**Lung Cancer Detection & stages Identification using CNN algorithm**

**ABSTRACT**

Identification of lung cancer is an efficient way to minimize the death rate and maximize survival rate of patients. It is an essential step to screen out the computed tomography (CT) images for pulmonary nodules towards the efficient treatment of lung cancer. However, robust nodule identification and detection is a most critical task due the complexity of the surrounding environment and heterogeneity of the lung nodules. The use of machine learning to detect, predict, and classify disease has grown exponentially in the past few years, especially for complex tasks such as lung cancer detection and recognition. Deep Convolutional neural networks (DCNN) have exploded in popularity for transforming the field of computer vision research. In this paper, we are using Deep Convolutional Neural Network for lung cancer classification using CT images based lung cancer image dataset consortium (LIDC) for detecting cancerous and noncancerous lung nodules for measuring the accuracy of classification better than existing methods.

**Keywords:** Lung Cancer, DCNN, Computed Tomography, computer vision

**CHAPTER 1**

**INTRODUCTION**

**1.1 Introduction:**

Lung cancer is one of the important reasons to increase death rate in the world, since every year it is seen that many deaths were occur due to lung cancer as compare to other types of cancer. Both men and women are being affected from this deadly disease. Hence suitable mechanism should be adopted to detect and identify this disease in the initial stage to save the life of large number of peoples suffering from lung cancer. If it is detected and identified in primary stage then survival rate of many number of patients can be improved.

Later after disease identification, by providing proper diagnosis can reduce the death rate of patients. So in order to avail a suitable and instantaneous outcome the importantly, applying recent techniques of machine leaning in the medical image processing field by enhancing the amount of duplication for the methods use can increase the accuracy of the classification. Therefore proper timely detection and identification in the prior stage will definitely improve the level of survival and can decrease the death rate.

The medical images taken in most of the earlier studies comprise of computed tomography (CT), magnetic resonance, and mammography images. The expert doctor of this domain uses these images for analysis to detect and identify the various levels of lung cancer by using suitable techniques. The different laboratory and clinical steps are being used including chemical treatment to destroy or stop the duplications of malignant cell, targeted therapy and also radiotherapy.

All these procedures adopted to identify and detect the cancer diseases are lengthy, costlier and more painful for the patients. Thus, to overcome all these problems suitable machine learning techniques for processing these medical images were used which comprise of CT scan images. CT scan images are preferred compared to other images because as comparable to other medical images such as MRI and X-Ray, CT images are less noisy. In the process of lung cancer classification, the images applied at the input layer of Deep Convolutional neural networks are classified into cancerous or non-cancerous at the output layer after processing in all hidden layers of the network.

DCNN is a deep learning algorithm that takes an input image, and then marks significance for each object in the image. The network further classifies each object in the image one from the other when it is trained precisely with more number of dataset. Deep learning methods needs minimum pre processing steps in comparable to the other image processing algorithms.

The objective of DCNN is to convert input images suitable for processing with minimum permissible loss of image features for achieving the best level of accuracy. To design and to attain better accuracy of classification in the DCNN, the parameters used are size of filters, more no of hidden layers and extracted number of feature maps. As the network layers are deeper, there is high detection level with high level of abstraction of features can be achieved. Deeper the network leads to increase in computation time due to more number of Convolutional operations.

The most suitable size of the Convolutional filters is 3 x 3 or 5 x 5. The performance of the network may minimize as the size of the Convolutional kernel increases.

**1.2 Lung cancer:**

Lung cancer is a type of cancer that begins in the lungs. Your lungs are two spongy organs in your chest that take in oxygen when you inhale and release carbon dioxide when you exhale.

Lung cancer is the leading cause of cancer deaths in the United States, among both men and women. Lung cancer claims more lives each year than do colon, prostate, ovarian and breast cancers combined.

People who smoke have the greatest risk of lung cancer, though lung cancer can also occur in people who have never smoked. The risk of lung cancer increases with the length of time and number of cigarettes you've smoked. If you quit smoking, even after smoking for many years, you can significantly reduce your chances of developing lung cancer.

Cancer is a noteworthy general heath issue worldwide with mortality rates increasing day by day. Lung cancer, among all other cancer types is the most common and deadly that occur both in men and women. Lung cancer, additionally known carcinoma is formation of malignant lung tumors (cancerous nodules) due to uncontrolled growth of cells in lung tissues. Eating tobacco and smoking are the leading risk factors for causing cancerous lung nodules. The survival rate of lung cancer patients combining all stages is very less roughly 14% with a time span of about 5-6 years.

The main problem with lung cancer is that most of these cancer cases are diagnosed in later stages of cancer making treatments more problematic and significantly reducing the survival chances. Hence detection of lung cancer in its earlier stages can increase the survival chances up to 60-70% by providing the patients necessary fast treatment and thus it curbs the mortality rate. Small cell lung cancer and non-small cell lung cancer are two main types of lung cancer classifications based on cell characteristics. The most commonly occurring is non-small cell lung cancer that makes up about 80-85% of all cases, whereas 15-20% of cancer cases are represented by small cell lung cancer.

Lung cancer staging depends on spread of cancer in the lungs and tumor size. Lung cancer is mainly classified into 4 stages in order of seriousness:

Stage I-Cancer is confined to the lung,

Stage II and III-Cancer is confined within the chest and

Stage IV-Lung cancer has spread from the chest to other parts of the body. Lung cancer diagnosis can be done by using various imaging modalities such Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and Chest X-rays. CT scan images are mostly preferred over other modalities because they are more reliable, have better clarity and less distortion. Visual interpretation of database is a tedious procedure that is time consuming and highly dependent on given individual. This introduces high possibility of human errors and can lead to misclassification of cancer. Hence an automated system is of utmost importance to guide the radiologist in proper diagnosis of lung cancer.

**1.2.1 Symptoms**

Lung cancer typically doesn't cause signs and symptoms in its earliest stages. Signs and symptoms of lung cancer typically occur only when the disease is advanced.

Signs and symptoms of lung cancer may include:

* A new cough that doesn't go away
* Coughing up blood, even a small amount
* Shortness of breath
* Chest pain
* Hoarseness
* Losing weight without trying
* Bone pain
* Headache

As cancer spreads, additional symptoms depend on where new tumors form. For example, if in the:

* lymph nodes: lumps, particularly in the neck or collarbone
* bones: bone pain, particularly in the back, ribs, or hips
* brain or spine: headache, dizziness, balance issues, or numbness in arms or legs
* liver: yellowing of skin and eyes (jaundice)

Tumors at the top of the lungs can affect facial nerves, leading to drooping of one eyelid, small pupil, or lack of perspiration on one side of the face. Together, these symptoms are called Horner syndrome. It can also cause shoulder pain

Lung cancer sometimes creates a substance similar to hormones, causing a wide variety of symptoms called paraneoplastic syndrome, which include:

* muscle weakness
* nausea
* vomiting
* fluid retention
* high blood pressure
* high blood sugar
* confusion
* seizures
* coma

## 1.2.2 What causes lung cancer?

Anyone can get lung cancer, but [90 percent](https://www.lung.org/lung-health-and-diseases/lung-disease-lookup/lung-cancer/symptoms-causes-and-risk-factors/what-causes-lung-cancer.html) of lung cancer cases are the result of smoking. From the moment you inhale smoke into your lungs, it starts damaging your lung tissue. The lungs can repair the damage, but continued exposure to smoke makes it increasingly difficult for the lungs to keep up the repair.

Once cells are damaged, they begin to behave abnormally, increasing the likelihood of developing lung cancer. Small-cell lung cancer is almost always associated with heavy smoking. When you stop smoking, you lower your risk of lung cancer over time.

Exposure to radon, a naturally existing radioactive gas, is the second leading cause, according to the [American Lung Association](https://www.lung.org/lung-health-and-diseases/lung-disease-lookup/lung-cancer/resource-library/lung-cancer-fact-sheet.html). Radon enters buildings through small cracks in the foundation. Smokers who are also exposed to radon have a very high risk of lung cancer.

Breathing in other hazardous substances, especially over a long period of time, can also cause lung cancer. A type of lung cancer called mesothelioma is almost always caused by exposure to asbestos.

Other substances that can cause lung cancer are:

* arsenic
* cadmium
* chromium
* nickel
* some petroleum products
* uranium

Inherited genetic mutations may make you more likely to develop lung cancer, especially if you smoke or are exposed to other carcinogens. Sometimes, there’s no obvious cause for lung cancer.

## 1.2.3 Stages of lung cancer

Cancer stages tell how far the cancer has spread and help guide treatment. The chance of successful or curative treatment is much higher when lung cancer is diagnosed and treated in the early stages, before it spreads. Because lung cancer doesn’t cause obvious symptoms in the earlier stages, diagnosis often comes after it has spread.

Non-small cell lung cancer has four main stages:

* **Stage 1:** Cancer is found in the lung, but it has not spread outside the lung.
* **Stage 2:** Cancer is found in the lung and nearby lymph nodes.
* **Stage 3:** Cancer is in the lung and lymph nodes in the middle of the chest.
* **Stage 3A:** Cancer is found in lymph nodes, but only on the same side of the chest where cancer first started growing.
* **Stage 3B:** Cancer has spread to lymph nodes on the opposite side of the chest or to lymph nodes above the collarbone.
* **Stage 4:** Cancer has spread to both lungs, into the area around the lungs, or to distant organs.

Small-cell lung cancer (SCLC) has two main stages. In the limited stage, cancer is found in only one lung or nearby lymph nodes on the same side of the chest.

The extensive stage means cancer has spread:

* throughout one lung
* to the opposite lung
* to lymph nodes on the opposite side
* to fluid around the lung
* to bone marrow
* to distant organs

At the time of diagnosis, [2 out of 3 people](https://www.cancer.org/cancer/lungcancer-smallcell/detailedguide/small-cell-lung-cancer-staging) with SCLC are already in the extensive stage.

## 1.2.4 Risk factors for lung cancer

The biggest risk factor for lung cancer is smoking. That includes cigarettes, cigars, and pipes. Tobacco products contain thousands of toxic substances.

According to the [Centers for Disease Control and Prevention (CDC)Trusted Source](https://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm" \t "_blank), cigarette smokers are 15 to 30 times more likely to get lung cancer than nonsmokers. The longer you smoke, the greater the chance of developing lung cancer. Quitting smoking can lower that risk.

Breathing in secondhand smoke is also a major risk factor. Every year in the United States, about 7,300 people who’ve never smoked die from lung cancer caused by secondhand smoke.

Exposure to radon, a naturally occurring gas, increases your risk of lung cancer. Radon rises from the ground, entering buildings through small cracks. It’s the [leading cause](https://www.epa.gov/radon/health-risk-radon) of lung cancer in nonsmokers. A simple home test can tell you if the level of radon in your home is hazardous.

Your risk of developing lung cancer is higher if you’re exposed to toxic substances such as asbestos or diesel exhaust in the workplace.

Other risk factors include:

* family history of lung cancer
* personal history of lung cancer, especially if you’re a smoker
* previous radiation therapy to the chest

## 1.2.5 Treatment for lung cancer

It’s usually a good idea to seek a second opinion before beginning treatment. Your doctor may be able to help make that happen. If you’re diagnosed with lung cancer, your care will likely be managed by a team of doctors who may include:

* a surgeon who specializes in the chest and lungs (thoracic surgeon)
* a lung specialist (pulmonologist)
* a medical oncologist
* a radiation oncologist

Discuss all your treatment options before making a decision. Your doctors will coordinate care and keep each other informed.

Treatment for non-small cell lung cancer (NSCLC) varies from person to person. Much depends on specific details of your health.

**Stage 1 NSCLC**: Surgery to remove a portion of the lung may be all you need. Chemotherapy may also be recommended, especially if you’re at high risk of recurrence.

**Stage 2 NSCLC**: You may need surgery to remove part or all of your lung. Chemotherapy is usually recommended.

**Stage 3 NSCLC:** You may require a combination of chemotherapy, surgery, and radiation treatment.

**Stage 4 NSCLC** is particularly hard to cure. Options include surgery, radiation, chemotherapy, targeted therapy, and immunotherapy.

Options for small cell-lung cancer (NSCLC) also include surgery, [chemotherapy](https://www.healthline.com/health/chemotherapy), and radiation therapy. In most cases, the cancer will be too advanced for surgery.

Clinical trials provide access to promising new treatments. Ask your doctor if you’re eligible for a clinical trial.

Some people with advanced lung cancer choose not to continue with treatment. You can still choose palliative care treatments, which are focused on treating the symptoms of cancer rather than the cancer itself.

## 1.2.6 Home remedies for lung cancer symptoms

Home remedies and homeopathic remedies won’t cure cancer. But certain home remedies may help relieve some of the symptoms associated with lung cancer and side effects of treatment.

Ask your doctor if you should take dietary supplements and if so, which ones. Some herbs, plant extracts, and other home remedies can interfere with treatment and endanger your health. Be sure to discuss all complementary therapies with your doctor to make sure they’re safe for you.

Options may include:

* **Massage**: With a qualified therapist, massage can help relieve pain and anxiety. Some massage therapists are trained to work with people with cancer.
* **Acupuncture**: When performed by a trained practitioner, acupuncture may help ease pain, nausea, and vomiting. But it’s not safe if you have low blood counts or take blood thinners.
* **Meditation**: Relaxation and reflection can reduce stress and improve overall quality of life in cancer patients.
* **Hypnosis**: Helps you relax and may help with nausea, pain, and anxiety.
* **Yoga**: Combining breathing techniques, meditation, and stretching, yoga can help you feel better overall and improve sleep.

Some people with cancer turn to cannabis oil. It can be infused into cooking oil to squirt in your mouth or mix with food. Or the vapors can be inhaled. This may relieve nausea and vomiting and improve appetite. Human studies are lacking and laws for use of cannabis oil vary from state to state.

## 1.2.7 Diet recommendations for people with lung cancer

There’s no diet specifically for lung cancer. It is important to get all the nutrients your body needs.

If you’re deficient in certain vitamins or minerals, your doctor can advise you which foods can provide them. Otherwise, you’ll need a dietary supplement. Don’t take supplements without talking to your doctor because some can interfere with treatment.

Here are a few dietary tips:

* Eat whenever you have an appetite.
* If you don’t have a major appetite, try eating smaller meals throughout the day.
* If you need to gain weight, supplement with low sugar, high-calorie foods and drinks.
* Use mint and ginger teas to soothe your digestive system.
* If your stomach is easily upset or you have mouth sores, avoid spices and stick to bland food.
* If constipation is a problem, add more high-fiber foods.

As you progress through treatment, your tolerance to certain foods may change. So can your side effects and nutritional needs. It’s worth discussing nutrition with your doctor often. You can also ask for a referral to a nutritionist or dietician.

There’s no diet known to cure cancer, but a well-balanced diet can help you fight side effects and feel better.

## 1.2.8 Lung cancer and life expectancy

Once cancer enters the lymph nodes and bloodstream, it can spread anywhere in the body. The outlook is better when treatment begins before cancer spreads outside the lungs.

Other factors include age, overall health, and how well you respond to treatment. Because early symptoms can be easily overlooked, lung cancer is usually diagnosed in later stages.

Survival rates and other statistics provide a broad picture of what to expect. There are significant individual differences, though. Your doctor is in the best position to discuss your outlook.

Current survival statistics don’t tell the whole story. In recent years, new treatments have been approved for stage 4 non-small cell lung cancer (NSCLC). Some people are surviving much longer than previously seen with traditional treatments.

The following are the estimated [five-year survival rates](https://www.cancer.org/cancer/lungcancer-non-smallcell/detailedguide/non-small-cell-lung-cancer-survival-rates) for NSCLC by SEER stage:

* Localized: 60 percent
* Regional: 33 percent
* Distant: 6 percent
* All SEER stages: 23 percent

Small-cell lung cancer (SCLC) is very aggressive. For limited stage SCLC, the five-year survival rate is [14 percent Trusted Source](https://www.cancer.gov/types/lung/hp/small-cell-lung-treatment-pdq#link/_25_toc). Median survival is 16 to 24 months. Median survival for extensive stage SCLC is six to 12 months.

Long-term disease-free survival is rare. Without treatment, median survival from diagnosis of SCLC is only two to four months.

The relative five-year survival rate for mesothelioma, a type of cancer caused by asbestos exposure, is 5 to 10 percent.

## 1.2.9 Facts and statistics about lung cancer

Lung cancer is the most common cancer in the world. According to the [American Lung Association](https://www.lung.org/lung-health-and-diseases/lung-disease-lookup/lung-cancer/learn-about-lung-cancer/lung-cancer-fact-sheet.html), there were 2.1 million new cases in 2018, as well as 1.8 million deaths from lung cancer.

The most common type is non-small cell lung cancer (NSCLC), accounting for 80 to 85 percent of all cases, according to the [Lung Cancer Alliance](https://lungcanceralliance.org/what-is-lung-cancer/types-of-lung-cancer/).

Small-cell lung cancer (SCLC) represents about 15 to 20 percent of lung cancers. At the time of diagnosis, 2 out of 3 people with SCLC are already in the extensive stage.

Anyone can get lung cancer, but smoking or exposure to secondhand smoke is linked to about [90 percent](https://www.lung.org/lung-health-and-diseases/lung-disease-lookup/lung-cancer/symptoms-causes-and-risk-factors/what-causes-lung-cancer.html) of lung cancer cases. According to the [Centers for Disease Control and Prevention (CDC)Trusted Source](https://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm" \t "_blank), cigarette smokers are 15 to 30 times more likely to get lung cancer than nonsmokers.

In the United States, each year about 7,300 people who never smoked die from lung cancer caused by secondhand smoke.

Former smokers are still at risk of developing lung cancer, but quitting can significantly lower that risk. Within 10 years of quitting, the risk of dying from lung cancer [drops by halfTrusted Source](https://www.cdc.gov/tobacco/data_statistics/sgr/50th-anniversary/pdfs/wynk-cancer.pdf).

Tobacco products contain more than 7,000 chemicals. At least 70 are known carcinogens.

According to the [US Environmental Protection Agency (EPA)](https://www.epa.gov/radon/health-risk-radon), radon is responsible for about 21,000 lung cancer deaths every year in the United States. About 2,900 of these deaths occur among people who have never smoked.

Black people are at [higher risk](https://www.lung.org/lung-health-and-diseases/lung-disease-lookup/lung-cancer/learn-about-lung-cancer/lung-cancer-fact-sheet.html) of developing and dying from lung cancer than other racial and ethnic groups.

Lung cancer remains a primary cause of death in both males and females worldwide(American Cancer Society, 2015). In order to improve the survival rate of lung cancer patients, it is essential to identify the characteristics of lung cancer. Lung cancer has been classified into two categories, namely non-small cell lung cancer and small cell lung cancer. Recent improvements in chemotherapy and radiation therapy (Baas et al, 2006) have resulted in the former being further classified into adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (Travis et al, 2015).

In clinical practice, it is often difficult to precisely differentiate adenocarcinoma and squamous cell carcinoma in terms of their morphological characteristics, thus requiring immunohistochemical evaluation. Cytodiagnosis is advantageous for the cytological evaluation of small cell carcinoma compared to histological specimen, often showing crushed small cell cancer cells. To obtain accurate diagnosis result, a combination of cytological evaluation and histopathological diagnosis is essential.

However, the cell morphology of these three types of cancer cells varies. Here we focused on the automated classification technique of cancer types using cytological images. Among the four major types of carcinoma, the large cell carcinoma is the easiest to detect because of its cell morphology. We therefore concentrate on the classification of adenocarcinoma, squamous cell carcinoma, and small cell carcinoma that are sometimes confused with each other in the cytological specimen.

In our previous studies (Teramoto et al, 2017; Teramoto et al, 2019), an automated classification scheme of lung cancer types in the cytology images using DCNN was developed. The original DCNN model was employed and trained using 15,000 cytological images, and the classification accuracy was found to be 71%. However, for clinical use, further improvement in the classification ability is necessary.

As an improvement method, increase in the number of data used for training can be mentioned. The lung cytological specimen has a three-dimensional structure in which cells overlap each other, acquisition of images by a uniform scanning method such as a whole slide scanner is difficult, and observation by a conventional microscope is mainstream. Therefore, collection of many cytological images is more difficult than that of histopathological images.

Transfer learning is a technique known for improving the performance of deep learning (Ravishankar et al, 2016). This is a method of diverting the majority of the deep learning architecture learned using a large number of natural images and performing another classification task. To further improve the accuracies, additional supervised classifiers (Amancio et al, 2014) have been utilized to support the DCNNs. In this study, we evaluated four pretrained DCNN architectures plus subsequent four supervised classifiers for comparison of classification accuracy of lung cancer tissue types.

Now-a- day’s lung cancer has become one of the leading causes of cancer related deaths of humans. So it is necessary for the trained radiologists to identify the cancer accurately in the lung in the early stage as possible to reduce the deaths of humans [2] .So this is very complicated job to detect the lung nodules which is affected by cancer and non-cancer in its early stages. In the past days, radiologist were manually analysing the CT scans images of lung, looking for the potential nodules and identify the cancerous and non- cancerous in those nodules.

This process needs a high knowledge about the lung nodules which is very tedious and time consuming. So this can resolved by introducing the computer aided diagnosis (CAD) system to detect the lung nodules and classification of nodules as either cancerous or non- cancerous. This system will be as second opinion to the radiologist to detect and analyse the nodules of the lung. US, lung cancer has become the cause for the cancer related deaths. Approximately 229,447 new cases of lung cancer were there and in that 159,124 related deaths. So early diagnosis of lung cancer can improve the effectiveness of treatment and increase the chance of the survival of patient’s.

The lung cancer can be identified by the non-invasive imaging modalities those are computed tomography (CT), contrast-enhanced computed tomography (CE-CT), low-dose computed tomography (LDCT) and positron emission tomography (PET)[3]. In the past years, CAD system has been developed for both the nodule segmentation and also classification of lung nodules as cancerous or non-cancerous. But this system for segmentation will generate a very good detection of lung nodules but however in the process they lead to many false positive, when actual positive compared to this false positive segmentation the ratio can be in hundreds[3].

So this system can lead many false positive, when dealing with lung cancer. It’s better to label anything that looks like nodule in the image, and the radiologists need to fallow the manual methods of detecting the lung cancer. Traditional algorithms of image processing used to detect unique features of images. So this requires the hand crafted features has to be created which learns the features at manual process. So in this method it is very complex to differentiate the features of cancerous and non-cancerous nodules.so the deep learning can avoid all these problems and are capable of tackling with the problems like image recognition, video recognition, speech recognition and natural language processing etc [7].

Manual feature extraction requires expert knowledge of the lung cancer to the designer. Deep learning will have the properties to learn all the features in the images. In particular, Convolutional Neural Network (CNN) extracts the features of input images by using one or more layers of convolution and subsampling or max-pooling layer that are in the hierarchical manner. The general CNN consists of three layers – Convolution, Max-polling and Fully connected. The classification of images is done by CNN by extracting features in each layer and producing a final model.

When other test image is given to model compares the features of both and classify the images to different classes by the accuracy. In this paper, we have used the CNN for the classification of CT scan images of lung cancer as cancerous or non-cancerous. Lung cancer is prominent cancer among both men and women, making up almost 25% of all cancer deaths [1]. 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 [2].

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 [3], [4], [5], and defines the types and subtypes of lung cancers [6]. For pathologists and other medical professionals diagnosing lung cancer and the types is a timeconsuming process. There is a significant change the cancer types are misdiagnosed, which directs to incorrect treatment and may cost patients' lives.

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 [7][8]. In previous research papers, most of the authors considered using x-rays, CT scans images with machine learning techniques such as Support Vector Machine (SVM), Random Forest (RF), Bayesian Networks (BN), and Convolutional Neural Network (CNN) for lung cancer detection and recognition purpose. Some papers also considered using histopathological images, but they distinguish between carcinomas and non- carcinomas images and with lower accuracy. This research paper has considered using Convolutional Neural Network (CNN) architecture to classify the benign, Adenocarcinoma, and squamous cell carcinomas. We have not found other papers using the CNN model to classify only the given three different histopathological images and the given model's accuracy.

**CHAPTER 2**

**LITERTURE SURVEY**

1. **Jiang, H., Qian, W., Gao, M., Li, Y. “ An automatic detection system of lung nodule based on multigroup patch-based deep learning network” IEEE Journal of Biomedical and Health Informatics,22(4):1227-1237.**

High-efficiency lung nodule detection dramatically contributes to the risk assessment of lung cancer. It is a significant and challenging task to quickly locate the exact positions of lung nodules. Extensive work has been done by researchers around this domain for approximately two decades. However, previous computer-aided detection (CADe) schemes are mostly intricate and time-consuming since they may require more image processing modules, such as the computed tomography image transformation, the lung nodule segmentation, and the feature extraction, to construct a whole CADe system.

It is difficult for these schemes to process and analyze enormous data when the medical images continue to increase. Besides, some state of the art deep learning schemes may be strict in the standard of database. This study proposes an effective lung nodule detection scheme based on multigroup patches cut out from the lung images, which are enhanced by the Frangi filter. Through combining two groups of images, a four-channel convolution neural networks model is designed to learn the knowledge of radiologists for detecting nodules of four levels.

This CADe scheme can acquire the sensitivity of 80.06% with 4.7 false positives per scan and the sensitivity of 94% with 15.1 false positives per scan. The results demonstrate that the multigroup patch-based learning system is efficient to improve the performance of lung nodule detection and greatly reduce the false positives under a huge amount of image data.

1. **Disha Sharma, Gagandeep Jindal, “Identifying Lung Cancer Using Image Processing Techniques”, International Conference on Computational Techniques and Artificial Intelligence, pp. 116-120, 2011.**

Lung cancer is the dangerous disease where more number of death occurs in both men and women hence identifying such a disease is a challenging task. The identification of lung cancer tumor in the early stage will save the life of more number of patients by proper prognosis and treatment hence to decrease the death rate and increase the survival rate its identification and classification is necessary.

Machine learning technique, opens the door to predict, identify and classify this disease however deep learning under machine learning brings a wide way to analyze and evaluate the features of tumor from its CT images. The system proposed in this paper provides a clear and accurate classification VGG-16 model and its advanced model where more number of hidden layers was utilized which is an Improved VGG-16 model. The system developed were trained using LIDC-IDRI CT image dataset and it is evident from the experiments that the classification accuracy of Improved VGG-16 model is 97% and VGG-16 model is 86% with a very less false positive rates of 0.0567 and 0.12.

1. **Farzad Vasheghani Farahani “Lung Nodule diagnosis from CT images Based on Ensemble Learning.” IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology , 2015.**

Early detection of cancer is the most promising way to enhance a patient's chance for survival. This paper presents a computer-aided classification method using computed tomography (CT) images of the lung based on ensemble of three classifiers including MLP, KNN and SVM. In this study, the entire lung is first segmented from the CT images and specific features like Roundness, Circularity, Compactness, Ellipticity, and Eccentricity are calculated from the segmented images.

These morphological features are used for classification process in a way that each classifier makes its own decision. Finally, majority voting method is used to combine decisions of this ensemble system. The performance of this system is evaluated using 60 CT scans collected by Lung Image Database Consortium (LIDC) and the results show good improvement in diagnosing of pulmonary nodules.

1. **Atsushi Teramoto at el. “Automated classification of benign and malignant cells from lung cytologicalimages using deep convolutional neural network” Informatics in medicine unlocked 16,2019,100205 elsvier.**

It is essential to accurately diagnose and classify histological subtypes into adenocarcinoma (ADC), squamous cell carcinoma (SCC), and small cell lung carcinoma (SCLC) for the appropriate treatment of lung cancer patients. However, improving the accuracy and stability of diagnosis is challenging, especially for non-small cell carcinomas. The purpose of this study was to compare multiple deep convolutional neural network (DCNN) technique with subsequent additional classifiers in terms of accuracy and characteristics in each histology.

Methods: Lung cancer cytological images were classified into ADC, SCC, and SCLC with four fine-tuned DCNN models consisting of AlexNet, GoogLeNet (Inception V3), VGG16 and ResNet50 pretrained by natural images in ImageNet database. For more precise classification, the figures of 3 histological probabilities were further applied to subsequent machine learning classifiers using Naïve Bayes (NB), Support vector machine (SVM), Random forest (RF), and Neural network (NN).

Results: The classification accuracies of the AlexNet, GoogLeNet, VGG16 and ResNet50 were 74.0%, 66.8%, 76.8% and 74.0%, respectively. Well differentiated typical morphologies were tended to be correctly judged by all four architectures. However, poorly differentiated non-small cell carcinomas lacking typical structures were inclined to be misrecognized in some DCNNs. Regarding the histological types, ADC were best judged by AlexNet and SCC by VGG16. Subsequent machine learning classifiers of NB, SVV, RF, and NN improved overall accuracies to 75.1%, 77.5%, 78.2%, and 78.9%, respectively.

Conclusion: Fine-tuning DCNNs in combination with additional classifiers improved classification of cytological diagnosis of lung cancer, although classification bias could be indicated among DCNN architectures.

1. **Xin-Yu Jin, Yu-Chen Zhang, Qi-Liang Jin “Pulmonary Nodule Detection Based on CT Images Using Convolution Neural Network.” 9Th International Symposium on Computational Intelligence and Design. 2016.**

Pulmonary nodule is a common lung disease, which can be prone to misdiagnosis and missed diagnosis. With the extensive application of CT technology, doctor's diagnostic efficiency has been greatly improved. However, the amount of CT image data is relatively large. Radiologists have to take a lot of time to read these images, and easy to overlook some minor lesions. Computer aided detection technology is an effective way to improve the efficiency and quality of the doctor's diagnosis.

This paper put forward a kind of lung segmentation method based on morphology and statistic of size of the image area, while effectively eliminating the influence of the trachea to pulmonary parenchyma image segmentation. We also propose a method of region of interest(ROI) extraction based on morphology and circular filter, reducing the number of false positive and trying to retain the integrity of the ROI form. Finally, we have realized a reliable lung nodules compute aided diagnosis application on CT image, using Convolution neural network.

1. **Ryota Shimizu, Shusuke Yanagawa, Yasutaka Monde, Hiroki Yamagishi,Mototsugu Hamada, Toru Shimizu, and Tadahiro Kuroda “Deep Learning Application Trial to Lung Cancer Diagnosis for Medical Sensor Systems” International Symposium on Computers and Communications, 2016.**

Personal and easy-to-use health checking system is an attractive application of sensor systems. Sensing data analysis for diagnosis is important as well as preparing small and mobile sensor nodes because sensing data include variations and noises reflecting individual difference of people and sensing conditions. Deep Neural Network, or Deep Learning, is a well-known method of machine learning and it is effective for feature extraction from pictures. Then, we thought Deep Learning also can extract features from sensing data. In this paper, we tried to build a diagnosis system of lung cancer based on Deep Learning.

Input data of the system was generated from human urine by Gas Chromatography Mass Spectrometer (GC-MS) and our system achieved 90% accuracy in judging whether the patient had lung cancer or not. This system will be useful for pre- and personal diagnosis because collecting urine is very easy and not harmful to human body. We are targeting installation of this system not only to gas chromatography systems but also to some combination of multiple sensors for detecting gases of low concentration.

1. **Po-Whei Huang, Phen-Lan Lin, Cheng-Hsiung Lee, C. H. Kuo, “A Classification System of Lung Nodules in CT Images Based on Fractional Brownian Motion Model”, IEEE International Conference on System Science and Engineering, July 2016.**

In this paper, we present a classification system for differentiating malignant pulmonary nodules from benign nodules in computed tomography (CT) images based on a set of fractal features derived from the fractional Brownian motion (fBm) model. In a set of 107 CT images obtained from 107 different patients with each image containing a solitary pulmonary nodule, our experimental result show that the accuracy rate of classification and the area under the Receiver Operating Characteristic (ROC) curve are 83.11% and 0.8437, respectively, by using the proposed fractal-based feature set and a support vector machine classifier. Such a result demonstrates that our classification system has highly satisfactory diagnostic performance by analyzing the fractal features of lung nodules in CT images taken from a single post-contrast CT scan.

1. **Vaishali C. Patil, Shrinivas R. Dhotre, “Lung Cancer Detection from Images of Computer Tomography Scan”, International Journal of Advanced Research in Computer and Communication Engineering , Vol. 5, Issue 7, July 2016.**

Lung cancer is the dangerous disease where more number of death occurs in both men and women hence identifying such a disease is a challenging task. The identification of lung cancer tumor in the early stage will save the life of more number of patients by proper prognosis and treatment hence to decrease the death rate and increase the survival rate its identification and classification is necessary. Machine learning technique, opens the door to predict, identify and classify this disease however deep learning under machine learning brings a wide way to analyze and evaluate the features of tumor from its CT images.

The system proposed in this paper provides a clear and accurate classification VGG-16 model and its advanced model where more number of hidden layers was utilized which is an Improved VGG-16 model. The system developed were trained using LIDC-IDRI CT image dataset and it is evident from the experiments that the classification accuracy of Improved VGG-16 model is 97% and VGG-16 model is 86% with a very less false positive rates of 0.0567 and 0.12.

1. **Ailton Felix, Marcelo Oliveira, Aydano Machado, Jose Raniery, “Using 3D Texture and Margin Sharpness Features on Classification of Small Pulmonary Nodules” ,29th SIBGRAPI Conference on Graphics, Patterns and Images, 2016.**

The lung cancer is the reason of a lot of deaths on population around the world. An early diagnosis brings a most curable and simpler treatment options. Due to complexity diagnosis of small pulmonary nodules, Computer-Aided Diagnosis (CAD) tools provides an assistance to radiologist aiming the improvement in the diagnosis. Extracting relevant image features is of great importance for these tools.

In this work we extracted 3D Texture Features (TF) and 3D Margin Sharpness Features (MSF) from the Lung Image Database Consortium (LIDC) in order to create a classification model to classify small pulmonary nodules with diameters between 3-10mm. We used three machine learning algorithm: k-Nearest Neighbor (k-NN), Multilayer Perceptron (MLP) and Random Forest (RF). These algorithms were trained by different set of features from the TF and MSF. The classification model with MLP algorithm using the selected features from the integration of TF and MSF achieved the best AUC of 0.820.

1. **Rotem Golan, Christian Jacob, Jorg Denzinger, “Lung nodule detection in CT images using deep convolutional neural networks” International Joint Conference on Neural Networks , 2017.**

Lung cancer is a hazardous disease that many deaths were occur in both men and women from this deadly disease. Hence suitable mechanism should be adopted to detect and identify this disease in the initial stage to save the life of large number of peoples suffering from lung cancer. The identification of lung cancer tumor in the early stage, proper prognosis and treatment will decrease the death rate and increase the survival rate. Machine learning technique used to predict, identify and classify this disease however deep learning under machine learning brings a wide way to analyze and evaluate the features of tumor from its CT images.

The system proposed in this paper provides an accurate classification AlexNet model and its advanced model where more number of hidden layers was utilized which is an Enhanced AlexNet model. The system developed were trained using LIDC-IDRI CT image dataset and it is evident from the experiments that the classification accuracy of Enhanced AlexNet model is 99% and AlexNet model is 97% with a very less false positive rates of 0.0196 and 0.0392

1. **Sri Widodo, Ratnasari Nur Rohmah, Bana Handaga, “Classification Of Lung Nodules And Arteries In Computed Tomography Scan Image Using Principle Component Analysis” 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, 2017.**

There is still a lack of a good method of diagnosingpulmonary nodules in CT Scan automatically, causing medicalstaff to observe a 2-D CT Scan data manually and interpretingdata one by one. This procedure is course less effective. Inaddition, lung specialists may differ in determining pulmonarynodules. The purpose of this research is to classify pulmonarynodules and artery automatically on chest Ct Scan image usingPrinciple Component Analysis (PCA).

This study includes 3steps. The first is lung organ segmentation using ActiveAppearance Model (AAM). The second step is segmentation ofcandidate nodules using morphological math. While the last stepis classification of pulmonary nodules and artery using PrincipleComponent Analysis method. The output from classificationprocess is image of nodule and artery. Results of testing, obtainedthe performance of classification system accuracy is 90%.

1. **Ravindranath K , K Somashekar, “Early Detection of lung cancer by nodule extraction – A Survey”, International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques , 2017.**

Early identification of lung cancer includes detection of uncertain nodules and classifying them into different condition of disease. The identification stage includes pattern matching and confirmation to increase accuracy, performed by fuzzy logic, support vector machine, statistical classifiers. The categorization stage involves matching characters (texture, shape and density) of the detected nodules to characters of normal cells (texture, shape and density) of nodules with known condition of disease (confirmed by sample extraction techniques).

The nodule detection is mainly considered as it plays an important role in cancer detection nodules extracted are classified using neural network classifiers to differentiate between normal and abnormal lung cancer.

1. **Rui Xu, Jiao Pan, Xinchen Ye, Yasushi Hirano, Shoji Kido, Satoshi Tanaka “A Pilot Study to Utilize a Deep Convolutional Network to Segment Lungs with Complex Opacities” Chinese Automation Congress , 2017.**

Lung segmentation on CT images is the first step of computer-aided system for lung diseases. Because of the complex opacities inside lung regions, the segmentation of lungs with severe diseases is not a trivial work. In this paper, we proposed a deep convolutional neural network based algorithm to segment lungs with complex opacities on CT images. This end-to-end network is composed of several convolutional layers, max-pooling layers and de-convolutional layers.

We evaluated the proposed method by using 7 CT images with mild lung diseases and 42 CT images with severe diffuse lung disease including 6 kinds of complex opacities that are consolidation, ground grass opacity, honeycombing, emphysema, nodular opacity and reticular ground grass opacity. The jaccard index of the proposed method is 0.944 ± 0.066, which is higher than the one of a widely used lung segmentation method based on basic image processing techiniques [3] (0.889 ± 0.180). This pilot study gave very promising results for the utilization of deep neural network to segment pathological lungs on CT images.

1. **Anna Poreva, Yevgeniy Karplyuk, Valentyn Vaityshyn, “Machine Learning Techniques Application for Lung Diseases Diagnosis” , 5th IEEE Workshop on Advances in Information, Electronic and Electrical Engineering , 2017.**

The article considers the basic methods of machine learning for applying them to the task of the lungs sounds classifying. A number of signal parameters were obtained on the basis of the lungs sounds set. The task of the study was to classify sounds using five different machine learning methods.

It was also necessary to determine from a number of signal parameters those that give the highest accuracy. Thus the seven most diagnostically valuable parameters of lung sounds were found. The results showed that two methods of machine learning - the method of reference vectors and the decision tree method - have the best accuracy. Thus this classification technique can serve as an auxiliary tool for a pulmonary physician to diagnosis.

1. **Pratiksha Hattikatti, “Texture based Interstitial Lung Disease Detection using Convolutional Neural Network”, International Conference on Big Data, IoT and Data Science , 2017.**

Large range of lung texture patterns of disease can be observed in CT scan images. These images are the intermixed of various patterns and hence it becomes very difficult for Radiologist to differentiate between them and diagnose the disease. One way of solving this issue is use of Convolutional neural networks (CNN). CNN is generally used for pattern classification and image recognition systems.

They have achieved less error on the database; image classification using CNN was surprisingly fast. Interstitial lung disease is a term which includes different types of lung disease. Interstitial lung diseases affect the interstitium i.e. the part of the lung's anatomic structure. Lung tissue characterization is essential parts of a computer aided diagnosis (CAD) system for detection of interstitial lung diseases (ILDs).

Thus using CNN, interstitial lung disease detection gives accurate result. The proposed system is consists of CNN having 7 layers with Local binary pattern (LBP) as feature extractor. The execution of classification exhibited the capability of CNNs in analyzing lung patterns. The CT Scan Images used in this study are officially verified by the certified Radiologist.

**CHAPTER 3**

**METHODOLOGY USED**

1. **Convolutional Neural Networks(CNN)**

A Convolutional Neural Network (CNN) is a type of feed forward neural network which is inspired by biological visual system models [15], where the individual neurons are lined in such a way that they respond to overlapping regions in its receptive field and continues to be reliable with the modern perceptive of the structure of the image system [21]. When neurons with the same parameters are applied on overlapping regions of the previous layer, at different locations, a form of translational invariance is obtained.

This allows CNNs to detect objects in their receptive field in a way that is invariant to their size, location, orientation, and other visual properties. Besides this restricted connectivity of CNNs minimizes the computational requirements of training compared to fully connected neural networks [26]. The architecture of a Convolutional neural network as shown in the figure.1 is a multi-layered feed-forward neural network, made by stacking many hidden layers on top of each other in sequence.

It is this sequential design that allows Convolutional neural networks to learn hierarchical features. The hidden layers are typically convolutional layers followed by activation layers, some of them followed by pooling layers.

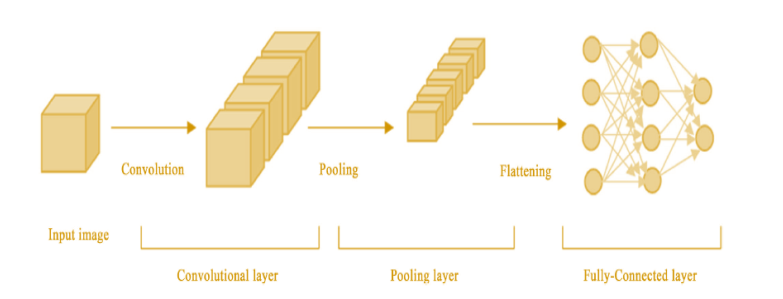


Figure 1: Architecture of CNN

Figure 1 Represents the Basic Blocks a Architecture of CNN comprise of three major layers such as an input, Convolutional, pooling and fully-connected layers.

1 Convolutional layer: This layer accepts the input images of specified size suitable for the network training, which is then, translated into feature maps by using filters or Convolutional kernels. The filters used in this layer are moved through the dimensions.

2 Pooling layer: The important function of this layer is to downsize the matrix and to minimize the parameters hence this layer does down-sampling from the feature maps of Convolutional layers. This layer calculates maximum value or weighted average by moving filters across the output of Convolutional layer.

3 Fully-connected layer: The objective of this layer is to classify the resulting images of the previous two layers into a label.

Since this layer utilizes the sofmax layer so as to find the probabilities of values in between 0 and 1. Besides it uses Batch normalization to enhance the training rate and to minimize over fitting. The Identification of lung cancer using Deep CNN, comprise of two categories. The first category does preprocessing functionalities suitable to train and process the images in DCNN, thus feature extraction can be performed and second category performs the classification of input CT images where it identifies the type of nodule as benign or malignant.

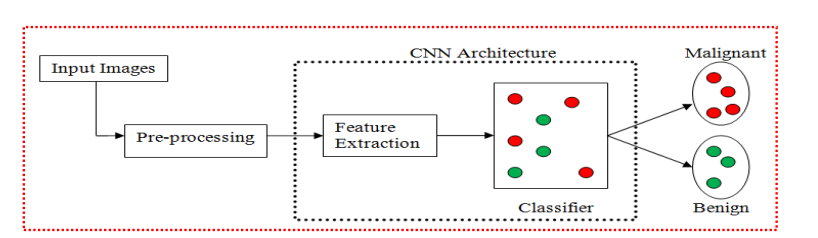


Fig.2. DCNN for lung cancer nodule identification.

The Architecture of Deep CNN as presented in Figure.4 with block diagram shown in figure.2 which comprise of 25 layers such as Eight convolution layer, eight ReLu layers, seven max- pooling layers and one fully-connected layer with softmax layer. From the Figure 1, it is seen that the first layer input layer consists of images of suitable network size. The convolution layer is the second layer which accepts the input images of size 256 x 256 size and then it converts the input images into feature maps by applying the Convolutional kernel of size 3 x 3.

Each Convolutional layer followed by activation rectified linear unit (Re-Lu) layers for converting image to feature maps. The next layer after Convolutional layer is max pooling layer, the size of the filter or kernel used is 2 x 2 and 2 pixels stride. Finally the outputs of previous two layers are applied to the fullyconnected layer to generate a 1024 output dimensions. The resultant outputs are then applied to another fully connected layer followed by softmax layer. The fully connected layer along with softmax layer gives the classification probability of benign or malignant cancer type. From the figure 3(a) and 3(b), the output sample of images after classification into malignant or benign are presented from the experimental work.

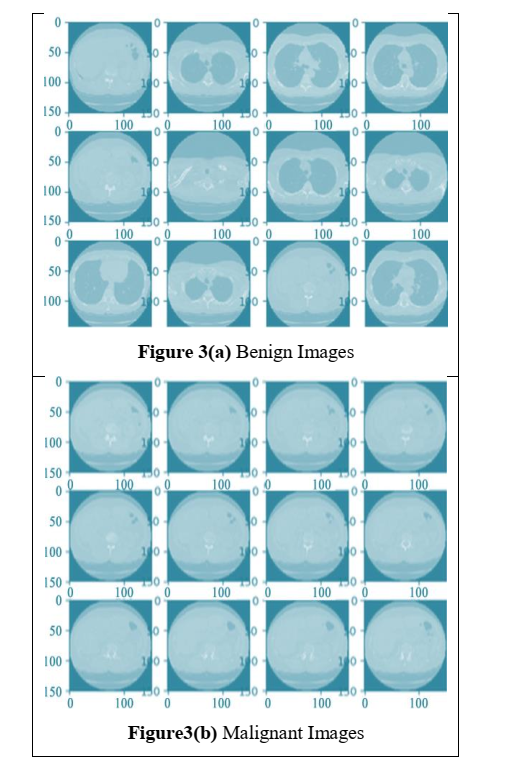


Figure.3 Experimental Images

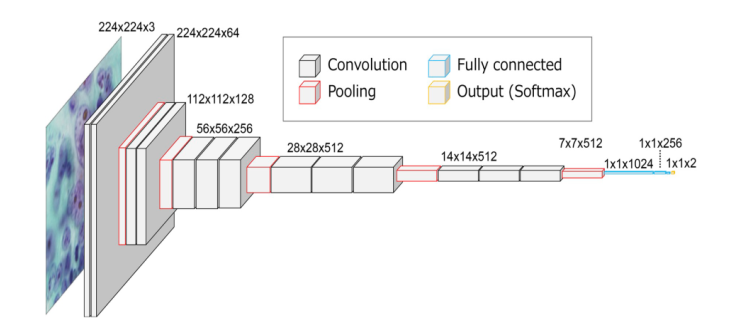


Fig.4. Architecture of Deep CNN

1. Training a deep CNN Back-propagation algorithms are used to train the Deep CNN using CT images of size 256x256x3. It consists of two phase, training phase and testing phase. In the first phase DCNNs are trained using CT images where 900 images are being used to train the network for the classification of lung as either cancerous or non-cancerous. In the testing phase an image unknown to the network is applied as input to classify as cancerous or noncancerous. For minimum loss of features images are trained and tested in the DICOM format itself by modifying the network parameters such that it can take DICOM images. The proposed designed network accuracy can be achieved by suitable evaluation.
2. Performance Measures parameters The performance of a medical image can be analyzed, by using performance evaluation parameters such as Accuracy, Loss and Computation Time.

• Accuracy: it is one of the important performance measure parameter to evaluate the model. It gives correctly classified number of pixels from the given image.

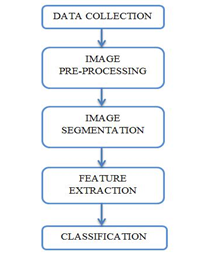
• Loss Function : The error in the neural network can be predicted by Loss which is calculated by Loss function. This is another performance measure parameter of the network.

• Computation Time: Time required for the process to complete its computation or its operations. If the process is simple then time taken for processing is less compared to the complex process whose computation time is more

**CHAPTER 4**

**PROPOSED METHOD**

The proposed system for lung cancer detection in CT images is shown with the help of a flowchart in figure 1.The methodology is carried out in five main steps and each step of this system is discussed in detail in section below.

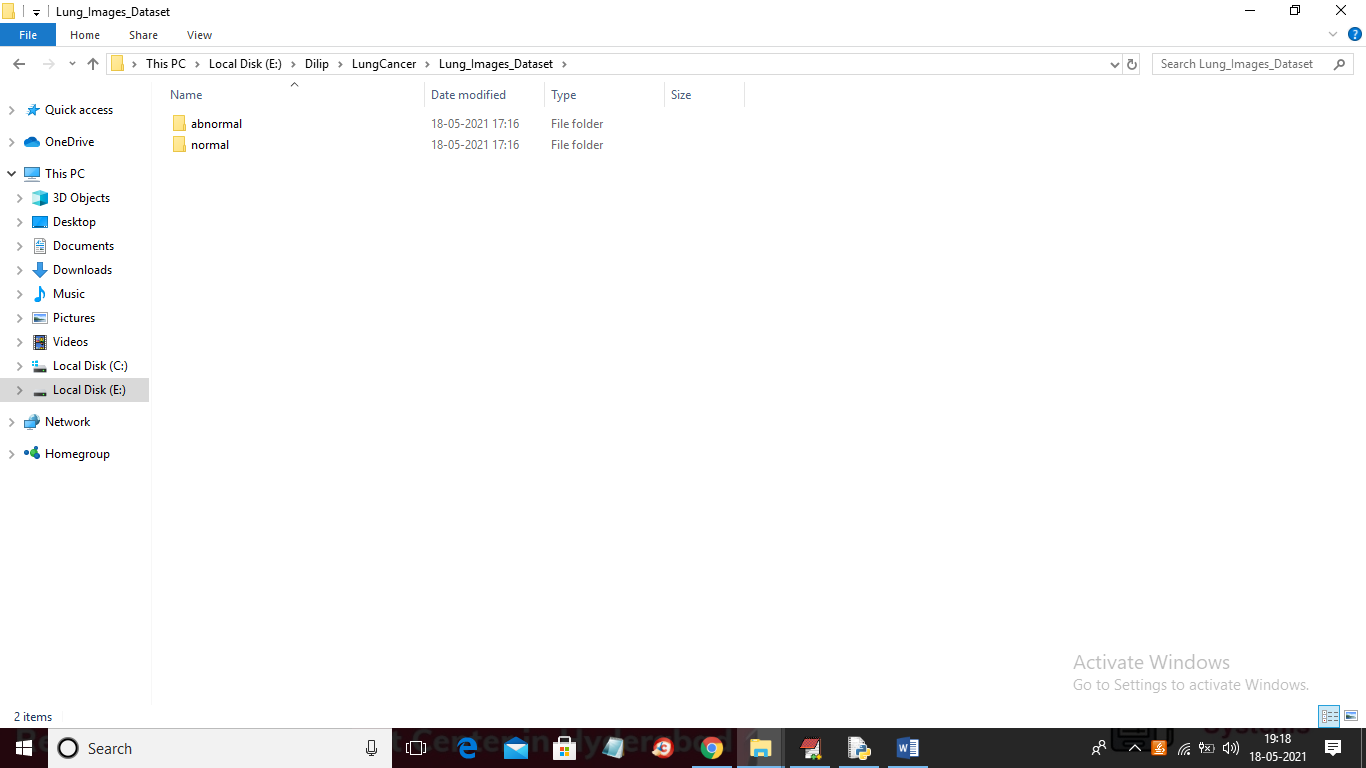
****

**Fig. 1.** Block diagram of the proposed system

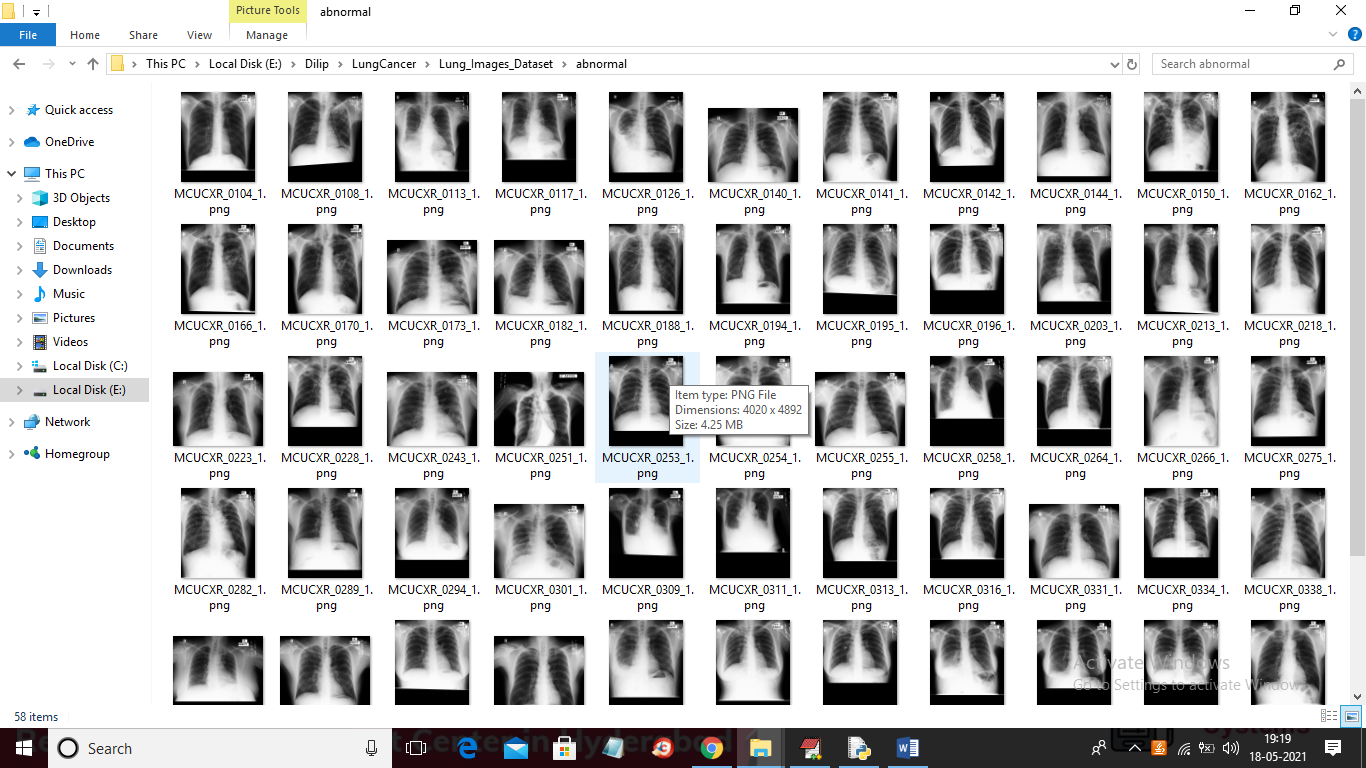
**CHAPTER 5**

**RESULT**

In this project we are using CNN algorithm to detect Lung cancer from CT-SCAN images and to train CNN we have CT-SCAN images dataset and this dataset saved inside ‘Lung\_Images\_Dataset’ folder and below screen shots showing images from dataset



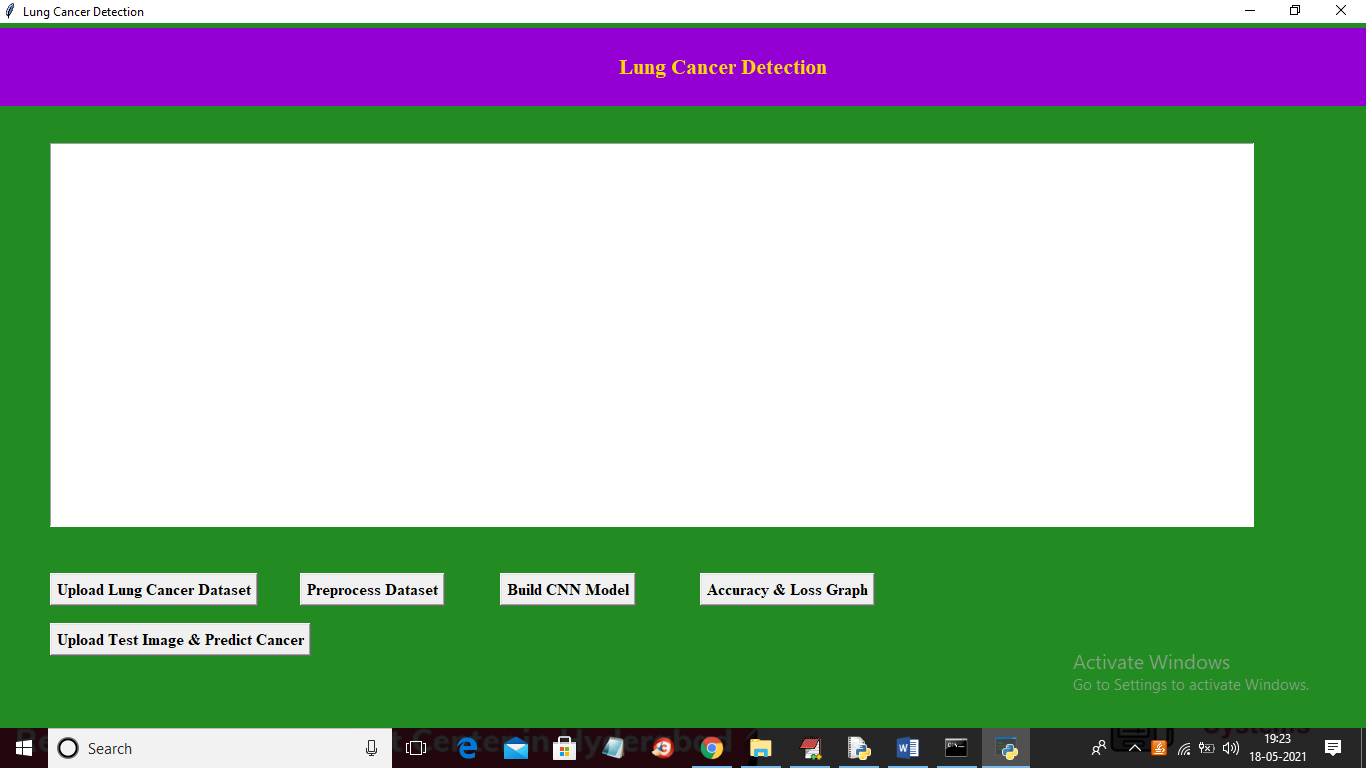
In above screen we have two folders where one folder contains NORMAL CT SCAN images and other folder contains ABNORMAL and you can go inside any folder to view those images



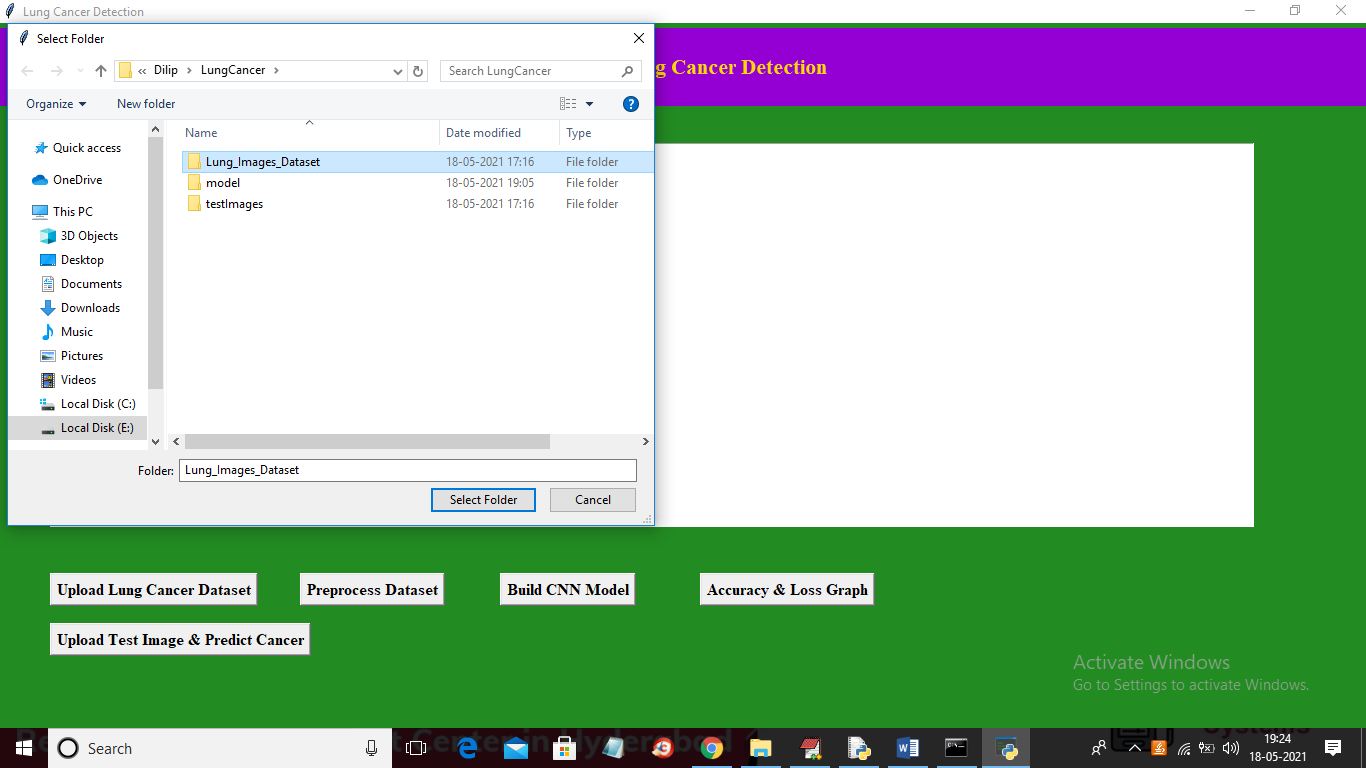
In above screen displaying few images from ABNORMAL folder and we used above images to train CNN algorithm and after training CNN we can upload test images and then CNN will predict whether CT\_SCAN contains normal or abnormal tumour

SCREEN SHOTS

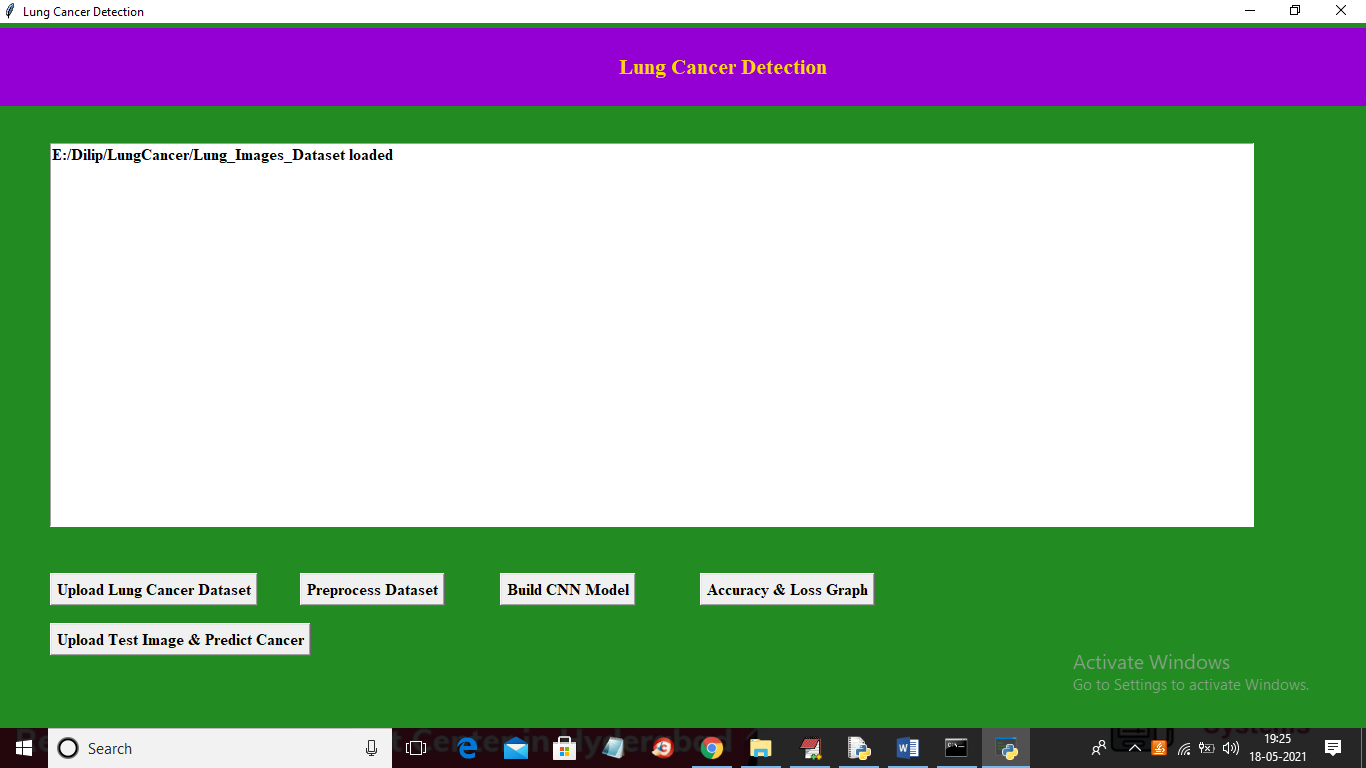
To run project double click on ‘run.bat’ file to get below screen



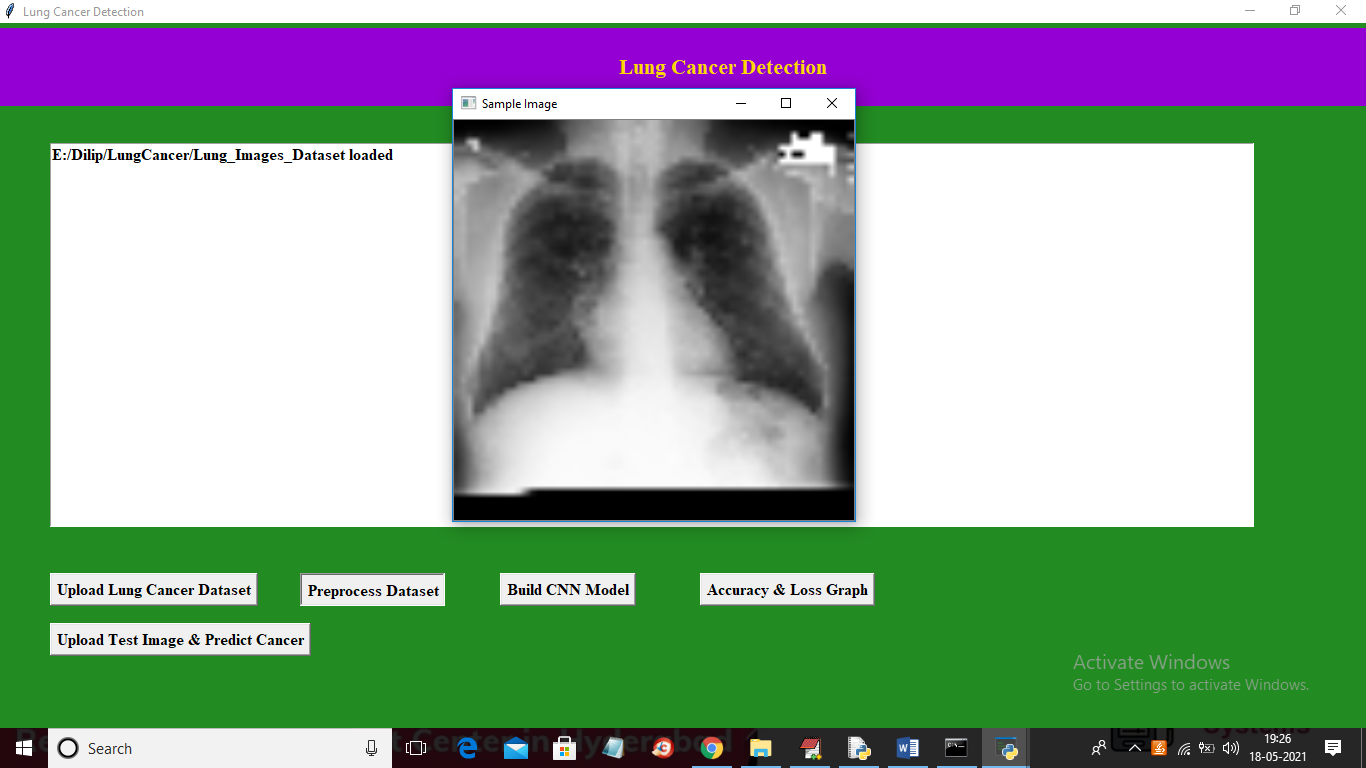
In above screen click on ‘Upload Lung Cancer Dataset’ button to upload CT-SCAN images



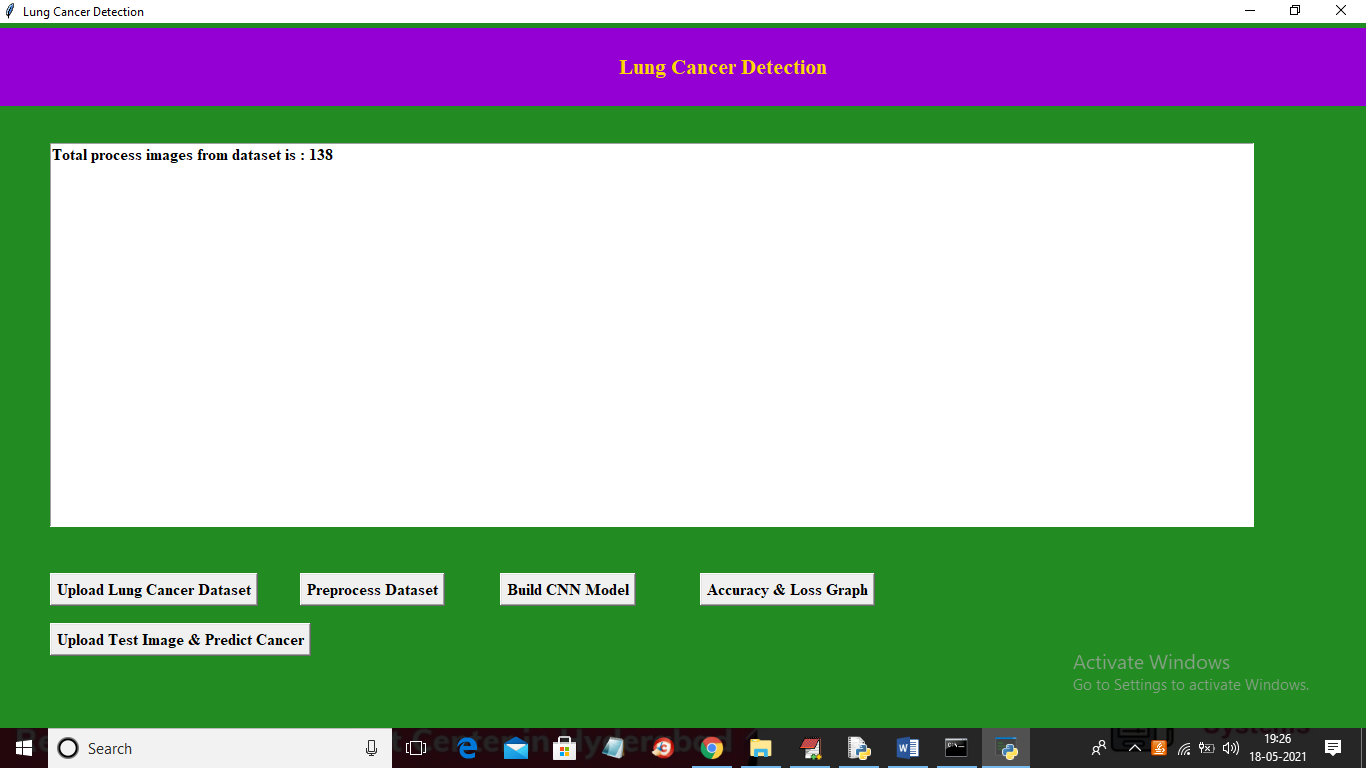
In above screen selecting and uploading ‘Lung\_Image\_Dataset’ folder and then click on ‘Select Folder’ button to load images and to get below screen



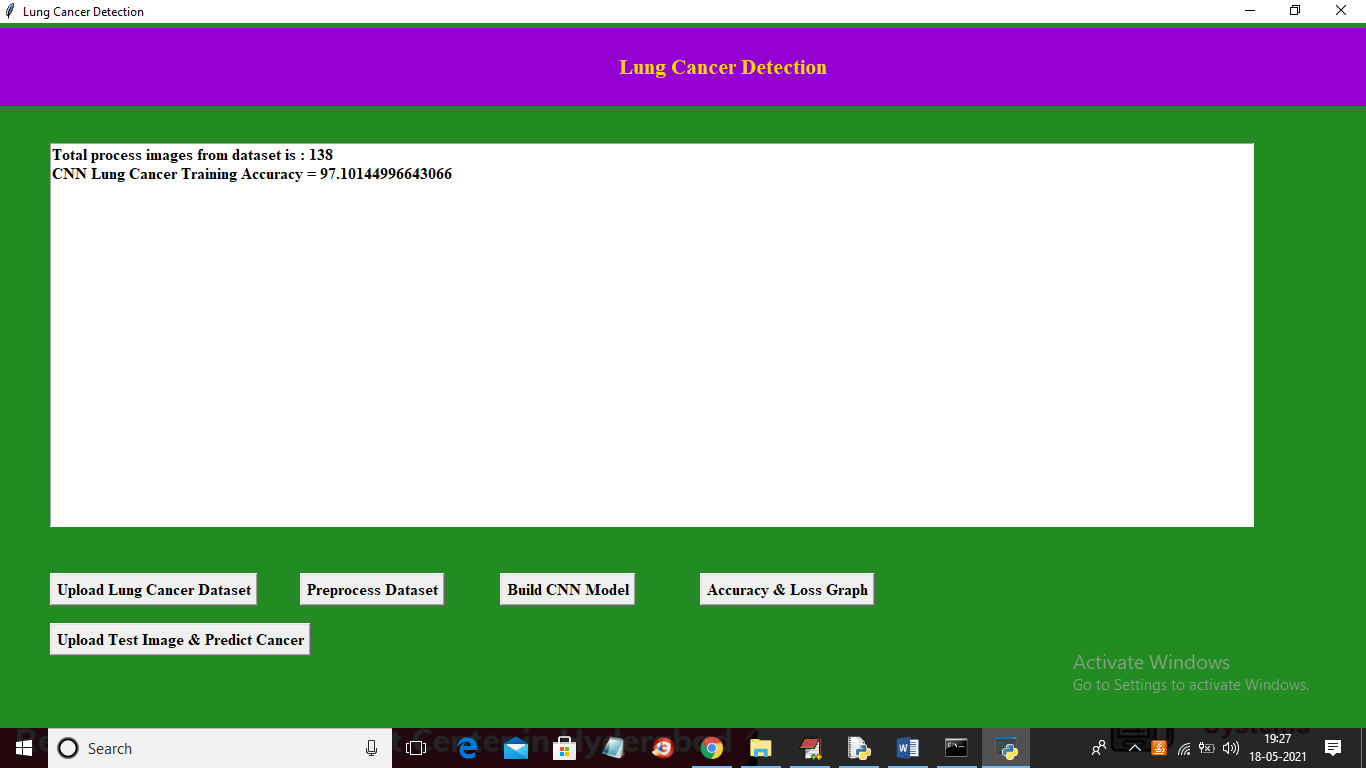
In above screen dataset loaded and now click on ‘Preprocess Dataset’ button to convert all images into colour format and resize them into equal sizes so CNN can accept those images



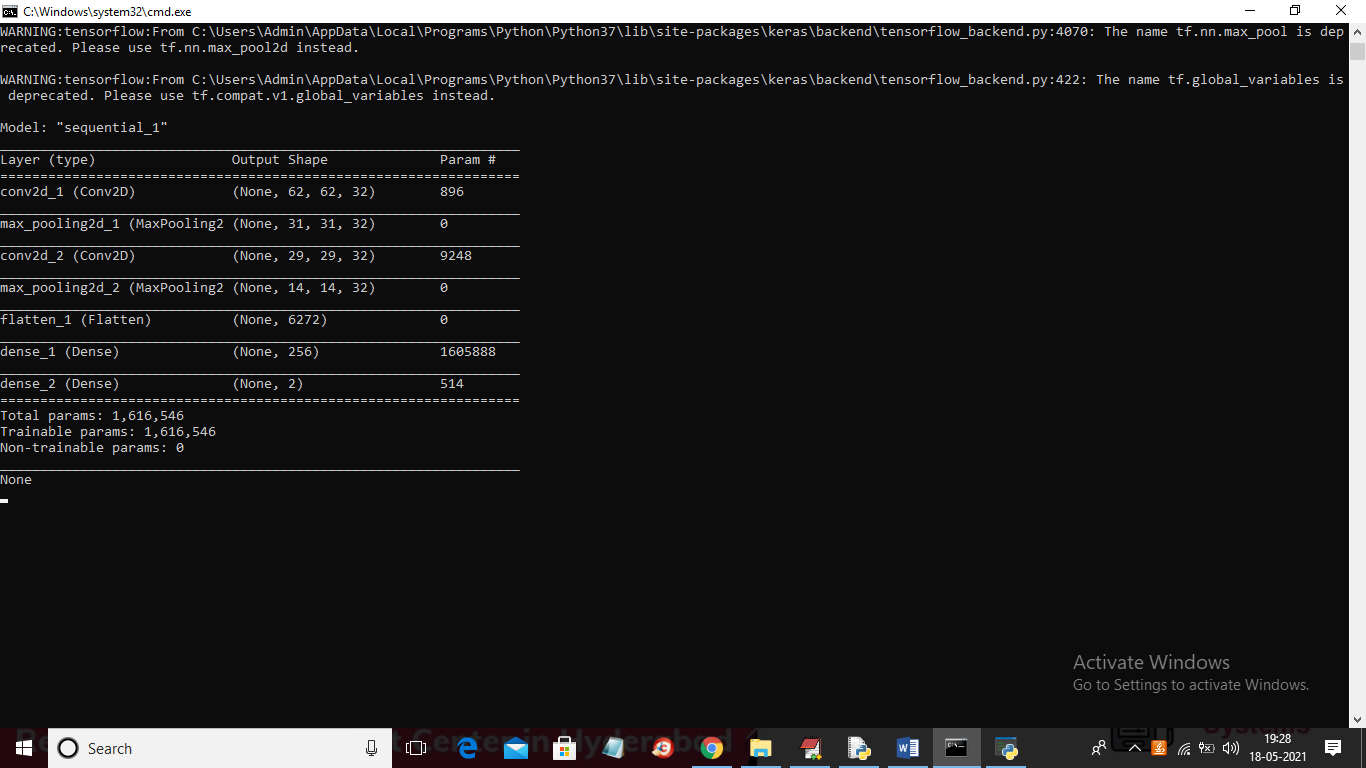
In above screen application process all images and then I am displaying one sample image to confirm all images loaded properly and now close above image to get below screen



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



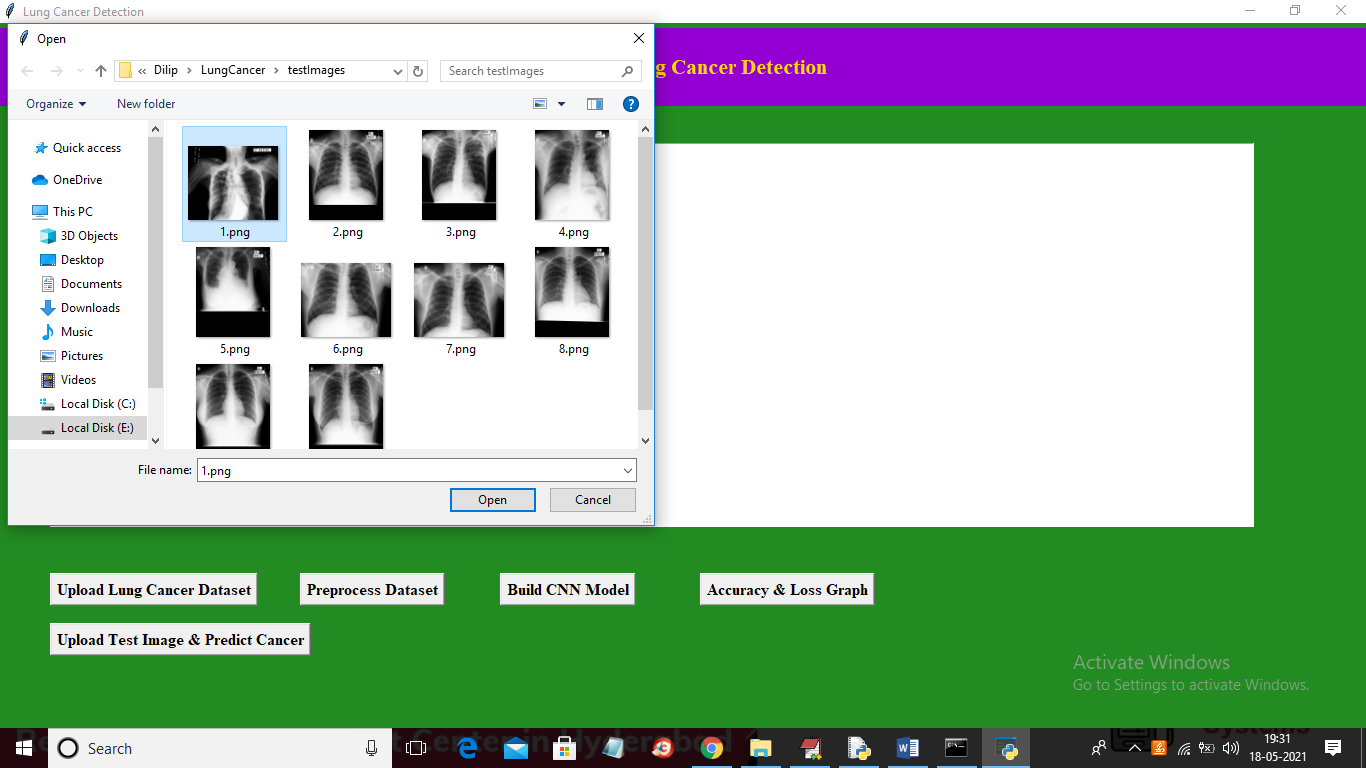
In above screen CNN model generated and we got its accuracy as 97% and in below screen we can see CNN architecture



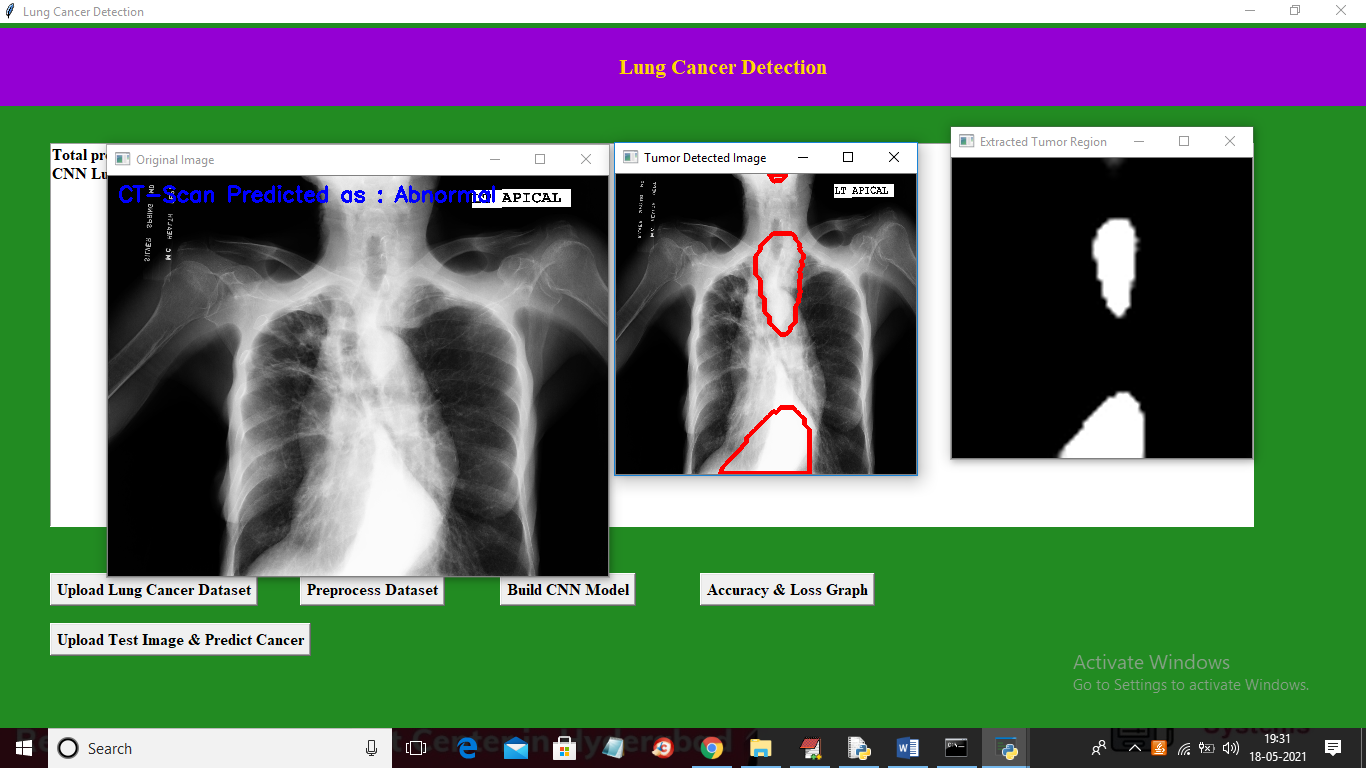
In above console to train CNN we created multiple layers where first layer process images of size 62 X 62 and second layer process 31 X 31 and goes on and now click on ‘Accuracy & Loss Graph’ button to get below graph



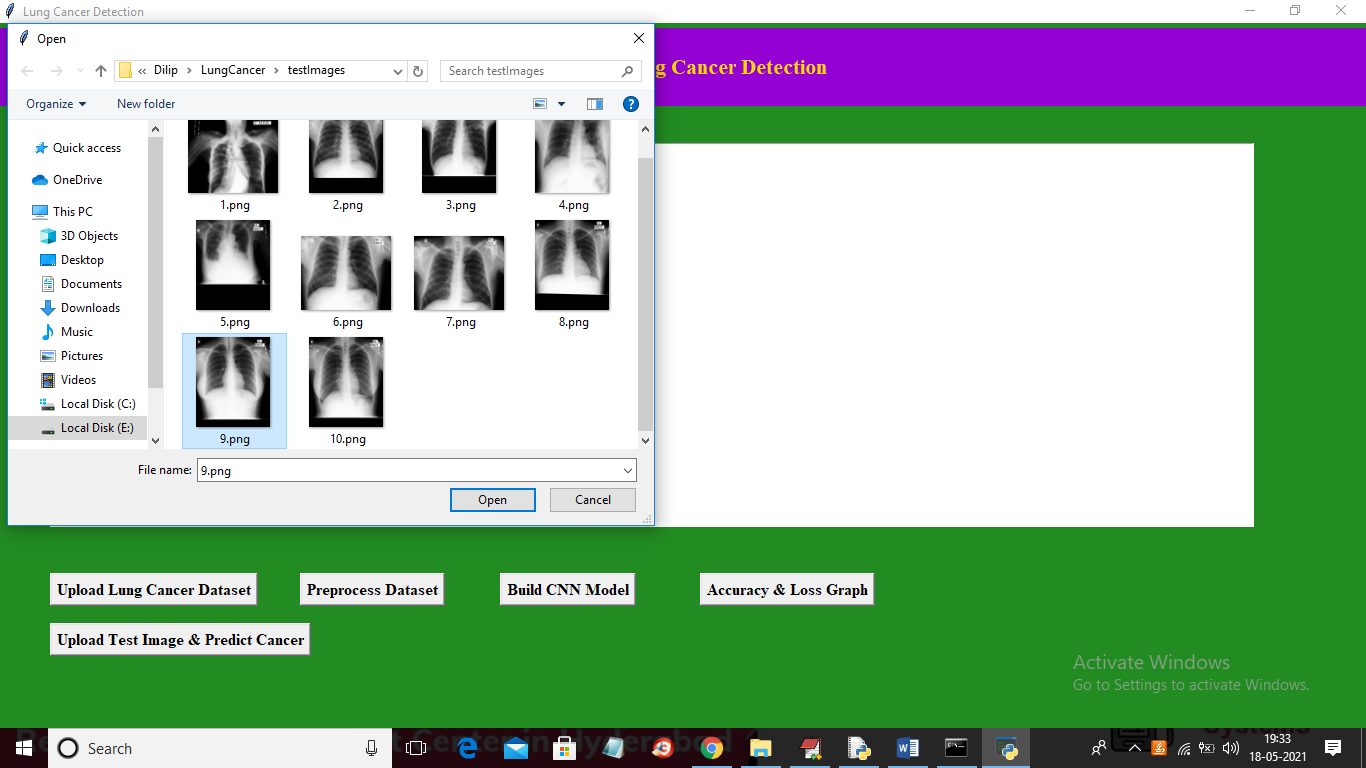
In above graph x-axis represents Epoch and Y-axis represents accuracy and loss values and in above graph we can see to train CNN we took 10 Epoch and at each increasing Epoch Loss values get decrease and accuracy gets increase and in above graph red line represents loss and green line represents accuracy. Now click on ‘Upload Test Image & Predict Cancer’ button to upload test image and then detect cancer



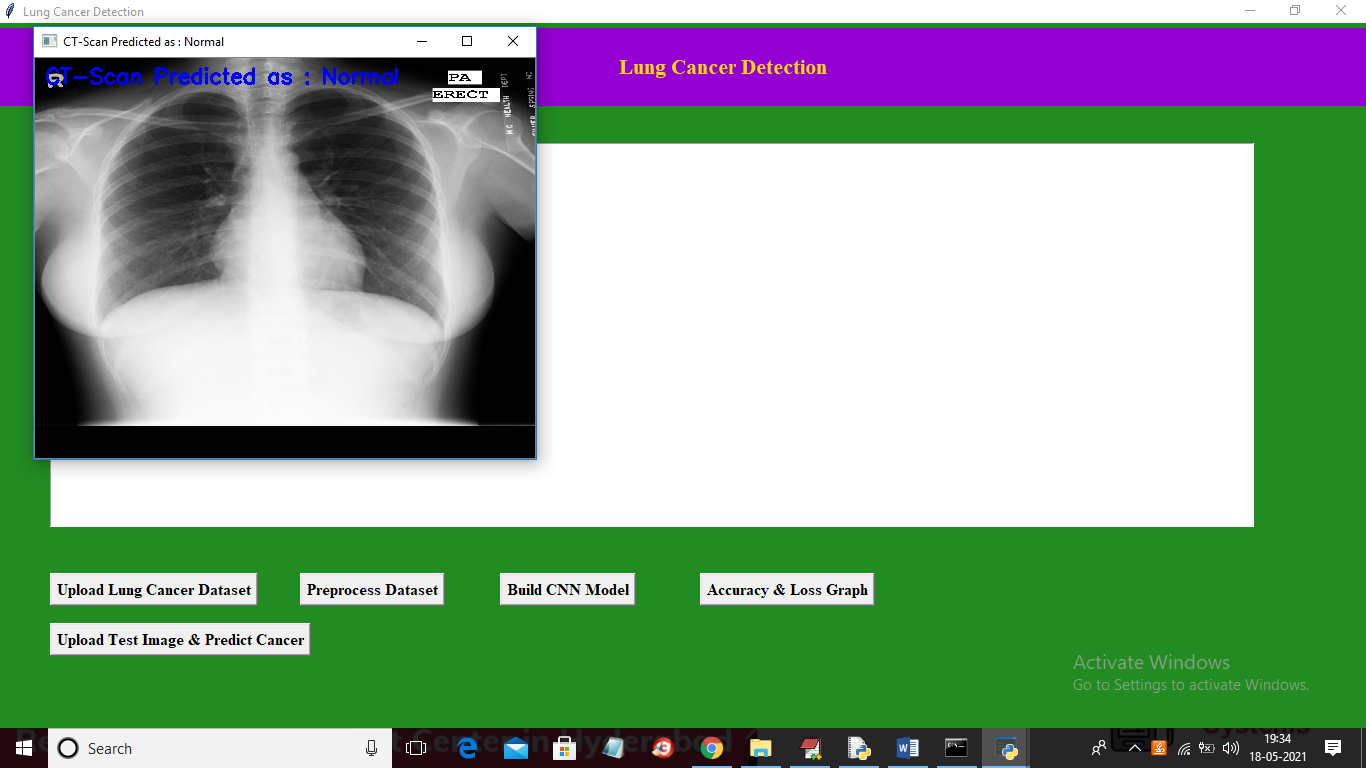
In above screen selecting and uploading ‘1.png’ file and then click on ‘Open’ button to get below result



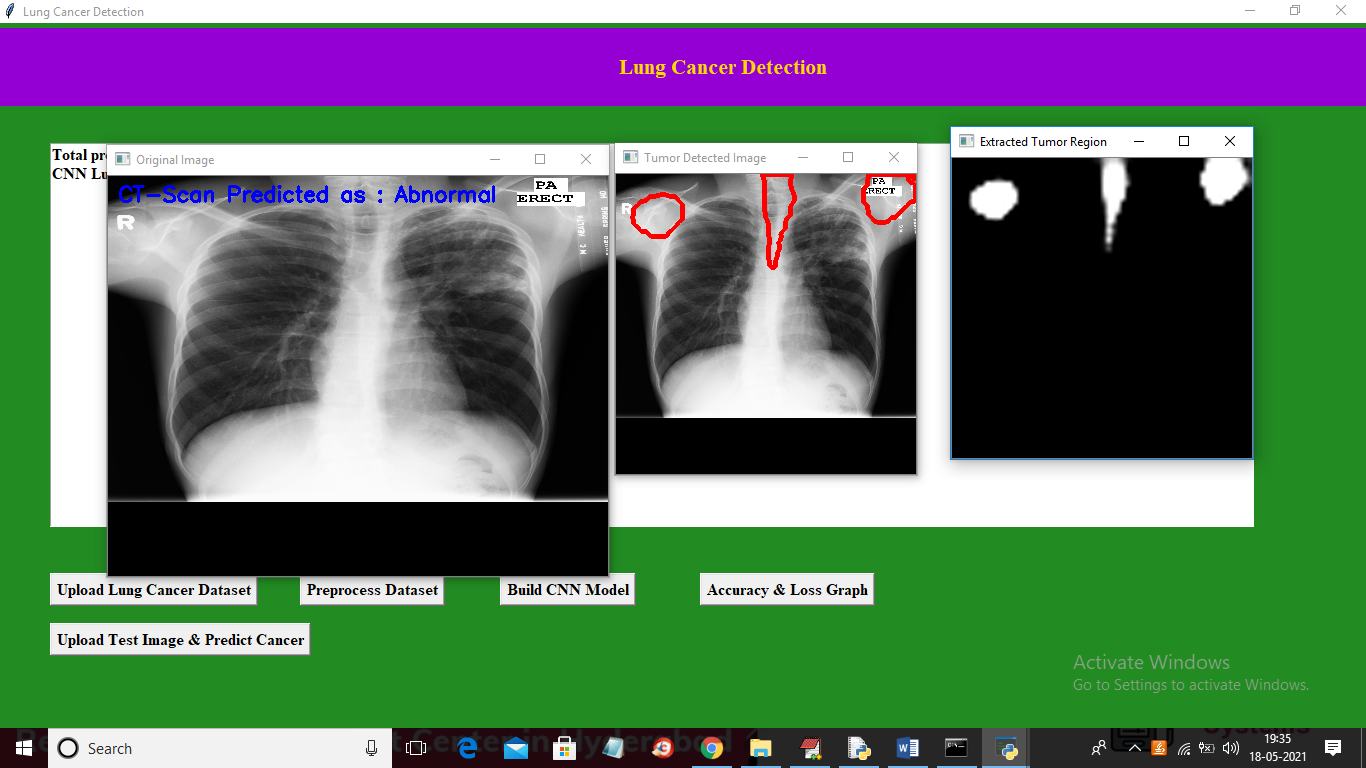
In above screen in first image in blue colour text we can see predicted result as CT-SCAN contains abnormality and in second image we are detecting places were abnormality detected and in third image we extracted all abnormality patches from original image and then displaying. Now test other image



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



In above screen CT-SCAN is predicted as NORMAL. Similarly you can upload and test other images



**CHAPTER 6**

**CONCLUSION**

In our research work, we have used deep Convolutional neural networks for classifying the ct images of lung nodules into cancerous (malignant) and non-cancerous (benign). Thus preprocessing has been done before applying input ct images to network model to make equal sizes and format of the images. The dataset used in our research work belongs to LIDC dataset. Hence we achieved an accuracy of 100% which is the better results comparable to previous research papers as mentioned. As a future work, the experiments could be performed by using Deep CNN architecture for other types of cancer.

**REFERENCES**

1. Jiang, H., Qian, W., Gao, M., Li, Y. “ An automatic detection system of lung nodule based on multigroup patch-based deep learning network” IEEE Journal of Biomedical and Health Informatics,22(4):1227-1237.
2. Disha Sharma, Gagandeep Jindal, “Identifying Lung Cancer Using Image Processing Techniques”, International Conference on Computational Techniques and Artificial Intelligence, pp. 116-120, 2011.
3. Farzad Vasheghani Farahani “Lung Nodule diagnosis from CT images Based on Ensemble Learning.” IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology , 2015.
4. Atsushi Teramoto at el. “Automated classification of benign and malignant cells from lung cytologicalimages using deep convolutional neural network” Informatics in medicine unlocked 16,2019,100205 elsvier.
5. Xin-Yu Jin, Yu-Chen Zhang, Qi-Liang Jin “Pulmonary Nodule Detection Based on CT Images Using Convolution Neural Network.” 9Th International Symposium on Computational Intelligence and Design. 2016.
6. Ryota Shimizu, Shusuke Yanagawa, Yasutaka Monde, Hiroki Yamagishi,Mototsugu Hamada, Toru Shimizu, and Tadahiro Kuroda “Deep Learning Application Trial to Lung Cancer Diagnosis for Medical Sensor Systems” International Symposium on Computers and Communications, 2016.
7. Po-Whei Huang, Phen-Lan Lin, Cheng-Hsiung Lee, C. H. Kuo, “A Classification System of Lung Nodules in CT Images Based on Fractional Brownian Motion Model”, IEEE International Conference on System Science and Engineering, July 2016.
8. Vaishali C. Patil, Shrinivas R. Dhotre, “Lung Cancer Detection from Images of Computer Tomography Scan”, International Journal of Advanced Research in Computer and Communication Engineering , Vol. 5, Issue 7, July 2016.
9. Ailton Felix, Marcelo Oliveira, Aydano Machado, Jose Raniery, “Using 3D Texture and Margin Sharpness Features on Classification of Small Pulmonary Nodules” ,29th SIBGRAPI Conference on Graphics, Patterns and Images, 2016.
10. Rotem Golan, Christian Jacob, Jorg Denzinger, “Lung nodule detection in CT images using deep convolutional neural networks” International Joint Conference on Neural Networks , 2017.
11. Sri Widodo, Ratnasari Nur Rohmah, Bana Handaga, “Classification Of Lung Nodules And Arteries In Computed Tomography Scan Image Using Principle Component Analysis” 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, 2017.
12. Ravindranath K , K Somashekar, “Early Detection of lung cancer by nodule extraction – A Survey”, International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques , 2017.
13. Rui Xu, Jiao Pan, Xinchen Ye, Yasushi Hirano, Shoji Kido, Satoshi Tanaka “A Pilot Study to Utilize a Deep Convolutional Network to Segment Lungs with Complex Opacities” Chinese Automation Congress , 2017.
14. Anna Poreva, Yevgeniy Karplyuk, Valentyn Vaityshyn, “Machine Learning Techniques Application for Lung Diseases Diagnosis” , 5th IEEE Workshop on Advances in Information, Electronic and Electrical Engineering , 2017.
15. Pratiksha Hattikatti, “Texture based Interstitial Lung Disease Detection using Convolutional Neural Network”, International Conference on Big Data, IoT and Data Science , 2017.
16. Wei Chen, Haifeng Wei , Jiawei Sun ,Xu Qiao , Boqiang Liu, “Hybrid Segmentation Network for Small Cell Lung Cancer Segmentation” IEEE Access, vol. 7, pp. 75591 - 75603, June 2019.
17. Moradi, P. & Jamzad, M., “Detecting Lung Cancer Lesions in CT Images using 3D Convolutional Neural Networks”, 4th International Conference on Pattern Recognition and Image Analysis, 2019.
18. Md. Sakif Rahman, Pintu Chandra Shill, Zarin Homayra, “A New Method for Lung Nodule Detection Using Deep Neural Networks for CT Images” , International Conference on Electrical, Computer and Communication Engineering , 2019.
19. Golan, R., Jacob, C. and Denzinger, J, “Lung Nodule Detection in CT Images Using Deep Convolutional Neural Networks”, International Joint Conference on Neural Networks, Vancouver, 24-29 July 2016, 243- 250

**APPENDIX**

**PYTHON**

\* One of the most popular languages is Python. Guido van Rossum released this language in 1991. Python is available on the Mac, Windows, and Raspberry Pi operating systems. The syntax of Python is simple and identical to that of English. When compared to Python, it was seen that the other language requires a few extra lines.

\*It is an interpreter-based language because code may be run line by line after it has been written. This implies that rapid prototyping is possible across all platforms. Python is a big language with a free, binary-distributed interpreter standard library.

\* It is inferior to maintenance that is conducted and is straightforward to learn. It is an object-oriented, interpreted programming language. It supports several different programming paradigms in addition to object-oriented programming, including functional and procedural programming.

\* It supports several different programming paradigms in addition to object-oriented programming, including practical and procedural programming. Python is mighty while maintaining a relatively straightforward syntax. Classes, highly dynamic data types, modules, and exceptions are covered. Python can also be utilised by programmes that require programmable interfaces as an external language.

**Python Features:**

**1) Easy:** Because Python is a more accessible and straightforward language, Python programming is easier to learn.

**2) Interpreted language:** Python is an interpreted language, therefore it can be used to examine the code line by line and provide results.

**3) Open Source:** Python is a free online programming language since it is open-source.

**4) Portable:** Python is portable because the same code may be used on several computer standard **libraries:** Python offers a sizable library that we may utilize to create applications quickly.

**6) GUI:** It stands for GUI (Graphical User Interface)

**7) Dynamical typed:** Python is a dynamically typed language, therefore the type of the value will be determined at runtime.

**Python GUI (Tkinter)**

\* Python provides a wide range of options for GUI development (Graphical User Interfaces).

\* Tkinter, the most widely used GUI technique, is used for all of them.

\* The Tk GUI toolkit offered by Python is used with the conventional Python interface.

\* Tkinter is the easiest and quickest way to write Python GUI programs.

\* Using Tkinter, creating a GUI is simple.

\* A part of Python's built-in library is Tkinter. The GUI programs were created.

\* Python and Tkinter together give a straightforward and quick way. The Tk GUI toolkit's object-oriented user interface is called Tkinter.

\* Making a GUI application is easy using Tkinter. Following are the steps:

1) Install the Tkinter module in place.

2) The GUI applicatioMakeske the primary window

3) Include one or more of the widgets mentioned above in the GUI application.

4) Set up the main event loop such that it reacts to each user-initiated event.

\*Although Tkinter is the only GUI framework included in the Python standard library, Python includes a GUI framework. The default library for Python is called Tkinter. Tk is a scripting language often used in designing, testing, and developing GUIs. Tk is a free, open-source widget toolkit that may be used to build GUI applications in a wide range of computer languages.

**Machine Learning**

\*Artificial intelligence (AI), which includes machine learning, enables computer systems to learn without being explicitly programmed. It has to do with statistics and applied mathematics. Mike Robert's definition of machine learning. As a computer gathers and learns from the data it provides, it may operate more correctly via machine learning.

\*For large classes of machine learning, many algorithms are used. We must provide algorithms with more precise data for them to complete certain jobs. In some circumstances, a computer will utilize data to gather information, check its output against the desired outcome, and make necessary corrections.

\*For instance, when someone texts on a phone, the phone learns about spelling errors and either autocorrects the offending word or suggests a replacement. For many top organizations, machine learning is a critical component of the creation of new products.

\*ML is an important factor in the operations of many companies, like Facebook and Google. Data science uses machine learning in many different ways. Data scientists rely on ML approaches to carry out their modeling. Regression and classification are of utmost relevance in data science; hence, the main tool utilized in ML is to accomplish such objectives.

\* ML applies applicable to practically all phases of data science and is most often associated with the data modeling phase. Python has been the primary computer language used for data processing. Several Python packages are used in ML settings. The three sections of Python are huge data, optimizing your code, and data files in memory.

**1.6 Types of Machine Learning**

There are three fundamental forms of machine learning: -supervising, semi-supervised, and machine learning

**a) Supervised Machine Learning**

\* That method looks for patterns in the labeled data set to obtain results. Data labeling in supervised learning requires human intervention. To train the algorithm with labeled inputs and the intended output, supervised ML requires human participation. ML under supervision is good for a task like;

**I.** Classify the data using a binary system into two groups.

**II. Multi-classification:** The division of data into more than two categories,

**III.** Modeling imaging continuous value using regression.

**IV. Assembling:** Compiling the estimates from many ML models to provide a precise estimate.

**b) Unsupervised Machine Learning**

\*This method searches for patterns in the data collection without relying on labeled data or human interaction. Data labeling is not necessary for this strategy. ML Unsupervised is effective for tasks like;

**I. Dimensionality reduction:** Reduce the number of variables in the data collection.

**II. Clustering:** Grouping the dataset based on similarities.

**III.** Association mining identifies the item or group of items that commonly appear together in data.

**IV.** Data point identification for anomaly detection in the data set

**c) Semi-supervised Learning**:

\*For this method, you require labeled data. As a consequence, human interaction is also necessary, but the process still moves forward. In this kind of learning, the algorithm is given a tiny quantity of labeled data by data scientists, and as a result, the algorithm gains knowledge about the data set's dimension, which it may then apply to mother del, unlabeled data.

\* There are several contexts in which semi-supervised machine learning (ML) may be used.

**I. Machine translation:** Language conversion using a learning system.

**II. Data labeling:** An algorithm trained on modest amounts of data will automatically apply data labels to enormous collections.

**1.7 Uses of Machine Learning**

\*Machine learning is used in many areas nowadays. The most well-known example is the machine learning recommendation engine that drives a book's news feed. This engine makes an effort to reinforce established patterns in a user's online activity inside a certain Facebook group.

\* The news is appropriately adjusted if a user alters the design and doesn't read anything from that particular group the following week. Applications of machine learning (ML) include business intelligence, human resource information systems, autonomous vehicles, and virtual assistants.

**Advantages:**

• ML helps enterprises in comprehending their clients. ML assists in improving goods in response to client demand by gathering the necessary user data and associating it with shifting behavior. Some companies' business models are heavily reliant on machine learning, such as Uber, which uses an algorithm to connect drivers and customers. To surface the advertising in searches, Google employs ML.

**Disadvantage:**

• ML might be expensive. High wages for machine learning are a result of data emotions command on the project. These initiatives also often demand expensive software infrastructure.

• In addition to that, when an algorithm is trained on a data set, ML bias might develop. That has flaws in it that might provide erroneous results.

**Steps to choosing the suitable ML model**

The issue is solved by selecting the best ML model, which might take some time. The steps are as follows:

1) For the difficulty with the pure date alignment, the input should be thought about.

2) Gather, label, and prepare the data as appropriate.

3) To put the right algorithms to use and test them to determine how well they perform.

**Libraries Used**

**Pandas:**

\* Pandas is a Python computer language library for data analysis and manipulation. It offers a specific operation and data format for handling time series and numerical tables. It differs significantly from the release3-clause of the BSD license. It is a well-liked open-source of opinion that is utilized in machine learning and data analysis.

**NumPy:**

\* The NumPy Python library for multi-dimensional, big-scale matrices adds a huge number of high-level mathematical functions. It is possible to modify NumPy by utilizing a Python library. Along with line, algebra, and the Fourier transform operations, it also contains several matrices-related functions.

**Matplotlib:**

\* It is a multi-platform, array-based data visualization framework built to interact with the whole SciPy stack. MATLAB is proposed as an open-source alternative. Matplotlib is a Python extension and a cross-platform toolkit for graphical plotting and visualization.

**Scikit-learn:**

\* The most stable and practical machine learning library for Python is scikit-learn. Regression, dimensionality reduction, classification, and clustering are just a few of the helpful tools it provides through the Python interface for statistical modeling and machine learning. It is an essential part of the Python machine learning toolbox used by JP Morgan. It is frequently used in various machine learning applications, including classification and predictive analysis.

**Keras:**

\* Google's Keras is a cutting-edge deep learning API for creating neural networks. It is created in Python and is designed to simplify the development of neural networks. Additionally, it enables the use of various neural networks for computation. Deep learning models are developed and tested using the free and open-source Python software known as Keras.

**h5py:**

\* The h5py Python module offers an interface for the binary HDF5 data format. Thanks to p5py, the top can quickly halt the vast amount of numerical data and alter it using the NumPy library. It employs common syntax for Python, NumPy, and dictionary arrays.