Lung Cancer Detection & Recommendation of Oncologist using Machine Learning

Pradnya Mehta
Assistant Prof.
Department of Computer
Marathwada Mitra Mandal's College of Engineering
Pune,India.

Nachiket Kelkar
UG Student
Department of Computer
Marathwada Mitra Mandal's College of Engineering
Pune,India.

Niraj Mate UG Student
Department of Computer
Marathwada Mitra Mandal's College of Engineering
Pune,India.

Atharv Kukade UG Student
Department of Computer
Marathwada Mitra Mandal's College of Engineering
Pune,India.

Abhijit Kulkarni UG Student
Department of Computer
Marathwada Mitra Mandal's College of Engineering
Pune,India.

Abstract: Lung cancer is one of the most prominent and deleterious forms of cancer and affects about 234,030 people every year on an average. The odds of a man developing lung cancer in his lifetime are 1 in 15. On a positive note, Lung Cancer death rates have significantly declined over the past decade due to early detection and treatment. Computed Tomography (CT) can be more efficient than X-ray. Hence, the proposed system uses CT images for detection of lung cancer. The proposed system contains several steps like image acquisition, pre-processing, binarization, thresholding, segmentation, feature extraction and detection of the presence and the stage of cancer if it is present. In first stage, Binarization technique is used to convert CT image to binary image and then compare it with threshold value to detect lung cancer. In second stage, segmentation is performed to segment the lung CT image and a strong feature extraction method has been introduced to extract some important features of segmented images like area, perimeter and roundness of Cancer cells. Extracted features are given to the system to detect the presence and hence the stage of existing cancer. The system then goes ahead and generates a report which is sent to the doctor directly for further analysis.

General Terms **- Machine Learning, Image Processing.**

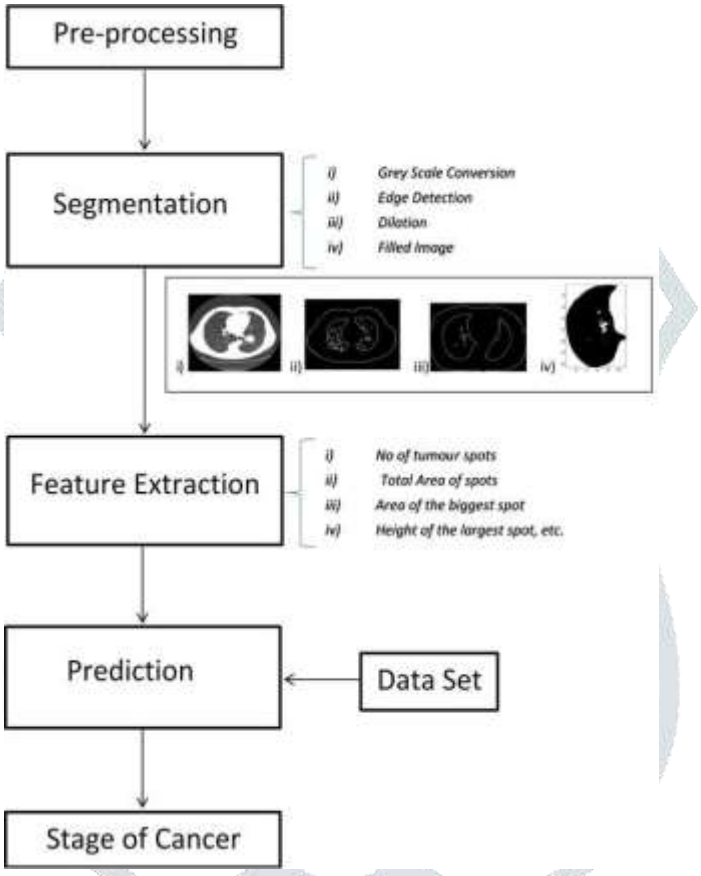
Keywords: **Preprocessing, Binarization, Segmentation and Feature Extraction.**

I. INTRODUCTION

With the rising trends in technology, almost everything is being done over the internet. Many healthcare related web applications and websites are coming forward like TreatHF, Preconception care, Mayo Clinic etc. This benefits many individuals seeking preliminary medical assistance at the tip of their fingers before receiving proper help from the appropriate medical practitioner. Data analytics and Machine Learning can also assist doctors by analyzing patient data and giving a prediction based on it. That being said, this rise in technology in no way indicates that it can replace the proper trained professionals in any way. All of these technologies can be seen more as tools to make the process of medical diagnosis and healthcare easier for the patients as well as the concerned physicians. With this paper we propose a web-based system for the detection of lung cancer using the CT scan image that is uploaded by the patient onto the site. On the server end this image goes through various steps and a report is generated based on it which is sent to the doctor. The purpose of this report is to serve as a reference to the doctor to aid with the diagnosis procedure and provide proper consultation to the concerned patient.

II. SYSTEM ARCHITECTURE

The purpose of the proposed system is to aid the physicians in the procedure for diagnosis of lung cancer. A large dataset containing various images for cancer research purposes is used to train the machine learning model in order to identify the stage of cancer if present. When an image is uploaded by the patient onto the system it undergoes various preprocessing steps in order to remove unnecessary noise. Then the image is converted to greyscale and edges are detected in order to extract the required features of cancerous cells. This is done using OpenCV (Open Source Computer Vision Library) which contains interfaces to support image processing operations. Various features are extracted from the image like the number of tumor spots, total area of the spots, area of the biggest spot, Roundness etc. The system architecture is as shown below:



Each module of proposed system architecture explained below:

a. Preprocessing

First step after taking the CT image as input is to apply pre-processing on it. The input CT image may contain errors related to the brightness values, geometry, and lack of contrast of the pixels. Following preprocessing steps applied to input CT image to convert the image to a form which is better suited for machine interpretation.

1. Grey Scale Conversion

A grayscale converted image

It is also called as color filtering which means to convert RGB image frames into HSV (Hue Saturation Value) image. The image is further converted in a gray scale image. Gray scale images are logically a matrix with each pixel represented as a discrete level out of 0 to 255 [10]. Gray Scale also helps to remove hardware errors if any.

2. Noise Removal

Noise Filtering will remove the unnecessary information from an image. It is also used to remove various types of noises from the images. Noise removal uses filters like low pass, high pass, mean, median etc.

b. Segmentation

1. Binarization

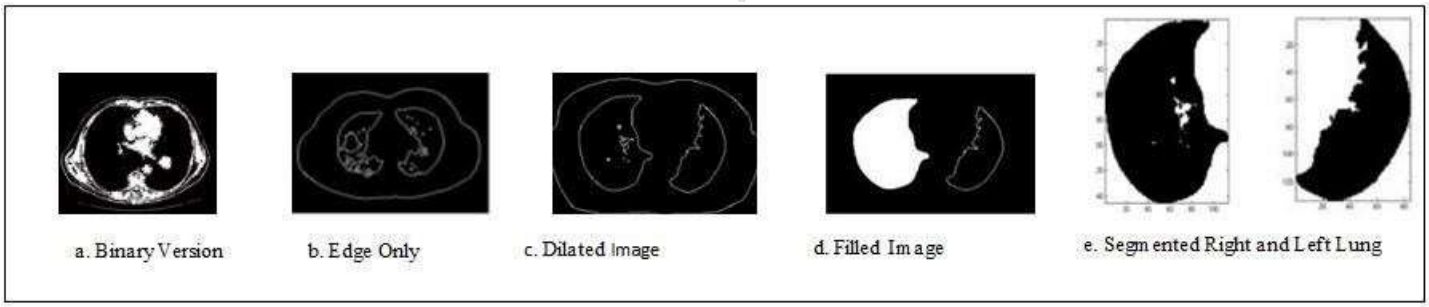
Binarization will convert the gray scale CT image to a binary image. The binary image contains only two pixels black and white. Here the tumors in the CT image lie on the higher side of the spectrum so the threshold value is set based on that. The pixels with tumor spots and other places having greater value than threshold will be converted to the value 1, while the others will be neglected and converted to 0.

2. Edge Detection

Edge detection methods transform original images into edge images. The edge representation of an image significantly reduces the quantity of data to be processed, yet it retains essential information regarding the shapes of tumor spots in the binarized CT image. Here the edges of the tumor spots will be detected and shown in the output of this stage.

3. Dilation and Filled Image

Dilation is used to add pixels at region of boundaries or to fill in holes which generate during erosion process. Dilation can also be used to connect disjoint pixels and add pixels at edges. The dilated image is further converted into a filled image which highlights the areas which are important.



c. Feature Extraction

This is a crucial and essential stage after image segmentation. The normality or abnormality of an image is determined by the final results of feature extraction stage.

Extracted features works as the base for classification process. Following features will be extracted:

1. Number of tumor spots.
2. Total area of the spots.
3. Area of the largest spot.
4. Height of the largest spot.

The features are defined as follows:

1. Number of tumor spots: It is a numerical value which is taken after calculating total number of tumor spots in the binarized CT image.
2. Total area of the spots: This value is retrieved by the summation of areas in the image which are marks as 1 in the binary CT image.
3. Area of the largest spot: This is the numerical value which is obtained after identifying the largest spot in the CT image.
4. Height of the largest spot: After the spot identification which is largest of all the spots in the CT image, height iscalculated.

d. Prediction

After the feature extraction we get the features of the CT image which helps us to do the prediction of stage of cancer. While predicting the stage of cancer we compare the extracted features of current CT image with the available features of various CT images in the dataset used.

III. LITERATURE REVIEW

In the papers referred for this survey, Convolutional Neural Network (CNN) based on deep learning technique is used for training patients' dataset. CNN having the advantage of both classification and prediction of input serves a robust and reliable algorithm to train against large datasets as CNN has N number of parameters to be manipulated for accurate training. However, this paper only classifies images into whether the cancer is present or absent [1]. Using neural networks for classification the design of the system was observed to be ideal for performing as a Computer Aided Detection (CADe) system for detection of lung cancer [2]. An optimized Lung cancer detection system approach shows that, Lung Cancer Detection from CT images can be done in four stages. In this case BAT Algorithms are used to improve performance system and ANN is applied to revamp accuracy of results [3]. For reference on using image processing techniques this paper was referred which presents in detail literature survey on various techniques that have been used in Pre-processing, nodule segmentation and classification [4]. In one of the proposed approaches, image preprocessing techniques used like median filter for noise removal, High boost operator for enhancement and marker-controlled watershed used for image segmentation then the feature extraction give suspicious region of interest of tumor. Finally describes the various classification techniques for detecting lung cancer [5]. Using the genetic approach, it can be observed that manual interpretations are time consuming and very critical, to overcome this difficulty the best features of Genetic Algorithm method and Naive Bayes Classification are taken to classify the different stages of the cancer images fast and accurately [6]. One of the proposed systems in a research paper that was referred is designed to detect lung cancer in premature stage in two stages. In first stage, Binarization technique is used to convert binary image. In second stage, segmentation is performed to segment the lung CT image. The proposed system is designed such that it can detect which lung is affected left lung or right lung specifically [7]. The different stages in a Computer Aided Lung Cancer Detection system are Enhancement, Segmentation and Feature extraction. There are different techniques for performing these stages. These different techniques have been explained in detail [8]. In a few of the papers referred, MATLAB was used to perform image processing tasks on the CT image and this research paper provides an alternative approach to it. In image processing steps, processes such as image pre-processing, segmentation and feature extraction have been discussed in detail [9].

IV. EXPECTED RESULTS

1. Stage of cancer.
2. Suggestion of name of Oncologists.

V. FUTURE WORK

In future scope this project, we can expand on the current system to tract the patient's consultations with the doctor and see the yielded results. In this way the system will become more involved with the patient by fetching more data on the patient through various other sources.

VI. DISCUSSION

The proposed system will be able to detect not only the presence of lung cancer in a patient's CT scan image but also classify it according to its stage and generate a detailed report for the concerned physician to see. A dataset of many cancer patients with varying levels of the disease is used to train the learning model. It aims to provide a web-based platform for patients to upload their image and to get their report generated and sent to their doctor to get further analysis.

VII. ACKNOWLEDGEMENT

We would like to thank Prudent software and grooming academy for their valuable guidance and continuous support.

VIII. REFERENCES

[1] Karan Sharma, Harshil Soni and Kushika Agarwal**, "**Lung Cancer Detection in CT Scans of Patients Using Image Processing and Machine Learning Technique", Springer 2018.

[2] "Lung Cancer Detection Using Digital Image Processing and Artificial Neural Networks." S.Kalaivani, Pramit Chatterjee, Shikhar Juyal, Rishi Gupta, ICECA 2017

[3] "An optimized Lung Cancer Classification System for Computed Tomography Images." Sheenam Rattan, Sumandeep Kaur, Nishu Kansal, Jaspreet Kaur IEEE 2017.

[4] "A Review on Image Processing Methods in Detecting Lung Cancer using CT Images." G.Niranjana, Dr.M.Ponnavaikko, ICTACC, 2017.

[5] "Detection of Lung Cancer Stages using Image Processing and Data Classification Techniques." Pooja R. Katre, Dr. Anuradha Thakare, I2CT 2017.

[6] "Lung Cancer Detection using Genetic Approach." Mansee Kurkure, Anuradha Tharkre, 2017

[7] "Detection of Lung Cancer from CT Image Using Image Processing and Neural Network," Md. Badrul Alam Miah, Mohammad Abu Yousuf, ICEEICT, 2015.

[8] "Computer Aided Lung Cancer Detection System." Sruthi Ignatious, Robin Joseph, GCCT 2015.

[9] "Lung Cancer Detection On CT images by using image processing" Anita Chaudhary, Sonit Sukhraj Singh, ICCS,

2012. [10]http://homepages.inf.ed.ac.uk/rbf/HIPR2/value.html accessed on 21-03-2017.

[11] Link used for data set: http://www.cancerimagingarchive.net/