International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 5, May 2015)

Automated Lung Nodule Detection and Classification using Morphological Segmentation and Texture features

Veena Puranikmath
Electronics and communication
Godutai engineering college for women
Kulburgi, Karnataka, India
Veenaip043@gmail.com

Dr. Lalita Y S
Electronics and communication
Appa institute of engineering
Kulburgi, Karnataka, India
Patil_lalita@gmail.com

Shivaganga Patil
Electronics and communication
Godutai engineering college for women
Kulburgi, Karnataka, India
shivagangabp@yahoo.co.in

***Abstract*:-Automated Lung Nodule analysis has been one of the most significant research areas in medical imaging. Chest CT scan images are difficult to analyze for nodules due to the presence of other artifacts like veins. Any computer aided technique is prone to misdetection due to visual similarity of the nodules with other visible artifacts in the CT scan images. Hence segmentation based nodule detection is not reliable. However nodes shows distinct features at higher image dimensions due to their inherent properties. Therefore we propose a novel work to first segment CT scan image to obtain all probable nodule candidates. Then we perform SVM based classification of the candidates based on image moments and texture features. Potential nodule candidates are marked and other candidates are omitted. The results show promising very low percentage of misdetection.**

Keywords: Lung Nodule Detection, Support Vector Machine, Morphology, GLCM

I. INTRODUCTION

Lung cancer is one of the major health concerns for many countries and especially in developing countries like India due lack of enough medical diagonestic tools for effective and early detection of the cancer. Computer Tomography (CT) is identified as the most reliable imaging tool for detection of lung cancer at an early stage.To add to it, analysis of Lung cancer in CT scan images need special skills by radiologists. Therefore computer aided diagnosis has been gathering momentum in this area. Major challenge in the screening for CT image is the presence of several other artefacts like blood vessels and bronchi, along with nodules. Also these artefacts often have similar features in terms of size and low level image features. Therefore more sophisticated image processing tools are needed for proper analysis of detected components in a CT image.

Several past works has been done towards lung nodule detection. The detection mechanism largely contains two phases: Detection and Classification. In detection stage possible nodule candidates in the CT images are identified using different threshold and segmentation techniques. Nodules are marked by trained and experienced radiologists. Marked nodules and detected candidates are compared. All the detected candidates are separated and their features are extracted. These features along with the known level are used to train a classifier which can then classify unmarked candidates from new image samples. The classifier proposed ranges from fuzzy classifier to neural network.

However nodule candidate classification is a binary classification problem. It is proved in the past work that support vector machines or SVM works best in the cases of binary classification problem. Also as the amount of nodule area and in fact other candidates are very small, there is a very high chance of their features being same or similar in the low level statistical domains. Several past works has focussed on extracting gray level concurrence matrix from the candidate and classify the candidates based on these features. One of the distinct problems with this approach is that it does not take into account of the fact that diagnosis of the nodules not only depends upon the nodule properties but at the same time also depends upon the properties of the neighbouring regions.

In this work we propose a pulmonary nodules detection technique in computed-tomography (CT) images. After extracting the probable nodule candidates (which also include blood vassals and bronchi) using a unique morphological segmentation technique we classify the candidates based on SVM classifier.

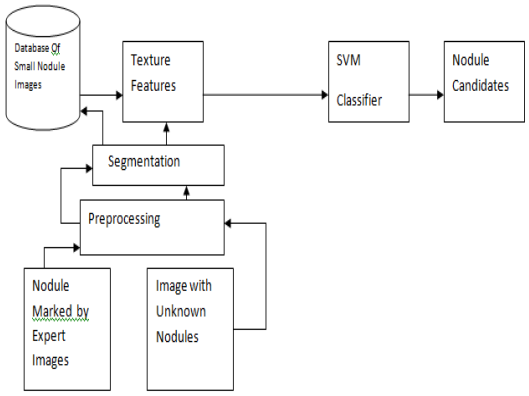


Figure 1.proposed system

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 5, May 2015)

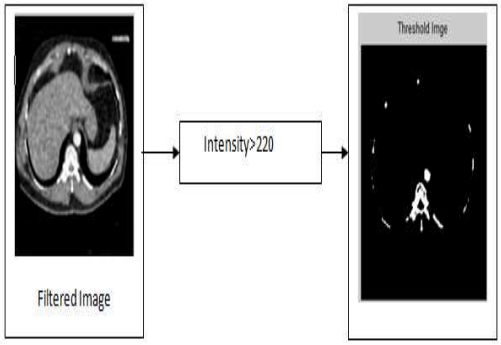


Figure2.thresholding stage

II. METHEDOLOGY

Pre-processing: Most of the past literature has focussed on smoothing the image with median filtering. The problem with median filtering is that in text marked images, several text artefacts is left out. Also median images tend to expand the texture towards outward direction which leads to misdetection. Thus we combine the results of median filtered image with Gaussian blurring in proposed work. First image is smoothed with median filter with kernel size 9x9. The raw CT image is also low pass filtered with Gaussian blurring by a Gaussian kernel of size 9x9 and sigma=3.5. We take the average of these two images to get a filtered image.

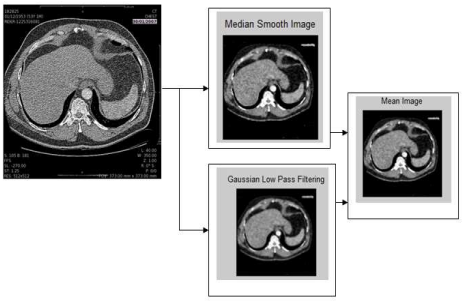


FIGURE 2: Pre-processing stage

By combining the blurring and smoothing processes, we get a more reliable image.

Segmentation: We adopt threshold based segmentation. Once threshold is applied we obtain a binary image with all probable candidates including nodules. However it can be seen that spinal cord area is also selected as one of the candidates due to high intensity. But the area of such an image is potentially high. Thus we calculate a bounding box of all the closure binary area and then we filter out the large areas. The area detected by bounding box is then superimposed on the original image in order to get the probable nodule candidates.

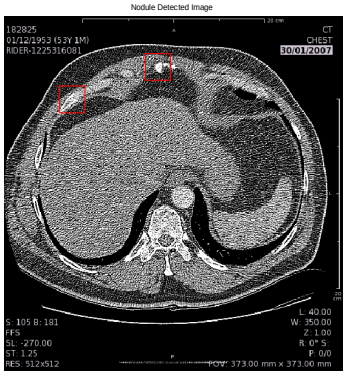


FIGURE 4: Probable Nodule Candidates

Training: Every area marked by segmentation process is extracted as independent image as shown in figure below. It can be seen that it is impossible to identify the nodule by comparing both the segments with intensity values. A nodule candidate will have certain correlation with its neighbours and intensity distribution in the nodule itself. Therefore we first label these images depending upon expert markings. We use these segments to create a database of positive nodules and negative (candidates which are not nodules). We extract GLCM features from these candidate images.

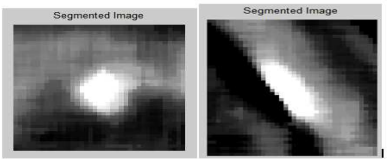


FIGURE 5: Segmented Image: (a) Actual Nodule (b) False Candidate

It can be seen that it is impossible to identify the nodule by comparing both the segments with intensity values. A nodule candidate will have certain correlation with its neighbours and intensity distribution in the nodule itself. Therefore we first label these images depending upon expert markings. We use these segments to create a database of positive nodules and negative (candidates which are not nodules). We extract GLCM features from these candidate images.

A **co-occurrence matrix** is one that is defined as a probabilistic distribution of index colors with respect to their neighbors. A co-occurrence matrix **C** is defined over an **n x m** image **I**, parameterized by an offset **(Δx, Dy)**, as:

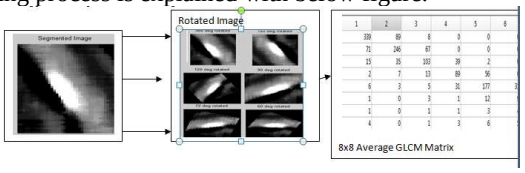
$$C_{\Delta x,\Delta y}(i,j)=\sum_{p=1}^{n}\sum_{q=1}^{m}\begin{cases}1, & \text{if } I(p,q)=i \text{ and } I(p+\Delta x,q+\Delta y)=j\\ 0, & \text{otherwise}\end{cases}$$

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 5, May 2015)

Any matrix can generate a co-occurrence matrix, though their main application is to measure the texture patterns in the images.

(Δx,Δy) parameters make the GLCM sensitive to image rotation. Therefore images are rototated in 6 equal angles and their GLCM values are aggregated in order to nullify the effect of rotation.

Training process is explained with below figure.



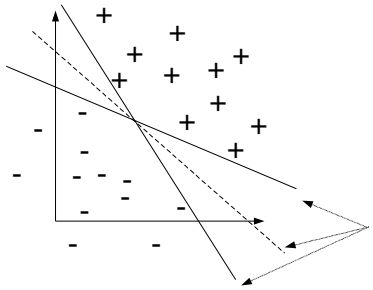
Several literatures extracts statistical features like Energy, Entropy, Dissimilarity, homogeneity features from this matrix and use them as features for classification.

However as SVM by itself creates a high dimensional feature space from low dimensionality data, instead of extracting statistical features from the matrix we reshape the matrix into a row matrix and use 64 features as feature for classifier.

Classification

Assume training data $D = \{(\vec{x}_i, y_i), i = 1...N\}$

with $y_i \in \{-1,+1\}$ is separable by a hyper plane.

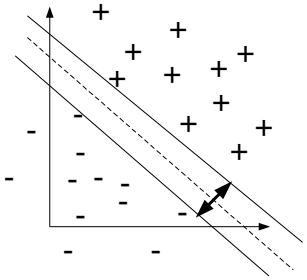


Question: What is the best linear classifier of the type

$$f(\vec{x}) = \vec{w}^T \vec{x} + b \quad (= w_1 x_1 + w_2 x_2 + w_3 x_3 + ... + w_n x_n + b)$$

.

While there can be an infinite number of hyper planes that achieve 100% accuracy on training data, the question is what hyper plane is the optimal with respect to the accuracy on test data?



Common sense solution: we want to increase the gap (margin) between positive and negative cases as much as possible. The best linear classifier is the hyper plane in the middle of the gap.

Given f(x), the classification is obtained as

$$\hat{\square} = sign(f(\square)) = \begin{cases} +1 & f(\square) > 0 \\ -1 & f(\square) < 0 \end{cases}$$

Note: Different **w** and **b** can result in the identical classification. For example, we can apply any scalar α such that:

$$\hat{\square} = sign(\alpha(\square\square + \square)) = sign(\square\square + \square)$$

Therefore there are many identical solutions.

+1 and -1 are the labels for Nodule and Non Nodule candidates respectively. The SVM first constructs a hyper plane with training data followed by classifying the unknown nodules.

III. RESULTS

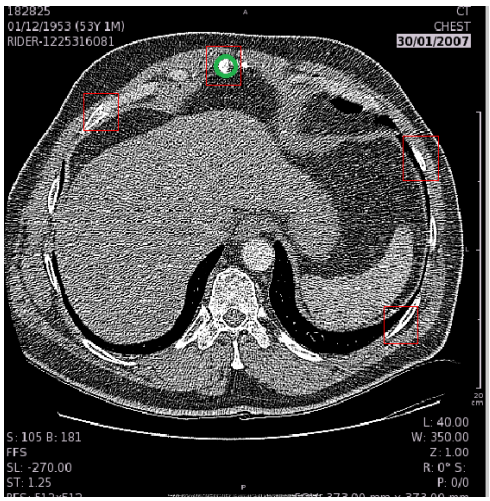


FIGURE 6 . Nodule marked after classification.

We perform the test on 50 images from Tata Cancer Research centre CT images. Also compared the results of proposed technique with

nearest neighbour classifier. It was evident that by taking statistical features from GLCM matrix, performance of the nearest neighbour classifier improved significantly. However SVM performance was observed to be consistent. Out of 50 images with nodules, our system detected all fifty nodules correctly where as miss classifying 8 candidates in total. False positive was about 6% where as false negative was zero. Involves matching these features to yield a result that is visually similar.

IV. CONCLUSION

The proposed system performed following tasks on CT scan images.

1) Feature extraction and database implementation.

2) Query image feature extraction: the first step in the process is the extracting image feature to a distinguishable extent.

3) Similarity measurement: the second step involves matching these features to yield a result that is visually similar.

4) Comparison of performance for different feature vector and matching techniques using following parameters.

We were able to successfully suppress the text part and filter the noisy images. The segmentation method in all cases identified nodule part along with other candidates. The false candidate selection was successfully eliminated using classification technique.

REFERENCES

[1]. Bach, P.; Mirkin, J.; Oliver, T.; Azzoli, C.; Berry, D.; Brawley, O.; Byers, T.; Colditz, G.; Gould, M.;Jett, J.; et al. Benefits and harms of CT screening for lung cancer. Context 2012, 307, 2418–2429.

[2]. Messay, T.; Hardie, R.; Rogers, S. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Med. Image Anal. 2010, 14, 390–406

[3]. Choi, W.J.; Choi, T.S. Genetic programming-based feature transform and classification for the automatic detection of pulmonary nodules on computed tomography images. Inf. Sci. 2012,212, 57–78.

[4]. S. Ozekes, O. Osman and O.N. Ucan. ”Nodule detection in lungs region that’s segmented using genetic cellular neural networks and 3D template matching with fuzzy rule based thresholding”, Vol. 9, No. 1, pp. 1-9, Feb.2008.

[5]. M.G. Nedo, M.J. Carreira, A. Mosquera, and D. Cabello, ”Computer Aided Diagnosis: A neural-network-based approach to lung nodule detection”, IEEE Transactions on Medical Imaging, Vol. 17, No. 6, pp.872-880, Dec. 1998