### **Pathology Image Analysis** **For Lung Cancer Prediction Using Deep Learning**

Major Project Report

Submitted By

**(BATCH NO: CSE\_05)**

GANTA SUCHENDAR REDDY (18UK1A0516)

KANKATI PRANITHA  (18UK1A0525)

MUSHIKA SAIRAM  (18UK1A0535)

JANAGANI VAMSHI KRISHNA  ( 18UK1A0576)

# **CHAPTER 1**

**INTRODUCTION**

#### **1.1. Introduction**

Lung cancer is one of the causes of cancer deaths. It is difficult to detect because it arises and shows symptoms in final stage. However, mortality rate and probability can be reduced by early detection and treatment of the disease. Best imaging technique CT imaging are reliable for lung cancer diagnosis because it can disclose every suspected and unsuspected lung cancer nodules. However, variance of intensity in CT scan images and anatomical structure misjudgment by doctors and radiologists might cause difficulty in marking the cancerous cell. Recently, to assist radiologists and doctors to detect the cancer accurately computer Aided Diagnosis has become supplement and promising tool. There have been many systems developed and research is going on detection of lung cancer.

However, some systems do not have satisfactory accuracy of detection and some systems still has to be improved to achieve highest accuracy tending to 100%. Image processing techniques and machine learning techniques have been implemented to detect and classify the lung cancer. We studied recent systems developed for cancer detection based on CT scan images of lungs to choose the recent best systems and analysis was conducted on them and new model was proposed. Small cell lung cancer and non-small cell lung cancer are common types of lung cancer. The general symptoms of lung cancer include coughing up blood, chest pain, weight loss and loss of appetite, shortness of breath and feeling weak.

#### **1.2 Objective of Project**

The goal of a CT screening program is to detect early lung cancer and facilitate curative treatment; however, primary prevention through smoking cessation or never starting is the best means to reduce the risk of dying of lung cancer. We need to get the word out to those at high risk who stand to benefit most from mortality reduction**.**

# **CHAPTER 2**

# **LITERATURE SURVEY**

#### **2.1 Existing problem**

Early detection of lung nodule is of great importance for the successful diagnosis / and treatment of lung cancer. Many researchers have tried with diverse methods, such as thresholding, computer- aided diagnosis system, pattern recognition technique, backpropagation algorithm, etc.

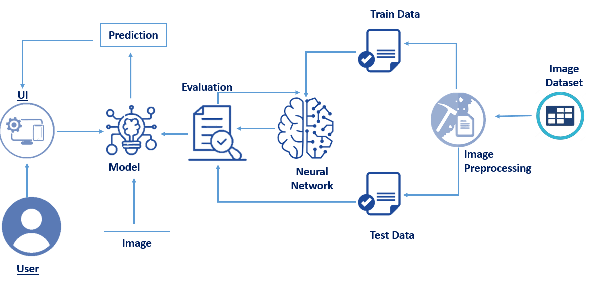
**2.2** **Proposed Solution**

Building a Web App which automatically generates the sequence of text by analysis the input text. Our model uses sliding window for analyzing large sequence of input text.

**CHAPTER-3**

**3.THEORTICAL ANALYSIS**

3.1 BLOCK DIAGRAM



3.2 HARDWARE/SOFTWARE DESIGNING

|  |  |
| --- | --- |
| REQUIREMENT | SPECIFICATION |
| Anaconda Navigator | You must have anaconda installed in your device prior to begin. |
| Spyder, Jupyter Notebook, Flask  Frame work | 1.One should have Spyder and Jupyter notebook.  2. One should install flask framework through anaconda prompt for running their web application  3.We need to build the model using jupyter notebook with all the imported packages. |
| Web browser | For all Web browsers, the following must be enabled:   * cookies * JavaScript |

Hardware Specifications:

|  |  |
| --- | --- |
| REQUIREMENT | SPECIFICATIONS |
| Operating system | Microsoft Windows  UNIX  Linux® |
| Processing | Minimum: 4 CPU cores for one user. For each deployment, a sizing exercise is highly recommended. |
| RAM | Minimum 8 GB. |
| Operating system specifications | File descriptor limit set to 8192 on UNIX and Linux |
| Disk space | A minimum of 7 GB of free space is required to install the software. |

**CHAPTER-4**

**EXPERIMENTAL INVESTIGATIONS**

**Project Flow:**

* User interacts with User interface to upload image
* Uploaded image is analyzed by the model which is integrated
* Once model analyses the uploaded image, the prediction is showcased on the UI

To accomplish this, we have to complete all the activities and tasks listed below

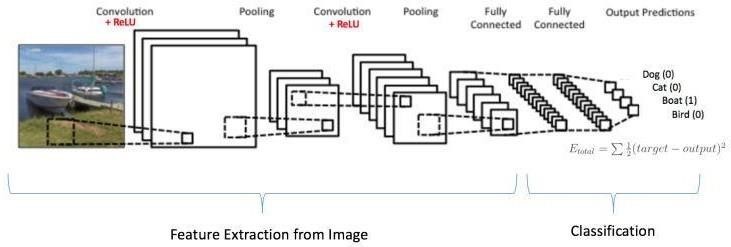
* Data Collection.
  + Collect the dataset or create the dataset
* Data Preprocessing.
  + Import the ImageDataGenerator library
  + Configure ImageDataGenerator class
  + Apply ImageDataGenerator functionality to Trainset and Testset
* Model Building
  + Import the model building Libraries
  + Initializing the model
  + Adding Input Layer
  + Adding Hidden Layer
  + Adding Output Layer
  + Configure the Learning Process
  + Training and testing the model
  + Optimize the Model
  + Save the Model
* Application Building
  + Create an HTML file
  + Build Python Code

### **CHAPTER-5**

**ANALYSIS**

### **Content diagram of Project**

CNN is a type of feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex, whose individual neurons are arranged in such a way that they respond to overlapping regions tiling the visual field. It consists of multiple layers of small neuron collections which process portions of the input image, called receptive fields. The outputs of these collections are then tiled so that their input regions overlap, to obtain a better representation of the original image; this is repeated for every such layer. A more detailed description of the layer of a Convolutional Neural Network can be found in the next section.

  
**Convolutional Layer:**

It is the core building block of a Convolutional Network, and its output volume can be interpreted as holding neurons arranged in a 3D volume. The CONV layer’s parameters consist of a set of learnable filters. Every filter is small spatially, but extends through the full depth of the input

volume. During the forward pass, we slide each filter across the width and height of the input volume, producing a 2-dimensional activation map of that filter. As we slide the filter, across the input, we are computing the dot product between the entries of the filter and the input. Intuitively, the network will learn filters that activate when they see some specific type of feature at some spatial position in the input. Stacking these activation maps for all filters along the depth dimension forms the full output volume. Every entry in the output volume can thus also be interpreted as an output of a neuron that looks at only a small region in the input and shares parameters with neurons in the same activation map. Three hyper parameters control the size of the output volume: the depth, stride and zero-padding

#### **ReLu:**

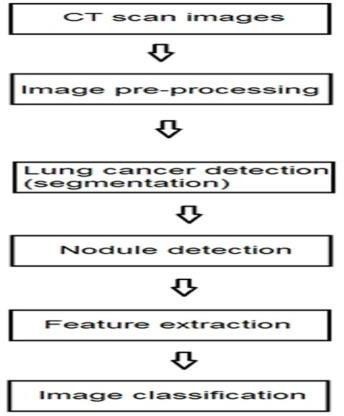
The rectified linear unit will apply an element wise activation function, such as the max(0,x) thresholding at zero. This leaves the size of the volume unchanged.

**Algorithms and Flowcharts**

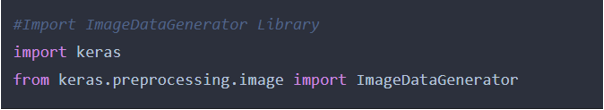
Generally, the structure of CNN includes two layers one is feature extraction layer, the input of each neuron is connected to the local receptive fields of the previous layer, and extracts the local feature. Once the local features is extracted, the positional relationship between it and other features also will be determined. The other is feature map layer, each computing layer of the network is composed of a plurality of feature map. Every feature map is a plane, the weight of the neurons in the plane are equal. The structure of feature map uses the sigmoid function as activation function of the convolution network, which makes the feature map have shift invariance. Besides, since the neurons in the same mapping plane share weight, the number of free parameters of the network is reduced. Each convolution layer in the convolution neural network is followed by a computing layer which is used to calculate the local average and the second extract, this unique two feature extraction structure reduces the resolution.

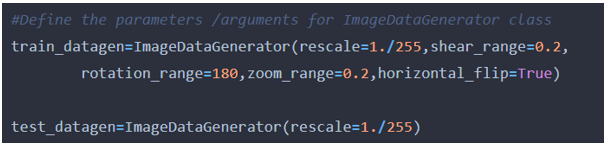
CNN is mainly used to identify displacement, zoom and other forms of distorting invariance of two-dimensional graphics. Since the feature detection layer of CNN learns by training data, it avoids explicit feature extraction and implicitly learns from the training data when we use CNN. Furthermore, the neurons in the same feature map plane have the identical weight, so the network can study concurrently. This is a major advantage of the convolution network with respect to the neuronal network connected to each other. Because of the special structure of the CNN’s local shared weights makes it have a unique advantage in speech recognition and image processing. Its layout is closer to the actual biological neural network. Shared weights reduces the complexity of the network.

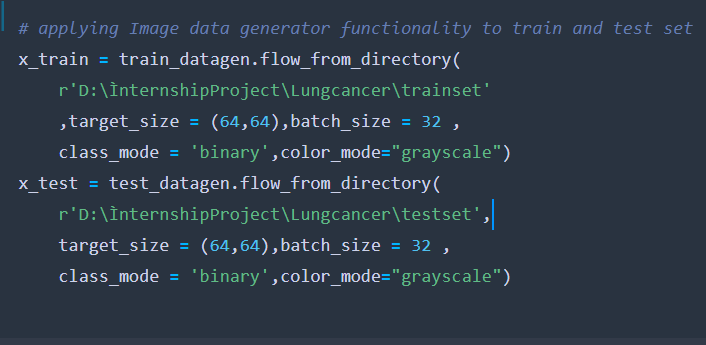
#### **FLOW CHART**



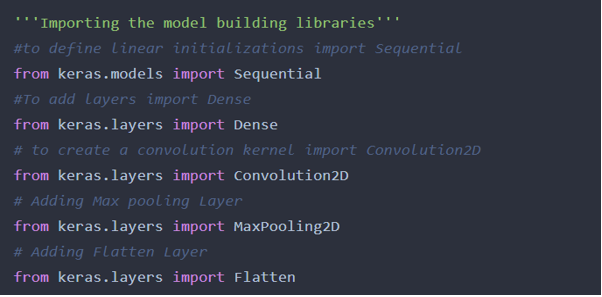
### Importing The ImageDataGenerator Library.



Define the parameters 



### **Import Libraries**



### **Initializing The Model**

Keras has 2 ways to define a neural network:

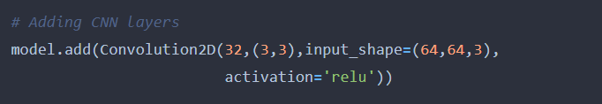
* Sequential
* Function API

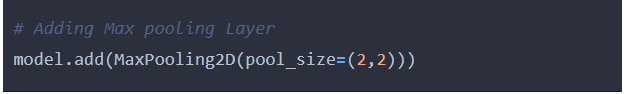


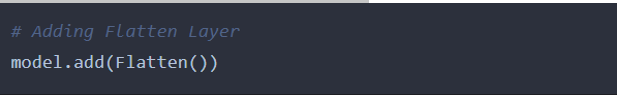
### **Add CNN Layers**

We will be adding three layers for CNN

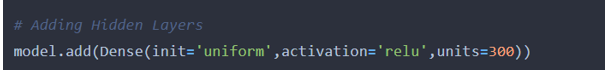
* Convolution layer
* Pooling layer
* Flattening layer

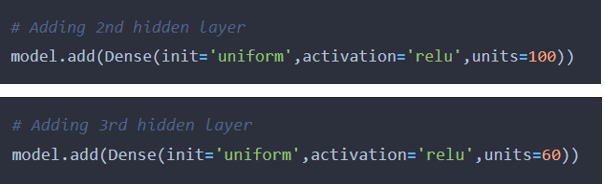






### **Adding Dense Layers**



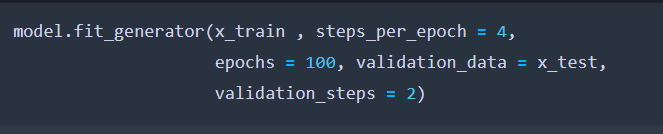


### **Adding the output layer**

### **Configuring The Learning Process**

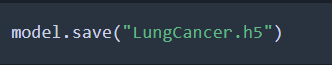
### 

### **Training The Model**



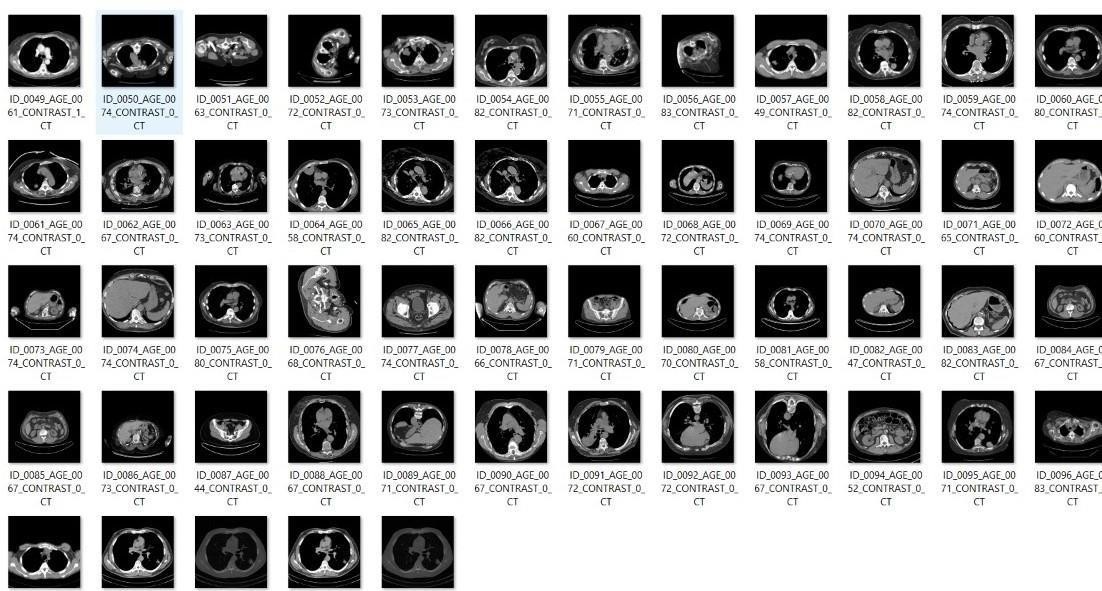
At this point, we have training data and a fully configured neural network to train with loaded data. All that is left is to pass the data to the model for the training process to commence, a process that is completed by iterating on the training data. Training begins by calling the fit ( ) method.

### **Save The Model**



**Train Data set:**

**CANCER Train Dataset**



**NON-CANCER Test Dataset**

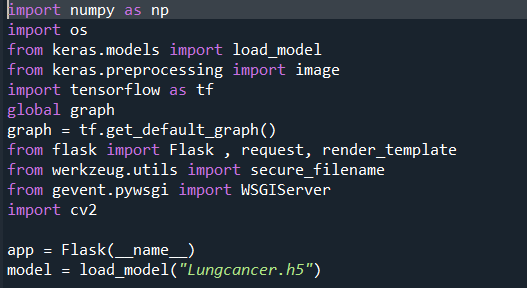
### TEST DATASET.jfif **Application Building**

In this section, we will be building a web application that is integrated to the model we built. A UI is provided for the uses where he has uploaded an image. The uploaded image is given to the saved model and prediction is showcased on the UI.

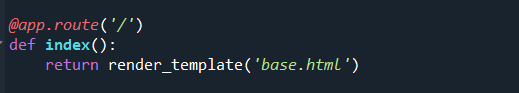
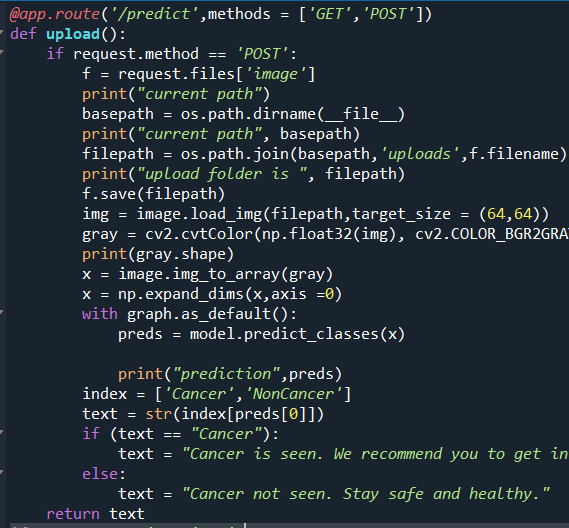
This section has the following tasks

* Building HTML Pages
* Building server-side scrip

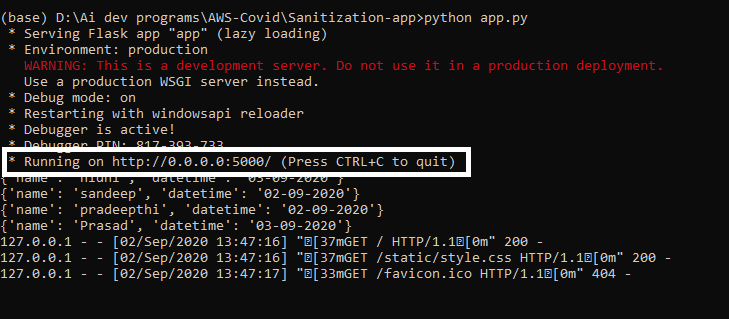
**Build Python Code:**



Render The HTML Page:

  
Bind the URL to perform some action:

G:\s4.png  
**Save the file: Run the App**

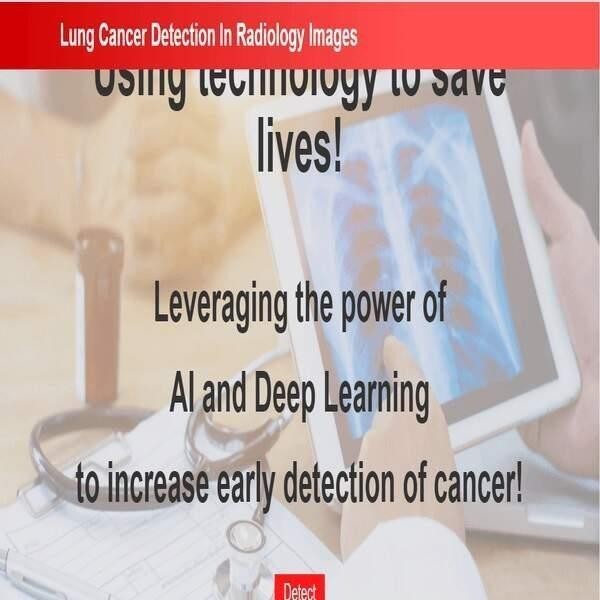


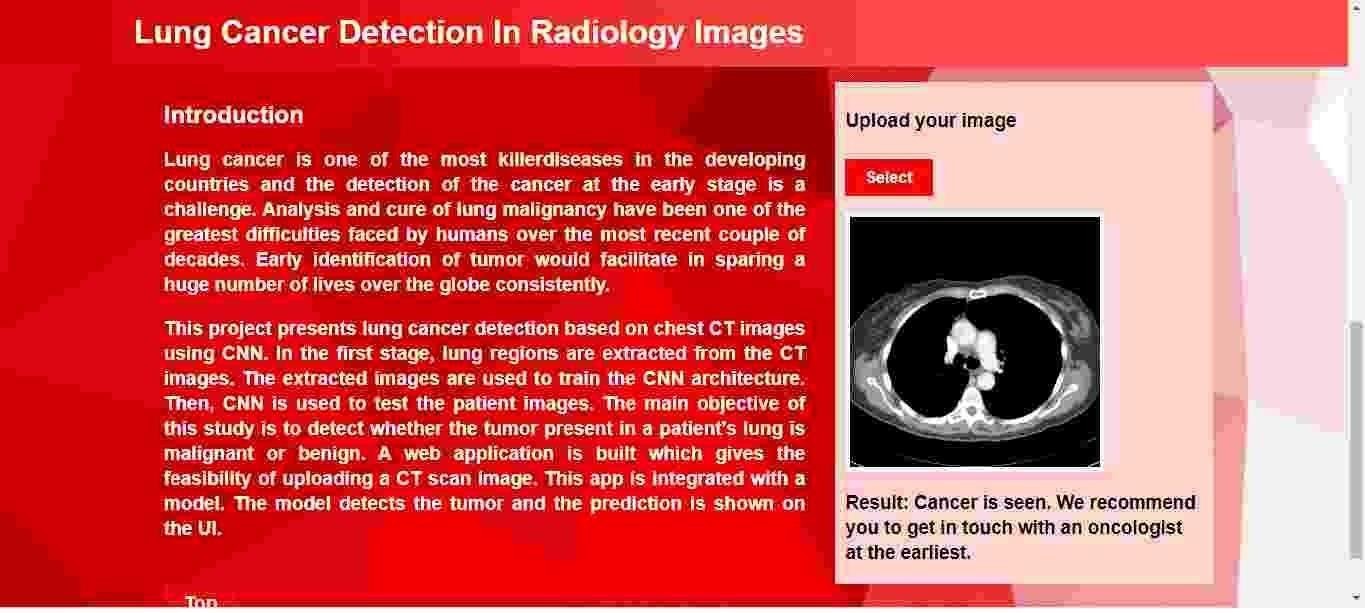
**CHAPTER 6**

**RESULTS**

Final output of the project:

The home page is be shown as:

  
Detect Pages Looks Like: cancer is predicted

  
  **CHAPTER 7**

**ADVANTAGES AND DISADVANTAGES**

**ADVANTAGES**

* Easily implemented low-cost assay
* High throughput
* Can be used to compare metastatic potential of cells
* Allows real time tracking of cells

**DISADVANTAGES**

* Low physiological relevance
* Can only assay single-cell motility
* Migration and invasion assays can results in conflicting data
* Lacks tumor complexity

**CHAPTER 8**

**CONCLUSION**

We obtained proof of concept that CNNs can be used as a tool to assist in the staging of patients affected by lung cancer. In this research, we used CNN classifier to determine whether a CT image of lung is cancerous or non-cancerous. Before using CNN, we preprocessed the CT image through a thresholding technique. We have performed a thorough experiment using LUNA 16 dataset. Our obtained detection accuracy is 80%, which is better than existing methods. In this work, we presented different deep CAD systems and models that pursue the common objective of alleviating the work of radiologists in lung nodule detection. It is shown that deep learning has achieved a level of precision that allows implementation, not only as a second opinion indiagnosis but as a powerful tool that can be considered by physicians in their work.

In the near future, the system will be trained with large datasets to diagnose the type of cancer with its size and shape. The overall accuracy of the system can be improved using 3D Convolutional Neural Network and also by improving the hidden neurons with deep network.

**CHAPTER-9**

**FUTURE SCOPE**

Further research and studies are to be conducted and validation of the proposed models of convolutional neural networks has to be performed. Validation of the proposed models is required for the practical application of these in the screening procedure of lung cancer and thereby increasing the detection in earlier stages. More research and trials are to be conducted utilizing the technological advancements and the doctors have to take up the challenge to improvise and implement them.

**CHAPTER-10**

# **BIBLIOGRAPHY:**

* + - <https://www.geeksforgeeks.org/>
    - <https://www.w3schools.com/>
    - <https://stackoverflow.com/>
    - <https://www.lucidchart.com/pages/>
    - [https://smartinternz.com/Student/guided\_project\_info/4652#](https://smartinternz.com/Student/guided_project_info/4652)