Detection and Segmentation of Lung Cancer using Statistical Features of X-Ray Images

Marwah Abdulmajeed Azeez¹, Sabah Noori Mazhir², Alyaa Hussein Ali³

¹Department of Physics, Baghdad University, Iraq
²Department of Physics, Baghdad University, Iraq
³Department of Physics, Baghdad University, Iraq

Abstract—The early detection of lung cancer is a difficult task because it is one of the important problems due to the installation of complex cancer cells. It is difficult to diagnose quickly through the x-rays only, therefore enlisted this techniques such as digital technologies, becoming a process of early detection of cancer easier complicated. X-ray images contain the noise so as to contain the bones of the rib cage, as well as the bones of the sternum, which is a major obstacle to distinguish cancer, therefore, to improve the images using a filter (median) as well as the way of image segmentation, the thresholding and edge detection methods. Dividing the lung into two parts (upper and lower) comparing the parts with each other to identify which part carry the tumor. The statistical features has been calculated using first Order histogram. In this research three images of one patient has been used after several stages of treatment.

Keywords—Cancer, X-ray image, segmentation, statistical feature

1. INTRODUCTION

Lungs are located on the lateral sides of the thoracic cavity and separated from each other by the mediastinum. As the left lung is physically smaller than the right lung This is because the right and left lungs exhibit some obvious structural differences since the heart projects into the left side of the thoracic cavit, the right lung is subdivided into the superior, middle, and inferior lobes by two fissures. The left lung has a medial surface indentation, called the cardiac impression that is formed by the heart [1].

Lung disease refers to disorders that affect the lungs, the organs that allow us to breathe. Breathing problems which are caused by lung disease may prevent the body from getting enough oxygen [2].

Lung cancer is a highly aggressive and frequently fatal malignancy that originates in the epithelium of the respiratory system. Smoking causes about 85% of all lung cancers. Metastasis, the spread of cancerous cells to other tissues, occurs early in the course of the disease, making a surgical cure unlikely for most patients [1].

X-ray imaging is the fastest, most common, and least expensive diagnostic. Production of digital X-rays from pictorial radiographs is becoming a common practice to maximize information and reduce the number of
rejected X-rays [3]. X-rays are among the oldest sources of electromagnetic (EM) radiation used for imaging. The best known use of X-rays is medical diagnostics [4].

2. METHODOLOGY

Analysis of tissue is a process which is very successful in the diagnosis of medical images of abnormal lung tissue and for this reason we need the help of x-rays and digital image processing, which is the way of modern diagnosis defect. The steps to detect the cancer are:-

1. Cancelling the rips which appear white on the lungs this can make mastics in detecting the cancer by subtracting the background of the lung image from the cancer.
2. Dividing the lung which carry the cancer into two parts (normal and abnormal). Comparing these with each other by the histogram plot and statistical features.
3. When we specify which part hold the cancer the image is processed by the following steps:-

A. Image segmentation

Image segmentation is the most difficult task in image processing. Segmentation refers to the grouping of image elements that exhibit “similar” characteristics, i.e. subdividing an image into its constituent regions or objects [5]. The thresholding process is the simplest way to segmented the x-ray image.

B. Median Filter

Median filtering is a nonlinear signal processing technique which is useful for noise reduction in images [6]. It is the best-known order-statistic filter, which replaces the value of a pixel by the median of the intensity levels in the neighborhood of that pixel.

C. Thresholding

Thresholding is one of the simplest segmentation methods. It can extract the object from the background by grouping the intensity according to the threshold value [7].

D. Edge detection

Edge detection is a type of image segmentation techniques which determines the presence of an edge or line in an image and outlines them in an appropriate way [8]. The main purpose of edge detection is to simplify the image data in order to minimize the amount of data to be processed [9].

E. Region of interest

A region of interest, is a selected subset of samples within a dataset identified for a particular purpose. The concept of an ROI is commonly used in many application areas. In medical imaging, the boundaries of a tumor may be defined on an image or in a volume, for the purpose of measuring its size [10]. And It’s property which are the statistical and Geometrical features.

3. FIRST-ORDER STATISTICS FEATURES

First-order statistics measure the likelihood of observing a gray value at a randomly-chosen location in the image. First-order statistics can be computed from the histogram of pixel intensities in the image. These depend only on individual pixel values and not on the interaction or co-occurrence of neighbouring pixel values [11].

Where, \( i \) is between \( 0, 255 \), \( p(i) \) the probability density of occurrence of the intensity, as determined from the histogram with total number of pixels in the image is given as [12].

\[
P(i) = \frac{H(i)}{NM} \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (1)
\]

Where

\( i=0,1,2,\ldots ,G-1 \)

\( G=\)gray level tone of an image.

\( N=\)number of cell in horizontal domine , \( M=\) number of cell in vertical domine. The features which obtained from the histogram.

• Energy

Provide the sum of squared elements. It is also known as uniformity or the angular second moment. Its range is from 0 to 1. Energy is \( E \) for a constant image [13].

\[
E = \sum_{i=1}^{G-1} \left(p(i)\right)^2 \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots (2)
\]
Entropy

Measures the randomness of a gray-level distribution. The Entropy is expected to be high if the gray levels are distributed randomly throughout the image [13]. It is given by

$$H = \sum_{i=1}^{G-1} p(i) \log_2[p(i)] \quad \ldots \ldots \quad (3)$$

mean

It provides the mean of the gray levels in the image. The Mean is expected to be large if the sum of the gray levels of the image is high [13]. It is given by

$$\mu = \sum_{i=1}^{G-1} i \cdot p(i) \quad \ldots \ldots \quad (4)$$

Variance

Variance tells us, how spread out the distribution of gray levels is. The Variance is expected to be large if the gray levels of the image are spread out greatly [13]. It is given by

$$\sigma^2 = \sum_{i=1}^{G-1} (i - \mu)^2 \cdot p(i) \quad \ldots \ldots \quad (5)$$

Standard Deviation

Standard Deviation shows much ‘variation’ or ‘dispersion’ exists from the average [13].

$$\text{Std} = \sqrt{\sigma^2} \quad \ldots \ldots \quad (6)$$

Skewness

Skewness of the histogram is also known as third moment. This measure is 0 for symmetric histograms, positive by histograms skewed to the right (about the mean) and negative for histograms skewed to the left. For smooth images this value comes to be negative. Following equation is used to calculate third moment[14].

$$\mu_3 = \sigma^3 \sum_{i=1}^{G-1} (i - \mu)^3 \cdot p(i) \quad \ldots \ldots \quad (7)$$

Kurtosis

The kurtosis is the flatness of histogram [12]. Kurtosis is the degree of peskiness of a distribution, defined as a normalized form of the fourth central moment $\mu_4$ of a distribution. There are several flavors of kurtosis commonly encountered, including the kurtosis proper [15,16].

$$\mu_4 = \sigma^4 \sum_{i=1}^{G-1} (i - \mu)^4 \cdot p(i) \quad \ldots \ldots \quad (8)$$

4. EXTRACTING THE LUNG DISEASE

In the sample images, the background illumination is clear in the center of the image than at the other regions. The morphological opening operation is used to estimate the background illumination. Morphological in erosion followed by a dilation, the opening operation has the effect of removing object that can not completely contain the structuring element. To create a more uniform background, subtract the background image, background, from the original image, after subtraction, the image has a uniform background but is now too dark. The thresholding has been applied with adding a pseudo –color image a matrix which is RGB to identify each object in the matrix image with a different color in the associated color map matrix.
Cancer Small cell carcinoma of the lung after Five weeks later Cases

Cancer Small cell carcinoma of the lung after Two months later Cases

Fig 1 Shows the left and right part of the lung for the normal and abnormal images with histogram of three cases of at different time

<table>
<thead>
<tr>
<th>Table 1</th>
<th>The statistical features for the special Cancer Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image No.1</td>
<td>mean</td>
</tr>
<tr>
<td>upper right side</td>
<td>36.6054</td>
</tr>
<tr>
<td>Bottom right patient</td>
<td>39.973</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>The statistical features for the special Cancer Case after Five weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image No.2</td>
<td>mean</td>
</tr>
<tr>
<td>upper right side</td>
<td>33.3142</td>
</tr>
<tr>
<td>Bottom right patient</td>
<td>37.6092</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>The statistical features for the special Cancer Case after Two months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image No.3</td>
<td>mean</td>
</tr>
<tr>
<td>upper right side</td>
<td>33.8003</td>
</tr>
<tr>
<td>Bottom right patient</td>
<td>31.3018</td>
</tr>
</tbody>
</table>

Three samples for the same patient have been taken for different time after the patient has received a treatment we have take an image for the patient after five weeks, then after two months. The statistical features for the three images are shown in table (1,2,3) figure (2) shows the (ROI) for the cancer.
4.1 Statistical features for the cancer

After the cancer detected and separated from the other lung texture. We calculate the statistical features.

<table>
<thead>
<tr>
<th>Image No.</th>
<th>RGB image</th>
<th>Threshold after ROI</th>
<th>Edge detection using prewitt</th>
</tr>
</thead>
<tbody>
<tr>
<td>image number 1</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>After Five weeks Image No.2</td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>After Two months Image No.3</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
</tbody>
</table>

**TABLE 4**

The Statistical Feature for the Image shows in figure (2) of Cancer Cases

<table>
<thead>
<tr>
<th>Image No.</th>
<th>mean</th>
<th>variance</th>
<th>energy</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>entropy</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image 1</td>
<td>8.8624</td>
<td>1.9143e+003</td>
<td>0.7213</td>
<td>6.7628e-005</td>
<td>3.5328e-008</td>
<td>0.1915</td>
<td>43.7527</td>
</tr>
<tr>
<td>After Five weeks No.2</td>
<td>14.1698</td>
<td>3.0648e+003</td>
<td>0.6212</td>
<td>2.5659e-005</td>
<td>8.3719e-009</td>
<td>0.2909</td>
<td>55.3606</td>
</tr>
<tr>
<td>After Two months No.3</td>
<td>7.8350</td>
<td>1.6763e+003</td>
<td>0.7102</td>
<td>8.8426e-005</td>
<td>5.2751e-008</td>
<td>0.2009</td>
<td>40.9426</td>
</tr>
</tbody>
</table>

![Graph1](image10.png) ![Graph2](image11.png)

**Fig 3 The Statistical Feature for the Image figure (2)**
5. CONCLUSIONS

This search aim to extract the cancer from the lung by using the segmentation method and the enhancement filter (median) then the region of interest is used to get the cancer out from the rest lung. The comparison of the histogram between the upper and lower part of the lung gives indication about which part carry the cancer, with the help of the statistical feature which has been calculated for the three images after five weeks and after two months table (1,2,3) shows the statistical features for the three images and the comparison between the upper and lower part of the lung, table (4) shown the first-order statistical features for the cancer. The mean value after two month and after the patient has received treatment is lower than before treatment also the skwness is high this mean’s that the texture of the lung become more symmetry than before, the energy increase and the entropy decrease this gives explanation about the number of gray level in the image (i.e) increase energy lower gray level number and the tissue of the lung become homogenous than before this also means that the lung tissue become smooth than before taken the treatment.

REFERENCES


[3] Dr. K. I. Jassam¹, and professor, M. Carr², "REMOVAL OF RANDOM NOISE FROM CONVENTIONAL DIGITAL X-RAY IMAGES", Researcher, the Institute of Islamic Medicine for Education and Research Panama City, FL¹, Department of Surveying Engineering University of Maine Orono, ME², USA Commission No: VII, PP: 113, 1992.


