JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

IDENTIFYING THE SIMILARITIES BETWEEN COVID 19 AND PNEUMONIA FROM CHEST X-RAY USING CNN MODELS

AINALA NAVYA DEVI [#1], D.D.D SURI BABU [#2], B.NANDAN KUMAR [#3]

[#1] M.Tech Student, Department of Computer Science and Engineering, DNR College of Engineering and Technology, Sri RamaPuram, Balusumudi, Bhimavaram - 534202.
[#2] Head & Assoc. Prof, Department of Computer Science and Engineering, DNR College of Engineering and Technology, Sri RamaPuram, Balusumudi, Bhimavaram - 534202.
[#3] Assistant Professor, Department of Computer Science and Engineering, DNR College of Engineering and Technology, Sri RamaPuram, Balusumudi, Bhimavaram - 534202.

ABSTRACT

In present days for identifying or predict any diseases, one should have proper diagnosis knowledge for predicting the disease which is present in that human body. In general for prediction of very serious diseases we try to use X-Ray, CT or MRI scan techniques for taking decision on that appropriate disease. These reports are diagnosed and given reports by using a medical person who has complete knowledge on that appropriate domain to find out the abnormality which is present in human beings. As we all know that india tops the world for having more deaths due to lung related diseases and in recent days COVID is one which cause problem for lungs and respiratory organs. After the second highest cause of deaths in India due to heart disease, this lung disease is one which is increasing its rank more and more. In general now a day's covid is one of the major diseases which is disturbing the humans. This covid 19 is also predicted using chest X-rays and there is quite similarity between COVID 19 and pneumonia when compared each other. Hence there should be very keen knowledge in order to differentiate each of them from X-ray sheets. Hence we try to design an application using Deep Learning Classifiers in order to find out the difference very accurately by taking some sample Chest X-Ray reports. Our proposed deep learning framework is divided into two types: One is generating multi class classifier for classification of normal, pneumonia and covid19. Next we try to present several X-Ray images and then apply binary classifier for identifying the difference between each and every patient record.

Keywords: Deep Learning, Lung Diseases-Ray, Pneumonia, MRI Diagnosis.

1. INTRODUCTION

According to the World Disease Research Fund (WCRF)'s lung disease statistics report 2019, few countries and continents had the very best rate of lung disease that is 56.7(Age standardized rate per 100,000) followed by small countries. If we take a worldwide report in 2012, more than two million people were affected with lung disease which resulted in 1.5 million deaths which depicts that lung disease.In 2019 survey we saw more number of deaths are seen due to covid which affected the lungs and respiratory organs. Therefore, on a mean the outcomes are worse within the developing world. Normally lung diseases are identified or classified into two ways: One is based on cell condition and other is finding the disease based on lungs origin. These diseases are also classified into two stages, one is disease which is found in primary stage is termed as primary lung disease and the disease which extends beyond the lungs is understood to be secondary lung disease. Some of the diseases which come under respiratory disease are pneumonia, tuberculosis and currently Coronavirus Disease 2019 (COVID-19).

As per the IRS (International Respiratory Societies [2]), report more than three hundred million people are continuously suffering from asthma disease and more than 2 million people die due to this lung diseases. From the recent analysis, we know the COVID-19 pandemic infected millions of people and healthcare systems and also there was great loss for the humans. In general these lung diseases are major cause of death and create disaster for the world. Normally early detection of lung disease plays an major role in the chance of disease recovery and there are very few recovery rates if they are early detected and treated. In the primitive days the lung diseases are detected via blood test, skin test and some X-ray and CT scan. The report need to be examined by the radiology department and the concern person who has enough knowledge will try to tell the report from the test sample, which is becoming a very complex task if the radiologist is not available all the time. Recently deep learning has gained a lot of user's attention towards medical domain for disease prediction and finding abnormality. Hence we try to use this deep learning technique on lung disease prediction and try to classify the abnormality which is present in the lungs using chest X-ray examination[3]-[8].In general there is very less members identified lung disease symptoms accurately prior and most of them failed to identify that accurately and they lead to several deaths.

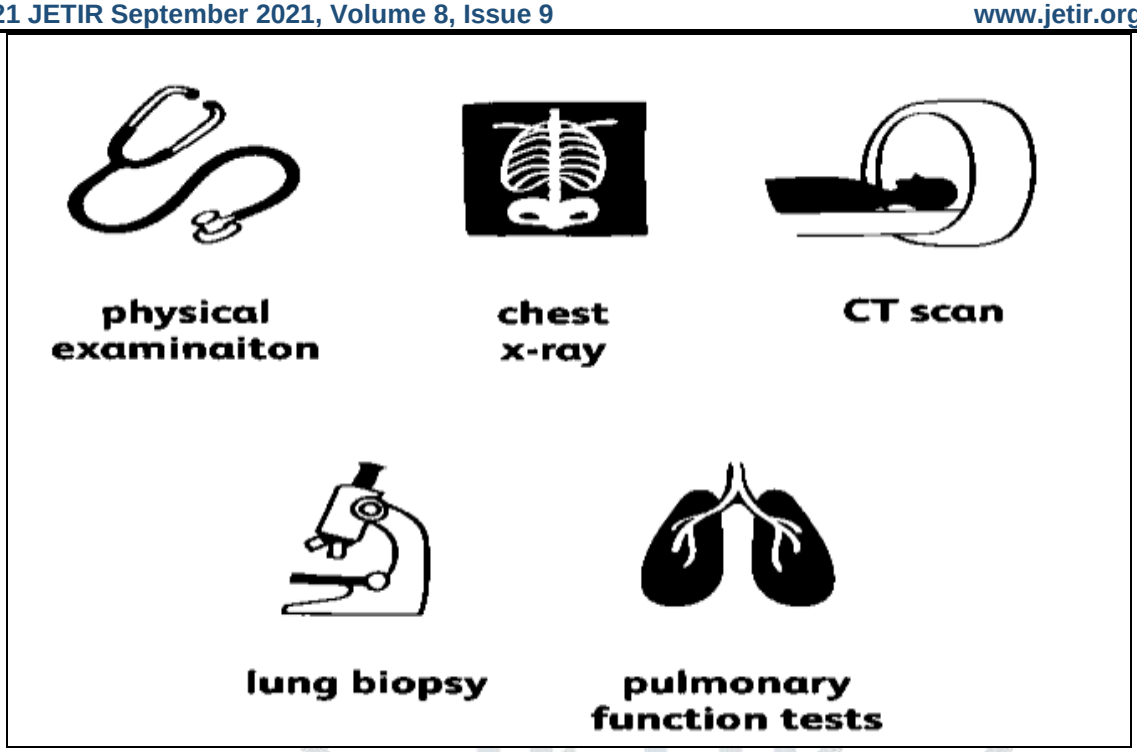


Figure 1. Represent the Medical Procedures for Lung Disease Prediction

From the above figure 1, there are several medical procedures for lung disease related predictions and following ways are physical examination, chest X-Ray ,CT Scan, Lung biopsy, Pulmonary tests.One among the best of all the methods is chest X-ray, so that from that report we can able to find out the present condition of lungs and its abnormality. If there is any abnormality present in the lungs we can able to identify that easily based on pre-trained knowledge given for the system.

2. LITERATURE SURVEY

Literature survey is that the most vital step in the software development process. Before developing the new application or model, it's necessary to work out the time factor, economy, and company strength. Once all these factors are confirmed and got approval then we can start building the application. The literature survey is one that mainly deals with all the previous work which is done by several users and what are the advantages and limitations of those previous models. This literature survey is mainly used for identifying the list of resources to construct this proposed application.

MOTIVATION

T. Panduranga Vital, M Murali Krishna, (2018) et.al [1] used some well-known classification algorithms such as Decision Tree, ADT, Naive Bayes, Bayes Net, K Star and Random Forest algorithms for detecting lung disease. In this current work the authors showed very high accuracy for prediction of lung related diseases and they achieved almost 96 % of accuracy compared with other algorithms and one main limitation or difficulty observed was that there was an opportunity of spurious relationships and the metric used was accuracy.

Timor Kadir, Fergus Gleeson (2018) et.al [2] used some well-known classification algorithms such as Convolutional Neural Networks [CNN] and Deep Learning for lung disease prediction. There are several risk models present in literature which are identified by the CADx tool. Here one main limitation they observed is it require a very large amount of data. The following are the several metrics used were nodule segmentation, texture feature extraction, risk score regression and risk score thresholding.

Majid Murtaza Noor, Vinay Narwal (2017) et.al [3] used some well-known classification algorithms such as Convolutional Neural Network for detecting lung disease. The authors used statistical probability and optimization techniques which is used to make the systems learn the knowledge from past examples and to detect hard to discern patterns from large noisy or complex data sets. The one main limitation observed in this paper is for small objects thus is not perfect and desired results are not obtained.

Puspanjali Mohapatra, Baldev Panda, Samikshya Swain (2019) et.al [4] used some well-known classification algorithms such as deep learning technology for detecting lung disease. The authors constructed a model which is in a position to achieve an accuracy of about 81 % for 120000 image samples. The disadvantage observed in this paper is it can only use high resolution images. The metrics utilized in this paper were accuracy, FI score, precision and recall.

Konstantina Kourou, Themis P. Exarchos, Konstantina P. Exarchos (2017) et.al [5] use Machine Learning ML methods, Artificial Neural networks ANN Bayesian networks Bn for detecting lung disease. the advantages observed in this paper are regression models with accuracy 84.6 %, 81.4% and 72.6 % respectively. Metric used is accuracy.

3. EXISTING SYSTEM AND ITS LIMITATIONS

In the existing system, there was no concept like lung disease prediction using CNN models. All the prediction is done using manual approach or by using primitive Machine Learning models. In the ML we can able to classify whether lung disease is present or not, but those models cannot classify the records with accuracy and parameters.

DISADVANTAGES OF THE EXISTING SYSTEM

In the existing or current clouds the following are the main limitations that are available

1. All the existing schemes are limited to the few classes classification only.

2. All the existing systems are failed to classify the chest xray images and then try to find out the disease symptoms.

3. All the existing ML approaches try to classify the patients information from the raw dataset

4. There is no accurate model to classify the real time chest x-ray for detecting and prediction of accuracy of that image.

4. PROPOSED SYSTEM AND ITS ADVANTAGES

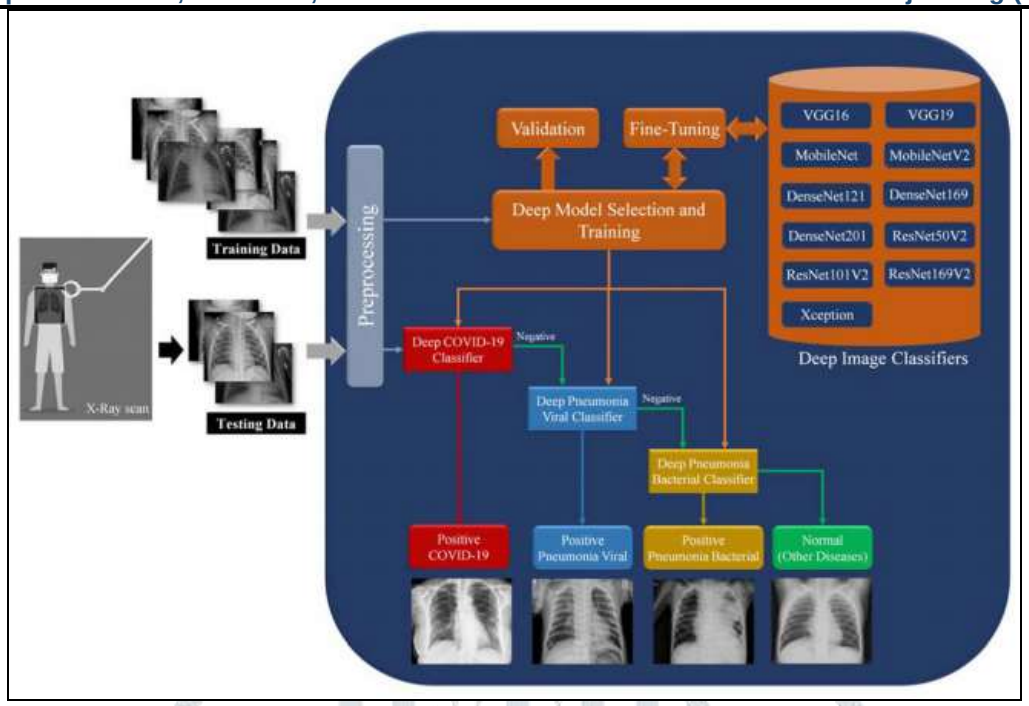
In this proposed work we try to design an application which can be used for prediction of lung related diseases from real world chest x-ray images. For training the system we try to collect the sample chest X-ray images which contain disease symptoms from KAGGLE website and then train the system. Once the system is trained now we can check the model performance by giving dynamic images and check the performance of each and every individual Model.Using Deep Learning to predict lung diseases from Chest X-rays can be a lifesaving factor for an individual suffering from the disease. This is possible as the results can be predicted with a high percentage of accuracy instantly. This paper presents an effective way for expert diagnosis of lung diseases using Deep Learning. It focuses on creating a system for assistance of Radiologists in detection of lung diseases. This will especially benefit rural areas where radiologists aren't easily available. We use two models like RESENT and Vgg19 for predicting the lung disease from chest x ray images and then tell which model gives high accuracy and performance.

ADVANTAGES OF THE PROPOSED SYSTEM

1. The proposed scheme is very accurate in classification of check x –ray images
2. The proposed system gives accurate recommendation for the doctors.
3. The proposed system is capable of classification of chest x-ray and find out the accuracy of that image.

5. PROPOSED CNN MODEL FOR CHEST DISEASE DETECTION

In this section we try to discuss about proposed CNN model which is used to detect chest disease detection using pre-trained CNN models such as RESNET and VGG19.



The Application is mainly divided into 4 modules. They are as follows:

A. Convolution Layer
B. Rectified Linear Unit (RELU) Layer
C. Pooling Layer
D. Fully Connected layer

Implementation is a stage where the theoretical design is converted into a programmatic manner. The application is mainly divided into following 4 modules. They are as follows:

1. Data Set Pre-Processing
2. Import Libraries
3. Labels Assignment
4. Apply Pre-trained CNN Models and Find Accuracy

1) DATA SET PRE-PROCESSING MODULE

Here we try to load the Lung related images which contain both disease symptoms and some are normal images. Here we can use either CT or MRI images as input dataset which contain several lung images in different slices. The dataset can be collected from several sources which are available in google.

2) IMPORT LIBRARIES MODULE

Here in this module we try to import several libraries which are present in python to load the lung images into the application and try to process those images for further identification of any diseases present in

lung or not. Here we try to load the numpy module as main library because the image is converted into numerical values by using that numpy module and once it is converted into numerical manner, then the end user can able to identify the presence of abnormality easily.

3) LABELS ASSIGNMENT MODULE

Here the input data is categorized into two types : One is test and another one is train dataset. Hence we try to categorize the data into THREE types and assign labels as test and train. Here training is nothing but sample images which are given to the system for learning the inner functionality properly to find out the abnormality present in lung images.Here one set is COVID19 , PNEUMONIA and NORMAL.

4) APPLY CNN MODULE

Here we try to apply pre-trained CNN model and then find out which one gives best accuracy for our current model. Here we try to use RESNET as pre trained CNN model for achieving accuracy in finding the lung disease from a sample CT or MRI slices which are given for the system as input. By applying this we can achieve high level of accuracy compared with primitive ML algorithms.Here we can able to calculate accuracy as well as prediction of COVID 19 present in the patient or not,

6. EXPERIMENTAL RESULTS

Implementation is a stage where the theoretical design is converted into a programmatic manner. In this proposed application we try to use PYTHON as a programming language in which Google Collaboratory or Jupiter Notebook as a working platform to process the current application.

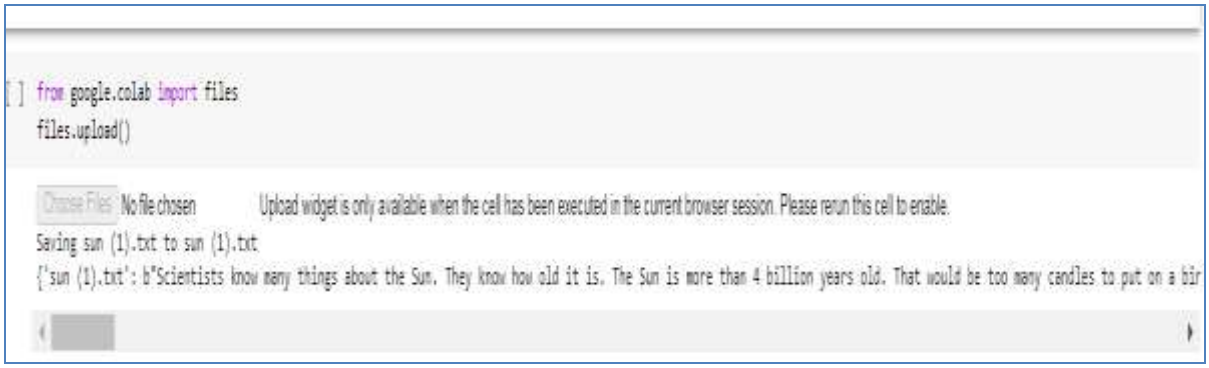
STEP 1: IMPORT THE LIB AND READING DATASET

Installing necessary libraries

```
!pip install gensim
# !pip install git+https://github.com/boudinfl/pke.git
# !python -m spacy download en
# !pip install bert-extractive-summarizer --upgrade --force-reinstall
! pip install bert-extractive-summarizer
# !pip install spacy==2.1.3 --upgrade --force-reinstall
# !pip install -U spacy
!pip install -U nltk
!pip install -U pywsd
!pip install flashtext
import nltk
nltk.download('stopwords')
nltk.download('popular')
!pip install spacy==2.1.3
!pip install transformers==2.2.2
!pip install neuralcoref

!python -m spacy download en_core_web_md
```

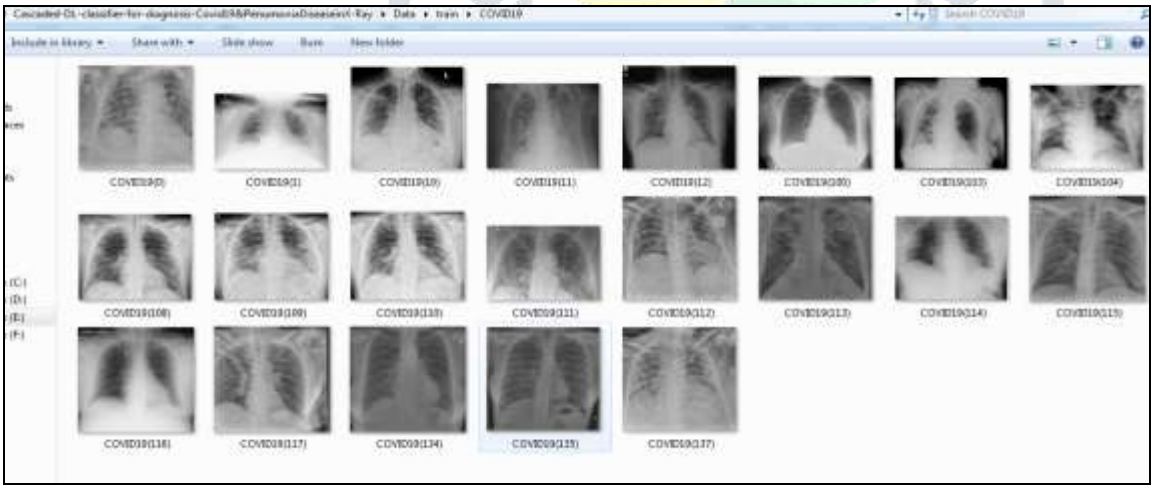
STEP 2: IMPORT THE TEXT DATASET AS INPUT



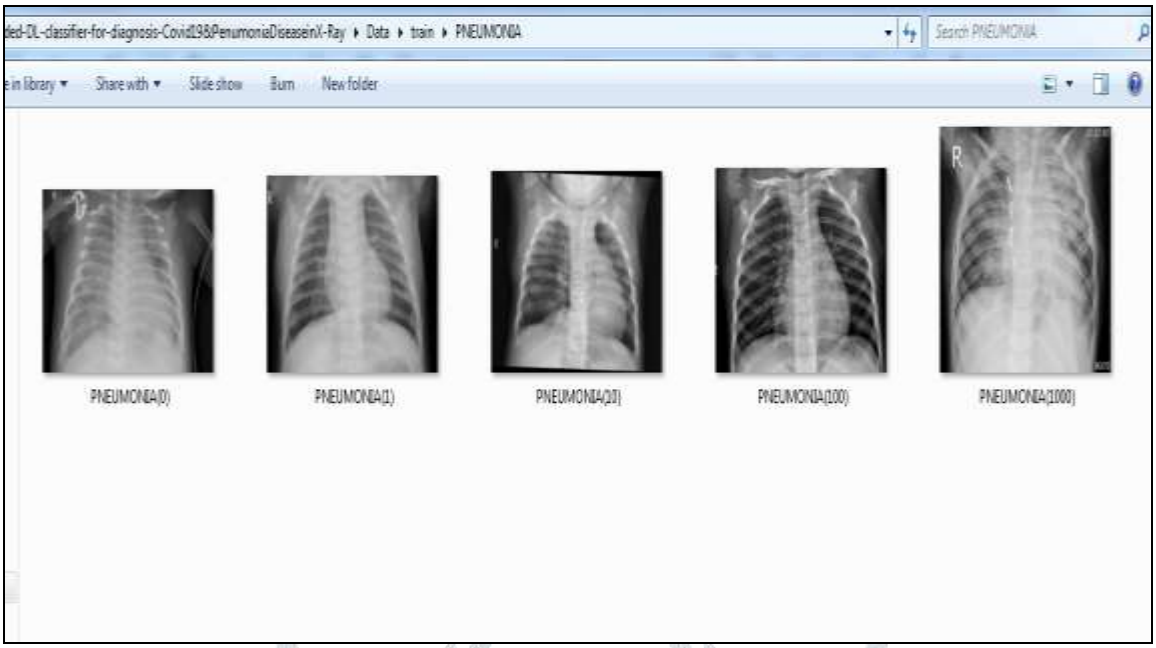
STEP 3: INSTALL NECESSARY LIBRARIES

```
    "cell_type": "code",
    "execution_count": 2,
    "metadata": {},
    "outputs": [],
    "source": [
     "import numpy as np\n",
     "import pandas as pd \n",
     "from keras.preprocessing.image import ImageDataGenerator,
load_img\n",
     "from keras.utils import to_categorical\n",
     "from sklearn.model_selection import train_test_split\n",
     "import matplotlib.pyplot as plt\n",
     "from sklearn.metrics import classification_report\n",
     "from sklearn.metrics import confusion_matrix\n",
     "from random import shuffle\n",
     "import random\n",
     "import os"
    ]
```

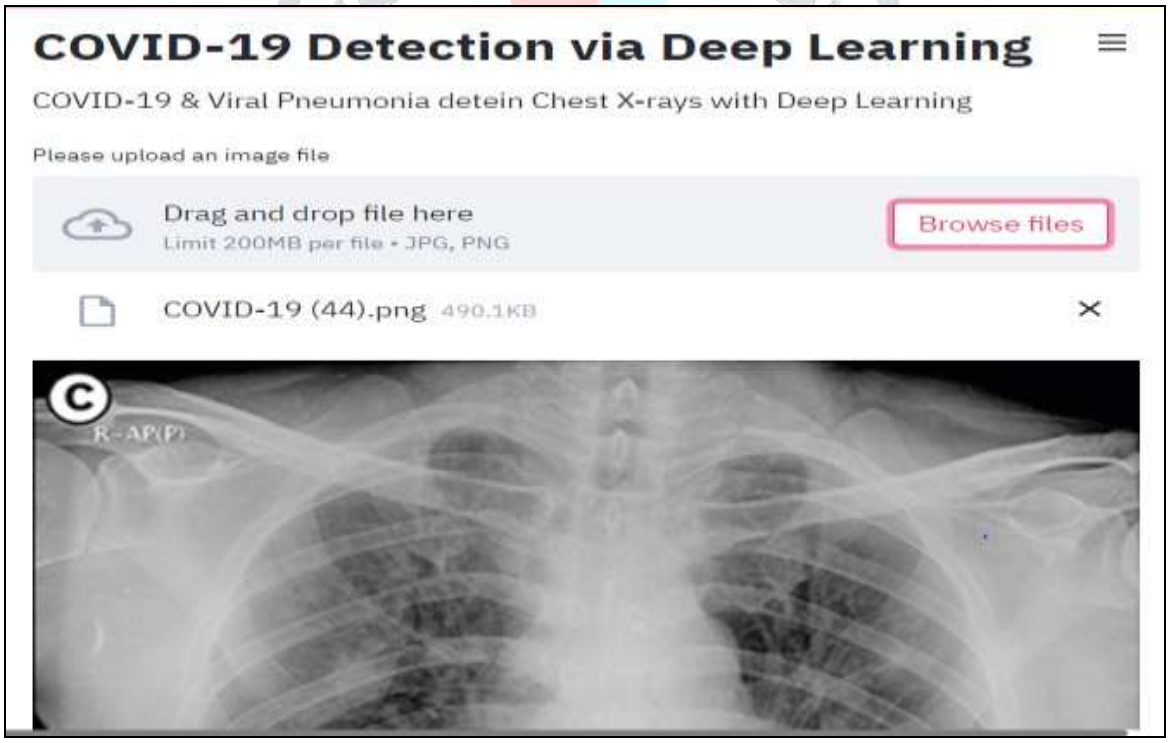
STEP 4: TRAIN IMAGES



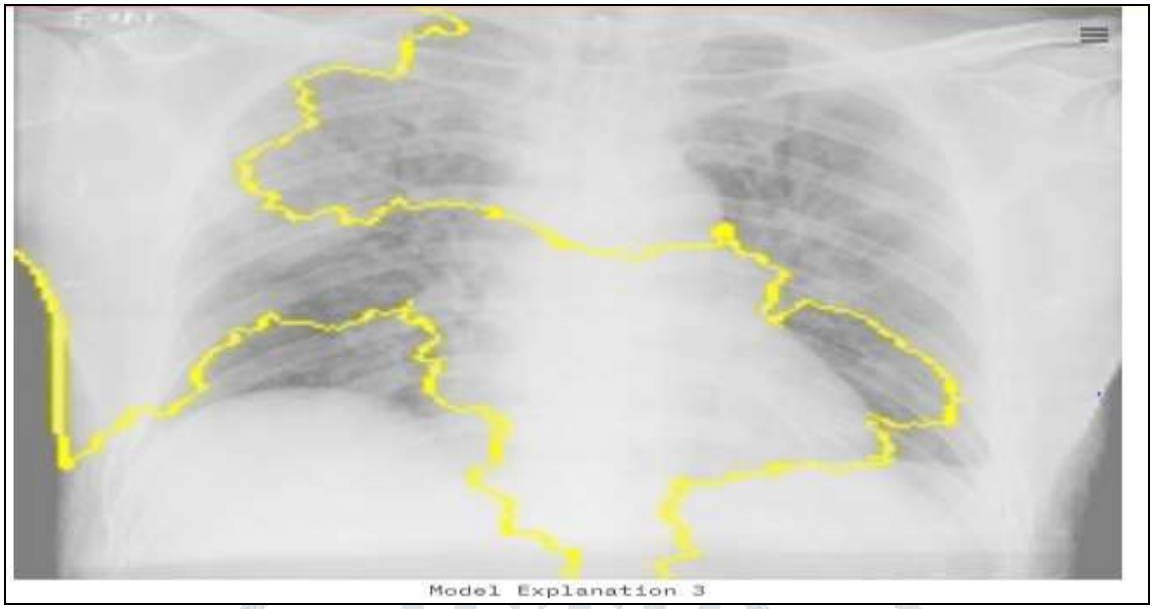
STEP 5: TRAIN IMAGES



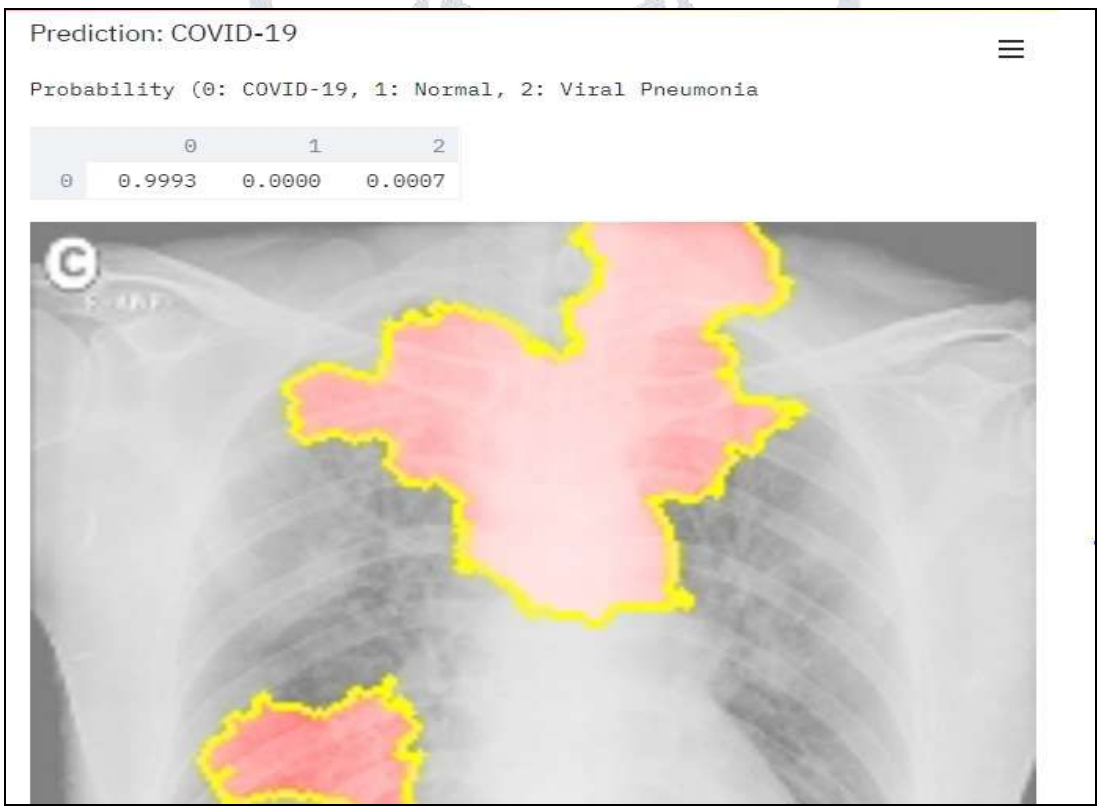
STEP 6: TEST A SAMPLE IMAGE



STEP 7: CNN MODEL IS TRAINED TO FIND ABNORMALITY



STEP 8: RESULT IS DISPLAYED WITH PREDICTION OF DISEASE NAME AND ACCURACY



7. CONCLUSION

In this proposed paper we for the first time designed an application to detect lung disease using Deep Learning Model. In this project we have used RESNET 50 model to detect the performance of lung detection. We have achieved more than 85 percent of accuracy by using deep learning model compared with primitive

machine learning algorithms for identifying the abnormalities present in the lung images. Here we can able to detect similarities between normal and covid infections from common chest X-ray images.

FUTURE WORK

In this study we concentrated only on one deep learning model for lung disease detection and in future we want to test the same mechanism on multiple deep learning models which can increase the performance of our models.

8. REFERENCES

1) 1. Vital, T. P., Krishna, M. M., Narayana, G. V. L., Suneel, P., & Ramarao, P. (2019). Empirical analysis on cancer dataset with machine learning algorithms. In Soft Computing in Data Analytics (pp. 789-801). Springer, Singapore.
2) Kadir, T., & Gleeson, F. (2018). Lung cancer prediction using machine learning and advanced imaging techniques. Translational lung cancer research, 7(3), 304.
3) Noor, M. M., & Narwal, V. (2017). Machine learning approaches in cancer detection and diagnosis: mini review. IJ Mutil Re App St, 1(1), 1-8.
4) Mohapatra, P., Panda, B., & Swain, S. (2019). Enhancing histopathological breast cancer image classification using deep learning. Int J Innov Technol Explor Eng, 8(7), 2024-2032.
5) Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal, 13, 8-17.
6) Rosenberger, A., Hung, R. J., Christiani, D. C., Caporaso, N. E., Liu, G., Bojesen, S. E., ... & Gomolka, M. (2018). Genetic modifiers of radoninduced lung cancer risk: a genome-wide interaction study in former uranium miners. International archives of occupational and environmental health, 91(8), 937-950.
7) Morrison, E. J., Novotny, P. J., Sloan, J. A., Yang, P., Patten, C. A., Ruddy, K. J., & Clark, M. M. (2017). Emotional problems, quality of life, and symptom burden in patients with lung cancer. Clinical Lung Cancer, 18(5), 497-503.
8) Mahale, A., Rawool, C., Tolani, D., Bathija, D., & Jewani, K. (2017). SVM classifier-based CAD system for Lung Cancer Detection. ijecs.
9) Rolfo, C., Mack, P. C., Scagliotti, G. V., Baas, P., Barlesi, F., Bivona, T. G., ... & Gandara, D. R. (2018). Liquid biopsy for advanced non-small cell lung cancer (NSCLC): a statement paper from the IASLC. Journal of Thoracic Oncology, 13(9), 1248-1268.
10) Le Marchand, L., Murphy, S. P., Hankin, J. H., Wilkens, L. R., & Kolonel, L. N. (2000). Intake of flavonoids and lung cancer. Journal of the National Cancer Institute, 92(2), 154-160.
11) Sasikala, S., Bharathi, M., & Sowmiya, B. R. (2018). Lung Cancer Detection and Classification Using Deep CNN. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(2S).

12) Akhtar, J. Non-Small Cell Lung Cancer Classification from Histopathological Images using Feature Fusion and Deep CNN.

13) Sahu, B., Panigrahi, A., & Rout, S. K. (2020). 10 DCNN-SVM: A NEW APPROACH FOR LUNG CANCER DETECTION. Recent Advances in Computer Based Systems, Processes and Applications: Proceedings of Recent Advances in Computer based Systems, Processes and Applications (NCRACSPA-2019), October21-22, 2019, 97.

14) Senthil, S., & Ayshwarya, B. (2018). Lung cancer prediction using feed forward back propagation neural networks with optimal features. International Journal of Applied Engineering Research, 13(1), 318-325.

15) S Venkata Lakshmi, Valli Kumari Vatsavayi. Query optimization using clustering and Genetic Algorithm for Distributed Databases. International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2016.

16) S Venkata Lakshmi, Valli Kumari Vatsavayi. Teacher-Learner & Multi-Objective Genetic Algorithm Based Query Optimization Approach for Heterogeneous Distributed Database Systems, Journal of Theoretical and Applied Information Technology, April 2017

17) K. Vanitha, K. Yasudha, S Venkata Lakshmi, The Development Process of the Semantic Web and Web Ontology, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol.2, No.7, 2011.

18) Vytarani Mathane, P V Lakshmi. Adaptive Security Framework for the Blockchain on IoT, International Journal of Innovative Technology and Exploring Engineering (IJITEE), July 2019.