Lung Cancer Diagnosis using Computer Aided Diagnosis System

Sarika Tale, PhD
Associate Proffesor
Digital Electronics and communication system
VIAT, Chickaballapura

Pushpek Kumar Singh
P.G Student
Digital Electronics and communication system
VIAT, Chickaballapura

ABSTRACT

Lung cancer has been the reason for fatality for many people in recent years. The manifestation of lung nodules is the prominent reason of lung cancer. Early detection of cancer facilitates early treatment which improves the chance of survival of patients. The most trivial way to discover lung cancer is by working with Computed Tomography (CT) image. Computer Aided Diagnosis (CAD) is a system that is designed by the integration of medical science and computers. A CAD system that is used for the diagnosis of lung cancer accepts lung CT images as input and depending on an algorithm helps doctors to implement image analysis. It includes three main steps to detect lung nodule: preprocessing, segmentation and classification of cancer nodule using support vector machine (SVM). With the aid of CAD, doctors can take the last word. Images include some redundant data and some feature that are critical for processing; pre-processing upgrade images by eliminating distortion and boost the important features. After pre-processing step the lung cancer nodule is drawn out. The obtained image through previous steps is used for training and finding the accuracy of the system in detecting cancer.

General Terms

Computer Aided Diagnosis (CAD), Computer Tomography (CT) images, image processing

Keywords

Segmentation, Dilation, Erosion, Support Vector Machine (SVM), Classifier.

1. INTRODUCTION

Lung cancer is the fundamental cause of cancer fatalities worldwide, and detection of cancer in its beginning stages is difficult because a sign of illness appears only in the later stages causing the death rate to be the maximum among all cancer types. It is the major cause of death compared to colon, and prostate cancers. There is a significant report indicating that the detection of cancer in early stages will decrease mortality rate. Thus the Computer Aided Diagnosis (CAD) system is very essential for early detection of lung cancer [4]. Radiologists use CAD systems for preprocessing of the images and finding the most probable locations for lung nodules. The first step in nodule detection is suppressing the background structure in lungs like ribs, bronchi and blood vessels. Then the nodules are classified based on size, contrast and shapes. Classification of nodules based on the above features produce a lot of false positives [3]. This paper proposes a CAD system which is fast and accurate for initial detection of lung cancer. The suggested system consist of some steps such as: collect lung CT scan image data,

preprocessing, segmentation, thresholding, erosion, dilation, creates a pattern from cancer tissue by using a bandpass filter, and classify images as lung cancer or healthy using support vector machine (SVM).

2. METHODOLGY

The proposed system consists of steps such as: collection of lung CT scan image, preprocessing, extraction of the lung region, feature extraction, create pattern from cancer tissue by using a band-pass filter, use pattern for training the cellular automata, classification of the images as lung cancer or healthy. Figure1 shows the overall architecture of CAD system:
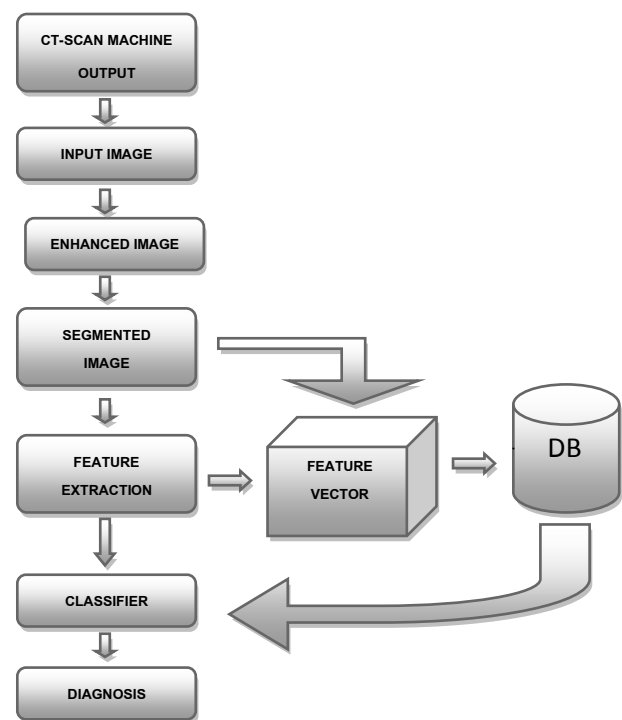

Fig: 1 Overall architecture

2.1 Database collection

The Database includes CT scan image of healthy and cancerous lung patient. Lung images were taken from the internet. Figure 2 shows the CT scan image of a normal and cancerous patient's chest.
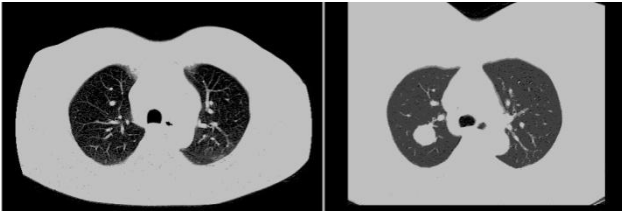
Fig: 2 CT scan image of healthy and cancerous lung patient

2.2 Image preprocessing

Image pre-processing is a process which eliminates image distortion and enhances important features. By using this process a corrected image produced. For applying preprocessing approach, MATLAB software was used.

2.3 Segmentation

Segmentation is an important step in image processing. By using segmentation, images are divided into regions. The purpose of segmentation is to change the image representation which makes analysis easy. We used segmentation methods in CT lung images to separate the lung tissues and bones in order to extract the region of interest (ROI) .

There are many methods for segmentation but thresholding is very useful among all. Thresholding has some benefits such as: fast processing, easy influence, resulting images do not keep weight space and another method is region growing. This algorithm starts from a pixel of an image and checks other pixels that are around start point. The non assigned pixel that are neighbors with region checked are similar to the region then the region growing frequently but if differences between neighbor pix and region is more than the threshold the process is stopped. [5] The region growing segmentation algorithm based thresholding is used on enhanced image.

2.4 Morphological image processing

Morphological image processing describes techniques that deal with the shape (or morphology) of features present in an image. Morphological operations are used to remove imperfections introduced during segmentation. Mathematical morphology is a tool of extracting image components that is useful in the representation and characterization of shape such as boundaries, skeletons etc. Goal of morphology operations are simplify image data, preserve essential shape characteristics and eliminate noise. There are two morphological operations dilation and erosion. Dilation expands the object, fill in small holes and connect disjoint objects. Erosion compresses the objects by eroding their boundaries. These operations can be custom built by the selection of the structuring element. The structuring element decides how the objects will be eroded or dilated [10] [11].

Notations:
Black pixel: value will be 0 in grayscale for an 8 bits/pixel indexed image

White pixel: value will be 255 in grayscale for an 8 bits/pixel indexed image

2.4.1 Dilation

Dilation is a process where the structuring element B is placed on the image A and it slides across the image similar to convolution. Steps involved in dilation:

1. If the white pixel in the image coincides with the origin of the structuring element then no need to change; check the next pixel.

2. If the black pixel in the image coincides with the origin of the structuring element then all the pixels covered by the structuring element are made black [11].

Notation: $A \oplus B$
The structuring element can take any shape.

2.4.2 Erosion

The erosion process is identical to dilation, but pixels are converted to 'white', not 'black'. Steps involved in erosion:

1. If the 'white' pixel in the image coincides with the origin of the structuring element then no need to change; check the next pixel.

2. If the 'black' pixel in the image coincides with the origin of the structuring element and at least one of the 'black' pixels falls over a white pixel then the 'black' pixel is converted to 'white' [11].

Notation: $A \ominus B$

2.4.3 Boundary Extraction

The boundary of a set A, denoted by β (A) is obtained by eroding A by B and then performing the differences between A and its erosion.

$$\beta(A) = A - (A \ominus B)$$

where B is a structuring element.

'–'is the difference operation on sets

2.5 Feature Extraction

Feature extraction is computer based systems which provide parameters. Computer system takes decision based on these parameters. After the features extraction, the diagnosis rules are designed to detect nodules in the lung. These features convey information regarding lung nodule. This information helps in distinguishing lung nodule as malignant or non-malignant. In this paper, the features extracted from the CT scan image are used as diagnostic indicators [13].

2.5.1 Wavelet Filter Banks

The discrete wavelet transform (DWT) is performed on the CT scan image to obtain the desired frequency component containing information about the lung nodule .Lung nodule details are present in the band pass frequency components. The two-dimensional discrete wavelet transforms (2D-DWT) uses iterative and recursive process to obtain spectral spatial frequency components. Figure 3 shows two-level decomposition. [14]. The image is decomposed into LL, LH, HL, and HH sub bands. A lung nodule characteristic lies between LL and HH sub band. The DWT decomposition is followed by a filtering operation with a sequence of band pass filters.
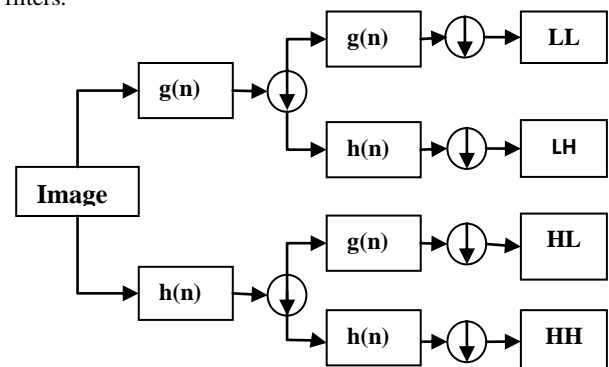

Fig: 3   2D -DWT Decomposition of image

2.5.2 Entropy

The randomness of a gray-level distribution is measured by entropy. Images with gray levels distributed randomly throughout tend to have high entropy. [12]

$$E = -\sum_{i}^{m}\sum_{j}^{n} P[i,j]\log P[i,j]$$

where P [i, j] is the probability of pixel intensity

The additional feature used as a classifier are:

2.5.3 Centroid (x-coordinate) and (y-coordinate)

It is a 1×2 vector which denotes the center of a mass nodule. The first element represent the horizontal coordinate (or x-coordinate), and the vertical coordinate (or y-coordinate) represent the second element [14].

2.5.4 Major and Minor Axis Length

The length (in pixels) of the major/minor axis of the ellipse with the same second moments as that of the mass nodule.

## 2.6 Morphological image processing

2.6.1 Support Vector Machine

The machine learning tool used for data classification is SVM. It constructs an N-dimensional hyper plane. The plane separates the data into two categories linear or non-linear. The training data is mapped into a kernel space .In non-linear cases mapping the given data into a different space are performed by SVM using kernel function. Kernel function acts upon the input data. The final classification is the result of final summation with an activation function. Kernel function are of different types which include polynomial, RBF, quadratic, Multi Layer Perceptron (MLP). The formulation of each kernel is performed by its own parameters like γ, σ, etc. The performance rate of the SVM is measured by varying above parameter.Table1 shows the comparison of various features present in the image of a cancerous and non cancerous patient.[15]

3. RESULTS AND DISCUSSION

The proposed computer aided diagnosis methodology is applied on images available in database. The database contains 20 cancerous patient images and 7 non cancerous patient images. These images are collected from the internet. The variation among the eight filter bank's energy for a malignant and non carcinogenic patient is indicated in graph1.

**Graph1: Variation in filter bank energy**
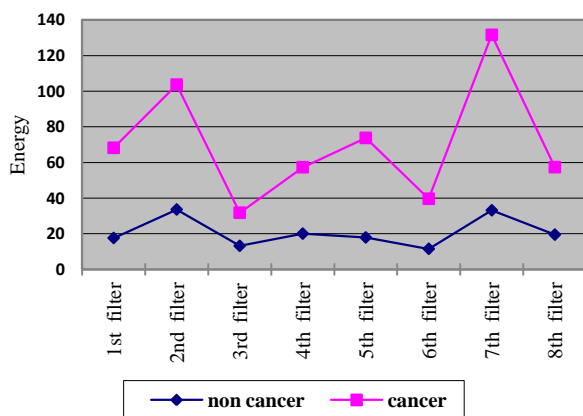
The features extracted from the image as shown in table 1 are given to the classifier for training purpose.

Table 1 Comparative analysis of various features in the image of a cancer and non cancer patient
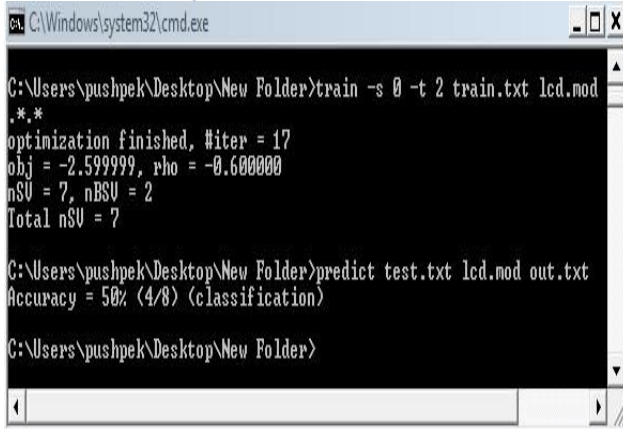
Classifier features	Cancer	Non – cancer
First filter bank output	68.235	17.6095
Second filter bank output	103.6555	33.5986
Third filter bank output	31.7993	13.1944
Fourth filter bank output	57.3288	20.062
Fifth filter bank output	73.7525	17.9108
Sixth filter bank output	39.6257	11.5385
Seventh filter bank output	131.591	33.1546
Eighth filter bank output	57.3699	19.5063
Global energy	212207.741	20272.5
Entropy	0.4301	0.8952
Centroid (x-coordinate)	790.4363	127.3557
Centroid (y-coordinate)	548.1718	92.1718
Major Axis Length	791.8709	228.4468
Minor Axis Length	429.7245	122.4592

The correlation between other feature extraction parameters is depicted in graph 2 for a healthy and malignancy influenced patient.

Graph 2: Correlation between other feature extraction parameters
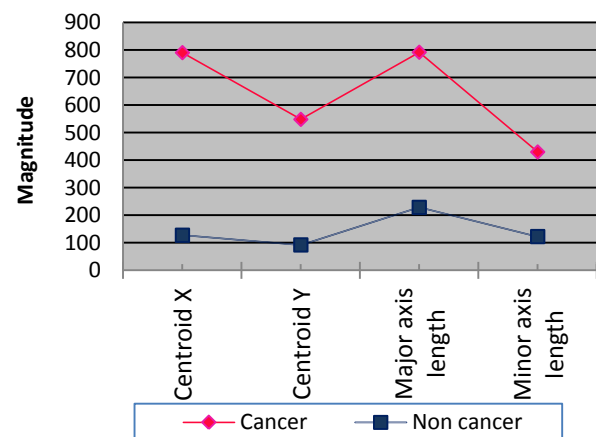
The diagnosis rules are then formulated from those images. These rules are used by the Support Vector Machine (SVM) for the learning process. The proposed CAD system will detect whether the supplied lung image is a cancer or not. The system is trained using 27 CT scan images and when it is tested using 8 CT Scan images the accuracy is found to be 50% as shown in Figure4.

Fig: 4 Output of CAD showing accuracy in detection

4. CONCLUSION

The Computer Aided Diagnosing (CAD) system process starts with detection of lung region by using basic image processing techniques like Binary to gray conversion,Denoising,segmentation,Thresholding,Erosion,Dilation,Lung Boundary Extraction to the CT scan images. After the detection of lung region the segmentation is carried out. With these segmented images 14 features are extracted such as 8wavelet filter bank energy, centroid(x and y coordinate), Major axis length, minor axis length, global energy and entropy. The diagnosis rules are formulated using these features. These rules and Support Vector Machine (SVM) are used for learning. The experimentation is performed with 27 CT scan images obtained from the reputed hospital. The result shows that the proposed CAD system is capable to detect the false positive nodules correctly. The usage of Support Vector Machine increase the accuracy of classification the cancer nodules. The accuracy of classifier can be enhanced by training the classifier with more features and databases of images.

5. FUTURE SCOPE

The wavelet and novel features extraction technique shows promising result. The system diagnosis capability can be enhanced by archiving images and patient records. The archives should be available for free to the trained engineers and open source communities. The trained doctors and engineers working together need discover some new features for better classification and prognosis of cancer in early stages. The above techniques can be applied on images from X-ray or MRI for comparison of result. The images which give better result should be considered. The full automation of the system can be achieved by integrating diagnosis of cancer stage with best possible oncological treatment.

6. ACKNOWLEDGMENT

7. REFERENCES

[1] Firmino, Macedo et al. "Computer-Aided Detection System for Lung Cancer in Computed Tomography Scans: Review and Future Prospects." BioMedical Engineering OnLine 13 (2014): 41. PMC. Web. 8 Aug. 2015.

[2] F. Qadir, M. Peer, and K. Khan, "Efficient edge detection methods for diagnosis of lung cancer based on two-dimensional cellular automata," Advances in Applied Science Research, vol. 3, pp.2050-2058, 2012.

[3] W. Yongjun, W. Na, Z. Hongsheng, Q. Lijuan, Y. Zhen, and W.Yiming, "Application of artificial neural networks in the diagnosis of lung cancer by computed tomography," Sixth International Conference on Natural Computation (ICNC), pp. 147-153, 2010.

[4] M. Gomathi and P. Thangaraj, "A computer aided diagnosis system for lung cancer detection using support vector machine," American Journal of Applied Sciences, vol. 7, p. 1532, 2010.

[5] Lung Cancer Diagnosis Using CT-Scan Images Based on Cellular Learning Automata Nooshin Hadavi[1], Md.Jan Nordin[2], Ali Shojaeipour[3]Center for Artificial Intelligence Technology (CAIT) Faculty of Information Science & Technology Universiti Kebangsaan Malaysia 43600, Bangi, Malaysian.hadavi@ftsm.ukm.my[1], jan@ftsm.ukm.my[2], a.shojaeipour@ftsm.ukm.my[3]

[6] A. Chaudhary and S. S. Singh, "Lung Cancer Detection on CT Images by Using Image Processing," in Computing Sciences(ICCS), International Conference on, pp. 142-146, 2012.

[7] S. S. Parveen and C. Kavitha, "A reviw on computer aided detection and diagnosis of lung cancer nodules," International Journal of Computers & Technology, vol. 3, pp. 393-400, 2012.L. B. Lusted, "Logical analysis in roentgen diagnosis," Radiology,vol. 74, pp. 178-93,1960.

[8] N. Asada, K. Doi, H. MacMahon, S. Montner, M. Giger, C. Abe, et al., "Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: pilot study," Radiology, vol. 177, pp. 857-860, 1990.J. Shiraishi, Q. Li, D. Appelbaum, and K. Doi

[9] L. B. Lusted, "Logical analysis in roentgen diagnosis," Radiology,vol. 74, pp. 178-93,1960. N. Asada, K. Doi, H. MacMahon, S. Montner, M. Giger, C. Abe, etal., "Potential usefulness of an artificial neural network fordifferential diagnosis of interstitial lung diseases: pilot study,"Radiology, vol. 177, pp. 857-860, 1990.

[10] Edge Improvement of Clustered Soybean Seeds by using Sobel Filter Sukhjinder Kaur Archna Mahajan Astt. Prof. (ECE) M.Tech (ECE) SSIET, Dera Bassi SSIET

[11] Sukhjinder Kaur, Archna Mahajan, "Edge Improvement of Clustered Soybean Seeds by using Sobel Filter" International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014 1614

[12] Chapter 7 - Morphological operation on binary images: users.utcluj.ro/~rdanescu/PI-L7e.pdf

[13] A Computer Based Feature Extraction of Lung Nodule in Chest X-Ray ImageNitin S. Lingayat and Manoj R. Tarambale.

[14] Detection of Lung Nodule Using Multiscale Wavelets and Support Vector Machine. K.P.Aarthy, U.S.Ragupathy

[15] Jeoung Hyun Kim et al. "Computer-aided detection system for masses in automated whole breast ultrasonography: development and evaluation of the effectiveness" Ultrasonography 2014; 33(2): 105-115.

[16] Segmentation of Lungs from CT Scan Images forEarly Diagnosis of Lung CancerNisar Ahmed Memon, Anwar Majid Mirza, and S.A.M. Gilani

[17] N. Asada, K. Doi, H. MacMahon, S. Montner, M. Giger, C. Abe, etal., "Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: pilot study,"Radiology, vol. 177, pp. 857-860, 1990.

[18] J. Shiraishi, Q. Li, D. Appelbaum, and K. Doi, "Computer-Aided Diagnosis and Artificial Intelligence in Clinical Imaging,"Seminars in Nuclear Medicine, vol. 41, pp. 449-462, 2011.

[19] J.-C. Chen, C.-M. Yeh, and J.-E. Tzeng, "Pattern differentiation of glandular cancerous cells and normal cells with cellular automata and evolutionary learning," Expert Systems with Applications, vol.34, pp. 337-346, 2008.

[20] M. S. AL-TARAWNEH, "Lung Cancer Detection Using Image Processing Techniques," Leonardo Electronic Journal of Practices and Technologies, pp. 147-158, 2012.

[21] D. Sharma and G. Jindal, "Identifying Lung Cancer Using Image Processing Techniques," International Conference on Computational Techniques and Artificial Intelligence (ICCTAI)

[22] N. Mesanovic, M. Grgic, H. Huseinagic, M. Males, E. Skejic, and M. Smajlovic, "Automatic CT Image Segmentation of the Lungs with Region Growing Algorithm," in 18th International Conference on Systems, Signals and Image Processing-IWSSIP,pp. 395-400, 2011.