

Largest Versus Smallest Nodules Marked by Different Radiologists in Chest CT Scans for Lung Cancer Detection

Preeti Aggarwal, Renu Vig, and H K Sardana

□

Abstract— In this paper, we present a novel approach to find and select texture features of solitary pulmonary nodules (SPNs) detected by computed tomography (CT) and evaluate the performance of grafted decision tree based classifier in differentiating benign from malignant as well as from metastasis SPNs. We compared the results of smallest as well as largest nodule of a patient visible in different slices of CT scan and conclude that by taking the slice of a patient with largest area nodule is better in classifying the SPNs in 3 classes as compared to considering the nodules with smallest area of the same patient. It also reflects that specificity as well as sensitivity is much better which can further assist the physician in yielding the right decision at right time in the detection and diagnosis of lung cancer. This study reveals that there could be a significant improvement in the field of lung nodule detection at an early stage of lung cancer and also ensures that unnecessary biopsies can be avoided using the proposed methodology and feature set.

Index Terms— Chest CT, Classification, Computer-aided-diagnosis (CAD), Lung Cancer, Texture Features.

I. INTRODUCTION

There are lots of work being done to develop computer assisted diagnosis and detection (CAD) technologies and systems to improve the diagnostic quality for pulmonary nodules in chest CT scan. CAD system can provide a “second opinion, —which might improve the radiologist’s performance. SPNs are common findings in thoracic imaging. 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 is useful in identifying and characterizing SPNs. There are many challenges in creating and evaluating such systems including the lack of ground truth provided by pathology reports, variability in the radiologists’ interpretation, and multiple instances per nodule caused by multiple CT slices intersecting a nodule.

In our work, using the National Cancer Institute (NCI) Lung Image Database Consortium (LIDC) [1] dataset,

Manuscript received December 08, 2012; Revised on January 08, 2013.

Preeti Aggarwal. She is working as Assistant Professor at UIET, Panjab University, Chandigarh, India (phone: 91-9872021863; E-mail: pree_agg2002@yahoo.com).

Renu Vig. She is with UIET, Panjab University, Chandigarh, India. She is the Director of the institute. (E-mail: renu_vig@hotmail.com).

H K Sardana. He is the Addl. Director, CSIO, Chandigarh (E-mail: hk_sardana@csio.res.in).

we propose a way which predicts that selecting the slices of a patient containing a nodule with maximum area marked by four different radiologists’ yields better classification results as compared to slices containing nodule with minimum area. This is because the bigger nodule features are more class discriminatory as compared to smallest one. This research is purely based on the experience and development of useful concepts in the field of medical imaging.

A. Lung Nodules

Lung cancer is one of the most lethal cancer types. Recent studies [2][3][4] report that lung cancer accounts for 32% and 25% of cancer deaths among men and women respectively, and causes 150,000 deaths a year in the United States. Lung nodule volumetric is used for nodule diagnosis as well as for monitoring tumor response to therapy. CT scan of chest is the better method to analyze these nodules for detection as well as for diagnosis. Due to multiple slices in CT, the physician has to see each and every slice for better understanding of each nodule, if present. This task is time consuming as well as not deterministic in any way. Similarly, to classify these SPNs into various classes like malignant, benign, metastasis etc. is not so easy. Nodules can be malignant i.e. cancerous, benign i.e. non-cancerous, metastasis i.e. the primary cancer is not lung cancer but due to the spreading of the cancer it has reached to lung. For example, breast cancer sometimes spreads so fast that it covers even the lungs. These are not only the fixed classes of lung cancer, according to the data and studies can vary. In this study from the diagnosis report available with LIDC data, nodules are classified in these three classes.

B. Perfect Segmentation of Nodules

Four experienced radiologists have marked the nodules consequently. The boundaries provided in the XML files are already marked using manual as well as semi-automated methods [1] [5]. Subsequently, segmentation results covered most of the nodule area and captured most characteristics of the borders.

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 1.

In Figure 1(a), a slice of a CT scan contains a single nodule where as in Figure 1(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 [6].

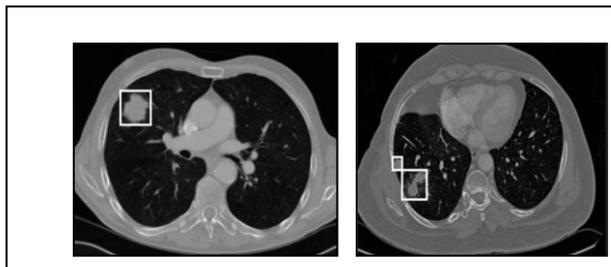


Fig 1(a, b): Nodules segmented from the original CT scan image

II. MATERIALS AND METHODS

CT scan of 80 biopsy confirmed 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 1733 nodules are marked in 80 patients considering each slice of a patient having area greater than all those marked by four different radiologists. Similarly 1741 nodules were extracted in the same way except with minimum area. As only 80 biopsy confirmed cases of different patients were available, hence these nodules were labeled as Malignant (M), Benign (B) and Metastasis (MT) There are 527 malignant, 518 benign cases, 688 metastasis cases available in maximum area database whereas in minimum area database 538 malignant, 523 benign and 680 metastasis cases available. All nodules vary in size and shape. The largest nodule is of size 66x51 and smallest nodule is of size 6x6. The complete flow of the study is shown in Figure 2 which is described in next section.

III. NODULE DETECTION FROM ANNOTATED CT IMAGES

In this study, the nodule and non-nodules are provided by TCIA, 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 almost available in each slice of patient. To make the task easier and effective slices were selected based on the annotations available by four different. Each slice is read independently to identify its area marked by all the four radiologists and only those slices per nodule is considered to be in the database whose area is either maximum or minimum. On the whole only those nodules whose area is more than 25 pixels are considered in this study. 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 0.5234mm to 0.8340mm to a side. Hence accordingly

area of each nodule can be calculated as follows in equation 1:

$$\text{Area} = (x * y * w) \quad (1)$$

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, a threshold on the area of nodules was kept which could be between 6.85 mm^2 to 17.39 mm^2

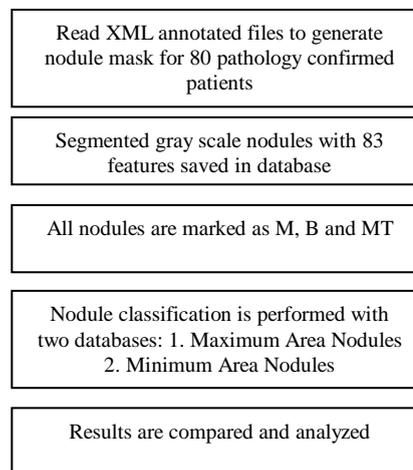


Fig 2: Nodules segmented from the original CT scan image

In Figure 2, it is clear that two databases were examined and compared for this study, one with maximum area nodules and other with minimum area nodules. This study reveals that largest nodules can assist to develop a better CAD system as it contains more context and content as compared to smaller one. This type of CAD system will definitely assist the physicians as second opinion for any future case without the need of further biopsies, if any case it can be avoided.

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. Generally, 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.

A complete list of features extracted for lung nodules in CT images is shown in Appendix A. In Table I, feature number

2 to 84, 83 features were extracted for all the nodules in two databases 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 and hence included in this study. 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 for identification purposes.

In this study also, we had conducted experiments, using all features versus reduced features. Figure 3 shows the desired results. Finally these features are nominated for classification purposes.

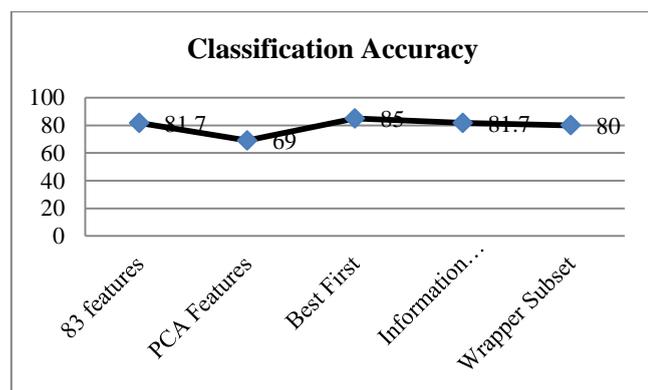


Fig 3: Comparison of full features with reduced data set

Figure3 shows that best first method provides the best classification accuracy with maximum area databases. The 10 best selected features by best first method are shown in Table II.

V. PERFORMANCE EVALUATION OF GRAFTED DECISION TREE CLASSIFIER FOR DIFFERENTIATING SPNS

Figure 4 shows the detailed accuracy of grafted decision tree classifier [14] showing true positive rate vs. false positive rate ratio and precision vs. recall for database1 (DBI) including nodules with maximum area and selected features by Best First method shown in section IV. Overall accuracy for the DBI is 85%. Similarly, Figure 5 shows the details for database2 (DBII) including minimum area nodules. Overall accuracy for the DB II is 82%.

Similarly, average of sensitivity and specificity for all the three classes are plotted against each other. 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.

TABLE II
REDUCED FEATURE SET USING BEST FIRST METHOD

Reduced Features	Detail of the Features
gabor1_mean	Mean of 5*5 Gabor filter with 0.3 frequency with degree 0
gabor3_std	Standard deviation of 5*5 Gabor filter with 0.5 frequency with degree 0
gabor4_mean	Mean of 5*5 Gabor filter with 0.3 frequency with degree 45
gabor8_std	Standard deviation of 5*5 Gabor filter with 0.4 frequency with degree 90
gabor11_std	Standard deviation of 5*5 Gabor filter with 0.4 frequency with degree 135
xcenter	X axis centroid
ycenter	Y axis centroid
minint	Minimum intensity of nodule for ground
minintBG	Minimum intensity of nodule back ground
circularity	Circularity is the shape feature of a nodule explains how much its circular is. 1 means more

Detailed Accuracy By Class using reduced features for database I

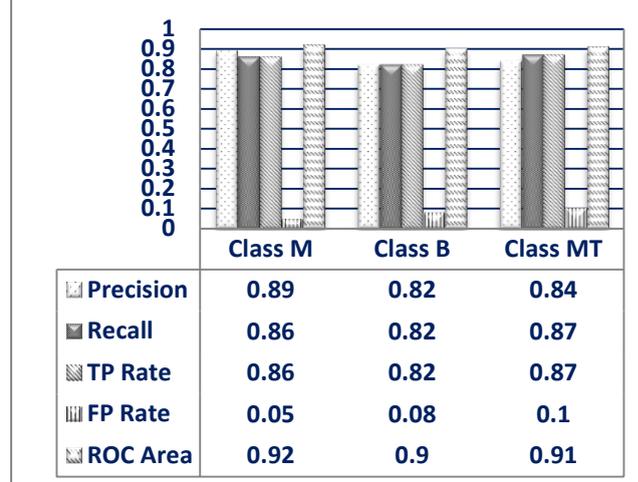


Fig 4: Detailed accuracy by class using reduced features for DB I

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.

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{2}$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \tag{3}$$

$$\text{Sensitivity} = \frac{TP}{(TP + FN)} \tag{4}$$

$$\text{Specificity} = \frac{TN}{(TN + FP)} \tag{5}$$

Equation 2-4 explains the values calculated for Precision, Recall, Sensitivity and Specificity. These terms are required to calculate especially in the field of medical science. Actually meaning of sensitivity is the ratio of number of cases correctly called positive and total number of positive cases whereas specificity is the ratio of number of cases correctly called negative and total number of negative cases.

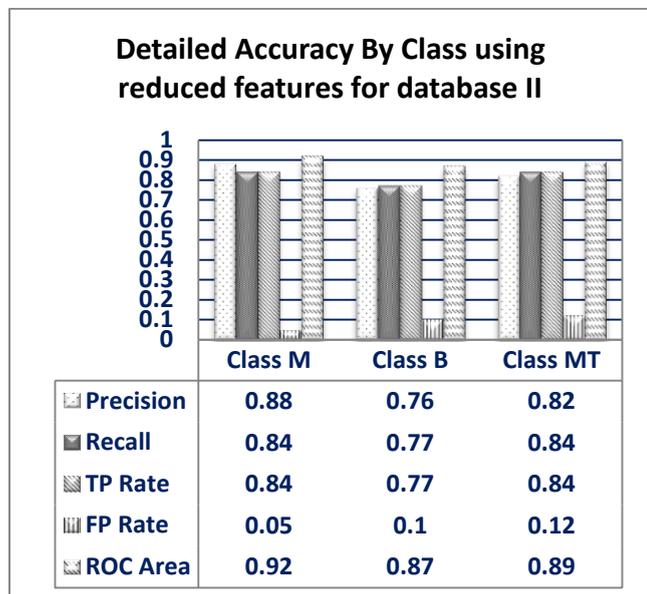


Fig 5: Detailed accuracy by class using reduced features for DB II

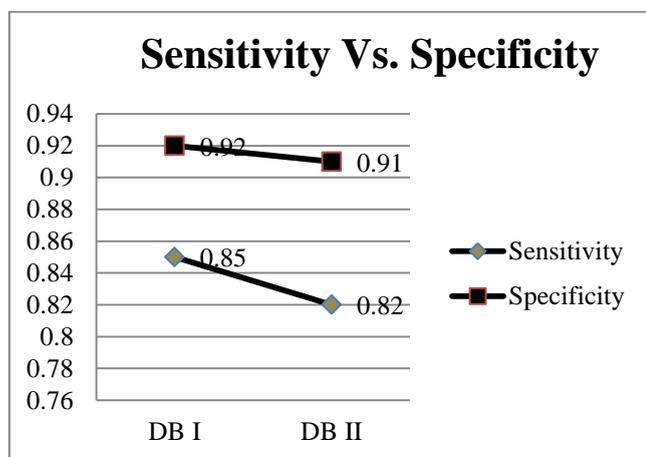


Fig 6: Sensitivity vs. Specificity comparison of DB I and DB II

In Figure 4 and 5, it can be seen that true positive rate as well as receiver operating characteristics (ROC) area, which is the curve between sensitivity and one minus specificity, is much larger as compared to false positive rate. But ROC area for DB I especially for class B and MT is higher than DB II which indicates that database of nodules with maximum area provides better classification than database of nodules with minimum area.

Figure 6, indicates that DB I is having higher sensitivity as well as specificity which clearly indicates that this study not only detects the disease as well as avoids unnecessary biopsies which is clearly the major objectives of all the physicians.

VI. CONCLUSION AND DISCUSSION

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. This is clearly indicated on our study. Moreover, the target of the paper is to declare that the size of the nodule really matters a lot for these types of tests. Larger nodule always contains more content as well as context which can really help in the detection and diagnosis of lung cancer. LIDC data is ambiguous in the sense that it contains redundant information about nodules as marked by four different radiologists. Every physician has his own view about the nodule's characteristics. Subsequently, in this paper, characteristics of both the maximum area as well as minimum area marked nodules are compared and concluded that larger nodule is far better in classifying the nodules. In this work, patient-wise diagnosis report is considered however there is a scope that nodule-wise diagnosis will be assembled and considered for more efficiency.

REFERENCES

- [1] 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.: The Lung Image Database Consortium (LIDC) and Image Database Resources Initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*; 38: 915–931, 2011.
- [2] K Doi; Current status and future potential of computer aided diagnosis in medical imaging, *British Journal of Radiology*, 78: S3-S19, 2005
- [3] L E Dodd, R F Wagner, S G Armato, et al. Assessment methodologies and statistical issues for computer-aided diagnosis of lung nodules in computed tomography: contemporary research topics relevant to the lung image database consortium, *Academic Radiology*, 11: 462-475, 2004.
- [4] Q Li, F Li, J Shiraishi, S Katsuragawa, et al. Investigation of new psychophysical measures for evaluation of similar images on thoracic computed tomography for distinction between benign and malignant nodules, *Medical Physics*, 30(10): 2584–2593, 2003.
- [5] 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, Clarke LP, et al.: The Lung Image Database Consortium (LIDC): Developing a resource for the medical imaging research community. *Radiology*; 232:739–748, 2004.
- [6] PreetiAggarwal, RenuVig, SonaliBhadoria, C.G.Dethe: Article: Role of Segmentation in Medical Imaging: A Comparative Study. *International Journal of Computer Applications*; 29(1):54-61, 2011.
- [7] Robert M. Haralick, K. Shanmugam, Its'hakDinstein: Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics Smc*; 3(6): 610-621, 1973.
- [8] FahimIrfanAlam, RokanUddinFaruqui: Optimized Calculations of Haralick Texture Features. *European Journal of Scientific Research*; 50(4):543-553, 2011.
- [9] Andrzyśiak T, Choras M: Image retrieval based on hierarchical Gabor filters. *International Journal of Applied Computer Science*; 15(4): 471-480, 2005.
- [10] Vibha S. Vyas and PritiRege: Automated Texture Analysis with Gabor filter. *GVIP Journal*; 6, Issue 1: 35-41, 2006.
- [11] Raicu DS, Varutbangkul E, Furst JD, Armato SG III: Modeling semantics from image data: Opportunities from LIDC. *International Journal of Biomedical Engineering and Technology*; 3: 83–113, 2010.

- [12] J. K. Kim and H. W. Park: Statistical textural features for detection of micro calcifications in digitized mammograms: IEEE Trans. Med. Imag; 18: 231-238, 1999.
- [13] OkyDwiNurhayati, Thomas Sri Widodo, AdhiSusanto, Maesadjitjokronagoro: First Order Statistical Feature for Breast Cancer Detection Using Thermal Images: World Academy of Science, Engineering and Technology; 46:424-26, 2010.
- [14] Pachghare, V.K.; Kulkarni, P., "Pattern based network security using decision trees and support vector machine," Electronics Computer Technology (ICECT), 2011 3rd International Conference on , vol.5, no., pp.254-257, 8-10 April 2011.

APPENDIX A

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

Feature Extraction Method	Feature No.	Feature Name in Database
Haralick Features [7] [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 [9] [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 [11]	46~56	area, xcenter, ycenter, perimeter, convexarea, solidity, extent, eccent, equidia, majoraxislen, minoraxislen, circularity, volume, perimeterequidia, sphericity, circularity2, roundness, compactness, concavity
GLDM Feature [12]	74~81	gldm1, gldm2, gldm3, gldm4, gldm5, gldm6, gldm7, gldm8
Intensity Features [11]	57~64	65~73 minint, maxint, meanint, sdint, minintBG, maxintBG, meanintBG, sdintBG, intdiff
First Order Statistics [13]	82~84	skew, kurt, stdd