learning algorithms. Firstly, import the NumPy and Pandas for linear algebra and data processing and reading the csv file. Then preprocessed the data and changed the categorical data to the numerical data. Secondly, split the data to train, test sets. Finally, the fit model into the algorithms. The algorithms we used are KNN algorithms and Decision Tree Classifier which are supervised machine learning algorithm used for solving both the regression and classification problems. A website is developed where we can predict the disease by using both symptoms and radiology images. The symptoms of the patient is given according

to the disease then it predicts the whether the disease is there or not based on the algorithms that are fitted with different models. In other way, radiology images are given as the input where the model is trained with the efficientnetb0 and fitted with different activation layers to predict the disease with the best accuracy for the images as dataset. Finally, the website is an interface that would be helpful for patients to predict the disease by the symptoms with a good accuracy and efficiently.

## Experiment Results

We use both Deep Learning and Machine Learning Algorithms to predict various lung diseases such as COVID-19, Tuberculosis, Pneumonia, and COPD. For Symptoms dataset i.e., CSV Data, For Covid-19 and COPD we got accuracies of 96.90%, 90.32% respectively using classification algorithm i.e., Decision Trees. And For Tuberculosis we use K-Means Clustering.

For Image Dataset, Which Contains Radiology Images of Lung Diseases like Covid, Pneumonia, Tuberculosis and COPD we obtained accuracy of 98.58% using EfficientNet B0, a deep learning algorithm. By Using Streamlit Package We build a website which helps the User/Patient in detecting chronic lung disease in an early stage, which in turn, enhances the chances of recovery and survival.

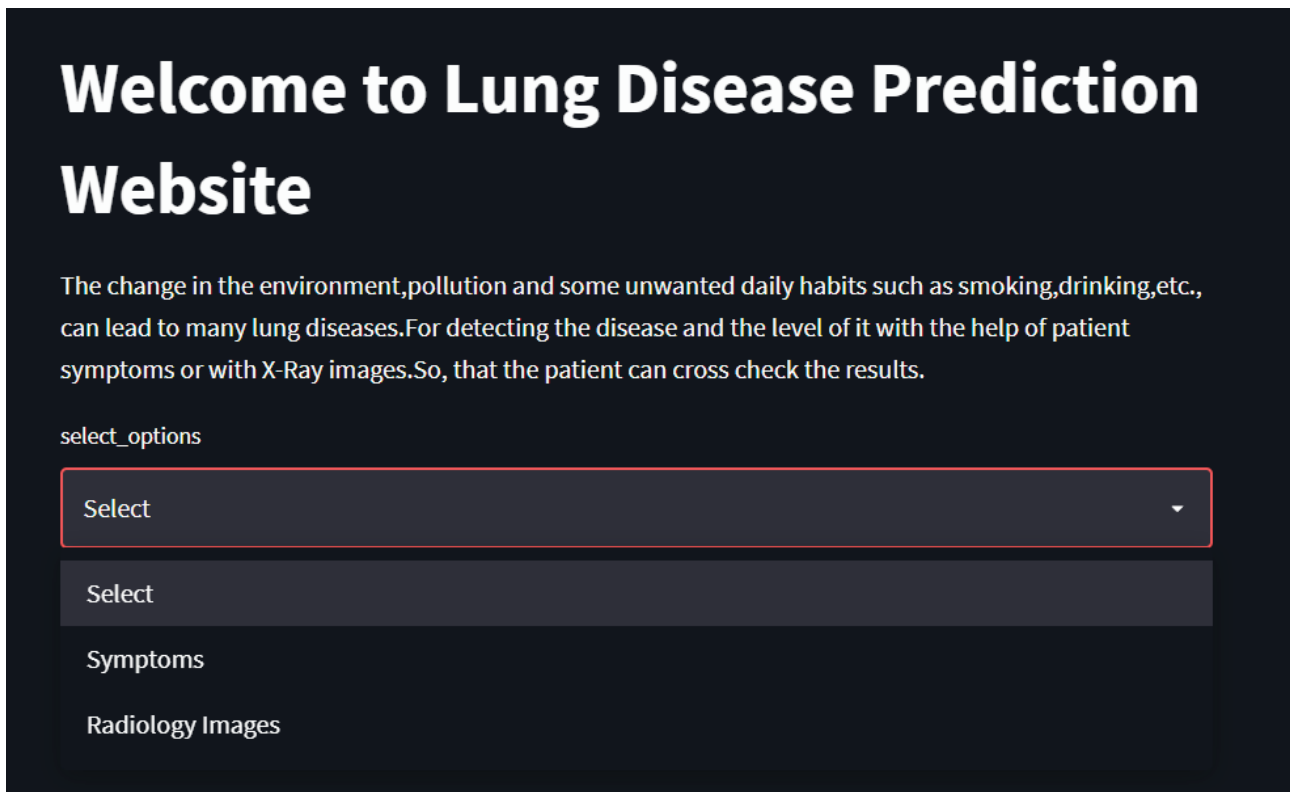


Figure 10: User Interface of website

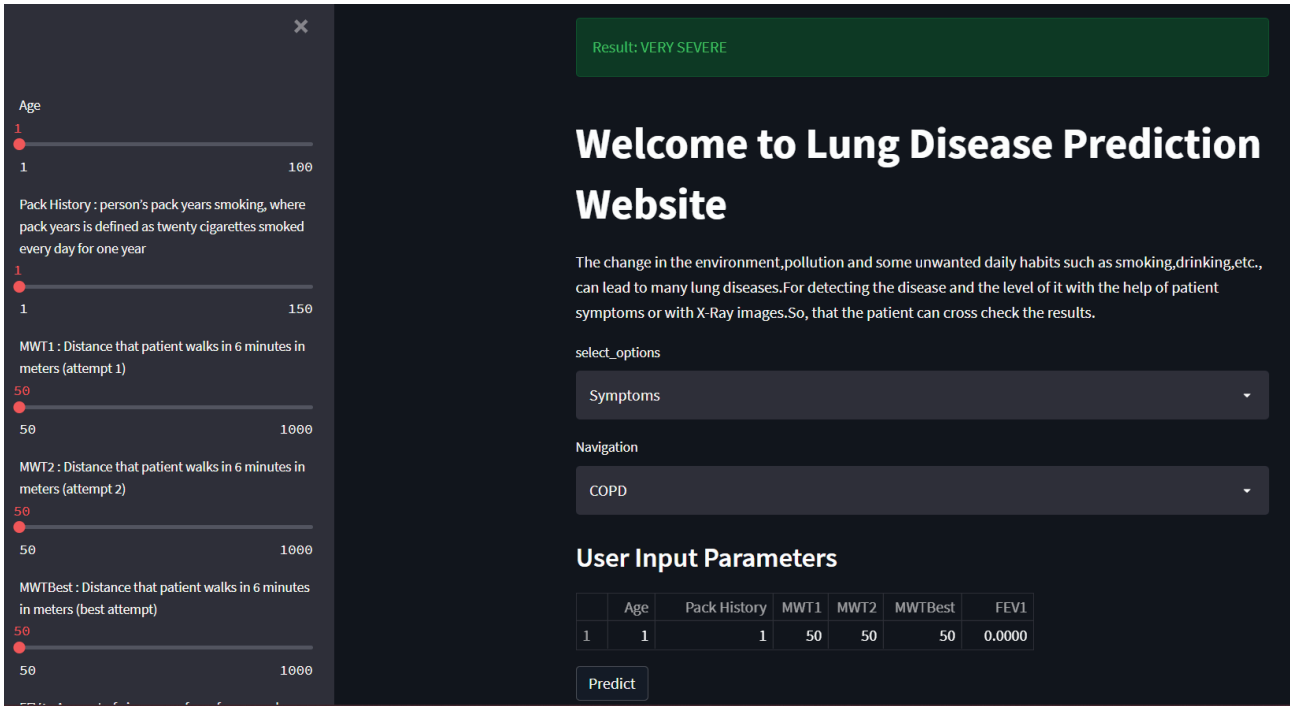


Figure 11: Prediction via Symptoms

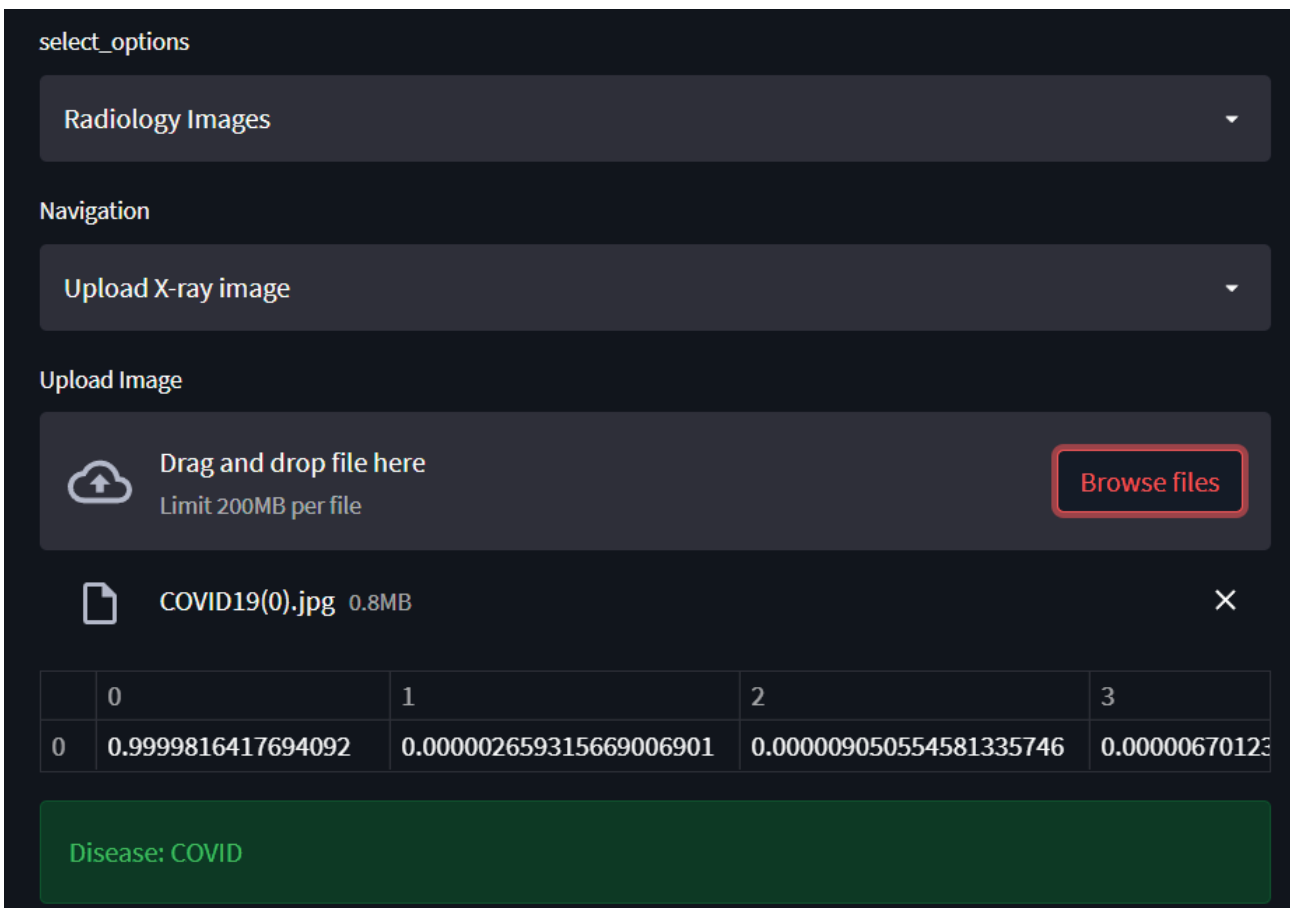


Figure 12: Prediction via Radiology Images

In this project we tried to develop a website for predicting and classifying lung diseases with a better accuracy which

helps in the prior treatment for the patient. We used both machine learning and deep learning algorithms for better classification and prediction.



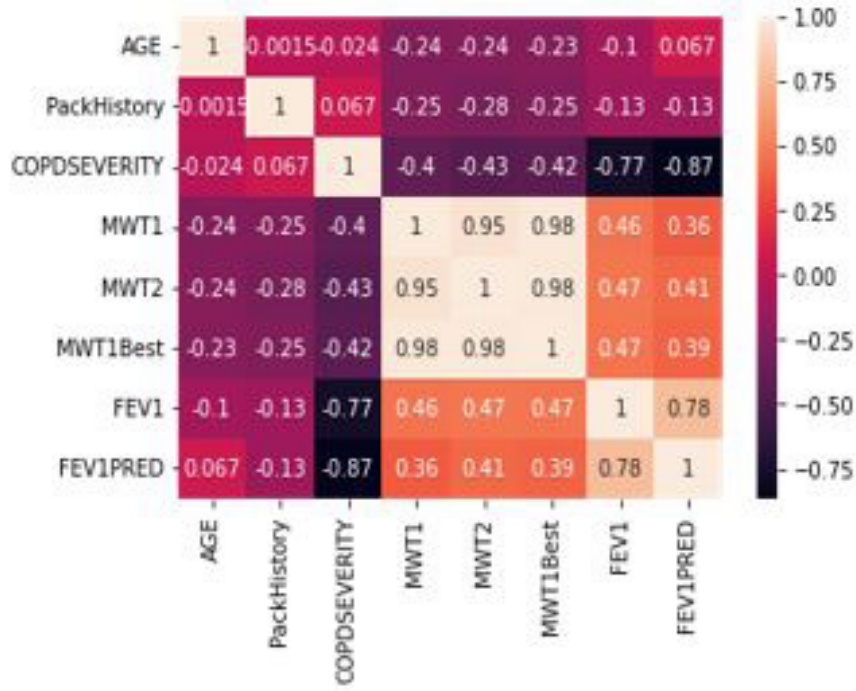


Figure 13: Heat Map for COPD Dataset model

Table 1: Accuracy Results

Disease	Model	Accuracy
Covid	Decision Trees	96.90%
Tuberculosis	K-Meansclustering	Error rate-0.33
COPD	Decision Trees	90.32%
Radiology images(Covid, Pneumonia Tuberculosis)	EfficientNetB0	Train-95.58%Test-86.18%

### Conclusion

The main aim of our research paper is to build a website to predict different types of lung diseases using symptoms and x-ray images of patients in real-time manner. The work done by us made the model to work in a better way such that it can predict the diseases using different patients x-ray images. In our model we use Machine Learning algorithms as Decision Trees, clustering for the prediction of symptoms and Keras which is a python library helps in deeplearning model, tensor flow for x-ray

images to predict the severity of the disease. We initially build a model and trained it such that it is capable of detecting and classifying the images and the symptoms in real-time. The model gave good accuracy but takes much time to train the data. Hereby, I conclude this paper by hoping that you got a fairknowledge, idea and understood the whole concept of designing the predicting lung disease using ML and DL in real-time, by using pre-trained models like Efficient Net B0 for better performance of models.

## References

1. Shimpy Goyal, Rajiv Singh (2021) Detection and classification of lung diseases for pneumonia and covid-19 using machine and deep learning techniques.
2. Asmaa Abbas, Mohammed M.Abdelsamea, Mohamed Medhat Gaber (2020) Classification of Covid-19 in chest x-ray images using DeTraC deep convolutional neural network.
3. Sema Candemir, Stefan Jaeger, Rahul k.Singh, Kannappan Palaniappan (2013) Lung Segmentation in chest Radiographs using Anatomical Atlases With Nonrigid Registration.
4. Alexanderos Karargyris, Les Folio, Fiona Callaghan, Zhiyun Xue. Automatic Tuberculosis Screening Using Chest Radiographs.
5. Stefanus Tao Hwa Kieu, Abdullah Bade, Mohd Hanafi Ahmad Hijazi, Hoshang Kolivand (2020) A survey of Deep Learning for lung disease detection on medical images- State of the art Taxonomy.
6. Siddhanth Tripathi, Sicnhana Shetty, Somil Jain, Vanshika Sharma (2021) Lung disease detection using deep learning.
7. Latheesh Mangeri, Gnana Prakasi, Neeraj Puppala (2021) Chest diseases prediciton from x-ray images using CNN model 12.
8. Anuradha D.Gunasinghe, Achala C.Aponsa, Harsha Thirimanna (2020) Early prediction of lung diseases.
9. Ishan Sen, Ikbal Hossain, Faisal Shakib, Asaduzzaman Imran (2020) Depth analysis of lung disease prediction using machine learningalgorithms.
10. Matthew Zak, Adam Krzyzak (2020) Classification of lung diseases using deep learning models”, International Conference on Computational Science.
11. Anuj Rohilla, Rahul Hooda, Ajay Mittal (2017) TB Detection in chest radiograph using deep learning architecture 6.
12. Peng Gang, Jiang Hui, Wei Zeng, S.Stirenko (2018) Deep learning with lung segmentation and bone shadow exclusive techniques for chest x- ray analysis of lung cancer”, ICCSEEA.

**Submit your manuscript to a JScholar journal and benefit from:**

- ¶ Convenient online submission
- ¶ Rigorous peer review
- ¶ Immediate publication on acceptance
- ¶ Open access: articles freely available online
- ¶ High visibility within the field
- ¶ Better discount for your subsequent articles

Submit your manuscript at  
<http://www.jscholaronline.org/submit-manuscript.php>