Lung Cancer Image Segmentation And Classification Using Soft Computing Techniques

G. Madhu Bala
Department of Computer Science
Periyar University
Salem-636011
madhumca93@gmail.com

I. Laurence Aroquiaraj
Department of Computer Science
Periyar University
Salem-636011
laurence.raj@gmail.com

***Abstract*- In this paper it shows current medical diagnosis, treatment, and surgery, medical imaging plays one of the most significant roles, since imaging devices such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and ultrasound diagnostics yield a great deal of information about diseases and organs. Lung cancer is the uncontrolled growth of abnormal cells that start off in one or both lungs, usually in the cells that line the air passages. The abnormal cells do not develop into healthy lung tissue; they divide rapidly and form tumors. In this paper, the noise that affects the features of CT images, the widespread use of CT image imaging requires the need for developing filter for decreasing noise. Some of the related work is given the most thresholding based segmentation methods attempt to segment the CT images. Most previous works are prepared to compare different thresholding based image segmentation algorithms based on characteristics such as correctness, stability with respect to parameter choice and stability with respect to image choice. Performance measure like precision, specificity and false positive rate is used to evaluate the accuracy. It is observed from the experimental evaluation that the performance of ANN is better than that of individual classifiers of Lung image data sets.**

Keywords- Image processing, Lung Cancer, Segmentation, ANN, SVM Classifier

I. INTRODUCTION

Cancer is a disease in which cells acquire genetic alterations and divide without control. Cell division of normal lung tissue is necessary to retain the structure and functionality of the organ. Normal cells undergo controlled transitions between resting and dividing states. Exposure to cigarette smoke, excess radiation and other environmental carcinogens along with genetic factors can cause malignant transformation (carcinogenesis) of normal cells. Malignant or cancerous cells grow and divide independent of the needs and limitations of the body, avoiding the resting state typical of normal cells. These cancerous cells have the ability to travel via the blood stream to other parts of the body where they continue to grow as metastases.

Lungs are a pair of large organs in your chest. They are part of your respiratory system. Air enters your body through your nose or mouth. It passes through your windpipe (trachea) and through each bronchus, and goes into your lungs. When you breathe in, your lungs expand with air. This is how your body gets oxygen. When you breathe out, air goes out of your lungs. This is how your body gets rid of carbon dioxide. Cancer begins in cells, the building blocks that make up all tissues and organs of the body, including the lungs. Normal cells in the lungs and other parts of the body grow and divide to form new cells as they are needed. When normal cells grow old or get damaged, they die, and new cells take their place. Sometimes, this process goes wrong.

New cells form when the body doesn’t need them, and old or damaged cells don’t die as they should. The buildup of extra cells often forms a mass of tissue called a growth or tumor. Tumors in the lung can be benign (not cancer) or malignant (cancer):

Lung cancer cells can spread by breaking away from a lung tumor. They can travel through blood vessels or lymph vessels to reach other parts of the body. After spreading, cancer cells may attach to other tissues and grow to form new tumors that may damage those tissues. When lung cancer spreads from its original place to another part of the body, the new tumor has the same kind of abnormal cells and the same name as the primary (original) tumor.

A. Preprocessing Techniques

Lung images are different to interpret, and a preprocessing of the image is necessary to improve the quality of the images and more reliable. Pre- processing is always a necessary whenever a pattern to be mined is noisy, inconsistent of incomplete and preprocessing significantly improves the effectiveness of the classification accuracy of the image. In the digitization process, noise could be introduced that needs to be reduced by applying some images processing techniques.

In addition at the time that the lung are taken the condition of illumination are generally different. The common characteristics of the images like as known noise, poor image contrast, in homogeneity, weak boundaries and unrelated parts will affect the content of the medical images, this problem rectified by preprocessing techniques. The preprocessing are fundamental steps in the medical image processing to produce better image quality for segmentation and feature extraction.

The aims of the image enhancement techniques are the quality improvement of a given image. The modified image demonstrates certain features that are better than in the original image essentially the contrast. Improve the ultrasound image for a better medical image interpretation, mainly sharpening boundaries between differently image regions (edges).

Preprocessing involves removal of artifacts and labels, filtering the images and removal of pectoral muscle are necessary. It includes gray-level and contrast manipulation, noise reduction, background removal, edge crisping and sharpening, filtering, interpolation and magnification, pseudo-coloring, and so on. Reducing the noise and blurring and increasing the contrast range could enhance the image.

In the dissertation, we used two types of filtering techniques and compared between them with some error measurements. They are

1. Wiener Filter
2. Median Filter
3. Max Filter

1.Wiener Filter

Wiener filter has been institute to be very powerful in eliminating noise from two dimensional signals without blurring edges. This creates it particularly appropriate for enhancing lung images. The wiener filter considers each pixel in the image in turn and looks at its nearby neighbors to decide whether or not it is representative of its surroundings. Instead of simply replacing the pixel value with the mean of neighboring pixel values, it replaces it with the median of thosevalues.Maintaining the integrity of the specifications.

$$\mathrm{w}(f_1,f_2) = \frac{H*(f_{1f_{2,}})S_{xx}(f_{1,f_2})}{H(f_{1,f_2})2s_{xx}(f_{1,f_2})+s_{nn}(f_{1,}f_2} \quad (1)$$

2.Median Filter

The best known order-statistics filter is the median filter, which, as its name implies, replaces the value of a pixel by the median of the gray levels in the neighborhood of that pixel:

$$f(x,y) = median(s,t)xy\{g(s,t)\} \quad (2)$$

3.Max Filter

Find the brightest points in an image.Finds the maximum value in the area encompassed by the filter.Reduces the pepper noise as a result of the max operation.The 100th percentile filter is max filter. Check the 50th percentile filter i.e the median filter.

$$F(x,y) = \frac{max}{s,t\epsilon s_{x,y}}\{g(s,t)\} \quad (3)$$

II. SEGMENTATION AND FEATURE EXTRACTION

Image segmentation is the process of partitioning the digital images into multiple segments. Segmentation is the process of straight forward approach. Segmentation is used to identify the object(or)other related information is called as segmentation. The goal of segmentation is to simplify and (or)change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects &boundaries (lines & curves) in images.

A. Thresholding

Thresholding is the simplest method of image segmentation for a gray image;thresholding can be used to create binary images. During the thresholding process, individual pixels in an image are marked as "object" pixels if their value is greater than some threshold value (assuming an object to be brighter than the background) and as "background" pixels otherwise.

Threshold algorithm

Input: Image after thinning border, threshold T

Output: Reconstruction of Lungs border

Procedure

For each pair of border pixels

Step 1: Compute d1

d1 = (x1-x2)2

Step 2: Compute d2

d1 = (x1-x2)2

Step 3: Compute ratio r

R = d2 / d1

Step 4:

If (r ≥ T)

Pixel is candidate for reconstruction

Else

Skip pixel

End if

B. Multi Threshold Otsu's Method

Otsu's method is used to automatically perform clustering-based image thresholding or the reduction of a gray level image to a binary image. The algorithm assumes that the image contains two classes of pixels following bi-modal histogram (foreground pixels and background pixels), it then calculates the optimum threshold separating the two classes so that their combined spread (intra-class variance) is minimal. The extension of the original method to multi-level thresholding is referred to as the Multi Otsu method.

In Otsu's method we exhaustively search for the threshold that minimizes the intra-class variance (the variance within the class), defined as a weighted sum of variances of the two classes:

$$\sigma_\omega^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \tag{4}$$

Multi threshold algorithm

1. Compute histogram and probabilities of each intensity level
2. Set up initial w_i and μ_i
3. Step through all possible thresholds t= 1… maximum intensity
    1. Update w_i and μ_i
    2. Compute $\sigma_b^2(t)$
4. Desired threshold corresponds to the maximum σ_b^2(t)

5. You can compute two maxima (and two corresponding thresholds). $\sigma_{b1}^2(t)$is the greater max and σ_{b2}^2 (t) is the greater or equal maximum

6. Desired threshold =$\frac{threshold1+threshold2}{2}$

Weights $w_i$are the probabilities of the two classes separated by a threshold t andσ_i^2 variances of these classes.

C. Adaptive Global Thresholding

Lung image is converted into gray image, and road region is segmented from that image using histogram analysis. Adaptive global thresholding is applied to remove the non-cancer pixels and segment-approximated cancer region from the lung image. The histogram of the lung image is analyzed and divided into four main sections to obtain the desired threshold value for segmentation.

This approach is used in the case where single value thresholding will not function properly, since the threshold value of pixel depends on its position within an image. Therefore, this technique is called as adaptive global thresholding. From this technique, approximated road regions are identified. The pixels that lie in that region are assigned to value 1 and all the remaining pixels are made to 0. Now, gray image is converted into binary image in which road regions appeared in white and all other pixels appeared in black.

D. Feature Extraction

This paper aims to focus on feature extraction. The first order statistical features such as Mean, Variance, Standard Deviation, Skewness, and Kurtosis and the second order statistical features are extracted from the texture description methods GLCM. Feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed then the input data will be transformed into a reduced representation set of features. Transforming the input data into the set of features is called feature extraction. If the features extracted are 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. The word “data” is plural, not singular.

A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, G, in the image. The matrix element P(i, j | Δx, Δy) is the relative frequency with which two pixels, separated by a pixel distance (Δx, Δy), occur within a given neighborhood, one with intensity i and the other with intensity j. One may also say that the matrix element P(i, j | d, θ) contains the second order statistical probability values for changes between gray levels i and j at a particular displacement distance d and at a particular angle (θ).

III. Lung Cancer Classification

A. Artificial Neural Networks

Artificial neural networks are very different from biological networks, although many of the concepts and characteristics of biological systems are faithfully reproduced in the artificial systems. Artificial neural nets are a type of non-linear processing system that is ideally suited for a wide range of tasks, especially tasks where there is no existing algorithm for task completion. ANN can be trained to solve certain problems using a teaching method and sample data. In this way, identically constructed ANN can be used to perform different tasks depending on the training received. With proper training, ANN are capable of generalization, the ability to recognize similarities among different input patterns, especially patterns that have been corrupted by noise.

A neural network for handwriting recognition is defined by a set of input neurons which may be activated by the pixels of an input image. After being weighted and transformed by a function (determined by the network's designer), the activations of these neurons are then passed on to other neurons. This process is repeated until finally, an output neuron is activated. This determines which character was read. Like other machine learning methods – systems that learn from data – neural networks have been used to solve a wide variety of tasks that are hard to solve using ordinary rule-based programming, including computer vision and speech recognition.

B. Support Vector Machine

The objective of any machine capable of learning is to achieve good generalization performance, given a finiteamount of training data, by striking a balance between the goodness of fit attained on a given training dataset and the ability of the machine to achieve error-free recognition on other datasets. With this concept as the basis, support vector machines have proved to achieve good generalization performance with no prior knowledge of the data. The principle of an SVM is to map the input data onto a higher dimensional feature space nonlinearly related to the input space and determine a separating hyper plane with maximum margin between the two classes in the feature space.

A support vector machine is a maximal margin hyper plane in feature space built by using a kernel function in gene space. This results in a nonlinear boundary in the input space. The optimal separating hyper plane can be determined without any computations in the higher dimensional feature space by using kernel functions in the input space. Table I:shows the classification accuracy for support vector machine and artificial neural network.

Table I. Classification accuracy for SVM and ANN

Data set index	**Database Name**	**Classification Accuracy**	
		SVM	**ANN**
1	GLCM_0	85	90
2	GLCM_45	80	88
3	GLCM_90	81	84
4	GLCM_135	79	82

IV. EXPERIMENTAL RESULTS ANDANALYSIS

Evaluation measuresare used to compare the actual result and the predicted result analysis. The accuracy of the system was analyzed using the confusion matrix. On the basis of the confusion matrix, performance values were calculated using the followings, where TPR= True positive rate, FPR= False positive rate, TNR= True negative rate, FNR= False negative rate.

Table II describes the overall Average values of Precision, TPR, TNR, FPR and FNR. It contains average precision, sensitivity or True positive rate, specificity or True Negative Rate, False positive rate, False negative rate for all classifiers.

Table II. Average value of performance measure for various classifiers

Classification algorithm	**Average Precision**	**AverageSensitivity or TPR**	**AverageSpecificity or TNR**	**Average FPR**	**Average FNR**
ANN	0.9786	0.9773	0.9314	0.0706	0.0208
SVM	0.7423	0.7643	0.7955	0.3231	0.1180

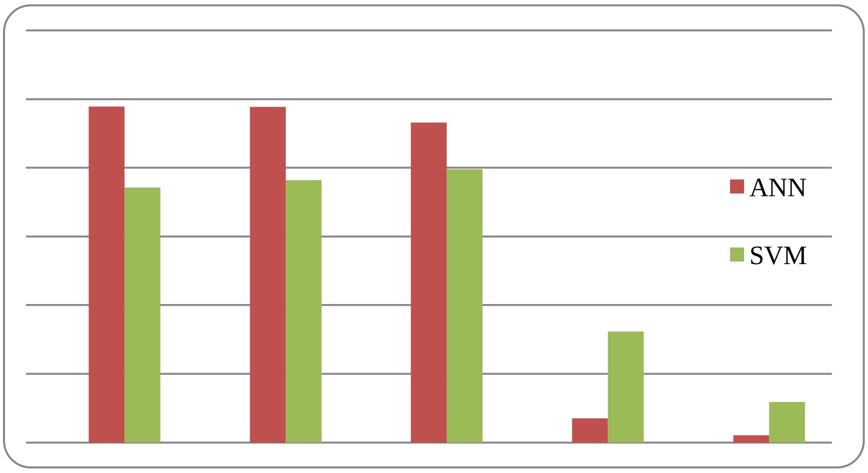

Figure 1. Average values of Precision, TPR, TNR, FPR and FNR for Classifiers.

Figure 1 shows the average value of precision true positive rate true negative rate false positive rate and false negative rate for classifiers.

Table III describes the overall Average values of Precision, TPR, TNR, FPR and FNR. It contains average precision, sensitivity or True positive rate, specificity or True Negative Rate, False positive rate, False negative rate for all classifiers.

Table III. Accuracy for classification

Classification algorithm	Accuracy
ANN	90
SVM	85

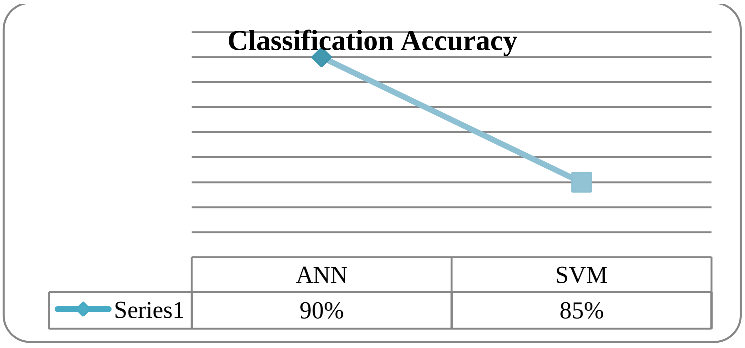

Figure 2. The accuracy of all classification methods

Figure 2 shows the accuracy of all classification methods, artificial neural network performance level about 90% of the lung cancer images, Support vector machine performance level about 85% of the lung cancer images.

V. CONCLUSION

In this paper lung CT scan images used to wiener filter, Max filter and Median filter is used to remove the noise and the validity measures PSNR , MSE used to compare with other noise removal methods used in this thesis. The lung segmented image extracted by using Otsu method, thresholding method, multi thresholding and adaptive global thresholding methods. The Artificial Neural Network (ANN) classifier and Support Vector Machines (SVMs) are used for image classification. The ANN gives better results compare than SVM. In future, some other filtering methods and segmentation methods may be used to improve the accuracy of the classification.

REFERENCES

[1] Mokhled S. Al-Tarawneh, “Lung Cancer Detection Using Image Processing Techniques”, ISSN 1583-1078, Issue 20, January-June 2012.
[2] Mr.Vijay A.Gajdhane , Prof. Deshpande L.M., “Detection of Lung Cancer Nodule on Computed Tomography Images by Using Image Processing”, Volume 3, Issue 7, July 2014.
[3] National Cancer Institute, "The cancer imaging archive”, "https:llwiki.cancerimagingarchive.net/display/Public/LIDC-lDRI, September 2011.
[4] D.Sharma & G.Jindal, “Identifying lung cancer using image processing techniques”. In International Conference on Computational Techniques and Artificial Intelligence, Vol. 17, pp. 872-880, ICCTAI’2011.
[5] A.Anam Tariq, M.U.Akram,& M.Y.Javed,. “Lung nodule detection in CT images using neuro fuzzy classifier”. In Computational Intelligence in Medical Imaging (CIMI), 2013 Fourth International Workshop, pp. 49-53, 2013.
[6] A.Chaudhary, & Singh,”Lung cancer detection on CT images by using image processing”. International Conference on Computing Sciences (ICCS), pp. 142- 146, 2012.

[7] Hashemi, A., Pilevar, A. H., & Rafeh, R. “Mass Detection in Lung CT Images Using Region Growing Segmentation and Decision Making Based on Fuzzy Inference System and Artificial Neural Network”. International Journal of Image, Graphics and Signal Processing 2013.
[8] Anand, S. V. “Segmentation coupled textural feature classification for lung tumour prediction”. In Communication Control and Computing Technologies (ICCCCT), 2010 IEEE International Conference, pp. 518-524, 2010.
[9] Lee, S. L. A., Kouzani, A. Z., & Hu,E.J.” A random forest for lung nodule identification”., pp. 1-5, 2008.
[10] Zhang, F., Song, Y., Cai, W., Zhou, Y., Shan, S., & Feng, D. “Context curves for classification of lung nodule images”. In Digital Image Computing: Techniques and Applications (DICTA), International Conference, pp. 1-7, 2013.
[11] S.Sivakumar, & C.Chandrasekar “Lung nodule detection using fuzzy clustering and support vector machines”. International Journal of Engineering and Technology, 5(1), 179-185, 2013.
[12] Jaffar, M. A., Hussain, A., Nazir, M., Mirza, A. M., & Chaudhry, A. “GA and morphology based automated segmentation of lungs from Ct scan images”. In Computational Intelligence for Modelling Control & Automation, 2008 International Conference, pp. 265- 270, 2008.
[13] Madero Orozco, H., Vergara Villegas, O. O., De Jesus Ochoa Dominguez, H., & Cruz Sanchez, V.G. “Lung Nodule Classification in CT Thorax Images Using Support Vector Machines”, 2013.
[14] PATIL, D. S., & Kuchanur, M. B.”Lung cancer classification using image processing”. International Journal of Engineering and Innovative Technology (IJEIT) Volume-2, 2012.
[15] Naresh, P., & Shettar, R. “Image Processing and Classification Techniques for Early Detection of Lung Cancer for Preventive Health Care”A Survey. Int. J. of Recent Trends in Engineering & Technology, 2014.
[16] Gajdhane, M. V. A., & Deshpande, L. M. “Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques”.
[17] Gomathi, M., & Thangaraj, P. “A computer aided diagnosis system for lung cancer detection using support vector machine”. American Journal of Applied Sciences, 7(12), 1532, 2010.
[18] Ilya Levner, Hong Zhangm,“Classification driven Watershed segmentation”, transactions on image processing VOL. 16, NO. 5, 2007.
[19] Sunil Kumar,“Lung Segmentation using Region Growing Algorithm”, International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, 2014.
[20] Fatma Taher, Naoufel Werghi and Hussain Al-Ahmad, “Bayesian Classification and Artificial Neural Network Methods for Lung Cancer Early Diagnosis”, 2012.