ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

An enhanced segmentation approach to detect lung cancer nodule from CT scan images

Mr.B. Muthazhagan, Suganya J
Associate Professor/IT, IV Year IT
Kings Engineering College, Kings Engineering College
muthazhagan@gmail.com, suganya.lakshmi3116@gmail.com

Abstract
Lung cancer seems to be the common cause of death among people throughout the world. Early detection of lung cancer can increase the chance of survival among people. The overall 5-year survival rate for lung cancer patients increases from 14 to 49% if the disease is detected in time. Although Computed Tomography (CT) can be more efficient than X-ray. However, problem seemed to merge due to time constraint in detecting the present of lung cancer regarding on the several diagnosing method used. Hence, a lung cancer detection system using image processing is used to classify the present of lung cancer in an CT-images. In this study, MATLAB have been used through every procedures made. In image processing procedures, process such as image pre-processing, segmentation and feature extraction have been discussed in detail. We are aiming to get the more accurate results by using various enhancement and segmentation techniques.

Keywords: Computed Tomography (CT), image processing, MATLAB, Enhanced Segmentation.

1. INTRODUCTION

This project deals with the prediction of lung cancer detection with the help of the Computed Tomography (CT) images. By improving the efficiency of the image segmentation algorithm we can detect tumor from the Lung Cancer Computed Tomography images. We are aiming to get the more accurate results by using various enhancement and segmentation techniques in the domain of Image Processing.

Lung diseases are some of the most common medical conditions worldwide. Tens of millions of people suffer from lung disease worldwide. Smoking, infections, and genetics are responsible for most lung diseases.

Some of the common lung disease can be divided in these groups:

- Lung Diseases Affecting the Airways;
- Lung Diseases Affecting the Air Sacs (Alveoli);
- Lung Diseases Affecting the Interstitium;
- Lung Diseases Affecting Blood Vessels;
- Lung Diseases Affecting the Pleura;
- Lung Diseases Affecting the Chest Wall.

Most occurred ones are: asthma, acute bronchitis, cystic fibrosis, emphysema, pneumonia, tuberculosis, emphysema, pulmonary edema, lung cancer, sarcoidosis, and different kind of pulmonary edemas.

The overall 5-year survival rate for lung cancer patients increases from 14 to 49% if the disease is detected in time. Although Computed Tomography(CT) can be more efficient than X-ray. However, problem seemed to merge due to time constraint in detecting the present of lung cancer regarding on the several diagnosing method used.

Imaging plays a vital role in the diagnosis of lung cancer, with the most common modalities including chest radiography, CT, PET, magnetic resonance imaging (MRI), and radionuclide bone scanning, but in this work, we primarily used CT images for analysis. X-Ray imaging will show most lung tumors, but CT is used because it is more sensitive in finding tumor size and the presence of lymph node metastases. However, with CT imaging, it is not always easy to distinguish the limits between tumor and normal tissue, especially when the dense pathology is present. Recent advances in Computed Tomography (CT) technology have enabled its use in diagnosing and quantifying different diseases. In particular, the expanding volume of thoracic CT studies along with the increase of image data, bring in focus the need for CAD algorithms to assist the radiologists. Several lung diseases are

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

diagnosed by investigating the patterns of lung tissue in pulmonary CT images, therefore segmentation and analysis is one of the important parts of CAD systems.

CT or computer axial tomography, uses special X-ray tube to obtain image data from different angles around the body, and then uses computer processing of the data to show a cross-section of body tissues and organs. Some of the basic ideas underlying CT are reconstruction from projections that means that the patient data is getting measured at many positions and angles. CT modalities can show various types of tissues, lung, soft tissue and bones, and using specialized equipment and expertise to create and interpret CT scans of the body, radiologists can more easily diagnose tumor, cancer or other lesion, and to measure its size, precise location, and the extent of the tumor's involvement with other nearby tissue. The images received from a CT scanner can reveal some soft tissue and other structures that cannot even be seen in conventional X-rays.

In imaging science, image processing is processing of images using mathematical operations by using any form of signal processing for which the input is an image, a series of images, or a video, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it. Images are also processed as three-dimensional signals with the third-dimension being time or the z-axis.

Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The acquisition of images (producing the input image in the first place) is referred to as imaging.

Closely related to image processing are computer graphics and computer vision. In computer graphics, images are manually made from physical models of objects, environments, and lighting, instead of being acquired (via imaging devices such as cameras) from natural scenes, as in most animated movies. Computer vision, on the other hand, is often considered high-level image processing out of which a machine/computer/software intends to decipher the physical contents of an image or a sequence of images (e.g., videos or 3D full-body magnetic resonance scans).In modern sciences and technologies, images also gain much broader scopes due to the ever growing importance of scientific visualization (of often large-scale complex scientific/experimental data). Examples include microarray data in genetic research, or real-time multi-asset portfolio trading in finance.

A lung cancer detection system using image processing is used to classify the present of lung cancer in a CT-images. In image processing procedures, process such as image pre-processing, segmentation and feature extraction have been discussed in detail. We are aiming to get the more accurate results by using various enhancement and segmentation techniques. [6] proposed a method in which the minimization is per-formed in a sequential manner by the fusion move algorithm that uses the QPBO min-cut algorithm. Multi-shape GCs are proven to be more beneficial than single-shape GCs. Hence, the segmentation methods are validated by calculating statistical measures. The false positive (FP) is reduced and sensitivity and specificity improved by multiple MTANN.

2. IMAGE ACQUISITION

First step is to acquire the CT scan image of lung cancer patient. The lung CT images are having low noise when compared to X-ray and MRI images; hence they are considered for developing the technique. The main advantage of using computed tomography images is that, it gives better clarity and less distortion. DICOM (Digital Imaging and Communications in Medicine) has become a standard for medical Imaging. Figure 5.2 shows a typical CT image of lung cancer patient used for analysis. The acquired images are in raw form. In the acquired images lot of noise is ob-served. To improve the contrast, clarity, separate the background noise, it is required to pre-process the images. Hence, various techniques like smoothing, enhancement are applied to get image in required form.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

Figure 2.1: Input CT Scan image

3. IMAGE PRE-PROCESSING

3.1 NOISE FILTERING

It is the initial step for detecting the lung cancer. In preprocessing step we can do two steps: They are Denoising and Weiner filtering. Image denoising algorithms may be the mostly used in image processing. Many methods, regardless of implementation, share the same basic idea noise reduction through image blurring. Blurring can be done locally, as 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.

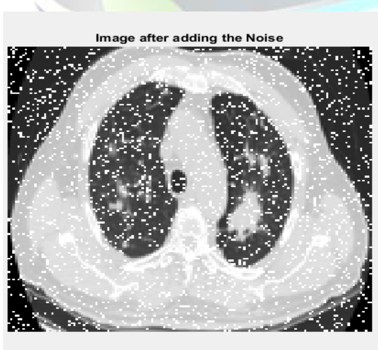

Figure 3.1: Image after adding the Noise

4. IMAGE SEGMENTATION

FUZZY SPATIAL ALGORITHM

Crisp spatial data types (for example, (Clementini & Di Felice, 1996b; Schneider, 1997; Schneider & Behr, 2006)) for points, lines, and regions have been formally defined on the basis of point sets and *point set topology* (Gaal, 1964). Our goal is to leverage their fuzzy counterparts for a definition of fuzzy spatial data types. In this section, we summarize some needed basic concepts and notations of fuzzy set theory and fuzzy topology.

Fuzzy set theory (Zadeh, 1965; Buckley & Eslami, 2002) is an extension and generalization of Boolean set theory. Let *X* be a classical (crisp) set of objects, called the *universe* (*of discourse*). Membership in a classical subset *A* of *X* can then be described by the
characteristic function c*A* : *X* !*f*0*;*1*g* such that for all
x 2 *X* holds:
c*A*(*x*) =
(
1 if, and only if, *x* 2 *A*
0 if, and only if, *x* =2 *A*
This function, which discriminates sharply between members and non-members of a set, can be generalized such that all elements of *X* are mapped to the real interval [0,1] indicating the *degree of membership* of these elements in the set in question. Hence, fuzzy set theory permits an element to have partial membership in a fuzzy set and/or multiple membership in several, different fuzzy sets. Larger values designate higher grades of set membership. Let *X* again be the universe. Then
μ˜*A* : *X* ![0*;*1]
is called the *membership function* of ˜*A*, and
the set˜*A*= *f*(*x;μ*˜*A* (*x*)) *j x* 2 *Xg* is called a *fuzzy set* in *X*. All elements of *X* receive a valuation with respect to their membership in ˜*A*. Those elements *x* 2 *X* that in the classical sense do not belong to ˜*A* get the membership value *μ*˜*A*(*x*) = 0; elements *x* 2 *X* that completely belong to ˜*A* get the membership value *μ*˜*A*(*x*)=1.

Next, we extend the set inclusion as well as the basic crisp set operations to fuzzy sets. We will comply with the definitions in Zadeh (1965). Let ˜*A* and ˜*B* be fuzzy sets in *X*. Then
(i) *:*˜*A* = *f*(*x;μ:*˜*A*(*x*)) *j x* 2 *X;μ:*˜*A* (*x*) = 1*¡μ*˜*A* (*x*)*g*
(ii) ˜*A μ* ˜*B* , 8 *x* 2 *X* : *μ*˜*A* (*x*) · *μ*˜*B*(*x*)
(iii) ˜*A* \ ˜*B* = *f*(*x;μ*˜*A*\˜*B*(*x*)) *j x* 2 *X* ^ *μ*˜*A*\˜*B* (*x*) = min(*μ*˜*A*(*x*)*;μ*˜*B* (*x*))*g*
(iv) ˜*A* [˜*B* = *f*(*x;μ*˜*A*[˜*B*(*x*)) *j x* 2 *X* ^ *μ*˜*A*[˜*B* (*x*) = max(*μ*˜*A*(*x*)*;μ*˜*B* (*x*))*g*
(v) ˜*A ¡* ˜*B* = ˜*A* *:*˜*B*
An a-*cut* or a-*level set* of a fuzzy set ˜*A* for a specified value a is the crisp set
A¸a = *fx* 2 *X j μ*˜*A*(*x*) ¸ a ^ 0 · a · 1*g*

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

A *strict* a-*cut* or *strict* a-*level set* of a fuzzy set ˜*A* for a specified value a is the crisp set
A>a = *fx* 2 *X j µ*˜*A*(*x*) > a ^ 0 · a < 1*g*
The strict a-cut for a = 0 is called the *support* of ˜*A*, that is, *supp*(˜*A*) = *A*>0. The a-cut for a = 1 is called the *core* of ˜*A*, that is, *core*(˜*A*) = *A*¸1. For a fuzzy set ˜*A* and a*;*b 2 [0*;*1] holds
(i) *X* = *A*¸0
(ii) a < b *)* *A*¸a *¶* *A*¸b
The set of all levels a *2* [0*;*1] that represent distinct a-cuts of a given fuzzy set ˜*A* is called the *level set*
L˜*A* of ˜*A*:
L˜*A* = *f*a 2 [0*;*1] *j 9 x* 2 *X* : *µ*˜*A* (*x*) = a*g*
Fuzzy (point set) topology (Chang, 1968; Liu & Luo, 1997) is a straightforward extension and generalization of ordinary point set topology and allows one to distinguish specific topological structures of a fuzzy set like its closure or interior.
For a non-empty, crisp set *X*, let [0*;*1]*X* = *f f j f* : *X* *!*[0*;*1]*g* be the set of all mappings from *X* to [0*;*1].
Let 1*X* be the function that corresponds to the whole set *X*, and 0*X* be the function that corresponds to the empty set ?. A *fuzzy topology* on *X* is a family *T µ* [0*;*1]*X* of fuzzy sets satisfying the following conditions:
(i) 1*X* 2 *T ;*0*X* 2 *T*
(ii) ˜*A* 2 *T ;* ˜*B* 2 *T)* ˜*A* \ ˜*B* 2 *T*
(iii) *S µ T)*
S˜*A* 2*S* ˜*A* 2 *T*
The pair (*X*, *T*) is said to be a *fuzzy topological space*. The elements of *T* are called *open fuzzy sets*1.
The family *T 0* of all *closed fuzzy sets*2 in a fuzzy topological space (*X*, *T*) is given by
T 0 = *f*:˜*A j* ˜*A* 2 *T g*
The *closure* of a fuzzy set ˜*A* in a fuzzy topological space (*X*, *T*) is the smallest closed fuzzy set containing ˜*A*, that is,
clT (˜*A*) = T
f˜*B j* ˜*B* 2 *T 0* ^ ˜*A µ* ˜*Bg*
The *interior* of a fuzzy set ˜*A* in a fuzzy topological space (*X*, *T*) is the largest open fuzzy set contained in ˜*A*, that is,
intT (˜*A*) = S
f˜*B j* ˜*B* 2 *T* ^ ˜*B µ* ˜*Ag*
We obtain an example of a fuzzy topological space if we equate *X* with the Euclidean plane R2. The fuzzy topological space (R2*;T*) .

Figure:4.1 output of segmented image

5. IMAGE EXTRACTION AND SELECTION

This stage is an important stage that uses algorithms and techniques to detect and isolate various desired portions or shapes of a given image. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant, then the input data will be transformed into a reduced representation set of features. The basic characters of feature are area, perimeter and eccentricity. These are measured in scalar. These features are defined as follows:
A) Area: It is the scalar value that gives actual number of overall nodule pixel in the extracted ROI. Transformation function creates an array of ROI that contains pixels with 255 values.
Area = A = (Ai,j , X ROI[Area] = I , Y ROI[Area] = j) Where, i, j are the pixels within the shape. ROI is region of interest. X ROI[ ] is vector contain ROI x position, Y ROI[ ] is vector contain ROI y position.
B) Perimeter: It is a scalar value that gives actual number of the nodule pixel. It is the length of extracted ROI boundary. Transformation function create array of edge that contain pixel with 255 values that have at least one pixel which contain 0 values.
Perimeter = P =(Pi,j , X edge[P] = i, Y edge[P] = j)
Where, X edge [ ] and Y edge[ ] are vectors represent the co-ordinate of the ith and jth pixel forming the curve, respectively.
C) This metric value is also called as roundness or circularity or irregularity complex (I) equal to 1 only for circular and it is less than 1 for any other shape.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

$$\text{Eccentricity} = \frac{\text{Length of Major Axis}}{\text{Length of Minor axis}}$$

Features estimated for separated nodule of given sample image above as shown in fig.3.2.4 has been found as follows:

1) **Area:** It is the simplest property and by its given size. Therefore, it is the total number of white pixels in the extracted area.

2) **Perimeter:** It is another simple property defined by the perimeter of the region. It is the length of extracted ROI boundary.

3) **Eccentricity:** It is used to decide the shape or circularity of the object. Area: 2291 Perimeter: 221 Eccentricity: 0.8289

Figure 5.1 extracted image

6. LUNG CANCER TYPE PREDICTION CLASSIFICATION

Support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification. The basic SVM takes a set of input data and for each given input, predicts which of two classes forms the input, making it a non-probabilistic binary linear classifier. SVM uses a kernel function which maps the given data into a different space; the separations can be made even with very complex boundaries. The different types of kernel function include polynomial, RBF, quadratic, Multi-Layer Perceptron (MLP). Each kernel is formulated by its own parameters like γ, σ, etc. Figure 6.1. shows maximum margin hyper planes. The original hyper plane algorithm is a way to create non linear classifier by applying the kernel trick to maximum margin hyper planes.

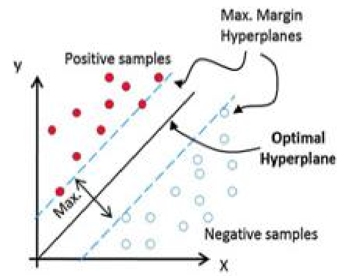

Figure 6.1 max margin

STAGING OF LUNG CANCER

Lung nodules are the smallest growths in the lung that measure between 5 mm to 25 mm in size. In abnormal images nodule size is greater than 25 mm.

Staging involves evaluation of a cancer size and its penetration into surrounding tissues as well as presence or absence of metastasis in the lymph nodes or other organs . Stages from I to IV in order of severity:

- Stage I : cancer is confined to the lung
- Stage II and III : cancer is confined to the chest
- Stage IV, cancer has spread from the chest to other parts of the body.

According to the medical field, non-small cell lung cancer is staged using TNM system (T for extent of primary tumor, N for regional lymph node involvement and M for metastasis). Table shows the criteria decided by doctors for the classification of lung cancer stages. Different types of stages are as shown in following figures according to their parameters

7. RESULTS

Finally the output will be predicted after the different iterations of the selected image and it is shown in figure 7.1.

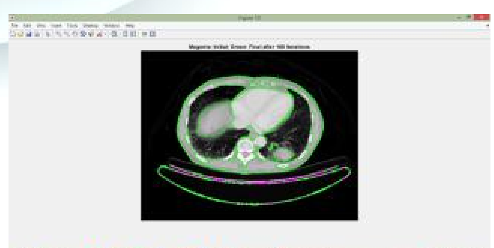

Figure 7.1 output predictions

8. CONCLUSION

In this paper, I achieved our purpose in developing an automatic CAD system for early detection of lung cancer by analyzing LUNG CT images using several steps. The approach starts by extracting the lung regions from the CT image using several image processing techniques, including bit image slicing, erosion, weiner filter. We introduced the using of bit plane slicing technique instead of the thresholding technique that is used in the first step in the extraction process to convert the CT image into a binary image. Bit-plane slicing technique is both faster and data- and user-independent compared to the thresholding technique. After the extraction step, the extracted lung regions are segmented using egion growing segmentation algorithm. Then, the initial lung candidate nodules resulting from the Region growing segmentation are analyzed to extract a set of features to be used in the diagnostic rules. After that to segment the lung region. In this region to detect cancerous region and to get the accurate result. The accuracy of 80% approximate the accuracy indicated by surgeons and radiologists for locating cancerous nodules during reading clinical CT images.

9. FUTURE WORK

In the segmentation part the threshold value can be implemented by the another efficient algorithm where it should satisfy the value of the threshold value of approximation

10. REFERENCES

[1] B. Muthazhagan and T. Ravi "An early diagnosis of Lung Cancer disease using data mining and medical image processing methods", Middle –East Journal of Scientific Research 24(10):32b3-3267,2016 ISSN 1990-9223.

[2] American Cancer Society, "Cancer Statistics, 2005", CA: A Cancer Journal for Clinicians, 55: 10-30, 2005, "http://caonline.amcancersoc.org/cgi/content/full/55/1/10".

[3] B.V. Ginneken, B. M. Romeny and M. A. Viergever, "Computer-aided diagnosis in chest radiography: a survey", IEEE, transactions on medical imaging, vol. 20, no. 12, (2001).

[4] 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 helicalCT images", Compute. Med. Image Graph, vol. 22, no. 2(1998),pp. 157-167.

[5] D. Lin and C. Yan, "Lung nodules identification rules extraction with neural fuzzy network", IEEE, Neural Information Processing, vol. 4,(2002).

[6] Christo Ananth, G.Gayathri, M.Majitha Barvin, N.Juki Parsana, M.Parvin Banu, "Image Segmentation by Multi-shape GC-OAAM", American Journal of Sustainable Cities and Society (AJSCS), Vol. 1, Issue 3, January 2014, pp 274-280

[7] A. El-Baz, A. A. Farag, PH.D., R. Falk, M.D. and R. L. Rocco,M.D.,"detection, visualization, and identification of lung abnormalities in chest spiral CT scans: phase I", Information Conference on Biomedical Engineering, Egypt (2002).

[8] Linda G. Shapiro and G.C. Stockman., Computer Vision: Theory and Applications. 2001: Prentice Hall.

[9] The DICOM Standards Committee. DICOM homepage. http: //medical.nema.org/, September 2004.

[10] B. Magesh, P. Vijaylakshmi, M. Abhiram, "Computer aided Diagnosis System for identification and classification of Lessions in Lungs", International Journal of Computer Trends and Technology- May to June Issue 2011.

[11] Rachid Sammouda, Jamal Abu Hassan, Mohamed Sammouda, Abdulridha Al-Zuhairy, Hatem abou ElAbbas, "Computer Aided Diagnosis System for Early Detection of Lung Cancer Using Chest Computer Tomography Images", GVIP 05 Conference, 19-21 December 2005, CICC, Cairo, Egypt.

[12] Yamomoto. S, Jiang. H, Matsumoto. M, Tateno.Y, Iinuma. T, Matsumoto. T, "Image processing for computer-aided diagnosis of lung cancer by CT (LSCT)", Proceedings 3rd IEEE Workshop on Applications of Computer Vision, WACV '96, pp: 236 –241, 1996.

[13] Yeny Yim, Helen Hong and Yeong Gil Shin, "Hybrid lung segmentation