Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/120220

Identification and Classification of Pulmonary Nodule in Lung Modality Using Digital Computer

Vijay Kumar Thalampathy Jeyapaul[1,*], *Lavanya Naidu* [1], *Khanna Nehemiah H*[2], *Ganapathy Sannasi*[3] *and Kannan Arputharaj*[1]

[1] Department of Information Science and Technology, CEG Campus, Anna Universtiy, Chennai 600025, India
[2] Ramanujan Computing Centre, CEG Campus, Anna University, Chennai 600025, India
[3] School of Computer Science and Engineering, VIT Chennai, India

Received: 14 Jan. 2018, Revised: 22 Feb. 2018, Accepted: 28 Feb. 2018
Published online: 1 Mar. 2018

Abstract: This paper proposes an intelligent approach for the development of a new support system to improve the performance of Computer Aided Diagnosis for automated pulmonary nodule identification on Computed Tomography images which is Digital Imaging and Communications in Medicine format. The first step in diagnosis of any abnormality in lung region, is to acquire a Computer Tomography image, a non-invasive procedure. The digital format of the image is highly portable, hence the extraction and sharing of valuable knowledge. The large number of related images pose a challenge in coherence and consequently arriving at conclusion. The CAD system has been designed and developed to segment the lung tumour region and extract the features which is of region of interest. The Detection process consists of two steps, namely Lung segmentation and Feature extraction. In segmentation of lung region K-means, Watershed and Histogram based algorithms is implemented. The extracted features and the label of the corresponding ROI were used to train a neural network . Finally , these properties are used to classify lung tumour as benign or malignant. The main objective of this method is to reduce false positive rate and to improve the access time and reduce inter-observer variability.

Keywords: DICOM, ComputedTomography, DigitalImages, Support Vector Machines, Backpropagation

1 Introduction

Cancer is the stage in human body when the cancer cells in the human body gets associated in large numbers and tend to grow uncontrollably. Lung Cancer is one of the leading cause of death in both men and women throughout the world as provided by World Health Organization.Treatment to cancer is subjective based on the physiology of the patient. The main cause for Lung Cancer is Tobacco Smoking. The incidence of this cancer is far low in Female when compared to Male. The data that is available throughout the world will help us to uncover valuable information regarding the pattern of incidence.

Lung cancer begins quietly. There are usually no symptoms or warning signs in the early stages. A spiral CT may pick up early lung cancers in some people.The US Preventive Services Task Force recommends annual screening for lung cancer with low-dose CT in adults between 55 and 80 years who have 30 ”pack-year ” smoking history and currently smoke or have quit within the past 15 years.

Oldest description of cancer was discovered in Egypt. Bone Cancer called osteosarcoma have been observed in mummies.

Studies have shown that the use of CT improves confidence in diagnosis, increases the number of ”definitive” lesions by 41% in patients with non-small cell lung cancer.

In computer vision image segmentation is the process of partitioning of a digital image into multiple segments. The goal of segmentation is to simplify and/or change the representation of an image into something more meaningful and easier to analyze.

Among the cancers, the lung cancer death is the most common one statistically, worldwide. Earlier the diagnosis better is the prognosis results. One of the main cause of this deadly disease is the tobacco smoking. Early detection using diagnostic tests promises to reduce mortality from lung cancer. The use of CAD system is

* Corresponding author e-mail: tjvijay@auist.net

imminent and highly-significant due to non-invasive procedure and reliability.

2 RELATED WORKS

There are many works on segmentation of Lung images which are found in the literature[1][2][9].

Sushmita et al [1] proposed one of the earliest active contours algorithm which used Greedy Snake for image segmentation. Boundary Identification in 2D images is implemented using Supervised algorithm. Statistical parameters from extracted image such as mean, standard deviation, skewness, kurtosis, and moments were quantified from the segmented ROI and the last resort of classification is done using back-propagation Neural Network. Sensitivity measurement is the predominant feature measured from the segmented slices that contains cancerous nodules.

Elizabeth et al [2] used a CT image which was pre-processed and segmented the lung parenchyma from each slice using a greedy snake algorithm. In the Lung Image Modality the normal positioning of the lung tissue lies in the range 2900 to 2500 HU, which is a prime factor used to segment the lung. The features are extracted using GLCM with labelling used to train the Neural Network. The accuracy achieved using the said algorithm is said to be 94.44 using RadialBasisFunction Neural Network.

Nameirakpam et al [3] proposed an image segmentation techniques using a clustering method called k-means followed by subtractive clustering algorithm. Prior to clustering the image, Partial Contrast Stretching technique is done to improve quality and contrast of the CT image. Subtractive Cluster is used to generate initial center and these centres are used k-means clustering . Median Filter is applied as the final step to remove any unwanted noise.

Baoan Han et al [4] implemented a watershed segmentation based on morphological gradient to overcome the problem of over-segmentation. Hence, to eliminate the darker pixels and noises in the gradient image, primitive image processing techniques such as the morphology erosion, dilation, opening and closing operations are done.

Rudwan et al [5] proposed local minima watershed transform, the image is first pre-processed using morphological operations as dilation, erosion, opening. The output of the image is further processed under watershed segmentation, the gradient magnitude function is used for the segmentation process. To solve the over-segmentation in Lung image modality, Radial Basis Function Neural Network will be trained andthe two local minima markers of each segment is identified .

Maryam Imami et al [6] implemented a technique using GLCM for Hyper-spectral image Classification. It is the combination of spatial and spectral features for classification improvement. The GLCM feature is used as powerful for the extraction of texture, shape and size from the neighbouring pixels. Thespatial features are extracted from the above techniques and combined with original spectral features for further evaluation. Spatial Distribution of grey-levels in image modality is quantified using GLCM matrix . Each pixel of image features such as correlation, variance, contrast, entropy is calculated. Gabor filter which acts as filter bank to acquire localization properties. Finally, the feature from GLCM, gabor filter is given as fusion to SVM and determines a classification map.

JinsaKuruvilla et.al [7] proposed a novel classification method in computed tomography (CT) images of lungs modality using artificial neural network classifier. In feed forward neural network triggering and flow of information takes place in forward direction only and with inclusion of feedback in back propagation neural networks, thirteen training algorithms was implemented. Traingda is a training function that updates weight and bias values accordingly depending on the gradient descent with adaptive learning rate which is observed to evaluate maximum classification accuracy .

Elmar et al [8] implemented automatic lung nodule segmentation and classification using SVM. ROI in the sample image modality is segmented using thresholding technique and morphological operations. Eventually from the ROI features are extracted through GLCM. Finally, classification of the image is done by identifying Support Vectors defines two hyperplanes, one through the support vector of one class and other that goes through the support vector of other class. Based on hyperplane, the classification output will be concluded as either malignant or benign.

DarmanayagamShiloahElizabeth ,Harichandran Khanna Nehemiah, Cyril Sunil Retmin Raj, Arputharaj Kannan [9] proposed a novel supervised approach for segmentation of Lung Parenchyma from Chest CT for Computer-Aided Diagnosis

Jiantao Pu et al [10] proposed and implemented an efficient lung segmentation algorithm called Adaptive Border Matching (ABM) which smoothes the border of lungs in a geometric way.They achieved an average over segmentation ratio and under-segmentation ratio of 0.43

Soleymanpour (2011) [11] proposed a hybrid algorithm for segmenting the lung region present in computed tomography lung images. In their method, initially the original images were enhanced by a combination of adaptive contrast equalization and non-linear filtering methods.

Jun Lai et al (2012) [12] proposed an effective segmentation method based on Primary Component Analysis (PCA) for segmenting the lungs from computed tomography lung images.

Sasidhar et al (2014) [13] proposed a hybrid algorithm that segments the lung regions which have nodules at the lung borders by using the combination of thresholding and morphological operators. This method removes the lung walls and increases the size of the image to visualize small nodules.

Stefano et al (2011) [14] described a new lung segmentation method and discussed about the most important problems in segmenting lung nodules in CT imaging occurring between nodules and other lung structures such as vessels or pleura. In their method, the problem of vessels attachments is addressed by proposing an automated correction method.

Xiuhua et al(2011) [15] proposed a methodology for accessing the performance of CAD system in detecting nodules. Their analysis was carried out for original CT images versus enhanced CT images. From their analysis, it is established that the image enhancement based on wavelet transform could improve the diagnostic accuracy of CAD for the malignant chest nodules.

Toshiro et al(2011) [16] proposed a new algorithm to detect the solid, non-solid vascularized and juxtapleural nodules present in computed tomography lung images. First, the algorithm separates the lung parenchyma and denser anatomical structures with the combination of coupled competition and diffusion process. Their experiments show that their algorithm is highly reliable in segmenting various types nodules present in the CT lung images. An Intelligent segmentation method for lung analysis was proposed by Khan et al(2014a) [17] using fuzzy bit plane thresholding. In another work by the same author fuzzy logic was used to improve the segmentation accuracy(Khan 2014b)[18].

Du Hongle et al (2009) [19] proposed an improved Fuzzy Support Vector machine through introduction membership to each data point for classification.

SrilathaChebrolu et al (2005)[20]used Baye?s theorem to classify the new instances of data. Each instance is a set of attribute values described by a vector, $X = (x_1, x_2 \ldots, x_n)$. Considering m classes, the sample X is assigned to the class Ci if and only if $P(X|Ci)P(Ci) > P(X|Cj)P(Cj)$ for all j in (1, m). Support Vector Machine can easily achieve the high classification accuracy for classifying the dataset (Gupta 2007).

Harikrishna Rai and Gopalakrishnan Nair (2009) [21] did Research in segmentation methods has resulted in the development of new techniques in CT Angiographic Images.

Research in segmentation methods has resulted in the development of new techniques Sluimer et al (2006)[21] for analysis of many different anatomical regions using image data acquired from a variety of modalities.

Several image processing techniques such as thresholding, edge detection, region growing, and morphological operations have been proposed earlier for the effective segmentation of medical images. Decision tree is also a classifier which is used to accurately classifying the dataset Sindhu et al (2012)[22].

Ganapathy et al (2014)[23] proposed a novel classifier for medical diagnosis decision support system. Particle Swarm Optimization is implemented for accurate detection.Particle Swarm Optimization is used to minimize the number of features extracted .Fuzzy temporal rules and PSO for effective classification of data items.The input dataset is obtained from UCI Machine Learning Repository.

F.V.Farahani et al [24] in their work focused to design a fuzzy rule based medical expert system for diagnosis of lung cancer. This proposed system consists of four modules. The working memory stores the domain knowledge, risk factors and symptoms of lung cancer. The fuzzy inference engine triggers relevant rules based on appropriate condition and provides the probability of disease as output.

Kathikeyan and Ramadoss[25] (2011) proposed a hybrid algorithm to separate the lung tissue from Chest CT images. Their method does an automatic segmentation of lung tissues by the usage of Fuzzy C-Means(FCM) clustering. Finally, morphological operations are used to smooth the irregular boundaries present in the segmented images.

3 SYSTEM ARCHITECTURE

The overall system architecture is referred in the Figure 1. It consists of two main tasks: Detection and Classification of Lung Image Modality. The objective of this work is to make a decision support system for the identification and classification of pulmonary nodule in the lung image as either malignant or benign tumour.

In order to overcome these issues, numerous methods have evolved in decision support system to reduce the false positive rate and hence to classify the nodule as cancerous or non-cancerous. In this proposed system, the objective is to provide fast identification and accurate automated segmentation and thereof reduce the false positive rate using classification techniques with lung Image datasets . The format of lung CT image database is DICOM format which is subjected to further processing to minimize the error rate.

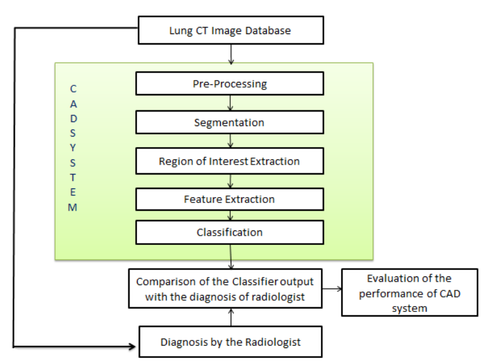

Fig. 1: System Architecture

3.1 PREPROCESSING

Removal of noise is done by the median filter technique which is a non-linear digital filtering technique. B = medfilt2(A, [m n]) performs median filtering in the given CT image where the median value in the m-by-n neighbourhood around the corresponding pixel in the input image modality .

3.2 SEGMENTATION

Three type of algorithms are used for image segmentation in-order to segment pulmonary nodules from the DICOM image to improve the accuracy rate. They are as follows

–K- Means Segmentation
–WaterShed Segmentation and
–Histogram Based Segmentation

3.2.1 K- Means Segmentation

In this research work K-Means algorithm has been implemented, which is a clustering technique. It groups k-clusters, in which a centroid value is calculated for the set of pixels and then clusters are updated. This process is completed when no number of pixels changes the clusters.

This algorithm is used here for grouping of n pixels with given five clusters and find the best results among the clusters which can resulted a pixels how much grouping for each iterations. Based on the particular iterations and resulted output, the image which results better segmentation and the nodules which is clearly identified is taken as output image from segmentation and further processing is developed. The distance between cluster and centroid value is given by Euclidean distance,

$$d_{ij} = ||X_j - V_i|| \tag{1}$$

where $x = x_1, x_2, ...x_n$ is the input data set and vi is the cluster centroid which is refer in the Equation 1. K-means algorithm doesn't uses any initial seed points, but the number of clusters which being used for segmentation are initially assigned by which the maximum and minimum ranges are calculated then mean value is re-updated for the given five iterations which is refer in Equation 2.

$$v_i = \frac{\Sigma_{j=1}^{n} u_{ij} x_j}{\Sigma_{j=1}^{n} u_{ij}} \tag{2}$$

The above Figure 2 shows the k-means output. When the clusters k=5 reaches to a particular state and hence no clusters are repeated, then the final resulted of the image is implemented above for K cluster input. By which one of the proper clustered output is taken for feature extraction.From the binary output of k-means, the eccentricity value lesser than 0.98 pixel nodules and its specific features are extracted shown in the Figure 3

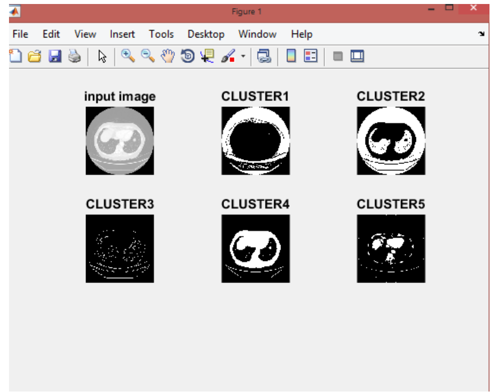

Fig. 2: K ? Means Segmentation

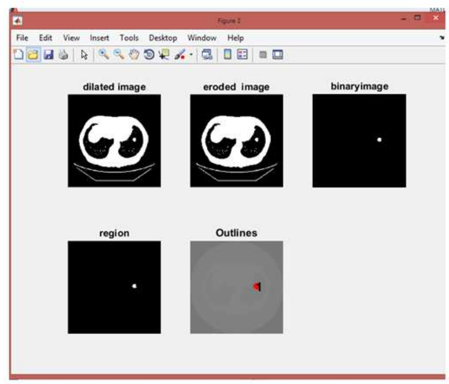

Fig. 3: Nodule Segment using K-Means

3.2.2 WaterShed Segmentation

Watershed algorithm computes a distance transformation from the threshold image. The distance transformation of the binary image is quantified using bwdist(BW). Negating the distance transform the two bright areas into catchment basins using $-bwdist(\ imimBW)$. Watershed function is then applied for a obtained binary image. Label matrix is compared with zero-valued elements to the watershed function and separates the object in the original image. The computed distance transformation from the complemented binary image, is modified to force the background to be its own catchment basin, the watershed transformation is computed. The new function label2rgb is used to display the segmented objects using different colors. To detect intensity of the image imextendedmin() and imimposedmin(). imextendedmin() function is a binary image. The imimposemin() function modifies the given input image to subsect only those region of interest found by the imextendedmin() function. The imimposemin() function also changes a region's pixel values to zero. The imposed minima was calculated by watershed transform which is contained in the Region -Of-Interest.

The mean and standard deviation value is calculated, based on which the distance transform binary image is calculated. Watershed function is applied which outputs a label matrix containing whole numbers that correspond to watershed regions. The irrelevant pixels are removed from the binary image by producing new binary image by the function area opening as shown in the Figure4.

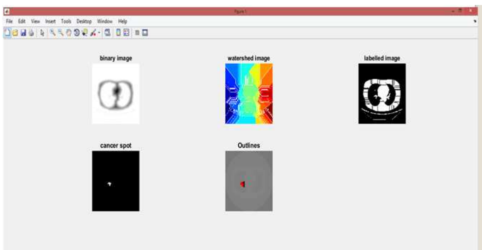

Fig. 4: WaterShed Based Segmentation.

3.2.3 Histogram Based Segmentation

Thresholding is set based on the histogram of the image using imhist() function. Based on the size of histogram image the mean, standard deviation and variance is calculated to set threshold value. Binary image is obtained by comparison between input image and threshold value. Finally, Morphological operations is performed to determine the region of interest. Mean value is calculated by the sum of row vector (the sum of each column) divided by the sum of column. Variance returns the sum of rows to the sum of columns with difference of mean values. Threshold value is determined by the sum of mean and standard value. Morphological operations is used for extracting shape information with the help of a structuring element followed by dilation, erosion and opening.

The implementation results are shown in Figure 5.

Dilation: I=$imdilate(I, strel('arbitrary', NHOOD))$ which dilates the input binary image and the argument SE is a structuring element object. Arbitrary creates a flat structuring element where NHOOD specifies the neighbourhood element.

Erosion: I=$imerode(I, strel('arbitrary', NHOOD)$ performs binary erosion, the input is the argument SE, a structuring element object or array of structuring element objects which is returned by the strel() function.

Opening: I=$bwareaopen(I, p)$ is a function that removes small irrelevant objects from the binary image under study. It also sheds out all connected components that have fewer than P pixels from the binary image, hence outputting another binary image.

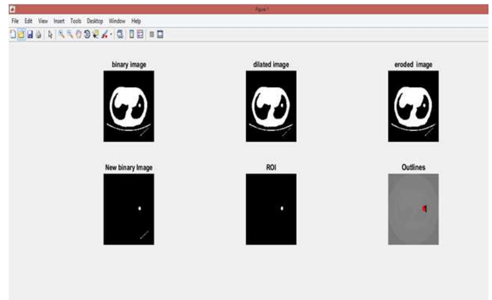

Fig. 5: Histogram Based Segmentation

4 FEATURE EXTRACTION

After the proper segmentation of Lung from the input CT image modality the next step is the feature extraction. In this process, from the region of interest quantification of the pixel density is calculated. The relationship among the neighbouring pixels and the tumour pixels were calculated using this feature extraction method. These quantified extracted features provides valuable insight from which invaluable decisions were made. It also helps in the reduction of inter observer-variability. Rules were framed to zero-in the classification of the nodules .

In this research work features were extracted by regionprops and GLCM matrix. Subsequently, labelling of the segmented image was done to obtain a label matrix, where the variable specifies the 8-connected objects in the neighbourhood . Then regionprops function in MATLAB was used to ascertain the set of properties specified by each connected component (object) from the given binary image.

The pulmonary nodule develops to be either a regular or irregular shape. Hence, the shape measurement parameter eccentricity was calculated. The output of which is a scalar value that quantifies the eccentricity of the ellipse that has the same second-moments as the region. The populated value ranges between 0 and 1, which are real numbers. The threshold value of the eccentricity was observed to be less than 0.98.

4.1 Grey Level Co Occurance Matrix

A statistical method to determine the spatial relationship among the pixels is the gray-level co-occurrence matrix, which is known by the other name, gray-level spatial dependence matrix. It calculates the recurrence of pixels, falling in a specified range in the neighbourhood with spatial relationship. The number of gray levels in the image determines the output obtained from the GLCM.

Contrast measures the local variations in the computed gray-level co-occurrence matrix(GLCM). It computes the measure of the intensity contrast between a pixel and its neighbourhood and the whole image. The value of Contrast is 0 for a constant image. It is given by the Equation : 3

$$\sum_{i,j} |i-j|^2 p(i,j) \tag{3}$$

where, 'i' and 'j' are x coordinate and y coordinate of the pixel , respectively. p(i,j) is the pixel value at the given location.

Homogeneity is the measurement of the closeness of the distribution of elements in the GLCM to its diagonal. The value ranges from 0 to 1. Homogeneity value is 1 for a diagonal GLCM. It is computed by the Equation : 4.

$$\sum_{i,j} \frac{p(i,j)}{1+|i-j|} \tag{4}$$

The randomness of a gray-level distribution in an image is given by Entropy. The entropy value is high if the gray levels are distributed randomly throughout the image. It is computed by the Equation 5.

$$E = -\Sigma_i^m \Sigma_j^n P[i,j] logP[i,j] \tag{5}$$

The quantitative measurement of pixel pair repetitions in the image is known as Energy. The main objective of Energy is to detect disorders in textures. The value of Energy parameter reaches to a maximum value that equal to one, the occurrence of which will occur when the gray level distribution has a constant or periodic form. It is computed from the Equation 6.

$$Energy = \Sigma_i \Sigma_j (g_{i,j})^2 \tag{6}$$

The features of the nodule are summarized as below

Energy	Entropy	Homogeneity	Contrast
9.74E-01	8.07E-02	9.96E-01	2.26E-01
9.87E-01	4.48E-02	9.98E-01	1.26E-01
9.96E-01	1.75E-02	9.99E-01	3.32E-02
9.97E-01	1.46E-02	9.99E-01	4.09E-02
9.81E-01	6.51E-02	9.96E-01	2.42E-01
9.94E-01	2.34E-02	9.99E-01	2.89E-02
9.89E-01	3.83E-02	9.98E-01	1.14E-01
9.99E-01	4.39E-03	1.00E+00	1.28E-02
9.99E-01	4.01E-03	1.00E+00	1.53E-04
9.95E-01	1.84E-02	9.99E-01	4.47E-02
9.99E-01	4.39E-03	1.00E+00	1.28E-02

Fig. 6: Features of Nodule

5 CLASSIFICATION

5.1 FEED FORWARD NEURAL NETWORK.

Feed Forward Neural Network is the prominent network architecture for the classification of labels. It is designed to have an input layer, one or two hidden layers and an output layer. The values are populated in the initial input layer and cascaded down to the next layer upon satisfying the triggering function and threshold value. Consequently, the output from the previous layers are populated as input to the current layer which are built with certain number of nodes. The output values from the output layer determines the classification of the labels which is application specific.

In this research work, the neural network consists of three layers, the input layer passes the output of input nodes to the hidden layer by the summation of weights between input and hidden nodes and the inputs in the input layer along with the bias acts as a threshold that serves to vary the activity of unit which is computed by the Equation 7.

$$I_j = \sum_i W_{i,j} O_i + \theta_j . \tag{7}$$

In this research analysis of 220 datasets , 150 datasets were used for training and 70 were used for testing. The Input layer was designed to have four nodes which represents the features extracted such as energy, entropy, homogeneity and contrast. The hidden layer consists of four nodes and single output node determines the nodule as malignant or benign. The sigmoid function is used as an activation function. The net input I_j to the unit j, then O_j, the output of unit j is given in the Equation 8.

$$O_j = 1/1 + e^{-} I_j \tag{8}$$

In the back propagation neural network the error-correction is propagated backward by updating the weights and biases for the network prediction. For a unit j in the output layer, the error in the node in the current layer, Err_j is computed by the Equation 9.

$$Err_j = O_j(1-O_j)(T_j - O_j) \tag{9}$$

where O_j is the actual output of unit j, and T_j is the known target value of the given training tuple.

To compute the error at the hidden layer unit j, the weighted sum of the errors of the units connected to the next layer, unit j is taken into consideration. The error calculation at the hidden layer unit j is given by Equation 10.

$$Err_j = O_j(1-O_j)\Sigma_k Err_k W_{jk} \tag{10}$$

where $w_j k$ is the weight of the inter-linking from unit j to a unit k in the previous layer, and Errk is the error at unit k. The weights and biases are updated to reflect the propagated errors which is given in the Equation 11.

$$\Delta Wij = (l) Err_j O_i w_{ij} = w_{ij} + \Delta Wij \tag{11}$$

The variable l is the learning rate, is a real value ranging between 0 and 1. The learning in Back propagation Neural Network is implemented using a method of gradient descent to search from a population of weights that fits the training data which try to minimize the mean squared distance between the network's class prediction and the known target value of the tuples.

Biases are updated after the output is obtained from the output layer during each iteration. It is computer using the below mentioned Equation 12, where $??_j$ is the change in bias $?_j$

$$\Delta\Theta_j = (l)Err_j\Theta_j = \Theta_j + \Delta\Theta_j \qquad (12)$$

5.2 SUPPORT VECTOR MACHINE

Support Vector Machine Classifier is also used for the classification of tumour as either benign or malignant. The trained features from GLCM output, xi and its associated label yi are given as input for the svmtrain function yi have two possible values -1 or +1. Kernel function in Support Vector Machine is implemented using svmtrain which is used to map the training data into kernel space. Linear kernel function uses a dot product and is calculated by the given Equation 13.

$$c = \Sigma_i \alpha_i k(s_i, x) + b \qquad (13)$$

where s_i are the computed support vectors, $?_i$ are the weights, b is the bias value assigned and k refers to the kernel function designed. In the case of a linear kernel, k is the dot product. Based on the value of c, x is classified as a member of the first group or otherwise as second group. For each vector x_i is given by,

For each vector x_i given by,

$w.x_i + b \geq 1$ for x_i class 1.

$w.x_i + b \leq -1$ for x_i class 0.

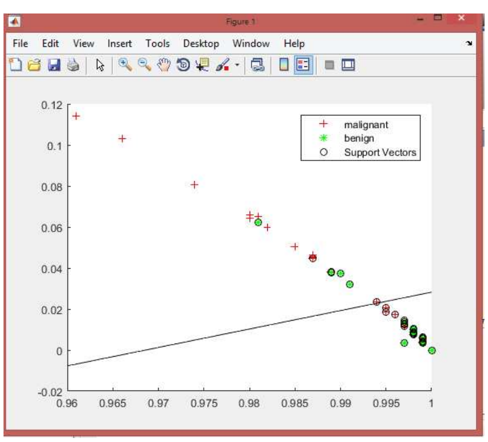

Fig. 7: Output of Support Vector Machine.

Therefore for the trained input features, its labels are assigned as input to the SVMStruct. It contains information about the trained SVM classifer. After normalization of data points, each row corresponding to a support vector is assigned to the normalized data space. The support vector points are shown in the output using showplot function, if the normalized input is true. SVMclassify classifies the tumour and plots the data in the figure using the Showplotproperty . This plotting is drawn to 2-D scale.

Table 1: Accuracy rate

Classification	**Accuracy**
BPNN.	98.721
SVM	73.33

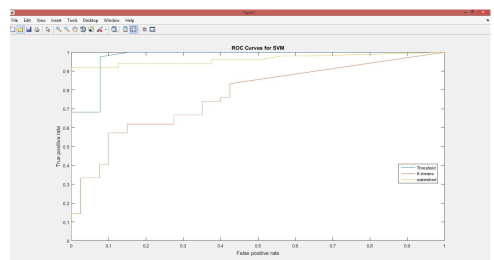

Fig. 8: ROC Classification using Support Vector Machine

6 Perspective

The CT lung image is being pre-processed by median filter and it was segmented using three algorithms.The nodule features were extracted by Region Measurements and Grey-Level Co-Occurrence Matrix. Feed-Forward Back Propagation neural network and Support Vector Machine were used to classify the nodules. FFBNN is implemented with variable learning rate, changing bias and weights. This provides accuracy of 98 percent. The accuracy of SVM classifier is 73 percent.In Future Work, SVM classification accuracy can be improved using support vectors with different kernel functions along with 3D hyperplane. This work implements linear kernel function in 2D hyperplane using dot product method for the support vectors which separates the binary class.

References

[1] Sushmita Mitra, B.Uma Shankar. Medical Image Analysis for Cancer Management in Natural Computing Framework. *Elsevier, Information Sciences*, pp: 111–131, 2015.

[2] Elizabeth. D.S, H.K. Nehemiah, C.S. Retmin Raj, A. Kannan. Computer Aided Diagnosis of Lung Cancer based on Analysis of the Significant Slice of Chest Computed Tomography Image. *IET image processing*, Volume 6, Issues 6, pp: 697?-705, 2012.

[3] NameirakpamDhanachandra, KhumanthemManglem and YambemJinaChanu. Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm *Elsevier, Procedia Computer Science*, pp: 764?771, 2015

[4] Baoan Han Watershed segmentation Algorithm Based on Morphological Gradient Reconstruction. *International Conference on Information Science and Control Engineering*, 2015.

[5] Rudwan A. Husain, Ali S. Zayed, Wesam M. Ahmed, and Hanan S. Elhaji Image Segmentation with Improved Watershed Algorithm Using Radial Basis Function Neural Networks. *International conference on Sciences and Techniques of Automatic control*, pp: 21-23, 2015.

[6] Maryam Imami, Hassan Ghassemian. GLCM, Gabor, and Morphology Profiles Fusion for Hyperspectral Image Classification. *ICEE*, 2016.

[7] JinsaKuruvilla, K. Gunavathi. Lung Cancer Classification using Neural Networks for CT Images. *Elsevier, Computer Methods and Programs in Biomedical*, pp: 202?-209, 2013.

[8] Elmar Rendon-Gonzalez and VolodymyrPonomaryov. Automatic Lung Nodule Segmentation and Classification in CT Images Based on SVM. *IEEE, InstitutoPolitecnico Nacional de Mexico*, 2016 .

[9] Jiantao Pu, Justus Roos, Chin Yi, A, Sandy Napel, Geoffrey Rubin, D, & David Paik, S 2008 , ?Adaptive Border Matching Algorithm: Automatic Lung Segmentation on Chest CT images?, *Computerized Medical Imaging and Graphics*,vol. 32, no. 6, pp. 452-462.

[10] Soleymanpour 'Fully Automatic Lung Segmentation and Rib Suppression Methods to Improve Nodule Detection in Chest Radiographs', *Journal of Medical Science and Signals.* 2011,no 3,pp. 191-199

[11] Jun Lai et al (2012) 'Automatic Segmentation for Pulmonary Vessels in Plain Thoracic CT Scans', , *Indonesian Journal of Electrical Engineering and Computer Science,* vol 10,No 4.

[12] Sasidhar et al, A Novel Method for Classification of Lung Nodules as Benign and Malignant using Artificial Neural Network ', Journal Of Engineering And Computer Science ISSN:2319-7242 Volume - 3 Issue - 8 August, 2014 Page No. 7641-7645.

[13] Stefano et al, ' Automated Segmentation Refinement of Small Lung Nodules in CT Scans by Local Shape Analysis' vol 58,Issue 12,2011.Page No : 3418-3428.

[14] Xiuhua et al, ' Computer-Aided Diagnosis for Early-Stage Lung Cancer Based on Longitudinal and Balanced Data ' 2013, doi.org/10.1371/journal.pone.0063559

[15] Toshiro et al, Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models', Image Analysis, Volume 15,Issue 1,2011, Pages 133?154.

[16] Khan et al, Intelligent Segmentation of Medical Images using Fuzzy Bitplane Thresholding', Measurement Science Review, Volume 14, No. 2, 2014

[17] Du Hongle el al Fast SVM incremental learning based on clustering algorithm' , *IEEE International Conference on Intelligent Computing and Intelligent Systems,* 2009,Vol 4.

[18] SrilathaChebrolu et al, Feature deduction and ensemble design of intrusion detection systems ', Computers & Security (2005) 24, Elsevier,pp 295-307.

[19] G.N.Harikrishna Rai and T.R.Gopalakrishnan, Gradient Based Seeded Region Grow method for CT Angiographic Image Segmentation', Computer Science and Networking, Volume 1, Issue 1, pp 1-6, 2009.

[20] Sluimer et al, Computer analysis of computed tomography scans of the lung', Transactions on Medical Imaging. vol 25, Issue 4,2006.

[21] Sindhu et al,'Decision tree based light weight intrusion detection using a wrapper approach ', vol 39,Issue 1,2012,pp 129-141.

[22] Ganapathy et al, An intelligent temporal pattern classification system using fuzzy temporal rules and particle swarm optimization, vol 39, Issue 2,Pages 283-302.

[23] Farahani et al, ," Lung nodule diagnosis from CT images based on ensemble learning ", Conference on Computational Intelligence in Bio Informatics and Computational Biology 2015

[24] Karthikeyan and Ramados Segmentation Algorithm for CT Images Using Morphological Operation and Artificial Neural Network, International Journal of Computer Theory and Engineering, Vol. 3, No. 4, August 2011

[25] Darmanayagam Shiloah Elizabeth, Harichandran Khanna Nehemiah, Cyril Sunil Retmin Raj, Arputharaj Kannan , " A Novel Supervised Approach for Segmentation of Lung Parenchyma from Chest CT for Computer-Aided Diagnosis." J. Digital Imaging vol 26, No 3, Springer, pp 496-509, 2013.

Vijay Kumar Thalampathy Jeypaul has completed his B.E in Computer Science and Engineering, M.E in Computer Science and Engineering from Thiagarajar College of Engineering, Madurai, India affiliated to Anna University, Chennai, India. Currently he is working as Teaching Fellow in the Department of Information Science and Technology, College of Engineering Guindy Campus, Anna University, Chennai, India. His areas of interest are Medical Image Processing, Soft Computing and Data mining.

Lavanya Naidu has completed her B.E in Electronics and Communication Engineering from Pallavan College of Engineering Kancheepuram, Affiliated to Anna University,Chennai,India and M.E in Multimedia Technology from College of Engineering, Guindy campus, Anna University, Chennai,India during 2017. Her areas of interest include Image Processing, Soft Computing and Multimedia Technology.

S. Ganapathy has completed his M.E. and PhD from Anna University, Chennai. Currently he is working as an assistant professor in VIT University-Chennai Campus. He has published more than 40 papers in journals and conference proceedings. His areas of interest are Wireless Sensor Network and Data Mining.

H. Khanna Nehemiah has completed his M.E. from University of Madras and PhD from Anna University, Chennai. Currently he is working as a Associate Professor in CEG Campus, Anna University, Chennai. He has guided and completed more than 10 PhD Scholars and published more than 50 research papers. His areas of interest are are Image include Wireless Sensor Networks and Medical Image Processing.

A. Kannan has completed his M.E. and PhD from Anna University, Chennai. Currently he is working as a Professor in CEG Campus, Anna University, Chennai. He has guided and completed more than 30 PhD scholars and published more than 300 research papers. His areas of interest are Database Management Systems, Artificial Intelligence, Wireless Sensor Networks and Data Mining.