Lung Tumor Detection Using Pixel Value Matching (PVM) Method

G.Vijaya, A.Suhasini, D. Dravida Selvi

Department of Computer Science & Engineering, Annamalai University, Chidambaram – 608002, India

ABSTRACT

Background: Lung tumor detection procedures have been earlier discussed using template matching concept but struggles with accuracy. Objective: This paper proposed a new method called pixel value matching method for tumor detection and it increase the accuracy with low time complexity. Results: The anticipated method composed of following steps: preprocessing by adaptive median filter along with sobel edge detection algorithm, segmentation was performed using Watershed segmentation method, pixel value matching and classification using Back propagation neural network (BPNN). Experimental results show that using BPNN we have reached the Accuracy of 72.5%. Conclusion: pixel value matching (PVM) method outperforms the existing cancer findings in a way such that the tumor can be identified easily through naked eyes.

INTRODUCTION

Cancer is a term used for diseases in which abnormal cells divide without control. There are more than 100 different types of cancer (medicine net). Lung cancer is the uncontrolled growth of abnormal cells in lungs; it divides rapidly and form tumors. Two types of lung cancer: 1.non-small cell lung cancer (NSCLC) 2.small cell lung cancer (SCLC). NSCLC can appear in any part of the lung. It tends to grow and spread quickly, which can make it harder to treat. SCLC often starts in the bronchi near the center of the chest, and it tends to spread widely through the body early in the course of the disease. The routine imaging work-up of suspected lung cancer should include postero anterior and lateral chest radiographs and computed tomography (CT) scan of the entire thorax and adrenal glands. Further imaging workup will be suggested by the patient's record, substantial findings, and lab findings. Of the many image modalities CT image can be helpful to view the inner organs of Lung pleasantly.

Related work:

A lot of approaches have already been presented to identify the tumor and to classify the tumor types. In (Mamta 2009), Adaptive Median Filtering Method for Impulse Noise Detection produces better filtering results with less distortion. The nodule detection using Template matching and thresholding concept was used in (Hosien 2014) and classify tumors by neural network classifier. In (Hong 2014) different types of sobel edge detection methodologies such as vertical/horizontal and Vivado_HLS tool are used to improve the output image vision. Automatically detects the lung nodules using morphological analysis and rule-base lung segmentation in (Serhat 2007). In (Mokheld 2012), efficient lung watershed segmentation technique is used to improve the detection results. In (Rekha 2014) various segmentation technique such as Segmentation Based On Edge Detection, Region Based Segmentation Method and Watershed Segmentation techniques are used for detection and evaluation of tumors.

The authors in (Rachana 2013) used normalized cross co-relation (NCC) technique to implement the template matching method. In (Amandeep 2014), the watershed transform was used in medical image segmentation; produces a complete division of the image in separated regions even if the contrast is poor, thus avoiding contour joining. It is appropriate to use this method to segment the high-resolution remote sensing image. In (Shankar 2014, Shireen 2007) Template Matching method was used for image segmentation. In (Vinod 2013) to extract the lung region, perform region growing segmentation algorithm and classify it as...
cancerous and non-cancerous tumor using artificial neural network method. Back propagation neural network approach is used to obtain good classification result in (Santhosh 2010) and it is used to determine the non-linear relationship that exists between the historical data supplied to the system during the training phase. In (Sudha 2012), efficient nodule detection was performed using thresholding and morphological operations. The authors in (Sonali 2012), signifies the over segmentation problem using watershed segmentation technique and extended edge detection. Template matching method was used to find the structures with similar properties of nodules in (Onur 2007).

**Outline of the Work:**

The proposed system identifies tumor in lung image consists of four core level: (i) enhanced the image (ii) Segmentation of extracted lung (iii) pixel value matching and Feature extraction (iv) classification using Back Propagation Neural Network. The first level starts with the elimination of noise using adaptive median filter. The second level segments the image using Watershed segmentation and the third level performed pixel value matching. Classification using BPNN, which indicates the types of images as benign or malignant. The rest of this paper is organized as follows. Section 2 describes a proposed approach. Performance evaluations are discussed in section 3. Section 4 describes the results and discussion. Finally, section 4 concludes the paper.

**Proposed system:**

The general overview of the proposed approach is illustrated in Figure 1. The input image is taken from Lung Image Database Consortium (LIDC) database for the detection of lung tumor. It's an image database that will assist users to provide an international research resource for the improvement, training, and assessment of Computer Aided Diagnosis (CAD) approaches in the detection of lung nodules in CT scan.

![Overall block diagram of the proposed work](image)

**Preprocessing:**

The image preprocessing stage starts with the elimination of noise and image enhancement. The acquired image may contain large amount of noise which must be removed for smooth processing. The original image is converted into gray scale image because colored image requires a lot of computational time. Enhanced the image, is to improve the interoperability of information included in the image for human viewers. The image enhancement and noise elimination is performed using Adaptive median filter.

![Noisy image and noise eliminated image](image)

**Adaptive median filter:**

The Adaptive Median Filter performs spatial handling to figure out which pixels in an image have been influenced by impulse noise (2). The Adaptive Median Filter specifies a pixel as noise, by matching every pixel in the image with its encompassing neighboring pixels. The size as well as the threshold of the neighboring pixel is customizable. A pixel that is not quite same as the dominant part of its neighbors, and being not structurally adjusted to those pixels to which it is compared, is marked as impulse noise. These noise pixels are then
swapped by the mean estimation of the pixels in the neighborhood. Figure 2 describes (a) salt and pepper noise image and (b) noise eliminated and smoothed image using Adaptive median filter.

The main purpose of Adaptive median filter is to remove impulse noise, smoothing of other noise and reduce distortion, like excessive thinning or thickening of object boundaries. Adaptive median filter changes size of $S_{x,y}$ (the size of the neighborhood) during operation. Adaptive median filter can handle up to 80% noises very well and preserves detail and smooth non-impulsive noise.

**Sobel edge detection:**

Sobel is a popular edge detection method which uses the derivative approximation to find edges (3&9). It returns edges at those points where the slope of the considered image is maximized. The horizontal and vertical gradient matrices are those, whose dimensions are of 3X3 for the Sobel method and it has been broadly used in the edge detection operations. Each direction of Sobel mask forms two new images. One image shows the vertical response and other image shows the horizontal response. The value of the threshold is used to detect edge pixels. Table 1 shows the horizontal and vertical masks of an image.

<table>
<thead>
<tr>
<th>Table 1: Masks of an Image.</th>
<th>Gx</th>
<th>Gy</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>+2</td>
<td>+1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 3 demonstrates the Sobel edge detection method for the given input image.

![Fig. 3: Sobel edge detection method.](image)

The gradient magnitude is computed by the formula

$$|G| = \sqrt{G_x^2 + G_y^2}$$

Approximate magnitude is computed by the formula

$$|G| = |G_x| + |G_y|$$

**B. Image segmentation:**

Image segmentation is an important process for most image analysis tasks. It divides the image into its constituent regions or objects. The main goal is the provision to change the representation of the image into something that is more meaningful and easy to evaluate and then used to locate object and boundaries in images. Image segmentation was performed using the watershed segmentation technique.
Watershed segmentation:

Watershed segmentation separate touching objects in an image (4, 8&14). The slope can be calculated for each pixel position in the image. Each of the connected regions must contain one local minimum in the equivalent slope image. Use internal markers to obtain watershed lines of the slope of the image to be segmented. Use the obtained watershed lines as external markers and each region defined by the external markers contains a single internal marker and part of the background. The problem of watershed lines can be reduced by partitioning the regions and panels into two parts: object (containing internal markers) and background (containing external markers).

Segmentation using the watershed transforms works well if we can identify, or "mark," foreground objects and background locations. Marker-controlled watershed segmentation follows this basic procedure:
1. Compute a segmentation function. This is an image whose dark regions are the objects we are trying to segment.
2. Compute foreground markers. These are connected regions of pixels within each of the objects.
3. Compute background markers. These are pixels that are not part of any object.
4. Modify the segmentation function so that it only has minima at the foreground and background marker locations.
5. Compute the watershed transform of the modified segmentation function.

Histogram:

Histogram is constructed by splitting the range of the data into equal-sized bins (called classes). Each bin is computed from the data set. Here vertical axis represents Frequency (i.e., pixel counts for each bin) and the horizontal axis represents Response variable.

Fig. 5: Segmented images.

(a) Left lung region
(b) Right lung region

Fig. 6: Pixel value matching image.

C. Pixel value matching:

PVM is a matching operation within the pixel value ranges. The fixed pixel range is from 80 to 150. Practically, in this existing research the pixel range (80-150 pixels) was demonstrated as the tumor area. The process is to extract the fixed pixel range from the segmented portion of the lung image. If the pixel range is equal to 80 to 150 ranges, it represented as 1 or white otherwise it represented as 0 or black. Finally the tumor area within this specified range, are represented as the white color. Figure 6 demonstrates the PVM method and the corresponding tumor inside the lung region.

Feature Extraction:

Image feature extraction stage is an important stage that uses algorithms and techniques to detect and isolate various desired portions or shapes (features) of a given image. The purpose of feature extraction is to reduce original dataset by measuring certain features that distinguish one region of interest from another. Features we have used for classification.

AREA - the number of non-black (0,0,0) pixels in the current image.
Area=rows\times columns
PERIMETER - the number of pixels that surround non black blobs.
Perimeter=2*rows+2*(columns-2)
COMPACTNESS - ratio of the square of the perimeter to the area; also known as "shape".
Compactness= \frac{4*3.14*\text{area}}{\text{perimeter}^2}

D. Classification:
In this work back propagation neural network classifier is used for classification. A neural network consists of units (neurons), arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer.

Back propagation neural network:
The back-propagation neural network used in this work has three layers of neurons, i.e., input, hidden, and output (15). Each node in the input and hidden layers is connected to each of the nodes in the next layer (hidden or output). All connections between nodes are directed (i.e. the information flows only one way). Each connection between nodes has a weighting factor associated with it. These weights are modified using the back-propagation algorithm, during the training process to produce "learning". In order to reduce the error rate, the weight associated with the output layer and the hidden layer is adjusted so that actual output meets the expected result.

Fig. 7: BPNN architecture.

Performance evaluation:
The performance of the classifier will be evaluated from the Accuracy, Sensitivity and Specificity. Confusion Matrix is a binary classification model classifies each instance into one of two classes: a true and a false class. It is a table with two rows and two columns that reports the number of false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN) values. This allows more detailed analysis than mere proportion of correct guesses (accuracy). For supervised learning with two possible classes, all measures of performance are based on four numbers obtained from applying the classifier to the test set.

To compute Accuracy, Sensitivity and Specificity:
Accuracy is the condition or quality of feature being true, correct, or exactly identified.
\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}
\]
Sensitivity measures the ratio of the proportion of people with disease who have a positive test result.
\[
\text{Sensitivity} = \frac{TP}{TP + FN}
\]
Specificity measures the ratio of the proportion of people without disease who have a Negative test result.
\[
\text{Specificity} = \frac{TN}{TN + FP}
\]
Where,

- true positive (TP) - Malignant patient correctly diagnosed as Malignant.
- true negative (TN) - Benign patient incorrectly diagnosed as Malignant.
- false positive (FP) - Benign patient correctly identified as Benign.
- false negative (FN) - Malignant patient incorrectly identified as Benign.
Table 2: Structure of confusion matrix.

<table>
<thead>
<tr>
<th>ACTUAL</th>
<th>PREDICTED</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>POSITIVE</td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>NEGATIVE</td>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>

Table 3: Performance Measures.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Values (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>72.5</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>79.8</td>
</tr>
<tr>
<td>Specificity</td>
<td>65.2</td>
</tr>
</tbody>
</table>

RESULTS AND DISCUSSIONS

The proposed method has been implemented using Matlab in which 80% of images were trained and 20% of images were tested and produced the efficient results with high accuracy. The LIDC (Lung Image Database Consortium) dataset was used in this study. Tumor was detected in the lung CT image based on the pixel value matching method. If the pixel range is from 80 to 150, that was represented as tumor. Area within this range is taken out and the corresponding features are extracted. Table 3 shows the test evaluated result produced as 72.5% of accuracy, and the Specificity and the Sensitivity as 79.8% & 65.2% respectively.

Conclusion and future work:

The effective Computer-Aided Diagnosis (CAD) system is used to assist physicians in the creation, modification, analysis, or optimization of a scheme. CAD system for effective lung tumor detection was proposed in this paper. The proposed method consists of four main steps: Noise elimination, Segmentation, Pixel Value Matching (PVM) and BPNN classification. Adaptive median filter is applied for noise elimination and watershed segmentation technique for segmentation. Also a PVM approach is employed to extract tumor area from the segmented image. Finally features are extracted from the tumor and they are classified into benign or malignant using BPNN classifier. The proposed CAD system has been evaluated against several images in the database and results show prominent progress.

The future enhancement was suggested to evaluate effective method by extending the matching technique with other features and to identify the effective features for further classification.

REFERENCES


