International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:2682
Vol.5, No.5, May 2016 **DOI:10.15693/ijaist.2016.v5.i5.61-64**

Lung Cancer Detection Using Bayesian Classifier

Hitha Rocky
Department of Computer Science and Engineering
Adi Shankara Institute of Engineering and Technology,
Kalady

Dr.Jereesh A.S
Assistant professor/Department of Computer Science and Engineering
Cochin University of Science and Technology,
Ernakulam

***Abstract*— Medical image processing is being widely used in the medical field for early detection of chronic diseases like lung cancer. Time is an important factor in the treatment of lung cancer. . If lung cancer is detected in time, the overall 5-year survival rate of cancer patients increases from 14 to 49% . Computed Tomography (CT) images being more efficient than X-rays are preferred for diagnosis. The challenge lies in choosing the best/most accurate segmentation and classification technique for isolating the cancer affected parts. Firstly, segmentation is used for selecting the region of interest for further processing. Secondly, a classifier is used for distinguishing between the diseased and non-diseased parts with the selected features as input. Segmentation can be done by simply considering the contrast variations in the image, but this will not work well with images of varying density.This paper tries to find out the best segmentation method for the image. The classification step is done using Bayesian Classifier**

Index terms -. Computed Tomography, Small Cell Lung Cancer, Non-Smal Cell Lung Cancer,Fuzzy C-Means

I. INTRODUCTION

Lung Cancer has the highest mortality rate amongst all other types of Cancers. Survival from lung cancer is directly related to its growth at its detection time. The earlier the detection is, the higher the chances of successful treatment are. The detection and localization of lung cancer in the micro invasive stages improves the chances of survival of a patient. However, the detection and localization of lung cancer in the micro invasive stages is very difficult.

Cells are the building blocks of tissues.They are always growing and replace the old ones.There are two types of tumors.One is benign and the othermalignant. Cells in benign cancer do not spread to other parts of thebody while the cancerous cells keep on dividing and spread into other parts of the body in malignant cancer. The spreading of tumor from one part of the body to another is called metastasis.

The incidence of lung cancer is strongly correlated with tobacco smoking, with about 90% of the cases arising as a result of tobacco use. Passive smoking, or the inhalation of tobacco smoke from other smokers, is also an established risk factor for the development of lung cancer. Research has shown that nonsmokers who reside with a smoker have a 24% increase in risk for developing lung cancer when compared with other nonsmokers.

.

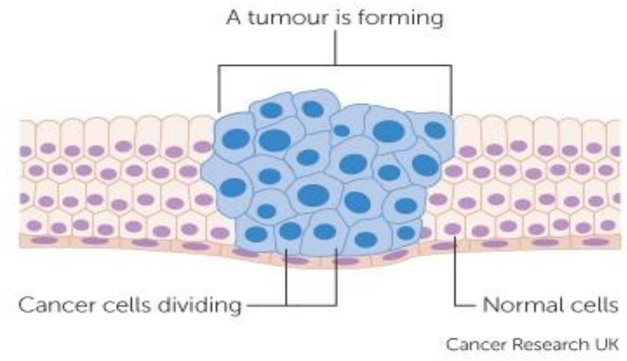

Figure 1. : normal and benign cells

There are two main types of lung cancer, small cell (SCLC) and non-small lung cancer (NSCLC).About 85% to 90% of lung cancers are non-small cell lung cancer, and only 10-15% is small cell lung cancer. Following are the subtypes of Non-small cell Lung cancer:

- Adenocarcinoma
- Squamous cell carcinoma
- Large cell carcinoma

Small cell Lung cancer spreads faster than non-small cell Lung cancer. In “limited stage”, cancer is found only in the chest whereas in the ”extensive stage” it spreads outside the chest region.

Selection of best imaging modality is the key to detection of the corresponding disease[1]. X-ray, MRI, PET scan and CT scans are the different imaging modalities currently used. Among the above, CT scan is used for the three-dimensional view of the Lung images[2].

Segmentation is the process of dividing an image into different regions .The pixels in each group will have similar characteristics. The criteria for grouping the pixels may depend on the properties of the pixels. Watershed segmentation is used for segmenting two touching objects [3]. For this, topological view of the object is used. Watershed segmentation controls over-segmentation using markers. For selecting a marker, preprocessing and finding the criteria that markers should satisfy also have to be considered. The main advantage of watershed is that, even if there are no strong edges between the markers, the watershed transform always detects a contour in the area. The modified adaptive fissure sweep first pre-processes the CT images to reduce noise. Wiener filters are used for noise removal instead of a median filter. The major challenge in this method is the variable shape and appearance, along with the low contrast and high noise associated with these images.Channeler Ant Model [4] is effective whenever complex connected structures are present in the image. Algorithm consists of selecting the number of ants. The maximum number of visits for each ant in the voxels determined primarily. Here selection of number of ants which is done manually is a challenging task.

Clustering is a method of grouping data objects into different groups, such that similar data objects belong to the same cluster and dissimilar data objects to different clusters [5][6]. The algorithm is formulated by modifying the distance measurement of the standard FCM algorithm to permit the labeling of a pixel to be influenced by other pixels and to restrain the noise effect during segmentation.FCM fails to segment images corrupted by noise, outliers and other imaging artifact.

Region Growing is an approach in which neighboring pixels are examined and added to a region class as long as no edges are found between them [7]. Choosing region membership is more difficult than applying edge detectors. It cannot search objects that span different disconnected regions

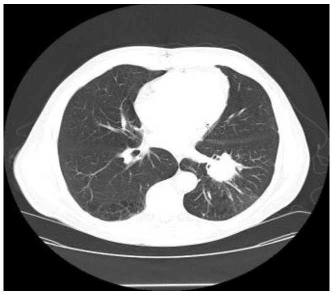

Figure 2. input image

In morphological operation, segmentation of the lung region is based on thresholding method [8]. The threshold is determined by analyzing the 2D region histogram, which shows distinct groups of pixels belonging to the thorax and background air.

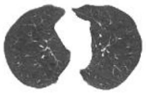

Figure 3. After morphological operation

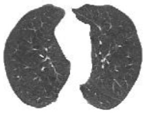

Figure 4. Segmented output

Second module is feature extraction, which is used as input to the classifier.High-level feature extraction concerns finding shapes in computer images. To be able to recognize faces automatically, for example, one approach is to extract the component features. Intensity based features, shape, texture and wavelet are the features extracted here.One of the techniques used to improve classification performance is the inclusion of clustering into the classification process.

In feature extraction, invariance properties are used so that the extraction process does not vary according to the chosen (or specified) conditions. That is, techniques should find shapes reliably and robustly, whatever the value of any parameter that can control the appearance of a shape.Otsu thresholding is used for segmenting the CT Lung image.Segmentation is used to extract features from the selected region of interest.Features like shape, size, mean, standard deviation and other statistical parameters are extracted from the segmented region for further investigation.

Classifier is used for categorizing an unknown pixel into its corresponding group. Bayesian Classifier is used for outputting the type of the input image. Classification based on Random forest clustering consists of training and testing set [9]. Here, the entire data is grouped together and then the dataset is grouped into nodules and non-nodules. It compared different ratio of training and testing set.

In neuro-fuzzy approach prior knowledge about the training data set can be encoded into the parameters of the neuro-fuzzy classifier. Moreover, the parameters obtained after the learning process can be easily transformed into structured knowledge in the form of fuzzy if-then rule [5].

One of the important problems in fuzzy clustering is how to design membership functions. In this method, based on the input and output of the fuzzy system, neural network is trained. The most common neural network model is the multilayer perceptron (MLP). This type of neural network is known as a supervised network because it requires a desired output in order to learn. The goal of this type of network [10] is to create a model that correctly maps the input to the output using historical data so that the model can then be used to produce the output when the desired output is unknown.

II.MATERIALS AND METHODS

A. *DATA*

The role of a classifier is to correctly identify the group of an unknown pixel.Bayesian classifier is based on Bayes theorem. The Bayesian Classification is a type of statistical method for classification. Itpredicts the class or group the given sample belongs to.For this, Bayesian classifier makes use of the class membership probabilities, such as the probability that a given sample belongs to a particular class. Bayes Rule is stated as follows: "Given a problem instance to be classified, represented by a vector x=(x1….xn) representing some n features (independent variables), it assigns to this instance probabilities

P = (Ck|x1……xn) for each of the K possible outcomes or classes."

The problem with the above formulation is that if the number of features n is large or if a feature can take on a large number of values, then basing such a model on probability tables is infeasible. Therefore reformulation of the model is good to make it more tractable. Using Bayes theorem[12], the conditional probability can be decomposed as

$$P(C_k|x) = \frac{p(C_k)\ p(x|C_k)}{p(x)} \qquad (1)$$

the above equation can be written as

$$\text{posterior} = \frac{\text{prior x likelihood}}{\text{evidence}}$$

III.EXPERIMENTAL RESULTS

The training set for the Bayesian classifier consists of70% of the total images and the testing set is 30% of the total images. The sensitivity, specificity andaccuracy are calculated as follows:

Sensitivity

It measures the proportion of actual positives which are correctly identified. That is the percentage of segmented slicescontaining cancerous nodule is correctly classified as cancerous:

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

Specificity

It measures the proportion of negatives which are correctlyidentified. The percentage of segmented slices without can-cerous nodule is correctly identified as non cancerous:

$$\text{Specificity} = \frac{TN}{TN+FP}$$

Accuracy

Accuracy is a statistical measure of how well a classifier correctly identifies or excludes a condition. The accuracy is theproportion of true results (both true positive and true negative)in the population.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Metric	Observed value
Sensitivity	89.4%
Specificity	94.1%
Accuracy	91.6%

Table1: Classification results using Bayesian Classifier for cancerous and noncancerous images

VI. CONCLUSION

Features are used as input for the classifier. Bayesian classifier is a promising method for correct classification.For this, Lung region is extracted from the original CT image. From the lung region, the ROIs were obtained. The nodules are evaluated based on the features such as mean, standard deviation, skewness, kurtosis, fifth and sixth central moment ,then subjected to classification to classify the input image. The project may be expanded by adding additional features like HOG. The existing classifiers can be compared and the better one may be used for improving the efficiency of the system.

REFERENCES

[1]. Awais Mansoor , Segmentation and image analysis of abnormal lungs at ct: Current approaches,challenges, and future trends, Radiographics ,2015.

[2]. Elisabeth Brambilla H.,Konrad Müller-HermelinkCurtis C, Harris William D,Travis,Pathology and genetics of tumours of the lung, pleura, thymus and heart, WHO Publications Center, 2004.

[3]. Lung lobe segmentation by anatomy-guided 3d watershed transform, Medical Imaging 2003: Image Processing, Vol. 4, No. 2, pp. 1482-1490,2003.

[4]. Piergiorgio Cerello and Sorin Christian Cheran, 3-d object segmentation using ant colonies,IEEE Nuclear science Symbosium Conference, ISSN :1082-3654, 2008.

[5]. S.Sivakumar and Dr.C.Chandrasekar,Lung nodule detection using fuzzy clustering and Support vector

machines,International Journal of Engineering and Technology (IJET) 5 (1), 179-185, 2013.

[6]. R.E. Wood, R.C.Gonzalez.,Digital image processing, ISBN number 9780131687288,3rd edition, Prentice-Hall,2008.

[7]. Atiyeh Hashemi., Mass detection in lung ct images using region growing segmentation and decision making based on fuzzy inference system and artificial neural network, I.J. Image, Graphics and Signal Processing, 2013, 6, 16-24.

[8]. Jayashree,P Sudha.V, Lung nodule detection in ct images using thresholding and morphological operation, International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319–6378, Volume-1, Issue-2, December 2012.

[9]. S.L.A. Lee., A random forest for lung nodule identification., Computerized Medical Imaging and Graphics 34 (2010) 535– 542,2010.

[10]. M.G. Penedo, Computer-aided diagnosis a neural-network-based approach to lung nodule detection,IEEE Transactions Medical Imaging . 1998 Dec;17(6):872-80

[11]. M. Madheswaran and D. Anto Sahaya Dhas, Classification of brain MRI images using support vector machine with various Kernels, Biomedical Research 2015; 26 (3): 505-513

[12]. Naïve Bayesian Classifier-English Wikipedia. https://en.wikipedia.org/wiki/Naive_Bayes_classifier.

Authors Profile

Hitha Rocky is currently doing her master's degree in Technology, specializing in Computer Science and Engineering at Adi Shankara Institute of Engineering and Technology, Kalady. Her areas of interest include image processing, Neuro-Fuzzy

Dr.Jereesh A.S is currently working at Cochin university of Science and Technology as Assistant Professor in Computer Science Department. Received Phd from NIT.His area of interest is Image Processing.