Pathology Image Analysis For Lung Cancer Classification

Using Ibm Watson Machine Learning

Developed By: G.Ruthvik, V.Bhavani, Sangeetha, Vamshi

Smart Bridge – Mini Project Report

1. INTRODUCTION

1.1 Overview

Automatic defects detection in CT images is very important in many diagnostic and therapeutic applications. Because of high quantity data in CT images and blurred boundaries, tumor segmentation and classification is very hard. This work hasintroduced one automatic lung cancer detectionmethod to increase the accuracy and yield and decrease the diagnosis time. The goal is classifying the tissues to three classes of normal, benign andmalignant. In MR images, the amount of data is too much for manual interpretation and analysis. During past few years, lung cancer detection in CT has become an emergent research area in the field ofmedical imaging system. Accurate detection of size and location of lung cancer plays a vital role in the diagnosis of lung cancer. The diagnosis method consists of four stages, pre-processing of CT images, feature, extraction, and classification, the features are extracted based on DTCWT and PNN. In the last stage, PNN employed to classify the Normal and abnormal.

1.2 Purpose

Cancer is a group of diseases characterized by the uncontrolled growth and spread of abnormal cells. If the spread is not controlled, it can result in death. Lung cancer was the most common cancer in worldwide, contributing 2,093,876 of the total number of new cases diagnosed in 2018.

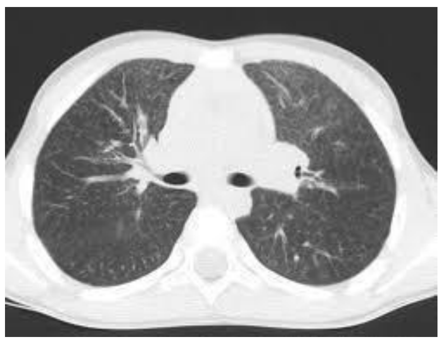
The incidence rate has been declining since the mid- 1980s in men, but only since the mid-2000s in women, because of gender differences in historical patterns of smoking uptake and cessation.

From 2005 to 2015, lung cancer incidence rates decreased by 2.5% per year in men and 1.2% per yearin women. Symptoms do not usually occur until the cancer is advanced, and may include persistent cough, sputum streaked with blood, chest pain, voice change, worsening shortness of breath, and recurrent pneumonia or bronchitis.

Cigarette smoking is by far the most important risk factor for lung cancer; 80% of lung cancer deaths in the US are still caused by smoking. Risk increases with both quantity and duration of smoking. Cigar and pipe smoking also increase risk. Exposure to radon gas released from soil and building materials isthought to be the second-leading cause of lung cancerin the US.

Other risk factors include occupational or environmental exposure to secondhand smoke, asbestos (particularly among smokers), certain metals (chromium, cadmium, arsenic), some organic chemicals, radiation, air pollution, and diesel exhaust. Some specific occupational exposures that increase risk include rubber manufacturing, paving, roofing, painting, and chimney sweeping. Risk is also probably increased among people with a history of tuberculosis. Genetic susceptibility (e.g., family history) plays a role in the development of lung cancer, especially in those who develop the disease at a young age.

We can cure lung cancer ,only if you identifying the yearly stage. So here, we use machine learning algorithms to detect the lung cancer. This can be made faster and more accurate. In this study we propose machine learning strategies to improve cancer characterization. Inspired by learning from CNN approaches, we propose new algorithm, proportion-PNN, to characterize cancer types.



2. LITERATURE SURVEY

2.1 Existing Problem

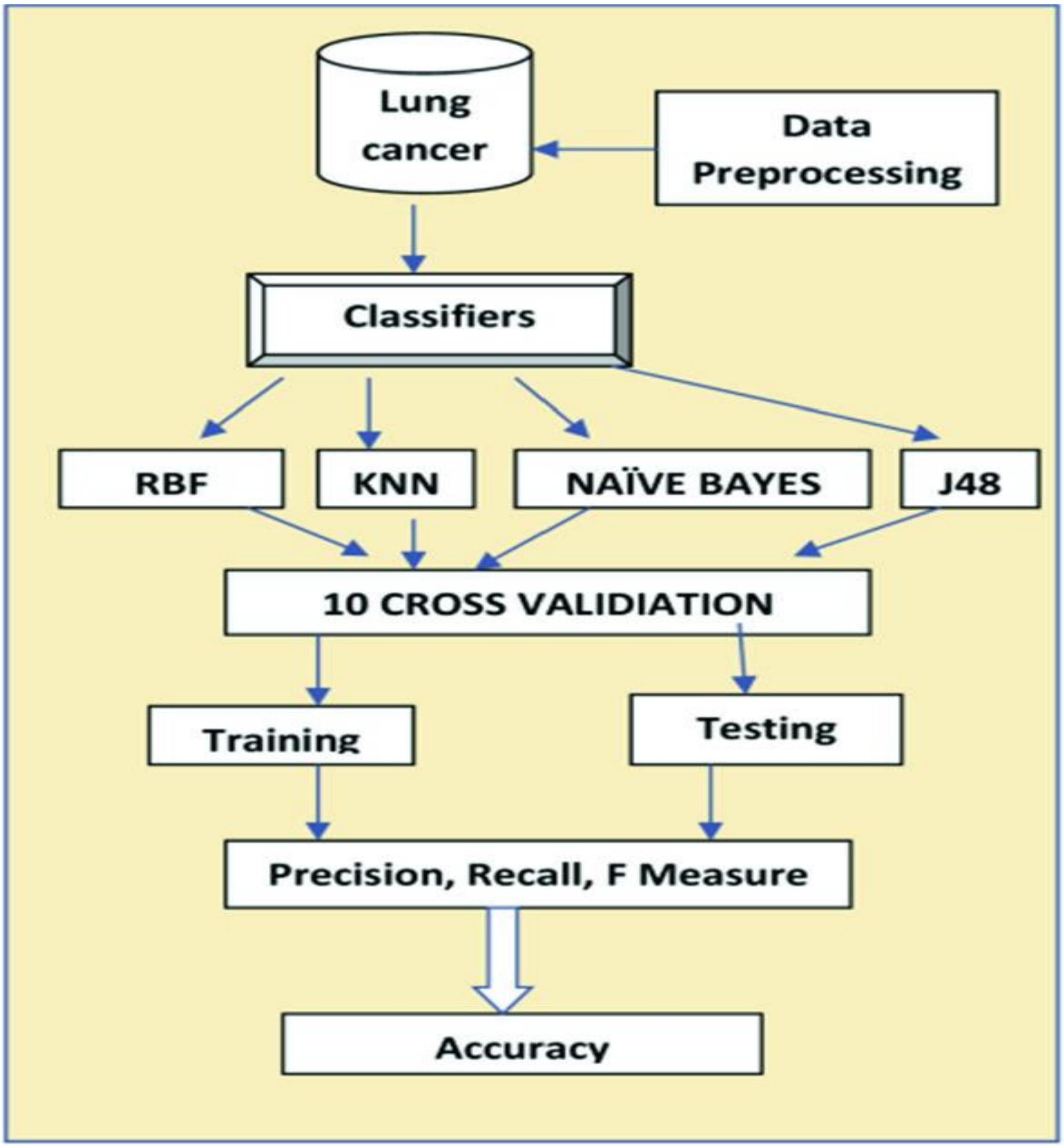
CNN is a class of deep neural network ,but it is done only with the collection of data and it is not labeled. It is most commonly applied to analyze visual imagery. CNN use relatively little pre-processing compared to other image classification algorithm. But ,it is difficult to get accurate results . Not applicable for multiple images for Lung detection in a short time .CNN uses relatively little pre-processing compared to other image classification algorithm

2.2 Proposed Solution

This summarizes the advances in machine learning applied to PNN for the development of lung cancer diagnosis.CNN is a class of deep neural network, but it is done only with the collection of data and it is not labelled. It is most commonly applied to analyze visual images.CNN use relatively little pre- processing compared to other images classification algorithm. since the CNN algorithm takes lot many images as a data to calculate. It creates lagging of time and doesn’t gives accuracy compared to PNN.

3. THEORITICAL ANALYSIS:

3.1 Block diagram



3.2 HARDWARE AND SOFTWARE REQUIREMENTS IN THE PROJECT:

For running a machine learning model on the system you need a system with minimum of 16 GB RAM in it and you require a good processor for high performance of the model.

In the list of **software requirements** you must have:

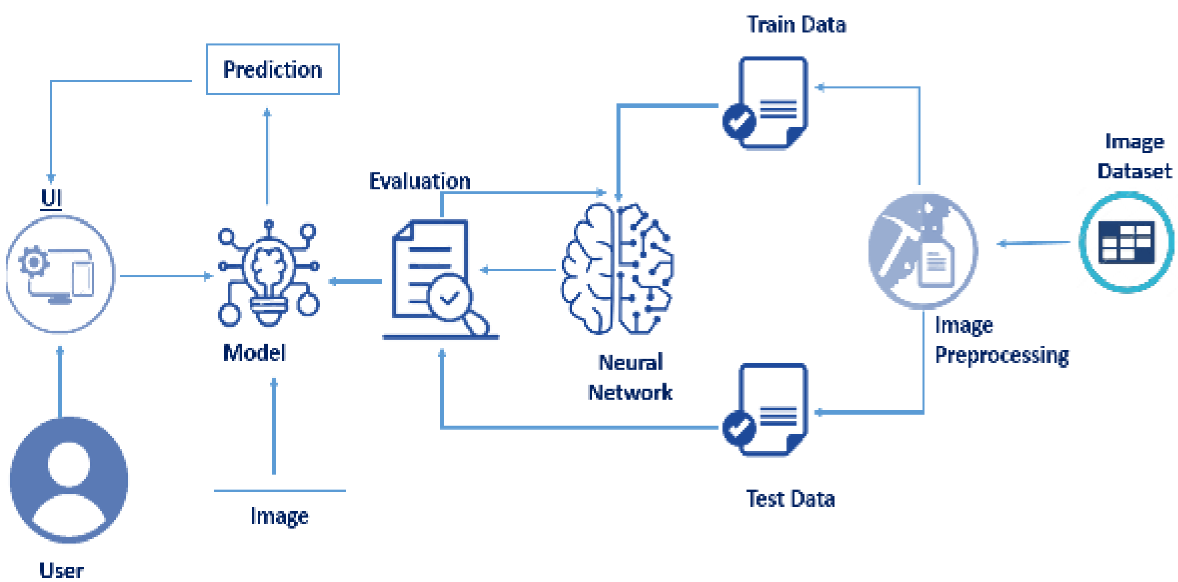
- Jupyter Notebook for programming, which can be installed by Anaconda IDE.
- Python packages
- A better software for running the html,js and css files for application building phase e.g.spyder.

4.EXPERIMENTAL INVESTIGATIONS:

4.1 Image Preprocessing:-

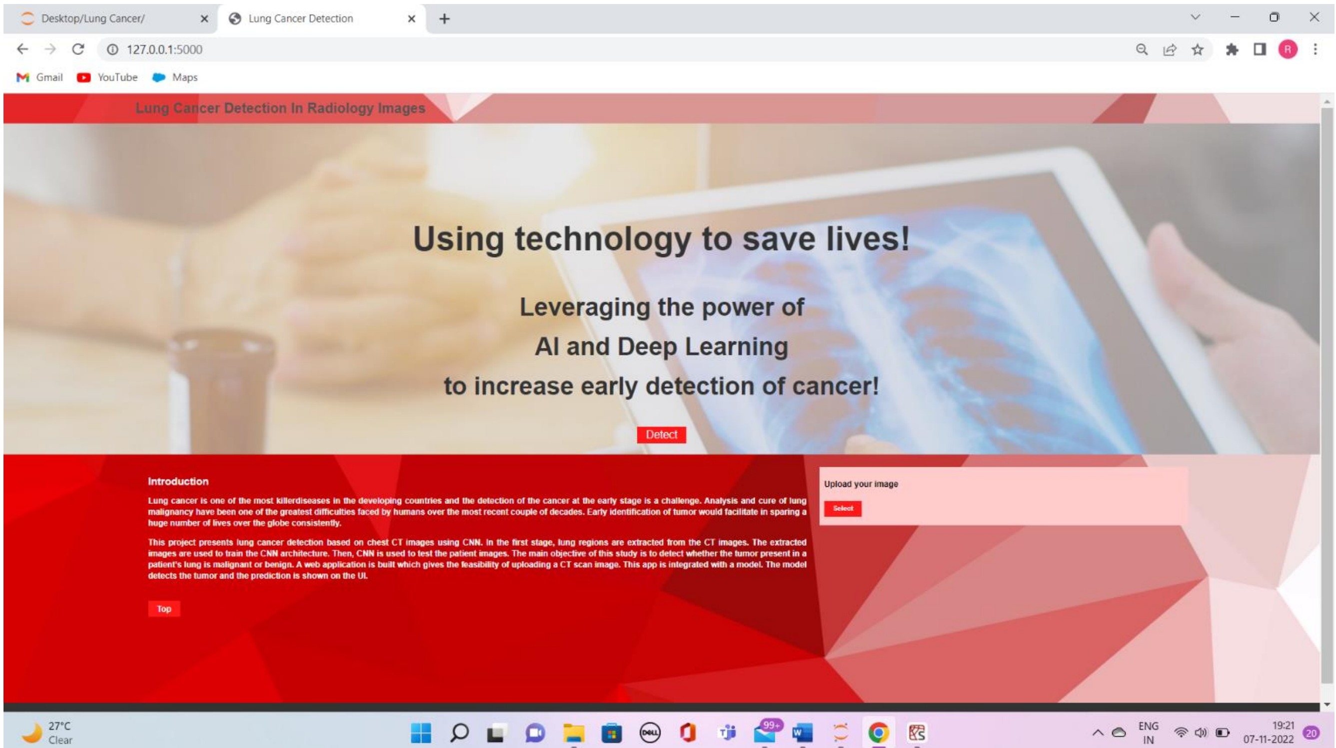
- Import ImageDataGenerator Library.
- Configure ImageDataGenerator Class.
- Applying ImageDataGenerator functionality to the trainset and test set.

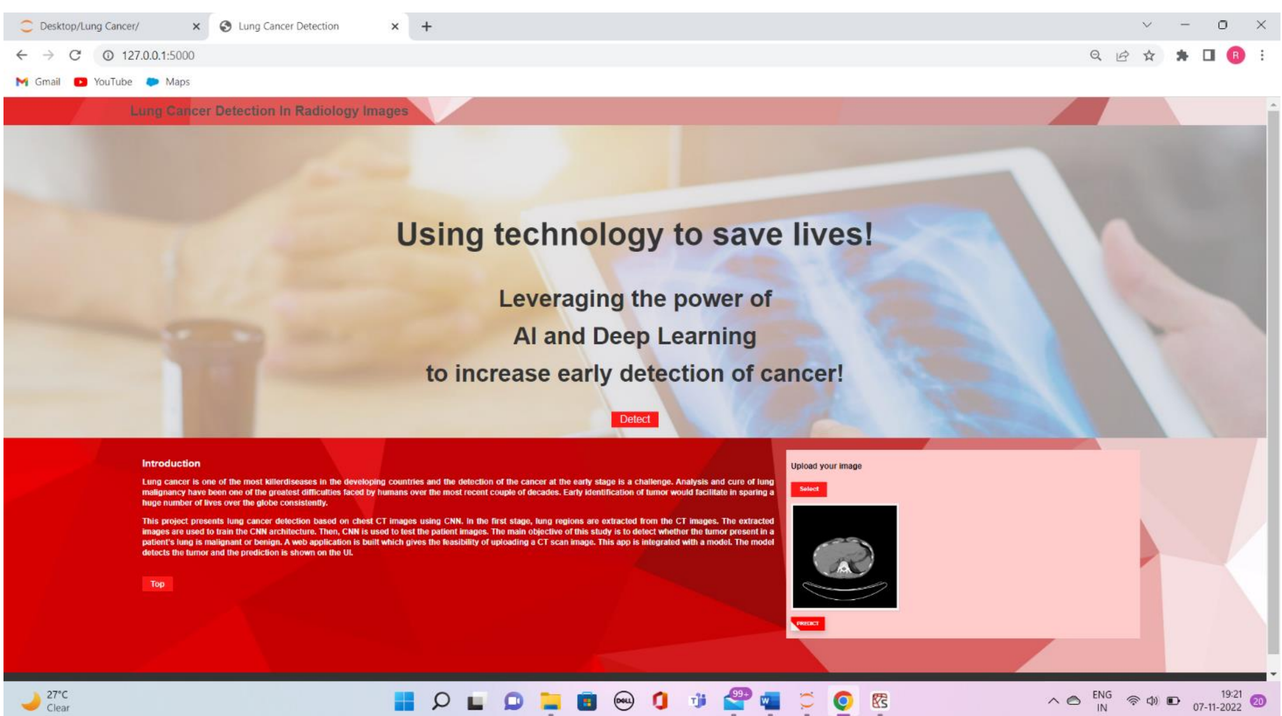
5. FLOW CHART:-

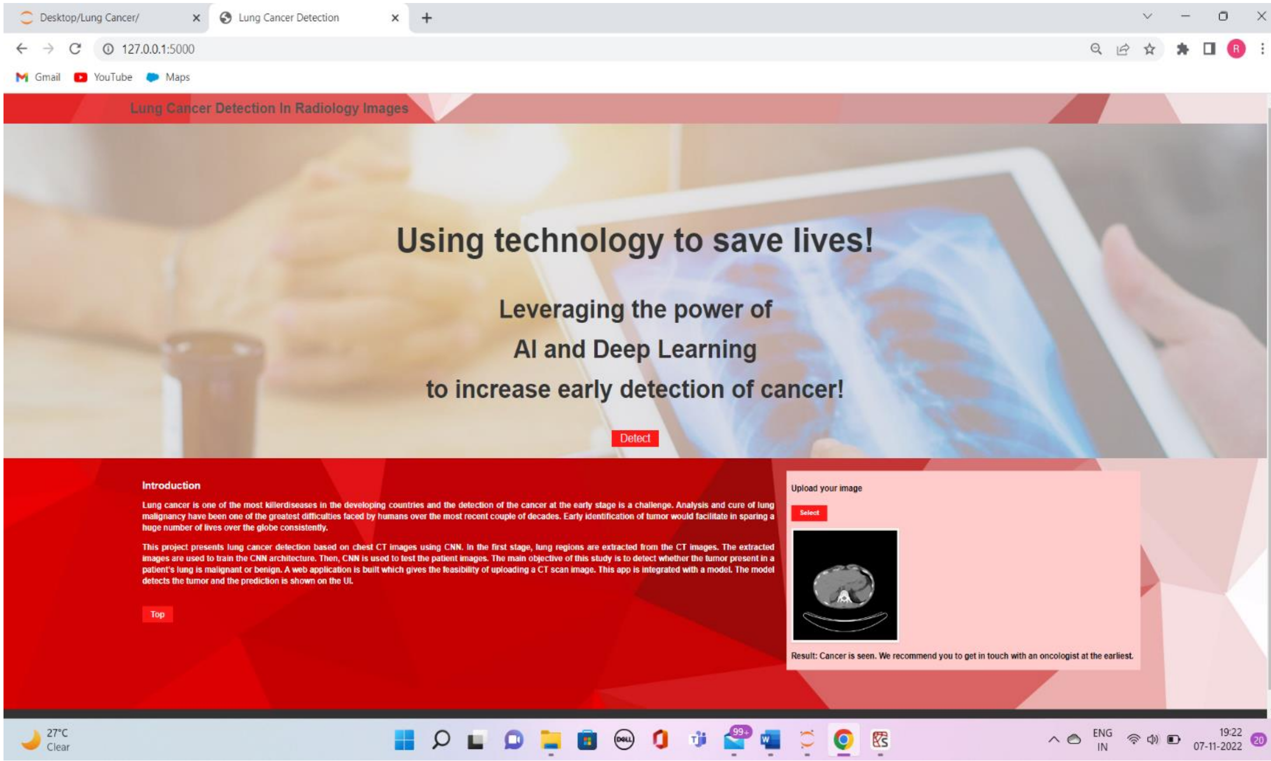


6. RESULTS:-

Final output of the project:






Desktop/Lung Cancer/
Lung Cancer Detection
127.0.0.1:5000
Gmail
YouTube
Maps
Lung Cancer Detection In Radiology Images
Using technology to save lives!
Leveraging the power of
AI and Deep Learning
to increase early detection of cancer!
Detect
Introduction
Lung cancer is one of the most killerdiseases in the developing countries and the detection of the cancer at the early stage is a challenge. Analysis and cure of lung malignancy have been one of the greatest difficulties faced by humans over the most recent couple of decades. Early identification of tumor would facilitate in sparing a huge number of lives over the globe consistently.
This project presents lung cancer detection based on chest CT images using CNN. In the first stage, lung regions are extracted from the CT images. The extracted images are used to train the CNN architecture. Then, CNN is used to test the patient images. The main objective of this study is to detect whether the tumor present in a patient's lung is malignant or benign. A web application is built which gives the feasibility of uploading a CT scan image. This app is integrated with a model. The model detects the tumor and the prediction is shown on the UI.
Top
Upload your image
Select
Result: Cancer is seen. We recommend you to get in touch with an oncologist at the earliest.
27°C
Clear
ENG
IN
19:22
07-11-2022

7. ADVANTAGES & DISADVANTAGES

ADVANTAGES

1. Healthy functional for the data with high dimension.
2. It is adequately grip the larger data sample.
3. The ability to integrate the multiple interpreters in an easy and step by step manner.
4. This is faster due to no training time.
5. Fully automatic, manual intervention is not required. Good classifier at feature detection.
6. Easy to implement and robust to irrelevant attributes.

DISADVANTAGES

1. It required the image intensity value.
2. Difficulty to develop the prototype formation.
3. It has voracious Characteristics.
4. KNN is less functional with noisy data.
5. Prone to overfitting due to complexity of model structure.
6. Performance degraded when features are highly correlated.

8. APPLICATIONS:-

In preparing this review several electronic databases were accessed including PubMed (biomedical literature), the Science Citation Index (biomedical, engineering, computing and physico-chemical literature), CiteSeer (computing literature), Google and Google Scholar (web-accessible scientific literature). Query terms included “cancer and machine learning”, “cancer prediction and machine learning”, “cancer prognosis and machine learning”, “cancer risk assessment and machine learning” as well as multiple sub-queries with specific types of machine learning algorithms. The relevance of the individual papers was assessed by reading the titles and abstracts and identifying papers that used recognizable machine learning methods as well as molecular, clinical, histological, physiological or epidemiological data in carrying out a cancer prognosis or prediction.

Papers that focused on diagnoses or simple tumor classifications were excluded as were papers that had coincidental appearances of the words “machine” or “learning” in their abstracts. A PubMed search of “cancer and machine learning” yielded 1585 results, while searches of “cancer prediction and machine learning” and “cancer prognosis and machine learning” yielded 174 and 240 hits respectively. A detailed review of these abstracts led to the identification of 103 relevant papers of which 71 could be accessed through various library holdings. Using CiteSeer, a search with the terms “cancer and machine learning” yielded 349 results, of which 12 (3.4%) were deemed relevant to cancer prognosis. Using Google Scholar, a search using “cancer prognosis and ‘machine learning’” yielded 996 results, of which 49 (4.9%) were judged relevant to cancer prognosis.

9. CONCLUSION-

Lung cancer is one of the most dangerous diseases and the most common cause of death, the severity of the disease liesin the difficulty of diagnosing it in the early stages. This paper tries to endeavor to investigate of three classifiers to find the best classifier could classify lung cancer in early stage. The informational indices included in this study were derived from UCI databases for lung cancer patients. The focus of this paper is on using WEKA Tool to investigate the accuracy of classification algorithms. The results show that the Support Vector Machine (SVM) give the best accuracy 95.56%, that can detect lung cancer in its early stages and save several lives and, K-Nearest Neighbor KNNIt gave less accuracy 88.40%.

FUTURE SCOPE:-

Several machine learning techniques have been used to detect, predict, compare, or classify lung cancer. However, there are lessons and research directions to advance research in lung and related diseases:

Several models have been used to detect, predict, compare or classify lung cancer using machine learning techniques. However, research is scarce in the use of modern soft computing to provide a high degree of accuracy. Soft computing such as ABC, PSO, Genetic algorithm, functional approximation etc., as a single or hybrid model to understand their strength and weakness in such domain..

Research works into the applications of machine learning in lung cancer are increasing. However, most of the works focused on early detection; there is a need to advance to severity level and other lung cancer components to support medical practitioners in the daily works. Most of the research works used CT scan images to detect, predict, compare or classify lung cancer using classical machine learning techniques, and there is a need to use other data set such as family history, personal attributes and X-ray images to understand if they can provide insights into the presence of lung cancer and related diseases.

10. BIBILOGRAPHY:-

[1] Arnaud A. A. Setio, Francesco Ciompi, GeertLitjens, Paul Gerke, Colin Jacobs, Sarah J. van Riel,Mathi"Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks"(2016).

[2]Junyuan Xie, Ross Girshick "Unsupervised DeepEmbedding for Clustering Analysis"(2016).

[3] Mario Buty1, Ziyue Xu1 , Mingchen Gao" Characterization of Lung Nodule Malignancy usingHybrid Shape and Appearance Features"(2017)

[4]Alan L. Yuille4 " Deep Supervision for Pancreatic Cyst Segmentation in Abdominal CT Scans"(2018)

[5]Kumar, D., Wong, A., Clausi, D.A.: Lung nodule classification using deep features in CT images. In: Computer and Robot Vision (CRV), 2015 12th Conference on. pp. 133–138. IEEE (2015).

APPENDIX :-

app.py

```
import numpy as np
import os
from tensorflow.keras.models import load_model
from tensorflow.keras.preprocessing import image
import tensorflow as tf
global graph
#graph = tf.get_default_graph()
from flask import Flask , request, render_template
from werkzeug.utils import secure_filename
from gevent.pywsgi import WSGIServer
import cv2

app = Flask(__name__)
model = load_model("Lungcancer.h5")

@app.route('/')
def index():
    return render_template('base.html')

@app.route('/predict',methods = ['GET','POST'])
def upload():
    if request.method == 'POST':
        f = request.files['image']
        print("current path")
        basepath = os.path.dirname(__file__)
        print("current path", basepath)
        filepath = os.path.join(basepath,'uploads')
        print("upload folder is ", filepath)
        f.save(filepath)

        img = image.load_img(filepath,target_size = (64,64))
        gray = cv2.cvtColor(np.float32(img), cv2.COLOR_BGR2GRAY)
        print(gray.shape)
        x = image.img_to_array(gray)
        x = np.expand_dims(x,axis =0)

        #with graph.as_default():
        preds = model.predict(x)

        print("prediction",preds)

        index = ['Cancer','NonCancer']
        text = str(index[int(preds[0][0])])

        if (text == "Cancer"):
            text = "Cancer is seen. We recommend you to get in touch with an oncologist at the
```

```
earliest."
        else:
            text = "Cancer not seen. Stay safe and healthy."

    return text
if __name__ == '__main__':
    app.run(debug = False)
```

main.js

```
$(document).ready(function () {
    // Init
    $('.image-section').hide();
    $('.loader').hide();
    $('#result').hide();

    // Upload Preview
    function readURL(input) {
        if (input.files && input.files[0]) {
            var reader = new FileReader();
            reader.onload = function (e) {
                $('#imagePreview').css('background-image', 'url(' + e.target.result + ')');
                $('#imagePreview').hide();
                $('#imagePreview').fadeIn(650);
            }
            reader.readAsDataURL(input.files[0]);
        }
    }
    $("#imageUpload").change(function () {
        $('.image-section').show();
        $('#btn-predict').show();
        $('#result').text('');
        $('#result').hide();
        readURL(this);
    });

    // Predict
    $('#btn-predict').click(function () {
        var form_data = new FormData($('#upload-file')[0]);

        // Show loading animation
        $(this).hide();
        $('.loader').show();

        // Make prediction by calling api /predict
        $.ajax({
            type: 'POST',
            url: '/predict',
            data: form_data,
            contentType: false,
```

```
            cache: false,
            processData: false,
            async: true,
            success: function (data) {
                // Get and display the result
                $('.loader').hide();
                $('#result').fadeIn(600);
                $('#result').text(' Result:  ' + data);
                console.log('Success!');
            },
        });
    });

});
```

base.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>Lung Cancer Detection</title>

            <link href="{{ url_for('static', filename='css/style.css') }}" rel="stylesheet">
            <link href="{{ url_for('static', filename='css/main.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>




</head>

<body>
            <header id="main-header">
                        <div class="container">
                                    <h1>Lung Cancer Detection In Radiology Images</h1>
                        </div>
            </header>

    <div class="parallax">
            <section id="showcase">
                        <div class="container">
                                    <h1 style=font-size:77px><br><br>Using technology to save
lives!</h1><h2 style=font-size:57px>Leveraging the power of<br>AI and Deep Learning<br>to
increase early detection of cancer!</h2>
```

```
                    <center> <a href="#section2" class="upload-label" style="color:white;font-
size:25px ;text-decoration:none">Detect</a></center><br>
                    </div>
          </section>
   </div>

   <div class="background" id="section2" style="height: 543px;">

          <div class="container">
                    <section id="main">
          <div class="justify">
                              <h1><b>Introduction</h1>
                              <p><h3 style=font-size:19px> Lung cancer is one of the most killerdiseases
in the
          developing countries and the detection of the cancer at the early
          stage is a challenge. Analysis and cure of lung malignancy have
          been one of the greatest difficulties faced by humans over the
          most recent couple of decades. Early identification of tumor
          would facilitate in sparing a huge number of lives over the globe
          consistently.</p>
                              <p>This project presents lung cancer detection based on chest CT images
using CNN. In the first stage,
                              lung regions are extracted from the CT images. The extracted images are
used to train the CNN architecture.
                              Then, CNN is used to test the patient images. The main objective of this
study is to detect whether the tumor
                              present in a patient’s lung is malignant or benign. A web application is built
which gives the feasibility of
                              uploading a CT scan image. This app is integrated with a model. The model
detects the tumor and the prediction
                              is shown on the UI. <br><br><left> <div class="upload-label">
             <left><a href="#" style="color:white; text-decoration:none">Top</a></left>
       </div></left><br></h3></p>

       </div>

                    </section>

                    <aside id="sidebar">
                    <div class="col-sm-6">
                                        <div>
                                                  <h4 style=font-size:19px>Upload your image</h4>
                              <form action = "http://localhost:5000/predict" id="upload-file"
method="post" enctype="multipart/form-data">
                                        <label for="imageUpload" class="upload-label">
                                                  Select
                                        </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="myButton" id="btn-
predict">Predict</button>
                                                </div>

                                    </div>

                                    <div class="loader" style="display:none;"></div>

                                    <h3>
                                                <span id="result"> </span>
                                    </h3>
                        </aside>
            </div>

    </div>

<script>
window.onscroll = function() {myFunction()};

var header = document.getElementById("myHeader");
var sticky = header.offsetTop;

function myFunction() {
  if (window.pageYOffset > sticky) {
    header.classList.add("sticky");
  } else {
    header.classList.remove("sticky");
  }
}
</script>

            <footer id="main-footer">
                        <p>Copyright &copy; 2020 RSS Pvt Ltd</p>
                        <script src="{{ url_for('static', filename='js/main.js') }}"
type="text/javascript"></script>
            </footer>
</body>
</html>
```

main.css

```css
.img-preview {
    width: 250px;
    height: 250px;
    position: relative;
    border: 5px solid #F8F8F8;
    box-shadow: 0px 2px 4px 0px rgba(0, 0, 0, 0.1);
    margin-top: 1em;
    margin-bottom: 1em;
}

.img-preview>div {
    width: 100%;
    height: 100%;
    background-size: 250px 250px;
    background-repeat: no-repeat;
    background-position: center;
}

input[type="file"] {
    display: none;
}

.upload-label{
    display: inline-block;
    padding: 6px 20px;
    background: #ff1a1a;
    color: #fff;
    font-size: 1em;
    transition: all .4s;
    cursor: pointer;
}

.upload-label:hover{
    background: #ff0000;
    color: #fff;
}

.loader {
    border: 8px solid #f3f3f3; /* Light grey */
    border-top: 8px solid #ff0000; /* Red */
    border-radius: 50%;
    width: 50px;
    height: 50px;
    animation: spin 1s linear infinite;
}

@keyframes spin {
    0% { transform: rotate(0deg); }
    100% { transform: rotate(360deg); }
```

}

style.css

```
body{
        background-image:url(bg.jpg);
    background-size: 400% auto;
        background-repeat: no-repeat;
        background-position:center;
        color:#555;
        font-family:Arial, Helvetica, sans-serif;
        font-size:16px;
        line-height:1.6em;
        margin:0;
}

.container{
        width:80%;
        margin:auto;
        overflow:hidden;
}

.justify{
    text-align:justify;
    text-justify: auto;
}

.parallax {
  /* The image used */
     background-image: url("doc.jpg");

  /* Set a specific height */
  min-height: 750px;

  /* Create the parallax scrolling effect */
  background-attachment: fixed;
  background-position: center;
  background-repeat: no-repeat;
  background-size: cover;
}

html {
  scroll-behavior: smooth;
}
```

```
#section2 {
  height: 500px;
  background: ;
}
div.background {
  background: url("static/bgg2.jpg");
  min-height: 5px;
 background-attachment: fixed;
  background-position: center;
  background-repeat: no-repeat;
  background-size: cover;
}


#navbar{
	background-color:#fff;
	color:#333;
}

#navbar ul{
	padding:0;
	list-style: none;
}

#navbar li{
	display:inline;
}

#navbar a{
	color:#fff;
	text-decoration: none;
	font-size:18px;
	padding-right:15px;
}

#showcase{
	min-height:300px;
	margin-bottom:30px;
}


#showcase h1{
    width: 100%;
	color:#333;
```

```
        font-size:40px;
        text-align: center;
        line-height: 1em;
        padding-top:10px;
}
#showcase h2{
    width: 100%;
        color:#333;
        font-size:30px;
        text-align: center;
        line-height: 1.6em;
        padding-top:10px;
}

#main{
        float:left;
        color:#fff;
        width:65%;
        padding:0 30px;
        box-sizing: border-box;
}

#sidebar{
        float:right;
        width:35%;
        background-color: #ffcccc;
        color:#000;
        padding-left:10px;
        padding-right:10px;
        padding-top:1px;
        box-sizing: border-box;
}


.img-preview {
    width: 10px;
    height: 10px;
    position: relative;
    border: 5px solid #F8F8F8;
    box-shadow: 0px 2px 4px 0px rgba(0, 0, 0, 0.1);
    margin-top: 1em;
    margin-bottom: 1em;
}
```

```
.img-preview>div {
   width: 10%;
   height: 10%;
   background-size: 100px 10px;
   background-repeat: no-repeat;
   background-position: center;
}

input[type="file"] {
   display: none;
}

.upload-label{
   display: inline-block;
   padding: 12px 30px;
   background: #39D2B4;
   color: #fff;
   font-size: 1em;
   transition: all .4s;
   cursor: pointer;
}

.upload-label:hover{
   background: #34495E;
   color: #39D2B4;
}

.myButton {
  border: none;
  text-align: center;
  cursor: pointer;
  text-transform: uppercase;
  outline: none;
  overflow: hidden;
  position: relative;
  color: #fff;
  font-weight: 700;
  font-size: 12px;
  background-color: #ff0000;
  padding: 10px 15px;
  margin: 0 auto;
  box-shadow: 0 5px 15px rgba(0,0,0,0.20);
}
```

```
.myButton span {
  position: relative;
  z-index: 1;
}

.myButton:after {
  content: "";
  position: absolute;
  left: 0;
  top: 0;
  height: 310%;
  width: 150%;
  background: #f2f2f2;
  -webkit-transition: all .5s ease-in-out;
  transition: all .5s ease-in-out;
  -webkit-transform: translateX(-98%) translateY(-25%) rotate(45deg);
  transform: translateX(-98%) translateY(-25%) rotate(45deg);
}

.myButton:hover:after {
  -webkit-transform: translateX(-9%) translateY(-25%) rotate(45deg);
  transform: translateX(-9%) translateY(-25%) rotate(45deg);
}

.loader {
    border: 8px solid #f3f3f3; /* Light grey */
    border-top: 8px solid #ff0000; /* Red */
    border-radius: 50%;
    width: 50px;
    height: 50px;
    animation: spin 1s linear infinite;
}

@keyframes spin {
    0% { transform: rotate(0deg); }
    100% { transform: rotate(360deg); }
}

#main-footer{
        background: #333;
        color:#fff;
        text-align: center;
        padding:1px;
        margin-top:0px;
```

```
}

@media(max-width:600px){
	#main{
		width:100%;
		float:none;
	}

	#sidebar{
		width:100%;
		float:none;
	}
}
```