

## Automated Detection of Cancerous Lung Nodule from the Computed Tomography Images

Mr. Vaibhav K. Likhitkar<sup>1</sup> Prof. Ujwalla Gawande<sup>2</sup> Mr. Kamal O. Hajari<sup>3</sup>

<sup>1</sup>Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India.

<sup>2</sup>Asst. Prof. Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India.

<sup>3</sup>Department of Information Technology Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India

---

**Abstract:** An automated Computer Aided Diagnosis (CAD) system is proposed in this paper for the detection of lung nodule from the computed tomography (CT) images. CT Images are more efficient and provide appropriate details on anatomical features of the lung nodule than X-ray. There are two types of lung nodule, cancerous and non-cancerous nodules. Lung nodules are classified with the help of CAD system. The CAD system can classify the lung nodule from CT images on the basis of nodule feature such as growth rate, density, shape and boundary of the nodule. These feature values are calculated in image processing tool by using image enhancement, segmentation, and feature extraction. The feature values are given as input to classifier to classify nodule.

**Keywords:** Lung nodule, cancerous lung nodule, non-cancerous nodule, computed tomography (CT), computer aided diagnosis (CAD) system.

---

### I. INTRODUCTION

There are several types of cancer from which lung cancer is becoming the prime factor in cancer deaths due to increasing rate of smoking and air pollution in different countries. According to survey report 31% of cancer deaths for males and 26% for females are caused by the lung cancer [1] there has been a great effort to improve the diagnosis and treatment of the lung cancer. Medical imaging is the main tool for cancer diagnosis and lot of researchers are focused on this area. Computed Tomography (CT) is considered to be the best modality for cancer diagnosis. Time and cost are two crucial factors in diagnosis of a lung cancer. The survival from lung cancer is directly related to lung nodule growth at its detection time. The earlier detection leads to the higher the chances of successful treatment. To overcome this scenario the automated detection of cancerous lung nodule system is required. It can eliminate the health risks and complications involved in invasive operations significantly and improve the chances for successful treatment of the lung cancer patient. It will help radiologist to improve the diagnosis efficiency by calculating the quantity of nodule growth in each stage accurately.

### II. TYPES OF LUNG NODULES

The Lung Nodules majorly classified into two types [2]:

1. Benign
2. Malignant

The fig. 2 shows the nodule growth from the benign stage to the last malignant stage. Benign is the primary stage of the nodules. It also shows the growth and boundary of each nodule at each stage.

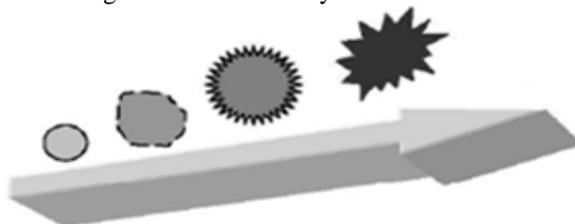


Fig.1 Nodule growth

If lung nodule area and density is Minimum then it is normal type of lung nodule. Small lung nodules are not diagnosis with the help of needle biopsy.

### **2.1 Benign**

Benign nodule is the type of non-cancerous lung nodule. If the nodule area is Minimum and density is Maximum then it is called benign nodule. Nodule sizes vary in range of 0.5 to 1. It has regular structure and hard core boundary without spine.

### **2.2 Malignant**

Malignant nodule is the cancerous lung nodule and it spread in affected area. It is most dangerous type of nodule. If the nodule area is Maximum and density is Minimum then it is called malignant nodule. Nodule sizes vary in range of 1 to 1.5. It has irregular structure and nodule boundary with spin. If lung nodule area is Maximum and density is Maximum then it is advanced lung cancer nodules.

## **III. LITERATURE SURVEY**

Lung cancer was not known before the advent of cigarette smoking. It was not even recognized as a disease until 1761 [3]. Different aspects of lung cancer were described further in 1810 [4]. In 1929 German; physician Fritz Lickint recognized the link between smoking and lung cancer [5]. The CT described by Bhandauria and Dewal[6] is widely used in diagnosing different diseases. The computer-aided detection (CAD) scheme is efficient for the better diagnosis for large volume of CT image data. Zhenghao Shi et al [7] used the Laplacian of Gaussian filter for enhancing nodules in a chest radiograph. It used the high intensity transmission and then applied a LOG filter to calculate the differences in contrast within inside and outside region of interest. Vivekanandan D et al [8] applied edge based segmentation using snakes on CT images. A snake was used as the energy function in term of internal energy and force that acts on image. Anita chaudhary et al [9] was performed segmentation by using thresholding and watershed segmentation techniques. The thresholding segmented image requires smaller storage space and it have fast processing speed and it is easy in manipulation. Watershed segmentation extracts the object from background with the help of seeds. But it is a bit difficult to differentiate region of interest so our implementation technique is more suitable and accurate. Jia Tong et al. [10] described a new CAD scheme for detecting lung nodules. In their approach the lung region was segmented from the CT data using adaptive threshold algorithm and then nodule features were extracted by Rule-Based Classification (RBC). The RBC was used to distinguish nodules type and they claim the accuracy of 85%. The alternative and promising method was suggested by Li Shen et al. [11] to detect lung cancer at an earlier stage. In their method nodule segmentation was performed in a nodule region to extract nodule boundary. They used thin-plate spline interpolation for nodule registration along the temporal dimension. Their resulting spatio-temporal model was used for distinguish between cancer and benign nodules. Maciej Dajnowiec et al. [12] proposed a method for automatic segmentation of the lung region. In his method optimum threshold value was calculated from the image of the data set. This threshold values were used to differentiate the organ and solid tissue from air region of the lung. After that the images was sequentially checked from the beginning until the two objects whose size exceeds 1% of the total pixel were found. It was considered as base slide. Then remaining region was obtained by applying region growing technique. This final result was multiplied with the original CT images to extract the lung portion of each slide. But it gives whole lungs left and right lobes with nodules region. But in our project; in the final step we get the lung nodule segmented part. CT technology was also adopted by K.Devaki et al. [13] to acquire high resolution of the lungs in a single breath hold. Analysis of these large volumes of image data was manually performed, which was time consuming task. So they used automated analysis of CT scan images by addressing segmentation result of various lungs anatomical structures. Noriyasu Homma et al [14] used N-quoit filter to find ROI from the CT images. To extract the lung nodule they binaries original image and then calculated mean, variance and entropy of intensity. After that they applied a Gabor filter to find orientation output for new feature. But drawback is that they find nodules manually and calculated their values. Thresholding classifier was proposed by FatmaTaher et al. [15] for the distribution of sputum pixels and non-sputum pixels in RGB space for extracting the sputum cell from the raw image. Qingxiang Zhu et al. [16] used a method to classify pulmonary nodules from blood vessels. This concept was based on the idea of low contrast of the intensity between nodules and vessels in CT images. Their method shows structure enhancement and active evolution, which can detect pulmonary nodules at a high accuracy. S.L.A. Lee et al [17] proposed random forests (RF) based classification aided by clustering (CAC) method. Lung nodules were classified by using clustering. In clustering the group of objects of same category belongs to same cluster. The cluster based classification improves the classification accuracy. The proposed method consists of training and test stages. The clusters were obtained by merging the nodule and non-nodule parts of training sets. Then each cluster was divided into two groups: nodule and non-nodule. Pei Xiaomin et al [18] used rule-based classifier. It required a large overtraining effect to classify the nodules. The rule-based classifier consists of two steps selection of a feature and cut off threshold. The optimal composite feature determines from separation metric and it was distinguish nodules from non-nodules. Separation metric maximizes due to the linear combination of simple

feature. This method proved that the overtraining in the process of threshold selection is eliminated by passing cut off threshold through one of the nodules. The result was obtained by calculating the ratio of number of non-nodules removed to the nodules sacrificed. Then determine k-ratio for k -possible threshold. If one of them achieved maximum value then it was defined as gain of feature. The optimal threshold for the feature was considered from corresponding feature. Wail A. H. Mousa et al [19] used support vector machines (SVM) for classification of lung nodules. They used features which were extracted from CT images to train the SVM classifier. SVM classifier maps the input into high dimensional feature space through nonlinear mapping. The optimal hyper plane was constructed to differentiate the nodules. By using the kernel mappings such as radial basis function a nonlinear hyper plane was constructed. The reduce dimensionality help to reduce the computational complexity.

#### IV. PROPOSED WORK

Proposed framework is model into four parts image enhancement, segmentation, feature extraction and classification. These four processes are described as follows.

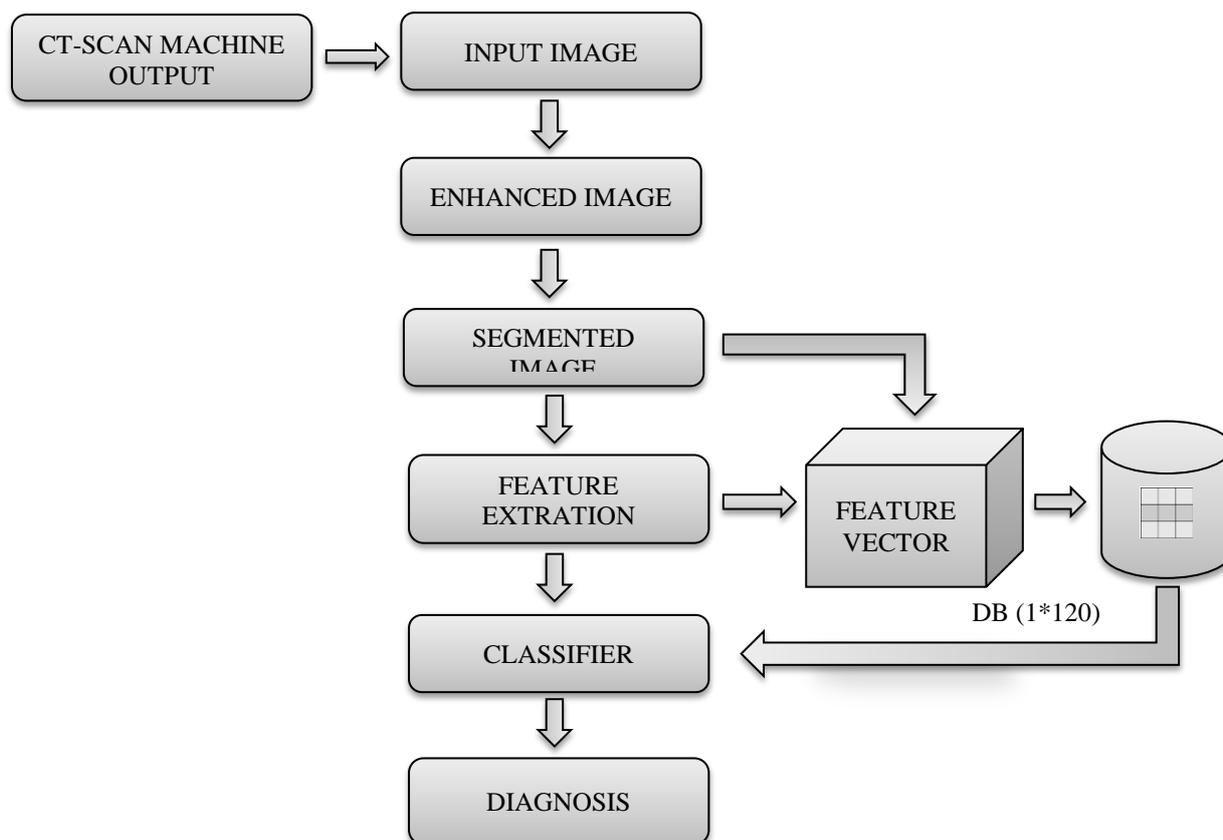


Fig.2 Overall architecture

#### 4.1 Input Image

The images collected from cancer hospital in the form of DICOM (Digital Image Communications in Medicine) image format which are of cancerous and non-cancerous patients. It is not just the image or file format. DICOM image provides all necessary tools for diagnostically accurate representation and processing of medical imaging data.



Fig.3 Input image

#### 4.2 Enhancement

Enhancement technique enhances the contrast of images. The contrast enhancement can limit in order to avoid the noise which is present in an image.

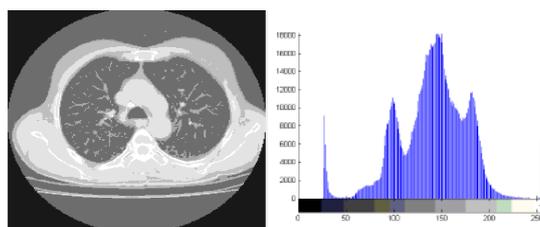


Fig.4 (a) Improved image Fig.4 (b) Histogram plot

We used histogram equalization to improve the contrast of images by transforming the values in an intensity of an image. So that the histogram of the output image approximately matches a specified histogram.

#### 4.3 Segmentation

In the image segmentation the CT image is sub-divided into small regions. Segmentation distinguishes the nodule from background part of the lung CT images

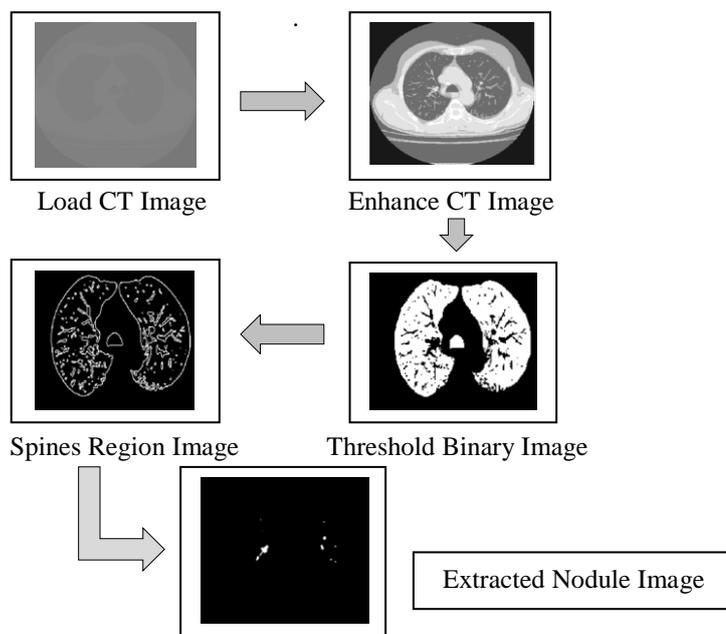


Fig.5 Segmentation method

The above diagram illustrates the same. We extract the lungs left and right region by using thresholding, imclearborder and bwareaopen functions. The thresholding method is based on a threshold value to turn a gray-scale image into a binary image. The threshold is used to isolate lung tissues and the imclearborder is used to remove artifacts attached to border. The bwareaopen is used to remove objects in less pixel size.

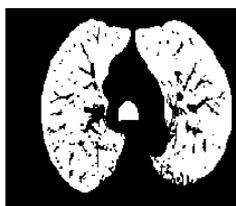


Fig.6 Threshold binary image

In second step, we used the gradient magnitude as the segmentation function to show the nodule and nodule spine borders. For this we used the sobel edge masks, and some arithmetic functions to compute the gradient

magnitude. The gradient value is high at the borders of the objects and low inside the objects. The output image fig. 7 shown below displays the spine border clearly. Spine border is used to find the nodules are cancerous or not.

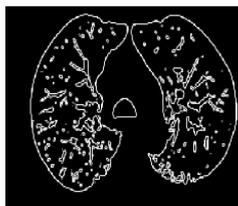


Fig.7Spines region image

Generally the cancerous nodules have spine border. The cancerous nodules are fully covered by the spines. To get the exact lung nodule image we perform further operation on previous output image of segmentation. The nodule image shows the cancerous and non-cancerous lung nodules are extracted from the CT scan image



Fig.8Nodule image

#### 4.4 Feature Extraction

Feature extraction is essential and very important step to extract region of interest (ROI). In our project implementation the nodule size, structure, volume and nodule spine values are considered as feature value. The fig.8 in above segmented method shows different nodule size, structure, and volume. The spine values are pre-calculated by using gradient magnitude method in our segmentation method. By using these two images feature values are calculated and are given as input to the feature vector of  $1 \times 120$  sizes. The following diagram shows vector of size  $1 \times 120$  of feature values.

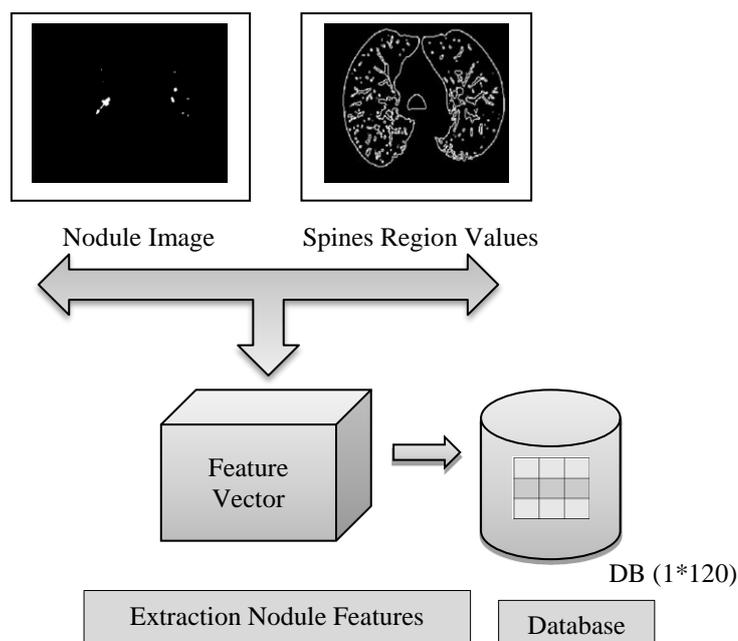
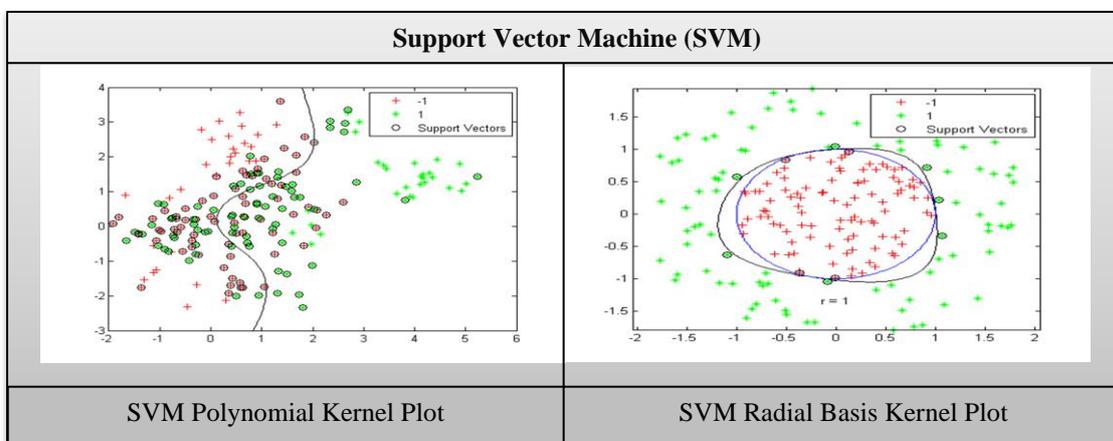


Fig.9 Feature vector

#### 4.5 Classification and Recognition

The Support Vector Machine (SVM) classifiers differentiate the non-cancerous (Benign) from cancerous (Malignant) lung nodules. It constructs a hyperplane in a high dimensional space, which can be used

for classification. The good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class. In general larger the margin lowers the generalization error of the classifier. The feature vector is given as input to the classifier. This method differentiates and identifies the non-cancerous (Benign) and cancerous (Malignant) lung nodules. Classification and recognition randomly divide database into 70% of the database for training and 30% for testing. Both subset have the random samples from the same distribution. In training data, where each row corresponds to an observation or replicate, and each column corresponds to a feature or variable. The classifier train on the training set applies it to the testing set and then measure performance by comparing the predicted labels and give decision as cancerous and non-cancerous.



#### IV. EXPERIMENTAL RESULTS

Table 1. Training and testing set

Classifiers	Training Images	Testing Images	Total
SVMRBF	400	100	500
SVMPOLY			

Table 2. FAR, FRR and GAR rate

Classifiers	FAR%	FRR%	GAR%
SVMRBF	4%	3%	96%
SVMPOLY	5%	6%	95%

The reference CT images database is created and on that database different classification techniques are applied such as support vector machine (SVM) with radial basis kernel and polynomial kernel and after several experimental results are obtained. The different training and testing are used while performing training and testing using classifier mention in table-II and the accuracy of the system is based on the performance evolution parameter i.e. false acceptance rate (FAR) is 4%, false recognition rate (FRR) is 3% and the genuine acceptance rate (GAR) is 96% for the SVMRBF classifier. The false acceptance rate (FAR) is 5%, false recognition rate (FRR) is 6% and the genuine acceptance rate (GAR) is 95% for the SVMPOLY classifier.

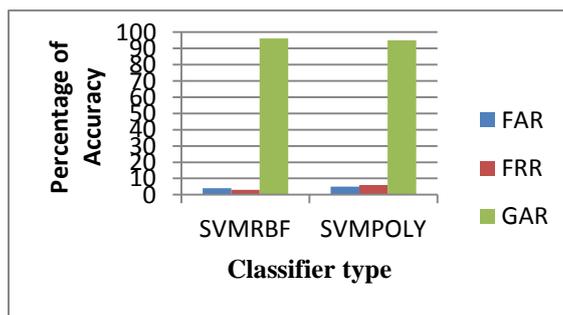


Fig.11 Graph for the classification accuracy

## V. CONCLUSION

The proposed research work is performed on computed tomography images i.e. CT scan images for the detection of Lung Nodule is cancerous or not. Existing approaches has certain drawbacks but there is need of precision in the Medical diagnosis domain, so there is a scope of improvement. Hence the different classification, Image enhancement technique are used for taking precise decision and proposed classification technique outperforms the existing available techniques.

## REFERENCES

- [1] AhmedinJemal, Rebecca Siegel, Elizabeth Ward, and YongpingHao", *Cancer journal for clinicians*,vol.58, pp.71-96, Dec 31, 2008.
- [2] Ann Leung,RobinSmithuis,"Solitary Pulmonary Nodule:Benign Versus Malignant Differentiation wit CT and PET-CT",May20,2007.
- [3] HanspeterWitschi, "A Short History Of Lung Cancer",*Toxicological Sciences*, vol. 64, pp.4-6,2001.
- [4] PoonamBhayan, Gagandeep Jindal "A Segmented Morphological Approach to Detect Tumor in Lung Images",*International Journal of Computer Science and Technology* vol. 2, pp. 499-503Issue 3, September, 2011
- [5] Robert N. Proctor "Golden Holocaust Origins of the Cigarette Catastrophe and the Case for Abolition",*Univerity of CalifornaPress*,2012.
- [6] Bhadauria H S, Dewal M L, "Comparison of Edge Detection Techniques on Noisy abnormal Lung CT Image before and after Using Morphological Filter", *International Journal of Advanced Engineering & Application*, pp.271-275,Jan.2010.
- [7] Zhenghao Shi, Jun Bai, Lifeng He, Tsuyoshi Nakamura, Quanzhu Yao, Hidenoriltoh "A Method for Enhancing Lung Nodules in Chest Radiographs by Use of LoGFilter",*Image and Signal Processing CISP '09. 2nd International Congress*,pp.1-4, Oct 17-19.,Tianjin,2009.
- [8] Vivekanandan D , Sunil Retmin Raj ,"A Feature Extraction Model for Assessing the Growth of Lung Cancer in Computer Aided Diagnosis",*IEEE-International Conference on Recent Trends in Information Technology, ICRTIT MIT*,pp.953-958,June 5,Chennai,India,2011.
- [9] Anita chaudhary, SonitSukhraj Singh,"Lung Cancer Detection on CT Images By Using Image Processing",*Computing Sciences (ICCS)*, *International Conference*pp.142-146,Sept.14-15, Phagwara,2012.
- [10] Jia Tong, Wei Ying, Wu Cheng Dong," A Lung Cancer Lesions Detection Scheme Based on CT Image",*2nd IEEE International Conference on Signal Processing Systems (ICSPS)*, 2010.vol1,pp.557- 560,July 5,Shenyang ,China 2010.
- [11] Li Shen, Wei Zheng, Ling Gao, Heng Huang, FilliaMakedon, and Justin Pearlman," Modeling Time-Intensity Profiles for Pulmonary Nodules in MR Images", *Proceedings of the IEEE Engineering in Medicine and Biology 27th Annual Conference*.pp.1359-1362 September 1,Shanghai,China,2005.
- [12] Maciej Dajnowiec, JavadAlirezaie,"Computer Simulation for Segmentation of Lung Nodulesin CT Images",*Systems, Man and Cybernetics, IEEE International Conference*,vol.5,pp.4491-4496,The Hague,2004
- [13] K.Devaki and V.MuraliBhaskaran," Study of Computed Tomography Images of the Lungs", *IEEE-International Conference on Recent Trends in Information Technology, ICRTIT MIT, Anna University*, pp.837-842,June 3,Chennai,India 2011.
- [14] Noriyasu Homma, Kazuhisa Saito, Tadashi Ishibashi, Madan M. Gupta, Zeng-GuangHou, and Ashu M. G. Solo ,"Shape Features Extraction from Pulmonary Nodules in X-ray CT Images",*Neural Networks IJCN. (IEEE World Congress on Computational Intelligence).IEEE International Joint Conference* , pp 3396-3400,June 1-8,Hong Kong,2008..
- [15] FatmaTaher, NaoufelWerghi and Hussain Al-Ahmad, "Thresholding approach for detection of sputum cell for lung cancer early diagnosis",*Image Processing (IPR 2012), IET Conference* , pp.1-6,July3 ,Londan ,2012.
- [16] Qingxizng Zhu, HongkaicXiong, Xiaoqian Jiang, " Pulmonary Blood Vessels and Nodules Segmentaton via Vessel Energy Function and Radius-Variable Sphere Model",*Healthcare Informatics, Imaging and System Biology(HISB),IEEE Second International Conference* ,pp.121,Sept27, San Diego,CA,2012.
- [17] S.L.A. Lee, A.Z. Kouzani, and G. Nasierding, E.J. Hu "Pulmonary Nodule Classification Aided by Clustering",*Systems, Man and Cybernetics,IEEE International Conference* ,pp.906-911,Oct. 11-14 San Antonio, TX, USA , 2009
- [18] Pei Xiaomin, GuoHongyu, Dai Jianping, "Computerized Detection of Lung Nodules in CT Images by use of Multiscale filters and Geometrical constraint Region Growing", *Bioinformatics and Biomedical Engineering ,4th International Conference* ,pp.1-4,June 18-20, Chengdu 2010.
- [19] Wail A.H Mousa and Mohammad A. U Khan, "Lung Nodule Classification Utilizing Support Vector Machines", *Image Processing.Proceedings.InternationalConference*, vol.3, pp.153 -156, 2002.