Identification Of Malignant and Benign Lung CT images using segmentation tools

1*Surekha Nigudgi,
Assistant Professor,ACS College Of Engineering,Bangalore, sursanju@gmail.com

2*Dr Channappa Bhyri,
Professor&Head of E&IE dept,PDA College Of Engineering,Kalaburgi, channubhyri@yahoo.com

Abstract

Discovering of cancer is the most extreme interesting investigation space for researcher in the early period. The method used to identify lung cancer with lung CT images in the begining stage so that the survival rate of the patient is high. The methods involve with several stages such as image aquisition,filtering techniques, segmentation techniques, morphological operations and classification. Here in this paper segmentation is perform to disintegrate an lung CT images where gradient magnitude technique is to extract certain essential features from the segmented images, further extracted features are considered for support vector method to classify whether it is a malignant or benign lung CT images.

Keywords: Morphological operations, Gradient magnitude technique, malignant, benign, SVM, Lung CT Images.

Introduction

Cancer is one of the crucial reasons for non-accidental death. One of the highest among all cancers is the lung cancer. Cancer is a disease caused when cells divide out of control and spread into neighboring tissues. Cells go unpredictable development widely and unbalance to form malignant tumor and damage the surrounding cells. Lung cancer when it diagnosed and treated in early stage there will be less mortality rate associated with this disease. Cigarette smoking is one of main factor, which accounts that the 85% of the patients are with the lung cancer[16]. Lung cancer, in nations like industrialized advances the social and cultural patterns have led to rising rates.[15]. Mortality rates of lung cancer among females have lesser than male, about 40 deaths per 100,000 among female and 90 deaths per 100,000 among males. These are similar when it is broken down into racial/ethical group[17]. Imaging modalities are categorized in which images are generated, ultrasound, radiation such as x-rays, MRI, Computed tomography to develop Computer aided diagnosis. Here in this paper, Lung CT images are used. CT scan lung images are used. CT is a computerized tomography imaging in which a of x-ray beam is positioned at a patient body and which is rotated around the body, to generate crossectional images or slices of the body. These slices are called as tomographic images which contain a detailed information for further analysis.[18].

The scope of this proposed methodology is to design a system in which the lung CT images are taken as inputs and gives particular output. The method used are efficient in terms of accuracy, specificity and sensitivity. The proposed methodology
consists of following steps such as: Acquisition of lung CT images, removing unwanted noises with preprocessing steps, contrast stretching is performed, morphological operations is performed, segmentation is done with further extracted features are considered for support vector method to classify whether it is a malignant or benign from lung CT images.

**Literature Survey**

Ashwini P*, Sherin Antony, Kanchana V et al[1], Authors in this paper they have proposed the solid feature extraction technique to extract essential features from segmented images. Later the extracted features they are considered for multi support vector machine to classify it as cancer or non-cancerous. Muzzamil Javaid et al[2], proposed a CAD lungs from CT scan are segmented using intensity thresholding, for better segmentation results histogram analysis is done. Morphological operations like opening to remove impurities for segmentation outcomes and closing is used for juxtapleural nodules for segmented lung regions. K-means clustering is connected for the beginning location and division of potential nodules. These sectioned potential knobs are at that point isolated into six bunches on the premise of their thickness and rate network with lung dividers. Timor Kadir et al[3] Author proposed an diagram of the most lung cancer forecast approaches proposed to date and highlight a few of their relative qualities and weaknesses. We talk about a few of the challenges within the improvement and approval of such strategies and diagram the way to clinical appropriation. Rachid sammouda et al[4], authors proposed an updated technique of hopfield fabricated neural organize classifier is identified and the method which is used to disintegrate lung areas which means segmentation technique is used it from human chest computer tomography pictures. The images are acquired utilizing this computer tomography imaging techniques from normal subjects and the techniques is used to identify the diseases and diagnose it. Ratishchandra Huidrom et al[5], proposed a method of segmentation which performs speed than the existing methods. Here the method uses optimal thresholding method the accuracy of the proposed method is speed as compared with the K-means method. Prenitha Lobo [et al], has presented segmentation and classification techniques used in detecting the lung cancer and and it has proposed an effective identification of lung tumor which gave 79.166% accuracy. Shahruk Hossain et al[7], has proposed an automated method for lung tumor detection and segmentation from lung ct images scan which is 3D from Radiomics dataset, further it also presents novel dilated hybrid -3D CNN for tumor segmentation. First segmentation of tumor is implemented and the slices are passed to the segmentation which extracts the features, secondly they are passed through the post processing block which cleans up through morphological operations. Muhammad Zia ur Rehman et al[8], has reviewed a systematic analysis of detection of lung cancer with a analysis of different trends and future challenges. A method has been propose to overcome these challenges and developing a method of computer aided detection (CAD) system. Authors also proposed that a team work is required among soft developers, technicians, physicians and other related stakeholders to understand deeply about the issues which is particular needs of a computer aided detection systems and where automatic techniques to overcome the upcoming challenges with image processing, low cost of implementation and there will low cost implementation and security software assistance. H. Mahersia, M. Zaroug et al[9], analysed and overviewed that current techniques and approaches have been investigated in the literature. It also gives the accuracy and performance of the existing
approaches. CAD systems for lung cancer have proposed in several studies. Siddharth Bhatia et al[11], proposed deep learning residual of techniques with preprocessing stage to highlight lung regions which are exposed to cancer and extract features of UNet and ResNet models, later these features are fed into multiple classifiers i.e XGBoost and Random Forest. The accuracy achieved is 84% as compared with the earlier. A. Asuntha1, A. Brindha et al[12], proposed a model and proposed using PSO, genetic optimization and SVM algorithm, these techniques are used for feature extraction and classification, super pixel segmentation and Gabor filter is used for demising the medical images, later comparison is done between three medical images.

Methodology

1. Lung CT Image Acquisition
2. Processing
3. Red Green Blue to Gray Conversion
4. Enhancement of Image (Contrast Stretching)
5. Segmentation
6. Morphological operations (Opening & Closing)
7. Feature extractions
8. SVM Classifier (Gradient magnitude Technique)
9. Classification
The above proposed methodology as shown in figure 1. For experimental 20 images are taken from which images are of normal, malignant and benign lung CT images.

Pre-processing:

The aim of the pre-processing is the improvement of the image that suppress the unwanted to remove artifact and denoise from the given CT images or enhances some important image features for further preprocessing.

The images which are acquired are converted to gray image and remove unwanted noises from median filter and it is enhanced by contrast stretching.
Segmentation is a process where it divides into smaller parts and as it increases accuracy, reliability, precision and decreases cost for computational detection[8]. The enhanced lung image is then segmented by thresholding technique(method); here in segmentation first the Lung image is divided into left lung and right lung; for particular left lung and right lung cancer is identified. For region of interest in this proposed system the affected region is extracted using median filtering and for that morphological operations is used and perform dilation and erosion. Here non-linear method is used and it is very effective removing noise and it preserves its edges[13], and in final stage features are extracted using SVM classifier with gradient magnitude technique. Where in 13 statistical parameters are identified out of which contrast, coorelation, energy, Homogenity, mean, standard deviation, entropy, RMS, varaiance smoothness IDM has been identified to classify it as normal, malignant and beginn. The formulae are as follows:

\[\text{image contrast} = \max(\text{grayImage}(:)) - \min(\text{grayImage}(:))\]

Homogeneity has the similarity between the largest element and smallest element in that region. \[\text{energy} = \lim_{t \to \infty} z(t)\]

\[\text{mean Intensity} = \text{mean(img(:))}\]

Standard deviation

\[\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}\]

Entropy

\[\text{Entropy} = -\sum_{N_{\text{gi}=1}} \sum_{N_{\text{gj}=1}} P(i,j) \log_2 [P(i,j)]\]
RMS:
\[ X_{\text{rms}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} |X^2_i|} \]

Variance: A variance is an image of the variances, that is the squares of the standard deviations, in the values of the input or output images.

variance =\[ \sum N_i = 1 \sum N_j = 1 (i - \mu)^2 P(i, j) \]

Smoothness:

It is used as a method of smoothing images, reducing the amount of intensity variation between one pixel and the next, resulting in reducing noise in images. \( Y = \text{filter2}(h, X) \) filters the data in \( X \) with the two-dimensional FIR filter in the matrix \( h \).

Kurtosis:
\[ \gamma = \frac{1}{N} \sum_{i=1}^{N} \frac{X^4 - u^4}{\sigma^4} - 3 \]

IDM (Inverse difference Moment) =\[ \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{1}{1 + (i-j)^2} P(i, j) \]

Classification:

Classification is a process where here it is into malignant, benign and normal from lung CT images, for this process support vector machine is used and it a machine learning algorithm, which can be used for classification to find the hyperplane which will help to differentiate data points. From these data points which are falling on either side of hyperplane where it can be classified to different classes. In this step the scope is to identify types of cancer is is a malignant benign and noemal based on sensitivity, specificity and accuracy.

Sensitivity = \( \frac{TP}{TP + FN} \)
TP = True Positive, FN = False negative
Specificity = \( \frac{TN}{TN + FP} \)
TN = True Negative, FP = False positive
Results:

Figure 8: Shows a set of malignant images

Figure 9: Shows a set of benign images

Figure 10: Shows a set of normal images
Table 1: Shows the support vector machine (SVM) parameters

<table>
<thead>
<tr>
<th>Features</th>
<th>Normal Images/1</th>
<th>Malignant Images/2</th>
<th>Benign Images/3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Contrast</td>
<td>0.1852</td>
<td>0.1898</td>
<td>0.1817</td>
</tr>
<tr>
<td>Correlation Index</td>
<td>0.0911</td>
<td>0.1316</td>
<td>0.0715</td>
</tr>
<tr>
<td>Energy</td>
<td>0.7967</td>
<td>0.8194</td>
<td>0.9204</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>0.9452</td>
<td>0.9509</td>
<td>0.9779</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0029</td>
<td>0.0017</td>
<td>0.001</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.0771</td>
<td>0.0801</td>
<td>0.0611</td>
</tr>
<tr>
<td>Entropy</td>
<td>3.3203</td>
<td>2.9404</td>
<td>2.4485</td>
</tr>
<tr>
<td>RMS</td>
<td>0.0772</td>
<td>0.0801</td>
<td>0.0611</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0059</td>
<td>0.0064</td>
<td>0.0037</td>
</tr>
</tbody>
</table>
### Table 2: List of parameters of correlation, entropy, Kurtosis and Skewness

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Normal</th>
<th>Malignant</th>
<th>Begign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoothness</td>
<td>0.9261</td>
<td>0.886</td>
<td>0.9365</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>7.7963</td>
<td>9.4719</td>
<td>54.422</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.7071</td>
<td>0.76</td>
<td>3.8035</td>
</tr>
<tr>
<td>IDM (Inverse Difference Moment)</td>
<td>0.9897</td>
<td>0.5833</td>
<td>5.047</td>
</tr>
<tr>
<td>Correlation Index</td>
<td>0.0911</td>
<td>0.1316</td>
<td>0.0715</td>
</tr>
<tr>
<td>Average</td>
<td>0.098</td>
<td>0.2117</td>
<td>0.147</td>
</tr>
<tr>
<td>Entropy</td>
<td>3.3203</td>
<td>2.9404</td>
<td>2.4485</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>7.7963</td>
<td>9.4719</td>
<td>54.422</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.7071</td>
<td>0.76</td>
<td>3.8035</td>
</tr>
</tbody>
</table>

**Discussion**

For experimentation 20 images are taken from which images are of normal lung CT images and images are of malignant and benign. The images acquired are converted to gray images and then unwanted noises is removed using median filter and then the image is enhanced by contrast stretching. This is the enhancement method which improves image by stretching the range of intensity values. The image is resized to 250*250. The resized image is segmented right and left to determine the cancer nodules in the lungs. This phase will help to identify the region in the lung nodule that can help to identify the normal, malignant and benign.

The 13 statistical parameters features for lung CT images are taken for experimentation using sequential forward selection algorithm in SVM classifier is applied. Here 2 classes (Class 1 as cancer in fig 11, class 2 as normal in fig 12) is used. An support vector machine is a classifier which differentiate the data to separate all data points of one class from those of other class by finding the best hyperplane. Performance is measured by accuracy, specificity, sensitivity as shown below;
Calculation:

Accuracy Analysis is performed for one parameter.

From the above Table2: 0.2 value is a threshold which is fixed by taking the average of normal, malignant and benign.
If 0.2> then it is malignant,
If 0.2< then it is a normal or benign.

Coorelation

A set of 20 images are taken for experimentation out of which normal, benign and malignant. Only 9 images are taken and are tabulated in table1 from which the specificity, sensitivity and accuracy are calculated one in normal is true negative and one in malignant also true negative and remaining 7 images are true positive so

Sensitivity = True positive /True pasitive +False Negative

=7/9=77%

Specificity= True Negative/True positive +false negative

=2/9=22%

Accuracy= True positive /Total number of images

=7/9=77%

sensitivity =77%
specificity = 22%
accuracy =77%.
Table 3: Performance of SVM classifier

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Accuracy</th>
<th>Specificity</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung Images</td>
<td>90%</td>
<td>22%</td>
<td>90%</td>
</tr>
</tbody>
</table>

**Conclusion**

A Set of 20 images of which normal, benign and malignant lung CT images are taken for experimentation. The 13 statistical parameters like contrast, correlation index, Energy, Homogeneity, mean, standard deviation, entropy, RMS, variance, smoothness, kurtosis, skewness, IDM (inverse, difference moment). The proposed methodology is to classify normal, benign and malignant tumors correctly by SVM classification and handle for better tumor classification [13]. The proposed method is providing 90% accuracy but, the 100% accuracy of the method can be concluded by considering more images.

**References**

1) Ashwini P*, Sherin Antony, Kanchana V Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham Campus, Mysuru, Karnataka, India "Categorizing the stages of lung cancer using Multi SVM Classifier" International journal of research in pharmaceutical sciences, 10(3), 2323-2328 ISSN: 0975-7538, 2019, Pharmoscope.org © 2019.


16) Suresh S. Ramalingam, MD1; Taofeek K. Owonikoko, MD, PhD2; Fadlo R. Khuri, MD3" Lung Cancer: New Biological Insights and Recent Therapeutic Advances" CA CANCER J CLIN 2011;61:91–112.

