



# International Journal of Advanced Research in Education and Technology (IJARETY)

Volume 11, Issue 6, November-December 2024

Impact Factor: 7.394



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# A Contemporary Technique for Lung Disease Prediction using Deep Learning

A.Lakshmipathi Rao<sup>1</sup>, Kurandasu Sai Sree<sup>2</sup>, Mareddy Chaitanya<sup>3</sup>, Chetan Nallagopula<sup>4</sup>

Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India<sup>1</sup>

Student, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India<sup>2,3,4</sup>

**ABSTRACT:** The impact of disease on health is escalating quickly as a result of environmental changes, climate change, adjustments in lifestyle, and other reasons. Health problems are now more likely as a result. In 2016, 3.4 million individuals died from chronic obstructive pulmonary disease (COPD), which is typically brought on by smoking and pollution, while asthma claimed the lives of 400,000 people. Particularly in developing and low-income nations, where millions of people struggle with poverty and air pollution, there is a very high risk of lung disorders. According to WHO estimates, illnesses linked to home air pollution, such as asthma and pneumonia, cause nearly 4 million preventable deaths each year. Worldwide, lung illness is a prevalent occurrence. These include pneumonia, asthma, TB, fibrosis, chronic obstructive pulmonary disease, and others. The early detection of lung illness is important. The method of discovering and classifying lung disorders into different groups using medical imaging has been improved because to the development of deep learning. For this, a variety of machine learning and image processing models have been created. Convolutional neural networks (CNNs), one type of existing deep learning approach, are used to forecast lung illness. CNN's basic version is inadequate. As a result, we propose an novel deep learning framework for predicting lung diseases based on the VGG16 Architecture. The goal of this research is to develop a VGG16 architecture-based lung disease detection model. Early detection and diagnosis of lung disease are essential in the medical field because doing so will make it easier to manage patients' future clinical care. The X-ray picture dataset obtained from the Kaggle source is subjected to the VGG16 Architecture. The dataset's sample and full versions are taken into consideration. The VGG16 Architecture beats existing techniques for both whole and sample datasets in terms of measures including precision, recall, F1 score, and validation accuracy. Therefore, the proposed VGG16 Architecture will make it easier for both professionals and clinicians to detect lung problems. This improvement has greatly helped the medical community's ability to treat patients quickly.

## I. INTRODUCTION

A lung disease prediction using an x-ray images project is all about detecting whether the given image has any lung disease or not. Here, a deep learning model and VGG16 model are implemented to predict the presence or absence of lung disease from the given image by utilizing various libraries in Python like NumPy, TensorFlow, etc. With a test accuracy percentage of 91%, the project was finished successfully. Machine learning is a branch of artificial intelligence in which computers use a range of statistical, probabilistic, and optimization approaches to "learn" from prior examples and detect difficult-to-find patterns in vast, noisy, or complex datasets. Machine learning is a key technique for developing powerful, automated, and objective algorithms for analyzing high-dimensional, multimodal biological data. In medical systems, machine learning plays a critical role. Disease detection can be added, allowing us to detect illnesses earlier and more correctly, perhaps saving many lives and reducing system load. One of the primary causes of mortality is lung disease. The diagnosis and prediction of lung illness have become an in research since it can help with patient care later on. Doctors can use Machine Learning-based decision support systems to help them with their diagnosis decisions. Patients' respiratory issues, as well as Corona, Tuberculosis, Pneumonia, and Lung Cancer, were all examined in the study. To review data and develop models for diagnosing patients, machine learning and deep learning are used. Combining patient data with data from chest X-rays, utilizing VGG16 with the well-known pre-trained model, and VGG16 for data in this fashion were among the approaches utilized in this study to diagnose lung disorders. Deep Learning were utilized to evaluate Project involves how lung disease prediction using x-ray images will predict through the binary classification model implemented, and various python libraries like Tensor Flow, Keras, NumPy, etc. are used. This research project will observe the prediction of lung diseases by using x-ray images and further the output will be predicted with the detailed example and a detailed source code. The implementation will be shown step by step.

## II. LITERATURE SURVEY

**K.R.Swetha et. al(2021)**Using big data for prediction analysis along with machine learning or deep learning techniques or algorithms is one the most active areas of research in order to improve the health and the medical science. There is a significant increase in the size of the medical data as well as the complexity in the diagnosis of various diseases. With this being said, the diagnosis or the prediction of many terminal or fatal diseases has seen huge success through deep learning. Among those fatal diseases, pneumonia is one of the greatest threats to the life of a man affecting the lungs leading to lung failure.To diagnose a man with pneumonia, the x- ray of chest is needed, and an expert in the prediction is also required. Hence, it is more convenient to build an automated predictor to predict the pneumonia using the big data deep learning methods. Among all the other techniques, CNN (Convolutional Neural Networks) stand tall and high in this prediction along with other classifiers. Also, pre-training the CNN models for very large datasets that is for big data of healthcare units stands a high chance for accurate classification. A CNN model which is pre-trained along with an efficient feature extraction technique and various classifiers to classify the positive from negative is considered to give highly accurate results. This research work represents the Prediction of Pneumonia using Big Data, Deep Learning and Machine Learning Techniques

**Dey et. al(2021)**The rise of the corona virus disease 2019 (COVID-19) pandemic has made it necessary to improve existing medical screening and clinical management of this disease. While COVID-19 patients are known to exhibit a variety of symptoms, the major symptoms include fever, cough, and fatigue. Since these symptoms also appear in pneumonia patients, this creates complications in COVID-19 detection especially during the flu season. Early studies identified abnormalities in chest X-ray images of COVID-19 infected patients that could be beneficial for disease diagnosis. Therefore, chest X-ray image-based disease classification has emerged as an alternative to aid medical diagnosis. However, manual detection of COVID-19 from a set of chest X-ray images comprising both COVID-19 and pneumonia cases is cumbersome and prone to human error. Thus, artificial intelligence techniques powered by deep learning algorithms, which learn from radiography images and predict presence of COVID-19 have potential to enhance current diagnosis process. Towards this purpose, here we implemented a set of deep learning pre-trained models such as ResNet, VGG, Inception and Efficient Net in conjunction with developing a computer vision AI system based on our own convolutional neural network (CNN) model: Deep Learning in Healthcare (DLH)-COVID. All these CNN model scatter to image classification exercise.We used publicly available resources of 6,432 images and further strengthened our model by tuning hyper parameters to provide better generalization during the model validation phase. Our final DLH-COVID model yielded the highest accuracy of 96% in detection of COVID-19 from chest X-ray images when compared to images of both pneumonia-affected and healthy individuals. Given the practicality of acquiring chest X-ray images by patients, we also developed a web application (link: <https://toad.li/xray>) based on our model to directly enable users to upload chest X-ray images and detect the presence of COVID-19 within a few seconds. Taken together, here we introduce a state-of-the-art artificial intelligence-based system for efficient COVID-19 detection and a user-friendly application that has the capacity to become a rapid COVID-19 diagnosis method in the near future.

**Patil(2021)**Background and objective:SAARS-COV-2 is are spiratory illness caused by the novel Corona virus (COVID-19) disease. The virus goes into the lungs through the respiratory tracks and damages the walls and linings of the air sacs in our lungs, as our body tries to fight it, our lungs become more inflamed and fill with fluid. This makes it harder to breathe. So, at early stages,deep learning applications can be used for screening and prediction at a rapid rate for diagnosing the lungs of patients.This paper uses Transfer learning methods. Four pretrained models were used in this study - VGG-16, VGG-19, Inceptionv3, Xception This paper addresses challenges while using pre-trained models in real-world.Also,high accuracies were achieved on these models.

**Yadav et. al (2021)**Pulmonary fibrosis is a severe chronic lung disease that causes irreversible scarring in the tissues of the lungs, which results in the loss of lung capacity.The Forced Vital Capacity (FVC) of the patient is an interesting measure to investigate this disease to have the prognosis of the disease. This paper proposes a deep learning-based FVC-Net architecture to predict the progression of the disease from the patient's computed tomography (CT) scan and the patient's metadata. The input to the model combines the image score generated based on the degree of honeycombing for a patient identified based on segmented lung images and the metadata. This input is then fed to a 3-layer net to obtain the final output. The performance of the proposed FVC-Net model is compared with various contemporary state-of-the-art deep learning-based models, which are available on a cohort from the pulmonary fibrosis progression dataset. The model showcased significant improvement in the performance over other models for modified Laplace Log-Likelihood (-6.64). Finally, the paper concludes with some prospects to be explored in the proposed study.

**Existing System**

- Chest imaging diagnostics is crucial in the medical area due to many serious lung diseases like cancers and nodules and particularly with the current pandemic of Covid-19. Machine learning approaches yield prominent results toward the task of diagnosis. Recently, deep learning methods are utilized and recommended by many studies in this domain.
- The research aims to critically examine the newest lung disease detection procedures using deep learning algorithms that use X-ray and CT scan datasets. However, different architectures of deep learning are used by many researchers but, Convolutional Neural Networks (CNN) are still state-of-art techniques in dealing with image datasets.

**Existing System Disadvantages**

- CNN do not encode the position and orientation of object.
- Lack of ability to be spatially invariant to the input data. Lots of training data is required.

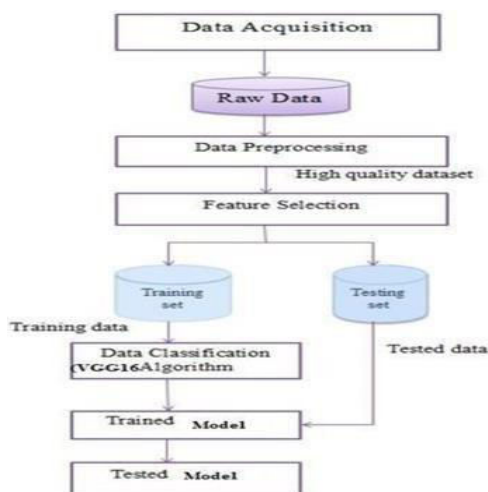
**Proposed System**

- There are many existing models which predict different diseases individually but we aim to predict different diseases in a single model with the highest accuracy. we have collected all the data from the Kaggle dataset, that data will be tested and then we will train all the classes.
- Then we will train the model step by step to complete the process, at the sequential process of model summary done layer by layer. So that we have created the VGG16 Architecture to train and test our multi Lung diseases like Covid-19, Pneumonia & Tuberculosis through their related x-ray images.
- Our model give a Training accuracy of 93.1% and testing accuracy of 91.1% in finding / predicting lung diseases.

**Proposed System Advantages**

- It is one of the popular algorithms for image classification.
- It is easy to use with transfer learning.
- The small-size convolution filters allows VGG to have a large number of weight layers, more layers leads to improved performance.

**System Architecture**



**Fig 1.1 System Architecture**

The data owner for this project must register all details before logging in. A document may be uploaded by the data owner. The data user may get a request from the data owner. A data user can utilise an uploaded document to search for a query. There is a download option for the file, which displays the encryption format. The cloud server receives requests from data users as well. A cloud server may have a login. The key will be accepted by it. The entire data set is visible to the cloud server as well. All user data is visible to the cloud server as well. The whole stored data set is visible to the cloud server. A key request from the user may be approved by the cloud server. After receiving the

request, the data owner might provide the user with a secret key. After then, the user can download a file. The user receives a notice that they have a block if they enter the incorrect keys.

### III. METHODOLOGY

#### Modules Name:

- Dataset
- Importing the necessary libraries
- Retrieving the images
- Splitting the dataset
- Building the model
- Apply the model and plot the graphs for accuracy and loss
- Accuracy on test set
- Saving the Trained Model

#### 1) Dataset:

In the first module, we developed the system to get the input dataset for the training and testing purpose. Dataset is given in the model folder. The dataset consists of 1075 Lungs X-ray images

#### 2) Importing the necessary libraries:

We will be using Python language for this. First we will import the necessary libraries such as keras for building the main model, sklearn for splitting the training and test data, PIL for converting the images into array of numbers and other libraries such as pandas, numpy, matplotlib and tensorflow.

#### 3) Retrieving the images:

We will retrieve the images and their labels. Then resize the images to (180,180) as all images should have same size for recognition. Then convert the images into numpy array. \

#### 4) Splitting the dataset:

Split the dataset into train and test. 80% train data and 20% test data.

#### 5) Building the model:

The concept of convolutional neural networks. They are very successful in image recognition. The key part to understand, which distinguishes CNN from traditional neural networks, is the convolution operation. Having an image at the input, CNN scans it many times to look for certain features. This scanning (convolution) can be set with 2 main parameters: stride and padding type. As we see on below picture, process of the first convolution gives us a set of new frames, shown here in the second column (layer). Each frame contains an information about one feature and its presence in scanned image. Resulting frame will have larger values in places where a feature is strongly visible and lower values where there are no or little such features. Afterwards, the process is repeated for each of obtained frames for a chosen number of times. In this project I chose a classic VGG16 model which contains only two convolution layers.

The latter layer we are convolving, the more high-level features are being searched. It works similarly to human perception. To give an example, below is a very descriptive picture with features which are searched on different CNN layers. As you can see, the application of this model is face recognition. You may ask how the model knows which features to seek. If you construct the CNN from the beginning, searched features are random. Then, during training process, weights between neurons are being adjusted and slowly CNN starts to find such features which enable to meet predefined goal, i.e. to recognize successfully images from the training set.

Between described layers there are also pooling (sub-sampling) operations which reduce dimensions of resulted frames. Furthermore, after each convolution we apply a non-linear function (called **ReLU**) to the resulted frame to introduce non-linearity to the model.

Eventually, there are also fully connected layers at the end of the network. The last set of frames obtained from convolution operations is flattened to get a one-dimensional vector of neurons. From this point we put a standard, fully-

connected neural network. At the very end, for classification problems, there is a softmax layer. It transforms results of the model to probabilities of a correct guess of each class.

**6) Apply the model and plot the graphs for accuracy and loss:**

We will compile the model and apply it using fit function. The batch size will be 10. Then we will plot the graphs for accuracy and loss. We got average validation accuracy of 93.00% and average training accuracy of 91.00%.

**7) Apply the model and plot the graphs for accuracy and loss:**

We got an accuracy of 91.00% on test set.

**8) Saving the Trained Model:**

Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle. Make sure you have pickle installed in your environment. Next, let's import the module and dump the model into .pkl file.

**Implementation**

**Intelligent routing scheme for traffic engineering(IRTE)** In this study, we propose an intelligent TE(IRTE) routing system in an SR setting. In an SR environment, the IRTE not only does bandwidth load balancing but also takes packet lengths into account for minimal control overhead. We suggest using the diversion routing technique to meet our goal of load balancing bandwidth. We install nodes with diversion routing capabilities and build the IRTE architecture using the SR controller. The matrix of information for every time interval. The controller computes the diversion routing paths for source hosts based on load balancing algorithms and network conditions after receiving their data transmission demand. For the IRTE, we create two brand-new overload optimisation and load balancing algorithms: improved ant colony optimization (IACO) and load balancing algorithm with diversion routings (LBA-DR). Using randomized rounding and linear programming, the LBA-DR creates the diversion routing path for every source node. After analysing this matrix, IACO offers a diversion routing path, which deviates from the conventional shortest routing method in most cases. On the one hand, this shows the model accuracy gradual improvement for the pneumonia disease with the functional model. The same dataset was used for this model, i.e., from Paul Mooney with 5,856 images of which 1,000 are normal X-rays and other 1,000 are infected chest X-rays. As the model has already been trained before, the starting accuracy is very good. There is a minor improvement after training for 15 epochs as is evident, Deep learning techniques have recently achieved an impressive result in the field of computer vision along with Medical Engineering. In this paper, we proposed and evaluated a deep convolutional neural network, designed for classifying the Chest Diseases. The proposed model consists of Convolutional layers, ReLU Activations, Pooling layer, and fully connected layer. Last full connected layer which consists of fifteen output units. Each output A publicly available dataset called Chest X-Ray 14 which consists of fifteen classes named Atelectasis, Cardiomegaly, Effusion, we developed a deep learning model, CXR Lung-Risk, to predict the risk of lung disease mortality from a chest x-ray.

**IV. ALGORITHM USED**

**Existing Algorithm**

**Convolutional Neural Networks(CNN) :**

A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data. There are other types of neural networks in deep learning. Neural networks accept an input image/feature vector (one input node for each entry) and transform it through a series of hidden layers, commonly using nonlinear activation functions.

Each hidden layer is also made up of a set of neurons, where each neuron is fully connected to all neurons in the previous layer. The last layer of a neural network (i.e., the "output layer") is also fully connected and represents the final output classifications of the network.

**Proposed Algorithm**

**VGG16 Architecture:**

VGG16 is a type of CNN(Convolutional Neural Network) that is considered to be one of the best computer vision models to date. The creators of this model evaluated the networks and increased the depth using an architecture with very small (3 × 3) convolution filters, which showed a significant improvement on the prior-art configurations.

We pushed the depth to 16–19 weight layers making it approx — 138 trainable parameters. VGG16 is object detection and classification algorithm which is able to classify 1000 images of 1000 different categories with high accuracy.

Experimental Results

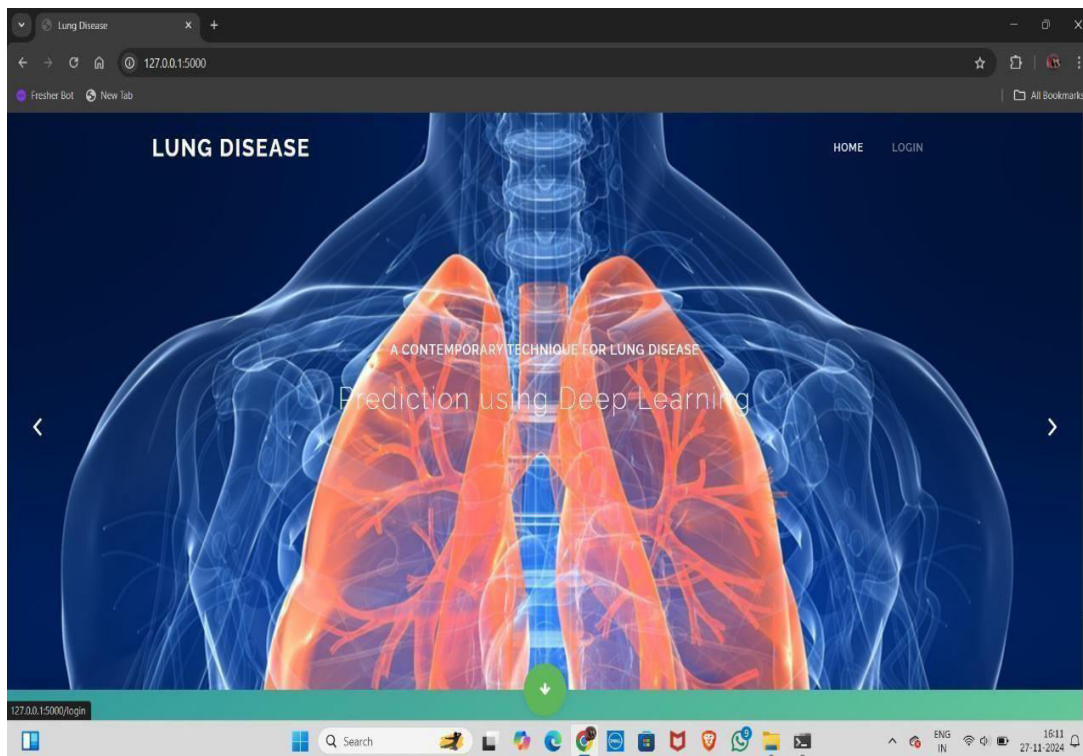


FIGURE:1

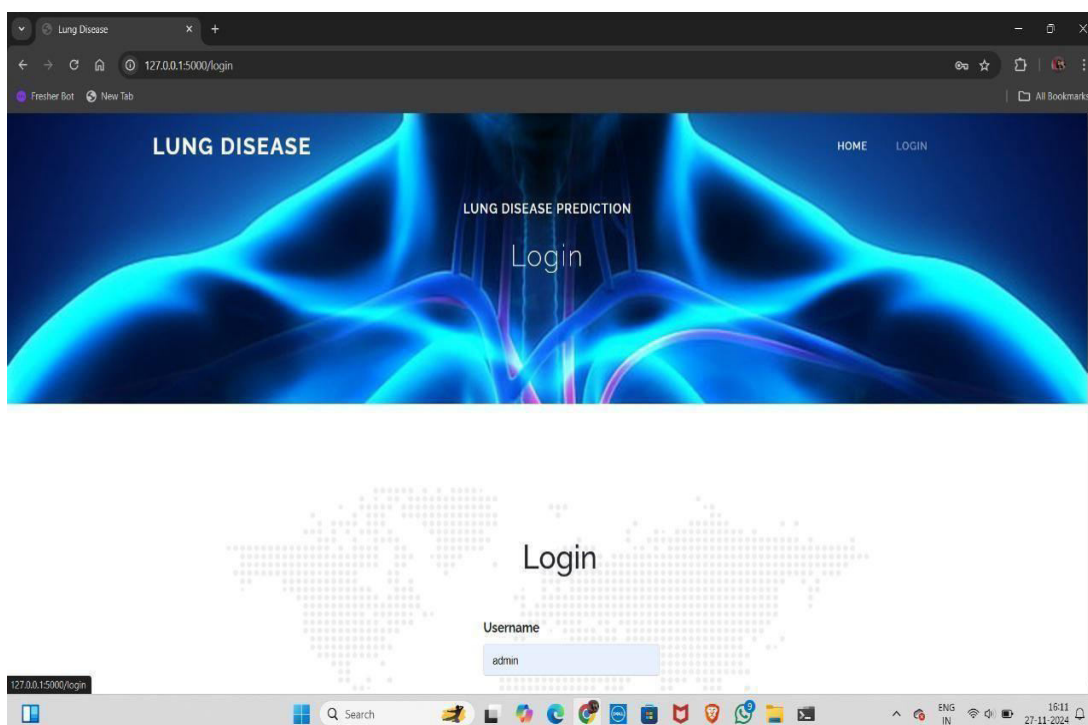


FIGURE:2

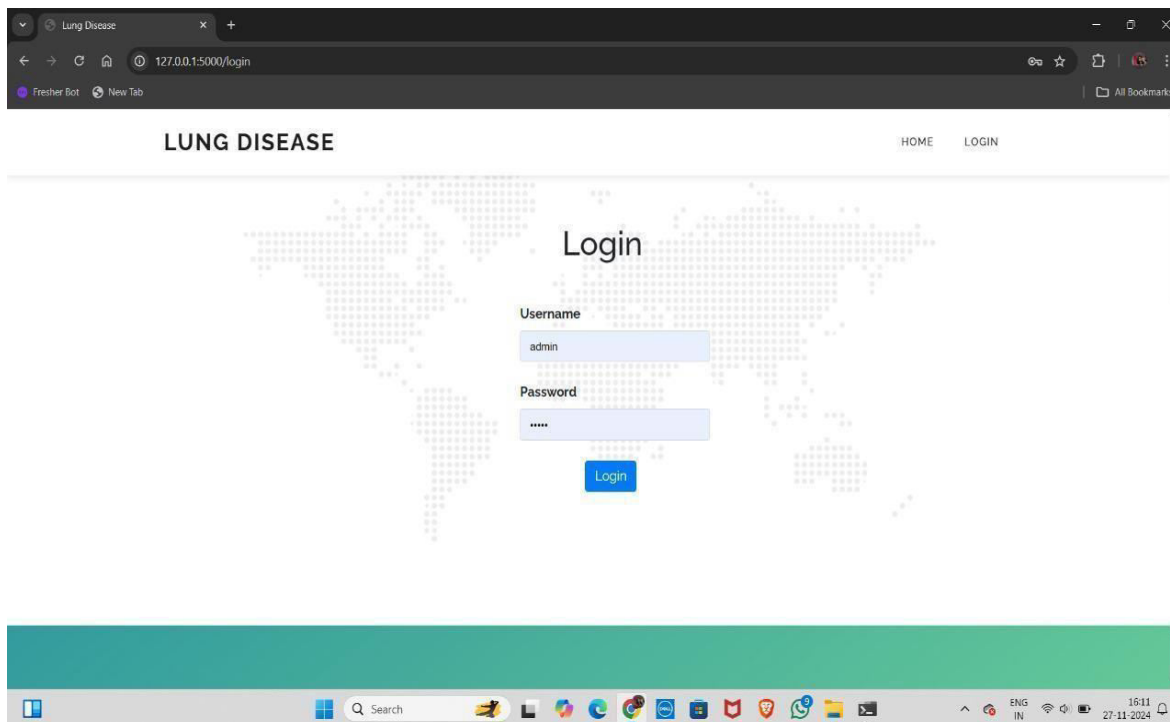


FIGURE:3

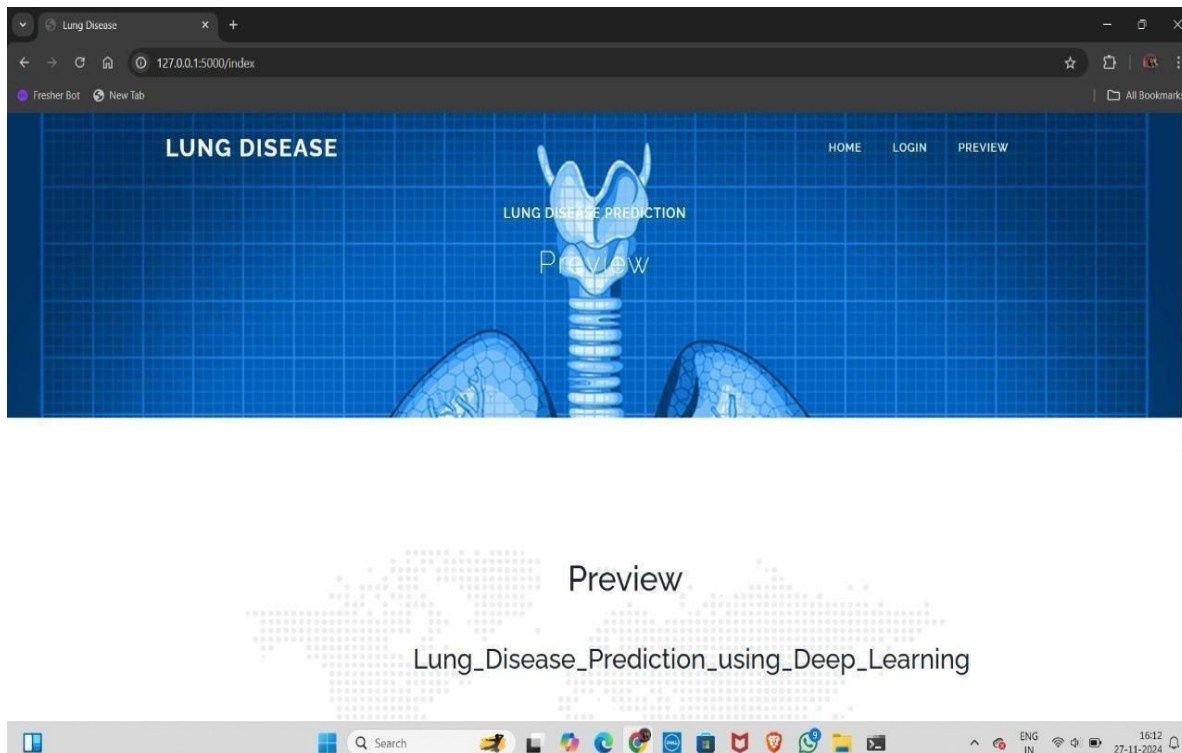


FIGURE:4



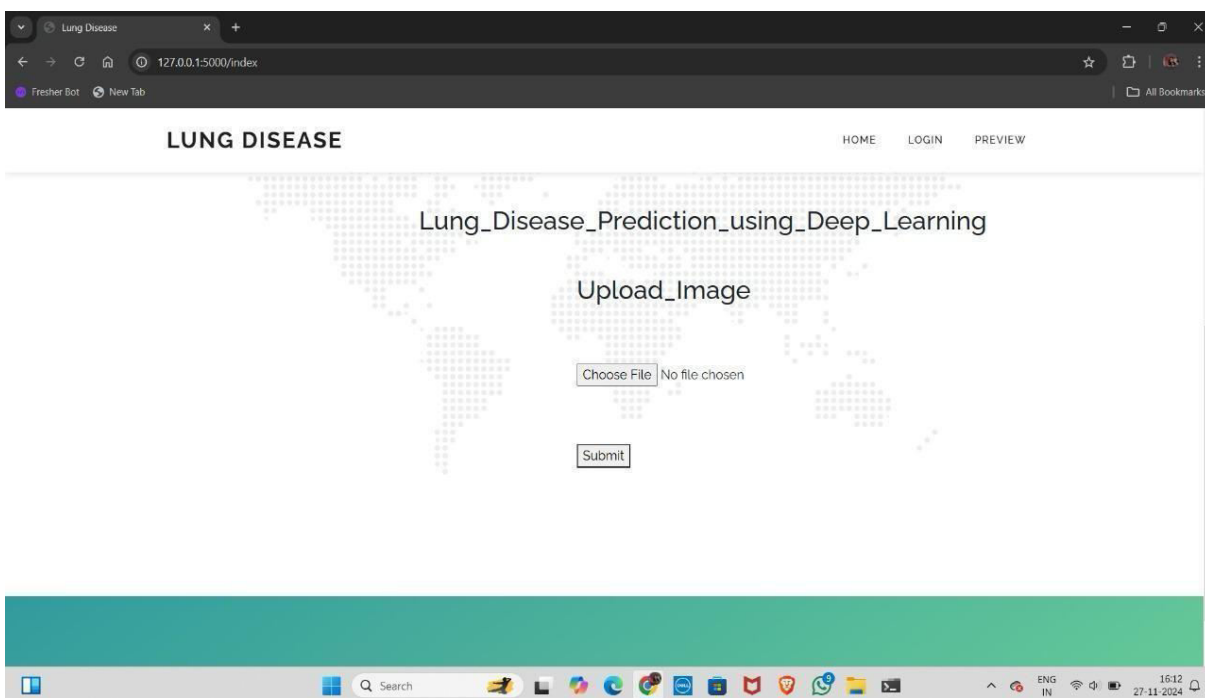


FIGURE:5

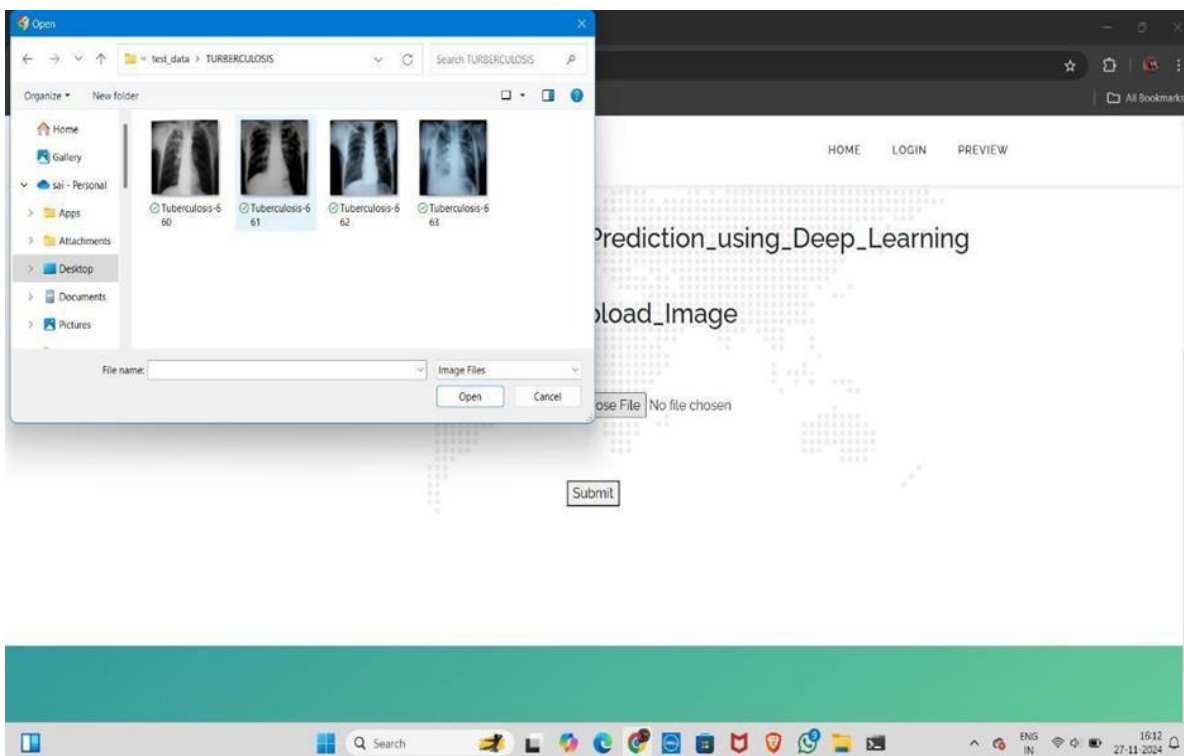


FIGURE:6

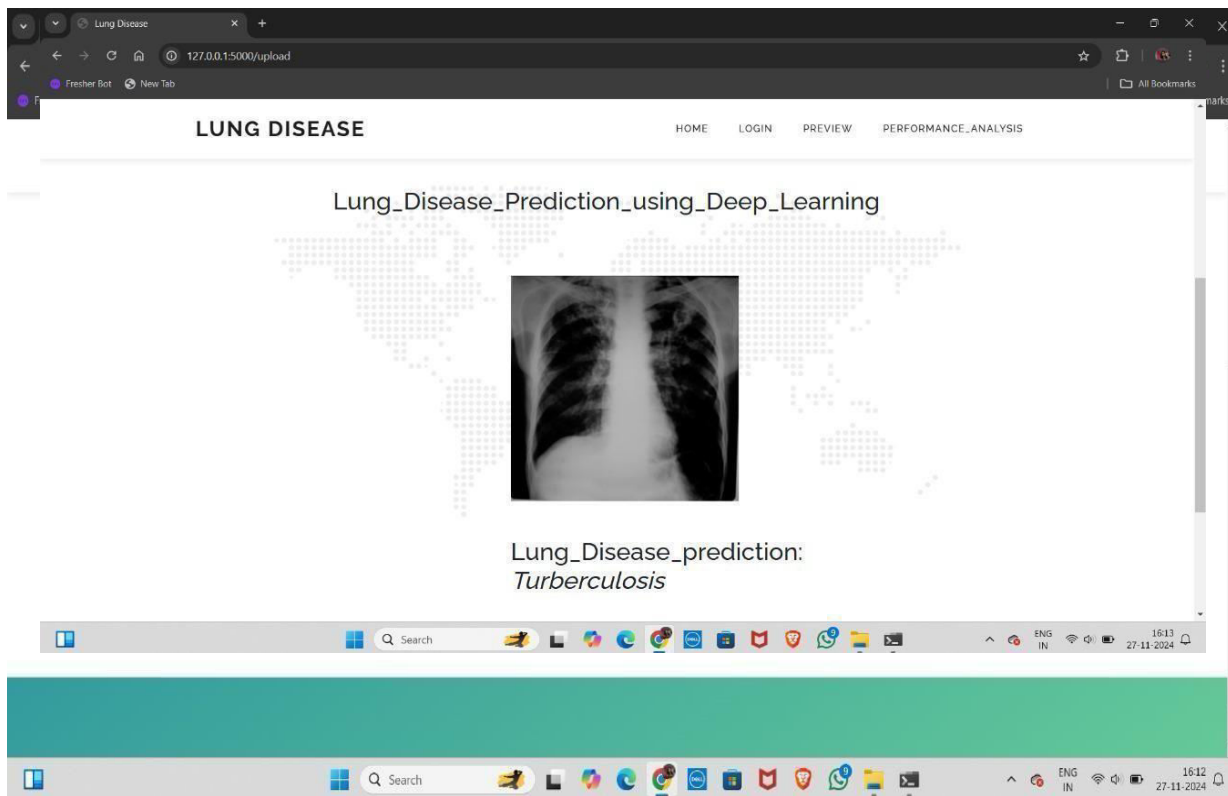


FIGURE:7

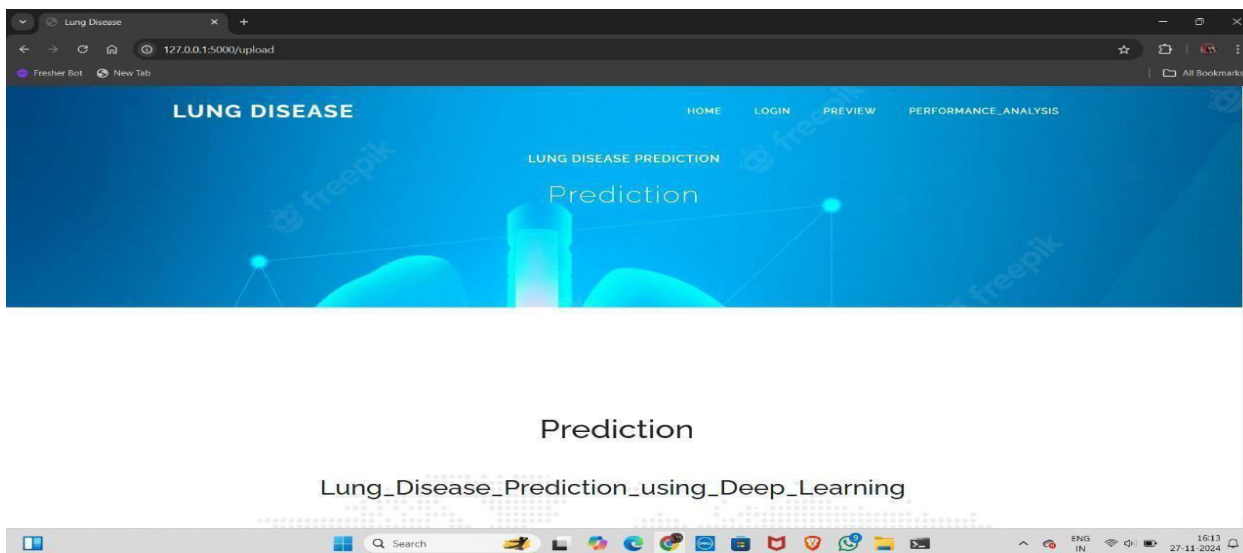


FIGURE:8

### V. CONCLUSION

In this project, the effect of the lungs of a modern patient on the various researchers and the damage to the lung is clearly explained by various researchers. Since these lung diseases have been cured the necessity of identifying this disease has become essential according to many researches. One of the main concerns of this research is to identify and select a proper datasets and technique to analyze lung diseases. Chest x-ray was selected based on the comparisons and discussions that were stated in this paper. Next a proper and suitable feature extraction algorithm was chosen since the

chest x-ray may contain lots of unnecessary data. This selection was based on advantages and disadvantages of using many common algorithms. Finally, a classification algorithm was also discussed based on their characteristic qualities. In short-term research, it was seen that VGG16 Architecture added additional benefits to predict the lung diseases in advance with better results. Ultimately, lung disease can be diagnosed.

## VI. FUTURE ENHANCEMENT

In the future, we hope to conduct a training with more data sets and change some parameters to faster the model. Some metric parameters of the metrics will also be tested. We can experiment on pre-trained model to improve the accuracy.

## REFERENCES

1. TensorFlow's main website, <https://www.tensorflow.org/>.
2. Swetha, K. R., et al. "Prediction of Pneumonia Using Big Data, Deep Learning, and Machine Learning Techniques." 2021 6th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2021.
3. Dey, Soumava, Gunther Correia Bacellar, Mallikarjuna Basappa Chandrappa, and Raj Kulkarni. "COVID-19 Chest X-Ray Image Classification Using Deep Learning." medRxiv (2021).
4. Tensorflow losses Module, [https://www.tensorflow.org/versions/r1.5/api\\_docs/python/tf/losses](https://www.tensorflow.org/versions/r1.5/api_docs/python/tf/losses)
5. CNN, [https://en.wikipedia.org/wiki/Convolutional\\_neural\\_network](https://en.wikipedia.org/wiki/Convolutional_neural_network).
6. HyperParameters, [https://en.wikipedia.org/wiki/Convolutional\\_neural\\_network#Hyperparameters](https://en.wikipedia.org/wiki/Convolutional_neural_network#Hyperparameters)
7. Swetha, K. R., et al. "Prediction of Pneumonia Using Big Data, Deep Learning, and Machine Learning Techniques." 2021 6th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2021.
8. Patil, Swati, and Akshay Golellu. "Classification of COVID-19 CT Images using Transfer Learning Models." 2021 International Conference on Emerging Smart Computing and Informatics (ESCI). IEEE, 2021.
9. Yadav, Anju, et al. "FVC-NET: An Automated Diagnosis of Pulmonary Fibrosis Progression Prediction Using Honeycomb and Deep Learning." Computational Intelligence and Neuroscience 2022 (2022).
10. Kogilavani, S.V., et al. "COVID-19 detection based on lung CT scan using deep learning techniques." Computational and Mathematical Methods in Medicine 2022 (2022).
11. Aggarwal, Karan, et al. "Has the Future Started? The Current Growth of Artificial Intelligence, Machine Learning, and Deep Learning." Iraqi Journal For Computer Science and Mathematics 3.1 (2022): 115-123.
12. Ahmed, Shaymaa Taha, and Suhad Malallah Kadhem. "Using Machine Learning via Deep Learning Algorithms to Diagnose the Lung Disease Based on Chest Imaging: A Survey." International Journal of Interactive Mobile Technologies 15.16 (2021).
13. Gowri, S., J. Jabez, J. S. Vimali, A. Sivasangari, and Senduru Srinivasulu. "Sentiment Analysis of Twitter Data Using Techniques in Deep Learning." In Data Intelligence and Cognitive Informatics, pp. 613-623. Springer, Singapore, 2021.
14. Srinivasulu, Senduru, Jeberson Retna Raj, and S. Gowri. "Analysis and Prediction of Myocardial Infarction using Machine Learning." 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2021.
15. Priya, T., and T. Meyyappan. "Disease Prediction by Machine Learning Over Big Data Lung Cancer." International Journal of Scientific Research in Computer Science, Engineering and Information Technology (2021): 16-24.
16. Battineni, Gopi. "Machine Learning and Deep Learning Algorithms in the Diagnosis of Chronic Diseases." Machine Learning Approaches for Urban Computing. Springer, Singapore, 2021. 141-164.
17. Jahnavi, Y., Kumar, P. N., Anusha, P., & Prasad, M. S. (2022, November). Prediction and Evaluation of Cancer Using Machine Learning Techniques. In International Conference on Sustainable and Innovative Solutions for Current Challenges in Engineering & Technology (pp. 399-405). Singapore: Springer Nature Singapore.
18. Saranya, S. S., Anusha, P., Chandragandhi, S., Kishore, O. K., Kumar, N. P., & Srihari, K. (2024). Enhanced decision-making in healthcare cloud-edge networks using deep reinforcement and lion optimization algorithm. Biomedical Signal Processing and Control, 92, 105963.
19. Singh, A., Kumar, T. C. A., Mithun, T., Majji, S., Rajesh, M., & Anusha, P. (2021, December). Image Processing Approaches for Oral Cancer Detection in Color Images. In 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 817-821). IEEE.



## International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 7.394