Fuzzy Based Tumor and Lymph Node Detection in Thoracic Images Using Soft Computing

N.Indhumathi, PG scholar

B.Kannan, Assistant Professor

V.S.B. Engineering College, Karur, Tamil Nadu

Abstract - The computed tomography (CT) scanners will generate hundreds of slices to visualize the condition of lung for patient. The analysis on slices-by-slices dataset is time consuming for radiologists. Therefore, automated identification of abnormalities on CT lung images is vital to assist the radiologists to make an interpretation and decision. Segmentation is used to identify the abnormality region from the image. Initially, the basic image processing techniques such as Median Filter, Erosion and Dilation are used to remove the noise and fill small gaps and holes. Next, the features such as GLCM, Local Binary Pattern (LBP) and Gray Level are extracted from the images. From the extracted features a set of diagnosis rules are created using Fuzzy logic to identify the cancer nodules clearly. The support vector machine is used for nodules classification because of its better classification rate. The main objective of this paper is to develop an automated system for identifying the lung lesions using the lung CT Thoracic images and classify the lesions as Benign or Malignant.

Keywords – **Gray-level co-occurrence matrix(GLCM), Local Binary Pattern(LBP), Support Vector Machine(SVM),CT Thoracic images, Abnormality region, Fuzzy Logic**

I. INTRODUCTION

Lung cancer is considered to be as the main cause of cancer in death worldwide, and it is hard to detect in its early stages because symptoms appear only at advanced stages causing the mortality rate to be the highest among all other types of cancer. More people die due to lung cancer compared to other cancers. There is significant evidence indicating that the early detection of lung cancer will decrease the death rate. The most recent estimates according to the data provided by world health organization indicate that approximately 7.6 million deaths worldwide each year because of this type of cancer. Furthermore, transience from cancer are expected to continue rising, to become approximately 17 million worldwide in 2030.

The most familiar cancer that occurs usually for men and women is lung cancer. According to the report submitted by the American Cancer Society in 2003, lung cancer would report for about 13% of all cancer diagnoses and 28% for all cancer deaths. The survival rate for lung cancer is analyzed in 5 years is just as 15%. However, only15% of diagnosed lung cancers are at this early stage. Unfortunately, clinical symptoms of lung cancer, such as shortness of breath, chronic cough, and hemoptysis, usually do not occur until the disease has reached a more advanced stage, when patient prognosis is especially poor.

The 5-year survival rate for lung cancer patients is only 13%.

II. RELATED WORK

The computer-aided diagnosis (CAD) system is used for early detection of lung cancer by analyzing chest 3D computer tomography (CT) images. The underlying idea of developing a CAD system is not to delegate the diagnosis to a machine, but rather that the machine algorithm acts as a support to the radiologist and points out locations of suspicious objects, so that the overall sensitivity is raised.CAD systems meet four main objectives, which are improving the quality and accuracy of diagnosis, increasing therapy success by early detection of cancer, avoiding unnecessary biopsies and reducing radiologist interpretation time.

A similar type of work is on lung tumor detection, which first detects all abnormalities, then extracts only those that are highly representative of lung tumors. By first segmenting the lung field, then the fuzzy-logic based approach is then used to detect the lung tumors but the detection performance is quite sensitive to the delineation accuracy of the lung field.

Another approach attempts to handle tumors lying close to the edge of lung fields by incorporating the location, intensity, and shape information [9], but the method could potentially result in a large number of false positives with the predefined SUV thresholds. To reduce the false positives detected in the mediastinum, learning-based techniques with tumor-specific features were proposed but the methods were based on empirical studies of SUV distributions and tumor sizes, and did not appear to consider abnormal lymph nodes in the thorax.

Another category of abnormality detection is to detect all instances from PET images, regardless of their types. Such approaches include a texture-based classification method, a water-shed based algorithm integrated with morphological measures and a region-based SUV threshold computed based on the object and background ratio.

III. METHODOLOGY

An overview of proposed method is presented in Figure 1.

Lung CT Image

↓

Preprocessing

↓

Segmentation

↓

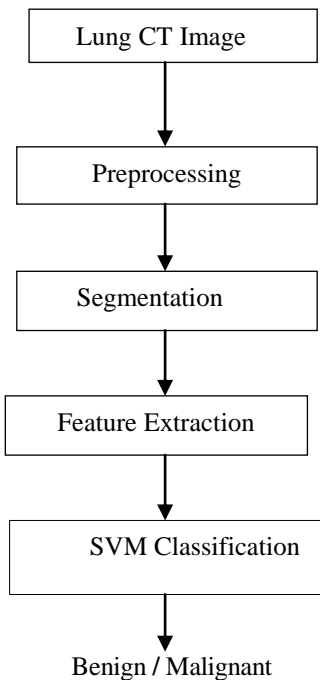Feature Extraction

↓

SVM Classification

↓

Benign / Malignant

Fig 1: An overview of proposed method

Preprocessing system

Preprocessing is the first step for detecting the lung image. In preprocessing step are done in two Steps. They are
Denoising
Median Filter

Denoising

Image denoising algorithms may be the oldest in image processing. Many methods, regardless of implementation, share the same basic idea in order to reduce the noise through image blurring. Blurring can be done in locally, as same in the Gaussian smoothing model or in anisotropic filtering by calculating the variations of an image.

White noise is one of the most common problems in image processing. Even a high resolution photo is bound to have some types of noise in it. For a high resolution photo is a simple box blur may be sufficient, because even a petite features like eyelashes or cloth texture will be represented in terms of a large group of pixels. However, current DirectX 10 class hardware allows us to implement high quality filters that run at acceptable frame rates. The main idea of any neighborhood filter is to calculate pixel weights depending on how the similar their colors are. Here two

such methods are described: the K Nearest Neighbors method and Median filters.

The input image is a normal RGB image RGB image is converted into gray scale image because the RGB format is not supported in Mat lab. Next the gray scale image contains noises such as white noise ,salt and pepper noise etc. This noise can be removed by using Median filter.

Median

The median filter is a nonlinear digital filtering technique, often used to remove noise from the image. Such noise reduction is a typical pre-processing step to improve image quality. Median filtering is the most widely used filter in digital image processing under definite conditions.

Segmentation

1) Otsu's Thresholding: Approximate segmentation of the lung portion in the CT image is performed by using Otsu's Thresholding method. LEVEL = GRAYTHRESH (I) computes a global threshold value that can be used to convert an intensity image to a binary image with the mat lab command IM2BW. LEVEL is a normalized intensity value that lies in the range [0, 1].

2) Morphological Opening Filter: The most basic morphological operations are dilation and erosion. The Dilation adds a pixels to the boundaries of an object in an image, while erosion removes the pixels on an object boundaries. The number of pixels can be added or removed from the objects in an image that depends on the size and shape of the structuring element in order to process the image. The morphological opening operation is erosion followed by dilation, using the similar structuring element for both operations.

Feature Extraction

In the pattern recognition applications and image processing applications, feature extraction is a particular form of dimensionality reduction.

When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (e.g. the same measurement in both feet and meters) then the input data will be transformed into a reduced representation set of features (also named feature vector). Transform the input data into the set of features is called *feature extraction*. If the features are extracted carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing the analysis of complex data one of the major problems stems from the number of

variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is the general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

Gray-level co-occurrence matrix

One of the simplest approaches for describing texture is to use statistical moments of the intensity histogram of an image or region . Using only histograms in calculation will result in measures of texture that carry only information about distribution of intensities, but not about the relative position of the pixels with respect to each other in that texture. Using the statistical approach such as co-occurrence matrix will help to provide valuable information about the relative position of the neighboring pixels in an image.

Given an image *I,* of size *N×N,* the co-occurrence, matrix P can be defined as

$$P(i,j) = \sum_{x=1}^{N} \sum_{y=1}^{N} \begin{cases} 1 & if I(x,y) = i \ and I(x+\Delta x, y+\Delta y) = j \\ 0 & otherwise \end{cases}$$

Here, the offset $(\Delta x, \Delta y)$, is specifying the distance between the pixel-of-interest and its neighbor.

Features	Explanation	Formula
Contrast	Returns a measure of the intensity contrast between a pixel and its neighbor over the whole image	$\sum |i-j|^2 . p(i,j)$
Energy	Returns the sum of squared elements in the GLCM	$\sum p(i,j)^2$
Correlation	Returns a measure of how correlated a pixel is to its neighbor over the whole image.	$\sum (i-\mu i)(j\mu j)$ $.p(i,j)/[\sigma i.\sigma j]$

Table: GLCM Features

Note that the offset $(\Delta x, \Delta y)$ parameterization makes the co-occurrence matrix sensitive to rotation. Choosing an offset vector, such that the rotation of the image is not equal to 180 degrees, will result in a different co-occurrence matrix for the same (rotated) image. This can be avoided by forming the co-occurrence matrix using a set of offsets sweeping through 180 degrees at the same

distance parameter Δ to achieve a degree of rotational invariance (i.e., [0 Δ] for 0° : **P** horizontal, [-Δ, Δ] for 45° : **P** right diagonal, [-Δ 0] for 90° : **P** vertical, and [-Δ −Δ] for 135° : **P** left diagonal).

Local Binary Pattern

Local Binary Pattern (LBP) is the simple and very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of every pixel and considers the result as a binary number. owing to its discriminative power and computational simplicity, LBP texture operator has become the popular approach in a range of applications. It can be seen as the unifying approach to the traditionally divergent statistical and the structural models of texture analysis. Perhaps the important property of the LBP operator in real-world applications is its robustness to monotonic gray-scale changes caused, for case, by illumination variations. Another main property is its computational simplicity, which makes it is possible to analyze images in challenging real-time settings.

The LBP value (LBP pattern, LBP code) of a pixel captures the structure of local brightness variations around it. Algorithmically, the value is computed by sampling circularly around the selected pixel and setting 1-bits in the LBP value for each sample that is brighter than the center.

$$LBP_{P,R} = \sum_{p=1}^{P} 2^{(p-1)} f_1 \left(g_p - g_c \right)$$

The LBP operator is unaffected by any monotonic gray-scale transformation which preserves the pixel intensity order in a local neighborhood. Note that each bit of the LBP code has the same significance level and that two successive bit values may have a totally different meaning. Really, The LBP code may be interpreted as a kernel structure index.

83	75	126
99	95	141
91	91	100

0	0	1
1		1
0	0	1
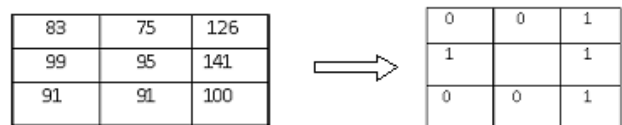
Binary: 00111001 decimal: 57

Gray Level features

These features are based on the differences between the gray-level in the candidate pixel and a statistical value representative of its surroundings.

Let I_H be the preprocessed image.

(x ,y) be the pixel in I_H.

S be the sub-image of I_H.

(s ,t) be the pixel in S

For the gray-level based feature extraction, 9*9 center point window is used.

$$F_1(x,y)= I_H(x,y) - \min\{I_H(s,t)\}$$
$$F_2(x,y)= \max\{I_H(s,t)\} - I_H(x,y)$$
$$F_3(x,y)= I_H(x,y) - \text{mean}\{I_H(s,t)\}$$
$$F_4(x,y)= \text{std}\{I_H(s,t)\}$$
$$F_5(x,y)= I_H(x,y)$$

Fuzzy logic

The main purposes of the recommended model are to diagnose the cancer diseases by using fuzzy rules with relatively small number of linguistic labels, reduce the similarity of the membership functions and preserve the meaning of the linguistic labels.

Modified fuzzy c-means algorithm (MFCM). The standard fuzzy c-means has various well-known problems, namely the number of the clusters must be specified in advanced, the output membership functions have high similarity, and FCM is unsupervised method and cannot preserve the meaning of the linguistic labels. On the contrary, the grid partitions method solves some of the previous matters, but it has very high number of the output clusters. The basic idea of the suggested MFCM algorithm is to combine the advantages of the two methods, such that, if more than one cluster's center exist in one partition then merge them and calculate the membership values again, but if there is no cluster's center in a partition then delete it and redefined the other clusters. Algorithm that illustrates the modified fuzzy c-means algorithm

Support vector machine

In machine learning, **support vector machine**(**SVMs**, also **support vector networks**) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, that used for classification and regression analysis. The basic SVM take a set of input data and predicts, for every known input, which of two possible classes forms the output, building it a non-probabilistic binary linear classifier. Given a set of training example, each noticeable data as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or other. An SVM model is a demonstration of the examples as points in space, mapped so that the examples of the split categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that similar space and predicted to belong to a category based on which side of the gap they fall on.

In addition to perform linear classification, SVMs can proficiently perform non-linear classification using what is called the kernel trick, implicitly map their inputs into high-dimensional feature spaces.

More formally, a support vector machine construct a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved through the hyper plane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

Fig: lung ct image   Fig: Preprocessed Image

Fig : Local Binary Pattern

IV.   RESULTS AND DISCUSSION

The experiments are conducted on the proposed computer-aided diagnosis systems with the help of real time lung images. This conducting tests data consists of 1000 lung images. These 1000 lung images are passed to the proposed system. The diagnosis rules are generated from those images and these rules are passed to the classifier for the learning process. After that, a lung image is passed to the proposed system. Then the proposed system will route through its processing steps and finally it will detect whether the supplied lung image is with cancer or not.

In view of the results obtained by the proposed CAD system, user has achieved the following. On one hand, user have developed an automatic CAD system for early detection of lung cancer using chest CT images in which a high level of sensitivity has been achieved, with a reasonable amount of false positives per image, (90% sensitivity with 0.05 false positives per image). This prevents the system from hindering the radiologist's diagnosis.

Image	Value	Classification
Image 1	0	Benign
Image2	1	Malignant
Image 3	1	Malignant
Image 4	1	Malignant
Image 5	0	Benign
Image 6	1	Malignant
Image 7	0	Benign
Image 8	1	Malignant

Table: Tumor Classification Using SVM

V. CONCLUSION

Proposed system helps physician to extract he tumor region and evaluate whether the tumor is benign or malignant. The computer tomography image is used in this paper. SVM provides the accuracy of 92.5%.The accuracy of system can be improved if training is performed by using a very large image database. The different basic image processing techniques are used for prediction purpose. In the first phase the image is denoised. In second phase lung region is separated from surrounding anatomy. In next phase ROI is extracted, after ROI extraction features extraction is performed by GLCM. Finally with the obtained texture features classification is performed to detect the occurrence

Reference:

[1] K. Suzuki, I. Horiba, N. Sugie, and M. Nanki(Nov. 1998), 'Noise reduction of medical X-ray image sequences using a neural filter with spatiotemporal inputs', in *Proc Int. Symp. Noise Reduction for Image .and Comm. Systems*, pp. 85–90

[2] S.G. Armato III, and H. MacMahon(2003), 'Automated lung segmentation and computer-aided diagnosis for thoracic CT scans,'*International Congress Series* pp. 977-982

[3] S. Edge, D. Byrd, C. Compton, A. Fritz, F. Greene, and A. Trotti, Eds(2010), 'AJCC Cancer Staging Handbook', 7th ed. New York: Springer

[4] I. Kapetanovic, S. Rosenfeld and G. Izmirlian(2004),'Overview of commonly used bioinformatics methods and their applications', *Ann N Y AcadSci*1020**,**10–21

[5] M. Kakar and D. R. Olsen, 'Automatic segmentation and recognition of lungs and lesions from CT scans of thorax,' Comput. Med. Imag

[6] S. G. Armato, M. L. Giger and H. MacMahon(2001),'Automated detection of lung nodules in CT scans: Preliminary results', Med. Phys., Vol.28

[7]J. Kuhnigk, V. Dicken, L. Bornemann, A. Bakai, D. Wormanns, S. Krass, and H. Peitgen(Apr.2006), "Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans," *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 417–434,

[8] W. Wever, S. Stroobants, J. Coolen, and J. Verschakelen(2009), 'Integrated PET/CT in the staging of nonsmall cell lung cancer: Technical aspects and clinical integration,'*Eur. Respir. J.*, vol. 33, pp. 201–212

[9] J. Lafferty, A. McCallum, and F. Pereira(2001), 'Conditional random fields: Probabilistic models for segmenting and labeling sequence data,'in*Proc. ICML*, , pp. 282–289

[10] A. Bardu, M. Suehling, X. Xu, D. Liu, S. Zhou, and D. Comaniciu(, 2010), 'Automatic detection and segmentation of axillary lymph nodes', in *MICCAI 2010, LNCS*, vol. 6361, pp. 28–36

[11] Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng(, 2011), "Discriminative pathological context detection in thoracic images based on multi-level inference," in *MICCAI 2011, LNCS*, vol. 6893, pp. 185–192

[12] YBoykov, O. Veksler, and R. Zabih, 'Efficient approximate energy minimization via graph cuts',*IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1222–1239, Dec. 2001

[13] Rezaul.K.Begg, MarimuthuPalaniswami and Brendan Owen (2005),'Support Vector Machines for Automated Gait Classification', IEEE Transactions on Biomedical Engineering, Vol.52,No.5

[14] Laura Auria1 and Rouslan A. Moro, "Support Vector Machines (SVM) as a Technique for Solvency Analysis", Berlin August 2008

[15] Armato SG III, Altman MB, LaRivière(2003) PJ. Automated detection of lung nodules in CT scans: effect of image reconstruction algorithm. Med Phys; 30:461–472.

[16] Weir HK et. al.,"Annual report to the nation on the status of cancer, 1975-2007"*Journal National Cancer Institute*, vol. 95 , No. 17.

[17] R.C. Gonzales and R.E. Woods, *Digital Image Processing*: Pearson Education, 2003.