Lung Cancer Prediction using Fuzzy Inference System

Kavita Kumari, Prabhakar Sharma

Department of Computer Science and Engineering, Raipur Institute of Technology, Raipur, Chhattisgarh, India

Abstract

Now a day’s lung cancer is the leading cancer among all human beings. Early detection of lung cancer can help in a sharp decrease in the lung cancer death rate, which accounts for more than 17% percent of the total cancer related deaths. Presence of lung cancer can be identified with the help of a CT image of lung. Doctor analyses the CT image and predicts the presence of cancer nodule. This manual identification may have the chances for false recognition. So there is a need of automated approach of lung cancer detection Image processing technique can be used for this purpose. In this paper we propose a Lung cancer identification system that uses a fuzzy inference system to spot the most prominent cancer cells. The approach has four stage to detect the existence of cancer nodule in lung. Pre-processing stage, Segmentation stage, feature extraction stage and fuzzy inference rules to identify lung cells. Pre-processing step includes image enhancement. Enhanced CT image of lung is then passed through segmentation phase. From the segmented output features are extracted to predict the existence of abnormality of lung. On these extracted features fuzzy rules are applied to identify the possibility of cancer cells.

Keywords: Lung nodule, Fuzzy Inference System, Image Segmentation.

I. INTRODUCTION

Cancer is now the biggest reason of death in the world. Lung cancer is one of the most common cancers in present days. Due to the lifestyle of people there is a steady increase in cancer patient. Pain, breathlessness, cough, weight loss and fatigue are the general symptoms of cancer. Survival from the disease is not easy if it is not identified at the early stage. Only 15% of lung cancer is recognized at the early stage [1].The reason for failure in detecting lung cancer in early stages is that there is only a dime-sized lesion growth called nodule, inside the lung, and by the time it is identified it is already too late for the patient. Also, these small lesions cannot be identified by X-rays and are only noticed by a CT scan. Even after the detection, it takes a considerable amount of effort and experience of radiologists to identify and mark the nodules as benign or as a probable case of malignancy. Considering the large number of cases encountered by radiologists every day there is a constant pressure on them to analyze a huge amount of data and make a decision as quickly as possible based on the analysis. Normally cancer can be divided into two groups. They are non-small cell lung cancer and small cell lung cancer.

Lung cancer is diagnosed from the CT image of lung. Normally a doctor analyses the CT image of lung and detect the presence of cancer in lung. In this manual diagnose approach may have the chance of false detection. False detection is due to the presence of air in bronchi, presence of ribs and blood vessels etc [2]. So there is a need to develop approach of automated identification of cancer. Image processing tools are the best tool for developing such a automated approach for lung cancer identification. CT image of lung is processed and finds whether the presence of cancer nodule is there or not. There are many image processing tools [3][4] for this purpose. This paper focuses to build an efficient and accurate automated approach for lung cancer detection. This paper proposed a Lung cancer detection system that uses a fuzzy inference system to identify the cancer cells. This approach has four stages to detect the presence of cancer nodule in lung. Pre-processing stage, Segmentation stage, feature extraction stage and fuzzy inference rules to identify lung cells. Pre-processing step includes image enhancement. Enhanced CT image is then passes through segmentation phase. From the segmented image some features are extracted to guess the existence of abnormality of lung. On these extracted features fuzzy rules are applied to identify the possibility of cancer cells.

II. RELATED WORK

In the past, several methods have been proposed to detect and classify lung cancer in CT images using different algorithms. For example, Camarlinghi et al. [5] have used three different computer aided detection techniques for identifying pulmonary nodules in CT scans. Abdulla and Shaharum [6] used feed forward neural networks to classify lung nodules in X-Ray images even if with only a small number of features such as area, perimeter and shape. Kuruvilla et al. [7] have taken six distinct parameters including skewness and fifth & sixth central moments extracted from segmented single slices containing two lungs along with the features mentioned in [6] and have trained a feed forward back propagation neural network with them to estimate accuracy for different features separately. In Bellotti et al. [8], the authors have projected a new computer-aided detection system for nodule detection using active contour based model in CT images. This paper reports a high detection rate of 88.5% with a average of 6.6 false positives (FPs) per CT scan on 15 CT scans.
Hayashibe et al [9] projected an automatic method based on the subtraction between two serial mass chest radiographs, which is used in the detection of new lung nodules. Kanazawa et al [10] presented a system which extract and analyze features of the lung and pulmonary blood vessel regions and then utilizes defined rules to perform identification. Then it was used in the detection of tumor candidates from helical CT images.

Naseer Salman [11] suggested that Marker-driven watershed segmentation extracts seeds indicating the presence of objects or background at particular image locations. The marker locations then set to be regional minima within the topological surface (typically, the gradient of the original input image). Then the watershed algorithm is applied. Mori et al [13] projected a method to extract bronchus area from 3-D chest X-ray CT images, which is used in a virtualized bronchoscope system.

III. METHODOLOGY

Step 1: Select Input Image:

Input image is selected from Database of CT scan lung cancer images taken from website of National Cancer Institute which contains Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI)[14].

All the images are in .dicom format (Digital Imaging and Communications in Medicine) so conversion has been performed to convert it into JPG format.

The image database contain 3 types of images:
1. Images having no nodule ( no cancer cell)
2. Images having nodule size <= 3mm.
3. Images having nodule size > 3 mm.

Image having nodule size less than 3 mm contains less probability of cancer and images having nodule size greater than 3 mm having higher probability of cancer.

Step 2: Enhancement image using Gabor Filter:

Gabor filter is named after its developer Dennis Gabor. It is a linear filter used for edge detection. In Gabor filter Frequency and orientation representations very much are identical to those of the human visual system. And it is found that it is particularly appropriate for texture representation and discrimination. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function multiplied by a sinusoidal plane wave. The Gabor filters are self-similar: all filters can be formed from one mother wavelet by dilation and rotation[13].

The image presentation based on Gabor function constitutes an excellent local and multi-scale decomposition in terms of logons that are concurrently (and optimally) localization in space and frequency domains. Gabor filter is a linear filter whose impulse response is defined by a harmonic function and multiplied by a Gaussian function. Because of the multiplication-convolution property (Convolution theorem), the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function [11].

Step 3: Segmentation Using Thresholding:

Thresholding is the most dominant tool for image segmentation. Thresholding operation converts the grey scale image into binary image. Thresholding operation selects a threshold value T and assigns two levels to the images that is one is above the threshold value and the other is below the threshold value. It can separate the object from the background by using the threshold value T Any point (x,y) for which f(x,y) > T is called an object point, otherwise the point is called a background point.

Select a global threshold value for the entire CT Lung image. By Applying the threshold value to the preprocessed image, it can be converted to binary and the threshold image is obtained.

Figure 1 Methodology
Global Thresholding: A simple algorithm:

1. Initial estimate of $T$
2. Segmentation using $T$:
   I. $G_1$, pixels brighter than $T$;
   II. $G_2$, pixels darker than (or equal to) $T$.
3. Computation of the average intensities $m_1$ and $m_2$ of $G_1$ and $G_2$.
4. New threshold value: $T_{\text{new}} = (m_1 + m_2)/2$
5. If $|T - T_{\text{new}}| > \Delta T$, back to step 2, otherwise stop.

Step 4: Morphological Operation:

To fill in holes and small gaps in the image morphological close operation is applied on the threshold image. Reserve the block whose area is the largest and set the others to zero using 8-connected neighbors. The binary lung mask is obtained using the defined step. Extract the lung boundary by setting a pixel to 0. If it is 4-connected neighbors are all 1’s, thus leaving only edge pixels. Original Lung CT image is multiplied with the lung masked image to get the final segmented lung area with gray level values as those of original image.

Dilation:

The value of the output pixel is the maximum value of all the pixels in the input pixel's neighborhood. If any of the pixels is set to the value 1, the output pixel is set to 1 in a binary image.

Erosion:

It is reverse of dilation i.e. the value of the output pixel is the minimum value of all the pixels in the input pixel's neighborhood. If any of the pixels is set to 0, the output pixel is set to 0 in a binary image.

Step 5: Select Region of Interest:

A part of an image that we want to filter or perform some other operation on is known as Region of interest. An ROI can be defined by creating a binary mask, which is a binary image that is the same size as the image you want to process. The pixels that define the ROI are set to 1 and all other pixels set to 0 in the mask image.

Here, our region of interest is suspected cancer area. Random selection is performed by user as region of interest. Then feature is extracted from selected region to identify probability of cancer in selected region.

Step 6: Feature Extraction

1. Texture Features:

   The most frequently used measures of texture, in particular of random texture, are the statistical measures planned by Haralick. Unlike Laws’ texture energy measures, some of Haralick’s measures may not be directly related to the intersecting structures, speculations, and node-like patterns of architectural distortion, but they may provide useful information regarding the statistical properties of the given ROI or image. Haralick’s texture measures are based upon the moments of a joint probability density function (PDF) that is estimated using the joint occurrence or co-occurrence of gray levels, called as the gray-level co-occurrence matrix (GCM), and may be computed for various directions and distances.

   GCMs are calculated with unit pixel distance for the angles of 0, 45, 90 and 135. So total 4 GCM features can be calculate as texture feature of image in different direction.

2. Local Binary Pattern (LBP):

   The local binary pattern (LBP) feature has a significant importance in the field of texture classification and retrieval. Ojala et al. proposed LBP s, which are converted to a rotational invariant version for texture classification. Various extensions of the LBP, such as LBP, variance with global matching, dominant

3. Area and Perimeter:

   Area: One of the main features of cancer cell is its area and perimeter. Segmented images have two values 1 and 0. Nodule part will be represented with value 1. Then area of the nodule can be calculated by finding number of pixel with value 1. Perimeter of the nodule means the number of pixels in the boundary region of the nodule.

   \[
   \text{Area} = \sum \text{White Pixel in image } I(x,y) \\
   \text{Perimeter} = \sum \text{Pixels in the Boundary of Image } I(x,y)
   \]

   Area of the objects in binary images is a scalar whose value corresponds roughly to the total number of on pixels in the image, but might not be exactly the same because different patterns of pixels are weighted differently.

   Perimeter: A pixel is part of the perimeter if it is nonzero and it is connected to at least one zero-valued pixel. The default connectivity is 4 for two dimensions, 6 for three dimensions.

   \[
   \text{Perimeter} = \sum \text{Pixels in the Boundary of Image } I(x,y)
   \]
IV. RESULT ANALYSIS

Figure 1.2 and 1.3 shows the output score for selected region on applying fuzzy rules. Figure 1.2 shows score 76 which means highest chance of cancer in selected region, whereas in figure 1.3 score is 16 which shows no possibility of cancer in selected cell.

V. CONCLUSION

The above figure shows that fuzzy interference system generates good output for suspected cell selection in lung cancer images. From figure 1.2 and 1.3 we can observed that if selected area is if more than 3mm then it gives accuracy more than 95%.

REFERENCES


