Patient-Wise Versus Nodule-Wise Classification of Annotated Pulmonary Nodules using Pathologically Confirmed Cases

Preeti Aggarwal
UIET, Panjab University, Chandigarh, India
Email: pree_agg2002@yahoo.com

Renu Vig [2], H K Sardana [3]
[2]UIET, Panjab University, Chandigarh, India, [3] CSIO, Chandigarh, India
Email: renuvig@hotmail.com, hk_sardana@csio.res.in

***Abstract*— This paper presents a novel framework for combining well known shape, texture, size and resolution informatics descriptor of solitary pulmonary nodules (SPNs) detected using CT scan. The proposed methodology evaluates the performance of classifier in differentiating benign, malignant as well as metastasis SPNs with 246 chests CT scan of patients. Both patient-wise as well as nodule-wise available diagnostic report of 80 patients was used in differentiating the SPNs and the results were compared. For patient-wise data, generated a model with efficiency of 62.55% with labeled nodules and using semi-supervised approach, labels of rest of the unknown nodules were predicted and finally classification accuracy of 82.32% is achieved with all labeled nodules. For nodule-wise data, ground truth database of labeled nodules is expanded from a very small ground truth using content based image retrieval (CBIR) method and achieved a precision of 98%. Proposed methodology not only avoids unnecessary biopsies but also efficiently label unknown nodules using pre-diagnosed cases which can certainly help the physicians in diagnosis.**

***Index Terms*— Computer aided diagnosis, lung cancer, CT, SPNs, Haralick, Gabor, classification, SVM, PCA**.

I. INTRODUCTION

Most lung cancer treatment methods rely on early detection of malignant tumors. SPNs are common findings in thoracic imaging. Computer tomography (CT) scan is being investigated for a variety of radiologic tasks involving lung nodules and lung malignancies. The volumetric CT technique has introduced spiral scans which shorten the scan time and, when used in thoracic imaging, reduce the artifacts caused by partial volume effects, cardiac motion, and unequal respiratory cycles. For these reasons, spiral CT [1] is useful in identifying and characterizing SPNs. However, it is still difficult for radiologists to distinguish the SPNs in various classes.

Diagnostic decision making in pulmonary medical imaging has been improved by computer-aided diagnosis (CAD) systems serving as second readers to detect suspicious nodules for diagnosis by a radiologist. Computer aided detection (CADe) and diagnosis (CADx) aims to augment the radiologist in meeting the increased demand for diagnostic imaging by serving as second reader. While increasing accurate in detection and diagnosis, CADx rarely offers supporting guidance about the rationale for diagnosis or supplies descriptive annotations about medically meaningful diagnostic characteristics [2][3].

It is reported that two radiologists working together outperform in comparison with any independent radiologist [3]. The CAD system can provide a "second opinion, "which might improve the radiologist's performance. Recent studies have focused on the role of CAD in differentiating and characterizing pulmonary nodules. A CAD system for lung nodule detection and retrieval typically consists of the following components: Data acquisition, preprocessing, feature extraction, feature selection, classification and retrieval of similar images in the database from the same class to which that query nodule belongs and assessment of outputs. Various algorithms were designed using CT scans for automatic detection and diagnosis of lung nodules.

One of the greatest difficulties facing automatic lung nodule segmentation algorithms is the absence of a reliable and unambiguous ground truth [4] [5]. Many algorithms are trained on data from the Lung Image Database Consortium (LIDC) [4], which provides a reference truth based on the contours marked by four radiologists. Armato et al. explored the possible reference truths that may be constructed from the sets of nodules detected by different radiologists on the same CT scans, and found significant variations. In this paper, we have extracted the nodules marked by four different radiologists using semi-automated system [4]. Some studies have been done on finding and selecting features and evaluating the performance of classifiers of lung nodule and tissues for CAD and CBIR purposes. In feature selection studies, most researches focused on differentiating the visual features of pulmonary nodules and tissues, and there were few considerations about the differentiating features for classifying benign from

doi:10.4304/jcp.8.9.2245-2255

malignant SPNs. In classifier construction studies for lung CAD, linear discriminate analysis (LDA) and artificial neural networks (ANN) were studied intensively. However, in LDA, the complex decision surface might not be linear. In ANN, it was difficult to determine the number of units in the hidden layer. In CAD for lung cancer detection and diagnosis, nodule [5] detection in chest CT is the most important and tedious task.

In this study, a three-dimensional volumetric boundary is completely defined by a series of two-dimensional regions of interest (ROIs) for each slice in the nodule [6]. Although data sets were smaller for other preliminary studies, the results were encouraging. The features were extracted from the image data, with the goal of quantifying the visual features radiologists typically use to determine malignant, benign as well as metastasis nodules (a tumor growth or deposit that has spread via lymph or blood to an area of the body remote from the primary tumor like neck cancer, breast cancer etc.).Selecting the right features and constructing the higher performance classifier of pulmonary nodules are very important in developing the qualified CAD and CBIR systems [6], sometimes known as content-based medical image retrieval (CBMIR). In this research, we extracted total of 83 features including well-known texture features like: 21 Haralick features based on co-occurrence matrices (a statistical-based method) [7][8] [9], 24 Gabor filter features (a transform-based method) [10] [11], 3 First-order statistical features [12], 19 Shape Features, 8 GLDM features (Gray level difference method probability density function) [13], 9 Intensity Features [14].

In this paper, biopsy report of 80 patients which was available online was considered to label the segmented nodules in two ways, patient-wise as well as nodule-wise as both methods have their own importance for a physician in diagnosis. We presented a method for selecting pattern features of pulmonary nodules of CT images and evaluate the performance of SVM classifier in differentiating benign (B), malignant (M) as well as metastasis (MT) SPNs for patient-wise database. PCA [15] has been performed to reduce the number of features in both the cases. The results of this research are not only helpful to improve CAD for detection and diagnosis on SPNs but also useful to build the highly efficient feature index of a CBIR system for CT images with pulmonary nodules. We discussed the impacts of kernel function selection on the performances of SVM-based classifier in differentiating SPNs. In nodule-wise study, the available diagnosis is only for 17 nodules which was extended to 121 nodules using CBIR, then grafted decision tree is used to classify the nodules in the three mentioned classes. Though the classification accuracy for two class classifiers is better than the three class classifier yet the later can also be used for differentiating SPNs as specificity increases with adding more classes which clearly indicates that this study can at least help in avoiding unnecessary biopsies. The results show that SVM analysis is a noble predictor of SPNs that whether it is malignant, benign or metastasis.

II. Materials and Methods

A. Materials

CT scan of 246 patients with solitary pulmonary nodules mostly less than 3 cm have been taken from The Cancer Imaging Archive/ Lung Image Database Consortium (TCIA/LIDC) was included in our study. All the images are of size 512*512 and each having 16 bit resolution. All images are in DICOM (Digital Imaging and Communication in Medicine) format which is well known standard used in medical field. Each patient file is associated with an XML annotated file having details of nodule boundaries as well as physician's annotation is associated. Total of 8050 nodules are marked in 246 patients considering each slice of a patient. As same nodule can appear in different slice of a patient and even more than one nodule can be there in a single patient [16]. Accordingly, considering all these facts, total of 489 unique nodules from 8050 nodules were extracted. The details of the procedure are explained in the following sections. As only 80 biopsy confirmed cases of different patients were available, hence out of 489 only 224 nodules were labeled as Malignant (M), Benign (B) and Metastasis (MT) and Indeterminate (I) using patient-wise report. There were 52 malignant, 95 benign cases, 64 metastasis and 13 Indeterminate cases were available. For nodule-wise, ground truth of 121 nodules were prepared from 17 nodules using CBIR and classified in three classes. The largest nodule is of size 66x51 and smallest nodule is of size 6x6.

B. Methods for Patient-wise Study

For patient-wise study, 246 annotated chest CT scan images in DICOM format are collected from LIDC. The flowchart of our CAD system is shown the Figure 1. Four radiologists reviewed each scan and based on that XML-based message system was developed to communicate the results of each reading. Each XML file contains the boundary of each nodule whose diameter is not more than 30 mm which is consistent with upper limits of nodule size found in literature [3] as well as the nine annotations (calcification, internal structure, subtlety, margin, texture, malignancy, lobulation, sphericity and spiculation). Furthermore, the minimum effective diameter for included nodules has been set at 3mm. These XML files are read to locate the boundaries of each nodule in 246 patients and generate the mask for the same. Total of 8050 nodules were marked in 246 patients. As there are multiple slices for each patient as well as four marking from different radiologist were available, thus to remove this redundancy unique nodule were extracted for each patient study. There can be more than one nodule per patient, those are also considered and explained in next section. Total of 489 unique nodules are extracted for 246 patients. For these 489 nodules, 83 various features were extracted and finally applying PCA these features have been reduced to eleven and twelve relevant features for two classes and three classes respectively. Only 80 biopsy confirmed cases were available therefore out of 489 only 224 nodules were finally labeled as Malignant (M), Benign (B), Metastasis (MT) and Indeterminate (I).

As there were only 13 instances of indeterminate class available, so those were not considered as this much number of instances are very less to train a classifier for that class. Hence, finally for classification 211 instances with three classes (M, B and MT) of nodules were considered. The detail of each step is shown in Figure 1 with output.

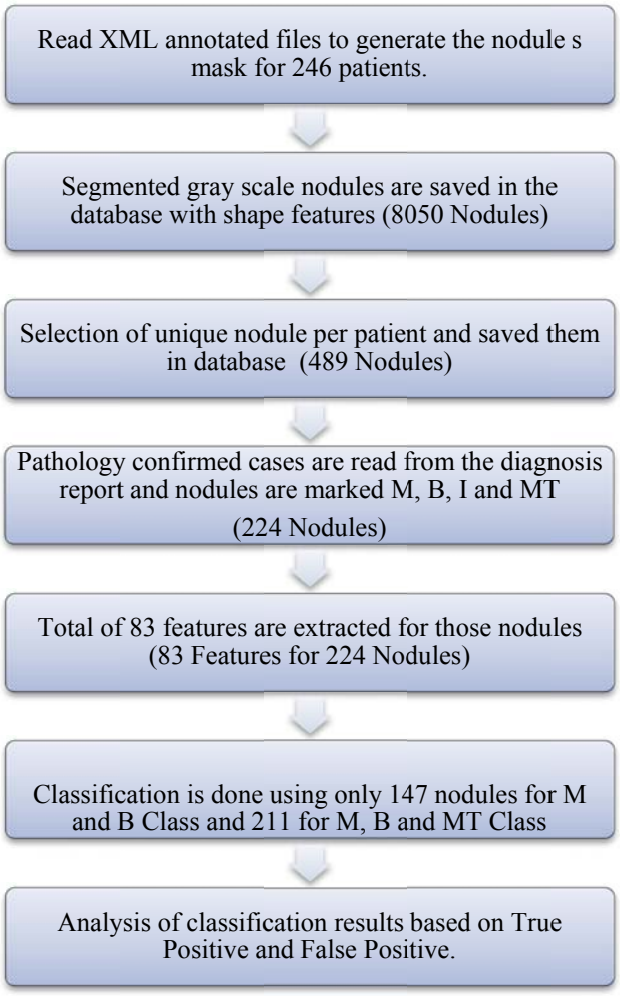


Figure 1: Flowchart of CAD system for classification of nodules

C. Methods for nodule-wise study

In nodule-wise study, Total of 1737 nodules are marked in 80 patients considering each slice of a patient having area greater than all those marked by four different radiologists. Out of 80 biopsy confirmed cases only 18 cases were available with single nodule. From these 18, only 17 cases were considered further to prepare the ground truth as diagnosis for one patient was unknown and this set will be referred to as the Diagnosed17. The classes assigned to these nodules were malignant, benign and metastases based on the diagnosis report available. Actually for these 17 patients, their patient-wise report is equivalent to nodule-wise report. Rest 62 patients were assigned the class based on the mean of malignancy rating provided by four different radiologists as no ground truth is available for these 62 patients with multiple nodules and this set will be referred as RadioMarked62. It contains 1677 nodules from 62 patients. As in patient-wise study, 83 features were extracted for each nodule. Four different "undiagnosed" query sets containing subsets of the LIDC Nodule Dataset were used, since neither computer-predicted nor radiologist-predicted malignancy ratings can be considered ground truth due to high variability between radiologists' ratings. Each of these query sets differed in diagnostic ground truth. The first query set (Rad210) used the radiologist-predicted malignancy, the second set (Comp210) used the computer-predicted malignancy, the third set (Comp_Rad_biopsy57) used only those nodules for which the radiologist, computer-predicted as well as biopsy confirmed malignancies agreed and the fourth set used only those nodules form which the radiologist- and computer-predicted malignancies agreed. For each query set, nodules with unknown malignancies were removed, and the set was balanced to contain all the three classes i.e. benign, malignant and metastases. The radiologist-predicted and computer-predicted contained equal number of nodules i.e. 210. and radiologist-computer-biopsy-agreement query set contained 57, and Rad_Comp92 contained 92 nodules after all modifications.

62 out of 80 biopsy confirmed cases with multiple nodules are assigned classes on the basis of radiologist's malignancy characteristics. The meaning and description of malignancy annotation feature of LIDC data is shown in Table1.

TABLE I.

MALIGNANCY RATINGS AND ITS MEANING IN LIDC DATASET

Diagnosis	Diagnosis at patient level as per LIDC diagnosis report	Class assigned in this work	Description
0	Unknown	I	In-determined
1	Benign	B	Non-Cancerous
2	Malignant	M	Cancerous
3	Metastases	MT	Cancer is spreading from other organ to lung.

Out of nine annotations only malignancy feature is used to assign the class to each nodule marked by radiologists as this is most promising feature to determine the malignancy of a nodule. Also, the other characteristics like margin, spiculation, and calcification are already involved in the medical definition of malignancy, so instead of considering all the nine only malignancy features is considered to assign the class as it approximately covers almost all the other features too. The method used to label each nodule is as follows

Nodules with malignancy rating >=3 assigned class Malignant (M) whereas
Nodules with malignancy rating <3 assigned class Benign (B)

Nodules are having multiple markings by four radiologists on different slices; therefore to reduce the variability among radiologists, the mean of the radiologists' ratings was used. In this way, 1677 nodules from 62 patients were assigned the malignancy class as above. These 1677 nodules contain multiple slices per nodule also and assigned to RadioMarked62 set, which further have been reduced to 210 and assigned to QueryNoduleSet210. If the same nodule appears in the multiple slices, then only those slices are considered in which nodule are having maximum area [17]. This method definitely reduces the database of nodules as well as makes the complexity of volumetric data simpler and effective to analyze.QueryNoduleSet210 further assigned to various categories like Rad210, Comp210 and Comp_Rad_biopsy210 as explained earlier.

III. Unique Nodule Detection from Annotated Images

In this study, the nodule and non-nodules are provided by The Cancer Imaging Archive (TCIA) [18], a large archive of medical images of cancer accessible for public download. The nodules in each patient are marked by four different radiologists and their location is saved as an XML file. Well-defined boundaries of each nodule are provided in XML files attached with each patient file which are read slice by slice and then surrounded by bounding box. The annotations available are very brief in majority of cases as they are filled out automatically by the machine. Most of the DICOM header information is hidden for ethical use.

As nodules in CT images are volumetric and to calculate the unique nodules in all patients, each slice is read independently to identify only unique nodules per patient. Only those nodules whose area is more than 25 are considered in this study. As per the literature [15] nodules with area less than 25 are mostly least important by physicians for consideration as a nodule. One patient can have more than one nodule, see Figure 2(b), hence those are also considered based on centroid method. The nodules whose centroid distance is more than 10 pixels and area greater than 25 are considered as separate/unique nodules, see equation 1:

$$\sqrt{(\boldsymbol{xcenter_1} - \boldsymbol{xcenter_2})^2 + (\boldsymbol{ycenter_1} - \boldsymbol{ycenter_2})^2} \quad \textbf{(1)}$$

Where $xcenter_1$ and $ycenter_1$ is the xcentroid and ycentroid respectively of one nodule and $xcenter_2$ and $ycenter_2$ is the xcentroid and ycentroid respectively of another nodule for the same patient. The size of pixel in each patient CT scan is not constant, which basically depend upon the CT scan machine. The pixel size varies from 0.5234 to 0.8340 to a side. Hence accordingly area of each nodule can be calculated as follows in equation 2:

$$\boldsymbol{AREA} = (x * y * w) \quad \textbf{(2)}$$

Where x and y denotes the height and width of a pixel respectively i.e. the actual size of the pixel and w denotes the number of pixels in a region. Hence, for calculating the number of unique nodules a threshold on area was kept which could be between 6.85 mm^2 to 17.39 mm^2 was considered .To find the unique nodules out of 489 nodules, this centroid method and area not less than 6.85 mm^2 was considered. As four radiologists have marked the nodules consequently it could be possible that the same nodule will fall in this category. In that case the nodule with maximum area is considered. Actually the boundaries provided in the XML files are already marked using manual as well as semi-automated methods [3] [5]. Subsequently, segmentation results covered most of the nodule area and captured most characteristics of the borders.

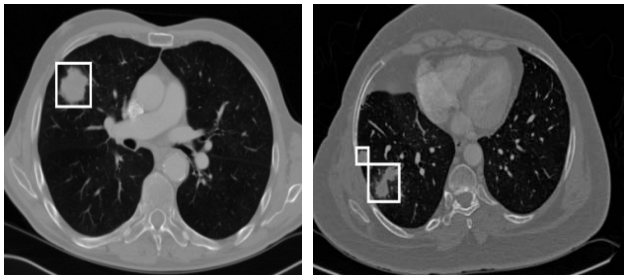


Figure 2: Nodules segmented from the original CT scan image

As we have used the extracted the boundaries of nodules which are marked by four different radiologists, thus our segmentation results are well approved by four radiologists and have taken as ground truth, see Figure 2.In Figure 2(a), a slice of a CT scan contains a single nodule where as in Figure 2(b), a slice from a CT scan contains two nodules. This method has also resolved the problems arise due to hard segmentation algorithms in our previous work [19].

IV. Feature Extraction and Selection

In image pattern recognition, feature extraction is the first step in image classification. The visual or low level features of lung nodules, such as the size, shape, and internal texture, intensity of ROI as well as background were considered in our study, as such characteristics would be considered by the radiologist when classifying a nodule as malignant or benign. Generally radiologist's primitive concern is whether the nodule is benign or malignant. Sometimes when malignancy is confirmed, then physicians are more interested to explore the form of cancer like whether it is a primary lung cancer or it is metastasis, which means that cancer is spreading in the body due to some secondary cancer like breast cancer, neck cancer etc. We performed specific feature extraction of lung CT images with nodules based on the parameters mostly suggested by physicians for identification of malignancy. Some features have good discriminative power, while other features contribute little to the classification. Therefore, the extracted features must be subjected to an optimal selection procedure before being used in classification. This selection procedure is further

described. A complete list of features extracted for lung nodules in CT images is shown in Table2.

All these features are normalized to the scale of 0-1 to maintain the consistency. Finally feature number 2 to 84, 83 features were extracted for 224 unique nodules and save in the database for classification. In case of Gabor features, the size of filter was tested for 3x3 as well as on 5x5 because the smallest nodule in the database is of the size 6x6 and largest as 66x51. The results for 5x5 were better than 3x3 in terms of classification accuracy for two classes as well as for three classes. In case of GLDM features, the values of the inter sample distance d is set at 11 as at this value the features contribute the highest classification accuracy. Feature number one is used to provide the name to each nodule like image1.dcm, image64.dcm. Each nodule is assigned a different name.

In the next section, these 83 features are transformed to a different set of reduced features by applying PCA and finally those features are nominated for classification purposes.

TABLE II.
A COMPLETE LIST OF FEATURES EXTRACTED FOR LUNG NODULES IN CT IMAGES

Feature Extraction Method	Feature No.	Feature Name in Database
Haralick Features [6][8]	2 ~ 21	inverse difference moment,autocorr, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, homogeneity, maximum probability, sum_of_squ, sum_avg, sum_var, diff_var, diff_entro, entropy, information measure of correlation1, information measure of correlation2, sum entropy, inverse difference normalized
Gabor Features[10]	22~45	gabor1_mean, gabor1_std, gabor2_mean, gabor2_std, gabor3_mean, gabor3_std, gabor4_mean, gabor4_std, gabor5_mean, gabor5_std, gabor6_mean, gabor6_std, gabor7_mean, gabor7_std, gabor8_mean, gabor8_std, gabor9_mean, gabor9_std, gabor10_mean, gabor10_std, gabor11_mean, gabor11_std, gabor12_mean, gabor12_std
Shape and Size Features [13]	46~56 74~81	area, xcenter, ycenter, perimeter, convexarea, solidity, extent, eccent, equidia, majoraxislen, minoraxislen, circularity, volume, perimeterequidia, sphericity, circularity2, roundness, compactness, concavity
GLDM Feature [12]	57~64	gldm1, gldm2, gldm3, gldm4, gldm5, gldm6, gldm7, gldm8
Intensity Features [13]	65~73	minint, maxint, meanint, sdint, minintBG, maxintBG, meanintBG, sdintBG, intdiff
First Order Statistics [11]	82~84	skew, kurt, stdd

A. PCA, a dimensionality reduction method for feature selection

High dimensional data analysis is becoming increasingly common as new problems are placing greater demands on computing resources. With high dimensional data, it is difficult to understand the any structure. A fast and simple algorithm for approximately calculating the principal components (PCs) of a data set and accordingly reducing its dimensionality is Principal Component Analysis (PCA). The objective of PCA is to perform dimensionality reduction while preserving as much of the randomness in the high-dimensional space as possible. PCA performs a rotation of the data that maximizes the variance in the new axes [20]. We enrolled 211 cases of SPNs in this study and extracted 83 features for each nodule. Thus, there was a high possibility of over fitting during the classification step due to the low number of samples relative to the number of features extracted, known as curse of dimensionality. For this reason, it was necessary to reduce the high dimensionality of the input feature vectors. For feature selection, a well-known dimensionality reduction method, PCA has been performed on all 83 features, which finally transform these features to twelve totally new features. As PCA is not basically feature selection method, it basically reduces the dimensionality using transformation of existing features especially in the field of CBIR [21], hence in our study is used for data pre-processing.

PCA is used to explore the usefulness of each feature and reduce the multidimensional features to simplified features with no underlying hidden structure. Besides, PCA is able to speed-up the computational time with the reduced dimensionality of the features while maintaining the classification accuracy [22] . The results of PCA in literature especially on biomedical imaging have shown tremendous results [21].

TABLE III.
LIST OF 12 PCA TRANSFORMED FEATURES FOR THREE CLASSES

New Name of Feature	Transformed Features
PCA_1	-0.17maxintBG-0.169gabor3_90_mean-0.169gabor5_0_mean-0.169gabor5_135_mean-0.169meanint
PCA_2	0.181majoraxislen+0.177perimeter+0.177perimeterequidia+0.176equidia+0.175gldm_pdf2_mean
PCA_3	-0.293Cluster Shade-0.281skewness+0.278sum_var+0.267sum_avg+0.262autocorr
PCA_4	0.355roundness+0.353compactness-0.32eccent+0.257extent+0.225solidity
PCA_5	-0.302Cluster Prominence-0.281sum_of_squ-0.277intdiff-0.266sdintBG-0.249sum_var
PCA_6	-0.324sdintBG-0.318stdd-0.311sdint-0.296intdiff+0.285Information measure of correlation2
PCA_7	-0.411Information measure of correlation1-0.367entropy-0.329Sum entropy+0.29 eccent-0.288roundness
PCA_8	-0.538kurtosis-0.244contrast-0.244diff_var-0.237Information measure of correlation1-0.22energy
PCA_9	-0.762xcenter-0.461ycenter+0.248kurtosis-0.162Maximum probability-0.161skewness
PCA_10	-0.744ycenter+0.32 Information measure of correlation1+0.255xcenter+0.242eccent-0.184Information measure of correlation2
PCA_11	0.368ycenter+0.325eccent+0.298kurtosis+0.287solidity-0.287concavity
PCA_12	-0.501xcenter-0.315kurtosis+0.273Maximum probability+0.234skewness+0.22 diff_var

In this study, 83 low level features of the 211 nodules are passes through PCA which finally produces 12 transformed features as shown in Table 3. As described in [20] [21], PCA actually find out the principal components in decreasing order, subsequently the first feature contains the highest principal components of the 83 features and so on. Finally for further study only these features have been used.

V. RESULTS

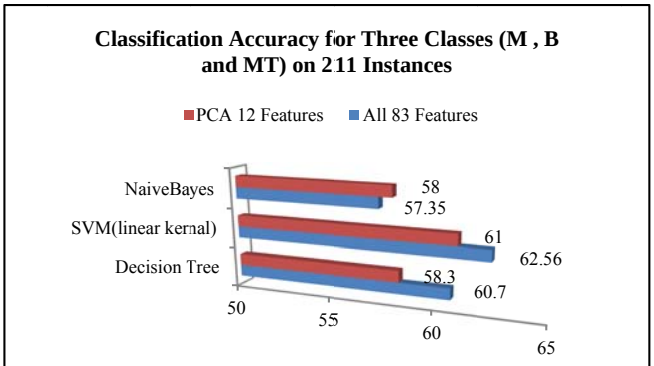
The results for each study are shown separately. First patient-wise results are explained followed by nodule-wise results.

Patient-wise Results

A. Selection of Best Performed Classifier on the Feature Set

After deducting the new set of features with the help of PCA, the results of which are shown in Table 2, the next step is to determine the relevance of each selected feature to the process of differentiating benign and malignant nodules and also the impact of these selected features to differentiate malignant, benign as well as metastasis SPNs. The number of unique nodule samples for differentiating malignant, benign as well as metastasis are 211. For classification purpose, well known supervised machine learning classification techniques decision tree, SVM as well as Naïve Bayes were implemented and tested [23]. As classification can be performed either with cross validation or holdout method, in this study as the number of samples are not much so we have adopted the former technique for classification. 10-fold as well as leave-one-out (LOO) cross validation is performed. In leave-one-out procedure one nodule was used for test purpose and the others were used for training the classifier. But the results of LOO were much better than 10-fold cross validation. Therefore, those were considered and explained in this study. Figure 3 shows the results of all the three classifiers on all the features as well as on PCA transformed features.

Every classifier has different nature to perform even on the same dataset. This is the reason for testing three classifiers on the same dataset so as to choose the best one for finding the accuracy of our feature data set. From Figure 3, we can see the highest accuracy is achieved using SVM with linear kernel. In SVM, the selection of appropriate kernel function is important, since the kernel function defines the transformed feature space in which the training set instances will be classified. It is common practice to estimate a range of potential settings and use cross-validation over the training set to find the best one. The use of linear kernel function in our study is also based on this assumption. As long as the kernel function is legitimate, a SVM classifier will operate correctly even if the designer does not know exactly what features of the training data are being used in the kernel-induced transformed feature space. For the implementation of this classification, LIBSVM [24] is used.



	Decision Tree	SVM(linear kernal)	NaiveBayes
PCA 12 Features	58.3	61	58
All 83 Features	60.7	62.56	57.35

Figure 3: Classification accuracy with all features and PCA features for three classes

Though the classification accuracy in this classification is not more than 62.56% using SVM with all the 83 features yet these results are encouraging as shown in Figure 4. It shows that the true positive rate in all the classes is high. The more number of features increases the classification accuracy as compared to PCA features in three classes. The reason behind is classifiers can suffer from the

problem of under fitting also sometimes when we increase the number of classes. In that case more number of features can help more in distinguishing the classes as compared to less ones. In literature too, SVM and Bayesian are the strongest classifiers to classify especially the medical data [25].
SVM is an emerging machine learning technology that has already been successfully used for image classification in both general as well as medical domain. It performs the classification between two or multiple classes by finding a decision surface that is based on the most informative points of the training set [26].

B. *Performance Evaluation of SVM-based Classifiers for Differentiating SPNs*

As from Figure 3, it is derived that SVM provides maximum classification accuracy with all the 83 features. Figure 4 shows the detailed accuracy of SVM classifier showing true positive rate vs. false positive rate ratio and precision vs. recall as well as sensitivity vs. specificity for each class corresponding to the classification accuracy as shown in Figure 3. These results show that most of the classifiers classify the benign class more accurately as compared to the malignant class. The reason behind this is the type of lung cancer as in some cases it can be adenocarcinoma and some cases it can be squamous cell carcinoma or any other type of cancer [1]. It is clear from the Figure 4 that recall, true positive rate as well as sensitivity values for each class is the same. Actually sensitivity and specificity are used to determine the effectiveness of a test, especially medical test in the diagnosis of a disease. Sensitivity refers to how good a test is at correctly identifying people who have a disease whereas specificity refers to how good a test is at correctly identifying people who are well.
Classification results may have errors if the classifier fails to identify an abnormality or identify an abnormality which is not present. These can be described by the following terms:

True Positive (TP): The classification result is positive in the presence of the clinical abnormality.
False Positive (FP): The classification result is positive in the absence of the clinical abnormality.
True Negative (TN): The classification result is negative in the absence of the clinical abnormality.
False Negative (FN): The classification result is negative in the presence of the clinical abnormality.

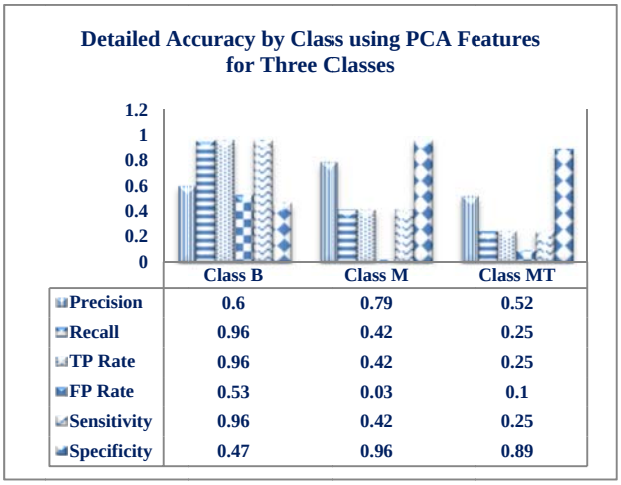


Figure 4: Detailed Accuracy with SVM using PCA features for three classes

TABLE IV.
FORMULAS USED FOR CALCULATING THE CLASSIFICATION ACCURACY

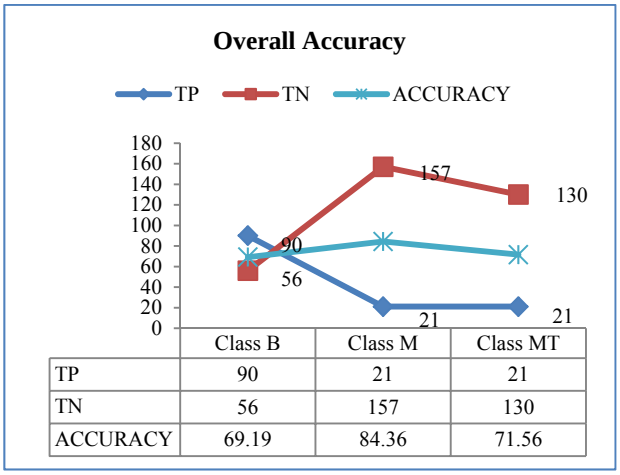
Precision=	TP / (TP+ FP)
Recall=	TP / (TP + FN)
Sensitivity=	TP / (TP+ FN)
Specificity=	TN / (TN+ FP)

From these results, it can be seen that class B has more sensitivity which indicates that more true positives were classified properly whereas specificity of class M is more that indicates that more true negatives were classified by SVM. The results of classification are purely dependent on the data and its ground truth labels. In this study as the labeled data was less, this result in decrease in the classification accuracy for three classes. Also, for unbalanced data sets, accuracy may not be good criteria for evaluating a model. The number of false positive nodules (FPNs) should be as less as possible as compared to true positive nodules (TPNs), which are clearly indicated in our results. From Table 4, the final accuracy for three classes from this study can be derived from equation 3;

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} * 100\ \% \qquad (3)$$

Figure 5 indicates the overall performance of the classification based on equation 3. It indicates the highest accuracy is achieved by class M which is most desirable by physician's for the detection and diagnosis of lung cancer.

To increase the classification accuracy from 62.56%, it was necessary to add more labeled data. For this, the classification model built was again tested on 264 unknown nodules to predict their class. After labeling, the whole data containing 475 labeled SPNs, out of which 173 were labeled as B, 226 were labeled as M and 76 were labeled as MT, were again tested with SVM and 10-fold cross validation.



	Class B	Class M	Class MT
TP	90	21	21
TN	56	157	130
ACCURACY	69.19	84.36	71.56

Figure 5: Overall Accuracy for whole labeled dataset

The classification accuracy increased to 82.32%; see Table 5 for confusion matrix

TABLE V.
CONFUSION MATRIX FOR CLASSIFICATION OF 475 LABELED NODULES USING SVM AND 10-FOLD CROSS VALIDATION

a	b	c	Classified as
167	**1**	**5**	**a=B**
20	**194**	**12**	**b=M**
44	**2**	**30**	**c=MT**

From Table 5, the Receiver-Operator-Characteristic curve (ROC) can be plotted for both the cases. Figure 6 shows the ROC curve using SVM classification to identify SPNs using the current feature set. Though sensitivity and specificity are the candidate measures for accuracy yet physicians rely more on ROC curves as it plots sensitivity against 1-specificity.The results clearly show that the sensitivity though decreases but specificity increases for detailed classification of SPNs.

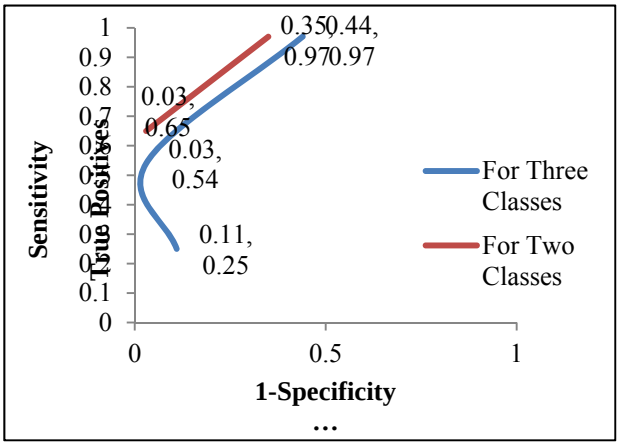


Figure 6: ROC Curve using SVM for identification of pulmonary nodules

Also, for more labeled data the peak is more closed to value 1 as shown by green star. It indicates that by increasing the database of labeled nodules physicians can definitely benefit the proposed methodology for the unknown cases in future. Having high sensitivity is not necessarily a good thing as compared to specificity because it is really important for a doctor to declare that the person is well and he is not suffering from any disease. This can definitely help in avoiding the unnecessary biopsies done for normal patients.

Nodule-wise Results

As ground truth for only 17 patients were available, there is a need to expand the diagnostically labeled database. In the absence of diagnostic information, labels can be applied to unlabeled data using semi-supervised learning (SSL) approaches. In SSL, unlabeled data is exploited to improve learning when the dataset contains an insufficient amount of labeled data [27]. CBIR can be used as a machine learning process that trains a system to classify images as relevant or irrelevant to the query. Using available datasets and by evaluating the method with a CAD application, we determined how to effectively expand the Diagnosed17 with CBIR and assist the physicians in the final diagnosis. Each nodule in the QueryNoduleSet210 was then used as a query to retrieve the ten most similar images from the remaining nodules in the Diagnosed17 using CBIR with Euclidean distance. The query nodule was assigned predicted malignancy ratings based on the retrieved nodules (e.g., if the maximum retrieved nodules belong to class malignant then the query nodule was assigned the class M), Figure 7. The newly identified nodule was considered candidates for addition to the Diagnosed17.

A. Diagnosed Subset Evaluation

In the current study, we adopted a semi-supervised approach for labeling undiagnosed nodules in the LIDC. CBIR was used to label nodules most similar to the query with respect to Euclidean distance of image features. Nodules to be added to the Diagnosed17 were selected from the candidates described above. For verifying the addition of a candidate nodule in the Diagnosed17, a reverse mechanism is adopted. Diagnosed17 nodules acted as query and nodules to be retrieved are from QueryNoduleSet210, see Figure 7.

The first three similar nodules are assigned the same malignancy as the query nodule if they were previously assigned as candidate nodules (i.e. if the query nodule is benign then the top three retrieved nodules are also assigned the class benign if previously are assigned as candidate nodule). Finally based on CBIR and CAD nodules are added in the Diagnosed17. With this mechanism Diagnosed17 in expanded to Diagnosed74, which means that now 74 nodules have the confirmed diagnosis and can be treated as LIDC ground truth. Predicted diagnosis with the pathologically-determined diagnosis, this process guarantees the accuracy of the CBIR-based diagnostic labeling.

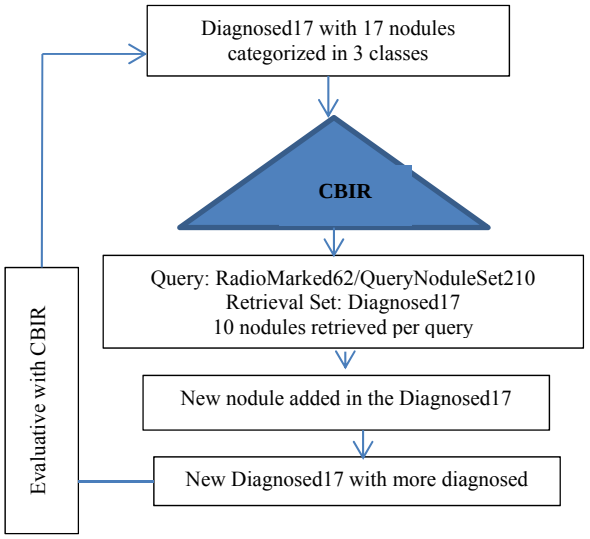


Figure7. Selection of candidate nodules using CBIR and Diagnosed17

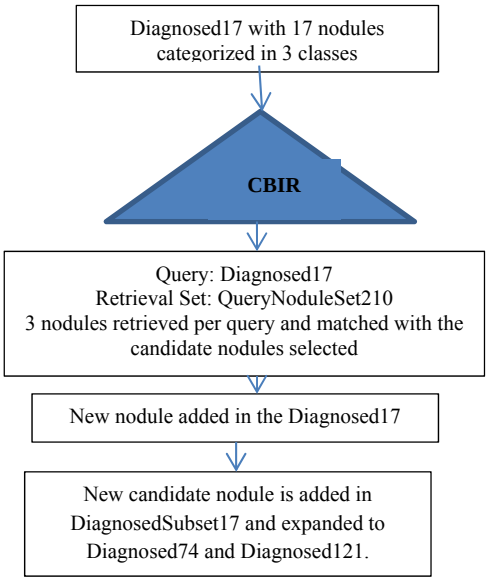


Figure8. Expansion of Diagnosed17 to Diagnosed74 and Diagnosed121

B. CBIR Mapping of Multiple Nodules Database with Single Nodule Database

An independent CBIR framework is implemented to increase the Diagnosed17 using CBIR from the QueryNoduleSet210. QueryNoduleSet210 is having multiple nodules per patient. 210 different nodules are present in this set. One by one each nodule is taken as query nodule and matched against Diagnosed17 using CBIR with Euclidean distance. As patient-wise diagnosis is available for QueryNoduleSet210, hence the top retrieved result is matched with this diagnosis. If top retrieved nodule class matches with the patient-wise diagnosis of query nodule then it is added in Diagnosed17 else discarded. With multiple iterations in this manner, Diagnosed17 is increased to Diagnosed121, see Figure 8. Predicted diagnosis with the pathologically-determined diagnosis, this process guarantees the accuracy of the CBIR-based diagnostic labeling.

C. Query and Retrieval Sets Concluded

In this CAD scenario, two ways process is implemented as discussed earlier. Once the nodules in Diagnosed17 were used as query and QueryNoduleSet210 was used for retrieving the nodules based on CBIR and Euclidean distance and expanded the ground truth to 74 nodules. Secondly, nodules in QueryNoduleSubset210 were treated as query and Diagnosed17 set was used to retrieve most similar nodules to assign the malignancy class accordingly and expanded the Diagnosed dataset to 121. Since neither computer-predicted nor radiologist-predicted malignancy ratings can be considered ground truth due to high variability between radiologists' ratings [28]. This mechanism guarantees the preparation of LIDC ground truth and accuracy of CBIR based diagnostic labeling. All the nodules can be classified in three class benign, malignant and metastases. Various query sets were formed and their precision are compared and shown in Figure 9.

Using the query and retrieval sets as described above, average precision after 3, 5, 10, and 15 images retrieved was calculated. A retrieved nodule was considered relevant if its diagnosis matched the malignancy rating (either radiologist-predicted, computer-predicted, or both) of the query nodule. Initial precision values were obtained by using the 17 nodules in the initial Diagnosed17 as the retrieval set. Then, nodules were added to this set as described earlier. Precision was recalculated, and the nodule addition process was repeated iteratively using the new Diagnosed17. In each subsequent iteration, only the newly added nodules in the Diagnosed17 were used to identify new candidates. This process repeated until no candidate nodules were added to the Diagnosed17 following an iteration. Various experiments were setup for the validation of nodules examined.

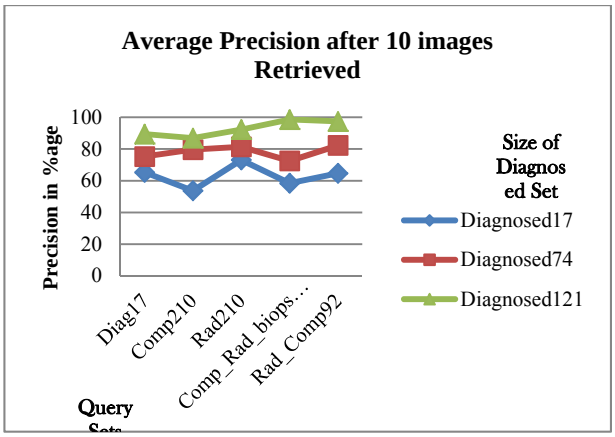


Figure9. Comparison of precision for different query sets at x-axis and different retrieval sets at y-axis.

Figure 9 shows that with five query sets and three retrieval sets Diagnosed17, Diagnosed74 and Diagnosed121, the precision increases respectively. Nodules in Comp_Rad_biopsy57 have provided the best

precision i.e. 98% which is the best precision achieved in the history of medical CBIR with best of our knowledge. The above results clearly support in avoiding unnecessary biopsies which is also one of the major aim of physician's in addition to save a patient having cancer.

VI. Conclusion and Future Work

Early detection of lung tumors (visible on chest film as nodules) may increase the patient's chance of survival. In this study, our focus is also on early detection of nodules using pre-diagnosed cases based on classification. The evaluation results in this study indicate that most of the selected features have important contribution in differentiating SPNs. Our classification results shows that adding more class definitely decrease the classification accuracy because actual nodules are not exactly spherical, circular, some true nodules can be missed as well as FPNs can be encountered. Figure 5 and Figure 9 shows that FPNs in all the three classifiers are less than TPNs in both patient-wise as well as nodule-wise study. No doubts that feature selection and dimensionality reduction have increased the overall accuracy in terms of time as well as classification. PCA features have not shown remarkable results in differentiating SPNs yet these features have important aspect in making the feature index for CBMIR.

Nodule-wise classification and its precision is more than patient-wise as patient-wise labeled nodules are biased as well as not that much accurate. CBIR is an effective method for expanding the Diagnosed Subset by labeling nodules which do not have associated diagnoses. As LIDC is having lack of ground truth, CBIR techniques works tremendously better to prepare the ground truth. By increasing the size of the Diagnosed Subset from 17 to 74 and finally to 121 nodules, CBIR expansion provides greater variability in the retrieval set, resulting in retrieved nodules that are more similar to undiagnosed queries. The proposed CBIR expansion method can be applied to differentiate benign, malignant as well as metastases nodules. In comparison to [29], our results outperforms as precision achieved in our nodule-wise study is 98% as compared to 90%. This clearly indicates the features extracted in our study are more discriminative as compared to [29]. The third class metastasis has not been introduced in the history of CBIR and medical imaging. An expanded set of diagnosed images is also useful for non-CBIR CAD systems, which require large datasets for robust and unbiased training and testing. In future studies, we will investigate using different distance metrics for nodule similarity when identifying candidates with the CBIR expansion method. We also plan to add more classes of malignancy as well as benign to further assist the physicians in more accurate diagnosis.

Using incremental learning approach to increase the labeled data, the same classification model can be used which no doubt will improve the classification accuracy. The results of this research are not only helpful to improve CAD for diagnosis on SPNs but also useful to build the highly efficient feature index of a CBIR system for CT images with pulmonary nodules.

Acknowledgment

The work was supported by The Cancer Imaging Archive/ Lung Image Database Consortium (TCIA/LIDC). The authors would like to thank Dr. Samuel G. Armato for providing the valuable information about LIDC-IDRI data and the access to the pre-diagnosed CT cases used in this study.

References

[1] Dodd LE, Wagner RF, Armato III SG, McNitt-Gray MF, Beiden S, Chan H-P, Gur D, McLennan G, Metz CE, Petrick N, Sahiner B, Sayre J. 2004. The LIDC Research Group: Assessment methodologies and statistical issues for computer-aided diagnosis of lung nodules in CT: Contemporary research topics relevant to the Lung Image Database Consortium. Academic Radiology; 11:462–475, 2004.

[2] Hyun-Chong Cho, LubomirHadjiiski, BerkmanSahiner, Heang-Ping Chan, ChintanaParamagul, Mark Helvie, and Alexis V. Nees. 2012. Interactive content-based image retrieval (CBIR) computer-aided diagnosis (CADx) system for ultrasound breast masses using relevance feedback. **Proceedings of SPIE,** Medical Imaging; 8315: 831509-1-831509-7, 2012.

[3] Hyun-Chong Cho, Lubomir Hadjiiski, Berkman Sahiner, Heang-Ping Chan, Mark Helvie, Chintana Paramagul, and Alexis V. Nees. 2011. Similarity evaluation in a content-based image retrieval CBIR, CADx system for characterization of breast masses on ultrasound images. Medical Physics: 38(4), 1820-1831. 2011.

[4] Armato SG III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, Zhao B, Aberle DR, Henschke CI, Hoffman EA, Kazerooni EA, MacMahon H, van Beek EJR, Yankelevitz D, et al. 2011. The Lung Image Da-tabase Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed refer-ence database of lung nodules on CT scans. Medical Physics; 38: 915–931, 2011.

[5] Armato SG III, Roberts RY, McNitt-Gray MF, Meyer CR, Reeves AP, McLennan G, Engelmann RM, Bland PH, Aberle DR, Kazerooni EA, MacMahon H, van Beek EJR, Yankelevitz D, Croft BY, Clarke LP. 2007. The Lung Image Database Consortium (LIDC): Ensuring the integrity of expert-defined "truth." Academic Radiology; 14: 1455–1463, 2007.

[6] Armato III SG, McLennan G, McNitt-Gray MF, Meyer CR, Yankelevitz D, Aberle DR, Henschke CI, Hoffman EA, Kazerooni EA, MacMahon H, Reeves AP, Croft BY, and Clarke LP, et al. 2004. The Lung Image Database Consortium (LIDC): Developing a resource for the medical imaging research community. Radiology; 232:739–748, 2004.

[7] Robert M. Haralick, K. Shanmugam, Its'hakDinstein. 1973. Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics Smc; 3(6): 610-621, 1973.

[8] Neeraj Sharma, Amit K. Ray, Shiru Sharma, K. K. Shukla, SatyajitPradhan, Lalit M. Aggarwal. 2008. Segmentation and classification of medical images using texture-primitive features: Application of BAM-type artificial neural network, J Med Phys; 33(3):119–126, 2008.

[9] FahimIrfanAlam, RokanUddinFaruqui. 2011. Optimized Calculations of Haralick Texture Features. European Journal of Scientific Research; 50(4):543-553, 2011.
[10] Andrysiak T, Choras M. 2005. Image retrieval based on hierarchical Gabor filters. International Journal of Applied Computer Science; 15(4): 471-480, 2005.
[11] Vibha S. Vyas and PritiRege. 2006. Automated Texture Analysis with Gabor filter. GVIP Journal; 6, Issue 1: 35-41, 2006.
[12] OkyDwiNurhayati, Thomas Sri Widodo, AdhiSusanto, MaesadjiTjokronagoro. 2010. First Order Statistical Feature for Breast Cancer Detection Using Thermal Images. World Academy of Science, Engineering and Technology; 46:424-26, 2010.
[13] J. K. Kim and H. W. Park. 1999. Statistical textural features for detection of micro calcifications in digitized mammograms. IEEE Trans. Med. Imag; 18: 231-238, 1999.
[14] Raicu DS, Varutbangkul E, Furst JD, Armato SG III. 2010. Modeling semantics from image data: Opportunities from LIDC. International Journal of Biomedical Engineering and Technology; 3: 83–113, 2010.
[15] Cheng Chen, Ozolek, J.A, Wei Wang, Rohde, G.K. 2011. A pixel classification system for segmenting biomedical images using intensity neighborhoods and dimension reduction. Biomedical Imaging: From Nano to Macro; 2: 1649-1652, 2011.
[16] Reeves AP, Biancardi AM, Apanasovich TV, Meyer CR, MacMahon H, van Beek EJR, Kazerooni EA, Yankelevitz DF, McNitt-Gray MF, McLennan G, Armato SG III, Henschke CI, Aberle DR, Croft BY, Clarke LP. 2007. The Lung Image Database Consortium (LIDC): A comparison of different size metrics for pulmonary nodule measurements. Academic Radiology; 14: 1475–1485, 2007.
[17] Preeti Aggarwal, Renu Vig, and H K Sardana, Largest Versus Smallest Nodules Marked by Different Radiologists in Chest CT Scans for Lung Cancer Detection, International conference on image engineering, ICIE-2013 organized by IAENG at Hong Kong.
[18] TCIA, The cancer imaging archieve. Available at http://cancerimagingarchive.net. (Accessed on July 17, 2010)
[19] PreetiAggarwal, RenuVig, SonaliBhadoria, C.G.Dethe. 2011. Article: Role of Segmentation in Medical Imaging: A Comparative Study. International Journal of Computer Applications; 29(1):54-61, 2011.
[20] Xue-min Mao, Chuan-xi Cai, Bing-yu Sun. 2011. Comparative research on methods of dimensionality reduction in high-dimension medical data. Fourth International Workshop on Advanced Computational Intelligence (IWACI); 4: 586-589, 2011.
[21] Dr. H. B. Kekre, Sudeep D. Thepade, Sudeep D. Thepade, Sudeep D. Thepade. 2010. CBIR Feature Vector Dimension Reduction with Eigenvectors of Covariance Matrix using Row, Column and Diagonal Mean Sequences. International Journal of Computer Applications; 3(12): 0975 – 8887, 2010.
[22] Tan ChuePoh, Nu FatehaMuhamadLani, Lai Wenf Kin. 2008. Multi-dimensional feature reduction of PCA on SVM classifier for imaging surveillance application. Proceedings of the 7th WSEAS, International Conference on Signal Processing, Robotics and Automation ISPRA: 192-197, 2008.
[23] S. B. Kotsiantis. 2007. Supervised Machine Learning: A Review of Classification Techniques. Informatica; 31: 249-268, (2007).
[24] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology; 2:27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[25] A. Farag and et al.. 2002. Detection and recognition of lung nodules in spiral CT images using deformable templates and Bayesian post-classification. ICIP; 5:2921–2924, 2002.
[26] Chapelle O, Haffner P, Vapnik V. 1999. SVMs for histogram-based classification. IEEE Trans Neural Networks 1999; 10(5):1055-64.
[27] Z.-H. Zhou, "Learning with Unlabeled Data and Its Application to Image Retrieval," PRICAI'06 Proceedings of the 9th Pacific Rim International Conference on Artificial Intelligence, 2006.
[28] W. H. Horsthemke, D. S. Raicu, J. D. Furst, and S. G. Armato III, "Evaluation Challenges for Computer-Aided Diagnostic Characterization: Shape Disagreements in the Lung Image Database Consortium Pulmonary Nodule Dataset," In J. Tan, New Technologies for Advancing Healthcare and Clinical Practices, IGI Global, Hershey PA, pp. 18-43, 2011.
[29] Anne-Marie Giuca, Kerry A. Seitz Jr., Jacob Furst, Daniela Raicu, Expanding diagnostically labeled datasets using content-based image retrieval, IEEE International Conference on Image Processing 2012, September 30 - October 3, Lake Buena Vista, Florida.

Preeti Aggarwal, Assistant Professor, Department of Computer Science and Engineering, UIET, Panjab University, Chandigarh is pursuing her research in the field of medical imaging. She has done M.E. (Computer Science and Information Technology). She has more than 15 research papers in the field of CBMIR.

Dr. Renu Vig, Professor and Director of UIET, Panjab University, Chandigarh. Her research includes pattern recognition, neural networks. She has number of publications in renowned journals.

Dr. H.K. Sardana, Addl. Director, CSIO, Chandigarh. He is Scientist-G and involved in various projects like rice classification, cephlogram landmark detection at CSIO. His research area includes pattern recognition, medical imaging, and sensors. He has number of publications in international journals and he is the reviewer of various journals.