Detection of Lung Cancer in CT Images Using Image Processing

Jagriti Shukla (22116037)
Electronics and Communication Department Indian Institute of Technology, Roorkee Roorkee, India
jagriti_s@ece.iitr.ac.in

Mahak Sahay(22116051)
Electronics and Communication Department Indian Institute of Technology, Roorkee Roorkee, India
mahak_s@ece.iitr.ac.in

Pragya Bhanot (22116068)
Electronics and Communication Department Indian Institute of Technology, Roorkee Roorkee, India
pragya_b@ece.iitr.ac.in

Alpari Sethy (22116011)
Electronics and Communication Department Indian Institute of Technology, Roorkee Roorkee, India
alpari_s@ece.iitr.ac.in

Abstract

Cancer is one of the most serious and widespread diseases that is responsible for a large number of deaths every year. Among all different types of cancers, lung cancer is the most prevalent cancer having the highest mortality rate. Computed tomography scans are used for the identification of lung cancer as they provide a detailed picture of the tumor in the body and track its growth. Although CT is preferred over other imaging modalities, visual interpretation of these CT scan images may be an error-prone task and can cause delays in lung cancer detection. Therefore, image processing techniques are used widely in medical fields for early-stage detection of lung tumors.

This paper presents a multi-step approach to identify and locate potential lung tumors. Initially, a predictive model is employed to assess the likelihood of tumor presence. Subsequently, basic image processing techniques, such as smoothing, binarization, morphological operations (including opening and complementing), and hole filling, are applied to pinpoint the precise location of tumors within the lung images.

Keywords—CT, Lung cancer probability, lung mask formation, and processing of the images

INTRODUCTION

Lung cancer is a significant global health concern, with increasing mortality rates. It is the most common type of cancer, affecting both men and women. This malignancy arises from the uncontrolled growth of cells in lung tissues, often leading to the formation of cancerous nodules or tumors.

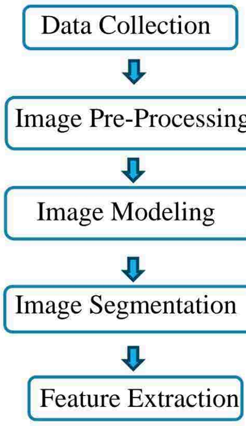
Tobacco use, particularly smoking, is a major risk factor for developing lung cancer. Unfortunately, many lung cancer cases are diagnosed at advanced stages, significantly reducing survival rates. The five-year survival rate for all stages of lung cancer is approximately 14%

Therefore, early lung cancer detection can improve survival rates by 60–70% by giving patients prompt, cancer classifications: small-cell lung cancer and non-small-cell lung cancer. While small cell lung cancer essential therapy, which lowers the death rate. Based on cell features, there are two primary categories of lung accounts for 15-20% of all cancer cases, non-small cell lung cancer is the most common, accounting for 80-85% of all instances. The spread of lung cancer determines the stage of lung cancer. In order of severity, lung cancer is primarily divided into four stages: Stage I: The lung is the only area affected by cancer; Stages II and III: The cancer is limited to the chest; and Stage IV: The lung cancer has spreadto other areas of the body. Numerous imaging modalities, including Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Chest X-rays, can be used to diagnose lung cancer. Because CT scan images are more dependable, clearer, and less distorted than images from other modalities, they are generally chosen. Visual database interpretation is a laborious, time-consuming process that greatly depends on the individual. This increases the likelihood of human mistakes and may result in incorrect cancer classification.

METHODOLOGY

Before resizing the lower pixel image to a larger one, we first apply a median filter to smooth the image. We resize the images because all images have different pixel values which leads to an impractical comparison. Next, using techniques like binary image generation, opening, complementing, and hole filling, we will create a lung mask from a non-cancer patient. Next, we apply the same operations to the lung image of a cancer patient. After that, we do morphological operations to the given images to obtain an image that is suitable for XOR operations. To acquire the final image, we did binary image conversion, opening, complementation, and clear boundary. We shall know the lung's border and the position of the tumor inside it thanks to the XOR procedure.

To determine the likelihood of a lung tumor or not, we first perform wiener filtering, which yields the magnitude spectrum. We can determine whether a tumor is there or not by examining it.



1. **Data Collection**

The first step is to obtain lung CT images of cancer patients and normal patients. We have taken images from the Lung Cancer Archive, where we have taken 10 images of patients in PNG formats. Healthy patients' CT scans are also taken from the web, for lung mask formation.

2. **Image Pre-Processing**

The objective of the image preprocessing stage is to suppress unwanted distortions present in the image and to enhance some features useful for further processing. It includes two main steps such as image smoothing and image enhancement. Image smoothing is done to remove any unwanted noise present in the image. CT scan images are prone to salt and pepper noise, hence median filtering is found to be quite an effective technique in eliminating this impulse noise while preserving the edges. Median filtering gives the best results for image smoothing as it removes noise without blurring the image.

Image enhancement technique improves the quality of digital images to produce better output for further processing. Contrast adjustment is done to enhance the image since image quality is affected by artifacts caused due contrast variations in the image. Contrast adjustment enhances the contrast of an image by transforming input pixel values to new values such that by default 1% of data gets saturated at low and high-intensity input image data. We have not implied it as images are already provided with contrast adjustment.

We have also resized the pixels of the various images we took. Additionally, Soble edge detection is applied to enhance edges, which are crucial features for subsequent segmentation and feature extraction.

3. **Image Modeling**

Image modeling refers to the process of building mathematical or statistical models to represent and understand images.

It involves extracting meaningful information from images, such as patterns, textures, and objects, and representing them in a way that computers can process.

To further enhance the feature extraction process from pre-processed CT scan images, K-Means clustering can be applied. By segmenting images into distinct regions based on pixel intensity and spatial proximity, K-Means can identify potential tumor areas. These segmented regions can then be used to extract relevant features, such as texture, shape, and size. This approach can improve the accuracy of subsequent machine learning models by providing more specific and informative features.

4. **Image Segmentation**

The process of separating the required region of interest from the image is known as segmentation. Mathematical morphological operations are powerful tools for acquiring lung regions from binary images. In our methodology, first, the preprocessed grayscale images were converted to binary images. Morphological opening

Operation (erosion operation followed by dilation) was performed on the binary image with disk structuring image. The opened image was then complemented and a clear border operation was performed to it. The lung masks were obtained by filling the holes and gaps present in the lungs. Finally, an exclusive OR operation was performed on lung mask output and clear border output to give us the segmented tumor region.

5. **Feature Extraction**

Feature extraction is the most essential step that transforms input data into required features. This stage extracts out significant features of the segmented region of interest and these features serve as input for the classification of CT scan images. The size and shape of the tumor present in the lungs is estimated by extracting three geometrical features. The features are mainly the area and perimeter of the cancerous lung nodule.

In precise:

i. Area: This is a scalar quantity that gives the total number of pixels acquired by a cancerous lung nodule. The area is evaluated from the binary image by taking a summation of pixel areas in the image to specify that regions are registered with value 1.

ii. Perimeter: This is a scalar quantity that gives the total pixels present at the border of the lung tumor. The perimeter is evaluated from the binary image by summing the pixels and these are also registered with value 1, at the outline of lung nodules.

RESULT

The proposed system was implemented in Python—this study's database.

Lung mask formation

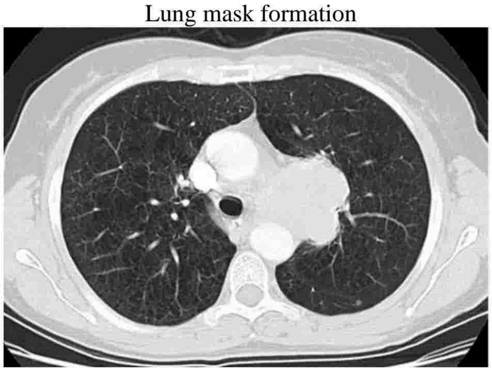


Fig1. Greyscale image of non-cancer patient

Next, a 3*3 median filter was applied to the lung CT images to eliminate salt and pepper noise from the image while preserving the edges. Figure 3 shows the contribution of median filtering towards the improvement of the image.

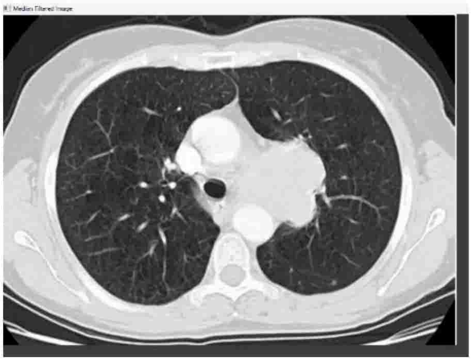


Fig 2. After Median Filtering

After that, we do Sobel Edge Detection.

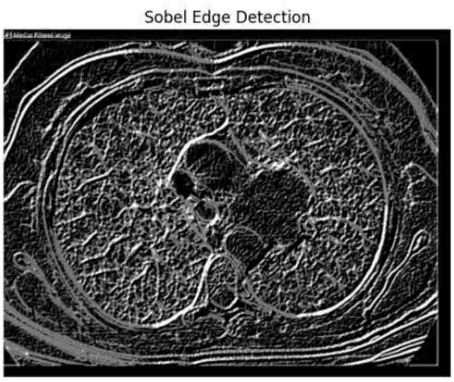


Fig 3. Image after Sobel Edge Detection

We cluster the pixels according to their pixel values.

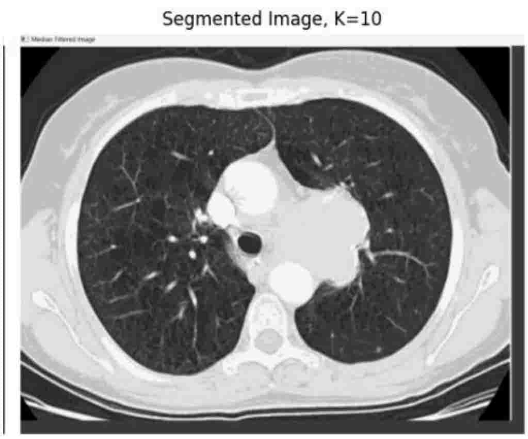


Fig 4. Clustered image with k = 10

Now, we convert the grayscale image to a binary image of the non-cancer patients.

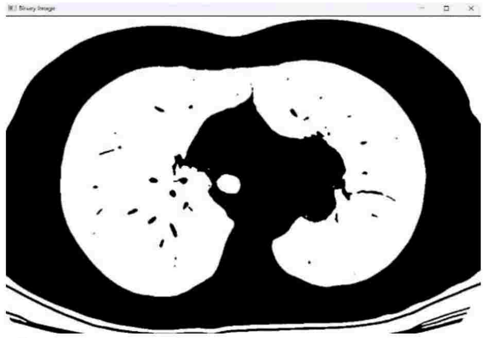


Fig 5. Binary image of non-cancer patient

After this, we complement the image.

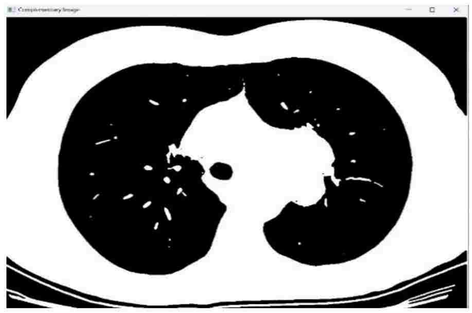


Fig6. Complement of an image of the non-cancer patient.

Filling holes in this image to get the final lung mask ready for XOR operation.

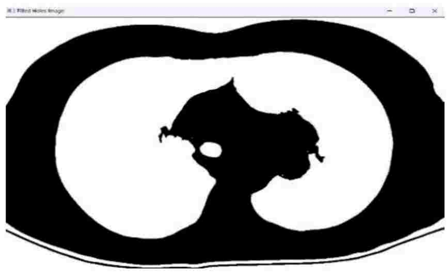


Fig 7. Hole filling in the image of the non-cancer patient

Processing of the image of having a lung cancer tumor.

A grayscale image of the cancerous patient after a median filtering.

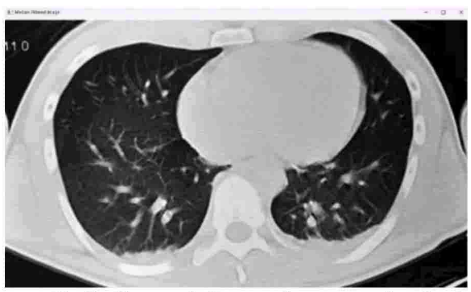


Fig 9 . Grayscale Image of cancer patient having tumor after median filtering

After that, we do Sobel Edge Detection.

Fig 10. Image after Sobel Edge Detection

We cluster the pixels according to their pixel values.

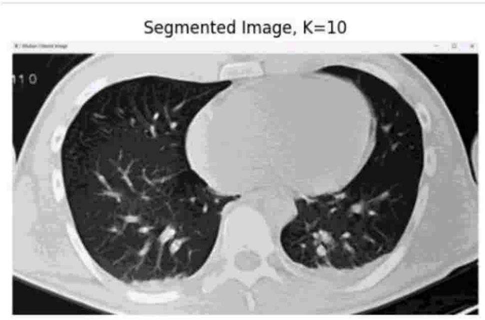


Fig 11. Clustered image with k = 10

Image after binary conversion

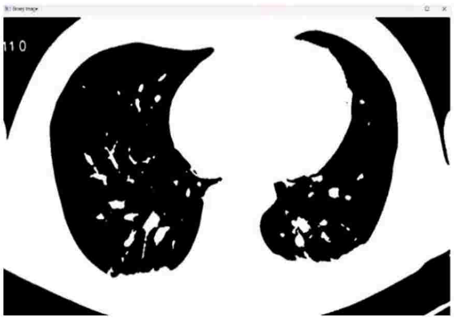


Fig 12. Grayscale image converted to Binary image

After performing a morphological opening operation, we get an image like this

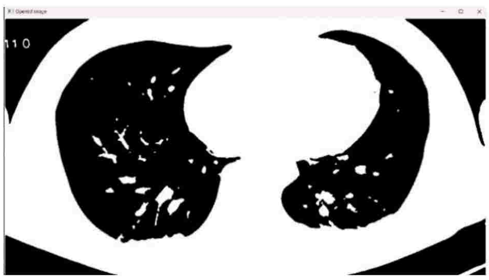


Fig 13. Image after Opening

After this, we perform the Complementary operation on the above image.

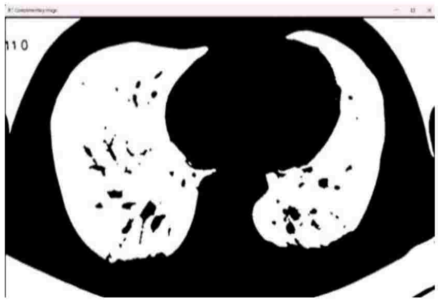


Fig 13. Image After doing Complementary

We performed the clear boundary operation after that and got our final image.

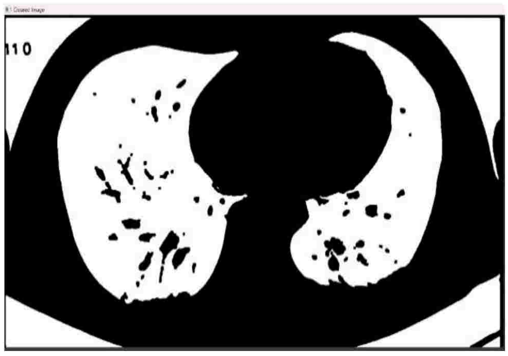


Fig 14. Final Image after clear Boundaries operation

Now, as said we will perform the XOR operation between the lung mask and our final cancer patient image to get the perimeter of the lungs and the denoted area of the tumor region in the lungs.

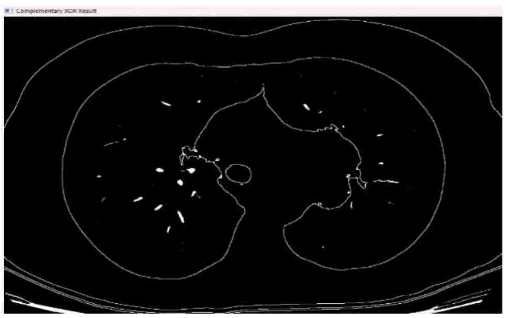


Fig 15. Final Result after XOR operation

Challenges and Limitations

- Due to anatomical differences of all human beings, while comparing the final output images of the non-cancer and cancer patients we got different images which is not desirable.

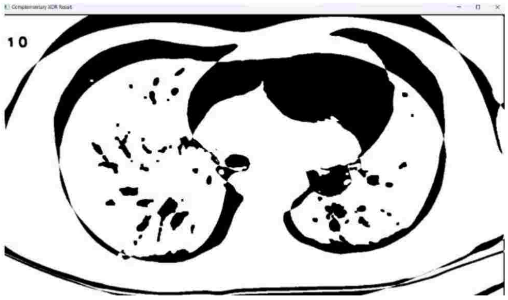


Fig 16. image after XOR operation on two different anatomical patients.

CONCLUSION

In this paper, a system for automatic detection of lung cancer in CT images was successfully developed using an image processing technique. The adopted methodology performs well in enhancing, segmenting, and extracting features from CT images. The median filtering technique effectively eliminated impulse noise from the images without blurring the image. Mathematical morphological operations enable accurate segmentation of lung and tumor regions. Two geometrical features i.e. area, and perimeter were extracted from the segmented tumor region. Our main purpose here was to identify the tumor region by comparing it with the normal lung mask of the non-cancerous patient and to get the approximate region of tumors in the lungs.

REFERENCE

[1] Anita Chaudhary, Sonit Sukhraj Singh, “Lung Cancer Detection on CT Images using Image Processing”, International Conference Computing Sciences, IEEE, 2012.

[2] Nooshin Hadavi, Md Jan Nordin, Ali Shojaeipour, “Lung Cancer Diagnosis using CT-scan Images based on Cellular Learning Automata”, International Conference on Computer and Information Sciences (ICCOINS), IEEE, 2014.

[3] K.Punithavathy, M.M.Ramya, Sumathi Poobal,"Analysis of Statistical Texture Features for Automatic Lung Cancer Detection in PET/CT Images", International Conference on Robotics, Automation, Control and Embedded Systems (RACE), IEEE, 18-20 February 2015.

[4] Anjali Kulkarni, Anagha Panditrao, “Classification of Lung Cancer Stages on CT Scan Images Using Image Processing”, International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), IEEE,2014.

[5] Gawade Prathamesh Pratap, R.P Chauhan, “Detection of Lung Cancer Cells using Image Processing Techniques”, 1st IEEE International Conference on Power Electronics, Intelligent control and energy systems, IEEE, 2016.

[6] Moffy Crispin Vas, Amita Dessai, “Classification of Cancerous and Non-cancerous Lung Cancer Nodule using Image Processing Technique” International Conference on Academic Research in Engineering and Management (ICAREM-17).

[7] Dataset- https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70224216

Contribution

- Pragya Bhanot- *Image pre-processing:* image resizing, median filtering, wiener filtering, FFT

Image segmentation: morphological operations (opening, border closing, XOR operation), lung mask-making, logic discussion and implementation

- Jagriti Shukla- *Image pre-processing*: image resizing, median filtering, wiener filtering

Image segmentation: morphological operations (opening, border closing, XOR operation), lung mask-making, logic discussion and implementation

- Mahak Sahay- *Image pre-processing*: Sobel Edge Detection, Histogram equalization, Thresholding calculations, and implementation for image conversion

Image modeling - k- means clustering, Logic discussion

- Alpari Sethy- *Image pre-processing*: Image thresholding, FFT,

  logic discussion and implementation, finding the sources, data collection, report-making