

## Detection of lung cancer using image processing techniques

Prathamesh Gawade\* and R.P. Chauhan

Department of Physics (Instrumentation), NIT, Kurukshetra, Haryana

©2016 ACCENTS

### Abstract

*The diagnosis of lung cancer at an early stage is of utmost importance if it is meant to degrade high mortality rate. The global lung screening program points to visualise positron emission tomography (PET) and computed tomography (CT) examinations amongst most aged groups at risk to enhance the early detection rate. Although use of invasive techniques, symptoms hardly appear until disease is advanced making it difficult for radiologist to identify lesions. Unfortunately, most lung cancer patients suffering at advanced stages result in dismal with five-year survival rate of 17.8% and for distant tumours, being only 4%. Genuine and precise information is the basis of disease control initiatives. More than 85% of the disease is related to tobacco consumption. In addition, genetic factors, exposure to environmental pollutants, second hand smoking inflate disease rapidly. Remedies including chemotherapy, radiotherapy, surgery, epidermal receptive drugs escalate survival rate and quality of life. This method is more about diagnosing at early and crucial stages with intelligent computational techniques with various distortion removals by segmentation techniques and algorithms which is the root concept of image processing. Detection of CT images obtained from cancer institutes is analysed using MATLAB.*

### Keywords

*Lung cancer, MATLAB, CT images, Distortion removal, Segmentation, Mortality rate.*

### 1.Introduction

Cancer can be explained as uncontrolled cell growth having ability to spread all over the body. Our body contains red blood cells (RBC's). The main function is to supply fresh oxygen (O<sub>2</sub>) to all parts of the body with the help of blood flow, due to which blood appears red [1]. In the lungs, tissue receives oxygen (O<sub>2</sub>) because of RBC's only.

The genetic content of erythrocytes has high concentration of haemoglobin. The cell membrane consists of proteins and lipids which is backbone of psychological cell function. They do not contain any important part of cell, which includes haemoglobin. Around 20 lakhs new RBC's are produced per second [2]. The cells are produced in the bone marrow and rotate throughout the body for about 4 months to & fro in arteries and veins. Each rotation takes about 20 seconds. Approximately 75% of the cells & majority concentration of blood in the human body are red blood cells [3, 4]. Young Dutch biologist Jan Swammerdam described it perfectly with an early microscope in 1658. Lung cancer, also called as carcinoma [5] is a malignant lung tumour defined by uncontrolled cell growth in tissues of the lung [6].

If left undiagnosed, this growth penetrates beyond the lung by the process of 'metastasis'-which produces agency (as cancer cells) from initial site of disease to another part of body [7].

Most cancers also known as primary lung cancers are carcinomas which start it in the lungs [8]. There are two main types are carcinoma 1) small-cell lung cancer and 2) non-small-cell lung cancer [9]. The most common symptoms are coughing (including blood coughing), weight loss, breath shortness, and minute as well as severe chest pains [10]. These cancer cells disturbs the cycle of RBC production and degradation. Internally they change the structure and composition of plasma membrane i.e. the outer structure of the cell in such a way that these RBC's do not die as per the corresponding lifetime. As a result the RBC count goes on increasing and more and more number of cell gets accumulated resulting in shortening of veins and arteries and ultimate bursting. These outcomes into blood through cough etc.

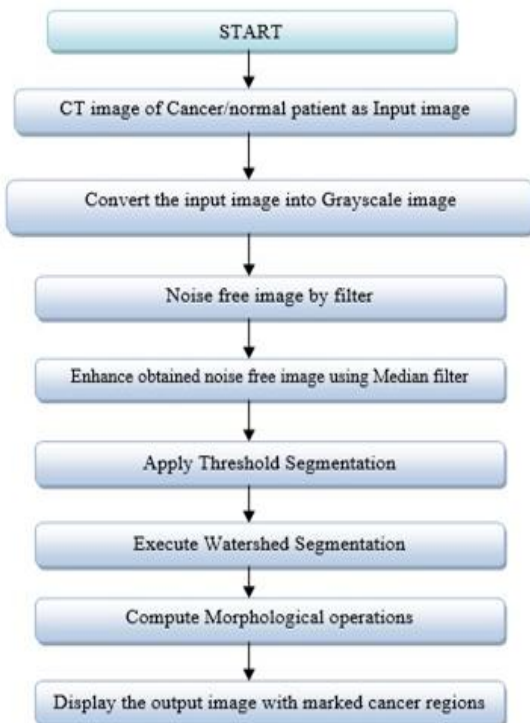
Image processing is used to analyse images at the lowest level provided any quality. These operations do not increase probability of image information content, but they decrease it if entropy is an

\*Author for correspondence

information measure. The main requirement of processing is to improve pixel intensity by converting from discrete to digital image, segmenting to pixels, carrying out mathematical operations on pixels, and reconstructing of image with better quality [11]. Pre-processing of CT images is the initial step in image analysis followed by segmentation process and ended with some morphological operations are applied to detect the cancer spots/cells in the image. Also it can be used to determine the amount of spreading of cancer i.e. what percentage of lung is affected with cancer. The morphological operations are basically applied by comparing the size and shape of the cancer cell with normal cell, and then the infected cells images displayed onto grey scale image with maximum intensity (255). The sequencing of the paper is as follows. Section II describes the prescribed methodology, in detail with mathematical operations performed on the image using various filtering methods followed by segmentation. Section III shows the results of the developed methodology using a MATLAB image. Finally, section IV defines the conclusion and application of the method in diagnostics of various diseases.

## 2.Methodology

The algorithm is proposed as is given in *Figure 1* as follows:



**Figure 1** Algorithm for cancer detection

The design is explained below as follows:

### 2.1Input image

For any type of cancer, firstly image of internal parts of the body should be obtained. CT scan also known as X-ray computed tomography makes use of X-ray for capturing the images from various angles and merge these images to generate cross sectional tomographic image of particular areas of scanned tissues i.e. it allows the person to see the status inside body without non-invasive techniques [12]. The lungs are the prime important organs of respiration in humans as well as other animals. In mammals and majority of the vertebrates, two lungs are located on either side of the heart near to backbone. Their role is to take oxygen from the atmosphere and transfer it into the bloodstream, and to give out carbon dioxide from the bloodstream into the atmosphere. Humans have 2 lungs, right & left. They are situated inside the thoracic cavity of chest. The right lung being bigger than the left, shares space with the heart. The lungs together weigh approximately 1.4 kg. Plural sac in which lungs are enclosed allows inner and outer walls to slip over each other without more friction. This sac encloses each lung and also divides each lung into sections called lobes.

The right lung has 3 lobes and the left has two. The lobes are further classified into bronchopulmonary segments and lobules. The lungs possess a unique blood supply, receiving deoxygenated blood from the heart for receiving oxygen (i.e. pulmonary circulation) and a distinguished supply of oxygenated blood (the bronchial circulation). The tissue of the lungs can be struck by a number of diseases, including pneumonia and lung cancer. Chronic diseases such as chronic obstructing pulmonary disease and emphysema (damaging alveoli's in the lungs) can be linked to smoking or exposure to harmful substances. Diseases such as bronchitis can also involve the respiratory tract. The image of affected lungs and normal lungs is quite different and easily differentiable. These CT images are converted to grayscale images.

### 2.2Grayscale image

In the computing world, a grayscale image is a digital image, in which the value of each pixel is an individual sample, i.e., it carries only intensity or amplitude. Images of this kind, also known as white (highest intensity) & black (lowest intensity) images, consist exclusive shades of gray [13]. Grayscale images are the result of measuring the intensity of light at each pixel in a single band of the light

spectrum. They can also be obtained from a full colour image. The reason behind choosing grayscale image is even minimum pixel intensity is also helpful in detecting changes in the cells. In fact a gray colour is one in which the R, G, B planes have equal amplitude, the brightness levels represented as a number from decimal 0 to 255. For every pixel in an RGB grayscale image,  $G = B = R$ . The intensity varies in proportion with the number representing the brightness levels of the RGB colours. Black is represented by  $R = G = B = 0$  and white is represented by  $R = G = B = 255$ .

### 2.3 High pass filter

As the name suggests, it passes the frequency above certain cut-off frequency and attenuates all the frequency below the cut-off frequency. A high pass filter is mainly used for sharpening images. It is done when contrast is enhanced between the adjacent areas with increase or decrease in brightness level. A high pass filter sets high threshold cut-off to obtain information of an image while cutting the low frequency data. The basis of the high pass filter is designed to increase the amplitude of the median pixel relative to adjacent pixels. The kernel array generally contains a single value at its center, which is completely surrounded by other values. The values may be defined in terms of positive or negative.

### 2.4 Median filtering

It is a nonlinear digital filter used to remove some noise in the image. To detect some edge in the image, firstly noise should be removed up to some threshold value and then edge removal is performed. Hence the median filter is placed before edge detector. Its main feature is it removes noise without edge removal. Median filter is same as that of averaging filter, in which each output image pixel is set corresponding to the average value of neighbouring pixel of the input image. The median filter is more sensitive to mean values and less sensitive to extreme values of pixel which helps in noise reduction.

### 2.5 Threshold segmentation

Threshold segmentation is one of the easiest segmentation methods. The pixels are divided depending upon their amplitude levels. There are various types of segmentation depending on parameters like threshold values of pixels, edge based, region based, clustering etc.

It relates grayscale image to binary image, also called as mapping. After this operation, image is divided into 2 pixel values only, 0 & 1. If there is an image which contains dark structure on bright background, then thresholding can be used to separate the structure. Also to set a particular threshold value, many sub algorithms can be used e.g. histogram estimation, optimal thresholding, iterative thresholding, K means clustering. In K means clustering, grayscale image is divided into K segments i.e. K-1 threshold values, thereby reducing variance. Many images which are made of pixels contains more than one value e.g. RGB. If we separate these pixel values for R, B & G, they are called channels.

### 2.6 Watershed algorithm

It can be explained using a practical idea. Consider a surface immersed into the lake with a hole at minima, so that water will start filling through that hole and will go on increasing. If 2 such surfaces very close to each other are placed then a point will come where the water will overlap and mix from both the surface. At that point only, dams are built so that water does not mix. These dams are watershed lines and by the process of filling water, surface separation can be done. There are many methods to carry out this algorithm. One of the most common watershed algorithms was introduced by F.Meyer called 'Meyer watershed algorithm'. This algorithm applies only to grayscale image [14].

### 2.7 Morphological operations

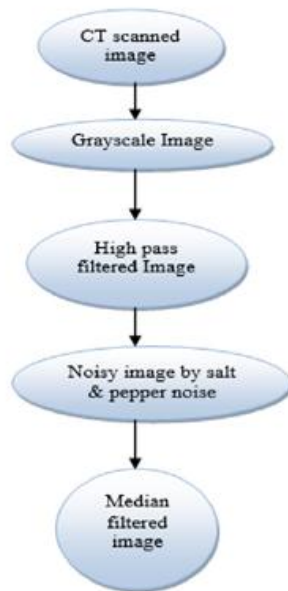
According to Wikipedia, mathematical morphology is a technique to evaluate segmented structures/images based on random functions, set theories etc. It is strictly applied to digital images only. E.g. binary morphological operations explores a particular grayscale image with simple, predefined shapes, and concluding how this shape fits into the image provided or what part of the image gets missed due to this shape.

## 3. Experimental result and analysis

For implementation of this above method, real time images of cancer patients have been obtained from institutes. Since cancer in CT image have more intensity level than background, it becomes very easy to extract image from CT image for further analysis. The working methodology is as follows:

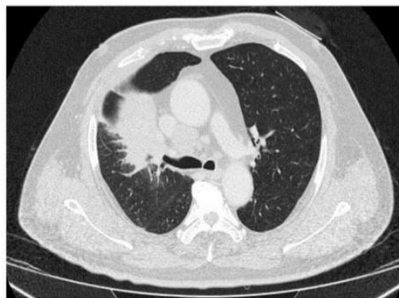
### 3.1 Processing grayscale image with median filter

The proposed work flow is given in *Figure 2*.



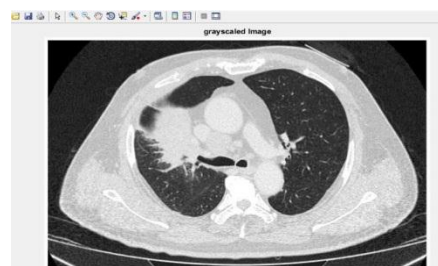
**Figure 2** Algorithm for preprocessing of CT scan image

The CT scanned Image with cancer infected lung obtained from the institute is given in *Figure 3*.



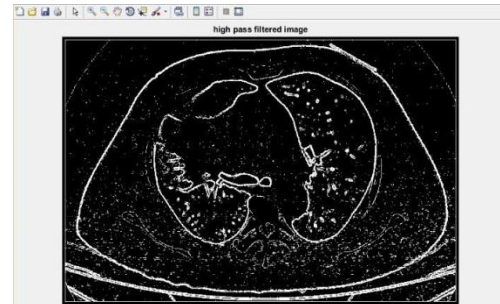
**Figure 3** CT Image of cancer infected lungs

The CT image is converted to grayscale image to perform mathematical operations shown in *Figure 4*.



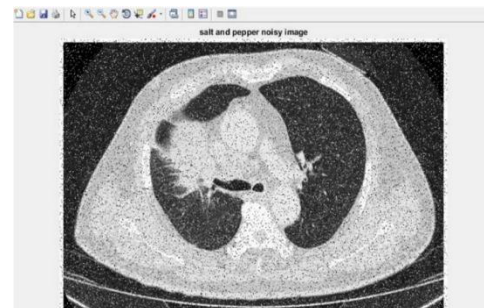
**Figure 4** Grayscale image

Now it is ready to carry out further mathematical operations, the image is passed through a high pass filter to enhance the information needed shown in *Figure 5*.



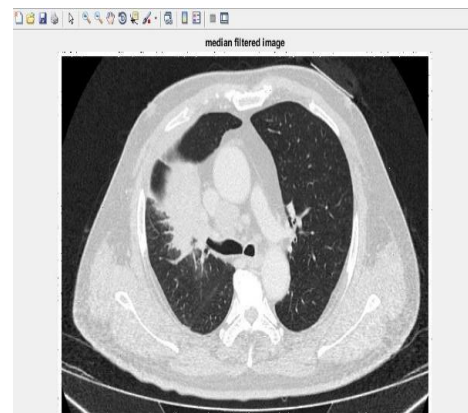
**Figure 5** High pass filtered image

Further enhancement of the image for detailed analysis is done using median filtering. For this we add salt and pepper noise to reduce noise from the image, so that the image is distortion less shown in *Figure 6*.



**Figure 6** Salt and pepper noisy image

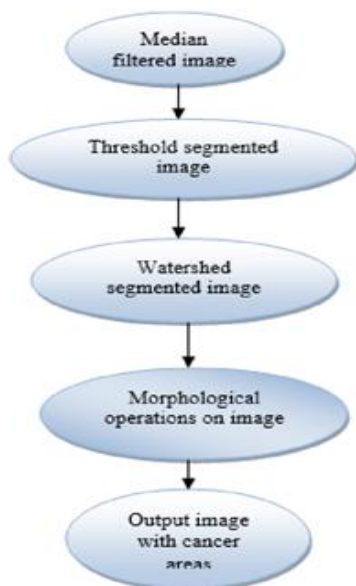
Final pre-processing is done by passing salt and pepper noise through a median filter which will allow removing noise completely from the image while restoring edges shown in *Figure 7*.



**Figure 7** Median filtered image

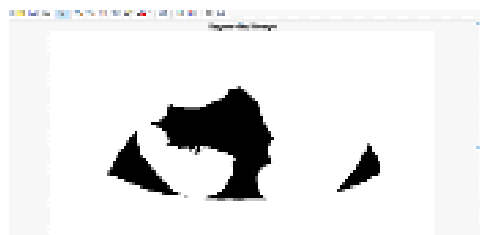


The median filtered passed through various segmentations shown in *Figure 8*.



**Figure 8** Segmentation process

The segmented image is shown as follows in *Figure 9*.



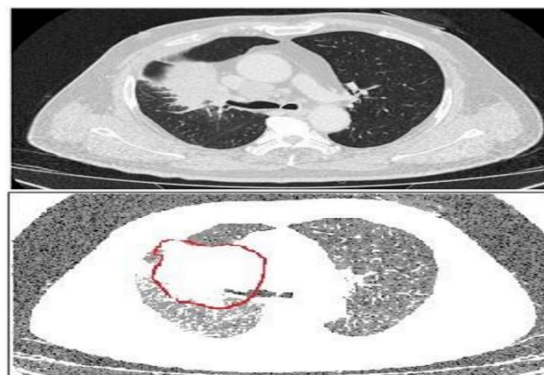
**Figure 9** Threshold segmented image

The heart of this process is watershed algorithm. It will identify the areas affected by cancer in the infected lung with the help of which we could differentiate cancerous lung tissue with normal lung tissue as shown in *Figure 10*.



**Figure 10** Watershed image

The output of the watershed image contains pixels of infected lung tissues which are not labelled. These pixels are in the form of watershed lines. *Figure 11* shows the output image after successful Morphological operations. This figure clearly represents left lung being more infected by cancer with background stating the cancer region as compared to right lung of the given CT scan image.



**Figure 11** Output image with input image after morphological operations

#### 4. Conclusion and future work

The above method is processed in two steps 1) Processing of noisy input image using filter and segmentation 2) Morphological operations on CT image. The cancer affected lungs region can be observed in the final output image to CT input image provided. The proposed method can also be applied to some other cancer types like breast cancer, skin cancer etc. Also it finds its application in the medical research as well.

#### Acknowledgment

A very grateful thank to Dr. Mayuresh Kiran M.B.B.S., M.D. (Pharmacology). Dy. General Manager-Medical Services, Centaur Pharmaceuticals, Mumbai who made significant contribution in improvement of the research paper by providing CT scan images for lung cancer affected patients.

#### Conflicts of interest

The authors have no conflicts of interest to declare.

#### References

- [1] Lipowsky R, Sackmann E. Preface to volume 1a from cells to vesicles: introduction and overview. Handbook of Biological Physics. 1995.
- [2] Lemjabbar-Alaoui H, Hassan OU, Yang YW, Buchanan P. Lung cancer: Biology and treatment options. BBA Reviews on Cancer. 2015; 1856(2):189-210.

- [3] Pierige F, Serafini S, Rossi L, Magnani M. Cell-based drug delivery. *Advanced Drug Delivery Reviews*. 2008; 60(2):286-95.
- [4] Villa CH, Anselmo AC, Mitragotri S, Muzykantov V. Red blood cells: supercarriers for drugs, biologicals, and nanoparticles and inspiration for advanced delivery systems. *Advanced Drug Delivery Reviews*. 2016.
- [5] Dubey AK, Gupta U, Jain S. Epidemiology of lung cancer and approaches for its prediction: a systematic review and analysis. *Chinese Journal of Cancer*. 2016; 35(1):71.
- [6] Scagliotti GV, De Marinis F, Rinaldi M, Crino L, Gridelli C, Ricci S, Matano E, Boni C, Marangolo M, Failla G, Altavilla G. Phase III randomized trial comparing three platinum-based doublets in advanced non-small-cell lung cancer. *Journal of Clinical Oncology*. 2002; 20(21):4285-91.
- [7] Laccetti AL, Pruitt SL, Xuan L, Halm EA, Gerber DE. Prior cancer does not adversely affect survival in locally advanced lung cancer: A national SEER-medicare analysis. *Lung Cancer*. 2016.
- [8] World Health Organization. The world health report 2000: health systems: improving performance. World Health Organization; 2000.
- [9] O'REILLY KM, McLaughlin AM, Beckett WS, Sime PJ. Asbestos-related lung disease. *Chest*. 2007.
- [10] Collins LG, Haines C, Perkel R, Enck RE. Lung cancer: diagnosis and management. *Am Fam Physician*. 2007; 75(1):56-63.
- [11] Hricak H, Thoeni RF. Anatomy of the prostate gland. In Pollack HM. *Clinical urography*. 2nd ed. Philadelphia, Pa: Saunders; 2001.

- [12] Larkins DB, Harvey W. Introductory computational science using MATLAB and image processing. *Procedia Computer Science*. 2010; 1(1):913-9.
- [13] Johnson S. Stephen Johnson on digital photography. O'Reilly Media, Inc.; 2006.
- [14] Gonzalez RC, Woods RE. *Digital image processing*. Prentice Hall; 2008.



**Prathamesh. P. Gawade.** I received my B.Tech degree from Pune University in Electronics Engineering in 2012, currently pursuing M.Tech degree from NIT kurukshetra. I am having 1 year of industrial experience on graduate level. I presented a technical paper in the area of Image Processing in various conferences held at IIT delhi, IIT BHU, IIT Guwahati & NIT kurukshetra. My area of interest includes Image Processing, Signal analysis & communication engineering.  
Email: pg1911991@gmail.com@gmail.com



**Dr. Rishi Pal Chauhan.** He is an Associate professor at NIT kurukshetra, PhD in physics and has teaching experience of 30 years. I am currently working in field of radiation physics, material science, and instrumentation and have more than 150 research papers published in SCI journals. Currently guiding 12 students for PhD degree, 40 students for M.tech degree and 20 students for M.Phil degree.