ISSN: 0975-766X
CODEN: IJPTFI
Available Online through *Research Article*
www.ijptonline.com

SVM BASED LUNG TUMOR SEGMENTATION USING OTSU'S THRESHOLDING

Ms. S. Parveen Banu*, Dr. N. Sakthieswaran, Dr. P. Suresh, Ms. Kanagalakshmi,
AP/ECE, Sree Sastha College of Engineering, Chennai.
AP/Civil, Regional Centre, Anna University, Chennai.
AP/ECE, National College of Engineering, Chennai.
AP/ECE, Sree Sastha Institute of Engineering and Technology, Chennai.
Email: jesima83@gmail.com

Received on 06-08-2016 *Accepted on 10-09-2016*

Abstract

Lung cancer is the abandoned growth of abnormal cells that instigate in one (or) both lungs; usually in the cells that line the air passages. The anomalous cells segregate rapidly and form tumors but not as a healthy lung tissue. There are three major types of lung cancer. They are NSCLC, SCLC and lung carcinoid tumors. Staging lung cancer is based on whether the cancer is local or has spread from the lung to the lymph nodes or other organs. In this paper, a novel approach is proposed to segment the lung cancer image. In preprocessing, first RGB is transformed into grayscale image and for the elimination of noise median filter is used. From the denoised image, we have segmented the lung region using Otsu's thresholding. Then the morphological operation is used to smoothen the images. Here for feature extraction, gray level co-occurrence matrix is used to extract the texture features. These features are imported as an input data to the support vector machine(SVM) classifier. The main aim of this paper is to widen a CAD system to recognize the lung tumor from the lung CT images and classify the tumor as normal or abnormal. In this research, MATLAB have been used through every procedures made. The proposed approach has produced efficient results with less false results and improves the accuracy.

Keywords: Computed Tomography (CT), Otsu's Thresholding, Morphological operations, SVM classifier.

I. Introduction

In recent years, medical imaging plays a vital role to improve public health with major advancements. Different imaging modalities are effectively utilized to achieve the information about the human body in order to diagnose, monitor, or treat medical conditions [1].It helps to identify the human diseases in earlier stages and makes the treatment easier. Among various types of cancer, lung cancer is the foremost reason of cancer deaths in universal. In the current year, it is predicted

that 85,920 men and 72,160 women deaths will occur due to this disease [2].Cigarette smoking, passive exposure to tobacco smoke and air pollution are prime factor of lung cancer.

There are three main types of lung cancer. They are NSCLC, SCLC and lung carcinoid tumor. Around 15% of lung cancer is NSCLC, 10-15% is SCLC and less than 5% is lung carcinoid tumors. In these, NSCLC is the most common type of lung cancer.[3]Lung cancer symptoms are cough, coughing up blood,fatigue,unexplained weight loss, shortness of breath etc.,[4]

Tumors are classified into two types. They are

- Benign tumor
- Malignant tumor

Benign tumors are non-cancerous. They tend to grow slowly and stay in one place, not spreading into other parts of the body. Benign tumors usually stay non-cancerous, except in very rare cases.

Malignant tumors are cancerous. Cancer can initiate in any one of the trillions of cells in our body. Usually in our body new cells will generate and old cells or damaged cell will die whenever it needs. When cancer occurs, this cyclic process will be affected. Unusual growth of new cells and survival of old or damaged cells will occur.

Staging helps to describe the location of the cancer, size of the tumor and if it has spread. Knowing the stage, doctors can able to know the chances of survival and also it helps to plan the best treatment. NSCLC staging uses the TNM system. Doctors commonly use the American Joint Committee on Cancer's (AJCC's) and the International Union for Cancer Control (UICC) to maintain the TNM classification system. [5]

CT scanning is often the best method to diagnose many cancers such as liver, lung and pancreatic cancers and also it allows the doctor to confirm the presence of a tumor and determine its size and location.

A computer aided system to diagnose lung nodule from the lung CT images automatically and semi-automatically is a challenging task in image processing. Several techniques are existed to detect the tumor in earlier stage and to help the radiologist to make easy decision. Some of the methods are discussed as follows.

Qiao Wei et.al.[12] segmented lung lobe with the implementation of two approaches. In this paper, fissure region, fissure location and curvatures are identified by using adaptive fissure sweeping and wavelet transform methods. The accuracy rate is 80% for locating the fissure regions.Khin Mya Mya Tun et.al.[16] implemented Otsu's thresholding for

segmentation purpose. For classification of tumor, GLCM feature extraction along with artificial neural network classifier is used. The accuracy rate of this existing system is 90%.Sunil Kumar et.al.[17] developed an algorithm to segment the lung parenchyma. In this paper, region growing method and edge detection technique is used.Ruchika et.al.[18] developed a solution for the image segmentation using mean shift algorithm. After this, morphological operations are performed to identify the edges of the lungs.Nastaran Emaminjad et.al.[19] developed a CAD scheme to segment lung tumors. Naive Bayesian network based classifier and multilayer perceptron classifier were used. Prediction scores of two classifiers were fused by the fusion method and AUC value was significantly increased.

II. Frame Work

The proposed method deals with four major steps. They are preprocessing, segmentation, feature extraction and classification. In preprocessing, salt and pepper noise is removed with the help of median filter. The segmentation is done by using the Otsu's Thresholding techniques. In this research, GLCM feature extraction technique is implemented. From these, features are extracted from the segmented image and it is fed as an input to the SVM classifier.Finally, the classification outputs are compared between the trained feature and test feature and the performance measurement parameter accuracy is measured.

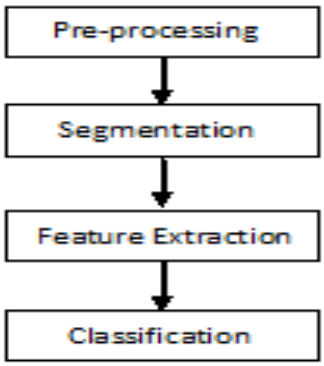


Fig. 1. Flow chart of the frame work.

III. Methodology

In this chapter, frame work of the paper is explained briefly. Here, the input image is the lung CT image in DICOM format and it is converted into gray image. And also it is resized into 256x256 pixels in size.

A. Preprocessing

Preprocessing of an image is an essential step whose aim is to improve image quality by suppressing the redundancy or by enhancing the image features for further processing. Usually, CT images contain salt and pepper noise, white noise etc.,.

This noise is removed by using median filter with 3x3 mask size. Median filter is a non-linear digital filtering technique and does a better job of preserving useful detail in the image [6].The input image and the denoised image is exposed in fig.2&3.

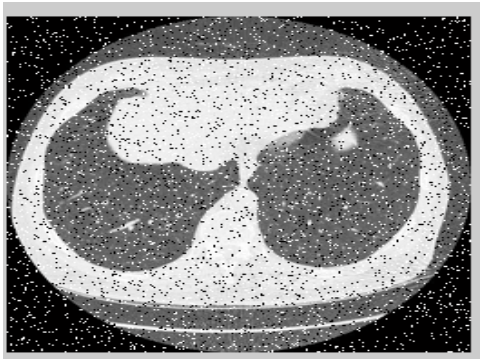


Fig.2.Noisy Input Image.

Fig.3.Denoised Image.

B. Segmentation

Segmentation is the important technique to convert a digital image into meaningful structure. In image processing, there are enormous types of segmentation techniques are found. But still, no suitable algorithm is found for the particular images. Thresholding is nothing but binarization of image by using threshold value T and it is the simplest approach in image segmentation. It has two types; one is the global thresholding and second is the local thresholding. For segmentation process we use Otsu's Thresholding [8] and some morphological operations.

Otsu's Thresholding

In Otsu's method [8],global image threshold and multilevel image threshold is achieved with the mat lab function gray thresh(I) and multithresh(A,N).In the global image threshold, only one threshold value is applied whereas in multilevel image threshold N threshold values are applied. In the proposed work, gray threshold is applied to segment the CT lung image with the threshold value of 0.566.

Morphological Operations

After the Otsu's thresholding, the resultant image is processed by applying some morphological operations. As a result, lung nodule is segmented by dilation and erosion operations. Dilation is the adding up of pixels in an image based on the structuring element whereas erosion is the reverse process of it (i.e.) eroding the pixels in an image. Both techniques works on the basis of the given structuring element. A structuring element [7]

The dilation of *A* using structuring element *B* is defined by:

$$A \oplus B = \bigcup_{b \in B} A_b$$

where A_b is the translation of *A* by *b*.

The erosion of the input image *A* using the structuring element *B* is defined by:

$$A \ominus B = \{z \in E | B_z \subseteq A\}$$

where B_z is the interpretation of *B* by the vector z, i.e.,

$$B_z = \{b + z | b \in B\}, \forall_z \in E$$

Using morphological operations with the structuring element of disk, the unwanted parts are removed and lung nodule is obtained.

C. Feature Extraction

Feature extraction is a necessary step which is used to classify the tumor. Some of the important features are calculated and it is provided as an input to the classifier.

GLCM

The GLCM is an earliest approach for the texture feature extraction proposed by Haralick et.al.[10].It is a powerful tool to extract second order statistical texture features.GLCM is a square matrix and the elements of the square matrix is estimated by considering the two neighboring pixel intensities. Entry of the elements represents how often a pixel with value I is adjacent to a pixel with value j.It is possible to calculate the adjacency in four directions (horizontal, vertical, left & right diagonal) as shown in fig.GLCM is the two dimensional matrix of joint probabilities $P_{d,\theta}$ (i.j) between pairs of pixels, separated by a distance 'd' in a given direction θ.

GLCM Features: Haralick suggested [9] 14 texture features. These features are calculated by the co-occurrence matrices obtained by using the four directions. In this paper, four main features are calculated and it is given as follows [14].

$$\text{Contrast} = \sum \tilde{\sum} (i-j)^2 p(i, j)$$

$$Correlation = \frac{\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i-\mu_i)(j-\mu j)p(i,j)}{\sigma_i\sigma_j}$$

$$\text{Energy} = \sum \sum (p(i, j))^2$$

$$Homogeneity = \sum_{i,j}\frac{P(i,j)}{1+|i-j|}$$

The glcm features value for various images is given below.

Images	Contrast	Correlation	Energy	Homogeneity
Image1	0.0020	0.9613	0.9451	0.9990
Image2	0.0005	0.9062	0.9943	0.9998
Image3	0.0005	0.9048	0.9940	0.9997
Image4	0.0008	0.9511	0.9836	0.9996

IV. Tumor Classification

After the feature extraction, the extracted features are given as an input to the classifier. Classification plays a key role to identify the tumor as normal or abnormal. In this paper, SVM classifier is used for the identification of tumor.

Support Vector Machine

Support vector machines (SVMs) are authoritative tools for data classification; it classify two category points by conveying them to one of two dislodge half spaces in any of the original input space of the problem for linear classifiers, or in a higher dimensional feature space for nonlinear classifiers. It is a supervised learning algorithm that is hypothetically simple, easier implementation, habitually faster & has better scaling properties for complicated problems. It is used mainly in classification problem but rarely in regression challenges. In this algorithm, each class of points is assigned to the closest of two parallel planes that are pushed apart as far as possible. In SVM, finding the right hyper plane to separate the two classes effectively is not a tough job and it can be done with the help of thumb rule. And also, margin (distance between nearest data item) should be high between the two data points. The reason for the higher margin is robustness and if lower margin hyper plane exists, then mis-classification will occur. Here two class svm classifier is used.

There is much more in the field of SVM kernel machines than we could cover here, including:

Regression, clustering, semi-supervised learning and other domains.

- Lots of other kernels, e.g. string kernels to handle text.
- Lots of research in modifications, e.g. to improve generalization
- Ability or tailoring to a particular task.
- Lots of research in speeding up training.

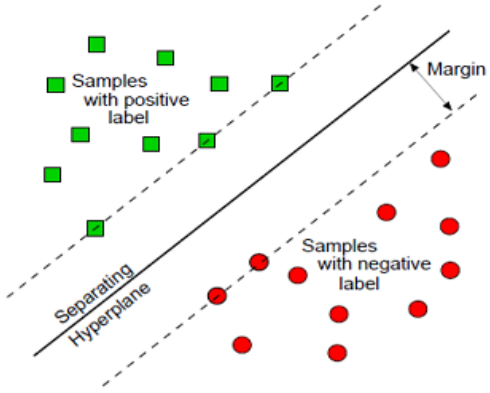


Fig. 1. Magnetization as a function

The hyper plane is defined by the equation

$$(w. x) + b = 0 \qquad (3.5)$$

Where, w= weight vector

x= feature vector

b= bias

Training the Classifier

In the training phase, known data is given and the classifier is trained. Here, six images are used for training out of which 3 are benign and 3 are malignant and are assigned class 0 for benign and class 1 for malignant

$$(w.xi) + b \geq +1 \text{ for } yi = +1 \qquad (3.6)$$

$$(w.xi) + b \leq -1 \text{ for } yi = -1 \qquad (3.7)$$

Testing of Data

In testing phase, unknown data are given and the classification is performed using trained classifier. Classification is done by using following decision function.

$$f(x, \{w, b\}) = \text{sign}(w.x+b) \qquad (3.8)$$

The sign of this function decides the class of the test image. Here, if it is positive, then result will be "malignant" and if it

is negative, then result will be "benign".

V. Results and Discussion

In this proposed method, 64 slices CT lung images of size 256x256 is used.CT scan images are used to analyze the anatomical structure of the body.Inorder to find the tumor, CT scan modality is one of the best modality among all.

In this research work, input image is resized and it is fed as an input to the median filter. Among various filter, median filter is the best filter to remove salt & pepper noise because they provide excellent noise reduction capabilities. After noise removal, Otsu's thresholding along with morphological operations are used to detect the cancer lung nodule alone. Then the features are extracted from these lung nodules by the GLCM algorithm. GLCM is a second order statistical measure which is used to calculate the texture features. Using GLCM, accuracy will be improved when compared with the conventional feature extraction method. These features are provided as an input vector to the SVM classifier. Finally, the classifier compare the test features with the trained features and classify the tumor as malignant is used to classify the tumor. The overall performance of the system is measured with the accuracy.

Accuracy refers to the closeness of true results to the total number of cases examined.

Accuracy = (TP + TN) / (TP + TN + FP + FN) * 100

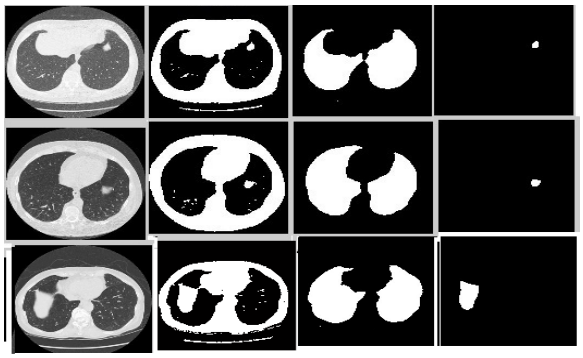
Where TP denotes True Positive; abnormal tumor detected correctly.

TN denotes True Negative; abnormal tumor detected incorrectly.

FP denotes False Positive; normal tumor detected correctly.

FN denotes False Negative; normal tumor detected incorrectly.

The accuracy rate of SVM classifier is 93%.The proposed system is trained with 64images and tested with 10 images. The results are shown in below fig.



(a) (b) (c) (d)

Fig.4.(a,b,c,d) Input image, Binary Image, Parenchyma Extraction, Lung Nodule.

Fig 4(a) is the input image after the removal of noise using median filter. Fig 4(b) is the binary image obtained after applying Otsu's thresholding. Fig 4(c) is the lung parenchyma extracted image. Finally Fig 4(d) is the lung cancer nodule image which is the result of the morphological operations (i.e.) dilation and erosion using the structuring element 'disk'.

VI. Conclusion and Future work

In image processing, medical imaging is the emerging out field in world wide. Now-a-days, identification of cancer disease becomes a serious problem consecutively to increase the survival rate. Because initial stage cancer patients will be recovered easily from it. So, the challenging work of the researchers is to find the cancer tumor at the initial stage itself. To solve this problem, so many research works are carrying out. However, still improving accuracy is a major task in this field .The proposed system helps the radiologist to evaluate whether the tumor is benign or malignant by providing more information using glcm feature extraction technique. In this research work, Otsu's thresholding along with SVM classifier is proposed for the segmentation and classification of lung CT images. To detect the tumor, various modalities are available in medical field. In future, the accuracy rate will be improved by combining the two different modality images using image registration technique. By this method, cancer tumor can be identified effectively and also there is a possible way to improve the accuracy.

References

1. www.fda.gov/Radiation-Emitting Products
2. www.cancernet/cancer-types/lung cancer-non-small-cell/statistics
3. www.cancer.org/cancer/lung cancer/
4. www.webmed.com/lung-cancer/
5. www.cancer.org/treatment/understanding your diagnosis/staging
6. https://en.wikipedia.org/wiki/Median-filter
7. https://en.wikipedia.org/wiki/Dilation-(morphology)
8. Rafael C.Gonzalez, Richard E.woods,"Digital Image Processing", Pearson Education.
9. R.M.Haralick,K.Shanmugam,andI.Dinstein(1973),"Textural Features of Image Classifiacation", IEEE,vol.3,pp.601-621
10. R.M.Haralick (1979),"Statistical and structural approaches to texture", proceedings of the IEEE, vol.67, pp.786-804

11. Yu-Qian,Z.,Wei-Hua,G.,Zhen-Cheng,C.,Jing-Tian,T.,,Ling-Yun,L(2005),"Medical images edge detection based on mathematical morphology", Proceedings of the IEEE Engineering in Medicine and Biology,27th annual conference shangai,china,pp.6492-6495
12. Qiao Wei,Yaoping Hu,Gary Gelfand and John.H.MacGregor(2009),"Segmentation of lung lobes in high resolution isotropic CT images",IEEE transactions on Biomedical Engineering,vol.56,No.5,pp no:1383-1393
13. A.Padma,R.Sukanesh(2011),"Automatic classification and segmentation of brain tumor in CT images using optimal dominant grey level run length texture features",IJACSA,vol-2,no.10,pp:53-59
14. A.Gebejes, R.Huertas (2013),"Texture characterization based on grey level co-occurrence matrix", 2nd international conference, ICTIC 2013.
15. Mohsen keshani,Zohreh Azimifar,Farshad Tajeripour,Reza Boostani(2013),"Lung nodule segmentation and recognition using svm classifier and active contour modeling: A complete intelligent system",Elseveir Ltd.,vol.43,issue 4,page 287-300
16. Khin Mya Mya Jun, Aung Soekhaing (2014),"Feature extraction and classification of lung cancer nodule using image processing techniques", IJERT, vol.3, issue 3, ISSN: 2278-0181
17. Sunilkumar, Ashokumar(2014), "Lung segmentation using region growing algorithm", IJARCSSE, vol.4, issue3, ISSN:2277 128x, pp:184-187
18. Ruchika, Ashimakalra(2015), "Detection of lung cancer in CT images using mean shift algorithm", IJARCSSE, vol.5, issue5, ISSN:2277-128x, pp:1147-1149
19. Nastaran Emaminejad, Wei Qian, Yubao Guan, Maxine Tan, Yuchen Qiu, Hong Liu and Binzheng(2015),"Fusion of quantitative image and genomic biomarkers to improve prognosis assessment of early stage cancer patients",IEEE transaction on biomedical engineering.

Corresponding Author:
Ms. S. Parveen Banu*,
Email: *jesima83@gmail.com*