

Detection of Lung Cancer in CT Images using Image Processing

Nidhi S. Nadkarni

dept. of Electronics and Telecommunication Engineering
Goa College of Engineering
Ponda-Goa, India
nidhinadkarni95@gmail.com

Prof. Sangam Borkar

dept. of Electronics and Telecommunication Engineering
Goa College of Engineering
Ponda-Goa, India
sangam@gec.ac.in

Abstract—Cancer is one of the most serious and widespread disease that is responsible for 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 identification of lung cancer as it provides detailed picture of tumor in the body and tracks 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 delay in lung cancer detection. Therefore, image processing techniques are used widely in medical fields for early stage detection of lung tumor. This paper presents an automated approach for detection of lung cancer in CT scan images. The algorithm for lung cancer detection is proposed using methods such as median filtering for image pre-processing followed by segmentation of lung region of interest using mathematical morphological operations. Geometrical features are computed from the extracted region of interest and used to classify CT scan images into normal and abnormal by using support vector machine.

Keywords—CT, Lung cancer, Region of interest, SVM.

I. INTRODUCTION

Cancer is a noteworthy general health issue worldwide with mortality rates increasing day by day. Lung cancer, among all other cancer types is the most common and deadly that occur both in men and women. Lung cancer, additionally known carcinoma is formation of malignant lung tumors (cancerous nodules) due to uncontrolled growth of cells in lung tissues. Eating tobacco and smoking are the leading risk factors for causing cancerous lung nodules. The survival rate of lung cancer patients combining all stages is very less roughly 14% with a time span of about 5-6 years. The main problem with lung cancer is that most of these cancer cases are diagnosed in later stages of cancer making treatments more problematic and significantly reducing the survival chances. Hence detection of lung cancer in its earlier stages can increase the survival chances up to 60-70% by providing the patients necessary fast treatment and thus it curbs the mortality rate. Small cell lung cancer and non-small cell lung cancer are two main types of lung cancer classifications based on cell characteristics. The most commonly occurring is non-small cell lung cancer that makes up about 80-85% of all cases, whereas 15-20% of cancer cases are represented by small cell lung cancer. Lung cancer staging depends on spread of cancer in the lungs and

tumor size. Lung cancer is mainly classified into 4 stages in order of seriousness: Stage I-Cancer is confined to the lung, Stage II and III-Cancer is confined within the chest and Stage IV-Lung cancer has spread from the chest to other parts of the body. Lung cancer diagnosis can be done by using various imaging modalities such Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and Chest X-rays. CT scan images are mostly preferred over other modalities because they are more reliable, have better clarity and less distortion. Visual interpretation of database is a tedious procedure that is time consuming and highly dependent on given individual. This introduces high possibility of human errors and can lead to misclassification of cancer. Hence an automated system is of utmost importance to guide the radiologist in proper diagnosis of lung cancer. The methodology developed for this system includes dataset collection, pre-processing, lung segmentation, feature extraction and classification.

II. METHODOLOGY

The proposed system for lung cancer detection in CT images is shown with the help of a flowchart in figure 1. The methodology is carried out in five main steps and each step of this system is discussed in detail in section below.

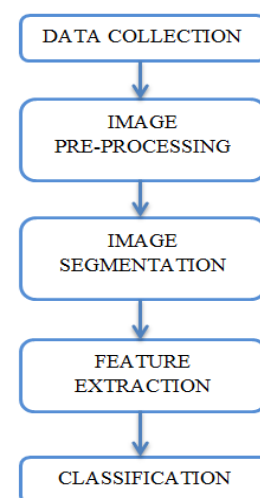


Fig. 1. Block diagram of the proposed system

A. Data Collection

The first step is to obtain lung CT images of cancer patients. For research work, the images have been downloaded from the Cancer Imaging Archive database. The images are stored in DICOM format. The image database contains Computed Tomography images of patients with and without lung cancer.

B. Image Pre-Processing

The objective of 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 unwanted noise present in the image. CT scan images are prone to salt and pepper noise, hence median filtering is found to be quite 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% data gets saturated at low and high intensity of input image data.

C. Image Segmentation

The process of separating out required region of interest from the image is known as segmentation. Mathematical morphological operations are powerful tools in acquiring lung region from binary images. In our methodology, first the preprocessed gray scale images were converted to binary images. Morphological opening operation was performed to the binary image with disk structuring element for removal of unwanted components from the image. The opened image was then complemented and clear border operation was performed to it. The lung masks were obtained by filling the holes and gaps present in the lungs. Finally exclusive OR operation was performed to lung mask output and clear border output to give us the segmented tumor region.

D. Feature Extraction

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

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

2. 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 registered with value 1, at the outline of lung nodule.
3. Eccentricity: This metric value is also referred to irregularity index (I) or circularity or roundness. For a circular shape eccentricity value is equal 1 and the value is less than 1 for any other shape.

$$\text{Eccentricity} = \text{length of major axis} / \text{length of minor axis}$$

E. Classification

The Classification stage involves labeling the CT scan images as normal and abnormal. In our method SVM algorithm will be used for detection of lung cancer in CT images. SVM classifiers are supervised learning models that analyze input data and classify them according to pattern. The SVM classifier builds a model by using training dataset and categorizes it into two classes. The SVM algorithm then assigns new examples of testing dataset to one of the two classes. SVM classifier thus finds the best hyper plane that separates the two groups and thus classifies the lung CT images. For the best hyper plane data points of one class are separated from the other by largest margin between the two classes.

III. RESULTS

The implementation of proposed system was done using MATLAB software. Database for this study was obtained from the Cancer imaging Archive (TCIA). Figure 2 shows the CT scan image of patient affected by lung cancer and its corresponding histogram. Histogram of an image is a graphical representation of an image which gives pixel distribution amongst its various gray levels.

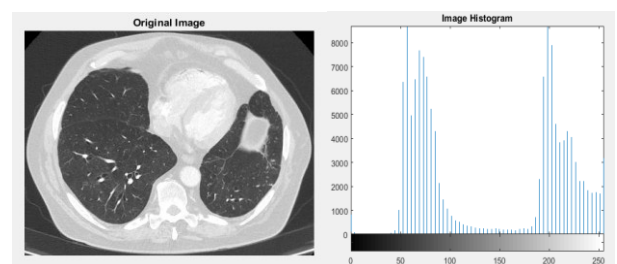


fig. 2. Cancerous lung CT image and histogram of image

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 improvement of the image.



fig. 3. Median filtering method for image smoothing

Further in the preprocessing stage image enhancement was done using contrast adjustment. In contrast adjustment image intensity values are mapped to full display range of the input data and the contrast of the image is enhanced. Figure 4 depicts the contrast enhanced CT scan image and its corresponding histogram plot.

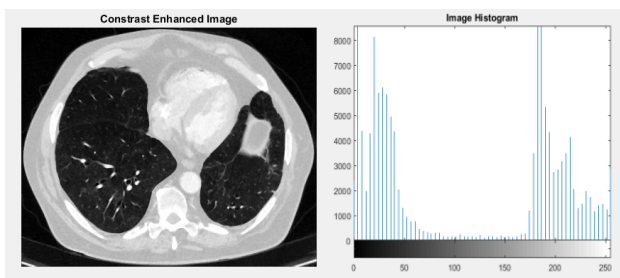


fig. 4. Contrast enhanced image and histogram of image

Segmentation technique divides input image into various parts and thus gives us the region of interest for further processing. To exact the region of interest, morphological operation based on structuring element were used. The lung masks and tumor region were obtained from CT image as shown in figure 5.

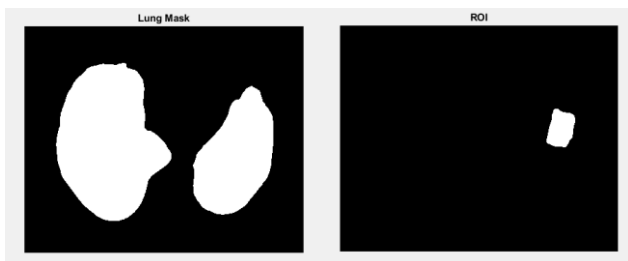


fig. 5. Lung mask and extracted tumor region

Next three geometrical features namely area, perimeter and eccentricity were extracted from segmented tumor region. These features serve as inputs to the SVM classifier for categorizing the images into normal (non-cancerous) or abnormal (cancerous). For the given sample CT scan image, the extracted lung tumor as the following three features:

- Area: 1793
- Eccentricity: 0.7319
- Perimeter: 161.6980

In the final stage, SVM classifier is used to determine whether the CT scan images are normal or abnormal.

IV. CONCLUSION

In this paper, a system for automatic detection of lung cancer in CT images was successfully developed using image processing technique. The adopted methodology performs well in enhancing, segmenting and extracting features from CT images. Median filtering technique was effective in eliminating impulse noise from the images without blurring the image. Mathematical morphological operations enable accurate segmentation of lung and tumor region. Three geometrical

features i.e. area, perimeter and eccentricity were extracted from segmented tumor region and fed to the input of the classifier for classification of lung CT images into normal and abnormal. Hence this proposed methodology helps in accurate and early stage detection of lung cancer.

ACKNOWLEDGMENT

First of all I would like to thank my project guide, Prof. Sangam Borkar, for his guidance and unconditional support. I would also like to express my special thanks to Dr. M. S. Krupashankar, Principal, Goa College of Engineering and Dr. H.G.Virani, Professor and HOD, Electronics and Telecommunication Engineering for allowing me to use college facilities that helped in this research project. I take this opportunity to thank all the staff of ETC department for their critical advice and timely support. Last, but not the least I want to thank God, my family and all my friends without whom this project work would not have been possible.

REFERENCES

- [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] Shubhangi Khoragade, Aditya Tiwari, "Automatic Detection of Major Lung diseases using Chest Radiographs and classification by Artificial Neural Network", 1st IEEE International Conference on Power Electronics, Intelligent control and energy systems, IEEE, 2016.
- [8] Pooja R. Katre, Dr. Anuradha Thakare, "Detection of Lung Cancer Stages using Image Processing and Data Classification Techniques", 2nd International Conference for Convergence in Technology (I2CT) 2017 .
- [9] S. Kalavani, Pramit Chatterjee, Shikhar Juyal, Rishi Gupta, "Lung Cancer Detection Using Digital Image Processing and Artificial Neural Network", International Conference on Electronics, Communication and Aerospace Technology ICECA 2017.
- [10] K.Gopi, Dr.J.Selvakumar, "Lung Tumor Area Recognition and Classification using EK-Mean Clustering and SVM", International Conference on Nextgen Electronic Technologies, IEEE 2017.
- [11] G.Niranjana, Dr.M.Ponnavaikko, "A Review on Image Processing Methods in Detecting Lung Cancer using CT Images", International Conference on Technical Advancements in Computers and Communications, IEEE 2017.

- [12] Sayali Satish Kanitkar, N.D.Thombare, "Detection of Lung Cancer Using Marker-Controlled Watershed Transform", International Conference on Pervasive Computing(ICPC) 2015.
- [13] Bhagyashri G. Patil, Prof. Sanjeev N. Jain, "Cancer Cells Detection Using Digital Image Processing Methods", International Journal of Latest Trends in Engineering and Technology (IJLTET), 2014.