www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 4 Issue 11 Nov 2015, Page No. 14983-14986

Lung cancer detection and classification using Support Vector Machine

Niranjan Shukla* Aakash Narayane* Aniket Nigade* Krishnakumar Yadav* Mrs Harshada Mhaske*

"Department of Computer Engineering"
Savitribai Phule Pune University,

Pune, India

,

Abstract— Lung cancer is that the most significant reason behind cancer death for each man and woman. Early detection is incredibly necessary to reinforce a patient's likelihood for survival of carcinoma. For early and automatic respiratory organ tumor detection, we have a tendency to purpose a system that relies on textural options. There are 5 main phases concerned within the planned CAD system. They're image pre-processing, segmentation, feature extraction, classification of carcinoma as benign or malignant. The respiratory organ parenchyma region is segmental as a pre-processing as a result of the tumor resides inside the region. This reduces the search area over that we glance for the tumours, thereby increasing process speed. This additionally reduces the prospect of false identification of tumor. The image pre-processing is done by using fuzzy filter. Segmentation is done by using water shade algorithm, Textural options extracted from the respiratory organ nodules victimisation grey level co-occurrence matrix (GLCM). Then finally for classification, SVM classifier is utilized. This classifier is utilized to classify the nodules as Benign or Malignant.

***Keywords*— Fuzzy Filter, Water shade Algorithm, Grey level Concurrence Matrix (GLCM), SVM Classifier.**

I. INTRODUCTION

Lung cancer is that the leading reason behind tumour-related deaths within the world. At identical time, it seems that the speed has been steady increasing. Carcinoma is caused by the uncontrolled growth of tissues within the respiratory organ. The yank cancer society estimates that 213, 380 new cases of carcinoma within the U.S are diagnosed and a hundred and sixty, 390 deaths as a result of carcinoma can occur in 2007.Tobacco smoking is the main behind all cases.

Lung cancer is that the growth of a tumour, known as a nodule that arises from cells lining the airways of the system. The detection of carcinoma has been a tedious task in medical image analysis over the past few decades. Within the health trade, chest X-rays are thought-about to be the foremost wide used technique for the detection of carcinoma.

The image processing consists of mainly four steps. These steps are 1. Image Enhancement 2.Segmentation 3.Feature Extraction 4.Result. In this paper, we use the CAD system for early detection of lung cancer.

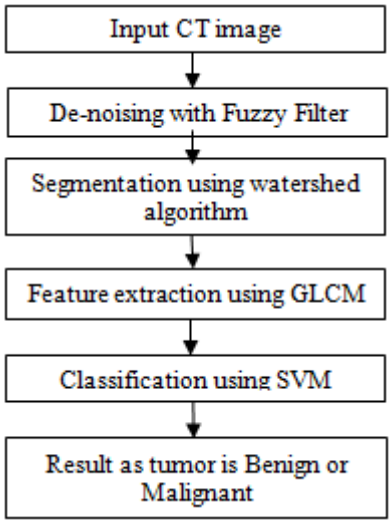

Figure 1: Entire diagnosis process of CAD system. [3]

In this paper lung Image is passed through different phases such as, De-noising with fuzzy filter, Image Segmentation using water shade algorithm, Feature Extraction using GLCM algorithm, and finally classifying data set of images using SVM classifier. Obtained result is Tumor which is Benign or Malignant. Step by step now we study this phases.

II. IMAGE DINOISING USING FUZZY FILTER

The Fuzzy filter is most commonly used in De-noising technique to filter the input CT-image. In that processing some steps are followed for filtering the image. These steps are shown in Fig 2:

DOI: 10.18535/Ijecs/v4i11.20

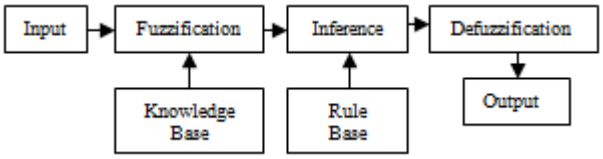

Figure 2 Fuzzy Filtering of image

Fuzzification: The art of converting standard expressions to fuzzy terms quantified by fuzzy membership functions.

Inference: logical thinking is main component of inference, which work primarily in one among 2 modes: forward chaining and backward chaining. Forward chaining starts with the renowned facts and assure new facts. Backward chaining is started with goals, and works backward to work out what facts should be declared in order that the goals may be achieved. The logical thinking engine uses IF-THEN rules. The final format of such rule is that if <logical expression> THEN <logical expression>.

Fuzzy filter works according to following steps:
Step 1: Image Acquisition using CT-scanner.
Step 2: Images are converted into gray scale images and Fuzzy filter is used to design the technique using FIS (Fuzzy Inference System) to improve the Image contrast.
Step 3: This algorithm is used to convert the image properties into fuzzy data and Fuzzy data into Defuzzification.
Step 4: Image denoising and enhancement done by fuzzy filter. All models are applied to noised images.
Step 5: The mean square error (MSE) is defined as:

$$MSE = (1/MN) \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [X(i,j) - X_c(i,j)]^2 \quad \ldots. (1)$$

Where, X (i, j) = original image.

X_c (i, j) =compressed image.

Step 6: PSNR represents a measure of the peak error and is expressed in decibels. It is defined by:

$$PSNR = 10\log_{10}((255^2) / MSE) \qquad \ldots..(2)$$

After performing Filtration on input image we get image without noise which is shown in the fig3.

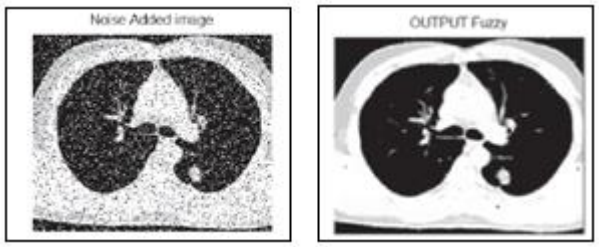

Figure 3: Result of denoising [1]

III. Segmentation using Marker Based Watershed algOrithm

The segmentation of lungs from CT images is a critical step in any Computer Aided Design system which leads to the early diagnosis of lung cancer and also in other pulmonary diseases. The segmentation of lung CT image is a very challenging problem due to in-homogeneity in the lung region and pulmonary structures having similar densities such as veins, arteries , bronchioles, and different scanning protocols and scanners.

The success of this technique can be measured in terms of accuracy, time complexity, processing time, and efficiency level. the tumor in lung form darker regions in CT images compared to other parts of the chest such as the heart and the liver.
Image segmentation is done by following techniques,
1. Marker based Watershed algorithm and thresholding.
The concept of watershed transformation was initially introduced as a tool for segmentation of grayscale images by S. Beucher and C. Lantu´ejoul in the late 70's.Now it is used as a fundamental step in many powerful segmentation procedures. Immersion simulations based watershed algorithms were proposed by F. Meyer and L. Vincent in the early 90's.
Watershed algorithm extracts the region of interest indicating the presence of objects or background at specific image locations.[2]
There are two ways for approaching segmentation of image. The first one is boundary based and detects local changes. The other one is region based and search for region and pixel based.
Algorithm for marker-based watershed algorithm:

Input:
Noise free Lung CT image from medical database.

output:
Segmented/ Partitioned image.

Step:
Begin

-Convert the CT image into binary image.
-Compute the Euclidean distance between each pixel of the binary image

-Identify the watershed regions from the image for each watershed region

-label each pixel with a value
-label with value 0 indicates it does not belong to unique threshold region

-Watershed pixels are pixels with value 0
-end for
-Create a sobel horizontal filter for edge detection.
-Filter the input image with the sobel operator to get the partitions.
end

DOI: 10.18535/Ijecs/v4i11.20

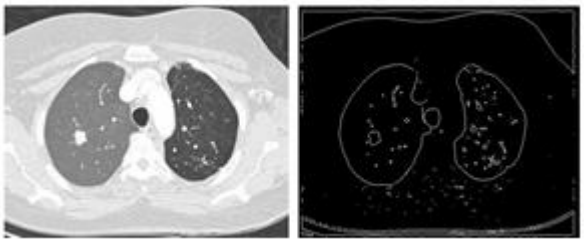

Figure 4: (a) Original images, (b) effects of watershed segmented

IV. FEATURE EXTRACTION USING GREY LEVEL CO-OCCURRENCE MATRIX

Gray Level Co-occurrence Matrix (GLCM) is one of the most popular ways to describe the texture of an image. The extracted ROI can be distinguished as either cancerous or not using their texture properties. A GLCM denote the second order conditional joint probability densities of each of the pixels, which is the probability of occurrence of grey level i and gray level j among a given distance 'd' and on the direction 'θ'.7 options area unit thought-about for planned technique.

1. Area: It provides the particular variety of pixels within the ROI.
2. Convex Area: It provides the quantity of pixels in convex image of the ROI.
3. Mean: it's the proportion of the pixels within the convex hull that also are within the ROI.

$$\text{Mean} = \mu_i = \sum_{i,j=0}^{N-1} i\,(P_{i,j}) \qquad \ldots. (1)$$

4. Energy: it's the summation of square parts within the GLCM and its price ranges between zero and one.

$$\text{Energy} = \sum_{K=0}^{N} P^2(i,j) \qquad \ldots\ldots (2)$$

5. Contrast: it's the live of distinction between AN intensity of constituent and its neighboring pixels over the total ROI. Where, N is that the variety of various grey levels.

$$\text{Contrast} = \sum_{i,j=0}^{N-1} P_{i,j}\,(i-j)^2 \qquad \ldots..(3)$$

6. Homogeneity: it's the live of closeness of the distribution of parts within the GLCM to the GLCM of every ROI and its price ranges between zero and one.

$$\text{Homogeneity} = \sum_{i,j} \frac{P(i,j)}{1+|i-j|} \qquad \ldots.. (4)$$

7. Correlation: it's the live correlation of constituent to its neighbor over the ROI.

V. IMAGE CLASSIFICATION USING SVM CLASSIFIER

SVM introduced by Cortes is mostly used for classification purpose. SVMs area unit economical learning approaches for training classifiers supported many functions like polynomial functions, radial basis functions, neural networks etc. it's thought-about as a supervised learning approach that produces input-output mapping functions from a labelled training dataset. SVM has vital mental capacity and thus is broadly speaking applied in pattern recognition. SVMs area unit universal approximators that rely upon the applied math and optimizing theory. The SVM is especially placing the biological analysis and capable to handle noise, massive dataset and enormous input areas.

The fundamental plan of SVM may be represented as follows:

a. Initially, the inputs area unit developed as feature vectors.

b. Then, by victimization the kernel perform, these feature vectors area unit mapped into a feature house.

c. Finally, a division is computed within the feature house to separate the categories of training vectors.

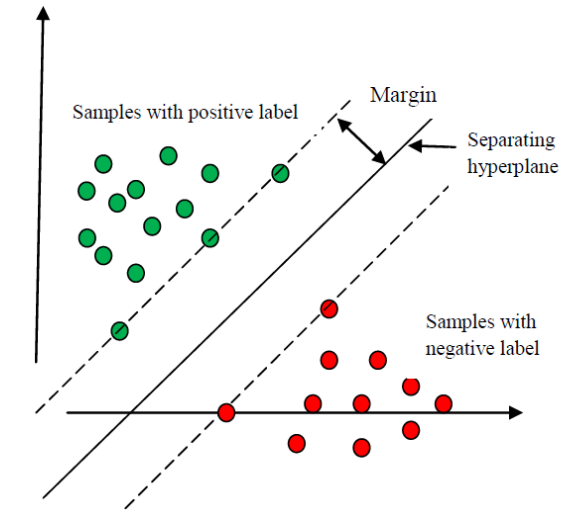

Figure 6: SVM classifier [5]

For binary classification SVM determines AN best Separating Hyperplane (OSH) that produces a most margin between 2 classes of knowledge. to form AN OSH, SVM maps knowledge into a better dimensional feature house and carries out this nonlinear mapping with the assistance of a kernel operate. Then, SVM builds a linear OSH between 2 categories of knowledge within the higher feature house. knowledge vectors that square measure nearer to the OSH within the higher feature house square measure referred to as Support Vectors (SVs) and embrace all knowledge necessary for classification. A kernel operates and also the parameters ought to be elite for constructing the support vector machine classifier. Here, 3 kernel functions square measure accustomed construct SVM classifiers:

a. Linear kernel operate
b. Polynomial kernel operate
c. Radial basis operate

The most used kernel operate for SVM is Radial Basis operate (RBF) owing to their localized and finite responses across the complete vary of real coordinate axis. The classification accuracy of RBF kernel was high; additionally, the bias price and also the error rate of RBF kernel were little compared to different kernels.

Cross-validation: Cross-validation could be a technique for associate degreealyzing however the results of a applied math analysis can generalize to an freelance information set. It's utilized in things; wherever the goal is prediction, and to estimate however accurately a prognosticative model can perform in apply. One spherical of cross-validation includes

dividing or partitioning a sample of information into complementary subsets, playing the analysis on one set (training set), and collateral the analysis on the opposite set (validation set or testing set). Multiple rounds of cross-validation are performed victimisation totally different partitions to scale back variability, and also the validation results are averaged over the rounds.

Confusion Matrix: Confusion matrix was wont to calculate the performance of the classifier. Figure one shows the confusion matrix. it's a particular table that helps to ascertain the performance of a learning formula. Every column of the matrix represents the expected category, and every row represents the particular category.

True Positive	False Negative
False Positive	True Negative

Figure 5: Confusion Matrix [4]

VI. CONCLUSION

In this paper, different phases of image processing were applied on Lung Nodules. From these different image processing techniques, the fuzzy filter will provide the efficient denoising. Segmentation done by marker based watershed algorithm, gives various region of image. GLCM is used to extract the different features of image and which takes less time for generating the result. This results are passed through SVM Classifier, which classifies the nodules as benign or malignant. SVM classifier provides 92.5% accuracy.

REFERENCES

[1] P. Yuvarani "Image Denoising and Enhancement for Lung Cancer Detection Using Soft Computing Technique"

[2] S. Beucher "The Watershed Transformation Applied to Image segmentation"

[3] Ms. Swati P. Tidke, Prof. Vrishali A. Chakkarawar "Classification of Lung Tumour Using SVM" (IJCER) Vol. 2 Issue.5.

[4] M.Gomathi, Dr. P. Thangaraj"An Effective Classification of Bening anf Malignant Nodules Using Support Vector Machine" (JGRCS)Vol 3 , No. 7, July 2012

[5] Mythily.A, Veena M.U "Segmentation and classification of lung tumour using Chest CT Image for Treatment Planning" ,(IJETT), Vol 7 Number 2-Jan 2014.

[6] Aneesh Agrawal, Abha Choubey and Kapil Kumar Nagwanshi (2011) "Development of Adaptive Fuzzy Based image Filtering Techniques for Efficient Noise Reduction in Medical Images", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (4) , 2011, 1457-1461.

[7] Florian Luisier and Michel Unser(2011) , "Image Denoising in Mixed Poisson- Gaussian Noise", IEEE Transactions on Image Processing,Vol.20,no.3.

[8] Mokhled S. AL-TARAWNE Leonardo Electronic Journal of Practices and Technologies ISSN 1583-1078 H ," Lung Cancer Detection Using Image Processing Techniques