Prediction of lung cancer using SVM and CNN

#1Mr. B. SOBHAN BABU, M.Tech., (Ph.D), Assistant Professor
Department of Information Technology ,Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru
#2D. Lakshmi manasa,#3 G. Gopi Krishna Reddy,#4. M. Nikitha,#5. B. Sai Kiran
Student, Department of Information Technology ,Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru

Abstract_ The leading cause of cancer-related death is lung cancer. The windpipe, main airway, or lungs can all be the site of lung cancer onset. It results from the unchecked growth and spread of some lung cell types. Lung cancer diagnosis rates are higher in people with lung conditions like emphysema and a history of chest pain. The main risk factor for developing lung cancer in Indian men is excessive tobacco use, which includes smoking cigarettes and beedis. However, Indian women do not smoke as frequently, suggesting that there may be additional risk factors. Exposure to radon gas, air pollution, and chemicals at work are additional risk factors.
Early lung cancer detection has become crucial and simple thanks to image processing and deep learning techniques. Images from lung patient Computer Tomography (CT) scans are used in this study to identify and categorise lung nodules as well as to determine their level of malignancy. In this project, the CNN algorithm is being used to identify lung cancer in CT-SCAN images, and the dataset of CT-SCAN images was used to train the CNN algorithm.

1.INTRODUCTION

Smokers account for 85% of all cases, making them the most common group. In recent years, a large number of computer-aided diagnosis (CAD) systems have been developed. Early lung cancer detection is essential for improving survival rates and preventing fatalities. Lung nodules are tiny masses of tissue that may or may not be cancerous, also known as malignant or benign nodules. Malignant tissues grow rapidly, can affect other body parts, and are harmful to health while benign tissues are typically non-cancerous and do not grow much. There are many different image types used in medical imaging, but CT scans are typically preferred due to their reduced noise.The best technique for feature extraction, object classification, and medical imaging has been demonstrated to be deep learning. Many researchers have developed various deep learning architectures to categorise lung cancer.

2. LITERATURE SURVEY

2.1 An Automatic Detection System of Lung Nodule Based on Multi-Group Patch-Based Deep Learning Network

AUTHORS: Hongyang Jiang, He Ma, Wei Qian, MengdiGao and Yan Li

Lung nodule detection that is highly effective greatly enhances the risk assessment for lung cancer. Finding the precise positions of lung nodules quickly is a significant and difficult task. For roughly two decades, researchers have conducted extensive research in this field. However, previous computer-aided detection (CADe) methods can require additional image processing modules, such as computed tomography image transformation, lung nodule segmentation, and feature extraction, to construct a complete CADe system, making them typically complicated and time-consuming. When the number of medical images continues to rise, it is challenging for these schemes to process and analyze enormous amounts of data. In addition, the database standard may be stringent in some modern deep learning schemes. Based on multigroup patches extracted from lung images and enhanced with the Frangi filter, this study proposes an efficient method for detecting lung nodules. A four-channel convolution neural networks model is made to learn what radiologists know about detecting nodules of four levels by combining two groups of images. With only 4.7 false positives per scan, this CADe scheme has a sensitivity of 80.06% and a sensitivity of 94%, respectively. Under a large amount of image data, the results demonstrate that the multigroup patch-based learning system is effective in improving lung nodule detection performance and significantly reducing false positives..

2.2 Deep residual learning for image recognition

AUTHORS: K. He, X. Zhang, S. Ren, J. Sun

Training deeper neural networks is more challenging. To facilitate the training of networks that are significantly deeper than those that have been utilized previously, we present a residual learning framework. We unequivocally reformulate the layers as learning lingering capabilities regarding the layer inputs, rather than learning unreferenced capabilities. We provide extensive empirical evidence demonstrating that these residual networks can gain accuracy from significantly increased depth and are simpler to optimize. We evaluate residual nets with up to 152 layers on the ImageNet dataset, 8 layers deeper than VGG nets [40] but still with lower complexity. On the ImageNet test set, an ensemble of these residual nets has an error of 3.57 percent. On the classification test for the ILSVRC 2015, this result placed first. Additionally, our research on CIFAR-10 with 100 and 1000 layers is presented. For a lot of visual recognition tasks, the depth of representations is extremely important. We outperform the COCO object detection dataset by 28% solely due to our extremely deep representations. Our submissions to the ILSVRC and COCO 2015 competitions were based on deep residual nets, and we also took first place in the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation..

2.3 Accurate Pulmonary Nodule Detection in computed Tomography Images Using Deep Convolutional Neural Networks

AUTHORS: Jia Ding, Aoxue Li, Zhiqiang Hu, Liwei Wang

Early discovery of pneumonic malignant growth is the most encouraging method for improving a patient's opportunity for endurance. A crucial step in the diagnosis of pulmonary cancer is the accurate detection of pulmonary nodules in computed tomography (CT) images. We propose a novel pulmonary nodule detection strategy based on deep convolutional neural networks (DCNNs) in this paper, inspired by the successful application of DCNNs in natural image recognition. For candidate detection on axial slices, we first introduce a deconvolutional structure to the Faster Region-based Convolutional Neural Network (Faster R-CNN). Then, at that point, a three-layered DCNN is introduced for the resulting bogus positive decrease. The experimental results of the LUng Nodule Analysis 2016 (LUNA16) Challenge demonstrate that the proposed method outperforms the best result on the LUNA16 Challenge leaderboard in terms of nodule detection (average FROC-score of 0.864), placing it first out of all the submitted results.

3.PROPOSED SYSTEM

Early lung cancer detection has become crucial and simple thanks to image processing and deep learning techniques. Images from lung patient Computer Tomography (CT) scans are used in this study to identify and categorise lung nodules as well as to determine their level of malignancy. In this project, the CNN algorithm is being used to identify lung cancer in CT-SCAN images, and the dataset of CT-SCAN images was used to train the CNN algorithm. The primary goal of this study is to investigate the performance of a classification algorithm to aid in the early diagnosis of lung cancer**.**

3.1 IMPLEMENTATION

- Upload Lung Cancer Dataset

In this module use upload dataset

- Preprocess Dataset

In this module data undergoes preprocessing

- Model Generation

In this module model generation is take place to predict desease.

- Build CNN Model

In this module CNN model is build.

- Accuracy & Loss Graph

In this module comparison graph is shown

- Upload Test Image & Predict Cancer

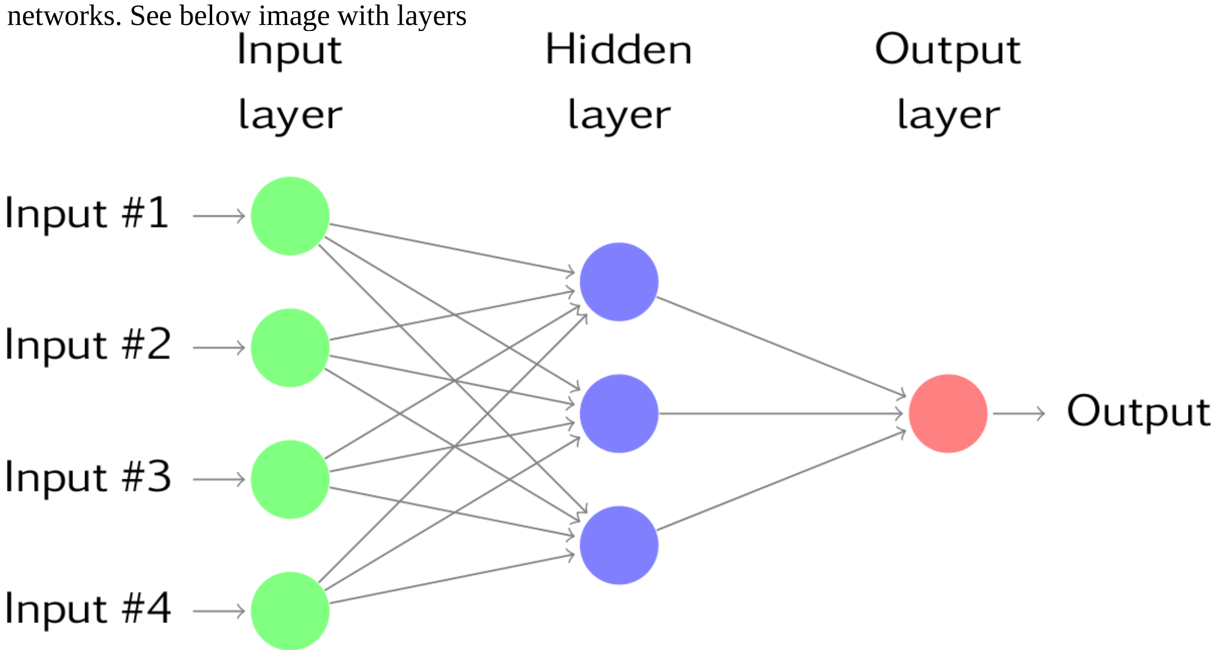
In this module, user uploads test image to predict diseases.

3.2 Convolutional neural network:

To demonstrate how to build a convolutional neural network based image classifier, we shall build a 6 layer neural network that will identify and separate one image from other. This network that we shall build is a very small network that we can run on a CPU as well. Traditional neural networks that are very good at doing image classification have many more parameters and take a lot of time if trained on normal CPU. However, our objective is to show how to build a real-world convolutional neural network using TENSORFLOW.

Neural Networks are essentially mathematical models to solve an optimization problem. They are made of neurons, the basic computation unit of neural networks. A neuron takes an input (say x), do some computation on it (say: multiply it with a variable w and adds another variable b) to produce a value (say; z= wx+b). This value is passed to a non-linear function called activation function (f) to produce the final output(activation) of a neuron. There are many kinds of activation functions. One of the popular activation function is Sigmoid. The neuron which uses sigmoid function as an activation function will be called sigmoid neuron. Depending on the activation functions, neurons are named and there are many kinds of them like RELU, TanH.

If you stack neurons in a single line, it's called a layer; which is the next building block of neural networks. See below image with layers



To predict image class multiple layers operate on each other to get best match layer and this process continues till no more improvement left.

Deep learning not only accelerates the critical task but also improves the precision of the computer and the performance of CT image detection and classification.

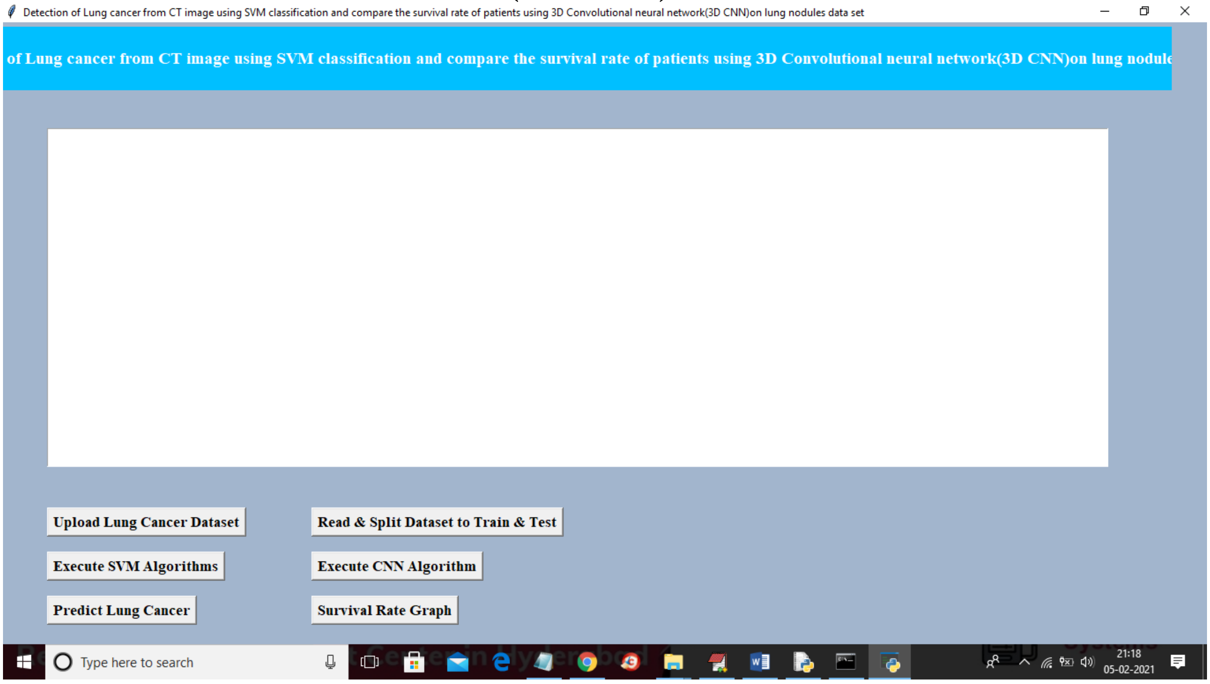
In this paper, the problem of classification of benign and malignant is considered. It is proposed to employ, respectively, the convolution neural network (CNN)and deep neural network (DNN).

The input data (image data) has a strong robustness on the distortion. The multiscale convolution image feature is generated by setting the convolution kernel size and parameter; the information of different angles is generated in the feature space.

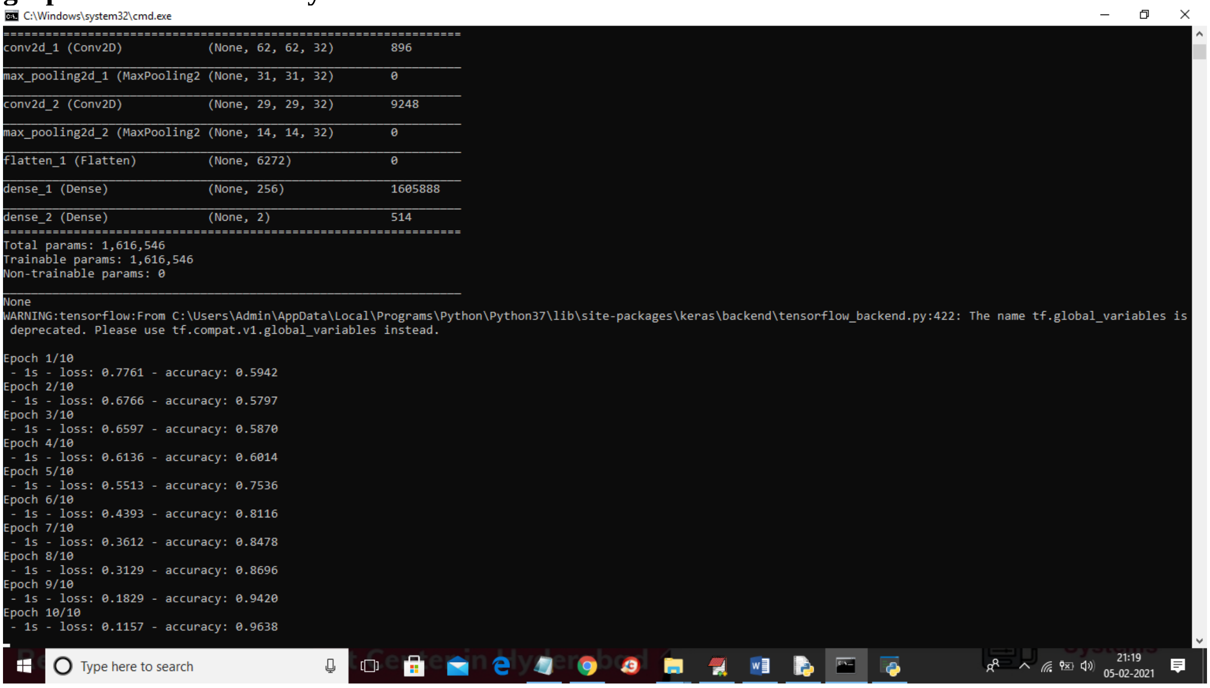
4.RESULTS AND DISCUSSION

ISSN NO: 2230-5807

In this project we are using same above Lungs dataset to train CNN and SVM algorithm and then calculate survival rate of patients by using both algorithms. If algorithm predicted 18 records correctly out of 20 records then survival rate will be (18/20 * 100) = 90%.

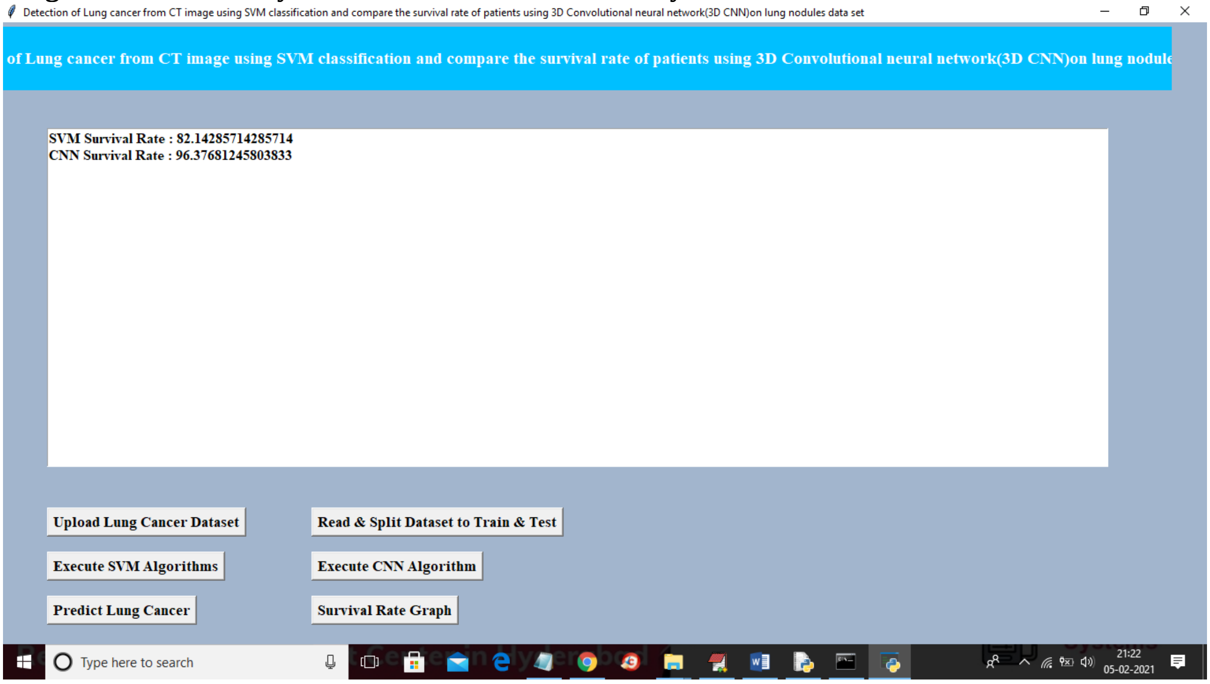


Similar to first two projects here also you upload 'Dataset' folder and then click on "read & split" button and then execute SVM and CNN and then predict cancer and go for survival rate graph. For CNN results you can refer to black console below

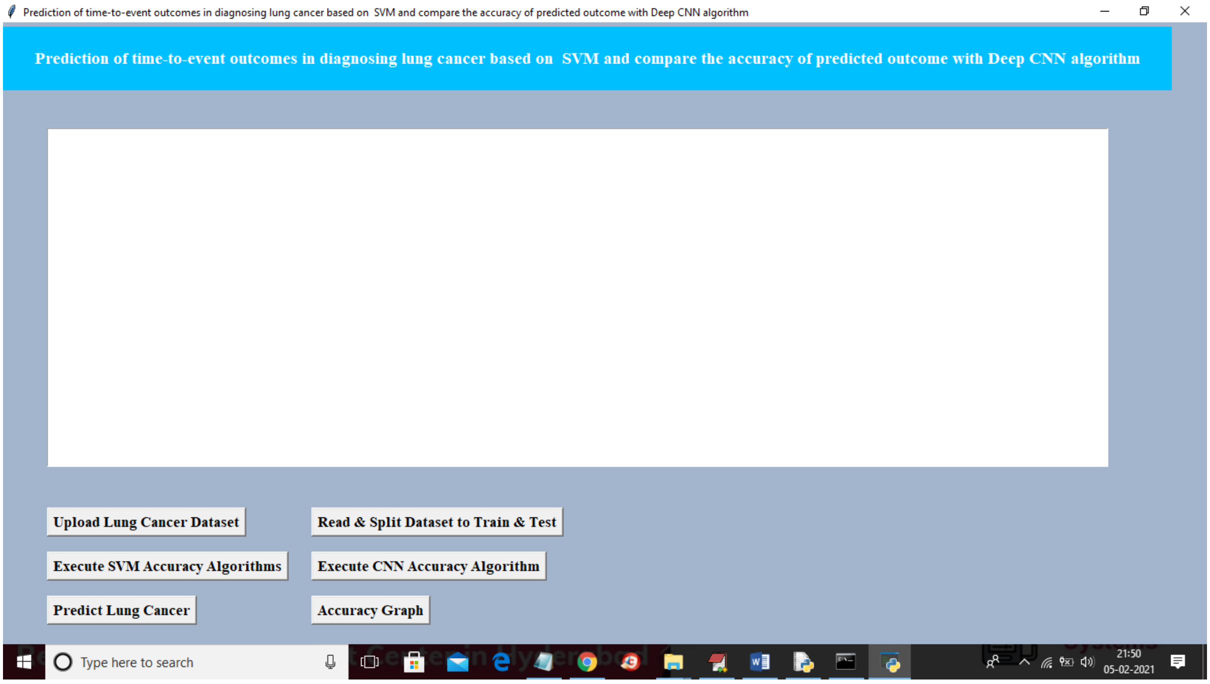


In above screen you can see for CNN we use multiple filters to filter dataset for better prediction result and in above screen in first layer CNN use 62 X 62 image size with 32 filters and in second layer for 31 X 31 image size also it uses 32 filters and for each filter we will have best image features and prediction accuracy will be better. In above screen to run CNN I used 10

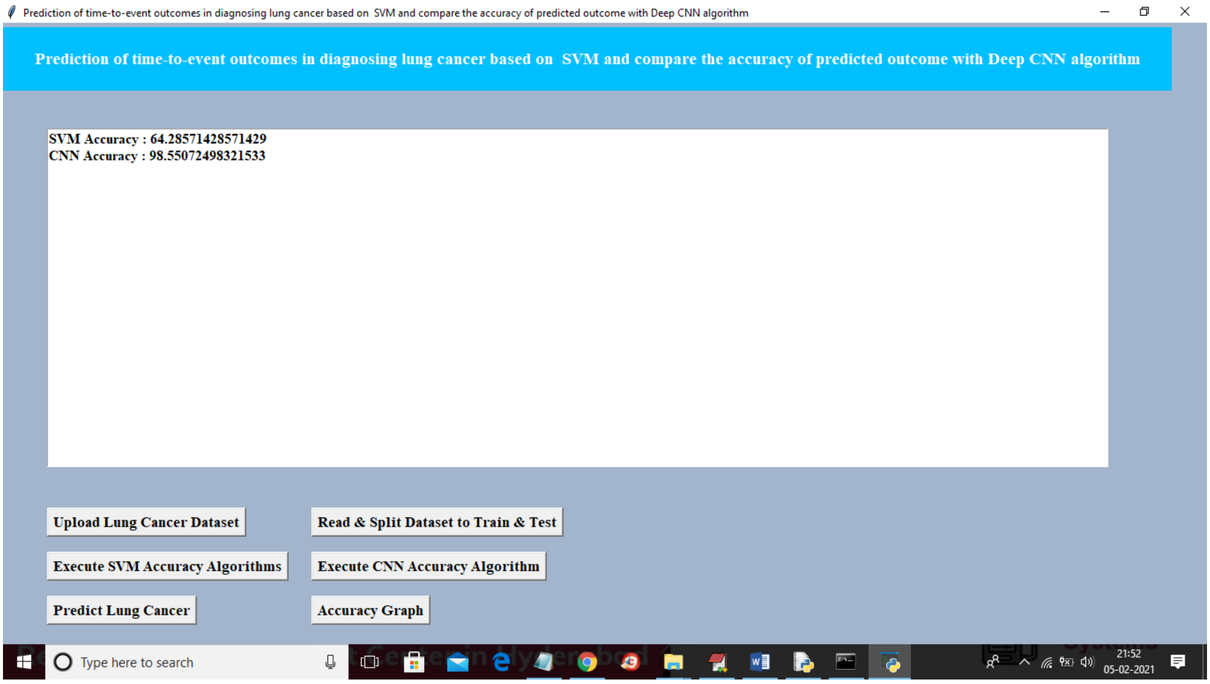
epoch/iteration and for each increase iteration accuracy get better and better and for last epoch we got 0.96% accuracy and below is the final accuracy result for both SVM and CNN



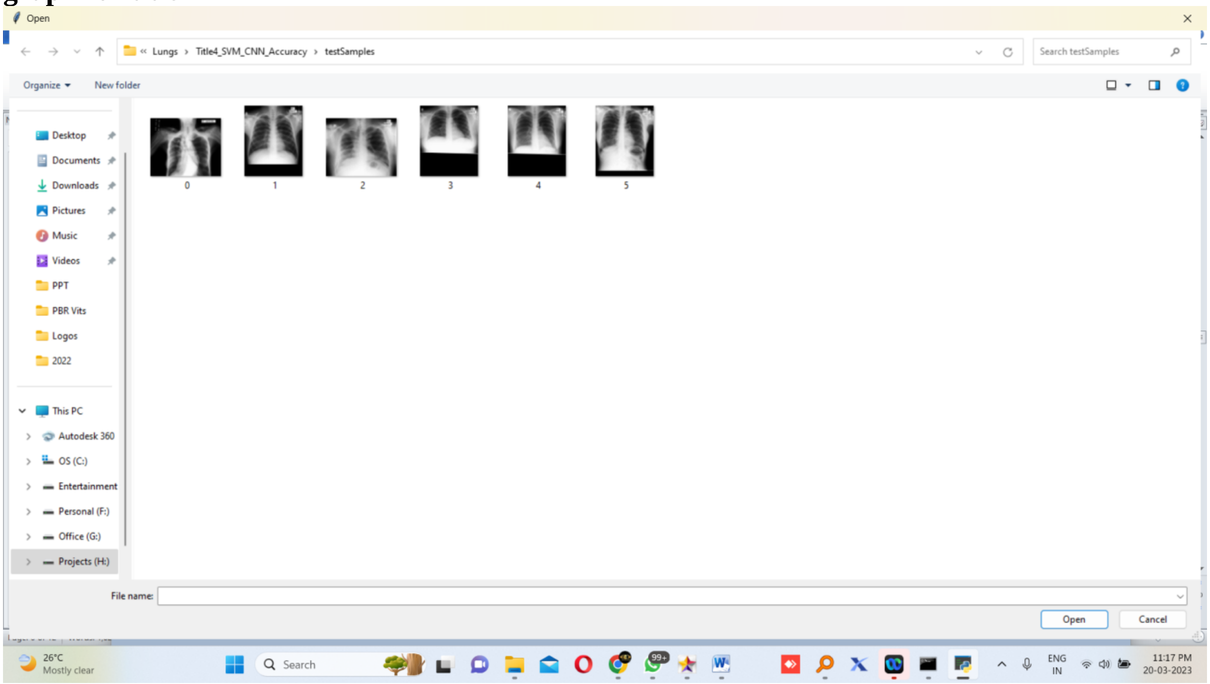
In above screen SVM survival rate is 82% and CNN survival rate is 96% and similarly you can go for predict button and graph button.



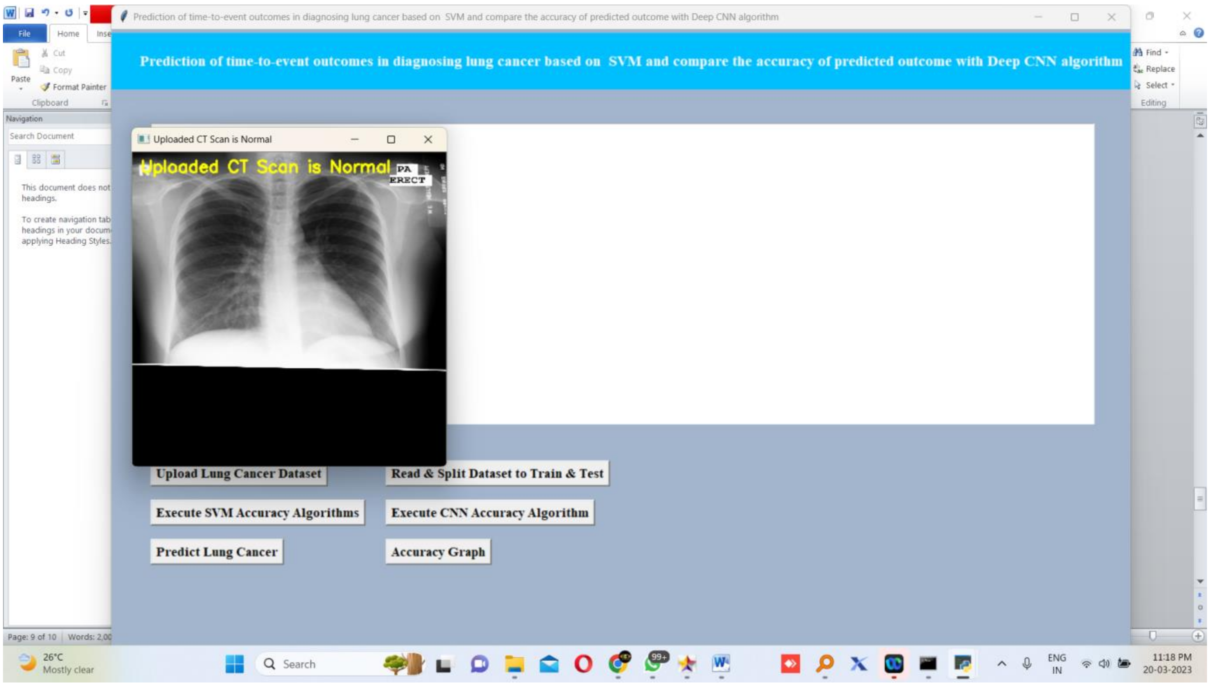
In above screen similar to first two projects upload dataset and then click on 'read and split dataset' button and then execute SVM with accuracy and CNN with accuracy and then you can go for predict lung cancer and accuracy graph



In above screen SVM accuracy is 64% and CNN accuracy is 98% and below is the comparison graph for title 4



In the above screen we are predicting lung cancer



In the above screen we are predicted as normal

5.CONCLUSION

In the past, a doctor would need to perform a number of tests to determine whether a patient had lung cancer or not. However, this was a lengthy process. A patient may occasionally be required to undergo pointless examinations or additional tests in order to diagnose lung cancer. There must be a preliminary test that alerts the patient and the doctor to the possibility of lung cancer in order to reduce process time and pointless examinations. Today, the prediction and classification of medical data heavily relies on machine learning algorithms.We can see the anticipated outcome because CT-SCAN shows abnormalities. In the second image, we are identifying the locations where abnormalities were found, and in the third image, we have extracted and displayed all of the abnormality patches from the original image.

REFERENCES

1. DSharma GJindal ‘‘Computer aided diagnosis system for detection of lung cancer in CT scan images International Journal of Computer and Electrical Engineering 3(5): 714-718 2011.
2. Janarthanan R Balamurali R Annapoorani A and Vimala V 2020 Prediction of rainfall using fuzzy logic Materials Today: Proceedings
3. DBhatnagarAKTiwariVVijayarajanAKrishnamoorthy ‘‘Classification of normal and abnormal images of lung cancer IOP Conference Series: Materials Science and Engineering Vol 263 2017