**LUNG CANCER DETECTION USING MACHINE LEARNING TECHNIQUES**

**Abstract:**

Lung cancer is one of the most lethal cancer types; thousands of peoples are infected with this type of cancer, and if they do not discover it in the early stages of the disease, then the chance of surviving of the patient will be very poor. For the suggested reasons above and to help in overcoming this terrible, early diagnosis with the assistance of artificial intelligence procedures most needed. Through this research, a Computer-aided system introduced for detecting lung cancer in a dataset collected from the Iraqi hospitals by using a convolutional neural network technique with AlexNet architecture for helping with the diagnosis of the patient's cases: normal, benign, or malignant. The proposed model gives high accuracy ups to 93.548%. The other performance metrics comes with high values such as 95.714% for sensitivity and 95% for Specificity.

**CHAPTER-1**

INTRODUCTION:

Lung cancer is one of the most well-known life-threatening illnesses in the globe(1). The up-to-date estimations are given by the "World Health Organization" (WHO) say that approximately 7.6 million mortality worldwide every year because of lung cancer. Furthermore, mortality due to this type of cancer is assumed to continue growing, to become almost 17 million worldwide in 2030(2, 3). According to the statistics of “The American cancer society”, lung cancer is the head cancer killer in people in the United States.(4) The overall number of the estimated new cases of all types of cancer in 2013 was 1660290 (854790 for men and 805500 for women), the number of lung cancer was 228199 incidences (118080 for men and 110110 for women). Where the total number of estimated death cases of cancer are 580350 cases (306920 for men and 273430 for women), in the case of the lung cancer the number of death cases was 159480(87260 for men and 732220 for women)(5). According to the Iraqi ministry of health in 2016, lung cancer is the second most widespread cancer type in Iraq. There are 2123 people who have lung cancer from two genders. This number represents about 8.31% of the total infections in the country. This portion indicates a small increase as compared to the ratio of the past year, which is about 8.1%. The rate of lung cancer represents approximately 13.27% of the total cancer cases, and this shows that lung cancer is the leading type in males. Also, there is a rise when compares to the ratio recorded in 2015, which approximately reaches 12.7%.(6, 7)**.** For females, lung cancer not the leading cancer type, it ranked fifth between other cancer types, in 2016 there is only 638 woman who has this disease, there are 638 women who diagnosed with lung cancer at 2016, this represents about 4.44% from the total cancer types infections. There is a slight rise if compared with the previous year, which represents approximately about 4.2%.(6, 7) Cancer is the fourth cause of death in the eastern Mediterranean region and is the third cause of mortality in Iraq,

and this rate is growing continuously. The chief and most notable of this raise is smoking. Other factors include pollution, unhealthy diet, endless exposure to manufacturing and agricultural carcinogens, and lowered physical motion(6). Total cancer mortality in Iraq in 2014 is (8211), approximately (4525) in males and (3959) in females. The most cancer sort deaths were the lung cancer with a total number of estimated deaths about (1339), (918) of them was for men and (421) of the total calculated estimate was for women, the entire portion of lung cancer

mortality amongst all other kinds of malignancy was 16.31%. During 2016, the whole number of cancer losses fell to 7568 cases, where the most significant portion of them was lung cancer, approximately 1257, which implies a total percentage of 16.61% from the total predicted mortality(6, 7). So for the reasons explained above, there is a necessity for implementing a CAD system for helping doctors in diagnosing lung cancer as possible as when in it is early stage, not only detecting the nodule but with high accuracy. Several studies applied artificial intelligence techniques for this purpose, for examples: using artificial neural network for detecting lung cancer as in (8, 9), or using support vector machine technique as in (10-12)*,* or applying

K-nearest neighbor as in (13), or using genetic algorithm for this operation as in (14, 15), also, fuzzy techniques are efficient when using to detect lung cancer as in (16-18), convolutional neural network can be used for this purpose as in (19-21). Artificial intelligence not only used in the area of lung cancer diagnosing, but it applied in all fields of biomedical engineering such as: diagnosis of breast cancer in (22-24), diagnosis of Heart disease in (25-27), also diagnosing and classification of diabetes in (28). In order to apply the above machine learning techniques, there is a need for using data as input to the algorithms which have been applied. Various methods are convenient for diagnosing lung cancer, particularly MRI, isotope, Xray, and CT. X-ray chest radiography and Computer Tomography (CT) are the two well-known imaging modalities

that are commonly utilized in the identification of different lung diseases(2, 29). In addition, there are many publicly databases used for the purposes of scientific research such as: ELCAP Public Lung Image Database, LIDC Database, and Data Science Bowl 2017.

The aim of our study is to implement a CAD system used as an assistant to doctors while deciding and diagnosing lung cancer, this system used for detecting and classifying the lung cancer cases if it normal, benign, or malignant with high accuracy. This done by applying convolutional neural network technique to a data set of lung cancer CT scans collected and diagnosed at the Iraqi hospitals.

**1.2 OVERVIEW**

Computer-aided system introduced for detecting lung cancer in a dataset collected from the Iraqi hospitals by using a convolutional neural network technique with AlexNet architecture for helping with the diagnosis of the patient's cases: normal, benign, or malignant.

**1.3 RELATED WORK**

In the proposed procedure, CNNs are applied to detect and classify lung cancer CT scans of the patients collected from hospitals. **Convolutional Neural Networks** is a sort of deep learning paradigm applied for processing data which has a grid pattern like images (31), it is all about using Deep Learning with Computer Vision. A good way to gain foreknowledge about this technique is to imagine a Neural Network Architecture also how it is practiced to

visual tasks i.e. Video and Images. Furthermore, the Convolutional Neural Networks is an important technique used for Object Recognition, create Facial Recognition, Self-Driving Cars. A **Convolutional Neural Network** is a Deep Learning algorithm that can take in image as input, with assigning importance learnable weights and biases to various objects inside this image and be capable of differentiating one from the other. In addition, the pre-processing

required for this technique is much lower if comparing with other classification algorithms. The role of the CNN is for reducing the images to a form that is easier to process but without losing features that are important for getting a good prediction (32). A typical CNN consists of three types of operation layers: the convolutional layer (CONV), the pooling layer (POOL), and finally the classifier layer (FC).

**1.4 PROJECT STATEMENT**

Lung cancer is the most perilous and widely spread cancer in the world according to stage of discovery of the cancer cells in the lungs. The motivation of this paper is to discover the cancer cells in the lungs at earlier stage. The lung cancer is detected using CAD System which is an interdisciplinary approach based on the techniques of Image Processing and Machine Learning. The forecasting of lung cancer is the most challenging problem, because of the structure of cancer cells, where most of the cells are superimposed. Recently, the image processing techniques are widely used in several medical areas for detection and treatment levels. The time factor is most important to determine the abnormality issues in the targeted image. Image quality and accuracy are the significant factors for quick identification of diseases. Image quality assessment and advancement are depending on the enhancement stage. For the suggested reasons above and to help in overcoming this terrible, early diagnosis with the assistance of artificial intelligence procedures most needed. Through this research, a Computer-aided system introduced for detecting lung cancer in a dataset collected from the Iraqi hospitals by using a convolutional neural network technique with AlexNet architecture for helping with the diagnosis of the patient's cases: normal, benign, or malignant. The proposed model gives high accuracy ups to 93.548%. The other performance metrics comes with high values such as 95.714% for sensitivity and 95% for Specificity.

**1.5 SCOPE OF THE PROJECT**

The scope of this project is to use Deep Learning techniques for early detection of lung cancer from CT scans and use the results in clinical diagnostics and cancer screening applications to support radiologist's diagnosis.

**1.6 PROPOSED SYSTEM**

Lung cancer is one of the most well-known life-threatening illnesses in the globe(1). The up-to-date estimations are given by the "World Health Organization" (WHO) say that approximately 7.6 million mortality worldwide every year because of lung cancer. Through this research, a Computer-aided system introduced for detecting lung cancer in a dataset collected from the Iraqi hospitals by using a convolutional neural network technique with AlexNet architecture for helping with the diagnosis of the patient's cases: normal, benign, or malignant. The proposed model gives high accuracy ups to 93.548%. The other performance metrics comes with high values such as 95.714% for sensitivity and 95% for Specificity.

**CHAPTER-2**

**Literature Survey:**

2.1.TITLE:

Are increasing 5-year survival rates evidence of success against cancer

AUTHOR:

Welch HG, Schwartz LM, Woloshin S.

CONTENT:

**Context:**Increased 5-year survival for cancer patients is generally inferred to mean that cancer treatment has improved and that fewer patients die of cancer. Increased 5-year survival, however, may also reflect changes in diagnosis: finding more people with early-stage cancer, including some who would never have become symptomatic from their cancer.

**Objective:**To determine the relationship over time between 5-year cancer survival and 2 other measures of cancer burden, mortality and incidence.

**Design and setting:**Using population-based statistics reported by the National Cancer Institute Surveillance, Epidemiology, and End Results Program, we calculated the change in 5-year survival from 1950 to 1995 for the 20 most common solid tumor types. Using the tumor as the unit of analysis, we correlated changes in 5-year survival with changes in mortality and incidence.

**Main outcome measure:**The association between changes in 5-year survival and changes in mortality and incidence measured using simple correlation coefficients (Pearson and Spearman).

**Results:**From 1950 to 1995, there was an increase in 5-year survival for each of the 20 tumor types. The absolute increase in 5-year survival ranged from 3% (pancreatic cancer) to 50% (prostate cancer). During the same period, mortality rates declined for 12 types of cancer and increased for the remaining 8 types. There was little correlation between the change in 5-year survival for a specific tumor and the change in tumor-related mortality (Pearson r=.00; Spearman r=-.07). On the other hand, the change in 5-year survival was positively correlated with the change in the tumor incidence rate (Pearson r=+. 49; Spearman r=+.37).

**Conclusion:**Although 5-year survival is a valid measure for comparing cancer therapies in a randomized trial, our analysis shows that changes in 5-year survival over time bear little relationship to changes in cancer mortality. Instead, they appear primarily related to changing patterns of diagnosis. JAMA. 2000.

2.2.TITLE:

Deep learning for lung Cancer detection and classification. Multimedia Tools Applications

AUTHOR:

Asuntha A, Srinivasan A

CONTENT:

Lung cancer is one of the main reasons for death in the world among both men and women, with an impressive rate of about five million deadly cases per year. Computed Tomography (CT) scan can provide valuable information in the diagnosis of lung diseases. The main objective of this work is to detect the cancerous lung nodules from the given input lung image and to classify the lung cancer and its severity. To detect the location of the cancerous lung nodules, this work uses novel Deep learning methods. This work uses best feature extraction techniques such as Histogram of oriented Gradients (HoG), wavelet transform-based features, Local Binary Pattern (LBP), Scale Invariant Feature Transform (SIFT) and Zernike Moment. After extracting texture, geometric, volumetric and intensity features, Fuzzy Particle Swarm Optimization (FPSO) algorithm is applied for selecting the best feature. Finally, these features are classified using Deep learning. A novel FPSOCNN reduces computational complexity of CNN. An additional valuation is performed on another dataset coming from Arthi Scan Hospital which is a real-time data set. From the experimental results, it is shown that novel FPSOCNN performs better than other techniques.

2.3.TITLE:

Future health prediction from multimedia and multimodal observations. Proceedings of the 23rd ACM international conference on Multimedia

AUTHOR:

Nie L, Zhang L, Yang Y, Wang M, Hong R, Chua T-S

CONTENT:

Although chronic diseases cannot be cured, they can be effectively controlled as long as we understand their progressions based on the current observational health records, which is often in the form of multimedia data. A large and growing body of literature has investigated the disease progression problem. However, far too little attention to date has been paid to jointly consider the following three observations of the chronic disease progression: 1) the health statuses at different time points are chronologically similar; 2) the future health statuses of each patient can be comprehensively revealed from the current multimedia and multimodal observations, such as visual scans, digital measurements and textual medical histories; and 3) the discriminative capabilities of different modalities vary significantly in accordance to specific diseases. In the light of these, we propose an adaptive multimodal multi-task learning model to co-regularize the modality agreement, temporal progression and discriminative capabilities of different modalities. We theoretically show that our proposed model is a linear system. Before training our model, we address the data missing problem via the matrix factorization approach. Extensive evaluations on a real-world Alzheimer’s disease dataset well verify our proposed model. It should be noted that our model is also applicable to other chronic diseases

2.4.TITLE:

. Cancer statistics, 2013. a cancer journal for clinicians.

AUTHOR:

Siegel, Rebecca, Naishadham, Deepa, Jemal, Ahmedin

CONTENT:

Each year, the American Cancer Society estimates the numbers of new cancer cases and deaths expected in the United States in the current year and compiles the most recent data on cancer incidence, mortality, and survival based on incidence data from the National Cancer Institute, the Centers for Disease Control and Prevention, and the North American Association of Central Cancer Registries and mortality data from the National Center for Health Statistics. A total of 1,660,290 new cancer cases and 580,350 cancer deaths are projected to occur in the United States in 2013. During the most recent 5 years for which there are data (2005-2009), delay-adjusted cancer incidence rates declined slightly in men (by 0.6% per year) and were stable in women, while cancer death rates decreased by 1.8% per year in men and by 1.5% per year in women. Overall, cancer death rates have declined 20% from their peak in 1991 (215.1 per 100,000 population) to 2009 (173.1 per 100,000 population). Death rates continue to decline for all 4 major cancer sites (lung, colorectum, breast, and prostate). Over the past 10 years of data (2000-2009), the largest annual declines in death rates were for chronic myeloid leukemia (8.4%), cancers of the stomach (3.1%) and colorectum (3.0%), and non-Hodgkin lymphoma (3.0%). The reduction in overall cancer death rates since 1990 in men and 1991 in women translates to the avoidance of approximately 1.18 million deaths from cancer, with 152,900 of these deaths averted in 2009 alone. Further progress can be accelerated by applying existing cancer control knowledge across all segments of the population, with an emphasis on those groups in the lowest socioeconomic bracket and other underserved populations.

2.5.TITLE:

Annual Report Iraqi Cancer Registry

AUTHOR:

Republic of Iraq , Ministry of Health\Environment, Board. IC.

CONTENT:

9 December 2019 ‒ Recently announced data from Iraq’s national cancer registry indicate that there are over 31,500 cancer and tumor-related cases in Iraq as of 2017‒2018. Cancer is considered one of the leading causes of mortality in the country, contributing to an estimated 11% of total deaths.

“These alarming figures highlight the need for WHO, the Ministry of Health and all partners to address cancer management as a key public health priority programme in the country. The cancer registration data for 2017‒2018 provides essential data that will contribute to the development of effective prevention and treatment plans for the disease,” said Dr Adham Ismail, WHO Representative in Iraq.

The national population-based cancer registry programme in Iraq, supported by WHO and the International Agency for Research on Cancer (IARC), collects, analyses and provides information on cancer-related cases and mortalities across the country. The data is collected from both public and private sector hospitals and laboratories countrywide.

To ensure accurate acquisition and management of data at both the central and governorate levels in Iraq, WHO recently trained seven master trainers and staff working in cancer institutions on the updated CanReg 5 program, an open-source tool to input, store, check and analyse cancer registry data

2.6.TITLE:

Lung Cancer Detection Using Artificial Neural Network. International Journal of Engineering Information Systems.

AUTHOR:

Nasser, Ibrahim M Abu-Naser, S. S

CONTENT:

In this paper, we developed an Artificial Neural Network (ANN) for detect the absence or presence of lung cancer in human body. Symptoms were used to diagnose the lung cancer, these symptoms such as Yellow fingers, Anxiety, Chronic Disease, Fatigue, Allergy, Wheezing, Coughing, Shortness of Breath, Swallowing Difficulty and Chest pain. They were used and other information about the person as input variables for our ANN. Our ANN established, trained, and validated using data set, which its title is “survey lung cancer”. Model evaluation showed that the ANN model is able to detect the absence or presence of lung cancer with 96.67 % accuracy.

2.7.TITLE:

Lung cancer detection by using artificial neural network and fuzzy clustering methods.

AUTHOR:

.Taher, Fatma Sammouda, Rachid

CONTENT:

The early detection of lung cancer is a challenging problem, due to the structure of the cancer cells, where most of the cells are overlapped with each other. This paper presents two segmentation methods, Hopfield Neural Network (HNN) and a Fuzzy C-Mean (FCM) clustering algorithm, for segmenting sputum color images to detect the lung cancer in its early stages. The manual analysis of the sputum samples is time consuming, inaccurate and requires intensive trained person to avoid diagnostic errors. The segmentation results will be used as a base for a Computer Aided Diagnosis (CAD) system for early detection of lung cancer which will improve the chances of survival for the patient. However, the extreme variation in the gray level and the relative contrast among the images make the segmentation results less accurate, thus we applied a thresholding technique as a pre-processing step in all images to extract the nuclei and cytoplasm regions, because most of the quantitative procedures are based on the nuclear feature. The thresholding algorithm succeeded in extracting the nuclei and cytoplasm regions. Moreover, it succeeded in determining the best range of thresholding values. The HNN and FCM methods are designed to classify the image of *N* pixels among *M* classes. In this study, we used 1000 sputum color images to test both methods, and HNN has shown a better classification result than FCM, the HNN succeeded in extracting the nuclei and cytoplasm regions.

2.8.TITLE:

Computer-aided detection of Pulmonary Nodules based on SVM in thoracic CT images.

AUTHOR:

Eskandarian P, Bagherzadeh J

CONTENT:

Lung cancer is one of the most lethal cancer types; thousands of peoples are infected with this type of cancer, and if they do not discover it in the early stages of the disease, then the chance of surviving of the patient will be very poor. For the suggested reasons above and to help in overcoming this terrible, early diagnosis with the assistance of artificial intelligence procedures most needed. Through this research, a Computer-aided system introduced for detecting lung cancer in a dataset collected from the Iraqi hospitals by using a convolutional neural network technique with AlexNet architecture for helping with the diagnosis of the patient's cases: normal, benign, or malignant. The proposed model gives high accuracy ups to 93.548%. The other performance metrics comes with high values such as 95.714% for sensitivity and 95% for Specificity.

2.9.TITLE:

Classification of X-rays using statistical moments and SVM

AUTHOR:

Ganesan S, Subashini T, Jayalakshmi K

CONTENT:

Due to collection of thousands of medical digital images every day in medical institutions, the usage of medical digital images has been increasing rapidly. Because of the increase in medical digital images, managing this data properly and accessing it accurately is a raising need. To overcome the difficulties of the manual classification automated method is proposed. In this work an effort has been made to automatically classify X-rays at the macro level (global level) using statistical moments and SVM classifier, six classes of X-ray images are taken namely chest, foot, spine, neck, head, and palm. Each class consists of 60 images and it is collected from IRMA database. Initially pre-processing is performed by using the M3 filter and its region of interest is found by applying Connected Component Labeling (CCL), feature is extracted by applying statistical moments. The extracted features are used for classification using Support Vector Machines (SVM) which gave an accuracy rate of 92.58%.

**CHAPTER-3**

**Hardware and Software requirements**

**3.1.REQUIREMENT ANALYSIS**

The project involved analyzing the design of few applications so as to make the application more users friendly. To do so, it was really important to keep the navigations from one screen to the other well ordered and at the same time reducing the amount of typing the user needs to do. In order to make the application more accessible, the browser version had to be chosen so that it is compatible with most of the Browsers.

**3.2.REQUIREMENT SPECIFICATION**

**3.2.1.Functional Requirements**

* Graphical User interface with the User.

**3.3.Software Requirements**

For developing the application the following are the Software Requirements:

1. Python
2. tkinter

**3.4.Operating Systems supported**

1. Windows 7
2. Windows XP
3. Windows 8

**3.5.Technologies and Languages used to Develop**

1. Python

**3.6.Debugger and Emulator**

* Any Browser (Particularly Chrome)

**3.7.Hardware Requirements**

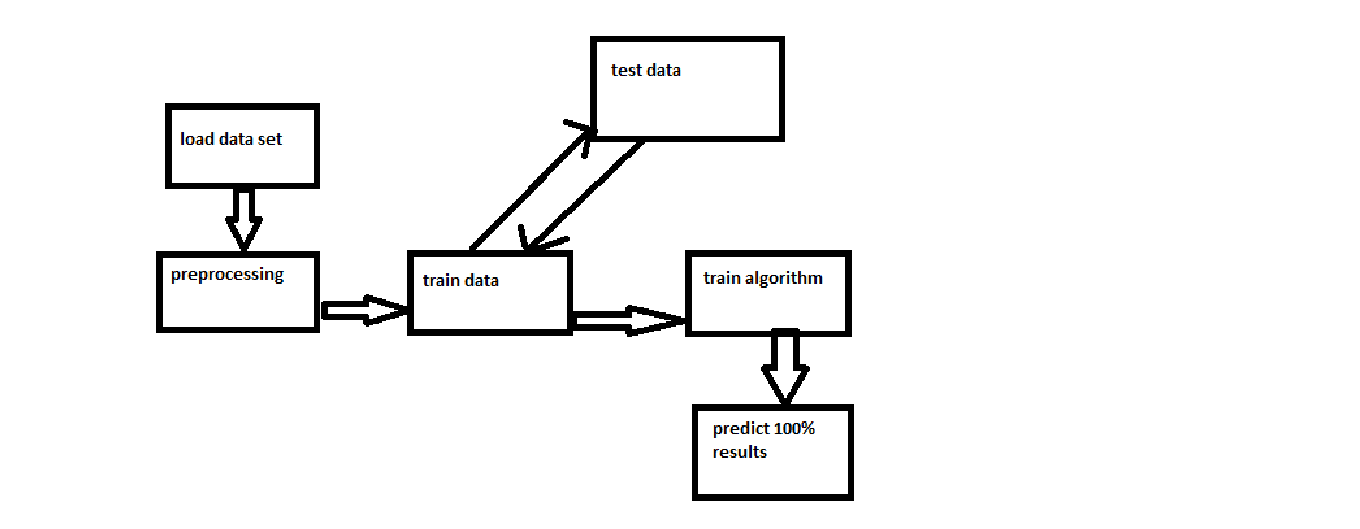
For developing the application the following are the Hardware Requirements:

* Processor: Pentium IV or higher
* RAM: 256 MB
* Space on Hard Disk: minimum 512MB

**CHAPTER-4**

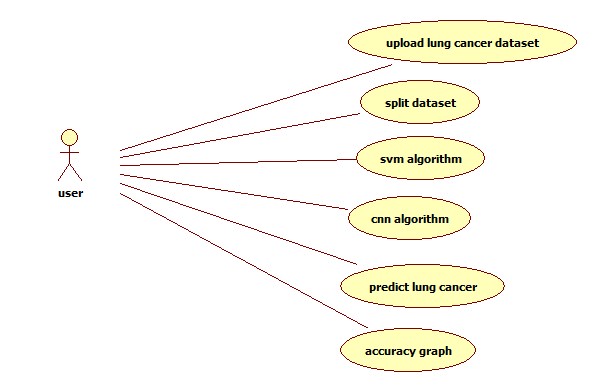
**4.1.Detailed Design**

**4.1.1.System Architecture**

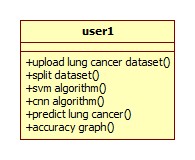


4.2.UML diagrams/ER diagram

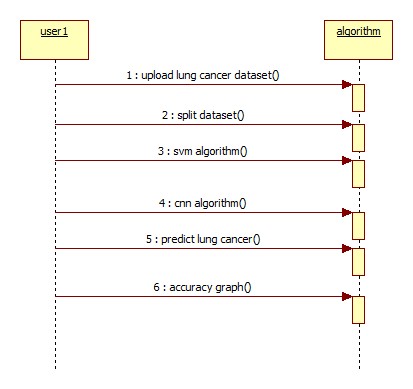
4.2.1.USECASE DIAGRAM:



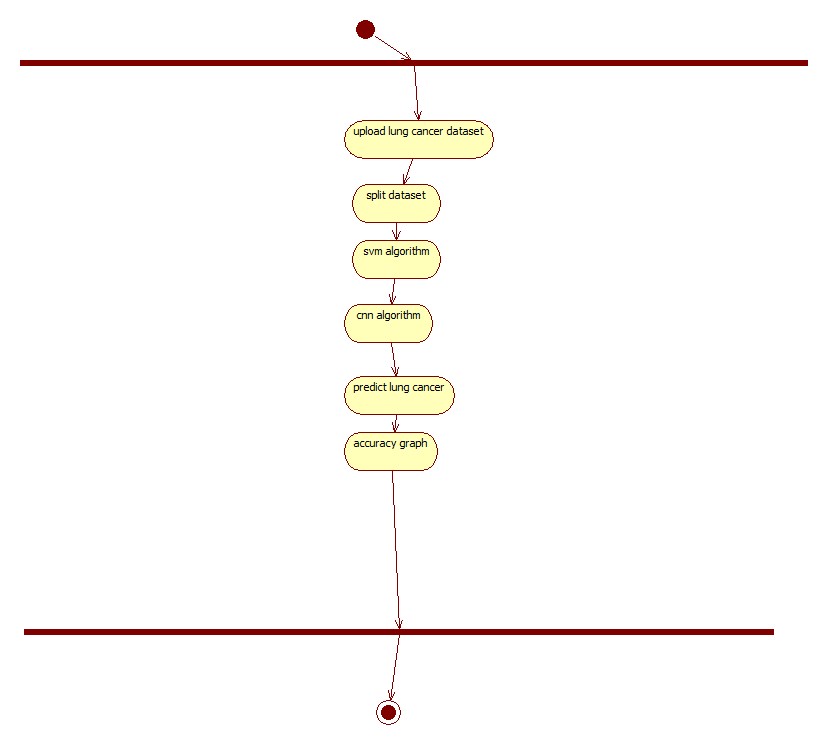
4.2.2.CLASS DIAGRAM:



4.2.3.SEQUENCE DIAGRAM:



4.2.4.ACTIVITY DIAGRAM:



**CHAPTER-5**

**5.1.Module description**

**5.1.1.Read lung cancer dataset:**

We are going to use lung cancer dataset. The dataset contains more than 100 images. It contains normal and abnormal lung disease images. With these images we will train the algorithm and test the algorithm.

**5.1.2.Split dataset:**

After loading the dataset we will split the data set into two parts one is for training the algorithm another one is testing the algorithm.

**5.1.3.SVM algorithm:**

SVM(support vector machine) algorithm is used to train the data. It will train and give accuracy of the algorithm lesser than CNN (Convolutional nueral network).

**5.1.4.CNN algorithm:**

CNN(Convolutional nueral network) is used to train with the data. It will give more accuracy than machine learning algorithms. Approximately 90% above.

**5.1.5.Predict lung cancer:**

In this module we will give input as the image to the algorithm. The algorithm will preduict either it is normal or abnormal.

**5.1.6.Accuracy graph:**

We will show accuracies of the algorithms in the form of graph.

**CHAPTER-6**

**6.1.Database Design/DFDs:**

In this project database is not required.

**CHAPTER-7**

**7.1.Implementation :**

from tkinter import messagebox

from tkinter import \*

from tkinter import simpledialog

import tkinter

from tkinter import filedialog

import matplotlib.pyplot as plt

import numpy as np

from tkinter.filedialog import askopenfilename

import pandas as pd

import os

import cv2

import numpy as np

from sklearn import svm

from sklearn.metrics import accuracy\_score

from sklearn.model\_selection import train\_test\_split

from sklearn.decomposition import PCA

from keras.utils.np\_utils import to\_categorical

from keras.layers import MaxPooling2D

from keras.layers import Dense, Dropout, Activation, Flatten

from keras.layers import Convolution2D

from keras.models import Sequential

main = tkinter.Tk()

main.title("Detection of Lung cancer from CT image using SVM classification and compare the survival rate of patients using 3D Convolutional neural network(3D CNN)on lung nodules data set")

main.geometry("1300x1200")

global filename

global classifier

global svm\_sr, cnn\_sr

global X, Y

global X\_train, X\_test, y\_train, y\_test

global pca

def uploadDataset():

global filename

filename = filedialog.askdirectory(initialdir=".")

text.delete('1.0', END)

text.insert(END,filename+" loaded\n");

def splitDataset():

global X, Y

global X\_train, X\_test, y\_train, y\_test

global pca

text.delete('1.0', END)

X = np.load('features/X.txt.npy')

Y = np.load('features/Y.txt.npy')

X = np.reshape(X, (X.shape[0],(X.shape[1]\*X.shape[2]\*X.shape[3])))

pca = PCA(n\_components = 100)

X = pca.fit\_transform(X)

print(X.shape)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, Y, test\_size=0.2)

text.insert(END,"Total CT Scan Images Found in dataset : "+str(len(X))+"\n")

text.insert(END,"Train split dataset to 80% : "+str(len(X\_train))+"\n")

text.insert(END,"Test split dataset to 20% : "+str(len(X\_test))+"\n")

def executeSVM():

global classifier

global svm\_sr

text.delete('1.0', END)

cls = svm.SVC()

cls.fit(X\_train, y\_train)

predict = cls.predict(X\_test)

svm\_sr = accuracy\_score(y\_test,predict) \* 100

classifier = cls

text.insert(END,"SVM Survival Rate : "+str(svm\_sr)+"\n")

def executeCNN():

global cnn\_sr

X = np.load('features/X.txt.npy')

Y = np.load('features/Y.txt.npy')

Y = to\_categorical(Y)

classifier = Sequential()

classifier.add(Convolution2D(32, 3, 3, input\_shape = (64, 64, 3), activation = 'relu'))

classifier.add(MaxPooling2D(pool\_size = (2, 2)))

classifier.add(Convolution2D(32, 3, 3, activation = 'relu'))

classifier.add(MaxPooling2D(pool\_size = (2, 2)))

classifier.add(Flatten())

classifier.add(Dense(output\_dim = 256, activation = 'relu'))

classifier.add(Dense(output\_dim = 2, activation = 'softmax'))

print(classifier.summary())

classifier.compile(optimizer = 'adam', loss = 'categorical\_crossentropy', metrics = ['accuracy'])

hist = classifier.fit(X, Y, batch\_size=16, epochs=10, shuffle=True, verbose=2)

hist = hist.history

acc = hist['accuracy']

cnn\_sr = acc[9] \* 100

text.insert(END,"CNN Survival Rate : "+str(cnn\_sr)+"\n")

def predictCancer():

filename = filedialog.askopenfilename(initialdir="testSamples")

img = cv2.imread(filename)

img = cv2.resize(img, (64,64))

im2arr = np.array(img)

im2arr = im2arr.reshape(64,64,3)

im2arr = im2arr.astype('float32')

im2arr = im2arr/255

test = []

test.append(im2arr)

test = np.asarray(test)

test = np.reshape(test, (test.shape[0],(test.shape[1]\*test.shape[2]\*test.shape[3])))

test = pca.transform(test)

predict = classifier.predict(test)[0]

msg = ''

if predict == 0:

msg = "Uploaded CT Scan is Normal"

if predict == 1:

msg = "Uploaded CT Scan is Abnormal"

img = cv2.imread(filename)

img = cv2.resize(img, (400,400))

cv2.putText(img, msg, (10, 25), cv2.FONT\_HERSHEY\_SIMPLEX,0.7, (0, 255, 255), 2)

cv2.imshow(msg, img)

cv2.waitKey(0)

def graph():

height = [svm\_sr, cnn\_sr]

bars = ('SVM Survival Rate','CNN Survival Rate')

y\_pos = np.arange(len(bars))

plt.bar(y\_pos, height)

plt.xticks(y\_pos, bars)

plt.show()

font = ('times', 14, 'bold')

title = Label(main, text='Detection of Lung cancer from CT image using SVM classification and compare the survival rate of patients using 3D Convolutional neural network(3D CNN)on lung nodules data set')

title.config(bg='deep sky blue', fg='white')

title.config(font=font)

title.config(height=3, width=120)

title.place(x=0,y=5)

font1 = ('times', 12, 'bold')

text=Text(main,height=20,width=150)

scroll=Scrollbar(text)

text.configure(yscrollcommand=scroll.set)

text.place(x=50,y=120)

text.config(font=font1)

font1 = ('times', 13, 'bold')

uploadButton = Button(main, text="Upload Lung Cancer Dataset", command=uploadDataset)

uploadButton.place(x=50,y=550)

uploadButton.config(font=font1)

readButton = Button(main, text="Read & Split Dataset to Train & Test", command=splitDataset)

readButton.place(x=350,y=550)

readButton.config(font=font1)

svmButton = Button(main, text="Execute SVM Algorithms", command=executeSVM)

svmButton.place(x=50,y=600)

svmButton.config(font=font1)

kmeansButton = Button(main, text="Execute CNN Algorithm", command=executeCNN)

kmeansButton.place(x=350,y=600)

kmeansButton.config(font=font1)

predictButton = Button(main, text="Predict Lung Cancer", command=predictCancer)

predictButton.place(x=50,y=650)

predictButton.config(font=font1)

graphButton = Button(main, text="Survival Rate Graph", command=graph)

graphButton.place(x=350,y=650)

graphButton.config(font=font1)

main.config(bg='LightSteelBlue3')

main.mainloop()

from tkinter import messagebox

from tkinter import \*

from tkinter import simpledialog

import tkinter

from tkinter import filedialog

import matplotlib.pyplot as plt

import numpy as np

from tkinter.filedialog import askopenfilename

import pandas as pd

import os

import cv2

import numpy as np

from sklearn import svm

from sklearn.metrics import accuracy\_score

from sklearn.model\_selection import train\_test\_split

from sklearn.decomposition import PCA

from keras.utils.np\_utils import to\_categorical

from keras.layers import MaxPooling2D

from keras.layers import Dense, Dropout, Activation, Flatten

from keras.layers import Convolution2D

from keras.models import Sequential

main = tkinter.Tk()

main.title("Prediction of time-to-event outcomes in diagnosing lung cancer based on SVM and compare the accuracy of predicted outcome with Deep CNN algorithm")

main.geometry("1300x1200")

global filename

global classifier

global svm\_acc, cnn\_acc

global X, Y

global X\_train, X\_test, y\_train, y\_test

global pca

def uploadDataset():

global filename

filename = filedialog.askdirectory(initialdir=".")

text.delete('1.0', END)

text.insert(END,filename+" loaded\n");

def splitDataset():

global X, Y

global X\_train, X\_test, y\_train, y\_test

global pca

text.delete('1.0', END)

X = np.load('features/X.txt.npy')

Y = np.load('features/Y.txt.npy')

X = np.reshape(X, (X.shape[0],(X.shape[1]\*X.shape[2]\*X.shape[3])))

pca = PCA(n\_components = 100)

X = pca.fit\_transform(X)

print(X.shape)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, Y, test\_size=0.2)

text.insert(END,"Total CT Scan Images Found in dataset : "+str(len(X))+"\n")

text.insert(END,"Train split dataset to 80% : "+str(len(X\_train))+"\n")

text.insert(END,"Test split dataset to 20% : "+str(len(X\_test))+"\n")

def executeSVM():

global classifier

global svm\_acc

text.delete('1.0', END)

cls = svm.SVC()

cls.fit(X\_train, y\_train)

predict = cls.predict(X\_test)

svm\_acc = accuracy\_score(y\_test,predict) \* 100

classifier = cls

text.insert(END,"SVM Accuracy : "+str(svm\_acc)+"\n")

def executeCNN():

global cnn\_acc

X = np.load('features/X.txt.npy')

Y = np.load('features/Y.txt.npy')

Y = to\_categorical(Y)

classifier = Sequential()

classifier.add(Convolution2D(32, 3, 3, input\_shape = (64, 64, 3), activation = 'relu'))

classifier.add(MaxPooling2D(pool\_size = (2, 2)))

classifier.add(Convolution2D(32, 3, 3, activation = 'relu'))

classifier.add(MaxPooling2D(pool\_size = (2, 2)))

classifier.add(Flatten())

classifier.add(Dense(output\_dim = 256, activation = 'relu'))

classifier.add(Dense(output\_dim = 2, activation = 'softmax'))

print(classifier.summary())

classifier.compile(optimizer = 'adam', loss = 'categorical\_crossentropy', metrics = ['accuracy'])

hist = classifier.fit(X, Y, batch\_size=16, epochs=12, shuffle=True, verbose=2)

hist = hist.history

acc = hist['accuracy']

cnn\_acc = acc[9] \* 100

text.insert(END,"CNN Accuracy : "+str(cnn\_acc)+"\n")

def predictCancer():

filename = filedialog.askopenfilename(initialdir="testSamples")

img = cv2.imread(filename)

img = cv2.resize(img, (64,64))

im2arr = np.array(img)

im2arr = im2arr.reshape(64,64,3)

im2arr = im2arr.astype('float32')

im2arr = im2arr/255

test = []

test.append(im2arr)

test = np.asarray(test)

test = np.reshape(test, (test.shape[0],(test.shape[1]\*test.shape[2]\*test.shape[3])))

test = pca.transform(test)

predict = classifier.predict(test)[0]

msg = ''

if predict == 0:

msg = "Uploaded CT Scan is Normal"

if predict == 1:

msg = "Uploaded CT Scan is Abnormal"

img = cv2.imread(filename)

img = cv2.resize(img, (400,400))

cv2.putText(img, msg, (10, 25), cv2.FONT\_HERSHEY\_SIMPLEX,0.7, (0, 255, 255), 2)

cv2.imshow(msg, img)

cv2.waitKey(0)

def graph():

height = [svm\_acc, cnn\_acc]

bars = ('SVM Accuracy','CNN Accuracy')

y\_pos = np.arange(len(bars))

plt.bar(y\_pos, height)

plt.xticks(y\_pos, bars)

plt.show()

font = ('times', 14, 'bold')

title = Label(main, text='Prediction of time-to-event outcomes in diagnosing lung cancer based on SVM and compare the accuracy of predicted outcome with Deep CNN algorithm')

title.config(bg='deep sky blue', fg='white')

title.config(font=font)

title.config(height=3, width=120)

title.place(x=0,y=5)

font1 = ('times', 12, 'bold')

text=Text(main,height=20,width=150)

scroll=Scrollbar(text)

text.configure(yscrollcommand=scroll.set)

text.place(x=50,y=120)

text.config(font=font1)

font1 = ('times', 13, 'bold')

uploadButton = Button(main, text="Upload Lung Cancer Dataset", command=uploadDataset)

uploadButton.place(x=50,y=550)

uploadButton.config(font=font1)

readButton = Button(main, text="Read & Split Dataset to Train & Test", command=splitDataset)

readButton.place(x=350,y=550)

readButton.config(font=font1)

svmButton = Button(main, text="Execute SVM Accuracy Algorithms", command=executeSVM)

svmButton.place(x=50,y=600)

svmButton.config(font=font1)

kmeansButton = Button(main, text="Execute CNN Accuracy Algorithm", command=executeCNN)

kmeansButton.place(x=350,y=600)

kmeansButton.config(font=font1)

predictButton = Button(main, text="Predict Lung Cancer", command=predictCancer)

predictButton.place(x=50,y=650)

predictButton.config(font=font1)

graphButton = Button(main, text="Accuracy Graph", command=graph)

graphButton.place(x=350,y=650)

graphButton.config(font=font1)

main.config(bg='LightSteelBlue3')

main.mainloop()

**CHAPTER-8**

**SOFTWARE ENVIRONMENT:**

**Python Introduction**

**Python** is a general purpose, dynamic, high level, and interpreted programming language. It supports Object Oriented programming approach to develop applications. It is simple and easy to learn and provides lots of high-level data structures.

Python is *easy to learn* yet powerful and versatile scripting language, which makes it attractive for Application Development.

Python's syntax and *dynamic typing* with its interpreted nature make it an ideal language for scripting and rapid application development.

Python supports *multiple programming pattern*, including object-oriented, imperative, and functional or procedural programming styles.

Python is not intended to work in a particular area, such as web programming. That is why it is known as *multipurpose* programming language because it can be used with web, enterprise, 3D CAD, etc.

We don't need to use data types to declare variable because it is *dynamically typed* so we can write a=10 to assign an integer value in an integer variable.

Python makes the development and debugging *fast* because there is no compilation step included in Python development, and edit-test-debug cycle is very fast.

Python 2 vs. Python 3

In most of the programming languages, whenever a new version releases, it supports the features and syntax of the existing version of the language, therefore, it is easier for the projects to switch in the newer version. However, in the case of Python, the two versions Python 2 and Python 3 are very much different from each other.

**A list of differences between Python 2 and Python 3 are given below:**

1. Python 2 uses **print** as a statement and used as print "something" to print some string on the console. On the other hand, Python 3 uses **print** as a function and used as print("something") to print something on the console.
2. Python 2 uses the function raw\_input() to accept the user's input. It returns the string representing the value, which is typed by the user. To convert it into the integer, we need to use the int() function in Python. On the other hand, Python 3 uses input() function which automatically interpreted the type of input entered by the user. However, we can cast this value to any type by using primitive functions (int(), str(), etc.).
3. In Python 2, the implicit string type is ASCII, whereas, in Python 3, the implicit string type is Unicode.
4. Python 3 doesn't contain the xrange() function of Python 2. The xrange() is the variant of range() function which returns a xrange object that works similar to Java iterator. The range() returns a list for example the function range(0,3) contains 0, 1, 2.
5. There is also a small change made in Exception handling in Python 3. It defines a keyword **as** which is necessary to be used. We will discuss it in Exception handling section of Python programming tutorial.

# Python Features

Python provides lots of features that are listed below.

#### 1) Easy to Learn and Use

Python is easy to learn and use. It is developer-friendly and high level programming language.

#### 2) Expressive Language

Python language is more expressive means that it is more understandable and readable.

#### 3) Interpreted Language

Python is an interpreted language i.e. interpreter executes the code line by line at a time. This makes debugging easy and thus suitable for beginners.

#### 4) Cross-platform Language

Python can run equally on different platforms such as Windows, Linux, Unix and Macintosh etc. So, we can say that Python is a portable language.

#### 5) Free and Open Source

Python language is freely available at [offical web address](https://www.python.org/" \t "blank).The source-code is also available. Therefore it is open source.

#### 6) Object-Oriented Language

Python supports object oriented language and concepts of classes and objects come into existence.

#### 7) Extensible

It implies that other languages such as C/C++ can be used to compile the code and thus it can be used further in our python code.

#### 8) Large Standard Library

Python has a large and broad library and prvides rich set of module and functions for rapid application development.

#### 9) GUI Programming Support

Graphical user interfaces can be developed using Python.

#### 10) Integrated

It can be easily integrated with languages like C, C++, JAVA etc.

# Python History and Versions

* Python laid its foundation in the late 1980s.
* The implementation of Python was started in the December 1989 by **Guido Van Rossum** at CWI in Netherland.
* In February 1991, van Rossum published the code (labeled version 0.9.0) to alt.sources.
* In 1994, Python 1.0 was released with new features like: lambda, map, filter, and reduce.
* Python 2.0 added new features like: list comprehensions, garbage collection system.
* On December 3, 2008, Python 3.0 (also called "Py3K") was released. It was designed to rectify fundamental flaw of the language.
* *ABC programming language* is said to be the predecessor of Python language which was capable of Exception Handling and interfacing with Amoeba Operating System.
* Python is influenced by following programming languages:
  + ABC language.
  + Modula.

Python programming language is being updated regularly with new features and supports. There are lots of updations in python versions, started from 1994 to current release.

# Python Applications

Python is known for its general purpose nature that makes it applicable in almost each domain of software development. Python as a whole can be used in any sphere of development.

Here, we are specifing applications areas where python can be applied.

#### 1) Web Applications

We can use Python to develop web applications. It provides libraries to handle internet protocols such as HTML and XML, JSON, Email processing, request, beautifulSoup, Feedparser etc. It also provides Frameworks such as Django, Pyramid, Flask etc to design and delelopweb based applications. Some important developments are: PythonWikiEngines, Pocoo, PythonBlogSoftware etc.

#### 2) Desktop GUI Applications

Python provides Tk GUI library to develop user interface in python based application. Some other useful toolkits wxWidgets, Kivy, pyqt that are useable on several platforms. The Kivy is popular for writing multitouch applications.

#### 3) Software Development

Python is helpful for software development process. It works as a support language and can be used for build control and management, testing etc.

#### 4) Scientific and Numeric

Python is popular and widely used in scientific and numeric computing. Some useful library and package are SciPy, Pandas, IPython etc. SciPy is group of packages of engineering, science and mathematics.

#### 5) Business Applications

Python is used to build Bussiness applications like ERP and e-commerce systems. Tryton is a high level application platform.

#### 6) Console Based Application

We can use Python to develop console based applications. For example: **IPython**.

#### 7) Audio or Video based Applications

Python is awesome to perform multiple tasks and can be used to develop multimedia applications. Some of real applications are: TimPlayer, cplay etc.

#### 8) 3D CAD Applications

To create CAD application Fandango is a real application which provides full features of CAD.

#### 9) Enterprise Applications

Python can be used to create applications which can be used within an Enterprise or an Organization. Some real time applications are: OpenErp, Tryton, Picalo etc.

#### 10) Applications for Images

Using Python several application can be developed for image. Applications developed are: VPython, Gogh, imgSeek etc.

There are several such applications which can be developed using Python

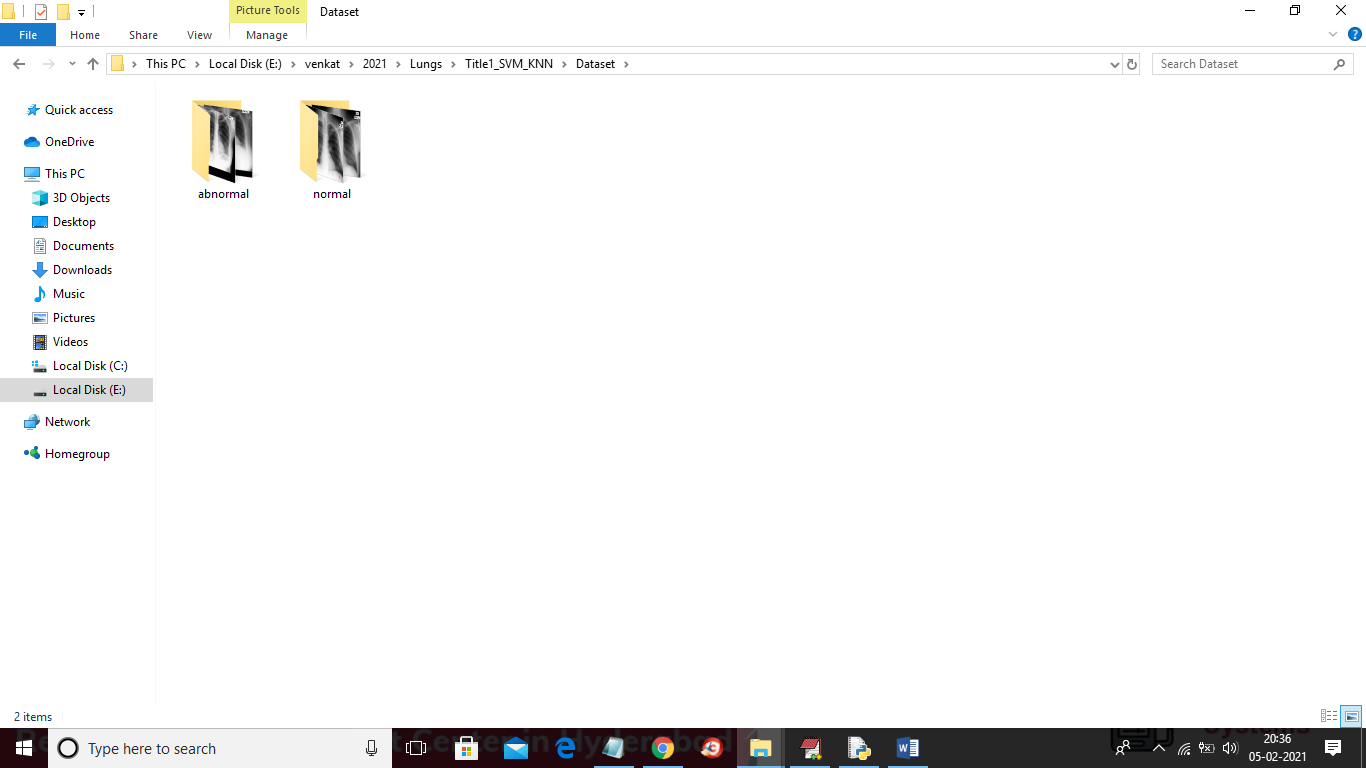
**CHAPTER-9**

**SCREEN SHOTS:**

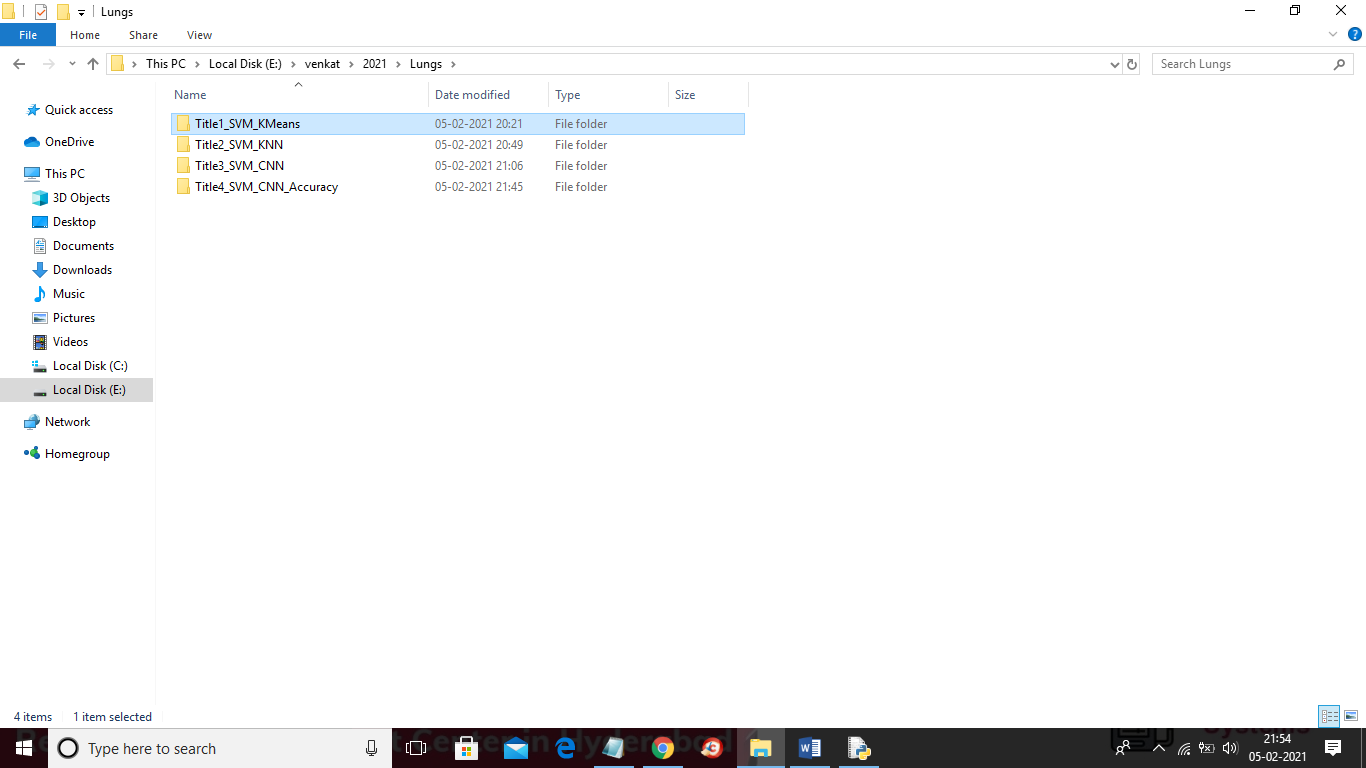
Title 1: output screens

Classification of Lung Cancer Nodules to Monitor Patients Health using Neural Network topology with SVM algorithm & Compare with K-Means Accuracy

In this project we are using CT Scan Lung Cancer Nodules dataset to predict patient health using SVM and KMeans algorithm and then comparing prediction accuracy between them. To implement this project we are using lung cancer images dataset and below screen showing dataset details and this dataset saved inside ‘Dataset’ folder.

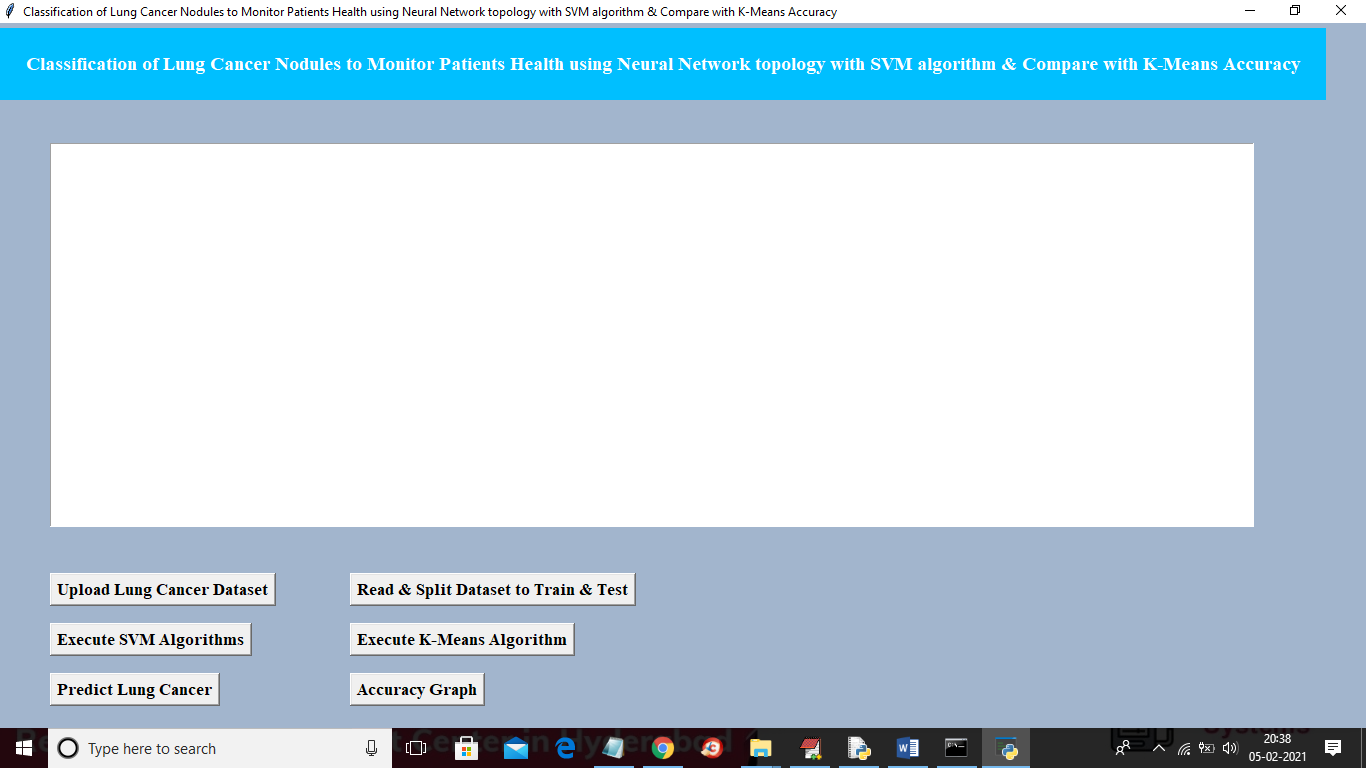


In above screen in dataset we have two types of images such as normal and abnormal and then SVM and KMEANS will get train on above dataset and when we upload new image then SVM will predict whether new image is normal or abnormal. To implement 4 titles we created 4 folders to separate algorithms for each title and you can run one by one

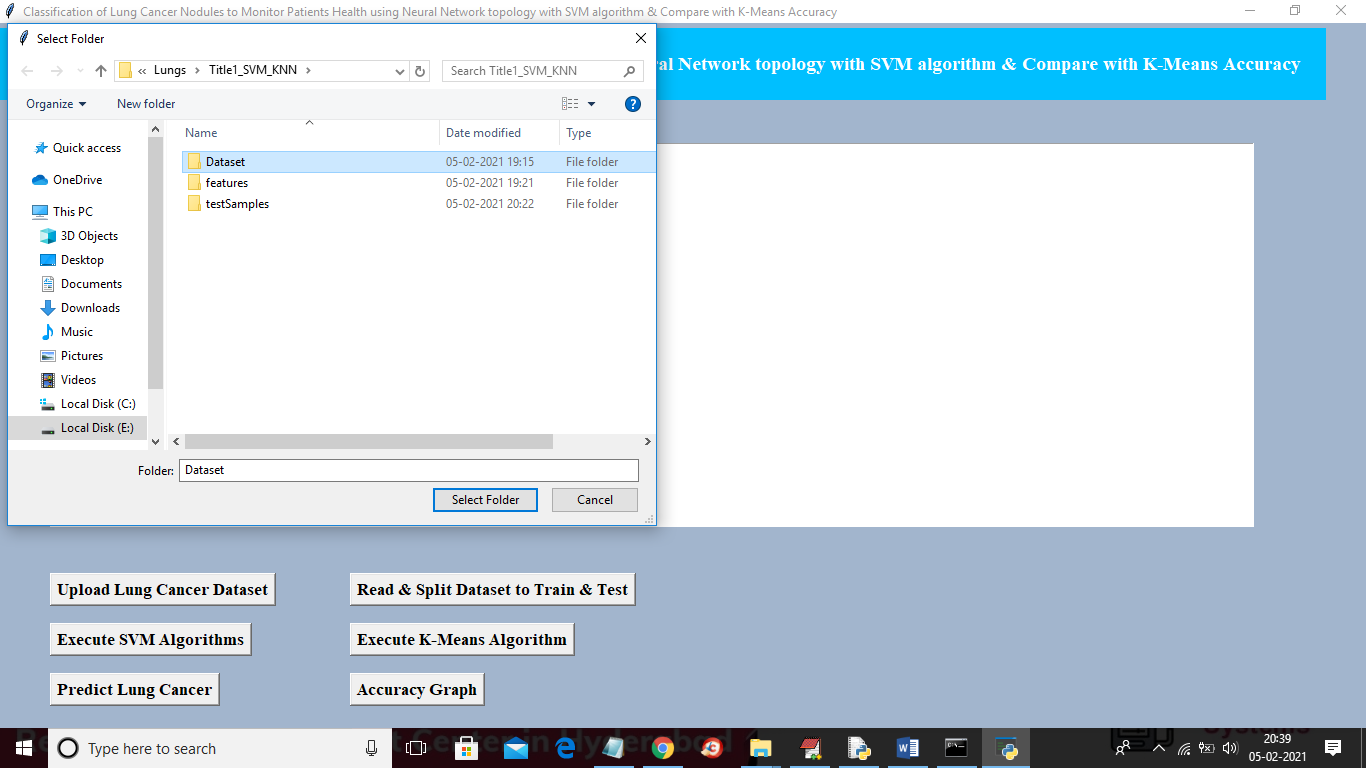


SCREENS SHOT

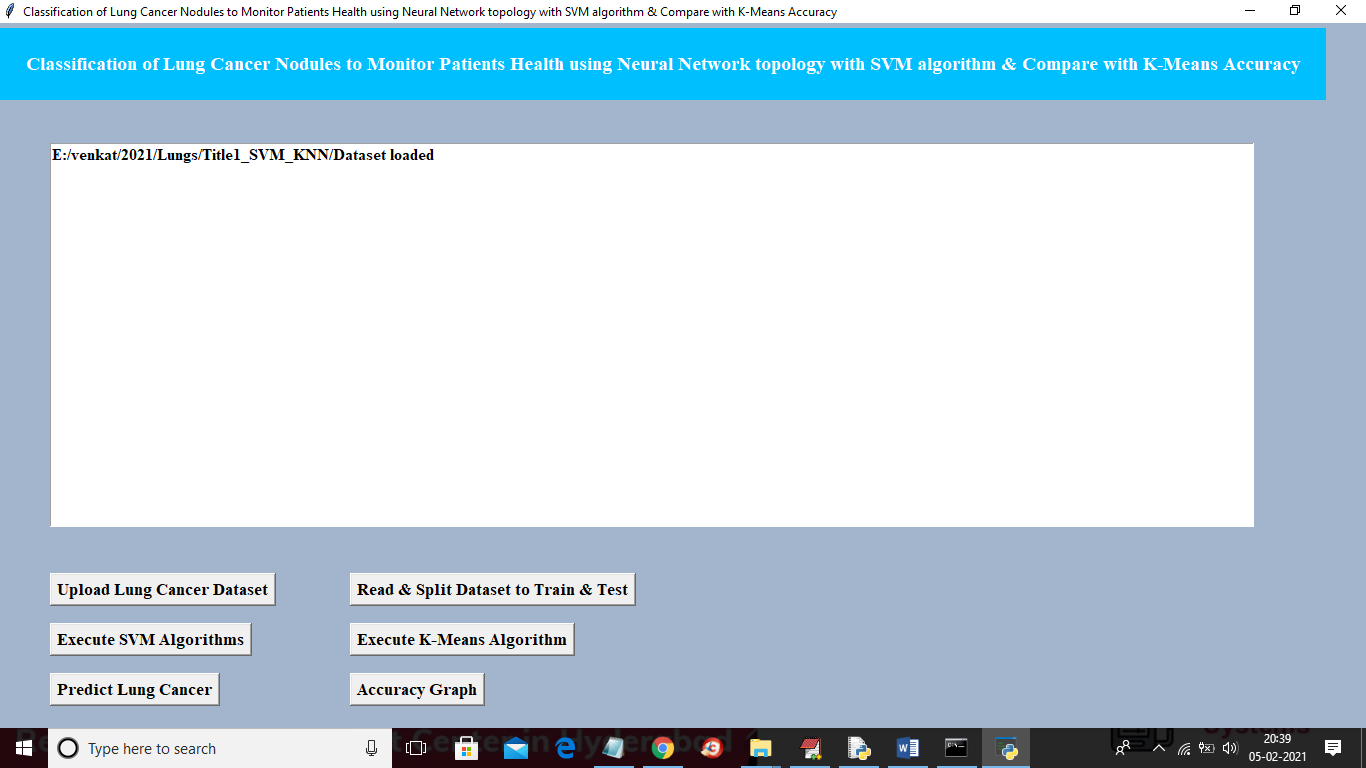
To run project double click on run.bat file from ‘Title1\_SVM\_KMeans’ folder to get below screen

le

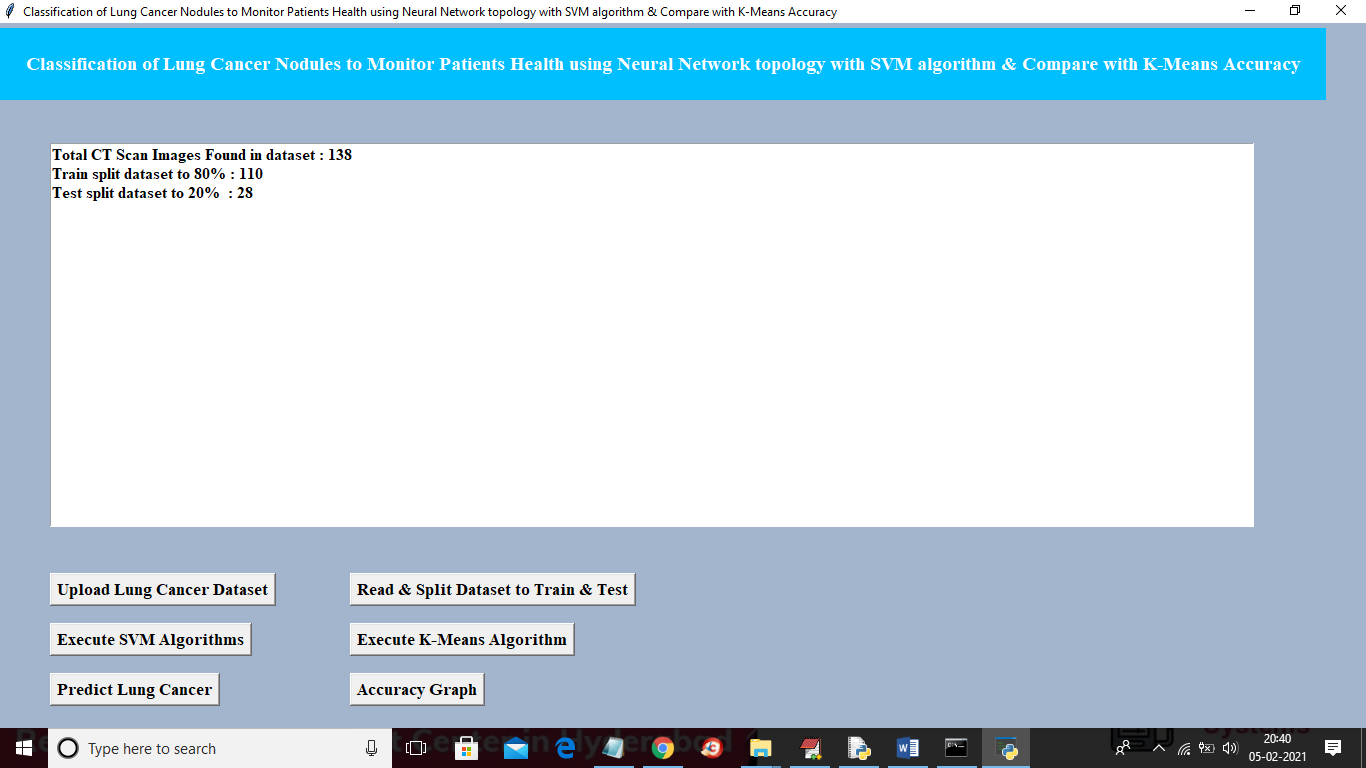
In above screen click on ‘Upload Lung Cancer Dataset’ button and then upload dataset folder



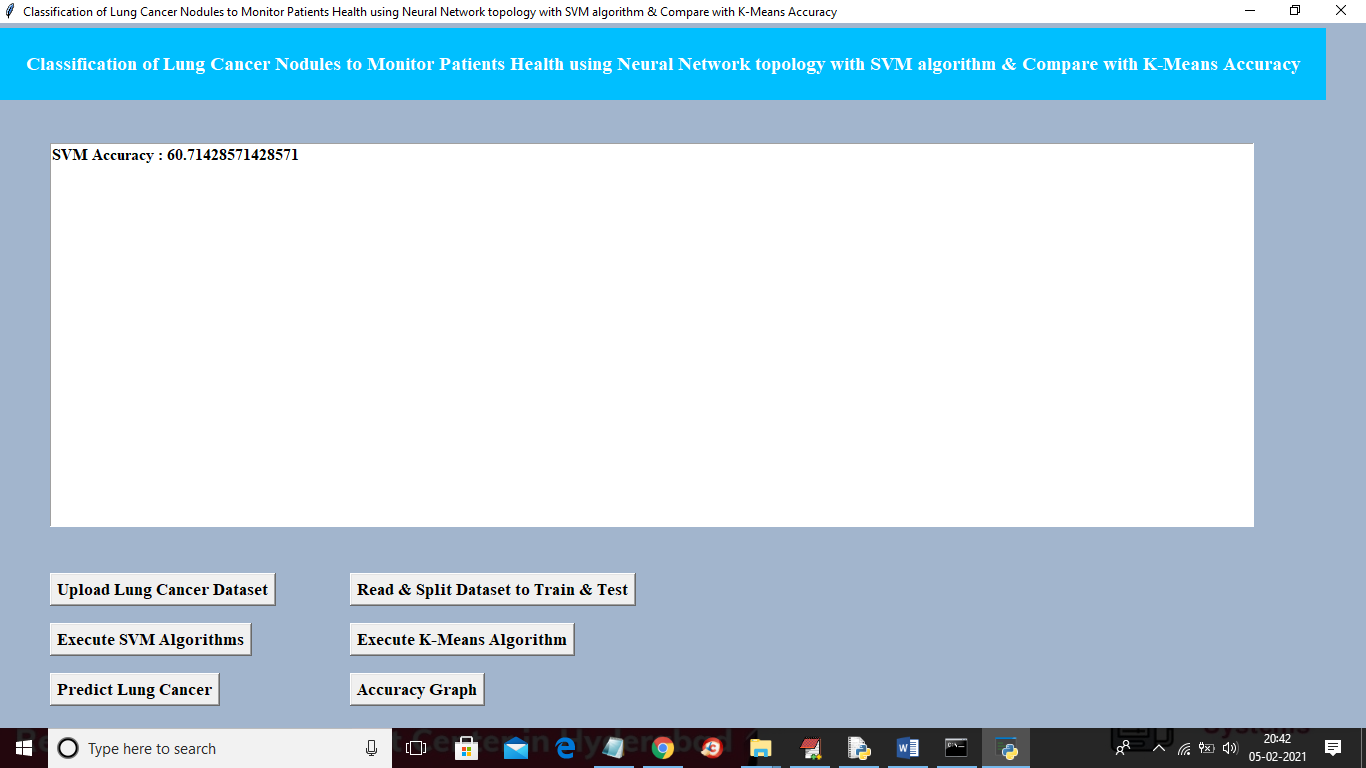
In above screen selecting and uploading ‘Dataset’ folder and then click on ‘Select Folder’ button to load dataset and to get below screen



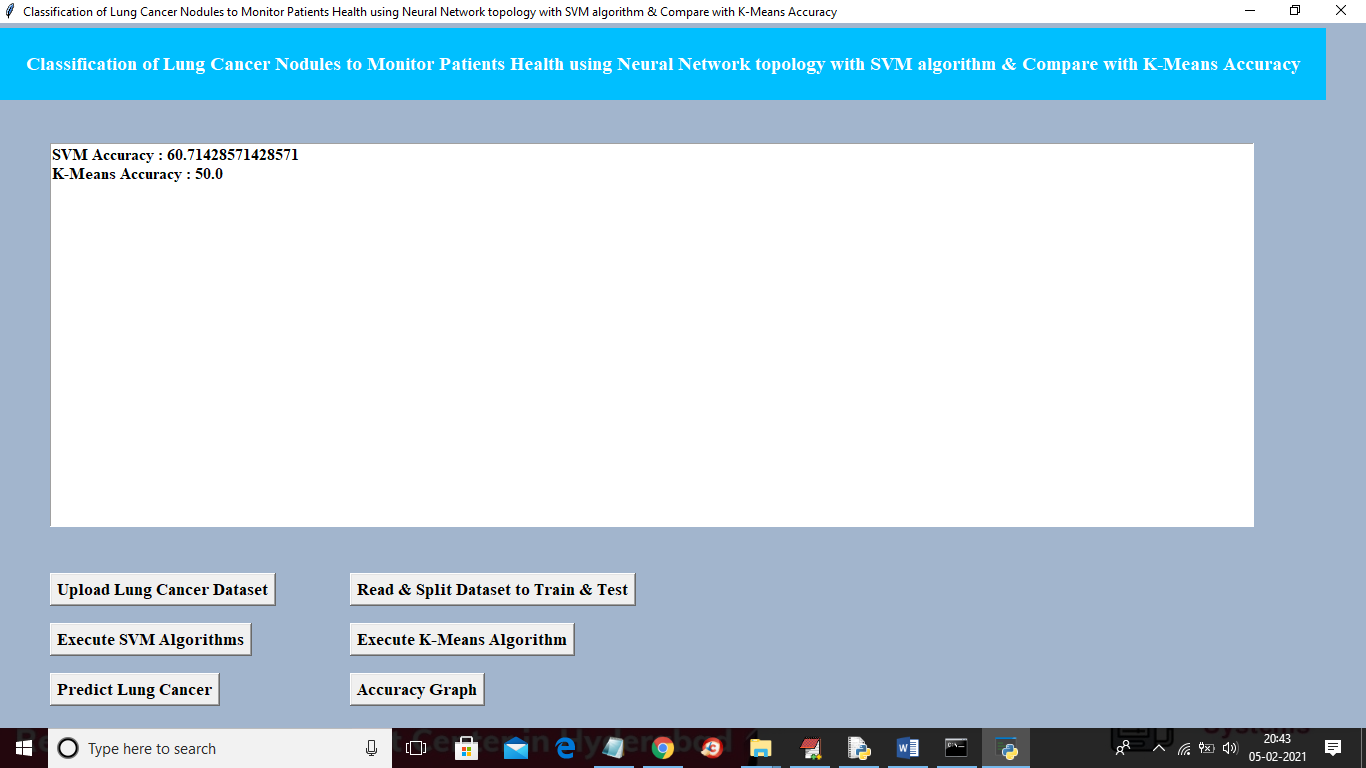
In above screen dataset loaded and now click on ‘Read & Split Dataset to Train & Test’ button to split dataset into train and test parts and application split 80% dataset for training and 20% dataset to test trained model



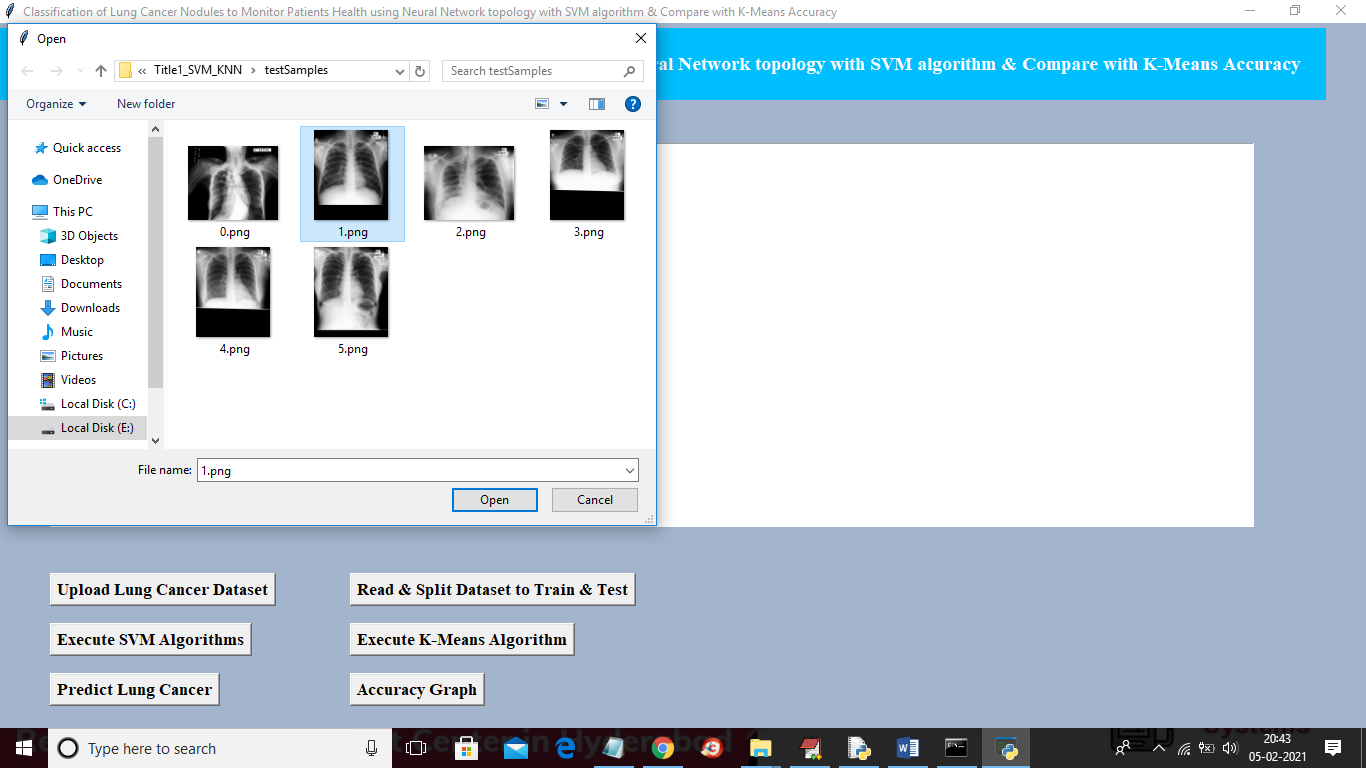
In above screen we can see dataset contains total 138 images and then application using 110 images for training and 28 images for testing and now data is ready and now click on ‘Execute SVM Algorithm’ button to run SVM on loaded dataset and to get below accuracy



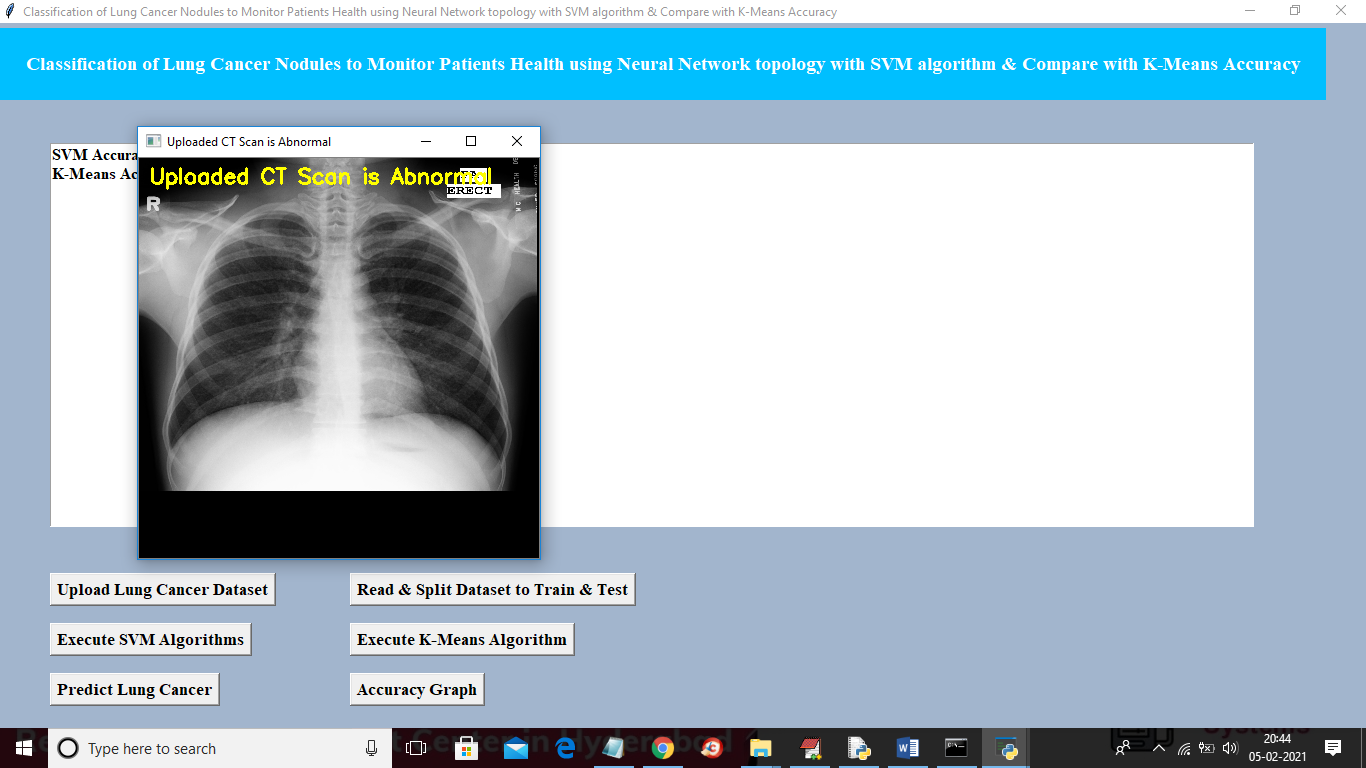
In above screen SVM accuracy is 60% and now click on “Execute K-Means Algorithm” button to run KMEANS algorithm on loaded dataset and to get below screen



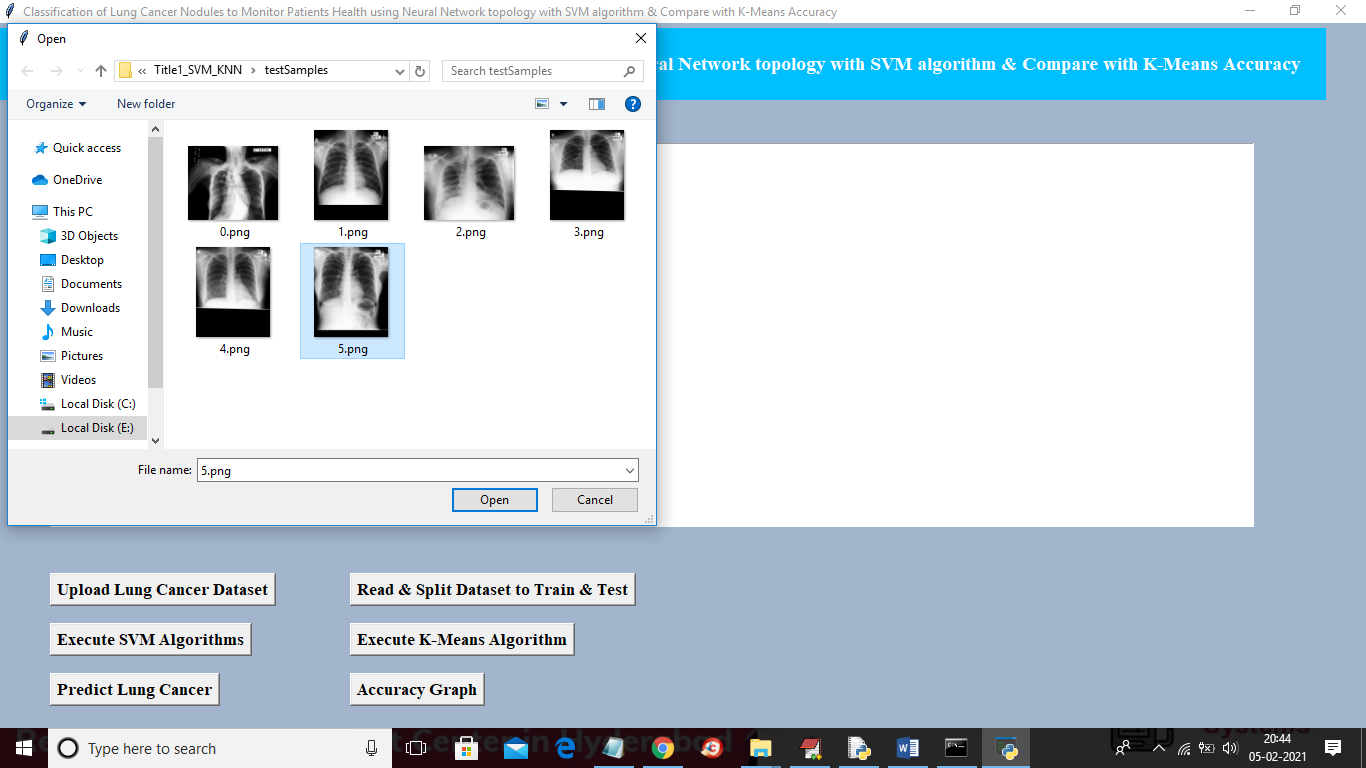
In above screen KMEANS got 50% accuracy and now click on ‘Predict Lung Cancer’ button to upload new test image and then application will give prediction result



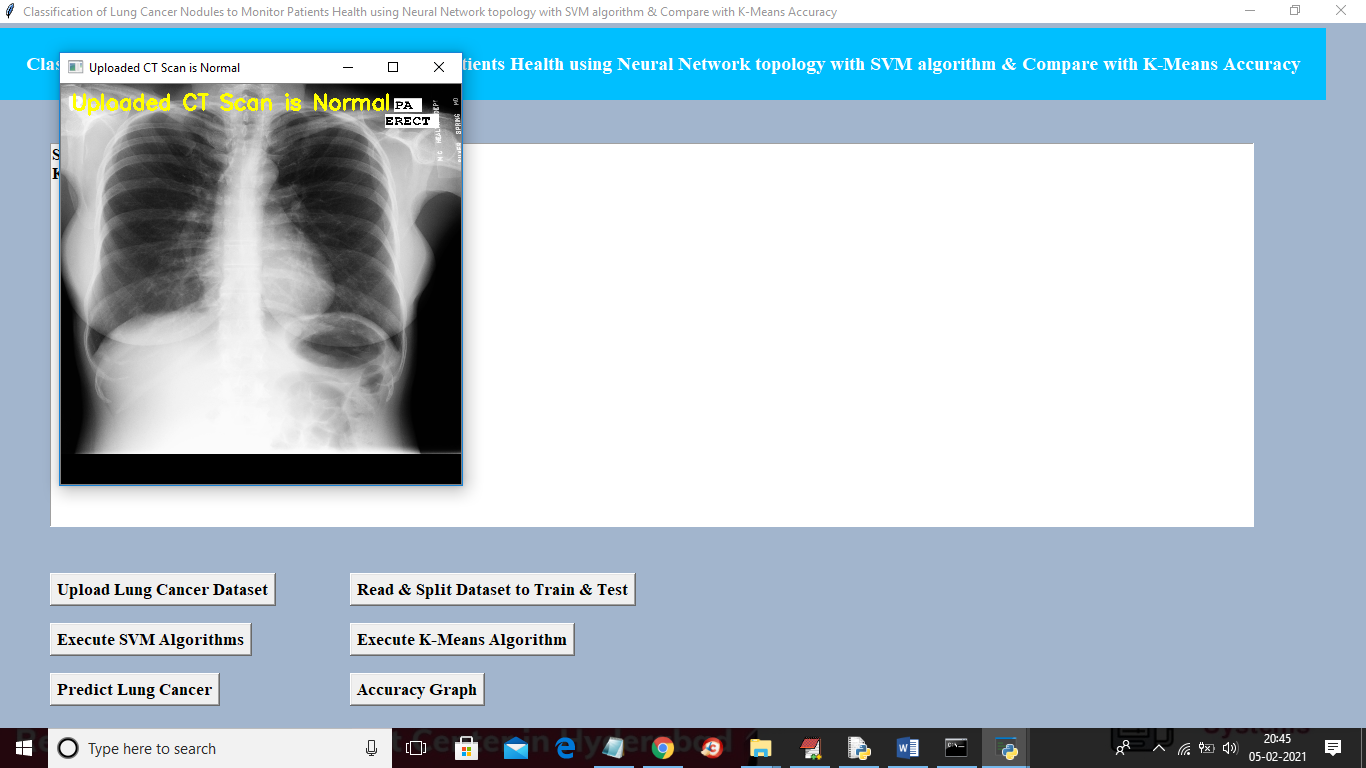
In above screen selecting and uploading ‘1.png’ file and then click on ‘Open’ button to get below result



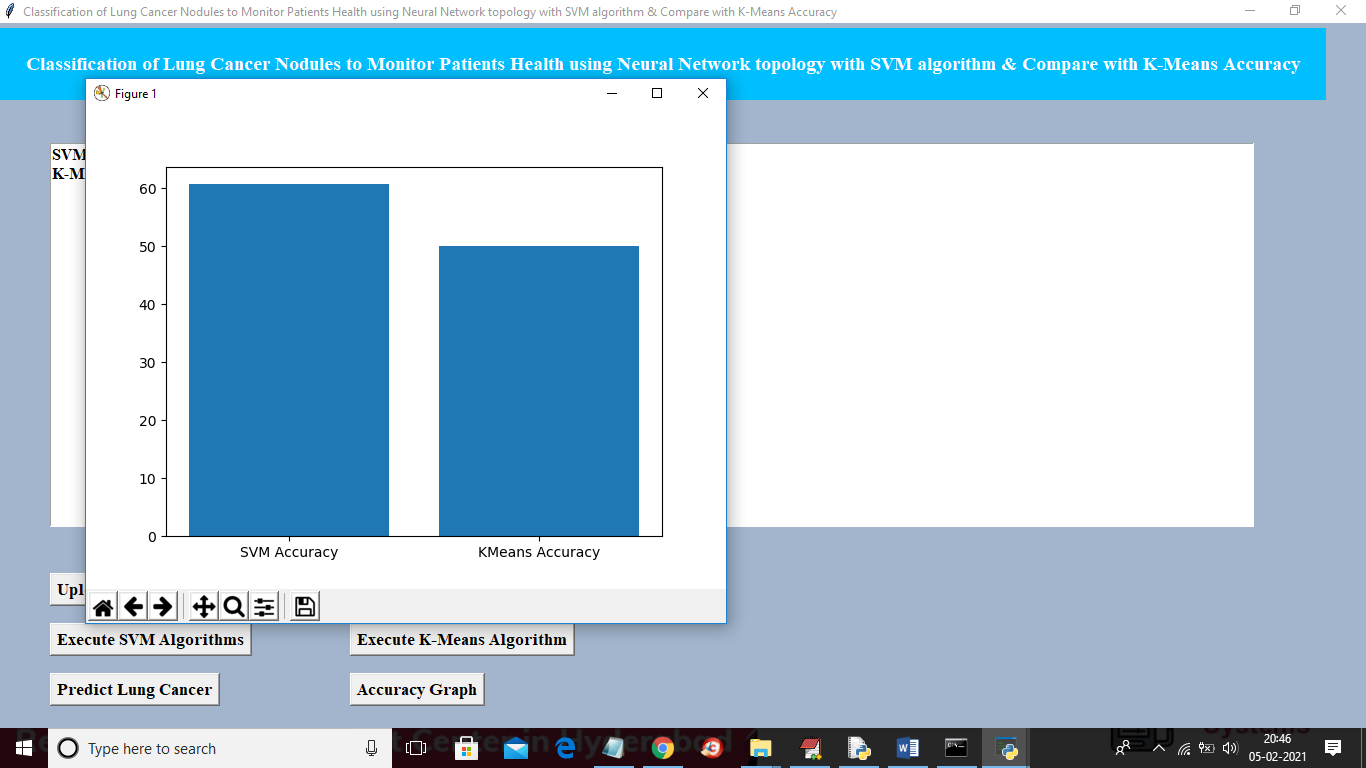
In above screen uploaded image predicted as Abnormal and now test with another image



In above screen uploading ‘5.png’ and below is the result



Above image predicted as Normal and similarly you can upload any image and perform prediction and now click on ‘Accuracy Graph’ button to get below graph



In above screen x-axis represents algorithm name and y-axis represents accuracy of those algorithms and from above graph we can conclude that SVM is better than KMEANS in prediction.

TITLE 2

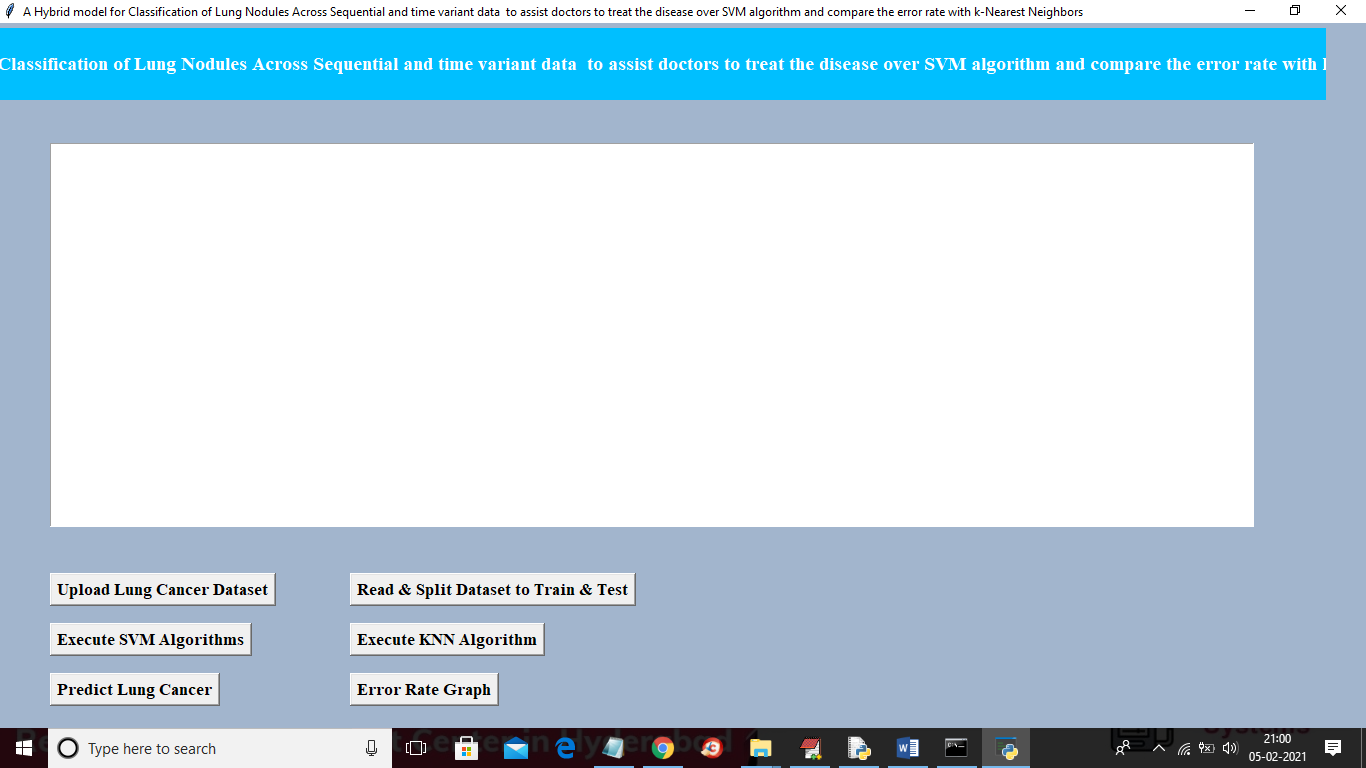
**A Hybrid model for Classification of Lung Nodules Across Sequential and time variant data to assist doctors to treat the disease over SVM algorithm and compare the error rate with k-Nearest Neighbours**

In this project we are using same above dataset to train SVM and KNN algorithm and then calculating error rate between this 2 algorithms and this error rate refers to wrong classification percentage. For example if application predicted 18 records correctly out of 20 records then error rate

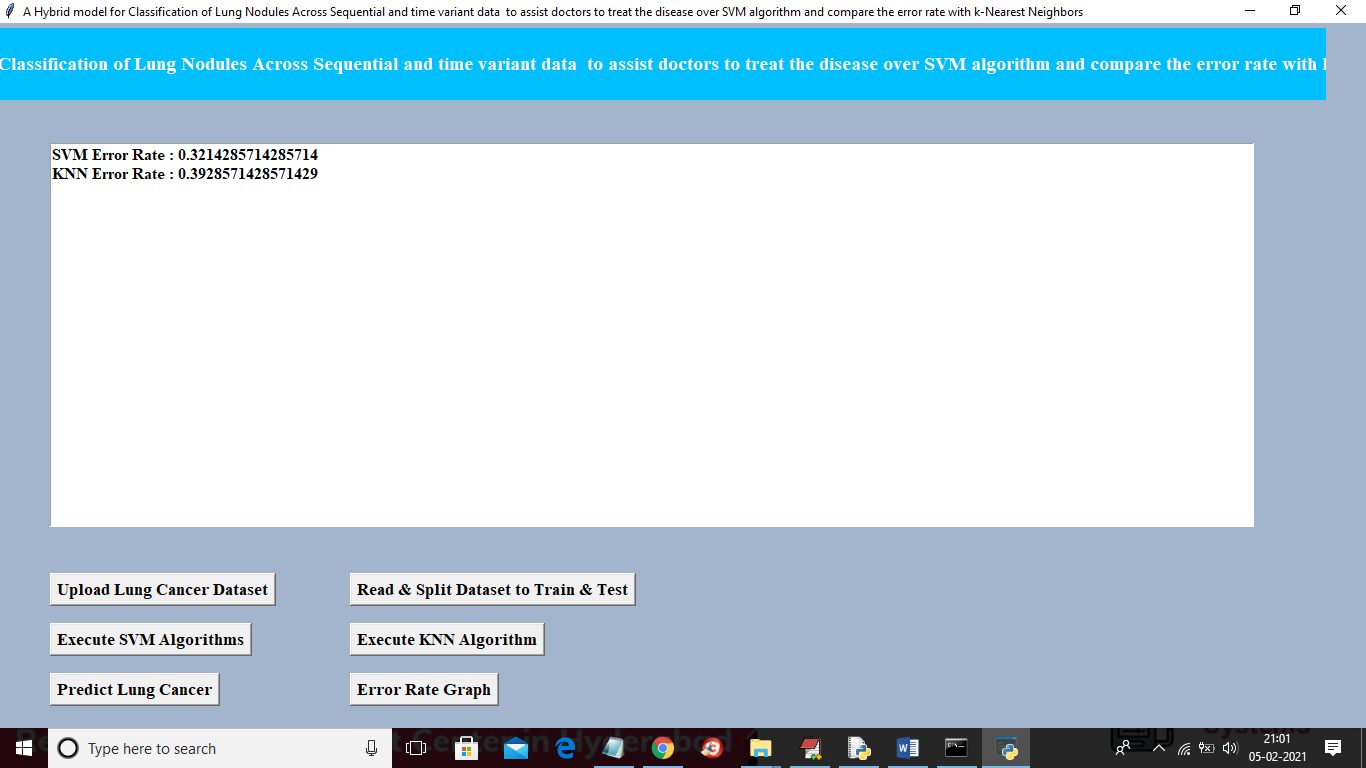
SCREEN SHOTS

will be (1 – (18/20) = 0.1

To run project double click on ‘run.bat’ file from ‘Title2\_SVM\_KNN’ folder to get below screen



In above screen upload dataset and then read dataset and then execute SVM and KNN and then you can predict and calculate error rate



In above screen we can see SVM error rate is 0.32% and KNN error rate is 0.39 and similarly like first project screen shots u can run prediction and graph button. In above screen datasets will be splitted randomly so for every run train and test data may change due to random selection so accuracy or error rate may vary

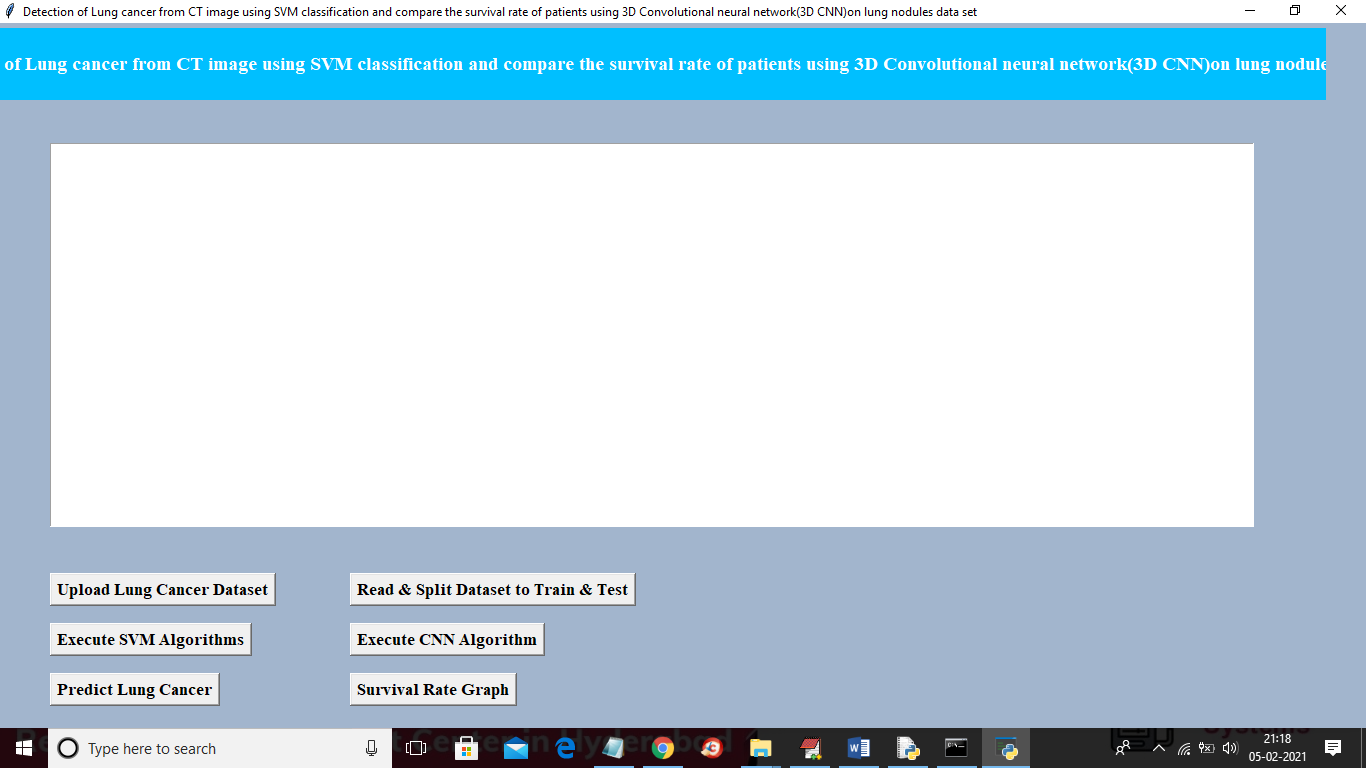
Title 3

**Detection of Lung cancer from CT image using SVM classification and compare the survival rate of patients using 3D Convolutional neural network(3D CNN)on lung nodules data set**

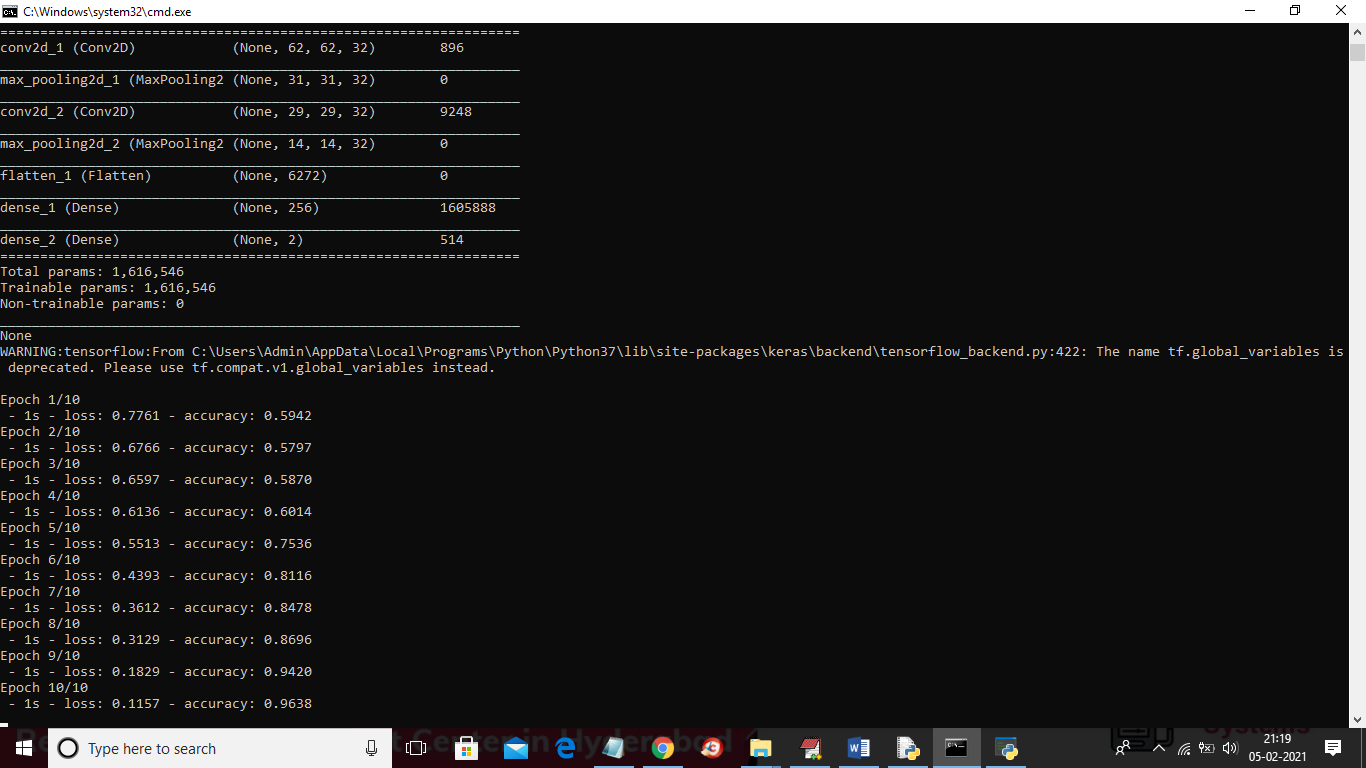
In this project we are using same above Lungs dataset to train CNN and SVM algorithm and then calculate survival rate of patients by using both algorithms. If algorithm predicted 18 records correctly out of 20 records then survival rate will be (18/20 \* 100) = 90%.

SCREEN SHOTS

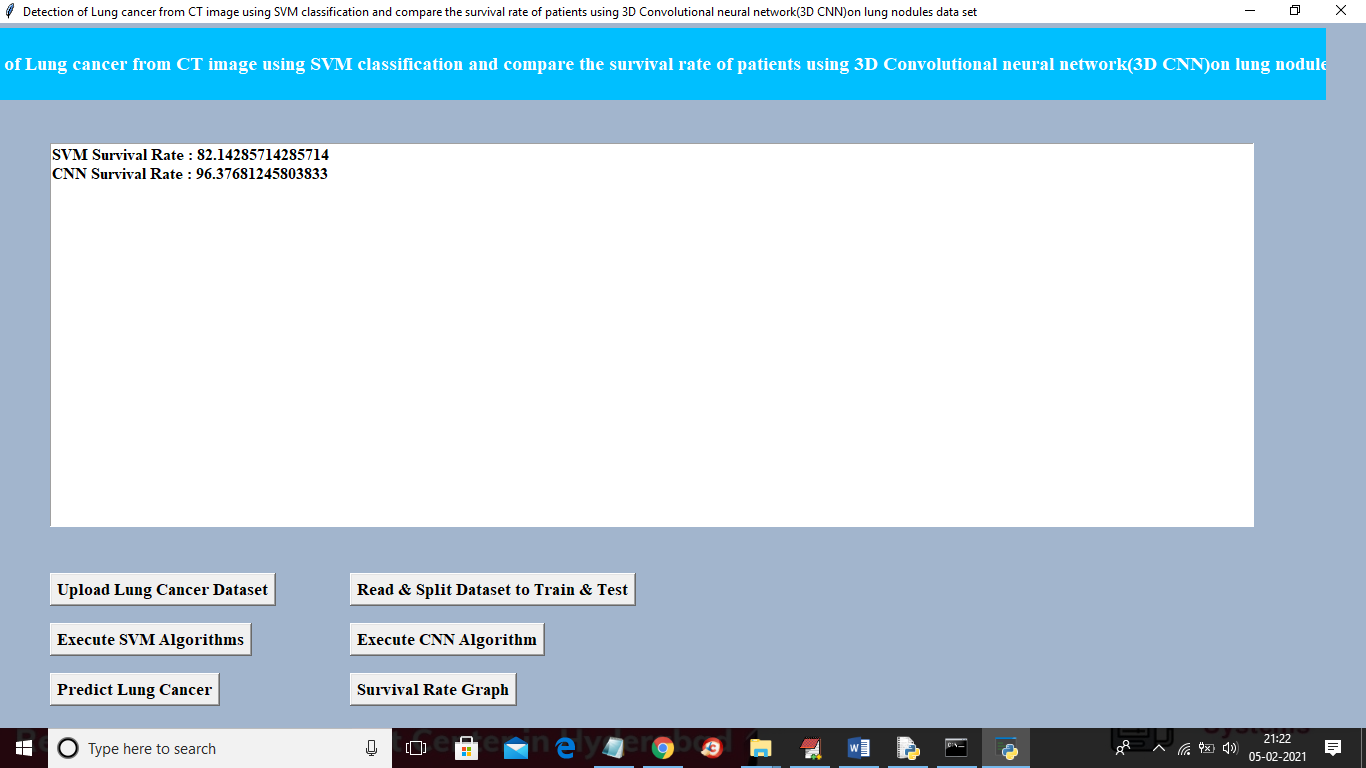
To run project double click on ‘run.bat’ file from ‘Title3\_SVM\_CNN’ folder to get below screen



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.

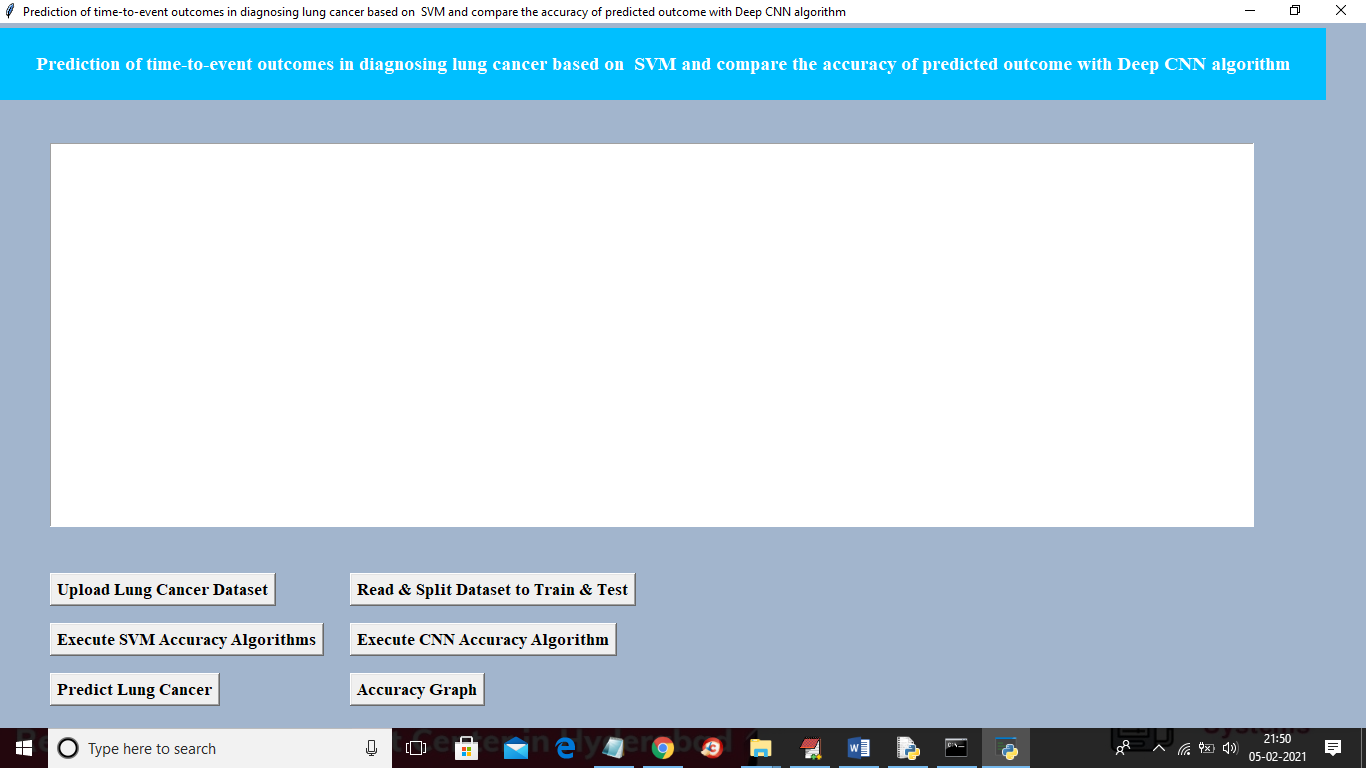
Title 4

**Prediction of time-to-event outcomes in diagnosing lung cancer based on SVM and compare the accuracy of predicted outcome with Deep CNN algorithm**

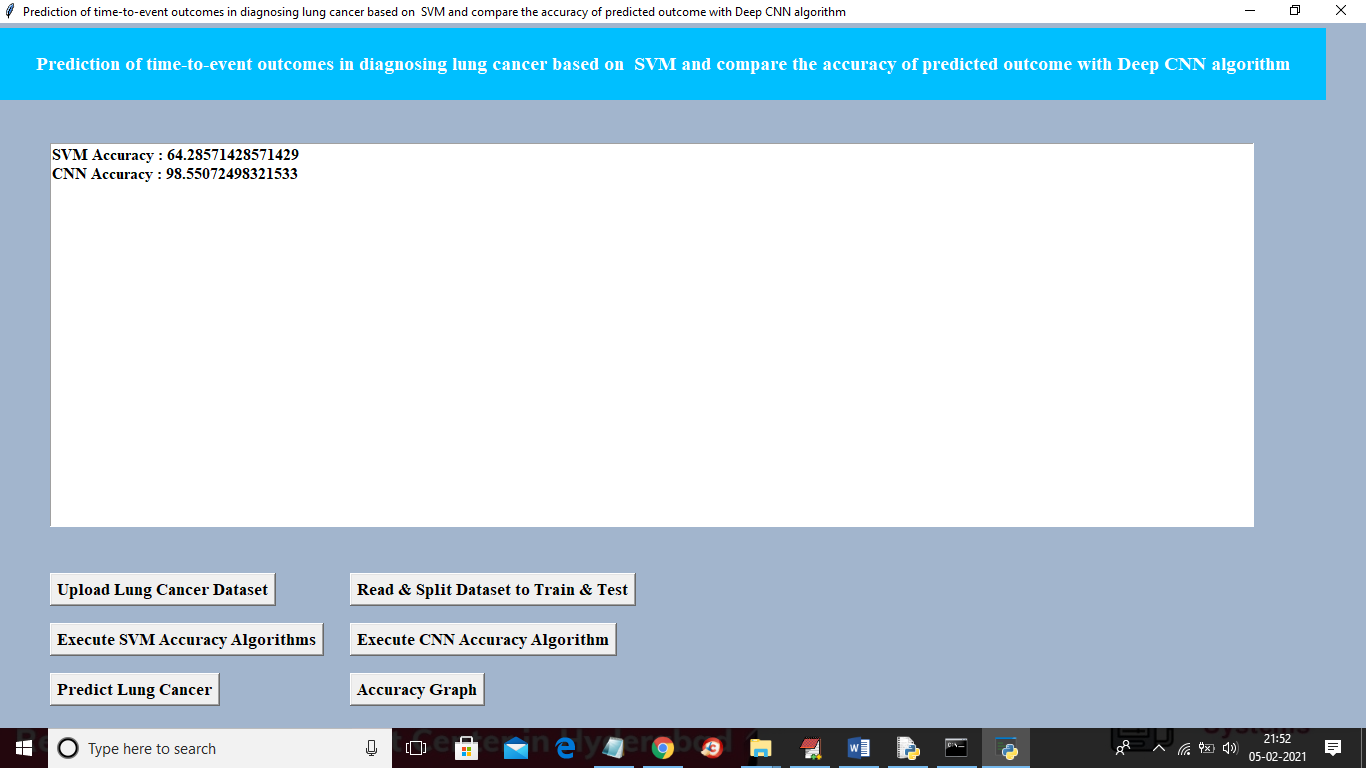
In this project we are training SVM and CNN with same LUNG dataset and then calculating and comparing accuracy of both algorithms

SCREEN SHOT

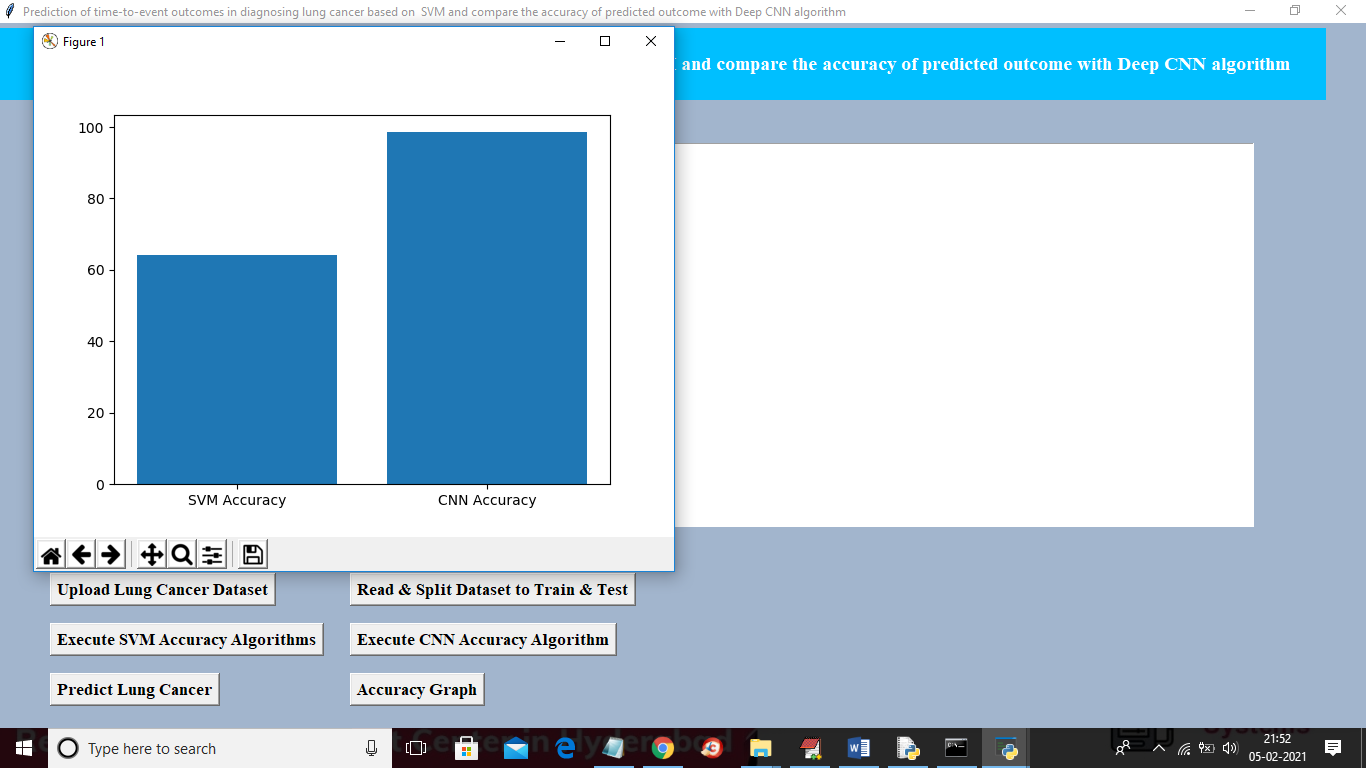
To run this project double click on ‘run.bat’ file from ‘Title4\_SVM\_CNN\_Accuracy’ folder to get below screen



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



**CHAPTER-10**

10.1.CONCLUSION AND FUTURE SCOPES.

This proposed study tries to defeat the problems faced in the early detection of lung cancer nodules before it gets worst. For this purpose, this study develops an effective computer-aided diagnosis scheme for early detecting of this lethal cancer. Chest tomography scans have been employed here as data input to the proposed model. This study's goal was to improve a CNN deep learning model able to detecting and classifying lung cancer nodules successfully. The obtained model gives high accuracy reaches 93.548% while applying on the dataset collected. In the future, this study can be improved by includes regularly enhancing the accuracy of the model by training it on bigger and more extensive datasets. Besides, various machine learning models can be combined to compare

**CHAPTER-11**

**REFERENCES**

Welch HG, Schwartz LM, Woloshin S. Are increasing 5-year survival rates evidence of

success against cancer? Jama. 2000;283(22):2975-8.

Asuntha A, Srinivasan A. Deep learning for lung Cancer detection and classification.

Multimedia Tools Applications. 2020:1-32.

Nie L, Zhang L, Yang Y, Wang M, Hong R, Chua T-S, editors. 2015, Beyond doctors:

Future health prediction from multimedia and multimodal observations.

Proceedings of the 23rd ACM international conference on Multimedia;

American Lung Association, URL www.lung.org.

Siegel, Rebecca, Naishadham, Deepa, Jemal, Ahmedin. Cancer statistics, 2013. a cancer

journal for clinicians. 2013;63(1):11-30.

Republic of Iraq, 2016,Ministry of Health\Environment, Board. IC. Annual Report Iraqi

Cancer Registry 2016.

Republic of Iraq, Ministry of Health\Environment, 2015., Board. IC. Annual Report Iraqi

Cancer Registry 2015.

Nasser, Ibrahim M Abu-Naser, S. S. 2019, Lung Cancer Detection Using Artificial

Neural Network. International Journal of Engineering Information

Systems.;3(3):17-23.

Taher, Fatma Sammouda, Rachid., editors. 2011, Lung cancer detection by using artificial

neural network and fuzzy clustering methods. 2011 IEEE GCC

Conference and Exhibition (GCC);: IEEE.

Eskandarian P, Bagherzadeh J, editors. 2015, Computer-aided detection of Pulmonary

Nodules based on SVM in thoracic CT images. 7th Conference on

Information and Knowledge Technology (IKT); 2015: IEEE.

Ganesan S, Subashini T, Jayalakshmi K, editors. 2014,Classification of X-rays using

statistical moments and SVM. 2014 International Conference on

Communication and Signal Processing;: IEEE.

Parveen SS, Kavitha CJIJoCA. 2014, Classification of lung cancer nodules using

SVM Kernels.;95(25).

Thamilselvan P, Sathiaseelan J. 2016, Detection and classification of lung cancer MRI images

by using enhanced k nearest neighbor algorithm. Indian Journal of Science Technology.;9(43):1-7.

Kurkure M, Thakare A, editors. 2016, Lung cancer detection using genetic approach.

2016 International Conference on Computing Communication Control

and automation (ICCUBEA);: IEEE.