Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/090124

Artificial Neural Network based Classification of Lungs Nodule using Hybrid Features from Computerized Tomographic Images

Sheeraz Akram[1,*], *Muhammad Younus Javed*[1], *Usman Qamar*[1], *Aasia Khanum*[2] *and Ali Hassan*[1]

[1] Department of Computer Engineering, College of Electrical and Mechanical Engineering, National University of Sciences and Technology, Islamabad, Pakistan
[2] Department of Computer Science, Forman Christian College, Lahore, Pakistan.

Received: 24 Mar. 2014, Revised: 24 Jun. 2014, Accepted: 25 Jun. 2014
Published online: 1 Jan. 2015

Abstract: An automated pulmonary nodule detection system is necessary to help radiologist to identify and detect the nodules at early stage. In this paper, a novel pulmonary nodule detection system is proposed using Artificial Neural Networks (ANN) based on hybrid features consist of 2D and 3D Geometric and Intensity based statistical features. The lung volume is segmented using thresholding, 3D connected component labeling, contour correction and morphological operators. The candidate nodules are extracted and pruned based on the rules that are built using characteristics of nodules. The 2D and 3D Geometric features and Intensity Based Statistical features are extracted and used to train a Neural Network. The proposed Computer-Aided Diagnostic (CAD) system is tested and validated using standard dataset of Lung Image Consortium Database (LIDC). The results obtained from proposed CAD system are good as compared to existing CAD systems. The sensitivity of 96.95% is achieved with accuracy of 96.68%.

Keywords: Computer-Aided Diagnostic (CAD) System, Artificial Neural Networks (ANN), Computerized Tomographic Images, Pulmonary Nodules Detection

1 Introduction

The human body suffers from different diseases. The cancer is dangerous disease for human life. The generic types of cancer in human body are Bladder, Breast, Colon and Rectal, Endometrial, Kidney, Leukemia, Lung, Melanoma, Non-Hodgkin Lymphoma, Pancreatic, Prostate and Thyroid cancer. The more number of people is suffering and died from lungs cancer than any other cancer [1, 2]. The survival rate of lungs cancer patient is only 14% but it could be increased up to 50% if there is an early detection of lungs cancer [2]. The survival rate is significantly improved but there is need to increase this survival rate more than the current value. This should be done without opening the patient body. The task is performed after having inner view of the human body. The multiple methods are used to take the images from inside the body like X-rays, CT scans, MRI etc. The CT scan is most recommended method which produces the 3D images of the lungs.

The obtained images are of not good quality. There is need of medical expert to give an opinion on the images obtained through the CT scans. The medical experts with same expertise are not available at every place. There is need of certain guidance for such medical experts. Even if medical experts are available, there are chances of human error due to resemblance of tissues, veins and small nodules presenting the image at the initial stage. To achieve this goal, the field of medical imaging introduced CAD (Computer-Aided Diagnostic) systems which help medical specialist to identify and categories the problem.
The lesions are produced with different body parts which cause the cancer. Such lesions are referred to as nodule if they causes cancer, otherwise non-nodule. In the design of a CAD system, the main task is to segment the volume of particular body part, like lungs volume should be separated from the complete image so that we can keep our focus on the object of interests. The next task is to separate the objects in lungs volume which are not part of lungs. These objects are unwanted lesions. These

* Corresponding author e-mail: sheerazakram@ceme.nust.edu.pk

unwanted lesions are potential nodules. The next step is to classify the potential nodules into nodules and non-nodule.

In proposed methodology, the 3D lung CT image is Thresholded; the background is removed from Thresholded image. The lung lobes contain holes those are filled using morphological operations. The contour correction is performed to include the juxta-pleural nodule. The candidate nodules are extracted using different levels of thresholding. The candidate nodules are pruned using rule based pruning method. The hybrid features are extracted from pruned candidate nodules and the feature vectors are formed using different features. The candidate nodules are Up-sampled to reduce the biasness. The Artificial Neural Network (ANN) classifier is trained using candidate nodules. The candidate nodules are tested and verified to classify as nodule and non-nodule using trained ANN classifier. The detail is given is later sections.

In this paper, the literature review is presented in next section. Section III describes the proposed method which contains preprocessing phase to extract the lung region, candidate nodule extraction, candidate nodule pruning, Hybrid Feature Extractions, Candidate Nodule Up-sampling, and Classification using Artificial Neural Networks (ANN). In Section IV, Results of proposed methodology are discussed in Result and Discussion section and the Conclusion section concludes all the work performed for proposed CAD system.

2 Literature Review

The lesions are detected automatically by scanning of the lungs. There are various methods to segment lung region, extract lesions and to classify these lesions as nodule and non-nodule. Ozekes et al. introduces nodule detection by calculating density value of each pixel, then rule-based lung region segmentation is performed, the Regions of Interest (ROIs) are extracted using 8-directional search. Subsequently preliminary classification is performed using Location Change Measurement and later nodules are searched using trained Genetic Algorithm from the images of ROIs in [3].

Ozekes et al. introduces the lung segmentation using Genetic Cellular Neural Network; the ROIs are extracted using 8-directional search. The nodules are detected by searching through 3D image with 3D template using convolution based filter. The Fuzzy Rule Based Thresholding is used to further refine the detected nodules in [4]. The Ye et al. introduces the 3D nodules extraction using anti-geometric diffusion, volumetric shape features, Gaussian filtering and multi-scale dot enhancement filtering. The 3D potential nodules are segmented. The 2D and 3D features are calculated from segmented nodules; the Rule-Based filtering and weighted SVM are used for Nodule classification in [5].

Retico et al. introduces the identification of the pleural region by Directional-gradient concentration (DGC) and morphological opening. The ROIs are extracted from segmented pleura region. The features are extracted and candidate nodules are classified using Feed-forward Neural Network in [6].

Sousa et al. introduces the identification of the lung parenchyma using region growing algorithm. The rolling-ball methodology is used to correct the boundaries of pleura. The region growing is used again to identify the lung nodule. The SVM is used to reduce the false positives in [7].

Lee et al. introduces the ensemble classification aided by clustering; the training dataset is classified using clustering and the nodule and non-nodules obtained from clustering are used for training of SVM in [8]. Maeda et al. introduces the usage of temporal subtraction of consecutive CT images to detect candidate nodules; the features of candidate nodules are calculated and the candidates nodules are refined using rule based feature analysis. The feature space is reduced using PCA and Artificial Neural network is used for nodule classification in [9]. Tan et al. introduces the isotropic resampling of CT image to change the resolution of image. The lung region is segmented and the nodule center is estimated using divergence of normalized gradient. The multi-scale nodule and vessel enhancement filtering is used to segment nodule clusters. The invariant, shape and regional descriptor are calculated. The mix of ANN, GA (FD-NEAT) and SVM is used for feature selection and nodule classification in [10]. Choi et al introduces the lung volume extraction using thresholding, contouring correction and morphological operation. The candidate nodules are extracted by multiple thresholding from lung volume. The extracted candidate nodules are refined. The features are extracted from candidate nodules. The Genetic Programming (GP) classifier is trained and used for the classification of nodules and non-nodules in [11]. Choi et al. introduces the hierarchical block classification approach using SVM for nodules classification. The CT image is split into blocks and the non-informative blocks are discarded. The block image is enhances and object is segmented in block image. The location of block is adjusted. The features are extracted from nodule candidate block images. The SVM is used to classify candidate nodules as nodules and non-nodules [12].

3 Proposed Method

In this paper, the methodology is proposed in which the preprocessing is performed on lung CT image to segment lung volume. The preprocessing step includes thresholding, background removal, hole-filling and contour correction of lung region. The candidate nodules are extracted using multiple levels of thresholding and candidate nodules are pruned to reduce the false positives. The features are extracted from candidate nodules and required features are selected to form feature vectors. The

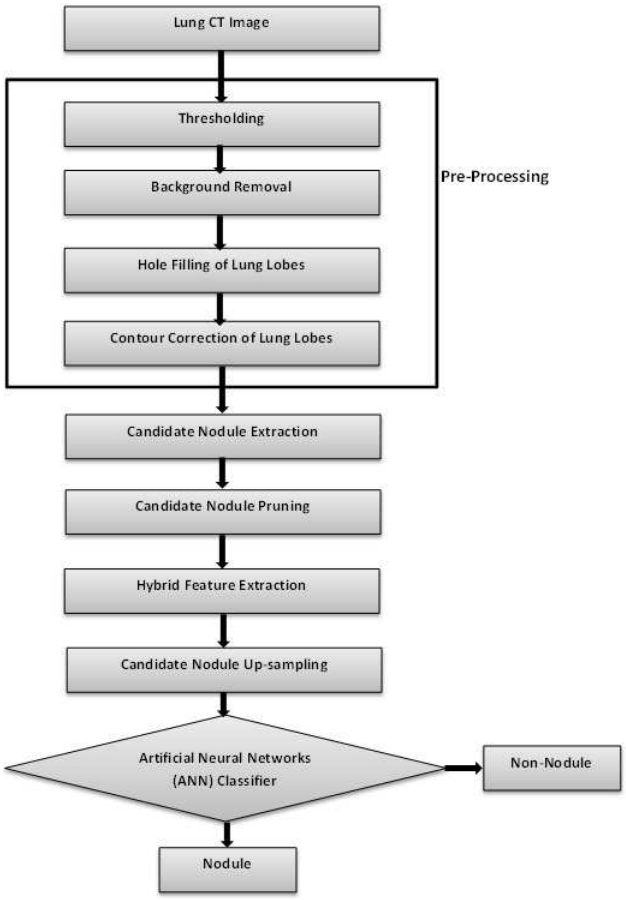


Fig. 1: Flow Chart of Proposed Pulmonary Nodule Detection and Classification System

number of candidate nodules is increased by up-sampling and ANN classifier is used for the classification of nodule and non-nodule. The block diagram of proposed methodology is given in Fig. 1. The detail of each step is given in later sections.

3.1 Preprocessing

The proposed methodology is perform and evaluated on Lung Image Database Consortium Database (LIDC) [13]. The lung CT image contains values in HU (Hounsfield units). The Lung images are 3D images. Each image contains slices ranges from 100 to 250. The size of each slice is 512 by 512. The following are steps for preprocessing of Lung CT Image:

3.1.1 Thresholding

The HU values in each Lung CT image is ranges from -2000 HU to +2000 HU. The lung area is low density area ranging from -1000HU to -450HU, can be called as non-body area. The CT scanner area is also part of non-body area of Lung CT Image. The body area contains the surrounding of lung lobes. The lungs are in non-body area, so we threshold it at -500 HU [14–16]. The voxels value below -500HU contains lungs and voxels values above -500HU contains body area as in Equation 1. The Fig. 2(b) shows the result of Thresholding.

$$\text{voxel value} := \begin{cases} \text{Min HU value in lung CT} & \text{voxel value} < -500HU \\ \text{Max HU value in lung CT} & \text{voxel value} \geq -500HU \end{cases} \quad (1)$$

3.1.2 Background Removal

The Thresholded Lung CT image contains body and non-body area. The black area is body voxels and white area is non-body voxels. The non-body area contains Lung lobes and CT scanner making a cylinder around the lungs and body area. There is need to remove this cylinder. The 3D connected component approach is applied [17, 18]. The non-body component touching the sides of Lung CT image is removed and voxels values are set to background values. i.e. the value of the body voxels. The Fig. 2(c) shows the result of Background Removal.

3.1.3 Hole Filling of Lung Lobes

The background removed image contains holes in Lung lobe; those are either nodules or vessels. It is important to include them in the Lung lobe region. The holes are filled using the hole filling morphological operators [18,19]. The Fig. 2(d) shows the result of Hole Filled image.

3.1.4 Contour Correction of Lung Image

The hole-filled Lung image may contains nodules or vessels at the border of lung lobe known as juxta-pleural. These juxta-pleural needs to be included in Lung area. Initially the lung contour is obtained and contour is corrected using chain codes [20]. The critical points are detected using Differential Chain coding. Later the required critical points are connected through straight lines in order to include the critical part at contour of lung lobe. The Fig. 2(e) shows the result of contour corrected image.

3.2 Candidate Nodule Extraction

The 3D lung mask is obtained from preprocessing step. This lung mask is used to extract the lung volume from Lung CT. The lung volume contains both vessels and nodules. The density of nodules, vessels and lungs is

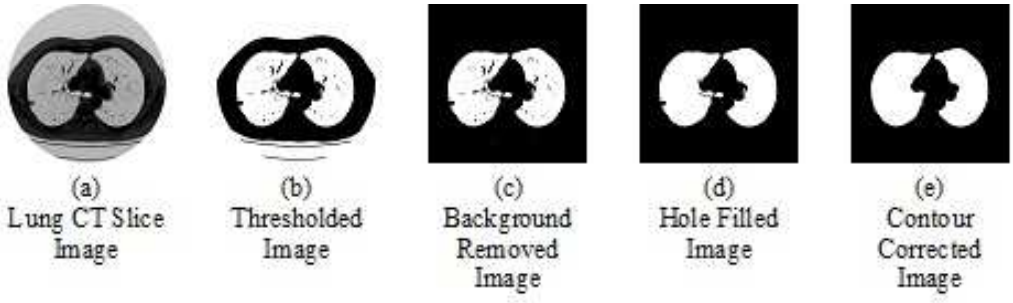


Fig. 2: Preprocessing for Lung Volume Segmentation

different from each other. In other words, the nodules and vessels are denser than lungs. The lung volume is extracted from Lung CT image using Lung mask. The optimal thresholding method is used to extract Regions of Interest (ROI). The Threshold is calculated on median slice. Multiple threshold values are calculated as the vessels and nodules have different density depending on the type of potential nodule.

3.3 Candidate Nodule Pruning

The extracted ROIs contain both nodules and vessels. The nodules in the dataset have diameter ranges from 3 mm to 30 mm. The ROIs having diameter smaller than 3 mm are ignored as noise and the ROIs having diameter greater than 30 mm are pruned as lesion/vessels. The property of elongation is used to detect the vessels in ROIs. The pruned nodules present in 3D lung image are mapped on to 2D image, shown in Fig. 3.

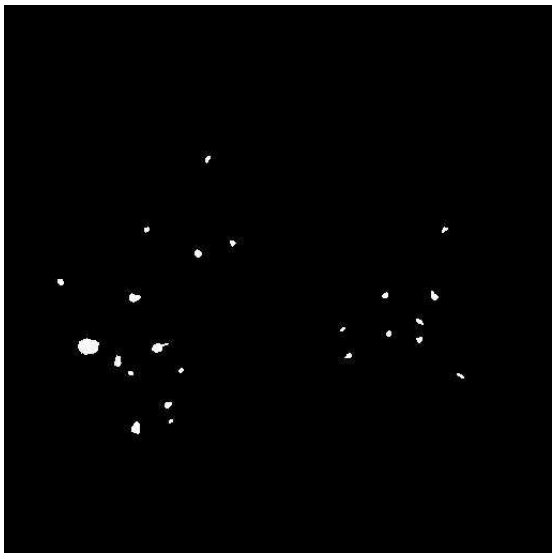


Fig. 3: Extracted Candidate Nodules

3.4 Hybrid Features Extraction

The candidate nodules extracted in previous step contains both nodules and non-nodules. These candidate nodules are 3D object. The Table 1 contains the features those are extracted from candidate nodules. The 2D Geometric features and Intensity based statistical features are extracted from the median slice of the segmented object. The 3D geometric features and intensity based statistical features are extracted from 3D segmented object.
The mathematical expressions to calculate geometric features are given in equations 2 to 9

$$Area(A) = \text{Number of Voxels in Median Slice of Segmented Object} \quad (2)$$

$$Diameter(D) = \text{Maximum Length of Bounding Box in Median Slice} \quad (3)$$

$$Radius(r) = \frac{D}{2} \quad (4)$$

$$Perimeter = \text{The number of voxels on the boundry of Segmented Object in Median Slice} \quad (5)$$

$$Circularity = \frac{A}{4\pi r^2} \quad (6)$$

$$Volume(V) = \text{Number of Voxels in Segmented Object} \quad (7)$$

$$Compactness = \frac{A}{\frac{4\pi r^3}{3}} \quad (8)$$

$$\text{Bounding Box Dimensions} = \text{The Dimensions of Smallest 3D Box Containing the Segmented Object} \quad (9)$$

Table 1: List of Features Extracted

2D Geometric Features	3D Geometric Features	2D Intensity Based Statistical Features	3D Intensity Based Statistical Features
-Area	-Volume	-Minimum Value Inside	-Minimum Value Inside
-Diameter	-Compactness	-Mean Inside	-Mean Inside
-Perimeter	-Bounding Box Dimension (3)	-Mean Outside	-Mean Outside
-Circularity	-Principal Axis Lengths (3)	-Variance Inside	-Variance Inside
	-Elongation	-Skewness Inside	-Skewness Inside
		-Kurtosis Inside	-Kurtosis Inside
		-Eigen Values (8)	

$$\text{Principal Axis Lengths} = \text{The Length of Principal Axis for 3D Segmented Object} \tag{10}$$

$$Elongation = \frac{\text{Maximum Principal Axis}}{\text{Minimum Principal Axis}} \tag{11}$$

The mathematical expressions to calculate intensity based statistical features are given in equations 12 to 15.

$$Mean(\bar{X}) = \frac{\sum_{i=1}^{n} x_i}{n} \tag{12}$$

$$Variance(s^2) = \frac{\sum_{i=1}^{n} (x_i - \bar{x^2})^2}{n-1} \tag{13}$$

$$Kurtosis = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^4}{(n-1)s^4} \tag{14}$$

$$Skewness = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{(n-1)^3} \tag{15}$$

3.5 Candidate Nodule Up-sampling

The Lung Image Database Consortium (LIDC) is a standard dataset which is used to develop and test Lung CAD systems. It consists of 84 CT Scans of different patients. The 47 Lung CT scans contain nodules; are used for experiment in this paper. The total 836 candidate nodules are obtained after pruning. The ground truth shows that there are 50 nodules and 786 non-nodules. This creates an un-balanced dataset for classification. We selected appropriate number of non-nodules dataset and then nodules are up-sampled by repeating the nodules candidate in order to make the dataset balanced. i.e. Number of nodules and non-nodules are equal.

3.6 Classification Using Artificial Neural Networks (ANN)

The Artificial Neural Networks (ANN) Feed-forward back propagation network is good for binary classification. The value of epochs and goal is set to 0.50 and 0.01 respectively. The 50% data is used for training of the classifier, 25% data is used for each testing and validation purpose. There is no biasness in the data. The candidate nodules are classified into nodules and non-nodules. The data set has features 2D Geometric Features, 3D Geometric Features, 2D Intensity Based Statistical Features, 3D Intensity Based Statistical Features. The features are divided into following sets:

–2D Geometric Features
–3D Geometric Features
–2D Intensity Based Statistical Features
–3D Intensity Based Statistical Features
–2D Geometric and Intensity Based Statistical Features
–3D Geometric and Intensity Based Statistical Features
–2D and 3D Geometric Features
–2D and 3D Intensity Based Statistical Features
–2D and 3D Geometric and Intensity Based Statistical Features

The dataset is divided into the following for training, testing and validation purpose:

–30-35-35, The 30% is used for Training, 35% is used for each Testing and Validation
–50-25-25, The 50% is used for Training, 25% is used for each Testing and Validation
–70-15-15, The 70% is used for Training, 15% is used for each Testing and Validation

All the candidate nodules are categorized into True Positive, True Negative, False Positive, and False Negative. The accuracy, sensitivity, specificity and AUC (Area under ROC Curve) is measured as in equation (16 - 18).

$$Accuracy = \frac{(TN+TP)}{(TN+TP+FN+FP)} \tag{16}$$

$$Sensitivity = \frac{TP}{(TP+FN)} \tag{17}$$

$$Specificity = \frac{TN}{(TN+FP)} \tag{18}$$

4 Results and Discussion

The standard dataset LIDC is used for training and validation purpose. The Lung CT scan is 3D image. Each

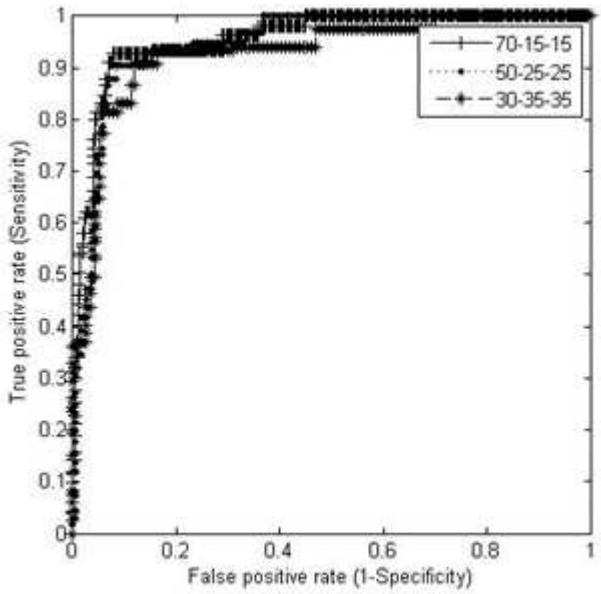


Fig. 4: ROC Curves for 2D Geometric Features

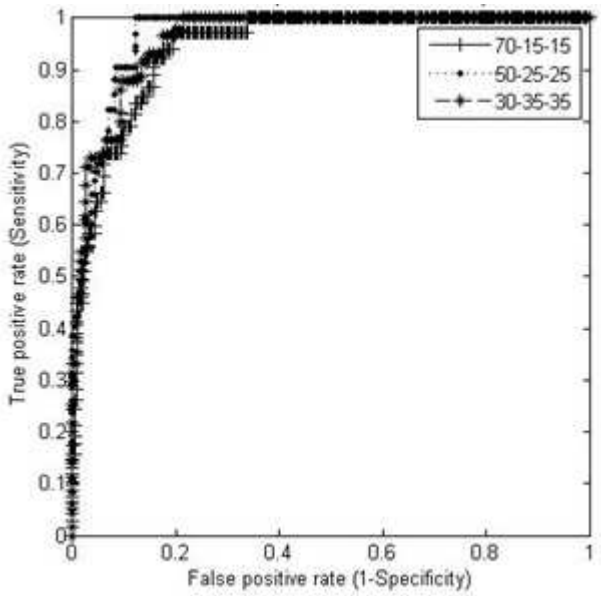


Fig. 5: ROC Curve for 3D Geometric Features

slice is of size 512 by 512. Each CT scan has around 150 slices and 4096 gray level values in HU. The pixel size in the database ranges from 0.5 mm to 0.76 mm, and the reconstruction interval ranges from 1 to 3 mm. This dataset is publicly available. The data is up-sampled to minimize the biasness of classifier. The results of Artificial Neural Network (ANN) for 2D Geometric Features, 3D Geometric Features, 2D Intensity Based Statistical Features, 3D Intensity Based Statistical Features, 2D Features (Geometric and Intensity Based Statistical), 3D Features (Geometric and Intensity Based Statistical), Geometric Features (2D and 3D), Intensity Based Statistical Features (2D and 3D), and 2D and 3D Features (Geometric and Intensity Based Statistical) are shown in the tables. The ROC curves are also drawn for each set of features.

The Table 2 shows the result of 2D Geometric features. With 50-25-25 training, testing and validation ratio, the 92.37% accuracy, 94.47% sensitivity and 90.21% specificity is achieved. The Fig. 4 shows the Receiver Operating Curve (ROC Curve) for 2D Geometric features for 30-35-35, 50-25-25 and 70-15-15 training, testing and validation ratio. The AUC is highest for 70-15-15 training, testing and validation ratio. The Table 3 shows the result of 3D Geometric features. With 50-25-25 training, testing and validation ratio, the 87.15% accuracy, 80.56% sensitivity and 92.82% specificity is achieved. The Fig. 5 shows the Receiver Operating Curve (ROC Curve) for 3D Geometric features for 30-35-35, 50-25-25 and 70-15-15 training, testing and validation ratio. The AUC is highest for 70-15-15 training, testing

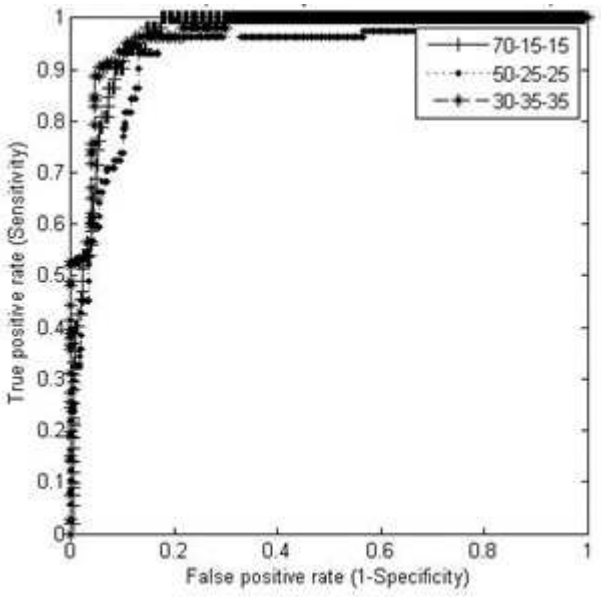


Fig. 6: ROC Curves for 2D Intensity Based Statistical Features

and validation ratio. The Table 4 shows the result of 2D Intensity based statistical features. With 50-25-25 training, testing and validation ratio, the 88.24% accuracy, 86.53% sensitivity and 89.90% specificity is achieved. The Fig. 6 shows the Receiver Operating Curve (ROC Curve) for 2D Intensity based statistical features for 30-35-35, 50-25-25 and 70-15-15 training, testing and

Table 2: Results of 2D Geometric Features

Training-Testing-Validation (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
30-35-35	91.07	90.91	91.24	0.96
50-25-25	92.37	94.47	90.21	0.96
70-15-15	93.22	95.73	90.76	0.97

Table 3: Results of 3D Geometric Features

Training-Testing-Validation (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
30-35-35	89.56	87.8	91.1	0.96
50-25-25	87.15	80.56	92.82	0.92
70-15-15	92.11	86.21	98.21	0.96

Table 4: Results of 2D Intensity Based Statistical Features

Training-Testing-Validation (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
30-35-35	86.97	81.82	92.22	0.92
50-25-25	88.24	86.53	89.90	0.94
70-15-15	89.79	86.61	93.52	0.96

Table 5: Results of 3D Intensity Based Statistical Features

Training-Testing-Validation (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
30-35-35	78.32	79.07	77.66	0.87
50-25-25	86.92	75	98.97	0.93
70-15-15	81.36	80.91	81.75	0.92

Table 6: Results of 2D Geometric & Intensity Based Statistical Features

Training-Testing-Validation (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
30-35-35	93.91	90.35	97.17	0.97
50-25-25	94.12	94.02	94.2	0.98
70-15-15	96.61	94.59	98.4	0.99

Table 7: Results of 3D Geometric & Intensity Based Statistical Features

Training-Testing-Validation (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
30-35-35	91.07	91.3	90.84	0.95
50-25-25	93.13	90.32	95.65	0.97
70-15-15	94.3	88.79	98.65	0.96

Table 8: Results of 2D & 3D Geometric Features

Training-Testing-Validation (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
30-35-35	90.33	88.36	92.31	0.96
50-25-25	94.15	91.41	96.92	0.97
70-15-15	93.16	87.96	97.62	0.96

Table 9: Results of 2D & 3D Intensity Based Statistical Features

Training-Testing-Validation (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
30-35-35	89.85	85.27	94.01	0.95
50-25-25	94.41	88.71	95.78	0.96
70-15-15	91.42	81.82	97.13	0.94

Table 10: Results of 2D & 3D Geometric & Intensity Based Statistical Features

Training-Testing-Validation (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
30-35-35	96.33	92.57	98.37	0.98
50-25-25	96.68	96.95	96.39	0.99
70-15-15	98.3	96.55	98.21	0.99

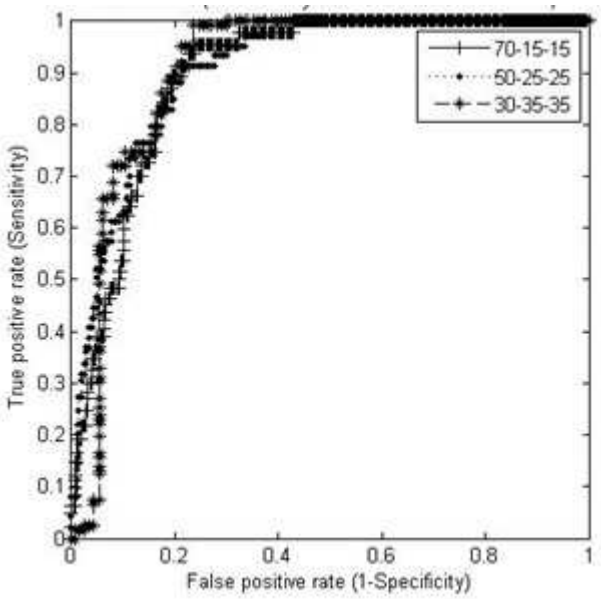


Fig. 7: ROC Curves for 3D Intensity Based Statistical Features

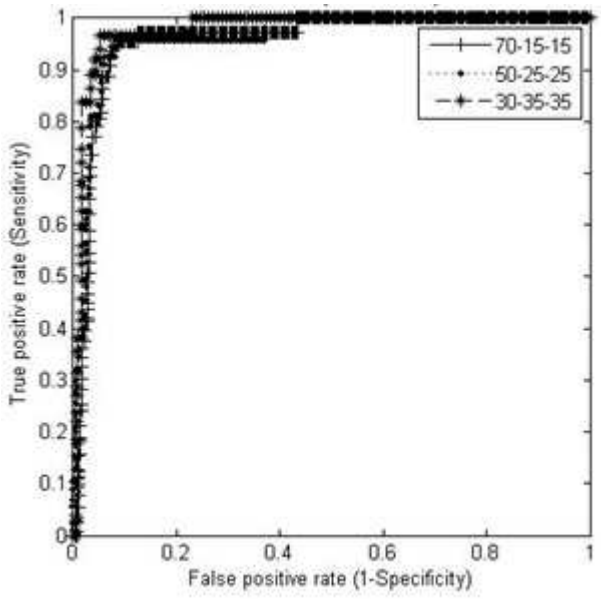


Fig. 9: ROC Curves for 3D Geometric and Intensity Based Statistical Features

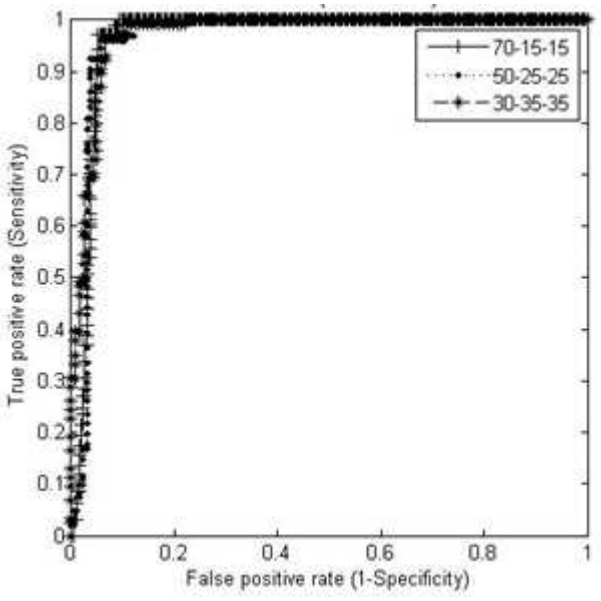


Fig. 8: ROC Curves for 2D Geometric and Intensity Based Statistical Features

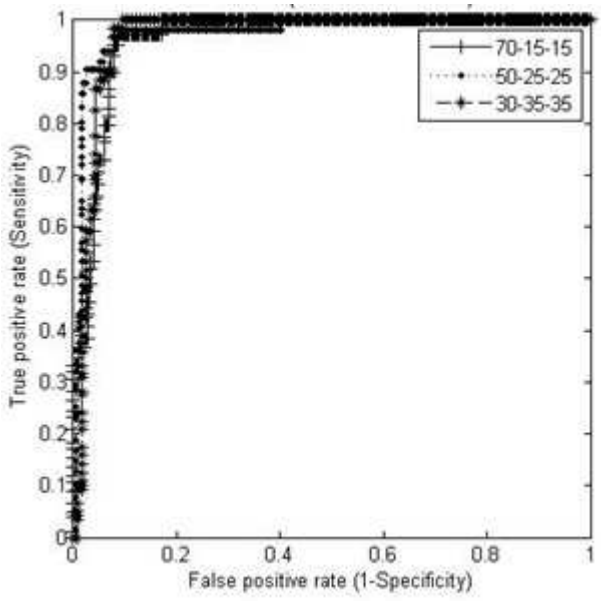


Fig. 10: ROC Curves for 2D and 3D Geometric Features

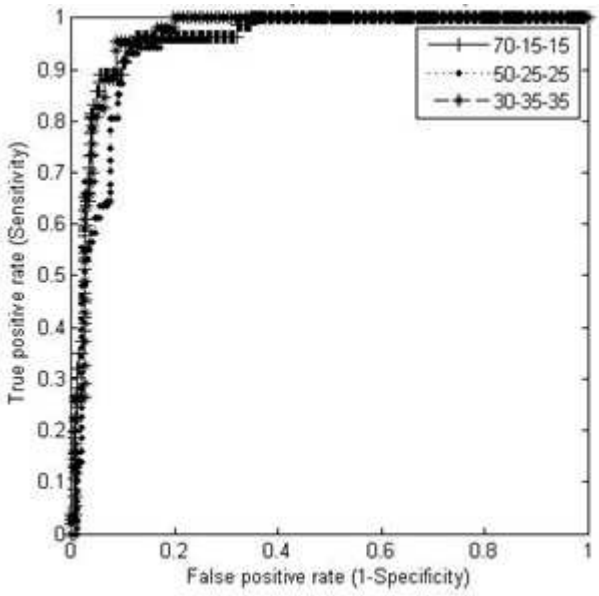


Fig. 11: ROC Curves for 2D and 3D Intensity Based Statistical Features

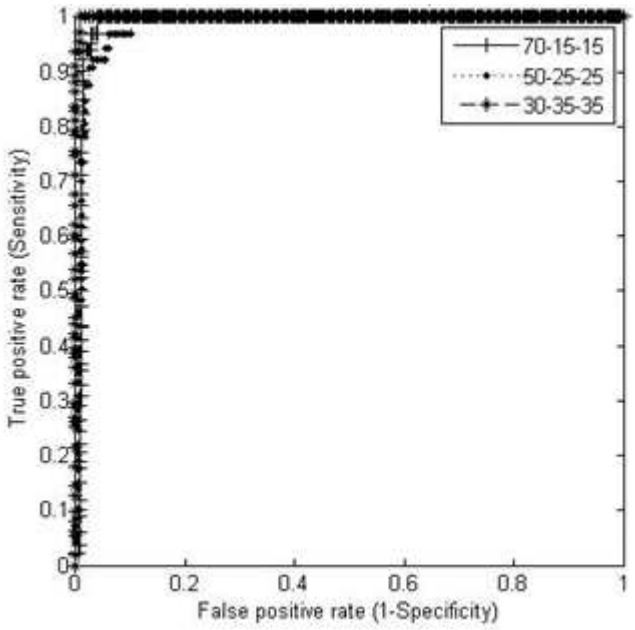


Fig. 12: ROC Curves for 2D and 3D Geometric and Intensity Based Statistical Features

validation ratio. The AUC is highest for 70-15-15 training, testing and validation ratio. The Table 5 shows the result of 3D Intensity based statistical features. With 50-25-25 training, testing and validation ratio, the 86.92% accuracy, 75% sensitivity and 98.97% specificity is achieved. The Fig. 7 shows the Receiver Operating Curve (ROC Curve) for 3D Intensity based statistical features for 30-35-35, 50-25-25 and 70-15-15 training, testing and validation ratio. The AUC is highest for 50-25-25 training, testing and validation ratio. The Table 6 shows the result of 2D Geometric and Intensity based statistical features. With 50-25-25 training, testing and validation ratio, the 94.12% accuracy, 94.02% sensitivity and 94.20% specificity is achieved. The Fig. 8 shows the Receiver Operating Curve (ROC Curve) for 2D Geometric and Intensity based statistical features for 30-35-35, 50-25-25 and 70-15-15 training, testing and validation ratio. The AUC is highest for 70-15-15 training, testing and validation ratio. The Table 7 shows the result of 3D Geometric and Intensity based statistical features. With 50-25-25 training, testing and validation ratio, the 93.13% accuracy, 90.32% sensitivity and 95.65% specificity is achieved. The Fig. 9 shows the Receiver Operating Curve (ROC Curve) for 3D Geometric and Intensity based statistical features for 30-35-35, 50-25-25 and 70-15-15 training, testing and validation ratio. The AUC is highest for 50-25-25 training, testing and validation ratio. The Table 8 shows the result of 2D and 3D Geometric features. With 50-25-25 training, testing and validation ratio, the 94.15% accuracy, 91.41% sensitivity and 96.92% specificity is achieved. The Fig. 10 shows the Receiver Operating Curve (ROC Curve) for 2D and 3D Geometric features for 30-35-35, 50-25-25 and 70-15-15 training, testing and validation ratio. The AUC is highest for 50-25-25 training, testing and validation ratio. The Table 9 shows the result of 2D and 3D Intensity based statistical features. With 50-25-25 training, testing and validation ratio, the 94.41% accuracy, 88.71% sensitivity and 95.78% specificity is achieved. The Fig. 11 shows the Receiver Operating Curve (ROC Curve) for 2D and 3D Intensity based statistical features for 30-35-35, 50-25-25 and 70-15-15 training, testing and validation ratio. The AUC is highest for 50-25-25 training, testing and validation ratio. The Table 10 shows the results of 2D and 3D Geometric and Intensity based statistical features. With 50-25-25 training, testing and validation ratio, the 96.68% accuracy, 96.95% sensitivity and 96.39% specificity is achieved. The Fig. 12 shows the Receiver Operating Curve (ROC Curve) for 2D and 3D Geometric and Intensity based statistical features for 30-35-35, 50-25-25 and 70-15-15 training, testing and validation ratio. The AUC is highest for 50-25-25 training, testing and validation ratio.

The Fig. [13-20] shows the scatter graphs of 2D Geometric Features, 3D Geometric Features, 2D Intensity based statistical features and 3D intensity based statistical features.

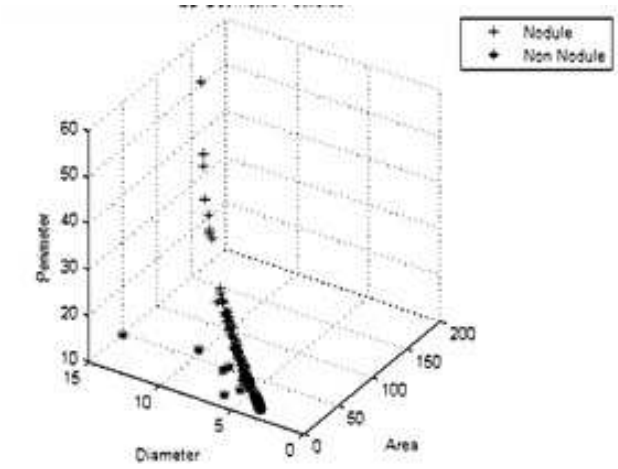


Fig. 13: Feature Space of 2D Geometric Features (Diameter, Perimeter, Area)

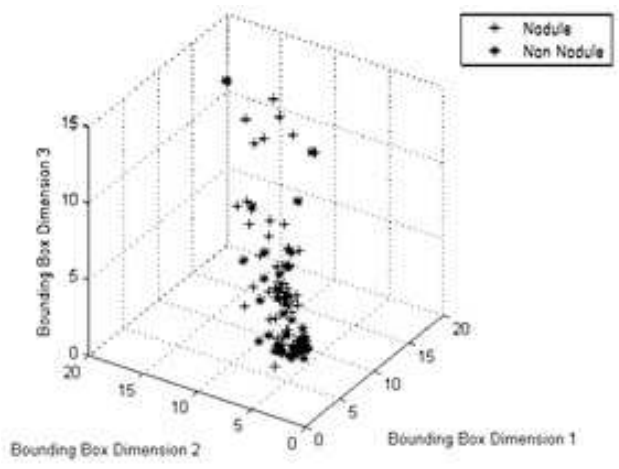


Fig. 15: Feature Space of 3D Geometric Features (Bounding Box Dimensions (3))

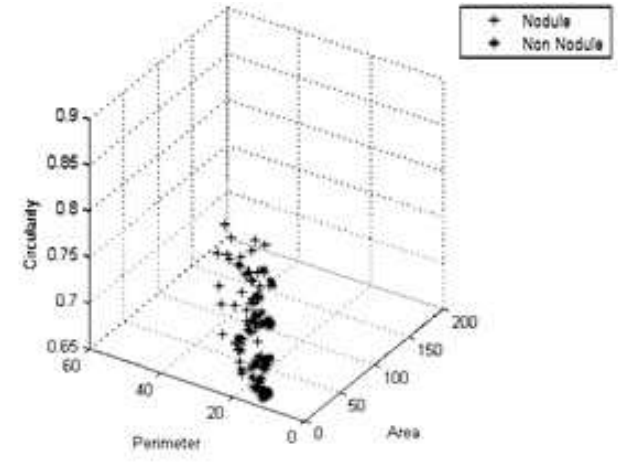


Fig. 14: Feature Space of 2D Geometric Features (Circularity, Perimeter, Area)

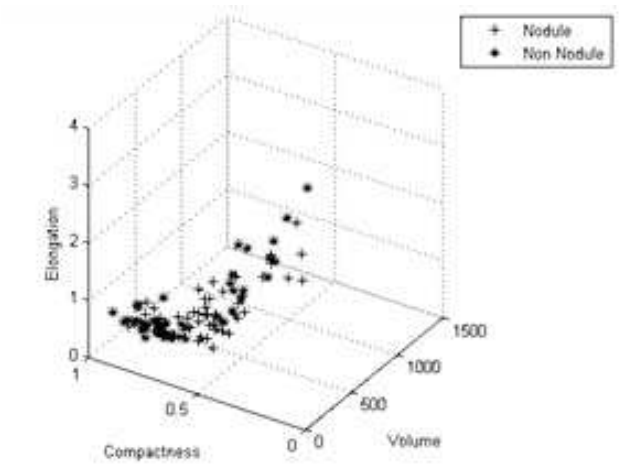


Fig. 16: Feature Space of 3D Geometric Features (Elongation, Compactness, Volume)

The Suzuki et al. (2003) [21] worked for nodules of size 8-20 mm with sensitivity 80.3%. The Rubin et al. (2005) [22] worked for nodule size ≥ 3 with sensitivity 76%. The Dehmeshki et al. (2007) [23] worked for nodule size 3-20 mm with sensitivity 90%. The SuarezCuenca et al. (2009) [18] worked for nodule sizes 4-27 mm with sensitivity 80%. The Opferand Wiemeker (2007) [24] worked for nodule size ≥ 4 mm with sensitivity 74%. The Sahiner at al. (2007) [25] work for nodule size 3-36.4 mm with sensitivity 79%. The Messay et al. (2010) [17] work for nodule size 3-30 mm with sensitivity 82.66%. The Choi et al. (2012) [11] work for nodule size 3-30 mm with sensitivity 94.1%. The Choi et al. (2013) [12] work for nodule size 3-30 mm with sensitivity 95.28%. The Ehmet et al. (2013) [26] works for nodule 2-20 mm with sensitivity of 89.60%. The Proposed Method work for nodule size 3-30 mm with sensitivity 96.95%, that is better than the earlier techniques. The performance comparison of earlier CAD systems and proposed CAD system is represented in Table 11.

5 Conclusion

In this paper, we have proposed a novel techniques for pulmonary nodule detection from Lung CT scan. The Lung image is Thresholded, background removed, 3D

Table 11: Performance Comparison of Different CAD Systems

CAD Systems	**Nodule Size**	**Sensitivity**
Suzuki et al. (2003) [21]	8-20 mm	80.3%
Rubin et al. (2005) [22]	$\geq$ 3 mm	76%
Dehmeshki et al.(2007) [23]	3-20 mm	90%
SuarezCuenca et al.(2009) [18]	4-27 mm	80%
Opferand Wiemeker (2007) [24]	$\geq$ 4 mm	74%
Sahiner at al. (2007) [25]	3-36.4 mm	79%
Messay et al.(2010) [17]	3-30 mm	82.66
Choi et al.(2012) [11]	3-30 mm	94.10%
Choi et al.(2013) [12]	3-30 mm	95.28%
Ahmet et al. (2013) [26]	2-20 mm	89.60%
Proposed Method	3-30 mm	96.95%

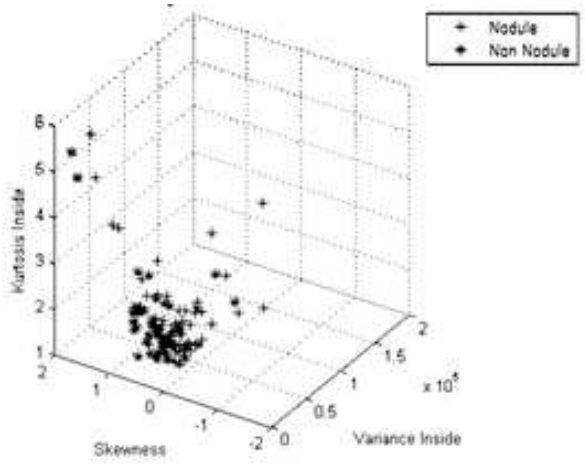


Fig. 17: Feature Space of 2D Intensity Based Statistical Features(Kurtosis Inside, Skewness, Variance Inside)

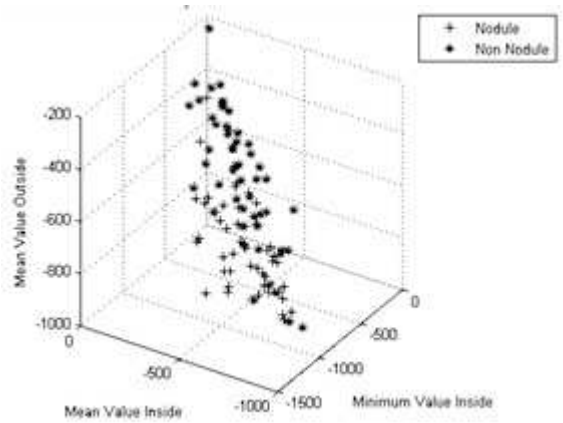


Fig. 18: Feature Space of 2D Intensity Based Statistical Features (Mean Value Outside, Mean Value Inside, Minimum Value Inside)

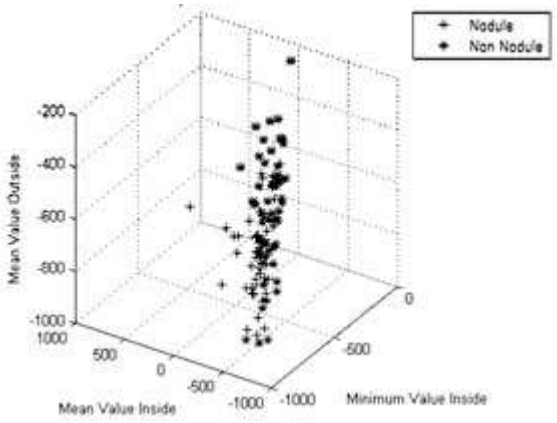


Fig. 19: Feature Space of 3D Intensity Based Statistical Features(Mean Value Outside, Mean Value Inside, Minimum Value Inside)

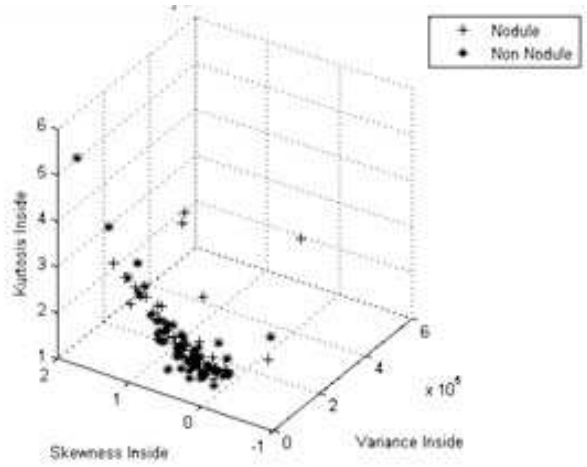


Fig. 20: Feature Space of 3D Intensity Based Statistical Features (Kurtosis Inside, Skewness Inside, Variance Inside)

connected component labeling, morphological operators, contour correction is performed to extract lung volume. The nodules are extracted using multi-level thresholding. The candidate nodules are refined. The 2D and 3D geometric and intensity based statistical feature, i.e. Hybrid features are extracted. The candidate nodules are used to train Artificial Neural Networks (ANN). The classifier is tested and validated using nodules in standard dataset of LIDC. The classifier achieves the sensitivity of 96.95% with accuracy of 96.68% that is better than the existing CAD systems.

References

[1] Greenlee RT, Murray T, Bolden S, Wingo PA. Cancer Statistics, 2000. CA Cancer J Clin 2000; **50**: 7-33

[2] Jung KW, Won YJ, Park S, Kong HJ, Sung J, Shin HR, Park EC, Lee JS. Cancer statistics in Korea: incidence, mortality and survival in 2005. J Korean Med Sci 2011; **43**: 1-11

[3] Ozekes S. Rule-based Lung region segmentation and nodule detection via Genetic Algorithm trained template matching. Istanbul Comm Uni J Sci 2007; **6**: 17-30

[4] Ozekes S , Osman O, Ucan ON. Nodule detection in a lung region that's segmented with using genetic cellular neural networks and 3D template matching with fuzzy rule based thresholding. Korean J Radiol 2008; **9**: 1-9

[5] Ye X, Lin X, Dehmeshki J, Slabaugh G, Beddoe G. Shape based computer-aided detection of lung nodules in thoracic CT images. IEEE T Bio-Med Eng 2009; **56**: 1810-1820

[6] Retico A, Fantacc M E, Gori I, Kasae P, Golosio B, Piccioli A, Cerello P, Nunzio G D, Tangaro S. Pleural nodule identification in low-dose and thin-slice lung computed tomography. Comput Biol Med 2009; **39**: 1137-1144

[7] Sousa J R F S, Silva A C, Paiva A C, Nunes R A. Methodology for automatic detection of lung nodules in computerized tomography images. Comput Meth Prog Bio 2010; **98**: 1-14

[8] Lee S L A, Kouzani A Z, Hu E J. Random forest based lung nodule classification aided by clustering. Comput Med Imag Grap 2010; **34**: 535-542

[9] Maeda S, Tomiyama Y, Kim H, Miyake N, Itai Y, Tan J K, Ishikawa S, Yamamoto A. Detection of lung nodules in thoracic MDCT images based on temporal changes from previous and current images. J Adv Comput Intell Intell Infor 2011; **15**: 707-713

[10] Tan M, Deklerck R, Jansen B, Bister M, Cornelis J. A novel computer-aided lung nodule detection system for CT images. Med Phys 2011; **38**: 5630-5645

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

[12] Choi W J, Choi T S. Automated Pulmonary Nodule Detection System in Computed Tomography Images: A Hierarchical Block Classification Approach. Entropy 2013; **15**: 507-523

[13] Reeves AP, Biancardi AM, Apanasovich TV, Meyer CR, MacMahon H, Beek EJ, Kazerooni EA, Yankelevitz D, McNitt-Gray MF, McLennan G, et al. The Lung Image Database Consortium (LIDC): a comparison of different size metrics for pulmonary nodule measurements. Acad Radiol 2007; **14**: 1475-1485

[14] Brown MS, McNitt-Gray MF, Mankovich NJ, Goldin JG, Hiller J, Wilson LS, Aberle DR. Method for segmenting chest CT image data using an anatomical model: preliminary result. IEEE T Med Imaging 1997; **16**: 828-839

[15] Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE T Med Imaging 2001; **20**: 490-498

[16] Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin 2009; **59**: 225-249

[17] Messay T, Hardie RC, Rogers SK. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Med Image Anal 2010; **14**: 390-406

[18] Surez-Cuenca JJ, Tahoces PG, Souto M, Lado MJ, Remy-Jardin M, Remy J, Vidal JJ. Application of the iris filter for automatic detection of pulmonary nodules on computed tomography images. Comput Biol Med 2009; **39**: 921-933

[19] Paik DS, Beaulieu CF, Rubin GD, Acar B, Jeffrey RB Jr, Yee J, Dey J, Napel S. Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT. IEEE T Med Imaging 2004; **23**: 661-675

[20] Jusoh NA, Zain JM. Application of freeman chain codes: An alternative recognition technique for Malaysian car plates. Int J Comput Sci Net Sec 2009; **9**: 222-227

[21] Suzuki K, Armato SG 3rd, Li F, Sone S, Doi K. Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography. Med Phys 2003; **30**: 1602-1617

[22] Rubin GD, Lyo JK, Paik DS, Sherbondy AJ, Chow LC, Leung AN, Mindelzun R, Schraedley-Desmond PK, Zinck SE, Naidich DP, et al. Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. Radiology 2005; **234**: 274-283

[23] Dehmeshki J, Ye X, Lin X, Valdivieso M, Amin H. Automated detection of lung nodules in CT images using shape-based genetic algorithm. Comput Med Imag Grap 2007; **31**: 408-417

[24] Opfer R, Wiemker R. Performance analysis for computer-aided lung nodule detection on LIDC data. In Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment; March 2007; San Diego, Calif, USA; **6515** of Proceedings of the SPIE: pp. 65151C

[25] Sahiner B, Hadjiiski LM, Chan H, Shi J, Cascade PN, Kazerooni EA, Zhou C, Wei J, Chughtai AR, Poopat C, et al. Effect of CAD on radiologists' detection of lung nodules on thoracic CT scans: observer performance study. In: Proceedings of SPIE 6515, Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment; March 2007; San Diego, Calif, USA; **6515** of Proceedings of the SPIE: pp. 65151D

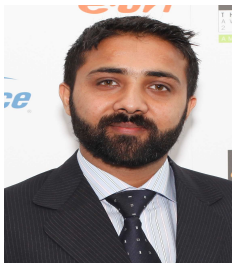
[26] Tartar A, Kilic N, Akan A. Classification of pulmonary nodules by using Hybrid features. Comput Math Meth Med 2013; **2013**: Article ID 148363, 11 pages

Sheeraz Akram completed his BCS (Honors) in Computer Science from International Islamic University Islamabad, Pakistan in 2004. He received his MS (Computer Science) from Lahore University of Management Sciences, Pakistan in 2006. Currently, he is a PhD scholar in Department of Computer Engineering, College of Electrical and Mechanical Engineering, National University of Sciences and Technology (NUST), Islamabad, Pakistan.

Muhammad Younus Javed completed his BSc in Electrical Engineering from UET Lahore in 1982. He received his MSc from the University of Dundee, UK in 1988 and PhD from the same university in 1991. He received ORS award from UK. He is serving at Department of Computer Engineering, College of Electrical and Mechanical Engineering, NUST, Islamabad, Pakistan since 1991 and presently he is Dean of Faculty of Engineering. He has been awarded "Best University Teacher Award" by the Higher Education Commission of Pakistan. In April 2010, he was declared "Best Researcher" of NUST, Islamabad, Pakistan. He has more than 250 national/international publications in well reputed journals and conferences.

Usman Qamar is currently an Assistant Professor at Department of Computer Engineering, College of Electrical and Mechanical Engineering, NUST, Islamabad, Pakistan. He is heading the "Data and Text Mining" Centre of the Department of Computer Engineering, College of Electrical and Mechanical Engineering, NUST, Islamabad, Pakistan. He has over 5 years of experience in data mining and data engineering both in academic and industry. His MPhil and PhD degrees are in the field of Data Mining. During his Post-Doc he was involved in various data mining projects for the industry.

Aasia Khanum holds a PhD degree in Software Engineering from NUST, Islamabad, Pakistan. She is working as an Associate Professor at Department of Computer Science, Forman Christian College, Lahore, Pakistan. She has published more than 30 papers in journals and conferences of international repute. Her research interests include application of intelligent techniques to multimedia processing, pattern recognition, data mining, software engineering, and cognitive computing.



Ali Hassan received his BE and MS in Computer Engineering from NUST, Islamabad, Pakistan in 2004 and 2007, respectively. He received his PhD in Electrical Engineering from the University of Southampton, UK, in 2012. He is currently working as Assistant Professor at NUST, Islamabad, Pakistan, College of Electrical and Mechanical Engineering, Department of Computer Engineering. His research interests include application of Machine Learning to Speech and Image Processing in the domains of Speech, Texture Classification and Biomedical Engineering.