Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

Classification of Lung Cancer Nodules using a Hybrid Approach

Varalakshmi.K
Department of Computer Science and Engineering SRM University
varalakshmi.kk@gmail.com

ABSTRACT

Medical Imaging plays an important role in the early detection and treatment of cancer. Computer Aided Diagnosis (CAD) system allows detection of lung cancer through analysis of chest CT images. Two problems have been focused; one is segmentation of organ of interest, which in case of Lungs is already a challenge. Second is classification, in which nodule features (like geometric properties, image intensity, shape and size) have to be taken into consideration. The objective of this study is identifying all nodules from the chest CT lung images and classifying these nodules into cancerous (Malignant) and non-cancerous (Benign) nodules, to reduce the false positive rate using Image processing techniques and Neural Network techniques. First, noise is removed from the image and then converted it to binary format. Then morphological operations are performed to extract the lung field. Features are extracted and these are fed to the neuro-fuzzy system for identifying true nodules.

Keywords: *Neural Networks, Fuzzy, CT image, segmentation, classification*

1. INTRODUCTION

Lung cancer is the most serious cancers in the world [1]. Of all types of cancer, lung cancer is the most common cause of death and accounts for about 28% of all cancer deaths. Overall, an estimated 12.7 million cancer deaths occurred in 2008, with 56% of new cancer cases and 63% of the cancer deaths occurring in the less developed regions of the world. The most commonly diagnosed cancers worldwide are lung (1.61 million, 12.7% deaths), breast (1.38 million, 10.9% deaths), colorectal cancers (1.23 million, 9.7% deaths).The most common cancer deaths are lung cancer (1.38 million, 18.2% of the total), stomach cancer (738, 000, 9.7%) and liver cancer (696,000 deaths, 9.2%).

Lung cancer most often occurs between ages 40-70, peaks in the 50-60 age range. Accounts that 40% patients of lung cancer is less than 50 years of age and 11% is less than 40 yrs of age and in younger age it is misdiagnosed as tuberculosis. The main cause of lung cancer is known to all- Smoking.

Fig 1: Original Lung CT image

1.1 Types of Lung Cancer

Generally Lung cancer is of Two types, i) Small Cell Lung Cancer ,ii) Non-Small Cell Lung cancer.In India,common histological type of lung cancer is squamous cell carninoma in both males and females.

a. Small Cell Lung Cancer

Small cell lung cancer is generally faster growing than non-small cell, but more likely to respond to chemotherapy. Small cell cancer is divided into "limited stage" (generally cancer confined to the chest) and "extensive stage" (cancer that has spread outside the chest).About 20% of all lung cancer cases are small cell lung cancer, meaning about 30,000 patients each year are diagnosed with this disease.At the time of diagnosis, approximately 30% of patients with small cell carcinoma will have tumor confined to the hemithorax of origin, the mediastinum, or the supraclavicular lymph nodes.

Table I: Types Of Lung Cancer

Non- Small Cell Lung Cancer	Small Cell Lung Cancer
Adenocarcinoma Squamous Cell Carcinoma Large Cell Carcinoma	Oat Cell Intermediate Combined

b. Non Small Cell Lung Cancer

Non-small cell lung cancer constitutes 75-80% of lung cancers. More than 70% of non small cell lung cancers are in stages III and IV.There are three types of NSCLC lung cancer. They behave in similar ways and respond to treatment differently than SCLC.

Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

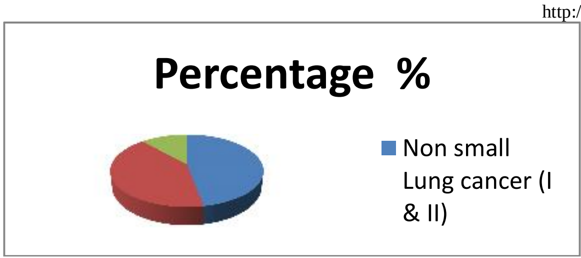


Fig 2: Distribution of Lung cancer

Adenocarcinoma: Adenocarcinoma is the most common type of NSCLC and the most common form of lung cancer among women. The incidence of this type of lung cancer is increasing. Adenocarcinoma often appears toward the outer edges of the lungs in the mucous glands that line the airways. Like other forms of lung cancer, adenocarcinoma may spread to other parts of the body.

Squamous Cell Carcinoma: The most common lung cancer in men is squamous cell carcinoma. It usually appears in the larger breathing tubes. Like other NSCLCs, squamous carcinoma is a relatively slow-growing cancer.

Large Cell Carcinoma: Large cell carcinoma occurs less often and has larger cells than other NSCLCs. It usually first appears in the smaller breathing tubes and may spread quickly. Large cell carcinoma is diagnosed after other types of lung cancer are ruled out. Large cell carcinoma tends to grow quickly and spread (metstasize) at an earlier stage than other forms of non-small-cell lung cancer.

1.2 Computed Tomography (CT)

Computer tomography (CT), sometimes called a CAT scan, uses special X-ray equipment to obtain image data from different angles around the body, and then uses computer processing of the information to show a cross-section of body tissues and organs. CT imaging can show several types of tissue—lung, bone, soft tissue, and blood vessels—with great clarity. Using specialized equipment and expertise to create and interpret CT scans of the body, radiologists can more easily diagnose cancer problems. The image allows a physician to confirm the presence of a tumor and to measure its size, precise location, and the extent of the tumor's involvement with other nearby tissue. The CT scan can reveal some soft-tissue and other structures that cannot even be seen in conventional X-rays. Using the same dosage of radiation as that of an ordinary X-ray machine, a CT scan can increase the clarity by about 100 times for an entire slice of the body. The CT scan is often performed after a chest X-ray.

1.3 Computer Aided Diagnostic (CAD)

Early detection of lung cancer is critical for improving survival of this disease because only 15% of lung cancers are found when they are localized. Since there are few or no symptoms in the early stages of the disease, the majority of lung cancers are diagnosed in the late stages of the disease. Symptoms of later-stage disease may include a persistent cough, sputum streaked with blood, chest pain, voice change, and recurrent pneumonia or bronchitis. The goal of a screening program is to find cancer at an early stage when there are fewer symptoms. Treatment at early stages of cancer can lead to more treatment options, less invasive surgery, and a higher survival rate.

Improving the ability to identify early-stage tumors is an important goal for physicians, because early detection of lung cancer is a key factor in producing successful treatments. Computer-aided diagnosis (CAD) involves the use of computers to bring suspicious areas on a medical image to a radiologist’s attention. CAD for cancer detection in medical images starts with a digital image. Here CT lung image is used to detect the lung cancer nodules. CAD System consists of the following stages,

i. Image Pre-processing – It not only removes the unwanted noise and background information but also enhances the image quality.
ii. Image Segmentation – Segmentation is the process of separating the lung lobes from CT image and it classifies the suspicious regions.
iii. Feature Extraction – After segregating lung lobes from lung CT image, based on the features such as area, perimeter, shape, size, color, etc. positive region will be extracted.
iv. Classification – After identifying the cancer nodules, the CAD system classifies the malignant nodules.

2. LITERATURE SURVEY

Detection of suspicious lesions in the early stages of cancer can be considered the most effective way to improve survival. To cope with this serious problem, simple X-rays have performed mass screening for lung cancer. However, it is known that the accuracy of this method is not sufficient for early detection of lung cancer [2], [4].

Thoracic Computed Tomography (CT) can display tiny nodule which may not be seen on chest radiographs, thus improves the diagnosis accuracy. Furthermore, CT plays an important role in lung cancer staging evaluation. Nodule detection on CT is further more challenging task due to low contrast, size and location variation. Especially in the central lung regions, lung nodules can go undetected because they are confused with the blood vessels imaged in cross section. So distinguishing the cancerous nodules from the blood vessels is the challenging task to perform.

Approaches of the detection of lung nodule have been reported such as feature based, neural based and template matching classification. In [5], used fuzzy clustering method to extract the organs of interest and the lung area was classified into two clusters, air cluster and other organs of cluster. Then they used similar features and rule based approach to classify the nodules and blood vessels. In [6], proposed a template-matching algorithm

Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

for the detection of nodules in chest helical CT images using nodular models with Gaussian distribution as reference images.

Accurate lung segmentation allows for the detection and quantification of abnormalities within the lungs. Van Rikxoort et al.[7] presented a fully automatic methods for segmentation of the lungs and lobes from thorax CT scans. Region growing approach is used to segment the images and then morphological operations are applied to smoothen the image. Multi-atlas segmentation is applied to detect the errors automatically in the resulting lung segmentation. 55 volumetric chest scan images are used here and time taken to perform the segmentation is 20 seconds per image.

In[9], Sobel Edge detection method is used to segment the CT lung image. Jun Lai et al.[8] proposed a fully automatic segmentation for pulmonary vessels in plain thoracic CT images. The frame size of CT image is 360x360. Fractional differential operator is used to enhance the quality of an image and OSTU thresholding method is used for segmentation. Vessel tree reconstruction algorithm [10] reduced the number of false positives generated by an existing nodule detection algorithm by 38%.

Tao Xu et al.[1] proposed an automatic global edge and region force (ERF) field guided method with non-linear exponential point evolution for lung field segmentation, by introducing global edge and force field information together with a new point evolution technique. Experimental results demonstrated that the proposed method is time efficient and improves the accuracy, sensitivity, specificity and robustness of the segmentation results, compared to the typical ASM and hybrid LSSP. Automatic initialization also has better performance than the original PIG based initialization. Although the experiments on chest radiographs, the proposed ERF technique can easily be adapted to other image segmentation applications

In this research, proposed a neuro-fuzzy approach is used to detect pulmonary nodules automatically in chest CT images. First, raw image is converted to gray scale image. Noise which is presented in an CT image will be removed. Then, lung tissue will be extracted from the original image. Classification is done based on the features such as shape, size, mean, intensity. Back propagation algorithm is used to discriminate the benign and malignant nodules.

3. Proposed System

System architecture of CAD for detecting the nodules at CT lung image is shown in figure-3 .First the original lung image is enhanced by the un sharp masking. Then, using a threshold technique the lungs were segmented. Using Morphological operations such as filling, closing the lung field is extracted from the original image. This is because to avoid the detection of nodules at the outermost portion of the lung image. Finally, various image features of each potential nodule were assessed to separate true nodules from false positive nodules. An artificial neural network was applied to determine the likelihood of the lesion being a true nodule on the basis of the image features.

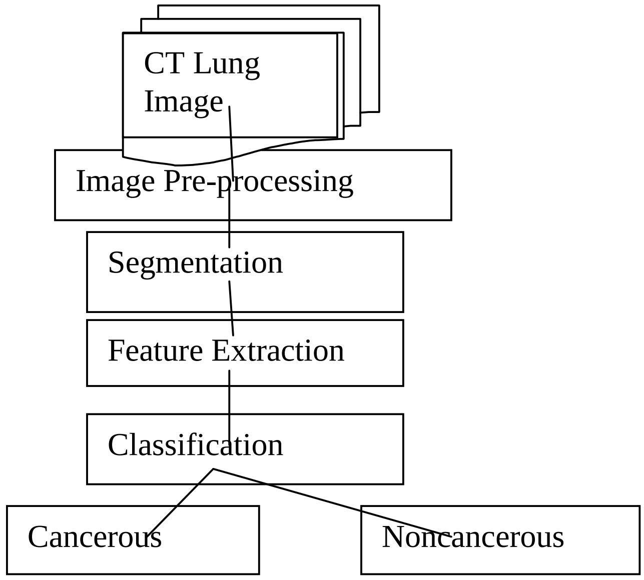


Fig 3: CAD System Architecture

3.1 Image Pre-processing

Chest radiographs inherently display a wide dynamic range of CT image intensities. In conventional, unprocessed images it is hard to "see through" the mediastinum and contrast in the lung fields is limited. A classical solution to this kind of problem in image processing is the use of histogram equalization techniques. A related technique is enhancement of high-frequency details (sharpening).

Here unsharp masking approach is used for enhancing images. The input image is passed through a block that extracts edges and features. The output is then scaled by an appropriate factor and added back to the original image. The edge extraction is often implemented as a linear high pass filter such as a discrete linear Laplacian operator.

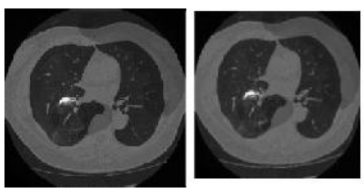


Fig 4: Image pre-processing a) Original Image b) Enhanced Image

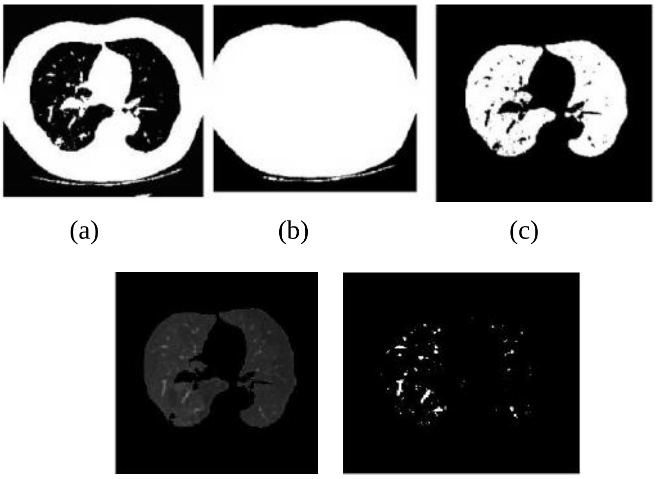
3.2 Segmentation

The second step in this CAD system is to identify the lung boundary, thereby removing other thoracic structures from the nodule candidate region. This process is typically accomplished using 2D gray level

Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

threshold techniques on each CT images. We have developed segmentation method based on threshold. A binary image is then constructed such that all pixels with the corresponding gray level greater than the selected threshold are 1 and all other pixels are pixels are zero. Then another filling operations automatically determines which pixels are in holes, and then changes the value of those pixels from 0 to 1 and fills the holes in that binary image. Then edge of the binary image is detected and which is subtracted from the filled image then we get the lung field.



(a) (b) (c)

(d) (e)

Fig 5: Segmentation a)Binary Image b)Image filling c)Image subtraction d) Lung Tissue e) Region of Interest

3.3 Feature Extraction

After extracting the region of interest, determine whether all the extracted regions are nodules or not. We have to analyze the features of each nodule to distinguish true nodules from the false positive nodules. Many features, such as size of area, circularity, mean value, variance, location, or gradient, etc, are usually applied in detection of medical images. We analyzed these features on CT images and found that five features, size of area, circularity, mean, skewness and kurtosis can be used as clues for discrimination explicitly.

Circularity is defined to be the ratio of the area of an object to the area of the circle with the same convex perimeter (4 pi *area/ convex perimeter).

3.4 Classification

The final procedure of the proposed system is to confirm the suspicious region and determine if it is a true nodule utilizing features obtained from previous stages. Neural networks and

Fuzzy logic have some common features such as distributed representation of knowledge, model-free estimation, ability to handle data with uncertainty and imprecision etc. Fuzzy logic has tolerance for imprecision of data, while neural networks have tolerance for noisy data. A neural network's learning capability provides a good way to adjust expert's knowledge and it automatically generates additional fuzzy rules and membership functions to meet certain specifications. This reduces the design time and cost. On the other hand, the fuzzy logic approach possibly enhances the generalization capability of a neural network by providing more reliable output when extrapolation is needed beyond the limits of the training data.

This model is organized layer by layer including input layer, fuzzification layer, rule inference layer, and defuzzification layer. The function of each layer is illustrated in the following sections.

a. Input Layer

We applied the features, size of area (A,), mean (M), kurtosis (K), skew ness (S) and circularity(C,), to the input layer. The input nodes, I1,I2,I3,I4 and I5, are values of A,M, K,S and C, calculated from sample image X respectively, and are normalized to the range between -1 and 1 to speed up the training process. These features are defined as following equations:

$$\text{Area (A)} = P_{roi}$$

$$\text{Mean (M)} = 1/N \sum_{(i,j)€ R} g(i,j)$$

$$\text{Skew ness (S)} = \frac{1/N \sum_{(i,j)€ R} [g(i,j) - \text{mean}]^3}{[\sqrt{\sum_{(i,j)€ R} [g(i,j)]^2 - [\text{mean}]^2}\,]^3}$$

$$\text{Kurtosis (K)} = \frac{1/N \sum (i,j)€ R\ [g(i,j) - \text{mean}}{[\sqrt{\sum (i,j)€ R\ [g(i,j)]^2 - [\text{mean}]^2}\,]^4}$$

$$\text{Circularity (C)} = Ix/A$$

where P_{roi} represents the number of pixels in the ROI. P(x, y) denotes the gray value of the pixel at position (x, y), I, is the size of area of the region A, within the circular Ix.

b. Fuzzification layer

In this layer, fourteen membership functions are set up for five input values. Both size of area A, and circularity C, have three membership functions, low, medium, and high since the nodules with low and medium size of area might possess high circularity and those with large size of area might possess high or median circularity. M has two membership functions, low and high, because the mean brightness of hue nodules may be larger than certain value. We propose that high membership function is formed by one positive sigmoid function, and median membership function is formed by the summation of two sigmoid functions, one positive and one negative. The symbol f, is a sigmoid function defined as

$$a = - Ii * W^{g}ji$$

$$f = 1/(1+\exp(a))$$

where Ii is the normalized feature value of either A,M,K,S or C, and W_{ij} is the weight parameter on the connection between input i and sigmoid function j , j = 1, . . . ,10.

c. Rule inference layer

After obtaining membership functions from the previous fuzzification layer, at most one hundred and sixty two inference rules can be generated by one of the membership values from each feature (3x3x3x3x2=18). For each rule, the minimum membership value is the output of the inference rule. The equation is as follows:

Rule k= min (f1, f2, f3, f4, f5)

Where f1, f2, f3, f4 and f5 is one of the membership value of A, M, S, K and C connected to the kth rule.

d. Defuzzification layer

In the defuzzification step, a weight parameter W fj is assigned for each rule. The weight (Wfj) is random values initially and is determined by back-propagation learning algorithm. Then we take the sum of product of each rule's output and their associate weights to compute the defuzzification output through sigmoid function (f(Ox)). The equation is computed as follows:

$$Ox = {}_{k=1\ \text{to}\ 162}\sum \text{Rule}_k\, W_k^{\ f}$$

$$F(Ox) = 1/(1+\exp(-Ox))$$

4. EXPERIMENTAL RESULTS

The CAD is developed and is applied to 15 clinical cases (512x512 pixels size). Totally 527 nodules were present in various sizes and the proposed system is able to detect 351 nodules and encountered 176 false positives in these cases. The size of nodules can detect by this system is between 3mm and 5cm in diameter within lung field. Most of the missed nodules are larger than 5cm in diameter and located on the lung wall. The overall performance of the proposed system is 89.3% correct detection and the false positive per slice image is as low as 0.3. This result shows that improvement of the detection rate with very low false positives.

5. CONCLUSION

For automatic detection of lung nodules, proposed a hybrid approach called neuro- fuzzy algorithm .The segmentation is achieved through by a series techniques including thresholding, median filtering, closing, and labeling. 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 size of area, circularity, skewness, kurtosis and mean and then subjected to classification to classify the nodules. Neural fuzzy model was designed to extract suitable diagnosis rules, and classified the true nodules from the ROIs. The final results of experiments showed that our system can detect small lung nodule accurately as high as 89.3%. Meanwhile, the false positive per image was reduced as low as 0.3. This CAD system was not able to detect 1mm size of nodules and also some 3mm nodules were missed. To improve the efficiency of the CAD system to 100%, in future planned to work on game theoretic framework with Active shape modelling.

REFERENCES

[1] H.K.Weir,Annual report to the nation on the status of cancer, 1975-2000, Journal National Cancer Institute, vol.95,no.17,pp.1276-1299,2003.

[2] Tao Xua, Mrinal Mandala, Richard Longb, Irene Chengc, Anup Basuc, An edge-region force guided active shape approach for automatic lung field detection in chest radiographs, Computerized Medical Imaging and Graphics,2012.

[3] D.Behara, Lung Cancer in India, Medicine update 2012, Vol.22, (401-407).

[4] T.Tanaka, K.Yuta, and Y.Kobayashi, A Study of the false-negative case in mass screening of lung cancer," Jay.J Thor. Med., vol.43,pp. 832-838,1984.

[5] J.Oda S.Akita, and K.Shimada, A study of false-negative case in comparative reading of mass-screening of lung cancer, Lung Cancer,vol.29,pp.271-278,1989.

[6] K.Kanazawa, Y.Kawata, N.Niki, H.Satoh, H.Ohmatsu, R.Kakinuma,M.Kaneko,N.Moriyama, and K.Eguchi, Computer -aided diagnosis for pulmonary nodules based on helical CT images, Comp. Med. Imag.Grap., Vol.22, no.2, pp.157-167, 1998.

[7] Yon bum Lee, Takeshi Hara, Hiroshi Fujitha, Shigeki Itoh, and Takeo Ishigaki, Automated detection of pulmonary nodules in helical CT images based on an improved Template matching technique, IEEE transactions on medical imaging, vol.20, No. 7, July- 2001.

[8] Bram van Ginneken, Bart M. ter Haar Romeny, and Max A. Wergeyer, Computer-Aided Diagnosis in Chest Radiography: A Survey, IEEE Transactions on Medical Image, Vol. 20, pp. 1228-1241, Dec (2001).

[9] E.M.van Rikxoort and B.van Ginneken, Automatic Segmentation of the lungs and lobes from thoracic CT scans,Diagnostic image Analysis group,Department of Radiology,Netherlands,2011.

[10] Jun Lai and Mei Xie, Automatic segmentation for pulmonary vessels in plain thoracic CT scans,

TELKOMNIKA,vol.10,No.4,August 2012,pp. 743-751.

[11] Disha Sharma and Gagandeep Jindal, Computer aided diagnosis system for detection of lung cancer in CT scan images,International Journal of Computer and Electrical Engineering,vol.3,No.5,Oct 2011,pp.714-718.

[12] Weiner Filtering,"http://www.owlnet.rice.edu".

[13] Gady Agam, Samuel G.Armato,III,and Changhua Wu,Vessel Tree Reconstruction in thoracic CT scans with application to nodule detection,IEEE transactions on Medical Imaging,vol.24,No.4,April 2005,pp.486-499.

[14] J. Breckling, Ed., the Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.