A Project Review-2 Report

on

LUNG CANCER DETECTION USING CNN

Submitted in partial fulfillment of the

requirement for the award of the degree of

Bachelor Of Technology

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of
Dr Alok Katiyar
Professor

Submitted By

Shivam Shukla 19SCSE1010408
Prabhat Mishra 19SCSE1010434

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
April, 2023

Table of Contents

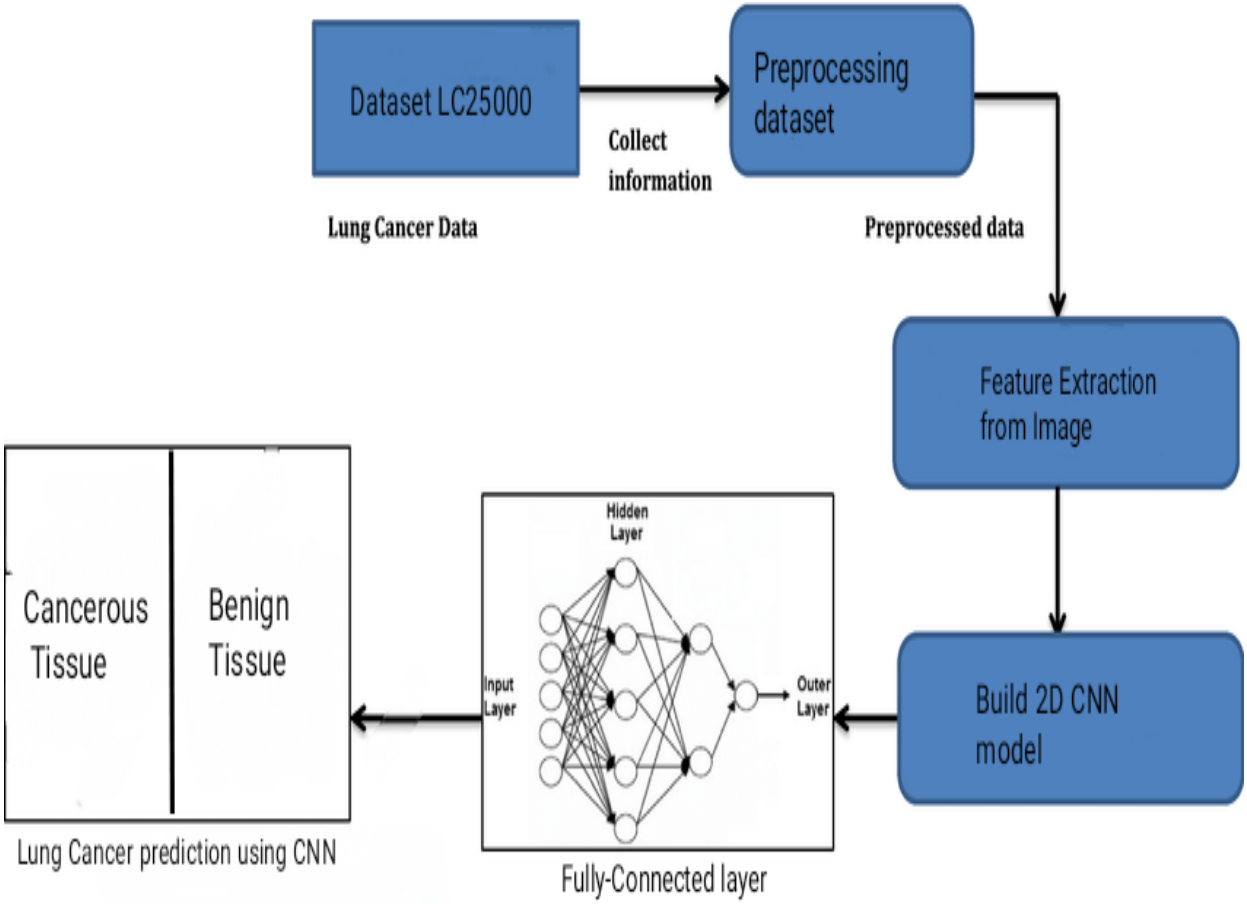
ABSTRACT

Lung Cancer is one of the leading life taking cancer worldwide. Early detection and treatment are crucial for patient recovery. Medical professionals use histopathological images of biopsied tissue from potentially infected areas of lungs for diagnosis.Most of the time, the diagnosis regarding the types of lung cancer are error-prone and time-consuming. The primary cause of death from lung cancer, about 80% is from smoking. Lung cancer in non-smokers can be caused by exposure to radon, second-hand smoke, air pollution, or other factors like workplace exposures to asbestos, diesel exhaust, or certain other chemicals lung cancers some people who do not smoke . Various tests like imaging sets (x-ray, CT scan), Sputum cytology, and tissue sampling (biopsy) are carried out to look for cancerous cells and rule out other possible conditions.

Convolutional Neural networks can identify and classify lung cancer types with greater accuracy in a shorter period, which is crucial for determining patients' right treatment procedure and their survival rate.

In the existing technique, the Support Vector Machine (SVM) based classification is performed for Lung Cancer Detection. It needs feature extraction output. Based on feature value, the classification output is generated and accuracy is calculated. The computation time is high and accuracy is low in SVM based cancerous and non-cancerous detection.Benign tissue, Adenocarcinoma, and squamous cell carcinoma are considered in this project work. The CNN model training and validation accuracy of 96.11 and 97.2 percentage can be achieved.

Lung Cancer Prediction Architecture



CHAPTER-1
Introduction

1.1 Introduction

Lung cancer is prominent cancer among both men and women, making up almost 25% of all cancer deaths . The primary cause of death from lung cancer, about 80% is from smoking. Lung cancer in non-smokers can be caused by exposure to radon, second-hand smoke, air pollution, or other factors like workplace exposures to asbestos, diesel exhaust, or certain other chemicals lung cancers some people who do not smoke . Various tests like imaging sets (x-ray, CT scan), Sputum cytology, and tissue sampling (biopsy) are carried out to look for cancerous cells and rule out other possible conditions. While performing the biopsy, evaluation of the microscopic histopathology slides by experienced pathologists is indispensable to establishing the diagnosis and defines the types and subtypes of lung cancers . For pathologists and other medical professionals diagnosing lung cancer and the types is a time-consuming process. There is a significant change the cancer types are misdiagnosed, which directs to incorrect treatment and may cost patients' lives.

1.2 Problem Statement

In the existing technique, the Support Vector Machine (SVM) based classification is performed for Lung Cancer Detection. It needs feature extraction output. Based on feature value, the classification output is generated and accuracy is calculated. The computation time is high and accuracy is low in SVM based cancerous and non-cancerous detection.

1.3 Formulation of Problem

Convolutional Neural networks can identify and classify lung cancer types with greater accuracy in a shorter period, which is crucial for determining patients' right treatment procedure and their survival rate.

Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that allows machines to learn without explicit programming by exposing them to sets of data allowing them to learn a specific task through experience. Some papers also considered using histopathological images, but they distinguish between carcinomas and non- carcinomas images and with lower accuracy. This project work has considered using Convolutional Neural Network (CNN) architecture to classify the benign,

Adenocarcinoma, and squamous cell carcinomas.

1.3.1 Tools and Technology Used

We have used Python language (v3.8) to Implement CNN algorithm,IDE as Visual Studio Code.Various Libraries such as Tensorflow,OpenCV,Matplotlib,Keras are used.

CHAPTER-2
Literature Survey

Since the year 2005 study, the death number of US residents have risen to 150,000 which is an estimated increase of one- forth i.e. 25% in 2018[1]. According to NAACR report , In USA, 234,000 people have been found and diagnosed with lung cancer which is also known as pulmonary cancer. In complementary to this the American cancer community performed a study in the year 2019 in which approximately 228,000 people below the age of 40 years were found with cancer with approx. same number of men and women and according to this study and research 141,000 people have died due to lung cancer[2].

According to the study, the number of the patients in which lung cancer is found and the number of people died because of the lung cancer is rising since last half of a decade. Basic diagnosis cannot be used for the early detection of Cancer.

Usually,detection of lung cancer can be done with the help of a highly qualified professional and with the help of symptoms like cough, pain in chest, fatigue, weight loss, memory loss,Broken bones,headaches,Alzheimer's disease and dementia,bleeding, swelling of the face,tone chance[3],joint pain.As soon as the patient is diagnosed and found to have lung cancer, the patient is exposed to diagnosis procedures such as biopsy, testing genes, bronchi scopey, fluid biopsy, and blood tests are used to detect lung cancer.

By using CT scan images a lung cancer detection program was performed to diagnose a disease using a large number of conventional methods such as removal of images, isolation,and feature selection and removal of cancer by using cancer classification. A separate system helps in removing other regions and reducing or completely removing complexity. Feature reduction process help in removing the need to compute the areas we don't need to compute and help us to save computational time and it also reduces data overload. With the help of a defined separator, picking different features which give correct and most information about the lung cancer we can use K-Neighbor, Machine learning based equipment such as SVM and other clever techs. Although the conventional technique of cancer prediction can easily be used to predict lung cancer, it takes a lot of time and the accuracy is not promising. In addition, the process can fail to define the minimum quality of the CT scan of the images so false positive elements are removed and can cause misalignment of many phases.

T.Atsushi, T.Tetsuya,and K.Yuka used the same Convolutional Neural Network on images of free cells to execute lung cancer type classification.They performed the activity on Small cell type of lung cancer cells, Squamous cell carcinoma, Adenocarcinoma images used as dataset images[4]. The CNN architecture of convolution and summarizing the features in the area of its map and 2 fully well

attached layers with dropout 0f 0.5 were used. This model showed an accuracy of 71 percent which looked like a failure and was quite low to be considered.

Detection of lung cancer in CT screening uses different algorithms for a most optimizec way of performance. During the study in detection of lung cancer 5 approaches are used which are intermediate compounding, mean integration and particle refinement and particle mixing technique are implemented to diagnose the tumor in the CT scan of the lungs. CT scan images can easily be used to detect and remove the cancerous part shown in the image and can be done with high accuracy[5].Adaptive filter are highly useful to enhance the quality of the image. Then the process of feature extraction begins and affected region is identified using the algorithm mentioned above. Thus, program presented detect the lung cancer with high accuracy of 95 percentage. Identification of the CT images of lungs is also very helpful. Primary step is to collect the LIDC IDRI data and used, Using multiple layers of deep learning , an accuracy of 83% can be achieved easily.

Although several mixes of techniques are used incorrect classification and management of data are still important issues present in lung cancer detection using machine learning[6].

Many researchers have tried to implement models to predict lung cancer using a variety of imaging and detection techniques using machine learning techniques. W. Rahane, H. Dalvi, Y. Magar, A. Kalane, and S. Jondhale used the algorithms of machine learning in Support vector machines and they used classification of images on CT scan images to provide better accuracy.

Sangamithraa and Govindaraju used the model of K mean clustering which is an unregulated learning algorithm to divide and combine. According to specific features, pixels are combined. A back-distribution network. For feature extraction such as entropy, correlation, homogeneity, PSNR, SSIM etc. we use Gray-level co-occurrence matrix (GLCM) method. This gives an accuracy of 97 percent. To remove the noise and improve the model we use the median filter method.

S.Sasikala,M. Bharathi,B.R. Sowmiya , studied the process using CNN on CT scan images for the early detection and diagnosis of lung cancer. Using the numeric computing platform MATLAB the obtained a huge amount of features from input data as the first phase and classification as the second phase. Their proposed system could classify the cancerous and non-cancerous cells with 96% accuracy.

W. Ausawalaithong, A. Thirach, S. Marukatat, and T. Wilaiprasitporn performed the study by using the same CNN method but with a different approach .By the help of transfer learning the performed the study and they used X-ray images of patient chest from various datasets[7].

3. METHODOLOGY USED

A simple linear bunch of sequential layers was used to create the ConvNets for the processing and feature extraction of image. Pre- Trained images were used for the process. They were passed through the layers of filters of kernel,pooling and neural networks.

Various functions were used to classify the feature value in the image as object.We tested our model in google colaboratory. Image of size (4000*4000)pixels were taken from the dataset and passed through the input layer.

This layer consisted of a neural network with some of the hidden layers and a fully connected layer with input layer.

A kernel matrix along with a activation function is attached to every layer of the network. Max pooling size was used to minimize the parameters of calculation for every next network layer.

An Accuracy and loss graph is obtained from using the algorithm.The image was loaded into the algorithm and we can see the following results based on the nature of image.

To train CNN we created multiple layers where first layer process images of size 62 X 62 and second layer process 31 X 31.

In Accuracy graph x-axis represents Epoch and Y-axis represents accuracy and loss values.To train CNN we took 10 Epoch and at each increasing Epoch Loss values get decrease and accuracy gets increase and in above graph red line represents loss and green line represents accuracy.

3.1 Improvement of algorithm.

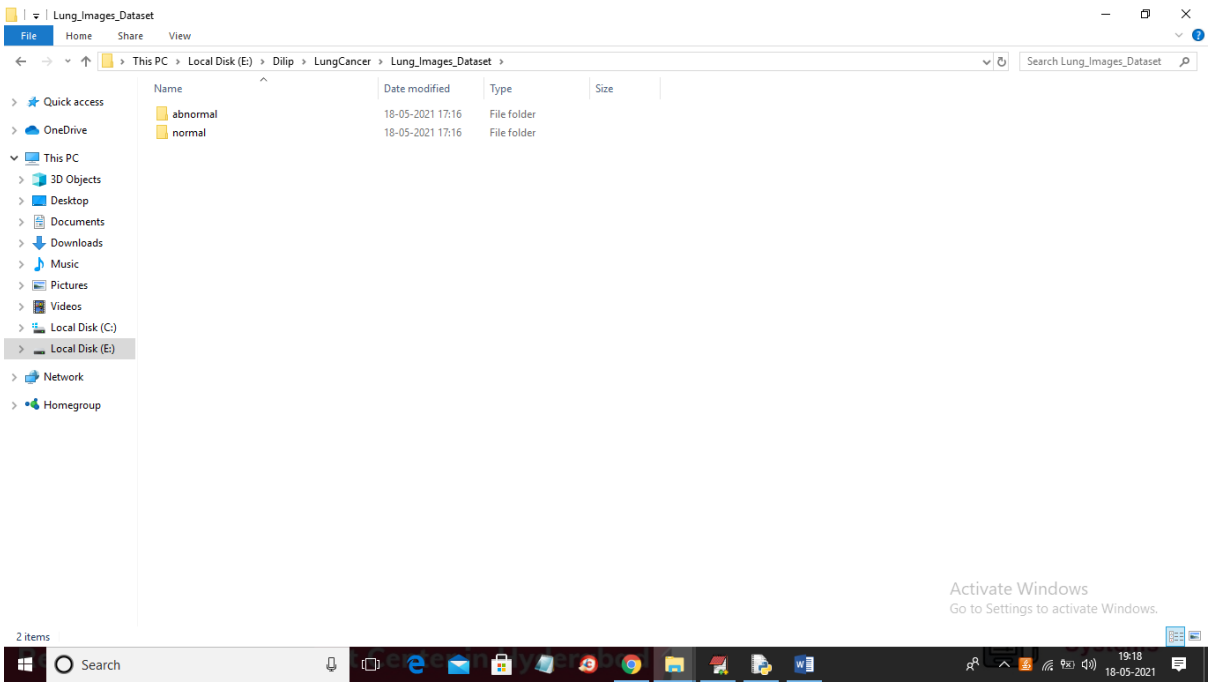
We tried to improve the CNN algorithm for better accuracy using various big pre-trained datasets.By using limited number of epochs we can find the number of epochs as after a limited numbers there will be no change in training loss and accuracy.

By using dropout layer we prevented overfitting.Moreover,using k-fold cross optimization we splitted our images in k equal parts and then trained them k times with a different set of training process.

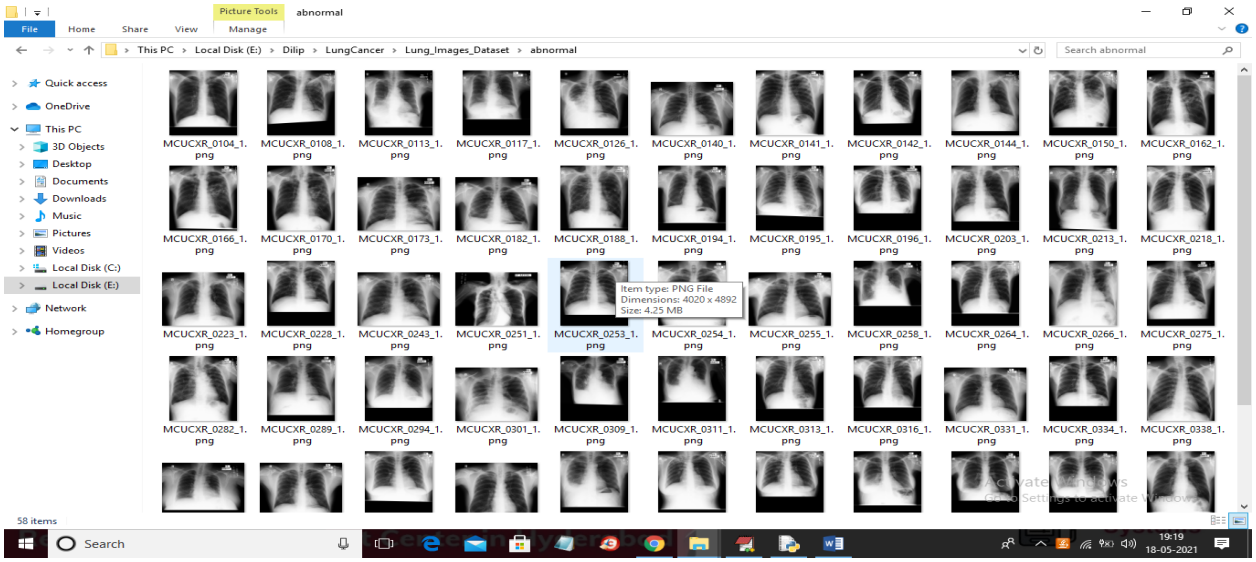
Cut Mix was used with the help of tensorflow and keras library to augment the images.It combined the random parts of two images which eventually increased the performance of the model.

4. SIMULATION

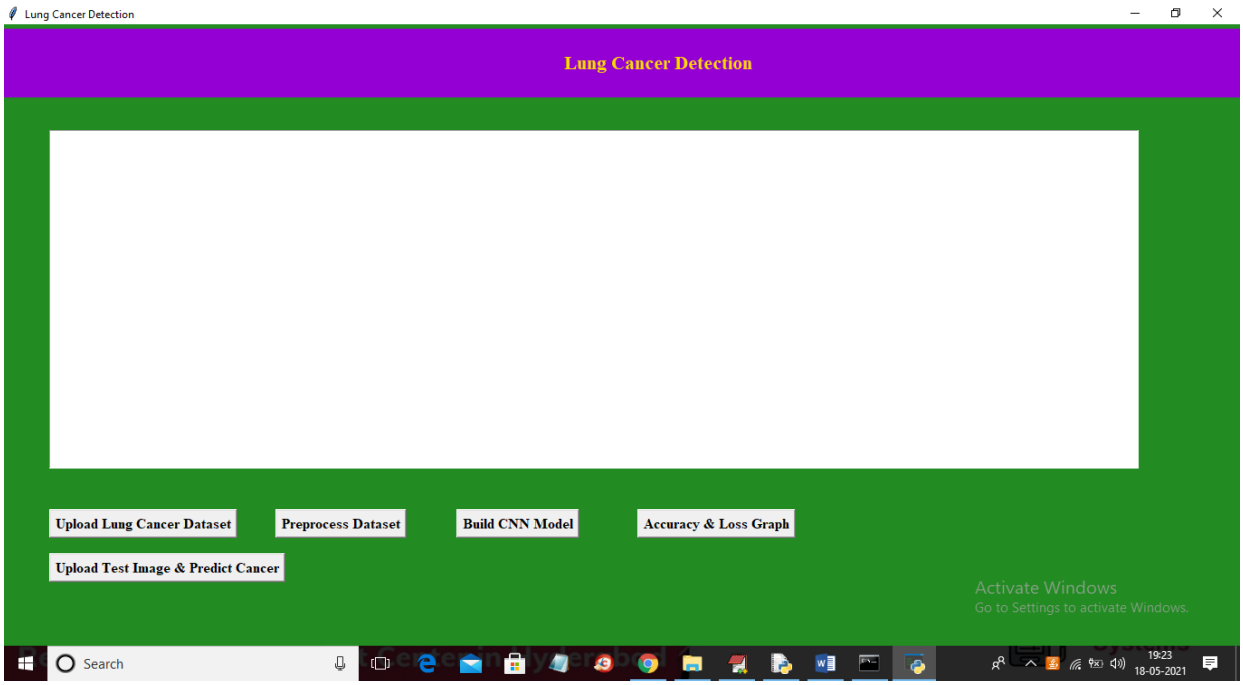
In this project we are using CNN algorithm to detect Lung cancer from X-ray images and to train CNN we have X-ray images dataset and this dataset saved inside 'Lung_Images_Dataset' folder and below screen shots showing images from dataset



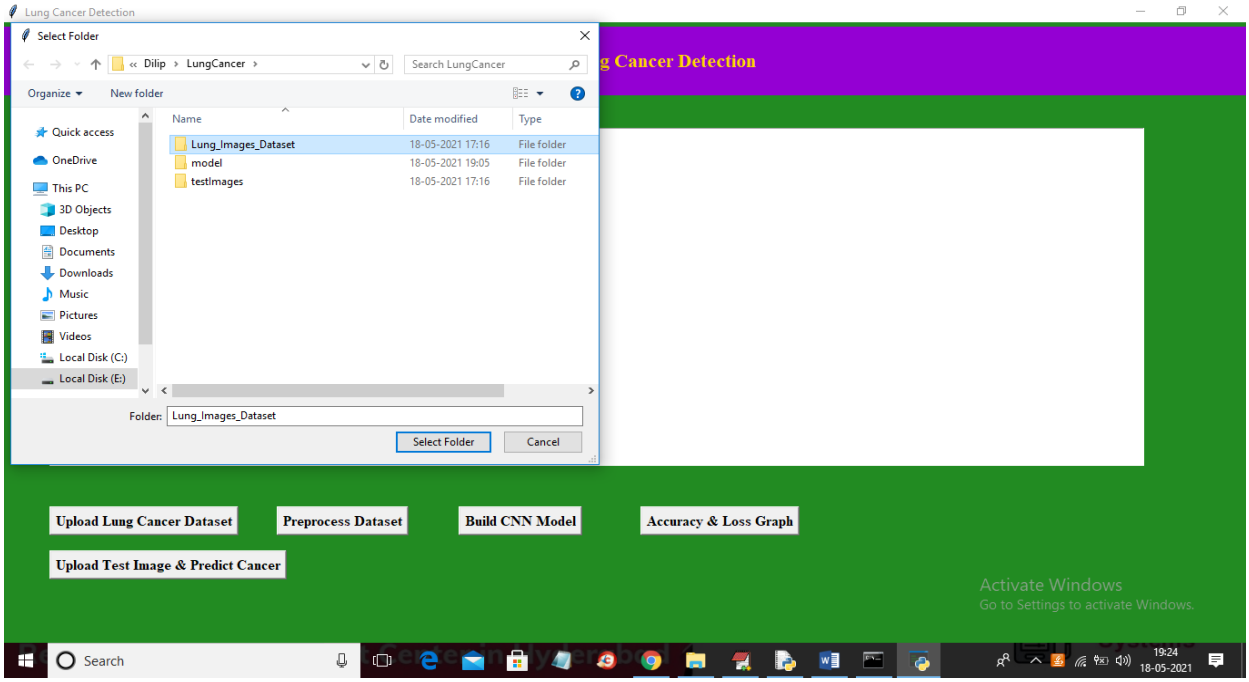
In above screen we have two folders where one folder contains NORMAL X-ray images and other folder contains ABNORMAL.



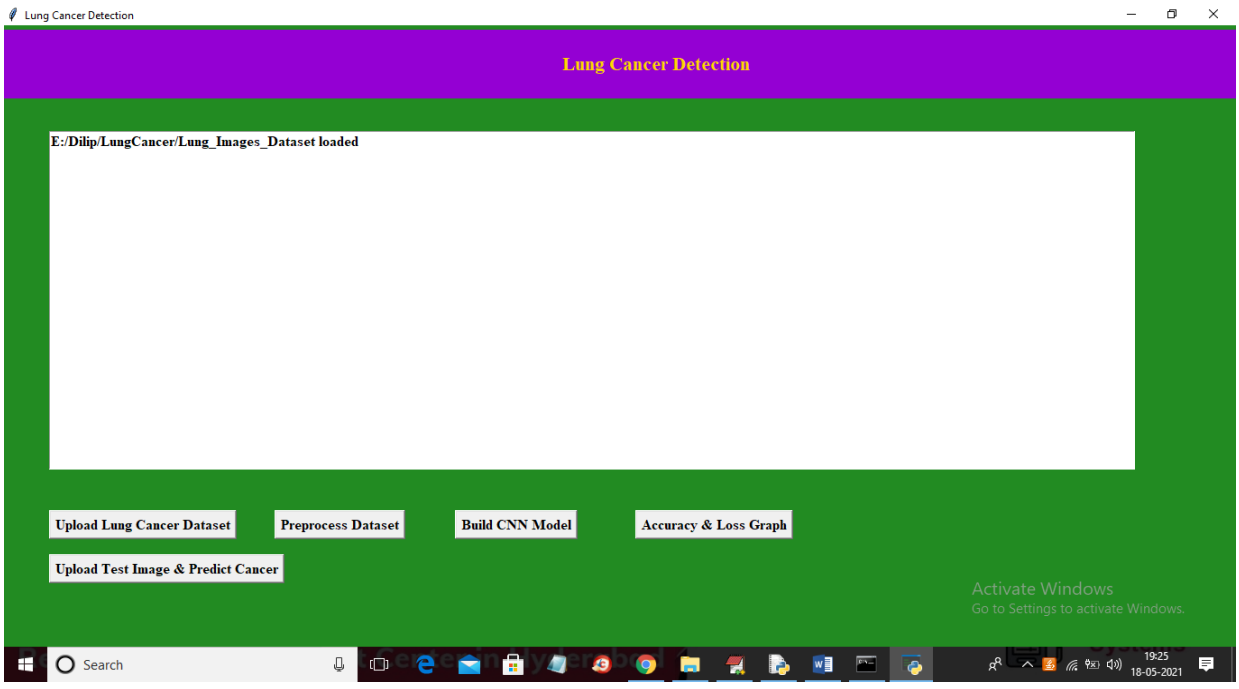
In above screen displaying few images from ABNORMAL folder and we used above images to train CNN algorithm and after training CNN we can upload test images and then CNN will predict whether X-ray contains normal or abnormal tumor.



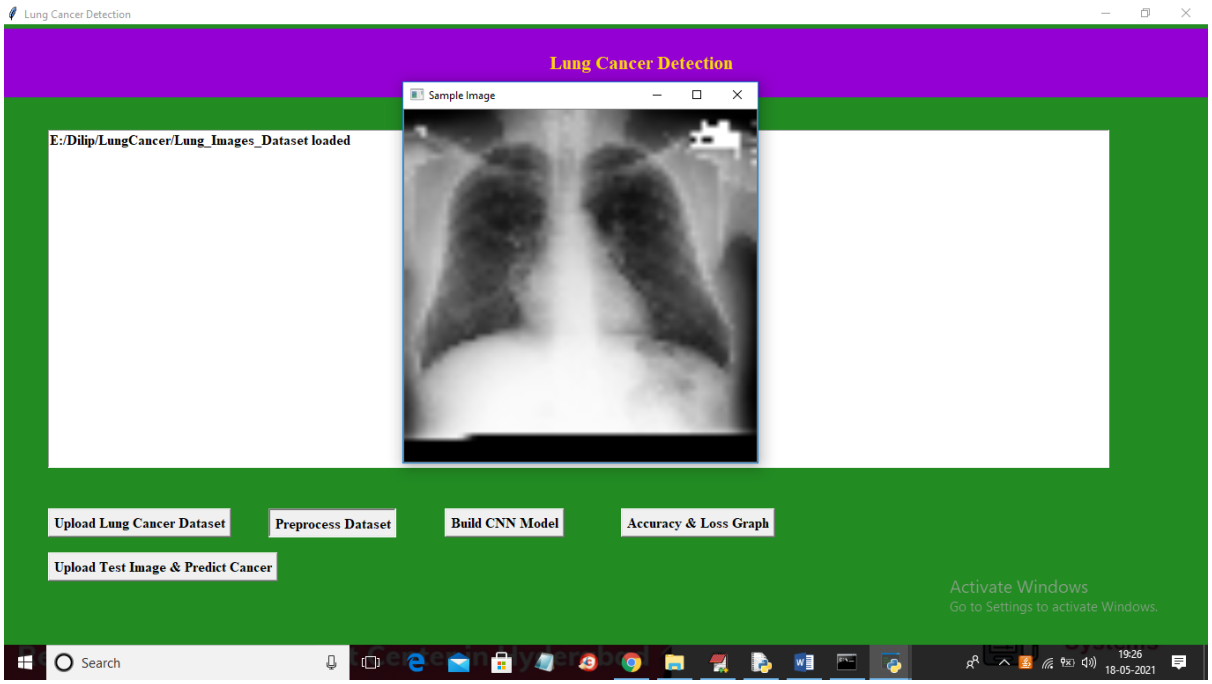
In above screen click on 'Upload Lung Cancer Dataset' button to upload X-ray images



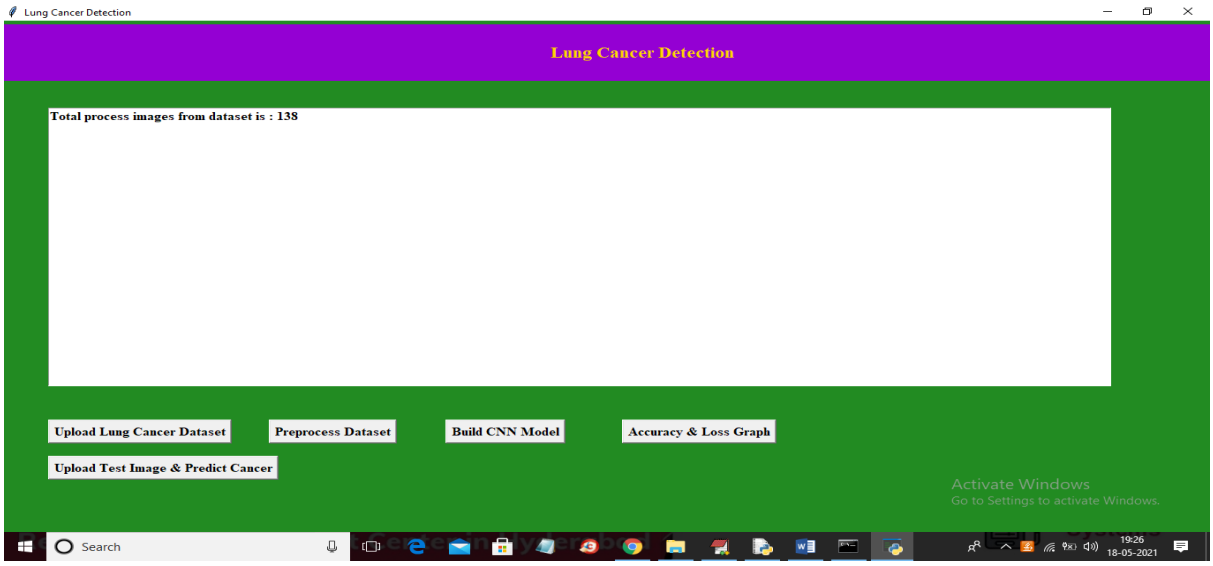
In above screen selecting and uploading 'Lung_Image_Dataset' folder and then click on 'Select Folder' button to load images and to get below screen



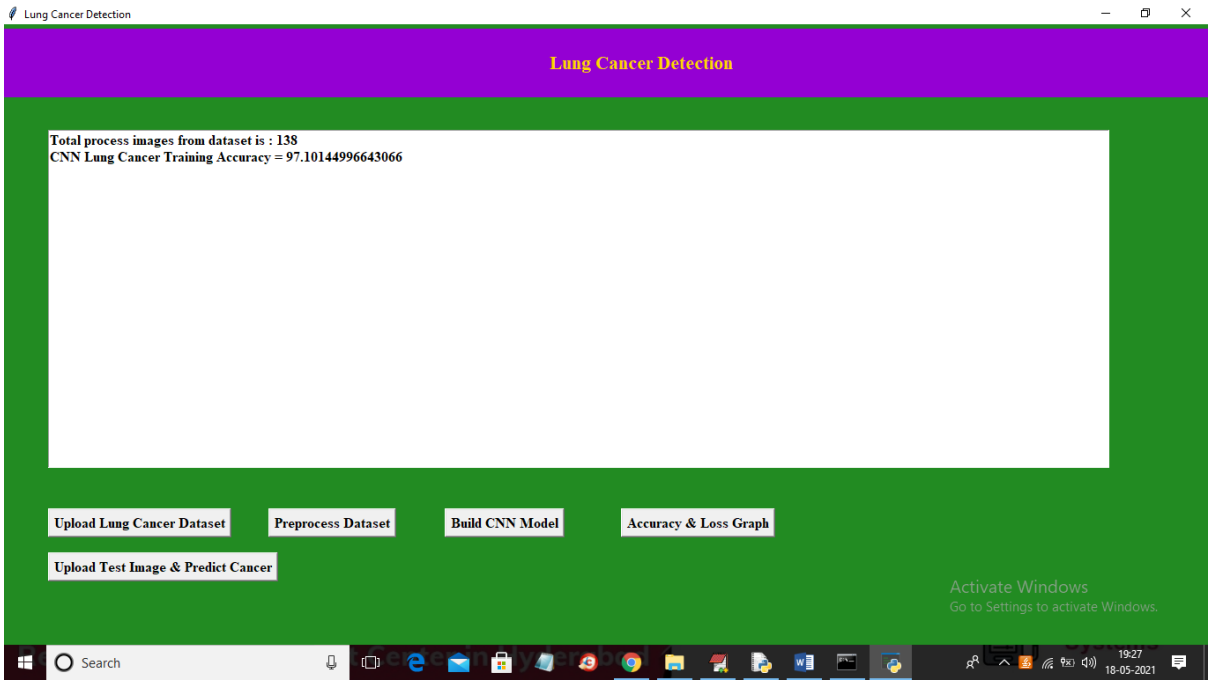
In above screen dataset loaded and now click on 'Preprocess Dataset' button to convert all images into color format and resize them into equal sizes so CNN can accept those images



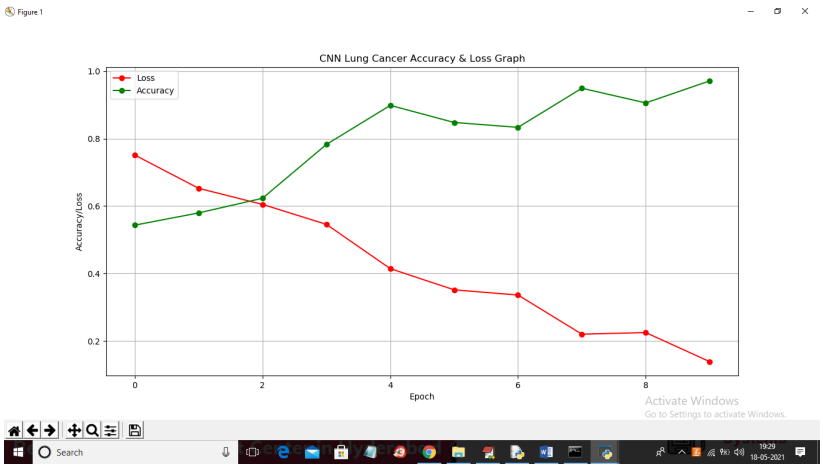
In above screen application process all images and then one sample image to confirm all images loaded properly and now close above image to get below screen



In above screen we can see dataset contains total 138 images and now click on 'Build CNN Model' button to train CNN algorithm on above images and then calculate prediction accuracy

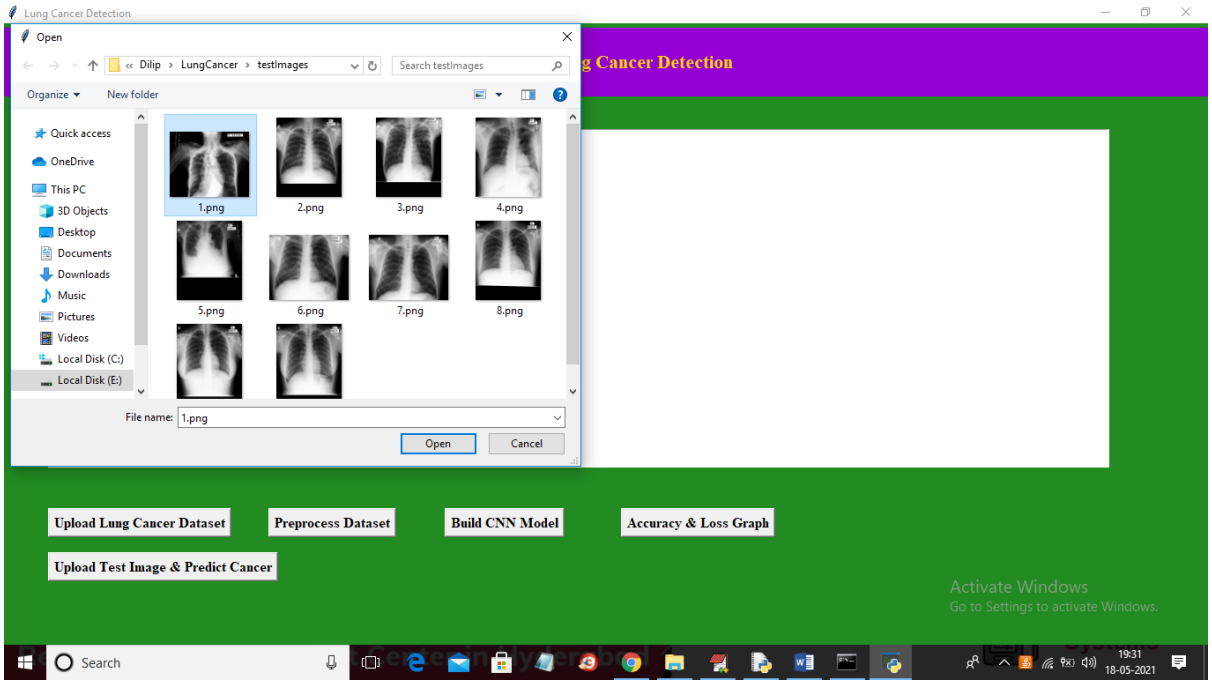


To train CNN we created multiple layers where first layer process images of size 62 X 62 and second layer process 31 X 31 and goes on and now click on 'Accuracy & Loss Graph' button to get below graph

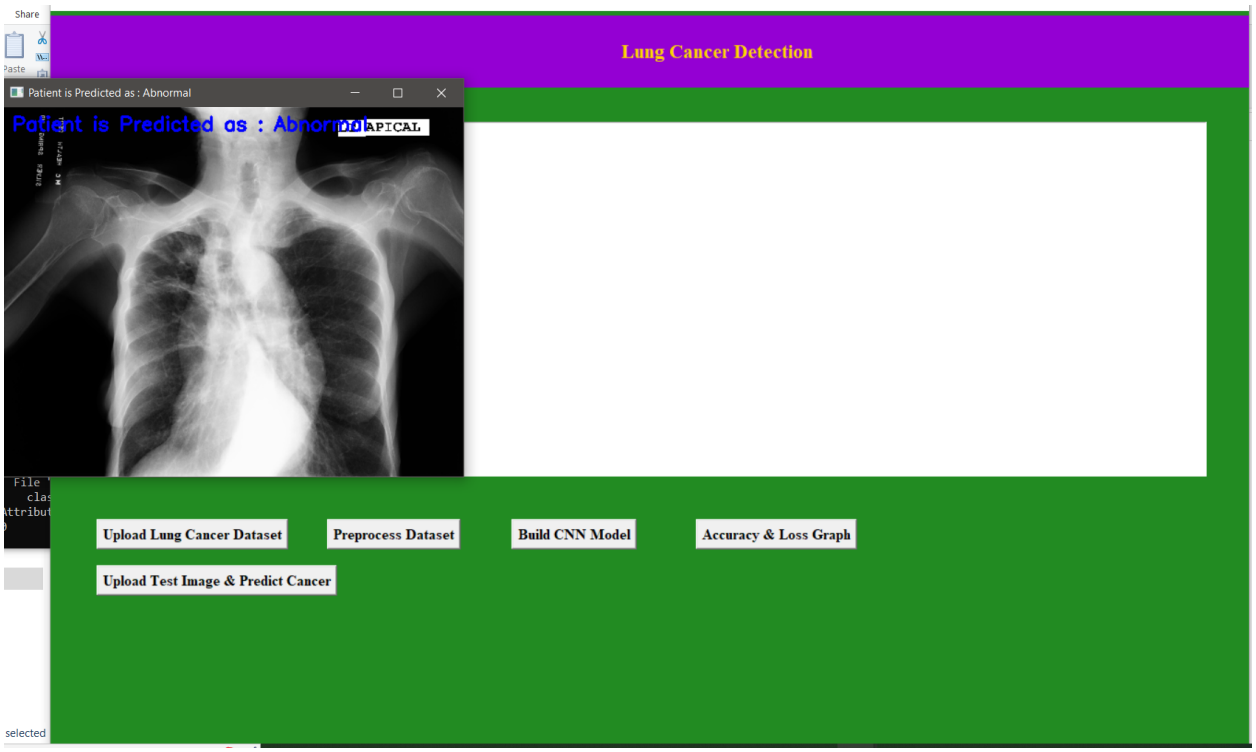


In above graph x-axis represents Epoch and Y-axis represents accuracy and loss values and in above graph we can see to train CNN we took 10 Epoch and at each increasing Epoch Loss values get decrease and accuracy gets increase and in above graph red line represents loss and green line represents accuracy.

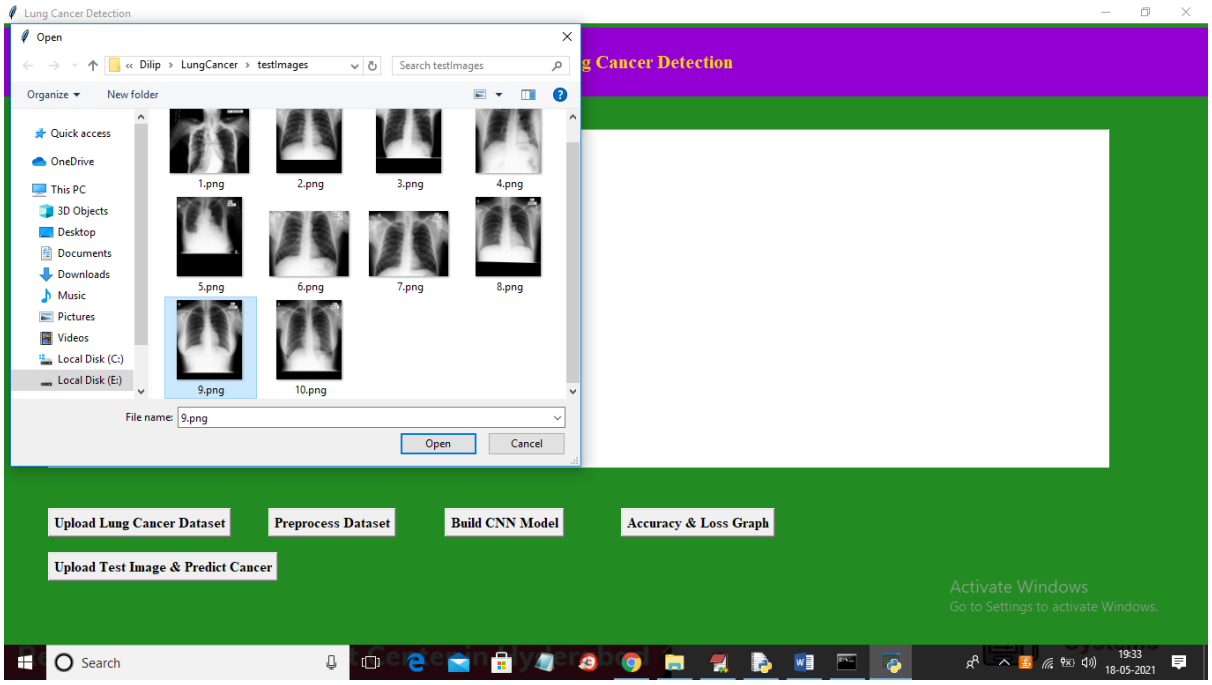
Now click on ‘Upload Test Image & Predict Cancer’ button to upload test image and then detect cancer



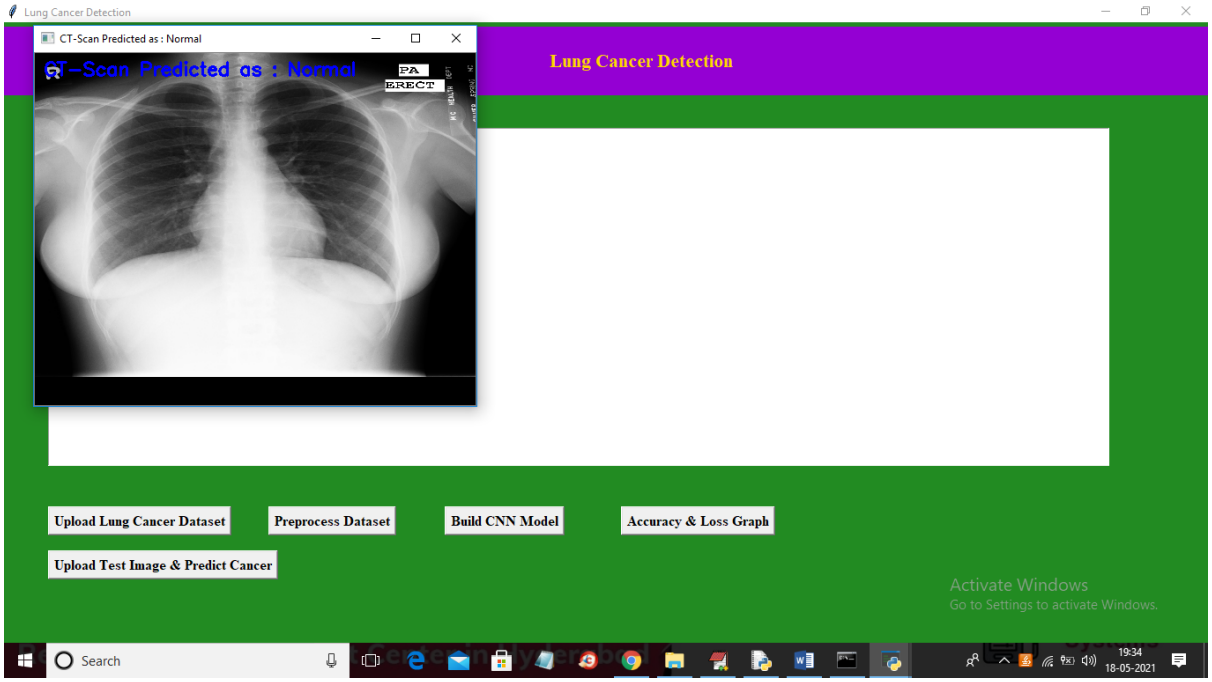
In above screen selecting and uploading ‘1.png’ file and then click on ‘Open’ button to get below result



In above screen in first image in blue colour text we can see predicted result as X-ray contains abnormality and in second image we are detecting places were abnormality detected and in third image we extracted all abnormality patches from original image and then displaying. Now test other image



In above screen selecting and uploading ‘9.png’ file and then click on ‘Open’ button to get below result



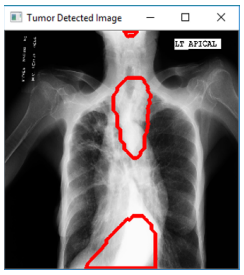
In above screen X-ray is predicted as NORMAL. Similarly we can upload and test other images.

4. RESULT AND FUTURE IMPROVEMENTS

This CNN model has proved to achieve a higher accuracy in limited number of epochs.But this model in some cases has failed to predict the cancer.So in the future, we will be using different optimization techniques and to provide patient with more information along with the result,such as Tumor region marked in circle,Black and White image of Tumor region.

A prototype image for the above improvement is shown below.

1.Tumor Region



2.Extracted Tumor-region for the same image

Furthermore a concept of 3-D CNN can be implemented to make the algorithm work on MRI images.

5. REFERENCES

[1] Liu Z, Wang J, Yuan Z, Zhang B, Gong L, Zhao L, Wang P (2018) Preliminary results about application of intensity-modulated radiotherapy to reduce prophylactic radiation dose in limited-stage small cell lung cancer. J Cancer 9(15):2625–2630.

[2] Balmelli C, Railic N, Siano M, Feuerlein K, Cathomas R, Cristina V, Gu¨thner C, Zimmermann S, Weidner S, Pless M, Stenner F, Rothschild SI (2018) "Lenvatinib in advanced radioiodine-refractory thyroid cancer: a retrospective analysis of the swiss lenvatinib named patient program. J Cancer 9(2):250–255.

[3] Manser R, Lethaby A, Irving LB, Stone C, Byrnes G, Abramson MJ, Campbell D (2013) Screening for lung cancer. Cochrane Database of System Rev 6(6):CD001991. https://doi.org/10.1002/ 14651858.cd001991.pub3 K

[4] Brock MV et al (2008) DNA methylation markers and early recurrence in stage I lung cancer. N Engl J Med 358:1118–1128

[5] Wang CC et al (2015) HOXA5 inhibits metastasis via regulating cytoskeletal remodelling and associates with prolonged survival in non-small-cell lung carcinoma.
PLoS ONE 10:e0124191
[6] Lee HY et al (2015) Differential expression of microRNAs and their target genes in non-small-cell lung cancer. Mol Med Rep 11:2034– 2040

[7] Manogaran G, Shakeel PM, Hassanein AS, Priyan MK, Gokulnath C (2018) Machine-learning approach based gamma distribution for brain abnormalities detection and data sample imbalance analysis.
IEEE Access. https://doi.org/10.1109/ ACCESS.2018.2878276

6.ACRONYMS

CNN	Convolutional Neural Network
SVM	Support Vector Machine
ML	Machine Learning
CT	Computed Tomography
AI	Artificial Intelligence
GLCM	Gray-level co- occurrence matrix