Lung Cancer Detection In Radiology Images

Using CNN

Developed by:Baireddy Manisha,Mamindla Shalini,Madishetti Manisha,Padakanti Ramya

Smart Bridge-Remote Summer Internship Program

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 detect the cancer accurately computer Aided Diagnosis has become supplement and promising tool. There has been many system developed and research 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 has 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 . Early detection improves the survival rate from 15% to 50% . However, there is a need to increase this survival rate more than the current value. Images generated by X-rays, Computed-Tomography (CT) scans, Magnetic Resonance Imaging (MRI) and others help in the early detection of lung cancer without surgery. The CT scan is the most recommended method which produces the 3D images of the lungs. Mortality rate can be reduced by early detection and treatment of the disease. The process of early detection of cancer plays an important role to prevent cancer cells from multiplying and spreading. Existing lung cancer detection techniques are not adequate for providing

accuracy. Hence, it is of importance to develop new methods for the early detection of lung cancer.

1.1.Overview :

In our project, we aim to predict the lung cancer with the help of CT images using Convolution neural network. Lung cancer is one of the dangerous and life taking disease in the world. However, early diagnosis and treatment can save life. Although, CT scan imaging is best imaging technique in medical field, it is difficult for doctors to interpret and identify the cancer from CT scan images. Therefore computer aided diagnosis can be helpful for doctors to identify the cancerous cells accurately. Many computer aided techniques using image processing and machine learning has been researched and implemented. The main aim of this research is to evaluate the various computer-aided techniques, analyzing the current best technique and finding out their limitation and drawbacks and finally proposing the new model with improvements in the current best model. The method used was that lung cancer detection techniques were sorted and listed on the basis of their detection accuracy. The techniques were analyzed on each step and overall limitation, drawbacks were pointed out. It is found that some has low accuracy and some has higher accuracy but not nearer to 100%. Therefore, our research targets to increase the accuracy towards 100%.

1.2.Purpose:

Our aim from the project is to make use of keras to extract the libraries for machine learning for the lung canacer detection in radiology images.The aim of this research was to detect features for accurate images comparison.

Secondly, to learn how to hyper tune the non cancer and cancer images to machine learning algorithm.

And in the end,the objective of this phase of the project is to develop a model which can tell us whether an individual has lung cancer or not based on radiology images.The objective is achieved through applying CNN.

2. LITERATURE SURVEY :

Several researchers has proposed and implemented detection of lung cancer using different approaches of image processing and machine learning.

- Aggarwal, Furquan and Kalproposed a model that provides classification between nodules and normal lung anatomy structure. The method extracts geometrical, statistical and gray level characteristics. LDA is used as classifier and optimal thresholding for segmentation. The system has 84% accuracy, 97.14% sensitivity and 53.33% specificity. Although the system detects the cancer nodule, its accuracy is still unacceptable. No any machine learning techniques has been used to classify and simple segmentation techniques is used. Therefore, combination of any of its steps in our new model does not provide probability of improvement.
- Jin, Zhang and Jin used convolution neural network as classifier in his CAD system to detect the lung cancer. The system has 84.6% of accuracy, 82.5% of sensitivity and 86.7% of specificity. The advantage of this model is that it uses circular filter in Region of interest (ROI) extraction phase which reduces the cost of training and recognition steps. Although, implementation cost is reduced, it has still unsatisfactory accuracy.
- Sangamithraa and Govindaraju uses K mean unsupervised learning algorithm for clustering or segmentation. It groups the pixel dataset according to certain characteristics. For classification this model implements back propagation network. Features like entropy, correlation, homogeneity, PSNR, SSIM are extracted using gray-level co-occurrence matrix (GLCM) method. The system has accuracy of about 90.7%. Image pre processing median filter is used for noise removal which can be useful for our new model to remove the noise and improve the accuracy.
- Roy, Sirohi, and Patdeveloped a system to detect lung cancer nodule using fuzzy interference system and active contour model. This system uses gray transformation for image contrast enhancement. Image binarization is performed before segmentation and resulted image is segmented using active contour model. Cancer classification is performed using fuzzy inference method. Features like area, mean, entropy, correlation, major axis length, minor axis length are extracted to train the classifier. Overall, accuracy of the system is 94.12%. Counting its limitation it does not classify the cancer as benign or malignant which is future scope of this proposed model

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.
The existing system is Time consuming process, and it's very difficult to detect it in its early stages as its symptoms appear only within the advanced stages. Implementing the system to automate the classification process for the first prediction of carcinoma.

2.2 Proposed Solution ：

Convolution Neural Network

Convolutional neural networks (CNN) are all the rage in the deep learning community right now. These CNN models are being used across different applications and domains, and they’re especially prevalent in image and video processing projects.A convolutional neural network (CNN) is a specific type of artificial neural network that uses perceptrons, a machine learning unit algorithm, for supervised learning, to analyze data. CNNs apply to image processing, natural language processing and other kinds of cognitive tasks.

A convolutional neural network is also known as a ConvNet.

Like other kinds of artificial neural networks, a convolutional neural network has an input layer, an output layer and various hidden layers. Some of these layers are convolutional, using a mathematical model to pass on results to successive layers. This simulates some of the actions in the human visual cortex.

CNNs are a fundamental example of deep learning, where a more sophisticated model pushes the evolution of artificial intelligence by offering systems that simulate different types of biological human brain activity.

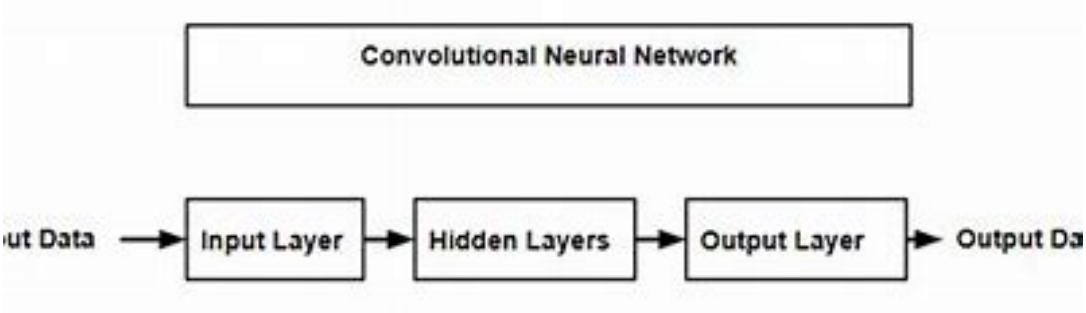
And also we have created an UI using the Flask for the lung cancer prediction,this UI will allow the users to predict the case status very easily and the User interface is user friendly not at least one complication in using the interface.

3.Theoritical Analysis:

Recently, image processing techniques are widely used in several medical areas for image improvement in earlier detection and treatment stages, where the time factor is very important to discover the abnormality issues in target images, especially in various cancer tumours such as lung cancer, breast cancer, etc. Image quality and accuracy is the core factors of this research, image quality assessment as well as improvement are depending on the enhancement stage where low pre-processing techniques are used .

The proposed technique is efficient for segmentation principles to be a region of interest foundation for feature extraction obtaining.
In recent years, advanced analysis of medical imaging using radiomics, machine, and deep-learning, including convolutional neural networks (CNNs), has been explored. These approaches offer great promise for future applications for both diagnostic and predictive purposes. CNNs are nonexplicitly programmed algorithms that identify relevant features on the images that allow them to classify an input object. They have been applied in various tasks such as detection (e.g., breast lesions on mammographic scans), segmentation (e.g., liver and liver lesions on computed tomography (CT)), and diagnosis (e.g., lung lesions on screening low-dose CT).

3.1.Block Diagram:



3.2 Software Designing

● Jupyter Notebook Environment
● Spyder Ide
●Deep Learning Algorithms(CNN)
●HTML
- Flask

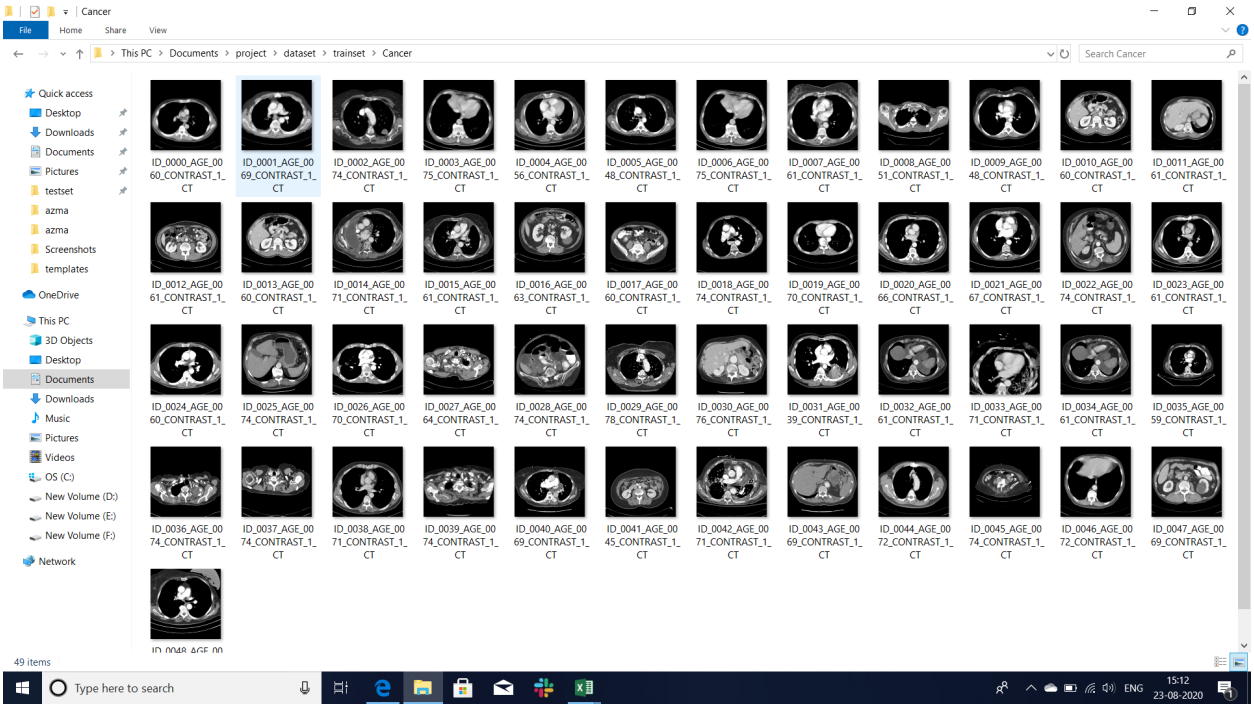
We developed this lung_cancer prediction by using the Python language which is a interpreted and high level programming language and usng the Deep Learning algorithms. For coding we used the Jupyter Notebook environment of the Anaconda distributions and the Spyder, it is an integrated scientific programming in the python language. For creating an user interface for the prediction we used the Flask. It is a micro web framework written in Python. It is classified as a microframework because it does not require par ticular tools or libraries. It has

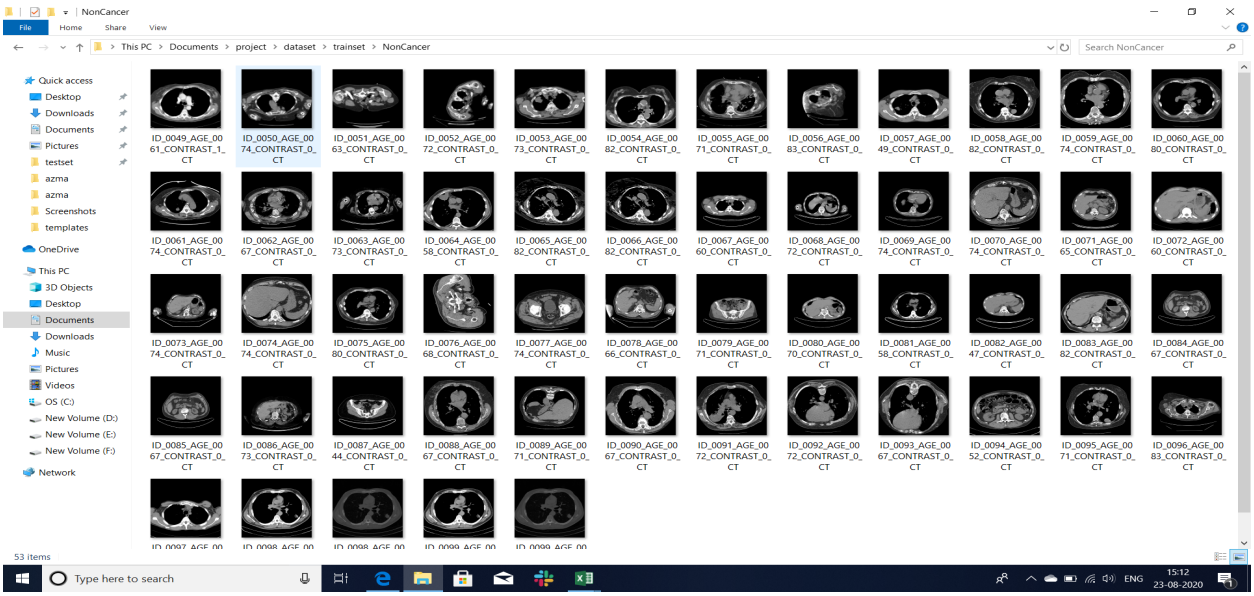
no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions, and a scripting language to create a webpage is HTML by creating the templates to use in th functions of the Flask and HTML.

4. EXPERIMENTAL INVESTIGATION

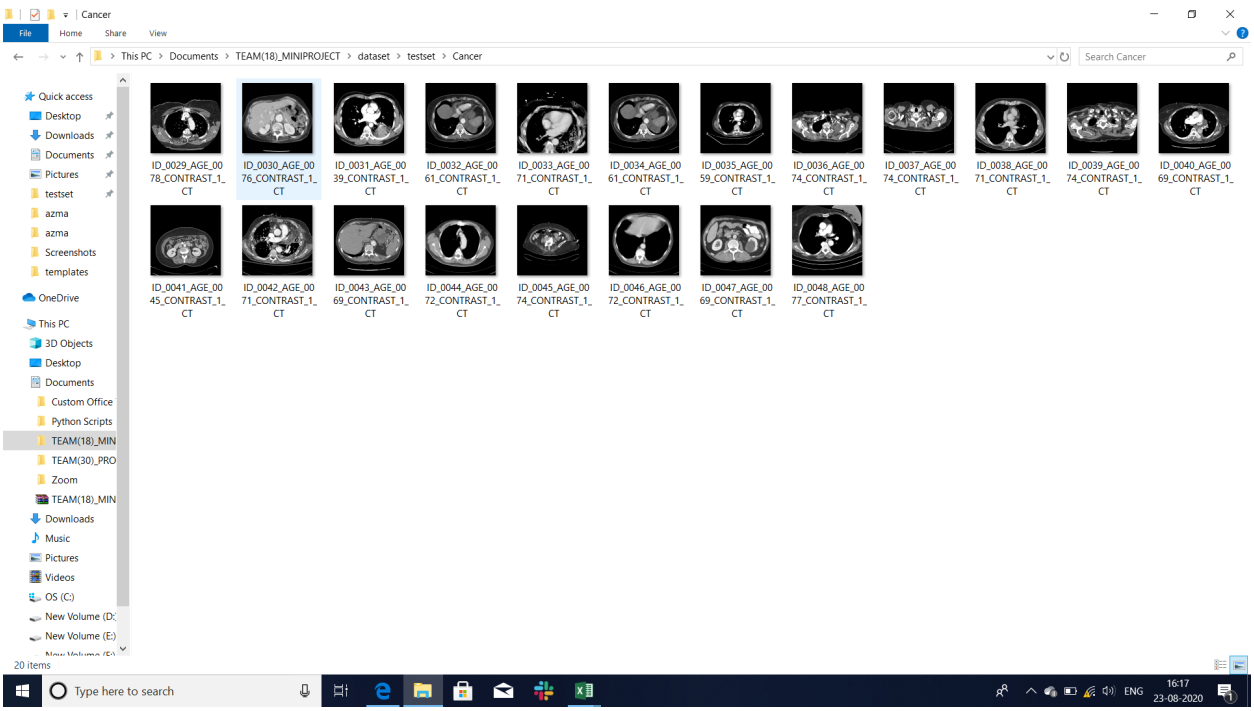
Our dataset consists of two sets of foleders they are trainset and testset.In train set we have two classes named cancer and non-cancer with total 102 images in train set where cancer images are 49 and in noncancer-53.In test set we have two classes named as cancer and non-cancer with total 40 images where 20 images are for cancer and other 20 are for non-cancer.

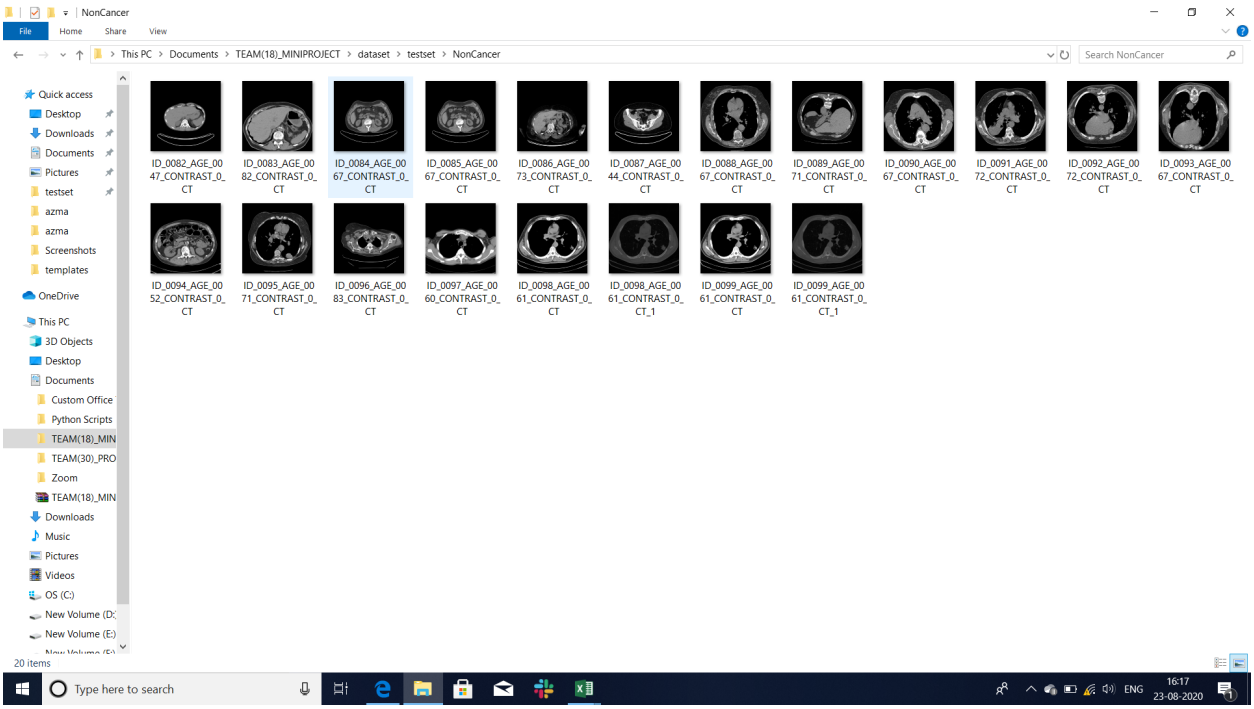
CNNs are a machine-learning technique based on an artificial neural network with deep architecture relying on convolution operations (the linear application of a filter or kernel to local neighbourhoods of pixel/voxels in an input image) and downsampling or pooling operations (grouping of feature map signals into a lower-resolution feature map). The final classification or regression task relies on higher-level features representative of a large receptive field that is flattened into a single vector. The development of an algorithm entails selection of the hyperparameters, training and validation, and testing. The *hyperparameters* include the network topology, the number of filters per layer, and the optimisation parameters. During the *training* process, the dataset of input images (divided into training and validation sets) is repeatedly submitted to the network to capture the structure of the images that is salient for the task. Initially, the weights for each artificial neuron are randomly chosen. Then, they are adjusted at each iteration, targeting minimisation of the loss function, which quantifies how close the prediction is to the target class. The performance of the trained model is then evaluated using an independent *test* dataset. This is also aimed at assessing whether an “overfitting” has occurred. The overfitting problem can arise in the case of limited datasets with too many parameters compared with the dataset size, in which case a model “memorises” the training data rather than generalising from them .





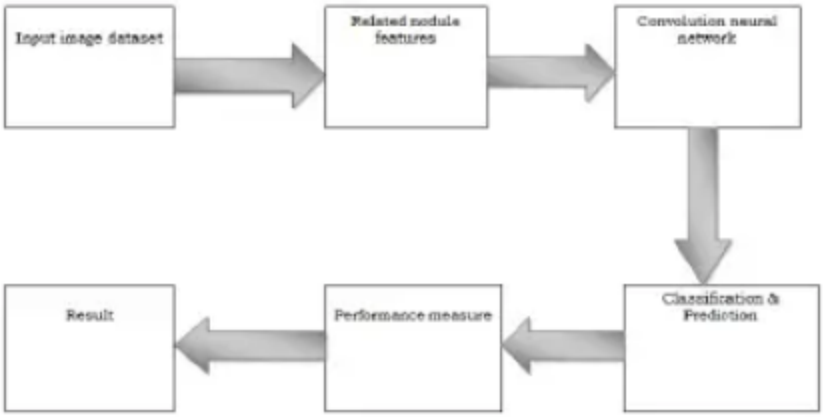
The above figures are of trainset





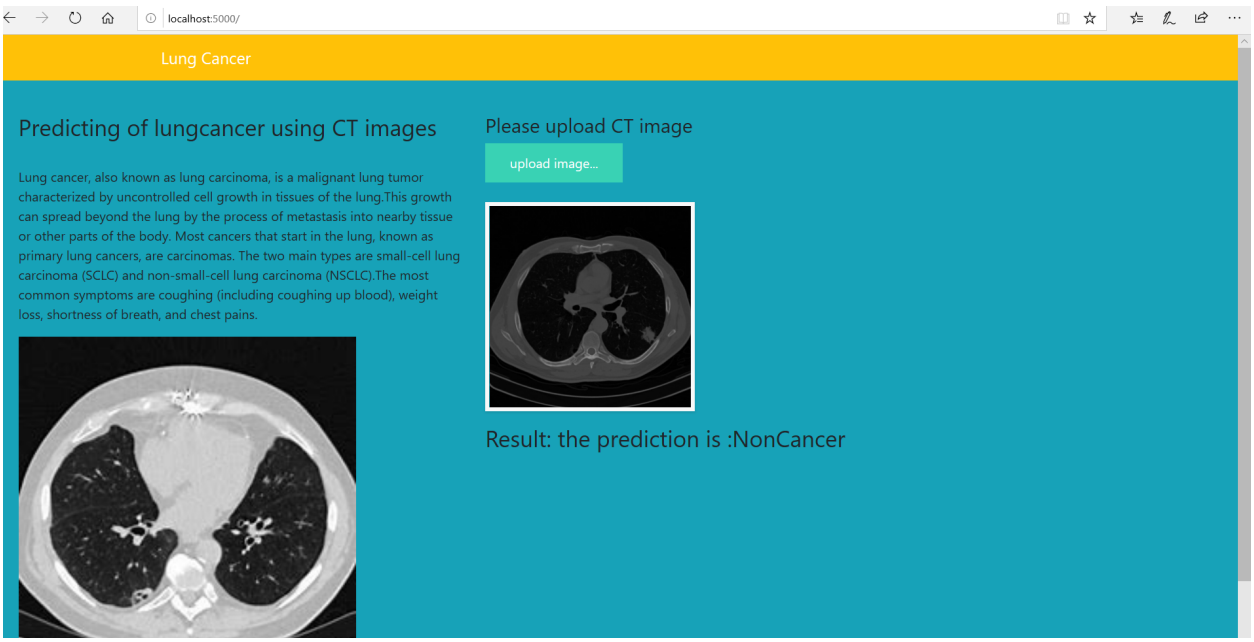
The above figures are of testset

5. FLOWCHART



6. RESULT

In this methodology,we can recognize and detect the lung cancer where it is present and analyze the condition of a person.



7.ADVANTAGES AND DISADVANTAGES

Advantages:

1.Highly detailed

Of all the internal imaging procedures available to physicians, the CT scan is the most

detailed, and can give a doctor the most complete picture of what's happening inside a patient's body. They are particularly useful and widely used in diagnosing cancer and CNN is an added advantage for the physicians to easily predict the lung cancer.

2.Precise

Because CT scan gives a doctor a very clear picture of where a tumor or other problem is located and whether it has spread, it can help her in planning a biopsy, surgery, radiation or other treatment with more precision.

3.Easily prediction can be done which results in faster treatment for the patient.

4.Once trained, the predictions are pretty fast.

Disadvantages:

1.Needs a clear image for prediction

2.Not robust to unclear images it may predict but it may be not robust in predicting.

3.But in order to predict perfectly the Classification of Images should be done with different Positions

8. APPLICATIONS

- Can also be used for different diseases to predict the disease in a short time.
- Not only in predicting the disease but also to predict anything via images this can be used.
- Used to analyze visual imagery.
- Decoding facial recognition
- It can be adapted to other tumour-and cancer diagnosis problems.

9.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.

10.FUTURE SCOPE:

The generality of our method also suggests that it can be adapted to other tumour-and

cancer diagnosis problems.The design of our project does not rely on any particualr feature of lungs or lung cancer,and so we can easily adapt our pipeline to other nodular cancers,or perhaps other diseases.

Radiologists do not arrive at a diagnosis of lung cancer from a single CT scan.They diagnose a particular type of lung cancer using a sequence of CT scans over afew months.They match the behaviour of the nodules over time with a particular subtype of lung cancer.To match the radiologist-level accuracy on the teask,we need to develop a time-varying model of lung cancer that can effectively include a progression of CT scans. In future, we will perform the experiments on a large amount of data and apply more features such as nodule size, texture and position for further improvement. We will also try to apply the state-of-the-art deep CNN methods for higher accuracy and use our method on other types of cancer detection.

11. BIBLIOGRAPHY

1.Mesleh AM. Jordan Journal of Biological Sciences, . 2017;10:239. [Google Scholar]

2. Anjali Sevani, et al. International Journal of Advance Engineering and Research Development, . 2018;5:792. [Google Scholar]

3. Abdillah B, et al. The Asian Journal, . 2016;893:1. [doi 10.1088/1742-6596/893/1/012063] [Google Scholar]

4. Vaishaw K, et al. An Innovative Approach for Investigation and Diagnosis of Lung Cancer by Utilizing Average Information Parameters, Elsevier Procedia of Computer Science, 2018;132:525. [Google Scholar]

5. Dimililer K, et al. Journal of Science and Technology, . 2017;7:131. [Google Scholar]

6. Sakthivel A, Nagarajan A. International Journal of EngineeringSciences and Research Technology, . 2018;7:727. [Google Scholar]

7. Firmino M, et al. BioMedical Engineering, . 2014;13:1. [PMC free article] [PubMed] [Google Scholar]

8. Al-tarawneh MS. Leonardo Electronic Journal of Practices andTechnologies, . 2012;p147 [Google Scholar]

9. Basak P, Nath A. International Journal of Innovative Research in Computer and Communication Engineering, . 2017;5:9708. [Google Scholar]

10. Subhasri P, Padmapriya A. International Journal of Applied Engineering Research, . 2015;10:1951. [Google Scholar]

APPENDIX

HTML:

```
    <html lang="en">
<head>
<meta charset="UTF-8">
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
   <meta http-equiv="X-UA-Compatible" content="ie=edge">
   <title>lungcancer prediction</title>
   <link href="https://cdn.bootcss.com/bootstrap/4.0.0/css/bootstrap.min.css"
rel="stylesheet">
   <script src="https://cdn.bootcss.com/popper.js/1.12.9/umd/popper.min.js"></script>
   <script src="https://cdn.bootcss.com/jquery/3.3.1/jquery.min.js"></script>
   <script src="https://cdn.bootcss.com/bootstrap/4.0.0/js/bootstrap.min.js"></script>
   <link href="{{ url_for('static', filename='css/main.css') }}" rel="stylesheet">
 <style>
 .bg-dark {
  background-color: #42660c!important;
 }
 #result {
  color: #0a1c4ed1;
 }
 </style>
</head>
<body style = "background-image:
url('https://th.bing.com/th/id/OIP.nTe0q5p6UdWbne-dEQiIBgHaHa?w=165&h=180&c=7
&o=5&dpr=1.25&pid=1.7'); background-size:cover;">
   <nav class="navbar navbar-dark bg-warning">
      <div class="container">
         <a class="navbar-brand" href="#">Lung Cancer</a>
      </div>
   </nav>
   <div class="navbar navbar-dark bg-info">
      <div id="content" style="margin-top:2em">
  <div class="container">
   <div class="row">
```

```
  <div class="col-sm-6 bd" >
    <h3>Predicting of lungcancer using CT images</h3>
    <br>
    <p>Lung cancer, also known as lung carcinoma, is a malignant lung tumor
characterized by uncontrolled cell growth in tissues of the lung.This growth can spread
beyond the lung by the process of metastasis into nearby tissue or other parts of the
body. Most cancers that start in the lung, known as primary lung cancers, are
carcinomas. The two main types are small-cell lung carcinoma (SCLC) and
non-small-cell lung carcinoma (NSCLC).The most common symptoms are coughing
(including coughing up blood), weight loss, shortness of breath, and chest pains.</p>
   <img
src="https://th.bing.com/th/id/OIP.nTe0q5p6UdWbne-dEQiIBgHaHa?w=165&h=180&c=7
&o=5&dpr=1.25&pid=1.7", style="height:450px"class="img-rounded" alt="Gesture">
  </div>
  <div class="col-sm-6">
   <div>
    <h4>Please upload CT image</h4>
  <form action = "http://localhost:5000/predict" id="upload-file" method="post"
enctype="multipart/form-data">
   <label for="imageUpload" class="upload-label">
    upload image...
   </label>
   <input type="file" name="image" id="imageUpload" accept=".png, .jpg, .jpeg">
  </form>

  <div class="image-section" style="display:none;">
   <div class="img-preview">
    <div id="imagePreview">
    </div>
   </div>
   <div>
    <button type="button" class="btn btn-info btn-lg " id="btn-predict">submit</button>
   </div>
  </div>
  <div class="loader" style="display:none;"></div>
  <h3>
   <span id="result"> </span>
```

```
  </h3>
 </div>
  </div>

   </div>
 </div>
 </div>
   </div>
</body>
<footer>
   <script src="{{ url_for('static', filename='js/main.js') }}" type="text/javascript"></script>
</footer>
</html>
```

APP.PY:

```
from flask import Flask,render_template,request
import os
from keras.preprocessing import image
from werkzeug.utils import secure_filename
from keras.models import load_model
import tensorflow as tf
global graph
graph = tf.get_default_graph()
import numpy as np
model = load_model("lungcancer.h5")
app = Flask(__name__)
@app.route('/')
def index():
   return render_template("cancer.html", methods = ['GET'])
@app.route('/predict',methods = ['GET','POST'])
def pred():
   if request.method == "POST":
      f = request.files["image"]
      print("hie")
      """take the path of  the current
```

```
        running prog,and concatenate to the
        folder where you wouldlike to save the file"""
        basepath = os.path.dirname(__file__)
        print(basepath)
        file_path = os.path.join(basepath,"uploads",secure_filename(f.filename))
        print(file_path)
        f.save(file_path)
        img = image.load_img(file_path,target_size = (64,64))
        x = image.img_to_array(img)
        x = np.expand_dims(x,axis = 0)
        with graph.as_default():
            p = model.predict_classes(x)
            print(p)
        index = ["Cancer","NonCancer"]
        text = "the prediction is :" +index[p[0]]
        return text
if __name__ == "__main__":
    app.run(debug = True)
```