International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Lung Segmentation and Tumor Identification from CT Scan Images Using SVM

Ganesh Sable[1], Harsha Bodhey[2]

[1]Professor, Savitribai Phule Women's Engineering College, Aurangabad, Maharashtra, India

[2]Student, Savitribai Phule Women's Engineering College, Aurangabad, Maharashtra, India

Abstract: ***Segmentation of images has become important and effective tool for many technological applications like lungs segmentation from CT scan images, medical imaging and many other post-processing techniques. Lung cancer is the primary cause of deaths for both sexes in most countries. Lung nodule, an abnormality which leads to lung cancer is detected by various medical imaging techniques like X-ray, Computerized Tomography (CT), etc. Detection of lung nodules is a challenging task since the nodules are commonly attached to the blood vessels. Many studies have shown that early diagnosis is the most efficient way to cure this disease. This paper presents an adaptive segmentation of the lungs and the lobes and also an efficient algorithm for tumor classification from CT scan images . In pre-processing, first the RGB image is converted into gray scale image and then the noise is removed from this image using wiener filter. After that we have segmented the lungs from the original CT image using thresholding. The lobes of the lung will be segmented using marker based watershed transformation. Through the project we have developed an algorithm for identifying the tumor from segmented lung images. For identification we have used GLCM features and SVM classifier together. At last the tumor present in the CT scan image is isolated using FCM approach. The results indicate a potential for developing an automatic algorithm to segment lung lobes and tumor classification for surgical planning of treating lung cancer. For this, we have collected online database of 51 patients from Lola 11. The proposed system is implemented in MATLAB software.***

Keywords: Lobe Segmentation, Watershed Transformation, Feature Extraction, GLCM (Gray Level Co-occurrence Matrix), FCM (Fuzzy C-means), SVM (Support Vector Machine).

1. Introduction

Lung cancer is one of the leading causes of death in India. It is very difficult for patients to detect the lung cancer until it reaches in advanced stage [1]. Pulmonary CT scan images have an important role in the diagnosis of several lung diseases such as lung cancer, old or new pneumonia, tuberculosis, emphysema and chronic obstructive lung diseases (COPD).The human lungs are subdivided into five lobes that are separated by visceral pleura called pulmonary fissure as shown in figure 1. There are three lobes in the right lung, namely upper, middle and lower lobe. The right upper and right middle lobe are divided by the right minor fissure whereas the right major fissure delimits the lower lobe from the rest of the lung. In the left lung there are only two lobes, the upper and the lower lobe, that are divided by the left major fissure. Lung lobe segmentation is relevant in clinical applications particularly for treatment planning. The location and distribution of pulmonary diseases are important parameters for the selection of a suitable treatment. A lobe-wise analysis shows the progression of the disease in more detail. Computed tomography (CT) allows visualization of the lungs within a few seconds. Since typical scans with high anatomical details contain over 400 slices with submillimeter resolution for each direction, manual segmentation is time consuming and there is demand for automatic lung lobe segmentation methods. The segmentation of pulmonary lobes is challenging because of anatomical variation and incomplete fissures. On the one hand, pathologies can deform the lobes and make the fissures unrecognizable. And on the other hand, even in patients with normal lung parenchyma the fissures are often not complete. The proposed system carries out lungs and lung lobes segmentation and tumor identification and extraction of tumor from chest CT scan images within few seconds.

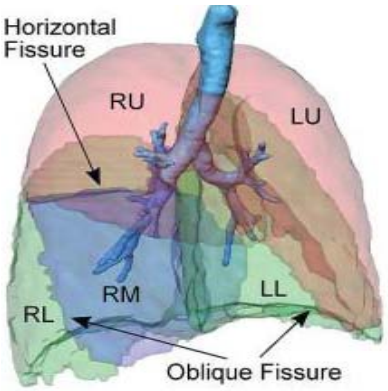


Figure 1: Lung lobes

Lung cancer is also called cancer of the bronchus which is produced as a result of uncontrolled growth of the lung tissues, especially the cells which line the air passages. The resulting cells will not develop into healthy ones; they divide to form tumors which are considered as the main cause of death from cancers.

1) Benign tumor
   Benign tumors are noncancerous cells; but they need to be treated because they might harm the neighboring tissues or other vital organs.

2) Malignant tumor
   Malignant tumors are cancerous cells and invade normal tissue or contain cancerous cells either from the lungs or other parts of the body.

2. Existing systems

Many authors explored the segmentation techniques in medical imaging depending on the region of interest till now [3]. Some of them use a semi-automatic algorithm and still need some user interaction, while others are fully automatic and the user has only a verification role. Various algorithms from different authors can be found for medical image segmentation such as region growing [2], thresholding [4]. A. Hoffman [5] have developed an automatic method for identifying lungs in 3D X-ray CT images. Zhang and Valentino [7] have suggested using artificial neural networks to classify each pixel in the CT slice into different anatomical structure. Some authors have proposed systems for nodule detection and classification using FIS [12], Neural network [13], Bayesian classifier [15].

3. Proposed Method

3.1 Introduction

Here, a new system is proposed for lung segmentation and tumor classification from CT scan images.

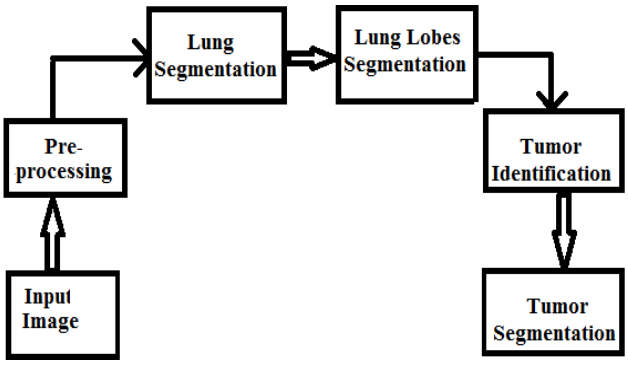


Figure 2: Block diagram of the proposed system

Above figure shows the block diagram of the proposed system. It consists of the following steps.

3.2 Pre-processing

Pre-processing includes following steps:

i. Input image
Here, the input images are chest CT scan images in JPEG format that contain tumors. First image selected from the file specified by the string filename. The user has to select the required lung CT scan image for further processing. Then each image is resized to 256*256.

ii. Wiener filtering
The input image is in RGB format. So it is first converted into gray scale image for further processing. Then wiener filter of mask size 3*3 is used to remove noise because it is one of the best methods to remove the noise from the CT images [14]; since these images usually contain artifacts or noise due to patient movements.

3.3 Post Processing

Post-processing includes following steps:

1) Lung Segmentation:
In this module we segment left and right lung from the CT image. First we have chosen the seed point in the CT image. From the point we found intensity value of the image. We compare the intensity value between the neighboring pixels and current pixel. If the neighbor pixels values are related to the seed value, it will segment lungs from the original image. These similarity pixels will be segmented from the CT image [14]. This process is continued until reach the last pixel. Finally the lungs will be segmented. Threshold value between 0 and 180 is selected.

2) Lobe Segmentation
Watershed transformation is a common technique for image segmentation. However, its use for automatic medical image segmentation has been limited particularly due to over segmentation and sensitivity to noise [11]. Employing prior shape knowledge has demonstrated robust improvements to medical image segmentation algorithms. We propose a novel method for enhancing watershed segmentation by utilizing prior shape and appearance knowledge. In watershed, internal markers to obtain watershed lines of the gradient of the image to be segmented. Use the obtained watershed lines as external markers. Each region defined by the external markers contains a single internal marker and part of the background. In watershed, regions without markers are allowed to be merged.

3) Feature Extraction
In this process, total 12 textural features of all images in the database are extracted using GLCM (Gray level co-occurrence matrix). Then these features are used for tumor classification. GLCM is simply a matrix that gives the sum of the number of times that the pixel with value *i* occurred in the specified spatial relationship to a pixel with value *j* in the input image. Texture feature calculations use the contents of the GLCM to give a measure of the variation in intensity at the pixel of interest. These GLCM features calculated for some of the images are shown in following table:

Table 1: GLCM features and their values

GLCM features	*Image 1*	*Image2*
autoc	32.9612	39.6644
contr	0.3473	0.3034
corrm	0.9726	0.9520
corrp	0.9726	0.9520
cprom	739.3252	177.3976
cshad	-5.9851	3.7745
dissi	0.2396	0.2379
energy	0.1737	0.2551
entro	2.1526	1.7496
homom	0.8943	0.8909
homop	0.8903	0.8876
maxpr	0.2924	0.4250

Like this, these GLCM features are calculated for all images in database.

GLCM Features

1. autoc (Autocorrelation):

$$\rho(x,y) = \frac{1}{(L_x - |x|)(L_y - |y|)} \int_{-\infty}^{\infty} I(u,v) I(u+x, v+y)\, du\, dv$$

$$\frac{1}{L_x L_y} \int_{-\infty}^{\infty} I^2(u,v)\, du\, dv \qquad |x| < L_x \text{ and } |y| < L_y.$$

2. Contr (Contrast): It is a measure of the intensity contrast between a pixel and its neighbor over the whole image.

$$\sum_{i,j=0}^{G-1} (i-j)^2 P(i,j)$$

3. corr (Correlation): It is a measure of gray level linear dependence between the pixels at the specified positions relative to each other.

$$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{\{i \times j\} \times P(i,j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y}$$

4. cprom (Cluster prominence):

$$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{i + j - \mu_x - \mu_y\}^4 \times P(i,j)$$

5. cshad (Cluster shade): It is a measure of skewness of the matrix.

$$\sum_{i,j=0}^{G-1} (i + j - \sigma_I - \sigma_J)^3 P(i,j)$$

6. dissi (Dissimilarity): It gives the measure of much dissimilar are of two neighboring pixels.

$$\sum_{i,j=0}^{N-1} P_{i,j} |i - j|$$

7. energy (Energy):

It is also known as uniformity of ASM (angular second moment) which is the sum of squared elements from the GLCM. Range = [0 1] Energy is 1 for a constant image.

$$\sum_{i,j=0}^{G-1} P(i,j)\, 2$$

8. entro (Entropy):

It is a measure of randomness. Entropy measures the loss of information or message in a transmitted signal and also measures the image information.

$$-\sum_{i=0}^{G-1} p(z_i) \log_2 p(z_i)$$

9. homom (Homogeneity):

It returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1 + (i-j)^2} P(i,j)$$

10. maxpr (Maximum probability):

This simple statistic records in the centre pixel of the window the largest P_{ij} value found within the window.

$$\max(i,j) P(i,j)$$

3.4 Tumor classification

After extracting GLCM features of images, tumor classification is carried out. For this, we have used SVM (Support vector machine) classifier. This classifier must be trained first. For training we have used 6 images out of which first 3 images are of 'benign' type and next 3 are of 'malignant' type tumor. The GLCM features are given as input to SVM. Each image is assigned a class i.e. for benign, class 0 and for malignant, class1.

Here, two class SVM classifier is used. An SVM classifies data by finding the best hyperplane that separates all data points of one class from those of the other class. The best hyperplane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyperplane that has no interior data points. The hyperplane is defined by the equation:

$(w.\ x) + b = 0$

Where, w= weight vector

x= feature vector

B= bias

The value of bias 'b' is found to be -0.3481.

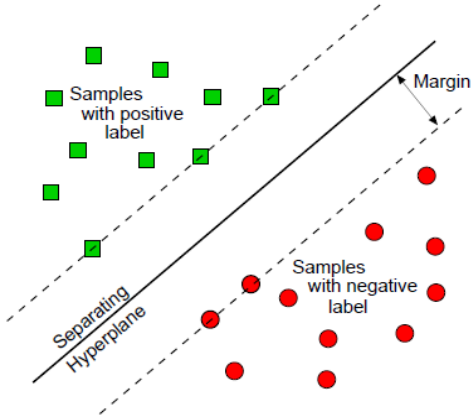


Figure 2: SVM classifier

The vectors closest to the boundaries are called support vectors and the distance between the support vectors and hyper plane is called margin [8]. SVM first maps the input feature vectors into higher dimensional feature space and then perform classification.

1. Training the classifier

In the training phase, known data is given and the classifier is trained. Here, six images are used for training out of which 3 are benign and 3 are malignant and are assigned class 0 for benign and class 1 for malignant. The training points satisfy the following conditions.

$(w.x_i) + b \geq +1$ for $y_i = +1$

$(w.x_i) + b \leq -1$ for $y_i = -1$

2. Testing of data

In testing phase, unknown data are given and the classification is performed using trained classifier. Classification is done by using following decision function.
f (x, {w, b})= sign (w.x+b)

The sign of this function decides the class of the test image. Here, if it is positive, then result will be 'malignant' and if it is negative, then result will be 'benign'.

3.5 Tumor segmentation

Clustering is a process for classifying objects or patterns in such a way that samples of the same group are more similar to one another than samples belonging to different groups. Here fuzzy C-means algorithm is used for tumor segmentation. Fuzzy clustering is basically a multi valued logic that allows intermediate values i.e., member of one fuzzy set can also be member of other fuzzy sets in the same image. In the proposed FCM, 3 clusters are taken and maximum number of iterations is 100. The algorithm is an iterative clustering method that produces an optimal c partition by minimizing the weighted within group sum of squared error objective function J_{FCM} [6].

$$J_{FCM} = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^q \, d^2 (x_k, v_i) \quad \ldots(1)$$

The membership function defines the fuzziness of an image and also to define the information contained in the image. These are three main basic features involved in characterized by membership function. They are support, Boundary. The core is a fully member of the fuzzy set. The support is non membership value of the set and boundary is the intermediate or partial membership with value between 0 and 1. This clustering algorithm allows one piece of data may be member of more than one clusters. It is based on reducing the equation 2,

$$Y_m = \sum_{i=1}^{N} \sum_{j=1}^{C} M_{ij}^{m} \left\| x_i - c_j \right\|^2 \quad \ldots. (2)$$

Where,
m- Any real number greater than 1.
Mij- Degree of membership of X; in the cluster j
xi- Data measured in d-dimensional.
Cj - Dimension centre of the cluster.
‖Xi-Cj‖²- Induced norm (Euclidean norm)
This system uses two level segmentation i.e. for two levels of thresholding, image is divided into 3 clusters and maximum number of iterations is 100.

3.6 System flowcharts

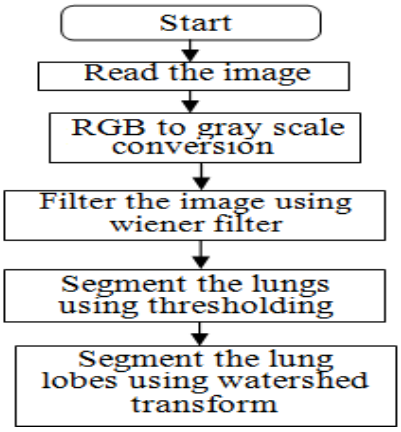


Figure 3: Flowchart for lobe segmentation

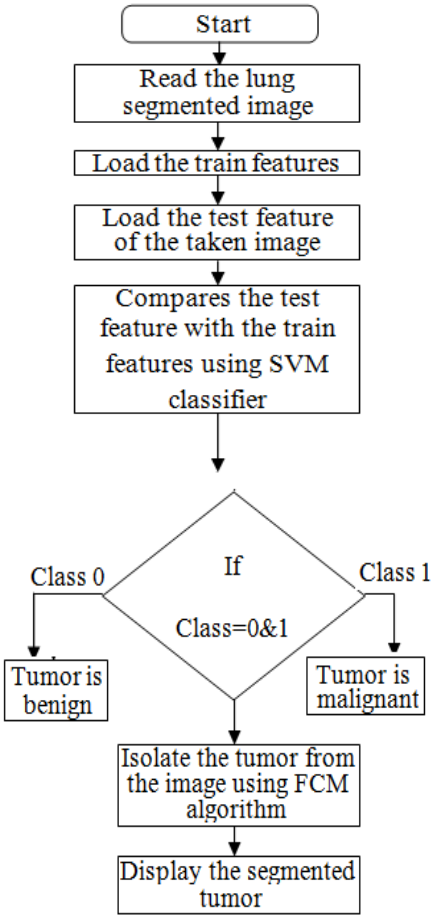


Figure 4: Flowchart for tumor classification and segmentation

3.7 Database

We have taken the database of 51patients which are lung CT scan images from web link http://lola11.com/Details. All these images are of size 256*256 and are in JPEG format. Some images from the database are shown in following figure:

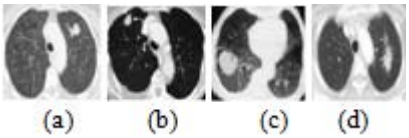
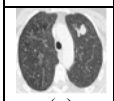
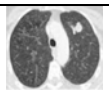
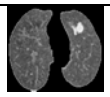
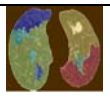
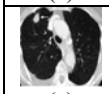
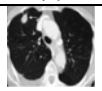
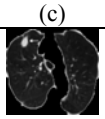
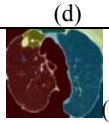
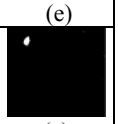
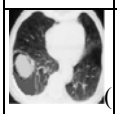
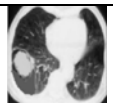
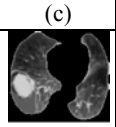
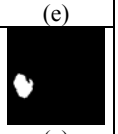
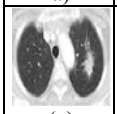
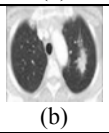
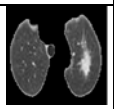
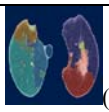
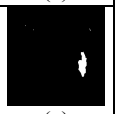


Figure 5: Images from database

4. Results of Experimentation

In this section, the results of the proposed system are shown for four images from the database.

Table 2: (a) Input image, (b) Wiener filtering, (c) Thresholding, (d) Watershed transformation, (e) FCM

Input image	Filtered image	Segmen-ted lungs	Lung lobes segme-nted image	Extracted tumor
 (a)	 (b)	 (c)	 (d)	 (e)
 (a)	 (b)	 (c)	 (d)	 (e)
 a)	 (b)	 (c)	 (d)	 (e)
 (a)	 (b)	 (c)	 (d)	 (e)

Similarly, the results for all 51 images in the database are obtained.

4.1 Classification Results and Analysis

When the above images were given to the SVM classifier for testing, we obtained the following outputs.

Table 3: SVM outputs and accuracy parameters

Image No.	*Expert's opinion*	*Experimental results*	*SVM output*	*Parameter*
1	Benign	Benign	Class 0	TP
2	Benign	Malignant	Class 1	FP
3	Malignant	Malignant	Class 1	TN
4	Malignant	Benign	Class 0	FN

Where,
TP- predicts benign as benign.
FP- predicts benign as malignant.
TN- predicts malignant as malignant.
FN- predicts malignant as benign.

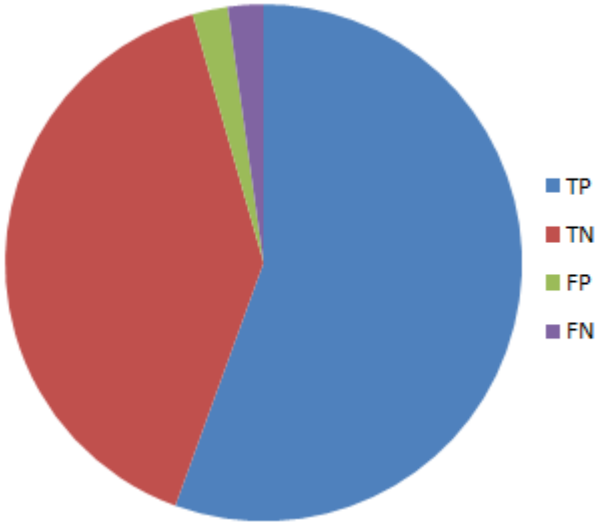
4.2 Performance measures

The following parameters are calculated on the basis of the results obtained for all the images in database.

1) Accuracy (AC) = (TP+TN) / (TP+TN+FP+FN)
= (25 + 18) / (45)
=0.9556
Accuracy in % = 95.56%

2) Sensitivity (SE) = TP / (TP+FN)
= 25 / (25+ 1)
= 0.9615
Sensitivity in % = 96.15%

3) Specificity (SP) = TN / (TN+FP)
= 18 / (18 + 1)
= 0.9473
Specificity in % = 94.73%



Graph 1: Accuracy parameters

Thus, from the results we found that our proposed system has achieved 95.56% accuracy, specificity of 94.73% and better sensitivity i.e. 96.15% that means our system has higher accuracy for the classification of benign type of tumor as compared to malignant tumor.

5. Conclusions

Thus it can be concluded that the proposed method performs well and is robust against anatomical variations of the lungs. The system gives results within few seconds. The SVM used for tumor classification has improved accuracy. Thus, this approach is a potential for developing an algorithm to segment lung, lobes and tumor identification for surgical planning of treating lung disease and it will assist radiologist as second opinion for the better diagnosis of lung cancer. The SVM classifier achieved an average accuracy of 95.56%.

6. Future Scope

The automatic lobe segmentation and tumor classification has very wide scope since it reduces manual work and also computational time. Also it can be useful for diagnosis of other lung diseases. Further it can be performed for 3D images in future. The million order dataset can be selected and image classification can be done on larger dataset. With increased size of dataset various issues such as uploading data, managing feature set, increased execution time of classification algorithms etc. can be considered. More image features can be extracted for better classification. Various combinations of pre-existing features can be used to correctly classify medical data.

References

[1] Bezdek J. C, "Pattern Recognition with Fuzzy Objective Function Algorithms", New York: Plenum Press, 1981.
[2] P. Croisille, M. Souto, M. Cova, S. Wood, Y. Afework, J. E. Kuhlman, E. A. Zerhouni, "Pulmonary nodules: improved detection with vascular segmentation and extraction with spiral CT", Radiology, Vol. 197, pp. 397-401, 1995
[3] D. L. Phan, C. Xu, J. Price, "A survey of current methods in medical image segmentation", Annual Review of Biomedical Engineering, 1998.
[4] M. S. Brown, M. F. McNitt-Gray, J. G. Goldin, R. D. Suh, J. W. Sayre, D. R. Aberle, "Patient-specific models for lung nodule detection and surveillance in CT images", IEEE Transactions on Medical Imaging, Vol. 20, pp. 1242-1250, 2001.
[5] E. A. Hoffmann, J. M. Reinhardt, "Automatic Lung Segmentation for Accurate Quantitation of Volumetric X-Ray CT Images", IEEE Transactions on Medical Imaging, Vol. 20, pp. 490-498, 2001.
[6] B. Zhao, G. Gamsu and M. S. Ginsberg, "Automatic detection of small lung nodules utilizing local density maximum algorithm", Journal of Applied Medical Physics, vol.4, No.3, 2003.
[7] D. Zhang, D. J. Valentino, "Segmentation of anatomical structure in X-Ray computed tomography images using artificial neural network", Proc. of SPIE, vol. 4684, pp. 1640-1652, 2005.
[8] R. Nithya, B. Santhi, "Mammogram classification using maximum Difference feature selection method", Journal of Theoretical and Applied Information Technology, vol. 33, pp. 197-204, Nov. 2011.
[9] Nitish Zulpe, Vrushsen Pawar, "GLCM Textural Features for Brain Tumor Classification", International Journal of Computer Science Issues, Vol. 9, Issue 3, pp. 354-359, May 2012.
[10] M. Gomathi, Dr. P. Thangaraj, "An effective classification of benign and malignant nodules using support vector machine", Journal of Global Research in Computer Science, vol. 3, pp. 6-9, July 2012.
[11] Bianca Lassen, Eva M. van Rikxoort, Michael Schmidt, Sjoerd Kerkstra, Bram van Ginneken, and Jan-Martin Kuhnigk, "Automatic Segmentation of the Pulmonary Lobes From Chest CT Scans Based on Fissures, Vessels, and Bronchi", IEEE Transactions on Medical Imaging, Vol. 32, No. 2, February 2013.
[12] Atiyeh Hashemi, Abdol Hamid Pilevar, Reza Rafeh, "Mass Detection in Lung CT Images Using Region Growing Segmentation and Decision Making Based on Fuzzy Inference System and Artificial Neural Network", International journal on Image, Graphics and Signal Processing, vol. 6, pp. 16-24, May 2013.
[13] Ada, Rajneet Kaur, "Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier", International journal of application or innovation in engineering and management, vol. 2, Issue 6, pp. 375-383, June 2013.
[14] T. Manikandan, D.Kesavaraja, "A study of different chest ct scan image segmentation for pulmonary lobes", International Journal of Research in Engineering & Advanced Technology, Vol. 1, Issue 5, pp. 1-8, Oct-Nov 2013.
[15] Anil M. Yametkar, R. D. Patane, "Lung Cancer Detection And Classification By Using Bayesian Classifier", Proceedings of IRF International Conference, Feb. 2014.
[16] S.Shaik Parveen, C.Kavitha, "Classification of Lung Cancer Nodules using SVM Kernels", International Journal of Computer Applications, Vol. 95– No.25, June 2014.
[17] R. C. Gonzalez, R. E. Woods, "Digital Image Processing", 3rd Edition, pp.738-756.

Author Profile

Dr. G. S. Sable, Professor and Head of Electronics & Telecommunication department at Savitribai Phule Women's Engineering College, Aurangabad, Maharashtra India. PhD from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra. M.E and B.E from J.N.E. College, Aurangabad. He has more than 50 publications to his credit and has been active in research and development. He has more than 14 years of teaching experience. He is a member of editorial advisory board of the different Journals. He is Authors of the book Microprocessor and Computer Organization for the second year CSE/IT Branch students and the member of the IEEE, ISTE, and IACSIT. He is Member of 32(6) a Malpractice Committee of Dr. Babasaheb Marathwada University, Aurangabad Maharashtra.

Harsha D. Bodhey, Student of M.E.(EC) 2ndYear, Savitribai Phule women's Engineering College, Aurangabad, Maharashtra, India. BE (EC) from SPWE College of Engineering, Aurangabad, Maharashtra in 2012. She has published two papers in international conference.