

# Automatic Detection And Classification Of Pulmonary Nodules On CT Images

Manikandan N, Usha Kingsly Devi K

**Abstract**— Computer-aided detection (CAD) systems are convenient for the automatic lung nodule detection in computed tomographic (CT) images, as the sheer volume of information present in CT datasets is overwhelming for radiologists to process. First, segmentation scheme is used as a preprocessing step for enhancement. Then, the nodule candidates are detected by Eigen value decomposition of hessian matrix and Multi-scale dot enhancement filtering. After the initial detection of nodule candidates using filtering technique, feature descriptors were extracted. The feature descriptor is refined using the process of wall detection and eradication. An Evolutionary Support Vector Machine (ESVM) is trained to classify nodules and non-nodules. The proposed CAD system is validated on Lung Image Database Consortium (LIDC) data. Experimental results show that the detection scheme achieves 98.3% sensitivity with only 11false positives per scan.

**Index Terms**— CT; Pulmonary nodule detection; CAD; Feature extraction;

## I. INTRODUCTION

In this modern era the total number of deaths caused by cancer [10] is increasing day by day. Lung cancer is the most common and fatal cancer in the world. Usually, lung cancer does not cause symptoms early in the disease process, and is mostly diagnosed at a late stage in a clinical setting, when the probability of cure is rare. At the time of diagnosis, most patients are already present with advanced disease. It is expected that screening can detect lung cancer at an early stage and reduce mortality. The goal of CAD [12] is to assist the radiologists in increasing the scanning efficiency and potentially improving nodule detection.

Generally the nodule detection system comprises of three steps namely lung segmentation, nodule candidate detection and classification. Several researchers have presented a variety of methods for segmenting [17] the lung volume from a pulmonary CT scan. The segmentation is usually carried out by thresholding. The various thresholding [2,5] schemes have been implemented. After thresholding, the lung volume is then extracted from the segmented images using 3D approaches [8].

Manuscript received May 18, 2014.

Manikandan N, Student, Applied Electronics, Regional Centre Anna University: Tirunelveli, Tamilnadu, India

Usha Kingsly Devi K, Assistant professor, Applied Electronics, Regional Centre of Anna University: Tirunelveli, Tamilnadu, India

The lung volume is segmented without artifacts by 3D-connected component labelling [6, 15]. The extracted lung volume needs to be refined to include juxta-pleural

nodules. Subsequently, due to the complexity of these approaches, several methods have been presented for refining a lung mask. Recently, the application of a chain code representation over a lung mask was also proposed an attempt to correct the contours.

In the segmented lung volume, nodule candidates have been detected using various methods. Here Eigen value decomposition of hessian matrix and Multi-scale dot enhancement filtering are applied to detect and segment nodule candidates. After nodule candidate's detection, there are many false positives that require elimination. False positives are eliminated by feature extraction and classification techniques.

The features are extracted from the detected nodule candidates. In this detection scheme Angular Histograms of Surface Normal feature [7] (AHSN) are extracted. Finally, nodules are detected with Evolutionary support vector machine classifier [20, 21] using the extracted feature, yielding minimal number of false positives.

## II. MATERIALS AND METHODS

### A. Proposed pulmonary nodule detection scheme

The proposed nodule detection scheme comprises of three main steps namely Lung volume segmentation, Feature extraction and ESVM [20, 21] based classification.

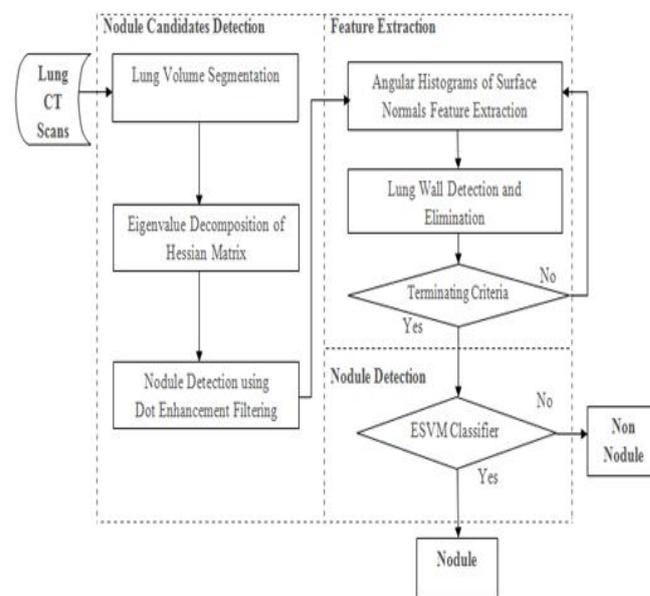


Fig -1: Overall pulmonary nodule detection scheme

a) Lung Volume Segmentation

Lung volume segmentation [17] is the pre-processing step in detection process. It comprises of three steps namely Thresholding, Lung region extraction and Contour correction [14].

(1) Thresholding

Thresholding is performed primarily in lung volume segmentation to discriminate low-density regions from high-density regions. The high-density regions principally comprise of the body surrounding the lung cavity, whereas the low-density regions enclose the lung cavity, the air neighboring the body, and other low-intensity areas. The 3D volume of a CT scan is indicated as  $Im(x, y, z)$ , where the x and y indices signify the slice coordinates, and z denotes the slice number. The volume consists of the total number of Z slices, and each slice has dimensions of  $X \times Y$ . A fixed threshold value has been used to segment the lung volume. As a result of thresholding initial lung mask  $M_i$  is obtained.

$$M_i(X, Y, Z) = Im(X, Y, Z) < -500 \text{ HU} \quad (1)$$

The chest wall, blood, and bone are so dense thus the threshold value is selected as -500 HU.

(2) Lung region extraction

After thresholding the lung region is extracted by 3D connected component labelling. A 3D-connected component labeling is applied to ensure region connectivity over the thresholded volume  $M_L$ . 3D-connected components are obtained by using an 18-connected neighborhood. The 18 connectivity voxels is shown in fig 2, here black point is the center point, and the 18 white points are the neighborhoods. The lung areas are chosen from labeled volumes based on size of their volume. The labeled volumes (L) are obtained. Air in the environs of the body is easily evacuated, because it is attached to the boundary of the volume. The largest and second-largest volumes in (L) as the lung region are selected. The computational difficulties are reduced by selecting the labeled volumes at the median slice. Since the median slice possess only a few other non-body components. During the volume selection utmost of the undesirable non-body components are neglected. Thus the air outside the body and gas in the intestine are removed. At this instant, the lung regions comprehend tiny holes, which are generally nodules are vessels. Then morphological hole-filling operation is performed, as these holes should be contained.

The extracted lung volume are combined as follows

$$M_L = L_1 \cup L_2 \quad (2)$$

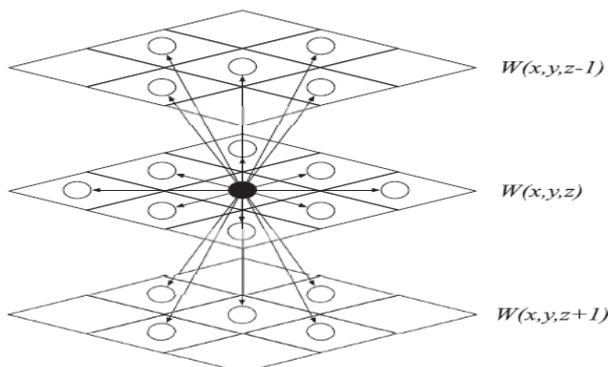


Fig-2: 18 Connectivity voxels

(3) Contour correction

The contour-refined lung volume is obtained using a contour correction method [6] based on chain code analysis. The extracted lung masks volume (SI) is not even and does not contain juxta-pleural nodules, which may impact the system performance. To obtain a even lung mask and to add juxta-pleural nodules in the lung volume, contour correction is performed to the initial lung mask. In this case, a chain code representation is used to remove the critical section. The eight chain codes considered are:  $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$  and  $315^\circ$ .

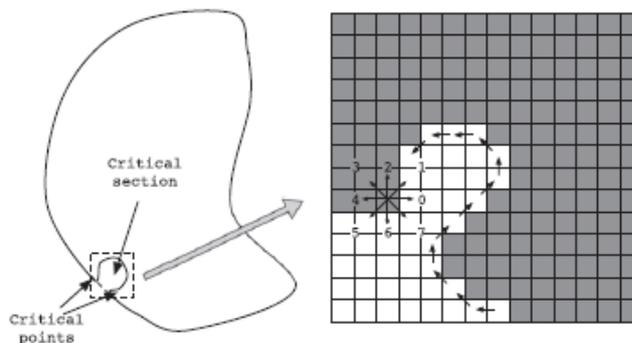


Fig-3: Contour correction using chain code representation

The chain code representation used to remove critical section is shown in fig.3. The noise in the contours of initial lung mask is eliminated by Gaussian smoothing filtering. The critical section is derived from its respective critical points. These are identified by determining the transition of the angular direction of the contour. If the span between a pair of critical points is smaller than the conventional nodule diameter, this pair is selected for critical section correction. The next step is then to unite respective pairs of critical points, and to stuff the critical sections.

b) Nodule candidate detection

The nodule detection is the vital step in the overall detection scheme, and the CAD systems performance mainly depends on the accuracy of nodule candidates detected. In this method nodule candidate detection method, local structure information of each voxel determined by Eigenvalue decomposition of Hessian matrix and nodule candidates are detected by multi-scale dot enhancement filtering.

(1) Eigenvalue decomposition of Hessian matrix

Local structure information is derived from the eigenvalue and eigenvectors which is obtained as a result of Hessian matrix decomposition. Gradient information relates the structure of objects in an image, identifying features or providing basic information for computer vision application. Hessian is a square matrix of second-order partial derivatives of a function. The Hessian matrix, also referred to as the second-moment matrix, it depicts the local curvature of a function of many variables. Gaussian smoothing over a number of scales is applied over the 3-D image to eliminate noise, prior to gradient calculation.

The Hessian matrix H is decomposed using eigenvalue decomposition, yielding three eigenvalues ( $\lambda_1, \lambda_2$  and  $\lambda_3$ ) and eigenvectors ( $e_1, e_2$  and  $e_3$ ). The Hessian matrix HM is decomposed by the following equation:

$$HM = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \lambda_3 e_3 e_3^T \quad (3)$$

Hence explicit structural information about the surfaceness, curvedness and pointedness are obtained.

(2) Multi-scale dot enhancement filtering

The dot enhancement filter is used to enhance spherical objects to identify nodules. The dot value for each and every voxel is defined as

$$Z_{dot}(\lambda_1, \lambda_2, \lambda_3) = \frac{|\lambda_1|^2}{|\lambda_2|} \quad (4)$$

Where  $\lambda_1, \lambda_2$ , and  $\lambda_3$  ( $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$ ) are three eigenvalues derived from the Hessian matrix.

The diameter of the nodules is assumed to be in the range of  $[d_0, d_1]$ , the N discrete smoothing scales ( $\sigma_n$ ) in the range of  $[d_0/4, d_1/4]$  is calculated as

$$\sigma_n = r^{n-1} \frac{d_0}{4} \quad (5)$$

Where  $r = (d_0/d_1)^{1/(N-1)}$ , and each scale has nodule diameter ( $4\sigma$ ).

Here the detection scheme involves five smoothing scales in the nodule diameter in the range of [3mm, 30mm]. By evaluating the dot value in dot enhanced image, the location of nodule candidates are found. The nodule candidates are detected by means of using a threshold value in the dot-enhanced image. The threshold is obtained by averaging local maximum dot values. The threshold value varies for each and every scale. The position of the nodule candidates are detected based on the local maximum dot value. The image section is derived as the nodule candidate from the identified position. The dimension of the image section is  $d_n \times d_n \times d_n$  in isotropic-sized voxels. The size of the image section in accordance with smoothing scale is given by relation

$$d_n = [4\sigma_n I + B] \quad (6)$$

Where I is the interpolation element of each direction for isotropic-sized voxels, B represents the boundary pixels around the desired object, and braces denote the ceiling function. The interpolated image section  $I_s$  is used as input for feature extraction.

c) Feature Extraction

Features are useful information that describes characteristics of the nodule candidates. In the detection system, these features are used to train the ESVM. The detected nodule candidates are considered as nodules or non-nodules using the extracted feature information. The shape based features [3] are extracted. Features are constructed from surface constituent (surface saliency and surface normal vector) that are obtained through eigenvalue decomposition of the Hessian matrix H. The shape based descriptor relates the shape of the desired object based on the orientation probability of the surface normal. Hence, the surface saliency and surface normal vector is obtained from the input image. Thus to compute the surface normal, the eigenvalue

decomposition of the Hessian matrix H is utilized to every voxel in the desired image.

The angular histograms of surfaceness information are obtained, to characterize the shape of desired object. These histograms indicate the angular direction of the surface normal vector on the respective surface saliency. The orientation of the surface normal vector is to be acquired before calculating the AHSN feature. The orientation is denoted by the altitude  $\theta$  and azimuth  $\phi$  in spherical coordinates:

$$\theta = \cos^{-1}(e^{(x)}) \quad (7)$$

$$\phi = \tan^{-1}\left(\frac{e_1^{(y)}}{e_1^{(x)}}\right) \quad (8)$$

The altitude  $\theta$  is diverse in the range  $[0, 180]$  degrees, and the azimuth  $\phi$  is motley in the range  $[0, 360]$  degrees.

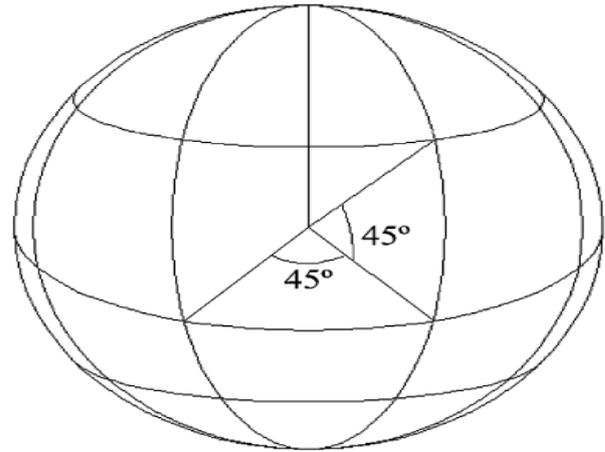


Fig-4: separating the altitude  $\theta$  and azimuth  $\phi$  into  $45^\circ$  in the spherical coordinates.

The shape features are extracted by arranging the formed surface normal vectors into bins which rifts the azimuth and altitude into  $45^\circ$  sections. Thereby an altitude orientation  $\theta$  histogram with  $n$  bins, with each bin covering  $(180/n)$  degrees is generated. Each sample in the image section is added to a histogram bin. The state of the histogram bin is weighted by its surfaceness saliency, and normalized by the sum of surfaceness saliencies in the image section. Likewise, the azimuth orientation  $\phi$  is quantized into  $n$  bins, with each bin covering  $(360/n)$  degrees, and each sample in the image section added to a histogram bin [19] is weighted and normalized. Thus, the dimension of the feature descriptor is  $2n$ , and the extracted AHSN feature is scale-invariant. Hence the shape of the desired object is derived using the AHSN feature descriptor.

(1) Depuration of feature descriptor

Wall detection and elimination technique is implemented to refine surface based feature descriptor. The presence of lung wall causes adverse impact in the nodule detection scheme. Wall elimination paves way for accurate nodule detection. The presence of walls may affect the shape feature descriptor; generally walls bear larger surface areas than other entities. The wall elimination method is used to detect non-isolated nodules.

Initially it is essential to detect and eradicate walls. Walls are detected by finding the local maxima on AHSN.

Consequently connected component labelling is applied to voxels having similar normal vector orientations to the peaks. The surfaces with similar normal vectors are reconstructed. The reconstructed surface is the wall which is eliminated, if the surface is larger than the other parts of the lung. The above step is repeated until there is no wall. Thus the wall elimination method effectively eliminates unnecessary walls near the desired object.

The wall identification and eradication algorithm is described in algorithm 1

Algorithm 1

```

1: procedure                               »Removing walls in
    WALL ERADICATION (IS)                nodule candidates
                                           Image section.
2: {Ssurface,                               »Hessian matrix
    Nsurface} ← STD(IS)                  decomposition
3: X ← AHSN (Ssurface,                       » Derive AHSN feature
    Nsurface)                             Descriptor
4: Nwall ← ∅
5: reprise
6: {θmax,                               »Identify the walls
    φmax} ← find high Peak(x)
7:                                           »Label connected
    la ← labelling( Nsurface                components that have
    θmax, φmax)                             alike orientations to
                                           Surface normal to
                                           { θmax, φmax }
8: for all l ∈ la do
9:   if region (l) > Tr then
10:  Nwall ← Nwall ∪ l
11: end if
12: end for
13:                                           » Derive AHSN feature
    X ← MaskedAHSN(Ssurface,                descriptor without wall
    Nsurface, Nwall)                       area
14: til there are no high
    Peaks in X
15: return X                               » Wall-eradicated
                                           AHSN Feature
                                           descriptor
16: end
    
```

d) Evolutionary support vector machine classifier

In order to classify the pulmonary nodules, Evolutionary support vector machine classifier is utilized. SVMs are supervised learning models with associated learning algorithms that evaluate data and distinguish patterns. Support Vector Machines (SVM) can be trained with different Kernel types along with various selection of parameters. A generic form of the SVM uses the Radial Basis Function (RBF) as the Kernel. For the traditional SVM, the values of control parameters such as box constraint C and kernel parameter  $\gamma$  must be specified. Contrary to the SVM, the input of ESVM is composed of the training data only where the values of control parameters would be tuned automatically by Genetic Algorithm(GA).

(1) Classifier training

During the classifier training stage, the dataset comprising of feature vectors is constructed. A balanced dataset is constructed to attain better training. The dataset is balanced by selecting N/2 nodules and N/2 non-nodules randomly from the detected nodule candidates. The balanced dataset is divided into training and testing datasets to validate the classifier. The training dataset,  $\mathbf{X} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ , is created by selecting N nodule candidates, with each training data pair consisting of an input feature vector and its corresponding known desired class.

High performance of ESVM takes place primarily from optimization of parameter setting of SVM.

Radial basis function kernel is given by

$$K_r(X_i, X_j) = \exp(\gamma \|X_i - X_j\|_2^2) \tag{9}$$

ESVM has a model selection tool using the RBF kernel for the box constraint C and kernel parameter  $\gamma$  and optimizes parameters (C,  $\gamma$ ) to improve performance. ESVM makes use of GA in optimizing system parameters by creating an efficient GA chromosome representation as well as an intelligent crossover operation. The procedure of ESVM is given as follows:

- 1) Initialize a random population of chromosomes. Each chromosome consists of a pair of values ( $\gamma$ , C).
- 2) Fitness for a chromosome is defined as the

$$Fitness = \frac{(sensitivity+specificity+accuracy)}{3}$$

- 3) A new generation is created by the following procedure.
  - a. The best chromosome is copied to the next generation
  - b. Replication with a probability Pr a chromosome is selected with a probability in proportion to its fitness value and it is copied to the next generation.
  - c. Crossover with a probability Pc, two chromosomes are selected with roulette wheel selection method in proportion to the fitness values. A new chromosome is created in the next generation by combining the values.

d. Mutation, a chromosome is selected with a probability  $P_m$  in proportion to its fitness value and a random change is made to create a new chromosome in the next generation.

- 4) Fitness values are calculated for the new generation. Step 3 is repeated till convergence.
- 5) The best chromosome of the final generation is selected as the parameter for the ESVM.

For ESVM training, the nonlinear separating hyperplane with maximal margin in high dimensional space is automatically adapted using the kernel. In this manner, the solution space is refined and converges to the optimal/near-optimal solution. Training is stopped if the optimized maximal margin hyperplane is obtained for all input training vectors. ESVM parameters are optimized as shown in Fig.5

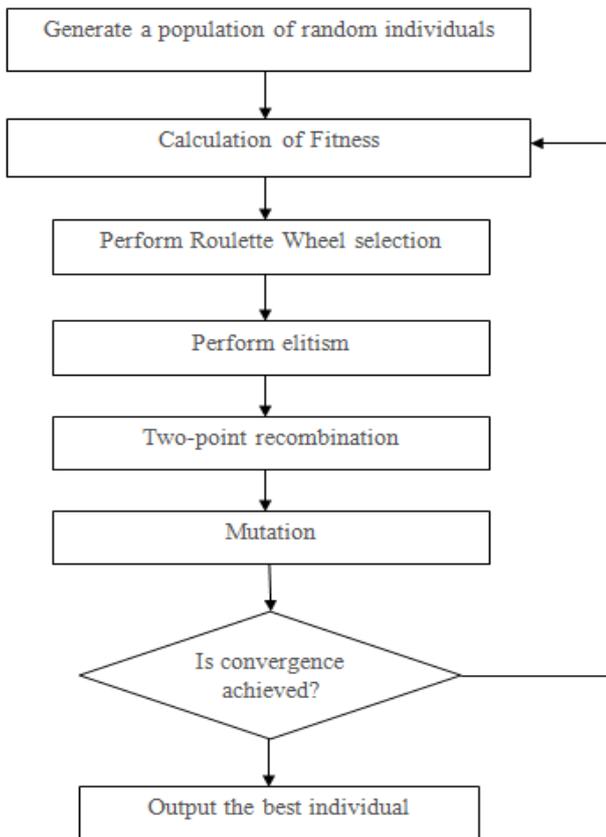


Fig-5: ESVM-Parameters optimization

(2) Nodule detection and classifier validation

After training to obtain the class predicted from the test data, the input features derived from nodule candidates is provided to the classifier. The trained classifier will predict a class by considering only the input feature vectors.

The ESVM finds the maximal margin hyperplane in the higher-dimensional space of the input feature vector in the training process. The nodules are then separated from non-nodules by the maximal margin hyperplane in the feature space. Hence the nodules are then separated from non-nodules.

The performance measures are given by,

$$TPR = \frac{TP}{TP+FN} \quad (10)$$

$$FPR = \frac{FP}{FP+TN} \quad (11)$$

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (12)$$

$$SPC = 1 - FPR \quad (13)$$

Where TP and FN are the number of nodules classified as True Positive and False Negative, respectively. SPC indicates the abbreviation of Specificity. The True Positive Rate (TPR) represents the number of correctly predicted positives divided by the total number of positive cases. The False Positive Rate (FPR) is the number of negative cases predicted as positive cases divided by the total number of negative cases. The accuracy is the proportion of true results in the population. Specificity (SPC) denotes the probability of a negative test given that the patient is well. Table 1 shows the overall performance of the nodule detection.

### III. RESULTS AND DISCUSSION

The above section presents the implementation specifics and a number of experimental results at each and every stage of the CAD system. Primarily the implementation details and results obtained during lung volume segmentation and nodule detection are discussed. Consequentially the implementation details and performance of the entire classification scheme is furnished.

#### A. Database and Imaging Protocol

The Lung Image Database Consortium [1,4] (LIDC) database is a publicly available database of thoracic CT scans that serves as a medical imaging research resource. The dataset comprises of 2114 slices, and the nodule diameter ranges from 3 mm to 30 mm. There were about 200 slices per scan, and each slice is 512 pixels × 512 pixels, with 4096 gray-level values in HU. The pixel size in the database ranged from 0.5 mm to 0.76 mm, and the reconstruction interval ranged from 1 mm to 3 mm. In the above dataset, four radiologists reviewed each scan and drew outlines for nodules 3.0 mm or larger in effective size. The ground truth was then established in a blind reading, which was followed by an unblinded reading sessions

#### B. Lung volume segmentation and nodule candidate detection

In the beginning, the lung volume is segmented by thresholding and 3-D connected component labelling. The results of each and every stage of the lung volume segmentation for distinct lung slices is shown in Fig. 6. The input CT images are shown in Fig.6(a). Initially low density regions are parted from high density regions based on the threshold values to derive the segmented lung volume. Fig. 6(b) shows the thresholded result of for every lung region. The thresholded outcome comprises of undesirable elements (air outside the body and gas in the intestine). In order to abstract lung region, 3-D connected component labelling to the thresholded image. The lung region extracted is shown in Fig.6 (c). The extracted lung regions have certain flaws such as holes and critical sections which is eradicated by means of contour correction. The holes are eliminated by hole filling

operation. Contour correction is done to embrace juxta-pleural nodules in the critical section. This modifies the critical section by joining pairs of critical points that are separated by a distance of less than 20mm. Contour corrected results for every input slice are shown in Fig. 6(d).

Sequentially the nodule candidate detection is executed on the segmented lung volume. The nodules are detected by enhancement of spherical objects by means of dot enhancement filtering. Totally five smoothing scales are used, 0.75, 1.33, 2.37, 4.21 and 7.5. The resulting image section of varying sizes 5,8,12,19, and 32 are obtained. The block is interpolated to isotropic voxel resolution of 1mm.

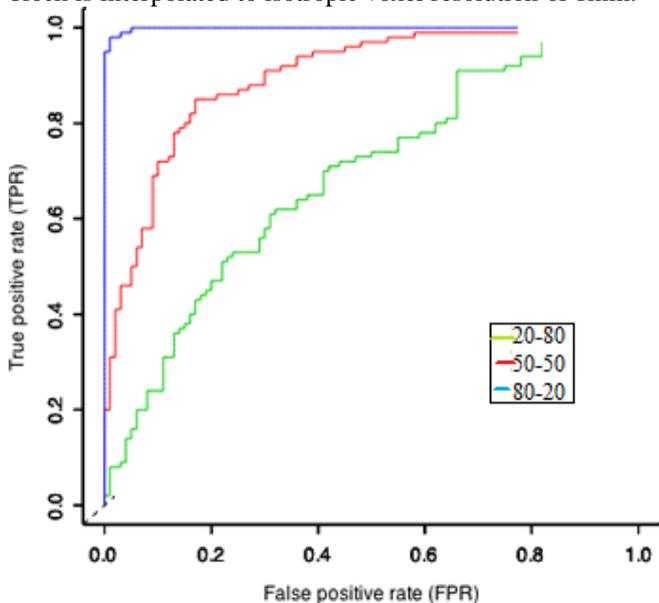


Fig-6: ROC curve of SVM

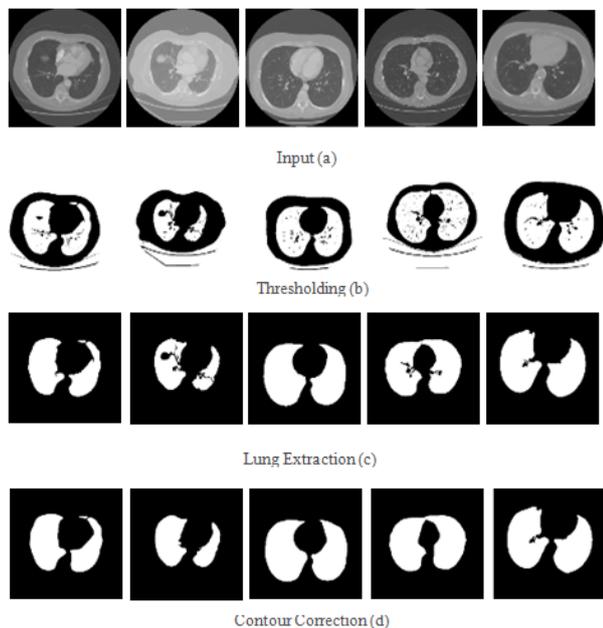


Fig-7: Results of lung volume segmentation

Hence the nodule candidates were detected with reduced number of false positives. A balanced dataset comprising of equal number of nodules and non-nodules is constructed.

C. Feature extraction and classification

The shape based feature descriptor is extracted from the image section, and walls are eliminated using feature refinement. AHSN features are extracted. The shape based

descriptor is more precise and constant as the Hessian matrix recapitulates the prevalent directions in a specified neighbourhood of a point. The dimension of the AHSN feature is not as much of conventional methods. The shape based feature gives better results after applying refinement technique.

The 3-D shape based feature descriptor is applied to the nodule detection scheme. The classification is performed by splitting the training and testing datasets of three different ratios. In ESVM the box constraint C and kernel parameter  $\gamma$  are obtained via genetic optimization. For the optimization the parametrical values for producing the new generation are set as:  $P_r=0.2$ ,  $P_c=0.6$  and  $p_m=0.2$ . Upon evolutionary computation the values of box constraint and kernel parameter values obtained in following intervals (3, 15) and  $\gamma$  (4,10). Fig-6 shows the receiver operating characteristics curve (ROC), indicating performance of the SVM with respect to three ratios of classification data. The inclusive performance of the nodule detection system is assessed for entire detected nodule candidates. Fig-8 shows the receiver operating characteristics curve (ROC), indicating performance of the ESVM with respect to three ratios of classification data. The above detection scheme reveals durable and valid performance in detecting nodules.

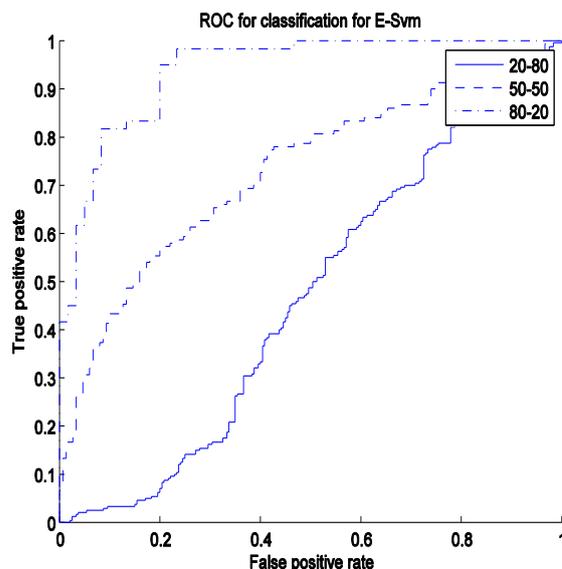


Fig-8: ROC curves of ESVM

Table-1: Overall performance results of the proposed CAD system

| Ratio | Accuracy(%) | Specificity(%) | Sensitivity(%) |
|-------|-------------|----------------|----------------|
| 20-80 | 51.4        | 21.4           | 78.7           |
| 50-50 | 61.3        | 38.6           | 84             |
| 80-20 | 87.5        | 76.6           | 98.3           |

System for different dimensions of AHSN features with ESVM classifier.

Table 1 denotes the performance of the nodule detection system on applying angular histograms of surface normal (AHSN) feature. The detection system achieves 11 FPs per scan, with 98.3% sensitivity.

Table 2 shows the performance comparison of CAD scheme using SVM and ESVM.

| parameter       | SVM  | ESVM |
|-----------------|------|------|
| Accuracy (%)    | 74.1 | 87.5 |
| Specificity (%) | 66.6 | 76.6 |
| Sensitivity (%) | 81.6 | 98.3 |

From these results, it is found the CAD system effectively reduces the number of false positive detections and maintains better sensitivity.

#### IV. CONCLUSION

In this paper, a computer-aided system for pulmonary nodule detection based on Evolutionary Support Vector Machine classifier is presented. The paper describes the outright design of the CAD system and illustrates a detailed performance analysis on publically available LIDC database. In order to detect the pulmonary nodules, the lung volume is segmented by thresholding and 3D-connected component labelling-based method. From the segmented lung volume, nodule candidates are detected by Eigen value decomposition of Hessian matrix and Multi-scale dot enhancement filtering. Next, the shape based feature descriptors were extracted from detected nodule candidates, and refined to eradicate walls. The refined feature descriptors were fed as input for ESVM to detect nodules. The detection system attains sensitivity of 98.3% at 11 false positives per scan.

#### REFERENCES

[1] S.G. Armato, G. McLennan, M.F. McNitt-Gray, C.R. Meyer, D. Yankelevitz, D.R. Aberle, C.I. Henschke, E.A. Hoffman, E.A. Kazerooni, H. MacMahon, A.P. Reeves, B.Y. Croft, L.P. Clarke, L.I.D.C.R. Group. Lung image database consortium: developing a resource for the medical imaging research community. *Radiology* 232; 2004: 739–748.

[2] M. Brown, M. McNitt-Gray, N. Mankovich, J. Goldin, J. Hiller, L. Wilson, D. Aberle, Method for segmenting chest CT image data using an anatomical model: preliminary results. *IEEE Transactions on Medical Imaging* 16; 1997: 828–839.

[3] J. Dehmeshki, X. Ye, X. Lin, M. Valdivieso, H. Amin. Automated detection of lung nodules in CT images using shape-based genetic algorithm. *Computerized Medical Imaging and Graphics* 31; 2007: 408–417.

[4] M.F. McNitt-Gray, S.G. Armato, C.R. Meyer, A.P. Reeves, G. McLennan, R.C. Pais, J. Freymann, M.S. Brown, R.M. Engelmann, P.H. Bland, G.E. Laderach, C. Piker, J. Guo, Z. Towfic, D.P.Y. Qing, D.F. Yankelevitz, D.R. Aberle, E.J.R. van Beek, H. MacMahon, E.A. Kazerooni, B.Y. Croft, L.P. Clarke. The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation. *Academic Radiology* 14; 2007: 1464–1474.

[5] T. Messay, R. Hardie, S. Rogers. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Medical Image Analysis* 14; 2010: 390–406.

[6] O. Osman, S. Ozekets, O.N. Ucan. Lung nodule diagnosis using 3D template matching. *Computers in Biology Medicine* 37; 2007: 1167–1172.

[7] D. Paik, C. Beaulieu, G. Rubin, B. Acar, R. Jeffrey, J. Yee, J. Dey, S. Napel. Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT. *IEEE Transactions on Medical Imaging* 23; 2004: 661–675.

[8] A. Retico, P. Delogu, M. Fantacci, I. Gori, A. Preite Martinez. Lung nodule detection in low-dose and thin-slice computed tomography. *Computers in Biology and Medicine* 38; 2008: 525–534.

[9] X. Ye, X. Lin, J. Dehmeshki, G. Slabaugh, G. Beddoe. Shape-based computer-aided detection of lung nodules in Thoracic CT images. *IEEE Transactions on Biomedical Engineering* 56; 2009: 1810–1820.

[10] Cancer Facts and Figs. *The American Cancer Society*. 2009.

[11] Wang Q., Kang W., Wu C. Wang B. Computer-aided detection of lung nodules by SVM based on 3D matrix patterns. *Clinical Imaging*; Doi:10.1016, 2012.

[12] T. Messay, R. Hardie, S. Rogers. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Medical Image Analysis* 14; 2010: 390–406.

[13] W.-J. Choi, T.-S. Choi. Genetic programming-based feature transform and classification for the automatic detection of pulmonary nodules on computed tomography images. *Information Sciences* 212; 2012: 57–78.

[14] J. Pu, D. Paik, X. Meng, J. Roos, G. Rubin. Shape “breaks-and-repair” strategy and its application to automated medical image segmentation. *IEEE Transactions on Visualization and Computer Graphics* 17; 2011: 115–124.

[15] D. Cascio, R. Magro, F. Fauci, M. Iacomi, G. Raso. Automatic detection of lung nodules in CT datasets based on stable 3D mass-pring models. *Computers in Biology and Medicine* 42; 2012: 1098–1109.

[16] J. Tangelder, R. Veltkamp. A survey of content based 3dshape retrieval methods. *Multimedia Tools and Applications* 39; 2008: 441–471.

[17] G. Medioni, M. Lee, C. Tang, A Computational Framework for Segmentation and Grouping. vol. 1. Elsevier Science. Amsterdam, the Netherland; 2000.

[18] B. Boser, I. Guyon, V. Vapnik. A training algorithm for optimal margin classifiers, in: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. ACM. Pittsburgh, Pennsylvania. USA; 1992: pp. 144–152.

[19] M. Scherer, M. Walter, T. Schreck. Histograms of oriented gradients for 3d object retrieval. in: *Proceedings of the WSCG, Citeseer, Plzen, Czech Republic*; 2010: pp. 41–48.

[20] T. Sun, J. Wang, X. Li, P. Lv, F. Liu, Y. Luo, Q. Gao, H. Zhu, X. Guo. Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. *Computer Methods and Programs in Biomedicine* 111; 2013: 519–524.

[21] Huang, H.-L., Chang, F.-L.: ESVM: evolutionary support vector machine for automatic feature selection and classification of microarray data. *Bio Systems* 90; 2007: 516–528.