

# **Lung Nodule Detection Based on Semi Supervised Classification**

S.Ramya Preethi<sup>1</sup>, Prof. R.Vijayalakshmi<sup>2</sup>, P.Deepa<sup>3</sup>

ME-CSE (Final Year), Muthayammal Engineering College, Rasipuram, India<sup>1</sup>

Professor and Head, Muthayammal Engineering College, Rasipuram, India<sup>2</sup>

Assistant Professor, Muthayammal Engineering College, Rasipuram, India<sup>3</sup>

**ABSTRACT:** Lung cancer is a major cause of cancer related deaths. Thus the identification of lung nodule is essential part for screening and diagnosis of lung cancer. The classification of four type of lung nodule in low dose computed tomography scans. i.e., Well-circumscribed, vascularized, juxta-pleural, and pleural-tail. Thus classification of lung nodule is on three stages by combining the lung nodule with surrounding anatomical structures. First stage an adaptive graph patch based division is used to construct concentric multi level partition use of super pixel formulation. The second stage of the method is feature set designed to incorporate intensity, texture, and gradient information for image patch feature description use of scale-invariant feature transform, local binary pattern provides texture description of objects and Histogram of oriented gradients represents the local portions of object. And third stage is to classify the lung nodule based on semi supervised classifier with respect to feature descriptors and performance is compared with supervised classification.

**KEYWORDS:** scale-invariant feature transform, local binary pattern, Histogram of oriented gradients, semisupervised classifier.

## **I. INTRODUCTION**

Lung Cancer remains a leading cause of cancer related of humanity. Its cure rate is very low because it is usually detected at very late stages. Lung Cancer is characterized by uncontrolled cell growth originating in the lungs. The most common cause of Lung Cancer is smoking however in some cases it can also be attributed to genetic factors, gas, asbestos, smoking and air pollution. Lung cancers are detected by chest radiograph and CT scan will detect the lung nodule.

Lung nodules are small masses of tissue in the human lung and are somewhat common. They appear as round, white shadows on a chest X-ray or computerized tomography (CT) scan. Most nodules (more than 60 percent) are not cancerous. The Lung nodule can be distorted by nearby anatomical structures, such as vessels and the neighbouring pleura. Inter parenchymal lung nodules are more possible to be malignant than those connected with surrounding structures. If a lung nodule is new or has transformed in size, shape or appearance, your specialist may suggest additional testing such as a CT scan, positron emission tomography (PET) scan, and bronchoscopy or tissue biopsy to determine if the nodule cancerous or not.



**Fig: 1 CT scan of lung nodule**

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

The nodules are divided into a four types. Well-circumscribed (W) with the nodule located centrally in the lung without any connection to vasculature; vascularized (V) with the nodule located centrally in the lung but closely connected to neighboring vessels; juxta-pleural (J) with a large portion of the nodule connected to the pleural surface; and pleural-tail (P) with the nodule near the pleural surface connected by a thin tail.

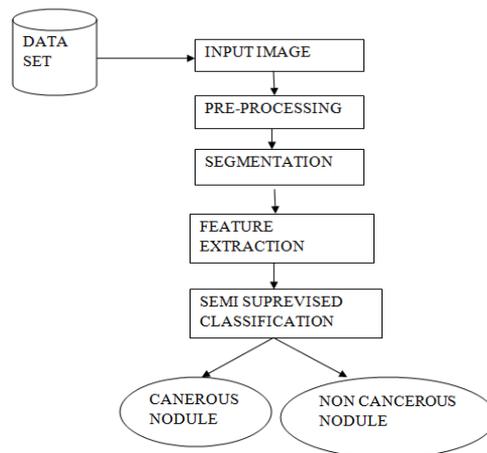


Fig: 2 steps to detect the lung nodule

## 1.1 LUNG NODULE DETECTION

The detection and segmentation of lung nodule will provide a limited data for lung nodule classification. The performance will be improved by better feature design and a classifier. The filter based feature extraction techniques are also used to remove the noise in the image. In that the anisotropic diffusion will reduce the noise in the image without removing significant parts of the image content, typically edges lines or other details that are important for the interpretation of the image.

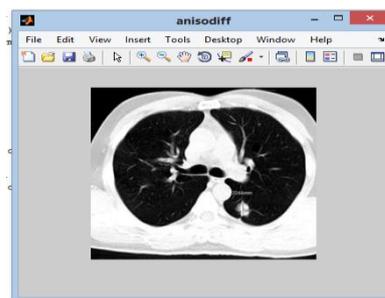


Fig: 3 Anisotropic Diffusion

## II. SEGMENTATION

FCM clustering algorithm will allows the one piece of data belong to two or more clusters and it will segment the images into four types of nodule with surrounding anatomical structures. Thus the lung nodules with surrounding anatomical structures are only present in certain type of nodule. For example the nodules are similar in shape and location in W and V. So it is difficult to classify the nodule. In order to classify the adaptive graph patch based division is used which segment the nodule images into number of patches and are used to construct the multilevel partition. In that the single patch will represent the one type anatomical structures. The superpixel formulation based on quick shift algorithm is used. It will provide the lung nodule images with surrounding anatomical structures.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

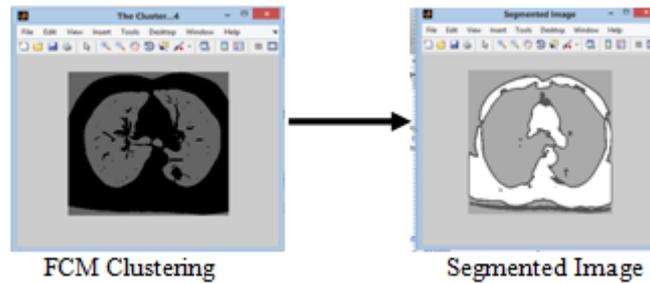


Fig: 4 Segmentation

### III. FEATURE EXTRACTION

Next to describe the patch numerically, and translate into feature vector. The feature set will extract the intensity, texture, and gradient information of the image. The Scale Invariant Feature Transform (SIFT) is invariant to image translation, scaling, rotation, illumination changes, to local statistical data. For texture the description of object Local Binary Pattern (LBP) which provides the rotation invariant property. Textural properties can be calculated from GLCM to understand the details about the image content. It will used to find contrast, correlation, dissimilarity, energy, entropy, and homogeneity. Histogram of Oriented Gradients will provide the gradient orientation in local portion of the image. It will show the effectiveness of individual image in order to get better the performance.

### IV. CLASSIFICATION

Next to feature extraction, the classifier will able to classify the nodule whether the nodule is cancer nodule or not. The number of classification techniques is used to classify the nodule with surrounding anatomical structures. Among that we propose the semi supervised classifier is to use the both of labeled and unlabeled images based on the contextual latent semantic analysis. It will calculate the probabilistic estimations for the rank based unlabeled relevant images.

#### 1.4.1 SUPERVISED CLASSIFICATION

Supervised classifications use only the labeled information to classify the images. The training data consist of a set of training examples. In supervised classification, each example is a pair consisting of an input object (typically a vector) and a required output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces a secondary function, which can be used for mapping new examples.

Here the lung nodules are classified use of CT images. First the patch based division is to segment the image into number of patches use superpixel formulation based on quick shift algorithm.

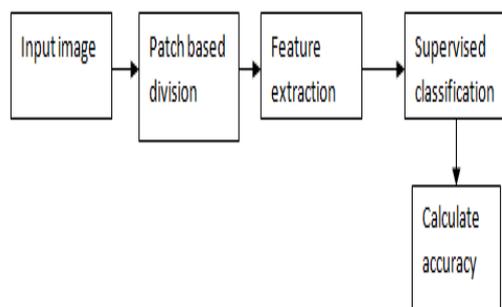


Fig: 5 Steps to detect the nodule based on supervised classification

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

## 1.4.2 SEMI SUPERVISED CLASSIFICATION

In semi supervised classifiers for lung nodule classification is to use unlabeled images into improve generalization deriving graph-based distances that emphasize label and unlabeled images classification propose a graph-based semi-supervised learning method exploiting the Unlabeled image classification . An adaptive graph patch-based division is used to construct concentric multilevel partition; then, a new feature set is designed to incorporate intensity, texture, and gradient information for image patch feature description, and then a contextual latent semantic analysis-based semi supervised classifier is designed to calculate the probabilistic estimations for the rank based unlabeled relevant images.

It computes pair-wise similarity of subjects while considering intrinsic geometry of data distribution. Pair-wise similarity between subjects encodes relationship between labeled and unlabeled data and it is shown to improve classification accuracy in presence of unlabeled data

Our proposed regularization term is inspired by Laplacian- SVM (lapSVM), and particularly the linear version of . In lapSVM, samples (labeled and unlabeled) are considered as nodes of a graph. Every two nodes are connected via an edge and there is a weight associated to the edge determining how similar the two nodes (samples) are. Properties of this graph are used to define a regularization for lapSVM.

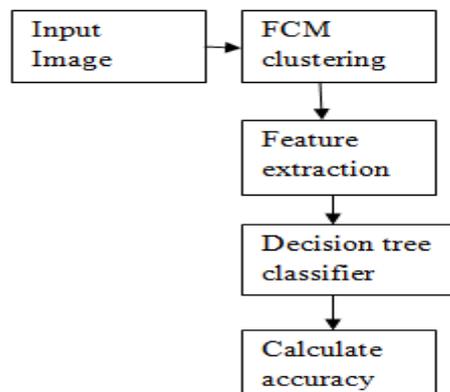


Fig 6 Steps for Semi Supervised Classification

## V. PERFORMANCE COMPARISON

We used the publicly available ELCAP (Early Lung Cancer Action Program) database for experiments. It will contain the 50 sets of low dose CT lung scan with 400 unduplicated lung nodules annotated at the centroid. Among that only 20 images are used for training. The training includes three tasks (1) anisotropic diffusion will remove the noise in the image, (2) Grouping of similar data use of FCM clustering, (3) the global dictionary for contextual analysis classification could be obtained through whole dataset, and more images were incorporated for testing. The performance of the nodule image classification was done with the overall classification rate, as:

$$\text{Classification rate} = N_{\text{right}} / N_{\text{all}}$$

where  $N_{\text{right}}$  and  $N_{\text{all}}$  are the numbers of correctly labeled images and all images. Also, to illustrate the performance on images from different types, recall and precision were computed:

$$\text{Recall} = TP / (TP + FN)$$

$$\text{Precision} = TP / (TP + FP)$$

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

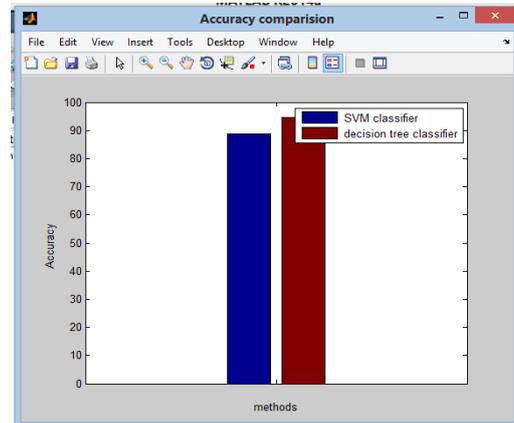


Fig 7 Accuracy Comparison

## VI. CONCLUSION

In this study a novel method for lung nodule image classification into four types, well-circumscribed, juxta-vascular, juxta-pleural and pleural-tail. The method includes three components: An anisotropic diffusion will remove the image noise without removing edges. And use of FCM clustering algorithm the image are segmented with surrounding anatomical structures. The feature descriptor such as SIFT, LBP, and HOG are designed to describe the processed image quantitatively according to the characteristics of each type of lung nodules use of GLCM matrix. For classification semi supervised classifier are used to classify the nodule and the performance are terms of sensitivity, accuracy, specificity, precision and recall.

## REFERENCES

- [1] prashant naresh, Dr. Rajashree shettar "early detection of lung cancer using neural network techniques", international journal of engineering research and applications, ISSN:2248-9622,vol.4, issue 8, august 2014 pp 78-83.
- [2] Fan Zhang, "lung nodule classification with multilevel patch based context analysis", IEEE transaction on biomedical engineering, vol. 61, no-4, april 2014.
- [3] Ms. Swati P. Tidke, "classification of lung tumor using SVM" international journal of computational engineering research" vol 2, issue 5 September 2012.
- [4] Wail A.H Mousa and Mohammad A. U Khan "Lung Nodule Classification Utilizing Support Vector Machines" IEEE international geosciences and remote sensing symposium" June 2002.
- [5] Anam Tariq , M. Usman Akram and M. Younus Javed " Lung Nodule Detection in CT Images using Neuro Fuzzy Classifier" IEEE 2013 Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI) – Singapore 2013.
- [6] M.Gomathi, Dr.P.Thangaraj "An Effective Classification Of Benign And Malignant Nodules Using Support Vector Machine" Volume 3, No. 7, July 2012 Journal of Global Research in Computer Science.
- [7] M. Gomathi and Dr. P. Thangaraj "Lung Nodule Detection using a Neural Classifier" IACSIT International Journal of Engineering and Technology, Vol.2, No.3, June 2010, ISSN: 1793-8236.
- [8] Pooja Kamavisdar1, Sonam Saluja2, Sonu Agrawal "A Survey on Image Classification Approaches and Techniques" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, January 2013
- [9] S. Liu, L. Zhang, W. Cai, Y. Song, L. Wen, and D. Feng, "A supervised multiview spectral embedding method for neuroimaging classification," in Proc. Int. Conf. Image Process., 2013, pp. 602–605.
- [10] J. Yao, A. Dwyer, R. M. Summers, and D. J. Mollura, "Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification," Acad. Radiol., vol. 18, no. 3, pp. 306–314, 2011.
- [11] F. Zhang, Y. Song, W. Cai, Y. Zhou, M. Fulham, S. Eberl, S. Shan, and D. Feng, "A ranking-based lung nodule image classification method using unlabeled image knowledge," in Proc. Int. Symp. Biomed. Imag., 2014.
- [12] F. Zhang, W. Cai, Y. Song, M.-Z. Lee, S. Shan, and D. Feng, "Overlapping node discovery for improving classification of lung nodules,"