C. Bhuvaneswari, P. Aruna, D. Loganathan / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 3, Issue 4, Jul-Aug 2013, pp.2517-2524

Advanced Segmentation Techniques Using Genetic Algorithm for Recognition of Lung Diseases from CT Scans of Thorax

C. Bhuvaneswari[1], P. Aruna[2], D. Loganathan[3]
[1](Head, Department of Computer Science, Theivanai Ammal College For Women (Autonomous), Villupuram,India)
[2] (Professor, Department of Computer Science and Engineering, Annamalai University, Chidambaram,India)
[3](Professor &Head, Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry,India)

ABSTRACT

In this study, texture based segmentation and recognition of the lung diseases from the computed tomography images are presented. The texture based features are extracted by Gabor filtering, feature selection techniques such as Information Gain, Principal Component Analysis, correlation based feature selection are employed with Genetic algorithm which is used as an optimal initialisation of the clusters. The feature outputs are combined by watershed segmentation and the fuzzy C means clustering. The images are recognized with the statistical and the shape based features. The four classes of the dataset of lung diseases are considered and the training and testing are done by the Naive Bayes classifier to classify the datasets. Results of this work show an accuracy of above 90% for the correlation based feature selection method for the four classes of the dataset.

Keywords – Features, Genetic Algorithm, Image Segmentation, Texture, Training.

I. INTRODUCTION

Lung diseases are leading cause for the most disabilities and death in the world. Radiologist diagnosis the chest CT and the success of the radiotherapy depend on the dosage of the drugs given and the doses that affect the normal tissues surrounding areas. The chest CT shows the first important modality of the assessment of the diseases. The CT image along with the symptoms of the diseases will give detailed assessment about the lung diseases.

The major causes of the lung diseases are caused by smoking, inhaling the drugs, smoke and allergic materials. The lung diseases are generally identified by the symptoms and the regular dosage of the antibiotics may cure the disease. If the antibiotics does not respond to the disease the computed tomography images assists in detecting the severity of the lung diseases. There are many types of the disease that causes the lung infection such as inflammatory lung diseases, chronic obstructive pulmonary disease(COPD),Emphysema, Chronic Bronchitis, pleural effusion,Intersitial lung diseases and lung carcinoma.

The datasets of the lung diseases considered in this study are the large cell lung carcinoma and small cell lung carcinoma. Lung cancer or lung carcinoma is currently the most frequently diagnosed major cancer and the most common cause of cancer mortality in males worldwide. This is largely due to the effects of cigarette smoke. An international system of tumor classification is important for consistency in patient treatments and to provide the basis of epidemiological and biological studies. In developing this classification, pathologists have tried to adhere to the principles of reproducibility, clinical significance and simplicity, and to minimize the number of unclassifiable lesions. Most of this classification is based on the histological characteristics of tumors seen in surgical or needle biopsy, and is primarily based on light microscopy, although immune histochemistry and electron microscopy findings are provided when necessary.

The methodology used in this work defined that the images are pre-processed for the removal of the noises and contrast enhancement is done for obtaining the enhanced images. Feature extraction is frequently used as a preprocessing step to machine learning where the Gabor filter is used in texture analysis. The feature selection method such as the Information Gain, correlation based feature selection, Principal Component Analysis with optimisation of the genetic algorithm are done. The feature outputs are combined by watershed segmentation and the fuzzy C means clustering combines the data that belongs to two or more clusters. The Naive Bayes classifier is used to classify the images and the results are shown with the performance measures.

The paper is organized as follows: Section 2 deals with the related works available in literature. Section 3 explains the methodology. In the section 4 experimental setup is detailed and section 5 deals with the performance analysis and section 6 deals with the findings of the study.

II. PREVIOUS STUDIES AND RELATED WORKS

Manish Kakar et al., [1] proposed a method based upon the texture features, as extracted from Gabor filtering, the FCM can be used for segmentation of CT of thorax given that the cluster centres are initialized by using a Genetic Algorithm. From the segmentation results, the accuracy of delineation was seen to be above 90%.For automatically recognizing the segmented regions, an average sensitivity of 89.48%was achieved by combining cortex-like, shape and position-based features in a Simple SVM classifier.

Ribeiro, et al., [2] proposed StARMiner (Statistical Association Rule Miner) that aims at identifying the most relevant features from those extracted from each image, taking advantage of statistical association rules. The proposed mining algorithm finds rules involving the attributes that discriminate medical image the most. The feature vectors condense the texture information of segmented images in just 30 features.

Bhuvaneswari et al., [3] proposed to extract features in the frequency domain using Walsh Hadamard transform and use FP-Growth association rule mining to extract features based on confidence. The extracted features are classified using Naïve Bayes and CART algorithms and the proposed method's classification accuracy measured. Investigate the efficacy of feature selection and reduction using Association Rule Mining (ARM) on medical images. Naïve Bayes and Classification and Regression Tree (CART) classifiers were used for evaluating the accuracy of proposed method.

Uppaluri et al. [4] have developed a general system for regional classification by using small areas that were classified into one of the six categories based upon 15 statistical and fractal texture features.

Shyu etal. [5] have developed a system that retrieves reference cases similar to the case at hand from a proven database. In their approach they have combined global and anatomical knowledge, combining features from several pathological regions and anatomical indicators per slice. The regions are however manually delineated rather then automatically detected. In all the studies mentioned above, some sort of grid over slices/ROI marking or marked pathologies beforehand are needed for training, thus a supervised approach is used.

III. METHODOLOGY

The methods employed for the processing of the work is dealt in detail in this section.

3.1 Preprocessing and feature extraction:

The removal of the noise and contrast enhancement is done by preprocessing .The noise of the images is removed by the median filter. The following features are extracted from the lung disease CT images. They are

Name	**Description**
Orientation	The angle between the *x*-axis and the major axis of the ellipse
Mean	Mean of the method response from the region
Variance	Variance of the method response from the region
CentroidX	*X* coordinate value of the centroid of the patch
CentroidY	*Y* coordinate value of the centroid of the patch
Area	Area of the patch
Entropy	Average, global information content of an image in terms of average bits per pixel.
Contrast Difference	difference between the lightest and darkest areas.
Homogeneity	The state or quality of being homogeneous, biological or other similarities within a group.

The other extracted features are median, standard deviation, root mean square, root mean square also known as the quadratic mean, is a statistical measure of the magnitude of a varying quantity. Histogram is used to graphically summarize and display the distribution of a process data set.

The Gabor filter is used to extract the texture features from the preprocessed image. The coding is implemented using the Matlab. The following are the steps for feature extraction

- Creates a flat, disk-shaped structuring element, of the radius R which specifies the radius which is a nonnegative integer.
- Performs the morphological bottom-hat filtering on the greyscale or binary input image, which returns the filtered image.
- The structuring element returned by the strel function must be a single structuring element object, not an array containing multiple structuring element objects.
- The Top-hat filtering and bottom-hat filtering are used together to enhance contrast in an image.
- Add the original image and the top-hat filtered image, and then subtract the bottom-hat filtered image.

By applying the above techniques the noise will be removed and contrast enhancement will be done.

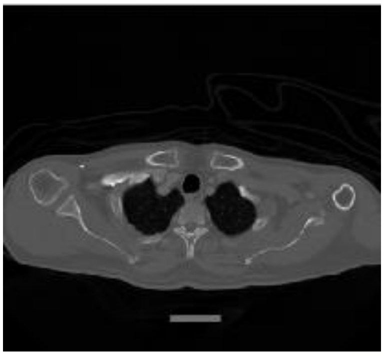

Fig 1. Original image

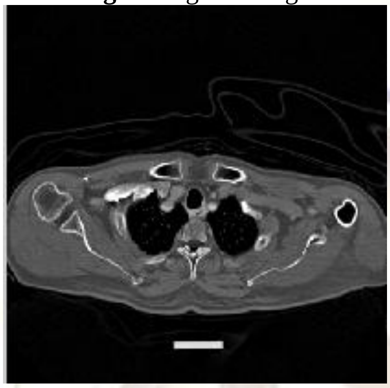

Fig 2. Preprocessed image

3.2 Gabor kernel filters

A complex Gabor filter is defined as the product of a Gaussian kernel times a complex sinusoid. The Gabor filter is widely used in image processing, especially in texture analysis. The function is based on 'Uncertainty Principle' and can provide accurate time-frequency location . The Gabor filters has optimal localization properties in both spatial and frequency domain. The Gabor function is a harmonic oscillator, made of sine wave enclosed in a Gaussian envelope. A 2-D Gabor filter over the image domain(x,y) is given by

$$G(x,y)=\exp\left(-\frac{(x-x_0)^2}{2\sigma_x^2}-\frac{(y-y_0)^2}{2\sigma_y^2}\right) \qquad (1)$$

$$X\exp\left(-2\pi i\left(u_0(x-x_0)+v_0(y-y_0)\right)\right)$$

Where

(x_0,y_0) is location in the image,

(u_0,v_0) specifies modulation which has

frequency $\omega_0=\sqrt{u_0^2+v_0^2}$ and

orientation $\theta_0=\arctan\left({}^{v_0}\!/_{u_0}\right)$

σ_x and σ_y are standard deviation of Gaussian envelope

Gabor filter calculates all the convolutions of the input image IMG with the Gabor-filter kernels for all combinations of orientations and all phase-offsets with the input image IMG. The result is a 4-dimensional matrix of which the first two indices are the image-coordinates, the third index is the phase offset, the fourth index is the orientation.

The following are the steps for the Gabor kernel calculation

- Calculate the ratio σ / λ from bandwidth then test if the σ / λ ratio.
- Creation of two (2n+1) x (2n+1) matrices x and y that contain the x- and y-coordinates of a square 2D-mesh.
- The wave vector of the Gabor function is calculated.
- Pre compute coefficients of the function.

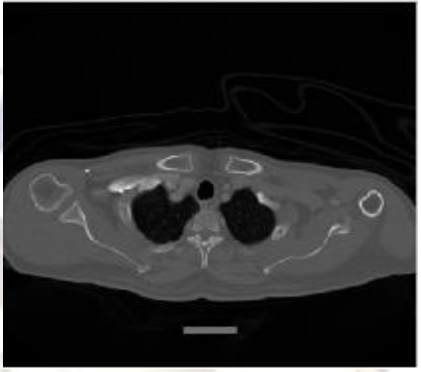

Fig 3. Gabor kernel filters images

3.3 Feature selection

Feature selection deals with selecting a subset of features, among the full features, that shows the best performance in classification accuracy. The best subset contains the least number of dimensions that most contribute to accuracy. This is an important stage in preprocessing. *Filter techniques* assess the relevance of features by looking only at the intrinsic properties of the data. Filter techniques can easily scale to very high-dimensional datasets, they are computationally simple and fast, and they are independent of the classification algorithm.

Three feature selection methods such as Information Gain, correlation based feature selection, Principal Component Analysis are employed

3.3.1 Information gain

One of the filter based univariate model search which is Fast, Scalable, Independent of the Classifier is Information gain method. It measures the number of bits of information obtained for category prediction, it measures the decrease in entropy when the feature is given or absent. The more space a piece of information takes to encode, the more entropy it occupies. The information gain of an attribute is measured by the reduction in entropy **IG(X) = H(D) − H(D|X)**. The greater the decrease in entropy when considering attribute X individually, the more significant feature X is for prediction.

The output of the Gabor kernel filter is given as an input where the REPTree is created. The REPTree is a fast decision tree learner which builds a decision/regression tree using information gain as the splitting criterion, and prunes it using reduced error pruning. It only sorts values for values for numeric attributes once.

3.3.2 Correlation-based feature selection (CFS):

Another filter based multivariate model search which is Models feature dependencies, Independent of the classifier, better computational complexity than wrapper methods is correlation based feature selection. CFS searches feature subsets according to the degree of redundancy among the features. The evaluator aims to find the subsets of features that are individually highly correlated with the class but have low inter-correlation. Correlation coefficients are used to estimate correlation between subset of attributes and class, as well as inter-correlations between the features.

CFS is used to determine the best feature subset and is usually combined with search strategies such as forward selection, backward elimination, bi-directional search, best-first search and genetic search.CFS first calculates a matrix of feature-class and feature-feature correlations from the training data.

Equation for CFS is given.

$$r_{zc} = \frac{k\overline{r_{zi}}}{\sqrt{k + k(k-1)\overline{r_{ii}}}} \qquad (2)$$

Where r_{zc} is the correlation between the summed feature subsets and the class variable

k is the number of subset features

r_{zi} is the average of the correlations between the subset features an the class variable

r_{ii} is the average inter-correlation between subset features.

The output of the Gabor kernel filter is given as an input where the numeric attributes are taken and the mean, standard deviation, weighted sum, precision of the attributes are calculated and the summary of the instances is calculated.

3.3.3 **Principal Component Analysis (PCA)**

PCA method is a global feature selection algorithm which identifies patterns in data, and expresses the data in such a way as to highlight their similarities and differences. It is a way to reduce the dimension of a space that is represented in statistics of variables (x_i, i = 1,2…. n) which mutually correlated with each other . PCA algorithm can be used to reduce noise and extract features or essential characteristics of data before the classification process.

The steps in the PCA algorithm are:

a) Create a matrix [X_1, X_2, ... X_m] which representing N2 X_m data matrix. Xi is the image of size N x N, where N2 is the total pixels of the image dimensions and m is the number of images to be classified.

b) Use the following equation to calculate the average value of all images

$$Y = \frac{1}{m}\sum_{i=1}^{m} X_i \qquad (3)$$

c) Calculated the difference matrix

$$\bar{X}_i = X_i - Y \qquad (4)$$

d) Use the difference matrix obtained previously to generated the covariance matrix to obtain the correlation matrix

$$\Sigma = \sum_{i=1}^{N} \bar{x}_i \bar{x}_i^{T} \qquad (5)$$

e) Use the correlation matrix to evaluate the eigenvector

$$\Sigma\phi_i = \lambda\phi_i \qquad (6)$$

Where ǿ is orthogonal eigenvector matrix, λ is the eigenvalue diagonal matrix with diagonal elements

f) If Φ is a feature vector of the sample image X, then :

$$y_n = \Phi^{T} \bar{X}_i \qquad (7)$$

With feature vector y is the n-dimensional.

The output of the Gabor kernel filter is given as an input where the numeric attributes are taken and the mean, standard deviation, weighted sum, precision of the attributes are calculated and the summary of the instances is calculated.

3.4 Genetic algorithm based Initialization

Genetic Algorithm (GA) is a well-known randomized approach. It is a particular class of evolutionary algorithms that makes use of techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover. In feature selection problems, each feature subset is represented by a binary string [13]. 1 of N^{th} bit means that the feature set contained feature X_n . A fitness function is a particular type of objective function that quantifies the optimality of a solution in a genetic algorithm. The wrapper approach is used in the experiment that will measure fitness function by the accuracy of learning algorithms.

A genetic algorithm mainly composed of three operators: selection, crossover, and mutation. In selection, a good string is selected to breed a new generation, crossover combines good strings to generate better offspring and mutation alters a string locally to maintain genetic diversity from one generation of a population of chromosomes to the next. In each generation, the population is evaluated and tested for termination of the algorithm. If the termination criterion is not satisfied, the population is operated upon by the three GA operators and then re-evaluated. The GA cycle continues until the termination criterion is reached. In feature selection, Genetic Algorithm is used as a random selection algorithm, Capable of effectively exploring large search spaces, which is usually required in case of attribute selection. For instance; if the original feature set contains N number of features, the total number of competing candidate subsets to be generated is 2N, which is a huge number even for medium-sized N.

The genetic algorithm approach is used along with the feature selection techniques where the feature subsets are reduced to the maximum to obtain the optimal solution. The purpose of using genetic algorithm in the feature selection methods in this study is that the huge datasets are reduced to the optimal size for appropriate segmentation of the images. The output obtained in the feature selection is then clustered by the fuzzy C means clustering.

3.5 Fuzzy c means clustering

Fuzzy clustering methods allow the pixel to belong to several clusters simultaneously, with different degrees of membership. The measure of dissimilarity in FCM is given by the squared distance between each data point and the cluster centre, i.e. the Euclidean distance between them and the distance is weighted by the power of the membership degree at that data point.The fuzzy *c*-means algorithm is very similar to the *k*-means algorithm:

- Choose a number of clusters.
- Assign randomly to each point coefficients for being in the clusters.
- Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than ε, the given sensitivity threshold) :
- Compute the centroid for each cluster, using the formula above.
- For each point, compute its coefficients of being in the clusters, using the formula above.
- The algorithm minimizes intra-cluster variance as well, but has the same problems as *k*-means; the minimum is a local minimum, and the results depend on the initial choice of weights.

3.6 segmentation

Image segmentation is the process of dividing an image into multiple parts which is typically used to identify objects or other relevant information in digital images. There are many different ways to perform image segmentation including Thresholding methods such as Otsu’s method, Clustering methods such as K-means and principle components analysis, Transform methods such as watershed ,Texture methods such as texture filters.

Clustering method such as K-Means is used to cluster the coarse image data. The steps are

- Read the Image by Inputting Colour Image.
- Convert Image from RGB Color Space to L*a*b* Colour Space and calculate the Number of Bins for coarse representation.
- The Window size for histogram processing and the Number of classes are given.
- Classify the Colours in a*'b*' Space Using K-Means Clustering label
- Using the Results from KMEANS every Pixel in the Image create Images that Segment the H&E Image by Color.
- Segment the disease into a Separate Image Output Segmented Image.

The watershed transform finds "catchment basins" and "watershed ridge lines" in an image by treating it as a surface where light pixels are high and dark pixels are low. Segmentation using the watershed transform identifies or mark foreground objects and background locations. Marker-controlled watershed segmentation follows this procedure:

1. **Compute a segmentation function:** The dark regions of the images are the objects that need to be segmented. Read in the Color Image and Convert it to Grayscale. Use the Gradient Magnitude as the Segmentation Function
2. **Mark foreground objects**: These are connected blobs of pixels within each of the objects.
3. **Compute background markers:** These are pixels that are not part of any object.
4. Modify the segmentation function so that it only has minima at the foreground and background marker locations.
5. Compute the watershed transform of the modified segmentation function and visualize the Result.

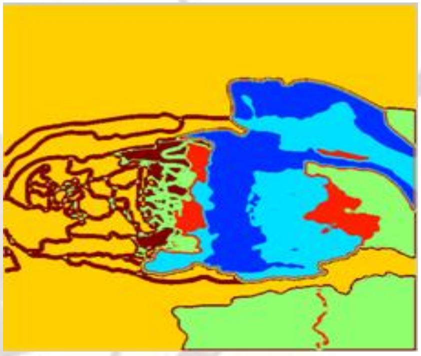

Fig 4. Segmented images

3.7 classifier

Bayesian classifiers assign the most likely class to a given example described by its feature vector. Naïve Bayes is used for classifying the extracted features in this study. The extracted features are classified to the most likely class. Learning in Naïve Bayes is simplified by assuming that the features are independent for a given class. The feature is classified as shown in equation (3):

$$P\left(X\middle|C\right)=\Pi_{i=1}^{n}P\left(X_i\middle|C\right) \qquad (8)$$

Where $X=(X_1,\ldots, X_n)$ is the feature vector and C is a class.

The choice of choosing this classifier is the feature extraction techniques used in this study prefers a suitable classifier that handles all type of value i.e., univariate wrapper, multivariate values and PCA method. The experiments and results shows the results of the work done by the classifier.

IV. EXPERIMENT AND RESULTS

In order to prepare the image for segmentation, pre-processing of the image was done by contrast enhancement and median filtering. Median filter was used for noise removal. Contrast enhancement was performed. The number of features reduced by feature selection methods with Genetic algorithm based Initialization for the optimization of results. Reducing the number of features of dataset is important .All methods were successful in reducing the number of features.

The Fuzzy c means clustering is done to cluster the images is done and segmentation of the images are done. The classification accuracy of datasets with 10-fold cross validation for finding the accuracy of the images are computed.

V. PERFORMANCE ANALYSIS

The correctly and incorrectly classified instances show the percentage of test instances. The percentage of correctly classified instances is often called accuracy or sample accuracy. Kappa is a chance-corrected measure of agreement between the classifications and the true classes. It is calculated by taking the agreement expected by chance away from the observed agreement and dividing by the maximum possible agreement. A value greater than zero means that the classifier is doing better than chance.

The mean absolute error is the sum over all the instances and their AbsErrorPerInstance divided by the number of instances in the test set with an actual class label.

MeanAbsErr = Sum(AbsErrPerInstance) / number of instances with class label.

Root mean squared error, Relative absolute error, Root relative squared error are used to assess performance when the task is **numeric prediction**. Root relative squared error is computed by dividing the Root mean squared error by predicting the mean of the target values .Therefore, smaller values are better and values > 100% indicate a scheme is doing worse than just predicting the mean .Coverage of cases and Mean relative region size shows the numeric level of the cases which gives absolute results.

Table 1 :Comparative Study and Summary of the classification of the images using classifier.

5.1 Performance measures

S. No	Evaluation of testing instances using Naive Bayes classifier	Feature selection methods (in percentage)		
		Correlation based feature selection	Information Gain	Principal Component Analysis
1	Correctly Classified Instances	10 (90.91)	6 (54.55)	6 (54.55)
2	Incorrectly Classified Instances	1 (9.09)	5 (45.45)	5 (45.45)
3	Kappa statistic	0.8493	0.375	0.3529
4	Mean absolute error	0.0455	0.3176	0.2845
5	Root mean squared error	0.2132	0.3989	0.5007
6	Relative absolute error	12.2596	86.4985	78.057
7	Root relative squared error	48.4879	93.096	117.9506
8	Coverage of cases (0.95 level)	90.9091	90.9091	54.5455
9	Mean relative region size (0.95 level)	25	88.6364	31.8182

The above evaluation is done for finding the appropriate result for the methods employed in this study.

The *True Positive (TP)* rate is the proportion of examples which were classified as class *x*, among all examples which truly have class *x*.The *False Positive (FP)* rate is the proportion of examples which were classified as class *x*, but belong to a different class, among all examples which are not of class *x*.The *Precision* is the proportion of the examples which truly have class *x* among all those which were classified as class *x*.

Table 2 . Detailed Performance Accuracy for classes using Naive bayes classifier for Correlation based feature selection, Information Gain, Principal Component Analysis

TP Rate			**FP Rate**			**Precision**			
C1	***C2***	***C3***	***C1***	***C2***	***C3***	***C1***	***C2***	***C3***	**Class**
1	0.67	1	0.17	0.13	0.63	0.83	0.67	0.38	1
0.5	0.5	0.75	0	0.29	0	1	0.5	1	2
0	0	0	0	0	0	0	0	0	3
1	1	0	0	0.22	0	1	0.5	0	4
0.91	0.55	0.55	0.08	0.18	0.17	0.92	0.46	0.47	Weighted avg.

Recall			**F-Measure**			**ROC Area**			
C1	***C2***	***C3***	***C1***	***C2***	***C3***	***C1***	***C2***	***C3***	**Class**
1	0.67	1	0.91	0.67	0.55	0.92	0.83	0.75	1
0.5	0.5	0.75	0.67	0.5	0.86	0.64	0.61	1	2
0	0	0	0	0	0	0	0.58	0.7	3
1	1	0	1	0.67	0	1	0.9	0.38	4
0.91	0.55	0.55	0.90	0.49	0.46	0.90	0.7	0.46	Weighted avg

Where ***C1***- Correlation based feature selection ***C2***- Information Gain ***C3***- Principal Component Analysis

VI. CONCLUSION

The accuracy was found to be 90.91%, 54.55% and 54.55% for correlation based feature selection, information gain, principal component analysis methods with genetic coding respectively. For automatically recognizing the segmented regions, the naive Bayes classifier is used. Performance measure shows the Correlation based feature selection more accurate results than the other two methods.

VII. ACKNOWLEDGEMENT

We would like to take this opportunity to thank Radiologist Dr. Ramesh Kumar M.D.,(R.D) ,Professor and Head, Department of Radiology Sri Manakula Vinayagar Medical college and Hospital Madagadipet for anonymizing the Dicom Images so that they could be used for analysis for providing patient database and helpful discussions regarding the lung diseases.

References

[1] Manish Kakara, Dag Rune Olsen, “Automatic segmentation and recognition of lungs and lesion from CT scans of thorax “, IEEE transactions on Computerized Medical Imaging and Graphics ,33 (2009) 72–82

[2] Ribeiro, M. X.; Balan, A. G. R.; Felipe, J. C.; Traina, A. J. M.; Traina Jr., C.” Mining statistical association rules to select the most relevant medical image features.” ,First International Workshop on Mining Complex Data (IEEE MCD’05), Houston, USA: IEEE Computer Society, 2005, p. 91–98

[3] C. Bhuvaneswari, P. Aruna, D. Loganathan“ Feature Selection Using Association Rules for CBIR and Computer Aided Medical Diagnostic”, International Journal of Computer & Communication Technology ISSN (PRINT): 0975 - 7449, Volume-4, Issue-1, 2013

[4] Uppaluri R, Hoffman EA, Sonka M, Hartley PG, Hunninghake GW, McLennan G.,”Computer recognition of regional lung disease patterns.” American Journal of Respiratory Critical Care Med ,1999;160:648–54.

[5] Shyu CR, Brodley CE, Kak AC, Kosaka A, Aisen AM, Broderick LS. “ASSERT: a physician-in-the-loop content based retrieval system for HRCT image databases”. Computer Visual Image Understanding, 1999; 75: 111–32.

[6] C. Brambilla and S. Spiro” HIGHLIGHTS IN LUNG CANCER” ,Copyright #ERS Journals Ltd 2001 ,European Respiratory Journal ,ISSN 0903-1936

[7] Armato SG, Giger ML, MacMohan H. “Automated detection of lung nodules in CT scans: preliminary results.” ,Med Phys 2001; 28: 1552–61.

[8] Lee Y, Hara T, FujitaH, Itoh S, Ishigaki T,”.Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique”. IEEE Trans Med Imaging 2001;20:595–604.

[9] McNitt-Gray MF, Har EM, Wyckoff N, Sayre JW, Goldin JG. “A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results.“,Med Phys 1999 ; 26: 880–8.

[10] Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B, Henschke CI. “Small pulmonary nodules: volumetrically determined growth rates based upon CT evaluation”,. Radiology 2000; 217: 251

[11] Zagers H, VroomanHA,Aarts NJM, Stolk J,Kool LJS, Dijkman JH, et al. “Assessment of the progression of emphysema by quantitative analysis of spirometrically gated

computed tomography images",. Invest Radiol 1996; 31: 761–7.

[12] Adelson, E. H. and Bergen, J. R. "Spationtemporal energy models for the perception of motion.", Journal of the optical society of america A, 2:284–299.

[13] Rajdev Tiwari and Manu Pratap Singh, "Correlation-based Attribute Selection using GeneticAlgorithm, *",International Journal of Computer Applications (0975 – 8887), Volume 4– No.8, August 2010:28-34*

[14]. I. H. Witten, E. Frank.." Data Mining: Practical machine learning tools and techniques.",2nd Edition, Morgan Kaufman, San Francisco, 2005.

[15] *H.D.Tagare, C. Jafe, J. Duncan*, "Medical image databases: A content-based retrieval approach", Journal of the American Medical Informatics Asssociation,4 (3),1997, pp. 184-198.

[16] Nassir Salman" Image Segmentation Based on Watershed and Edge Detection Techniques"*The International Arab Journal of Information Technology, Vol. 3, No. 2, April 2006, pp.104-110*

[17] A. Jain and D. Zongker, "Feature Selection: Evaluation,Application, and Small Sample Performance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, Vol. 19, No. 2, pp. 153-158.