International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015
ISSN 2091-2730

A SUPERVISED LUNG NODULE CLASSIFICATION METHOD USING PATCH BASED CONTEXT ANALYSIS IN LDCT IMAGE

V.G Aswathy[1], Mrs.T.Johncy Rani[2]

[1]PG Scholar, [2]Asistant professor, ME, department of ECE, Arunachala college of Engineering for women (Anna University Chennai), Tamilnadu, India

aswathyvg24@gmail.com

ABSTRACT- In our daily life, cancer is well-known disease that causes of death in both men and women and understand about the survival rate of lung cancer which is extremely poor. To increase this survival rate of cancerous patient, it is primarily to detect at premature stage which enables many new options for the cancer treatment without risk. This project deals with one of the efficient method to classify four types of lung nodules. That is well-circumscribed, juxta pleural, pleural tail and vascularised. This proposed method was based on patch based contextual analysis by combining the lung nodule and surrounding anatomical structures in LDCT image and it has three main stages: an adaptive patch-based division is used to construct nodule patch and context patch; then, a new feature set is designed to identify the intensity, texture, and gradient information, and then a contextual latent semantic analysis-based classifier and the SVM classifier are designed to calculate the probabilistic estimations for the relevant images. Our proposed method was evaluated on a publicly available dataset and clearly demonstrated promising classification performance.

KEYWORDS- Patch, LDCT, context, latent schematic classifier, SLIC, superpixel, fs3

INTRODUCTION

The lung is an important organ that performs multiple functions every second of our lives. Lung cancer is a disease characterized by uncontrolled cell growth in tissues of the lung and it is the major cause of cancer related deaths in human worldwide. Approximately 20% of the lung nodule represents lung cancer [1]. Hence it was important to identify whether it is malignant or non malignant [2]. Lung nodules are small masses in the human lung, small structures that are roughly spherical. These structures are called pulmonary nodules; however, they can be distorted by surrounding anatomical structures such as vessels and the adjacent pleura [3]. Pleura form an envelope between the lungs and chest wall. Lung nodules are divided into different types according to their virtual positions. At present, the classification from Diciotti et al. is the most popular approach and it divides nodules into four types: well-circumscribed (W) with the nodule located centrally in the lung without any connection to vasculature; vascularized (V) with the nodule located centrally in the lung but closely connected to adjacent vessels; juxtapleural (J) with a large portion of the nodule connected to the pleural surface; and pleural-tail (P) with the nodule near the pleural surface connected by a thin tail.

The difficulty of early detection for this disease is a main reason why lung cancer has the highest mortality rate. Like most cancers, survival rate depends on how early cancer is detected. Unfortunately, it is a long and difficult process for the physician to detect the presence of this disease. One of the most important and difficult tasks the radiologist has to carry out consists of the revealing and diagnosis of cancerous lung nodules from chest radiographs. Some of these lesions may not be detected due to the fact that they may be invisible by the underlying anatomical structure, or the low-quality of the images or one-sided and variable decision criterion used by radiologists. Computed tomography (CT) Offers higher resolution and faster acquisition times. In current clinical practice, however, interpretation of CT images is challenging for radiologists due to the large number of cases. Where the radiologists fail to diagnose small lung nodules in as many as 30% of positive cases. In recent research, digital image processing techniques have been used in developing CAD systems for locating suspected nodules [4], but too many false-positive (FP) classifications/chest radiograph are made.

RELATED WORK

Data Mining and Image processing plays very crucial role in healthcare industry especially for disease diagnosis. Data Mining is very beneficial for finding hidden information or pattern form the huge databases, some widely used data mining techniques are classification, prediction, association analysis, pattern matching and clustering. Image Processing plays significant role in cancer detection when input data is in the form of images; some techniques used in Image Processing for information retrieval are Image acquisition, Noise Removal, Segmentation, and Morphological operations etc.

Farag[5] et al present a feature based extraction to classify lung nodules in low-dose CT slices (LDCT) into four categories: juxta, well-circumscribed, vascularized and pleural-tail, based on the extracted information. The Scale Invariant Feature Transform (SIFT) and an adaptation to Daugman's Iris Recognition algorithm are used for analysis. The SIFT descriptor results are anticipated to lower-dimensional subspaces using PCA and LDA. Iris Recognition algorithm publicized improvements from the original Daugman binary iris code. But here the larger nodule database cannot be generated and it only purposeful on identifies the nodules located in the intersections among dissimilar types.

Abdullah et al. [2012][6] stated that the segmentation of the lung region due to the curb regarding on the similarities of the intensity in the X-ray image. As for lung cancer nodule detection process, it does not seem to be the problem because of the deficient of the similar intensity due to the lung segmentation done. It can be used in the lung cancer application, the system can also be used in the application such as the detection and classification of breast tumour in mammography images a propos on the higher discrepancy of intensity present.

Another alternative would be making the methods themselves publicly accessible. The increased availability of open-source software (such as the Image Segmentation and Registration Toolkit (ITK) is an encouraging movement in that direction. Okada *et al. [7]*presented an automated method to approximate solid nodules as well as ground glass opacities by ellipsoids using anisotropic Gaussian fitting. The volume of the nodule was estimated by the volume of the ellipsoid.

M.F.McNittGray 1999[8] reported on some of initial studies in the classification problem were the patients with a solitary nodule were imaged using high decree computed tomography. Quantitative channel of texture were extracted from these images using co-occurrence matrices. These matrices were twisted with different combinations of gray level quantization, distance between pixels and angles. The derived measures were input to a linear discriminant classifier to predict the classification (benign or malignant) of each nodule. We suggest, however, that improved performance could be obtained by better feature design and a more advanced classifier.

METHEDOLOGIES

In this work, propose a new image classification method for lung patches, based on an LSA classifier. Scrutiny of primary lung tumours and nodules is significant for lung cancer staging, and a computerized system that can perceive both types of abnormalities and it will be helpful for clinical routine.

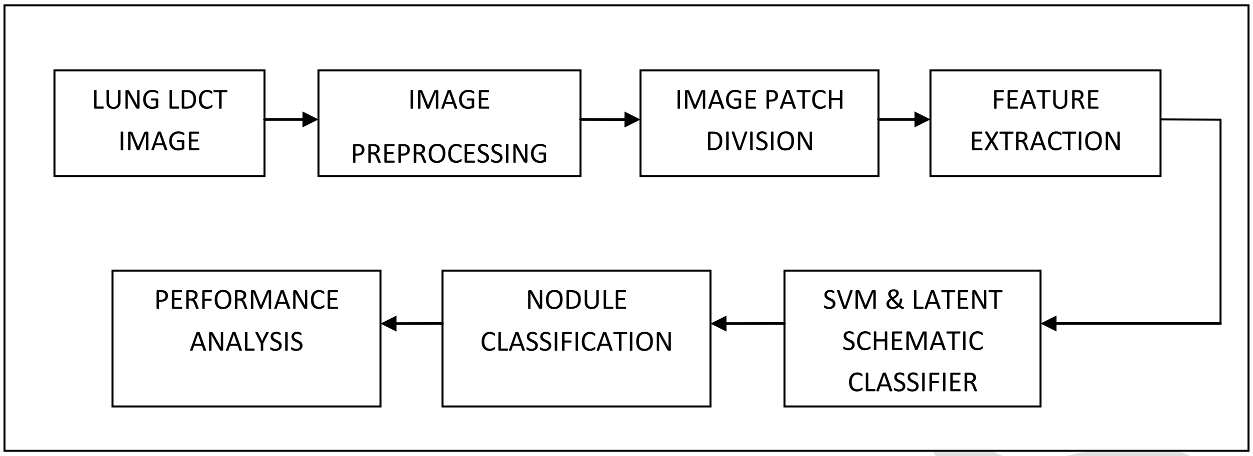


Fig 1: Lung nodule classification systems

LUNG NODULE DETECTION

Nodule detection is an image processing problem. The task is to find positions (and shape) of specific pathological structures in the lungs called nodules. A nodule is a small, round lesion in the lungs, or worm-shaped lesion connected to pleura (the lung boundary) with radio density greater than lung parenchyma. Lung nodules detection in complicated, Nodules in LDCT images show up have relatively low contrast white circular shape and it also overlap with shadows, vessels and ribs.

INPUT LUNG LOW DOSE (LD) CT IMAGE

The lung LDCT images having low noise when compared to scan image and MRI image[9]. So we can take the LDCT images for detecting the lungs. The main pro of the computer tomography image having better clarity, low noise and distortion. The mean and Variance can be easily calculated. The calculated value is very closer to the original value. LDCT is highly effective spot tiny lung nodules. It is also a primary of clinicians for early detection of lung cancer. The LDCT gathers a complete 3D volume of a human thorax in a single breath-hold and it provides very high spatial, temporal and contrast resolution of anatomic structures. The National Lung Screening Trial (NLST) research in 2011 shown, the LDCT screening can avoid more than 8000 lung cancer deaths per year.

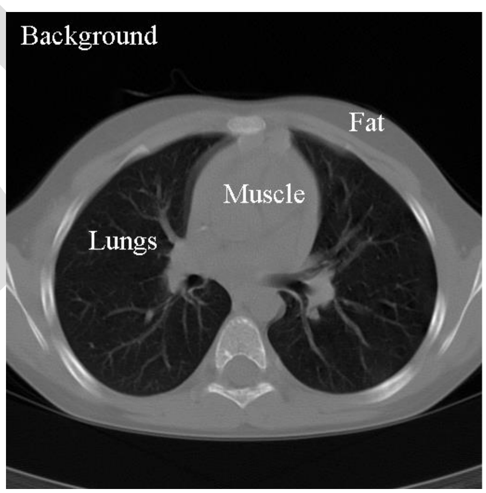


Fig 2: Input LDCT image

PREPROCESSING

Preprocessing is the initial step for detecting the lung cancer. In preprocessing step we have done two steps [10]. They are:

1. Denoising

2. Wiener Filter

3. Iterative segmentation

Denoising: Image denoising algorithms may be the mostly used in image processing. Many methods, despite of execution, share the same basic idea noise lessening through image blurring. Blurring can be done in the vicinity, as in the Gaussian smoothing model or in anisotropic filtering by calculating the variations of an image. White noise is one of the most common problems in image processing. The LDCT image is a gray scale image that contains noises such as white noise, salt and pepper noise etc. This can be removed by using wiener filter from the extracted lung image.

Noise removal y=im2single(y1),where y1 is the input image

Wiener Filter: The goal of the Wiener filter is to filter out noise that has corrupted a signal. It is based on a statistical approach. Typical filters are designed for a desired frequency response. However, the design of the Wiener filter takes a different approach. One is assumed to have knowledge of the spectral properties of the original signal and the noise, and one seeks the linear time-invariant filter whose output would come as close to the original signal as possible. Wiener filters are characterized by the following:

1. Assumption: signal and (additive) noise are stationary linear stochastic processes with known spectral characteristics or known autocorrelation and cross-correlation

2. Requirement: the filter must be physically realizable /causal (this requirement can be dropped, resulting in a non-causal solution)

fy=wiener2(y1,[3 3]) this equation represents filtering of the lung LDCT image from noise

IMAGE PATCH DIVISION

The current approaches are usually based on patches with fixed size and shape[11], such as dividing the images into square patches or into circular sectors based on radial partitions with predefined number of pixels in these areas. In our system we have to form the patches according to local anatomical structure and pixel values and based on super pixel formulation and SLIC using an improved quick shift clustering method.

Simple linear iterative clustering super pixels

Due to small size of lung nodule, quick shift clustering method applied only to amplified image. In our proposed method the image is first amplifier with nearest neighbour interpolation with local intensity information. Amplify the image twice or thrice according to input image size

fy2=imresize(fy,[30 30])
fy=imresize(fy,[64 64])
fy=imresize(fy,[128 128])
fy=imresize(fy,[256 256])

Two parameters are introduced in quick shift 1) *kernelsize*: size of kernel used to estimate density 2) *maxdist*: maximum distance between points in feature space. The maximum distance should set small multiple of the kernel size. In our system we have to choose kernel size as 2 and maxdist as 20.

[Iseg labels map gaps E] = vl_quickseg(fy, ratio, kernelsize, maxdist);

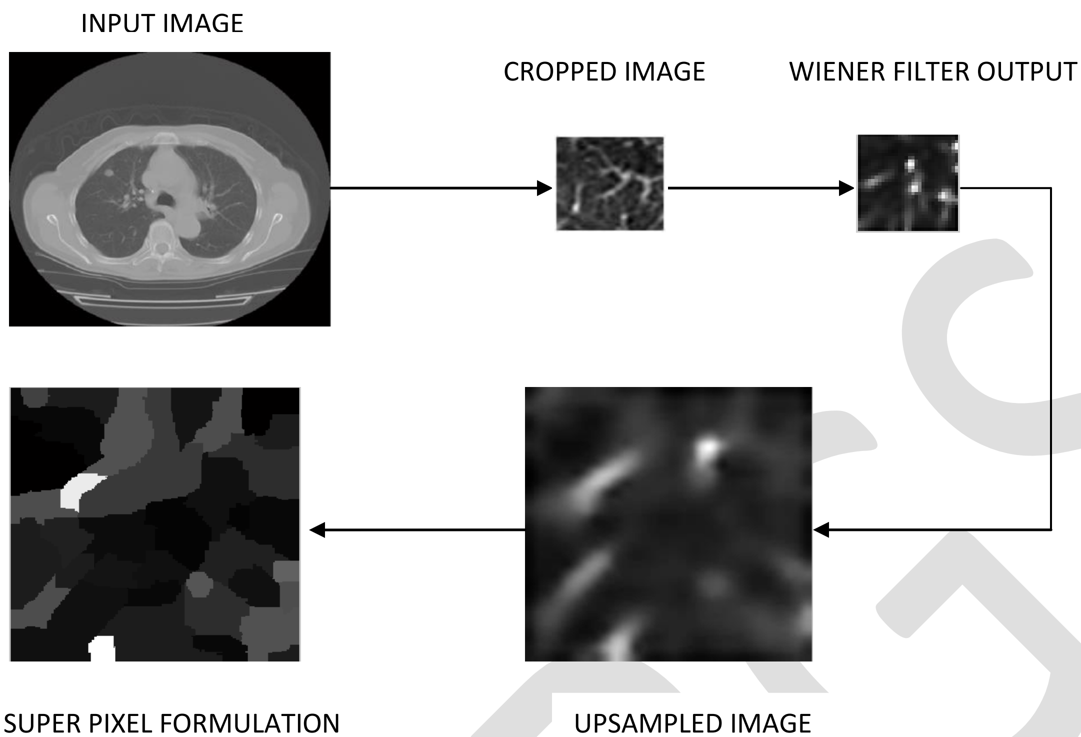


Fig 3: Superpixel formulation

LEVEL PARTITION OF PATCHES

Next we have to divide the image patches into multiple concentric levels. The patch contains the nodule segment and the patch with surrounding structure. The patch that contains the nodule centroid is nodule patch. In our method we construct two concentric levels based on distance between each patch with nodule patch.

FEATURE EXTRACTION

The features extraction is especially effective by using algorithms and applying methods to detect and separate several shapes from an LDCT image. The concern features to be extracted by domain-specific knowledge using image processing tools in MATLAB.

Features, characteristics of the objects of interest, if selected carefully are representative of the maximum relevant information that the image has to offer for a complete characterization an image. Feature extraction methodologies analyze objects and images to extract the most prominent features that are representative of the various classes of objects. Features are used as inputs to classifiers that assign them to the class that they represent. Based on our visual analysis the lung nodules, we suggest that intensity, texture, and gradient can characterize the various nodules and the diverse contextual structures. In this Work SIFT[12], HOG[13] and MR8+ LBP[14] (Local Binary Pattern) features are extracted. From each of the patch determine fs3 feature(texture, intensity and gradient).

SIFT(scale invariant feature transform)

Sift, the overall descriptor determines texture, gradient and intensity features from each of the patches i.e., from nodule patches and contextual patches. It generates a 128 length vector near the centroid of each patch. SIFT (pao) -calculated by selecting one key point near the centroid. It describe the local features of image.

Function varargout=siftvarargin)

SIFT is invariant to image rotation and scaling. But partially invariant to change in illumination .it was highly distinctive.

MR8 + Local Binary Pattern (LBP)

The combination of MR8+LBP used here for attainment better off texture description of patches. Maximum Response 8 (MR8) filter bank which is composed by 38 filters. The MR8 bank contains an edge filter at 3 scales, and a bar filter at the same 3 scales and we use only 8 filter response.

The local binary pattern (LBP) feature has emerged as a silver lining in the field of texture classification and retrieval. Ojala et al. proposed LBPs, which are converted to a rotational in-variant version for texture classification. Various extensions of the LBP, such as LBP variance with global matching, dominant LBPs, completed LBPs, joint distribution of local patterns with Gaussian mixtures, etc., are proposed for rotational invariant texture classification. The LBP operator on facial expression analysis and recognition is successful. Xi Li et al.[15] pro-posed a multistage heat-kernel-based face representation as heat kernels is known to perform well in characterizing the topological structural information of face appearance. Furthermore, the LBP descriptor is incorporated into multiscale heat-kernel face representation for the purpose of capturing texture information of the face appearance.

The Lung image is divided into several regions from which the LBP feature distributions are extracted and concatenated into an enhanced feature vector to be used as a image feature descriptor.

HOG

Histogram of oriented gradients are feature descriptors, has a set of feature vectors. Each feature vector is computed from a block placed across the source image. Each element of a vector is a histogram of gradient orientations.

The algorithm of finding the HOG descriptors [16] consists of the following steps

- Compute gradients for each pixel of an image
- Perform binning of gradients orientation
- Collect the histogram within a call of pixels
- Weight the histogram by blocks and cells for local normalization of the contrasts
- Normalize the histogram

Assuming that the center of patch p_{ao} is c_{pao} , we built 8 coordinate systems that share the same origin c_{pao} but have different initial orientations (0 degree). Two of them are shown with (x0, y0) and (x1, y1). Contra-rotating the first coordinate system (i.e., (x0, y0)) by 45 degree generates the next one (i.e., (x1, y1)). Instead of predefining the initial orientation of the first coordinate system, we set it as the direction from the centroid of the patch to the centroid of lung nodule. Next, for each coordinate system, patch pao is divided into 9 cells, within which gradient orientations of the pixels in 9 undirected histograms are counted to encode the gradient distribution.

```
HOG = vl_hog(im2single(fy2), CELLSIZE) ;
HOG1=[];
HOG1=[HOG1 hist(HOG(:,:,1:30))];
```

NODULE CLASSIFICATION

For any classification problem, a given image feature is considered to be good only if it has enough information to distinguish classes. A single feature by itself is insufficient for classification; several features are used by various classification algorithms. Sequence of stages progressed from image preprocessing, image cropping and image patch division and to end with the image classification. The final result is to determine the normality or abnormality of an image. To predict the probability of lung cancer presence; two approaches are SVM and latent schematic classifier [17], which are based on data powerfully related to lung anatomy and lung CT imaging. It involves SVM for lung nodule patches and pLSA analysis for context patches, Estimates probability by applying k means clustering strategy

$$\mathscr{P}_{\text{level-nodule}}(t_{pt}|\mathrm{I}) = \mathrm{P}_{\text{SVM}}(t_{pt}|\mathrm{I})$$

$$P(tpt|\mathrm{I}) = \lambda * \mathrm{P}_{\text{level-nodule}}(tpt|\mathrm{I}) + (1-\lambda) * \mathrm{P}_{\text{level- context}}(tpt|\mathrm{I})$$

Where λ **E (0, 1)**

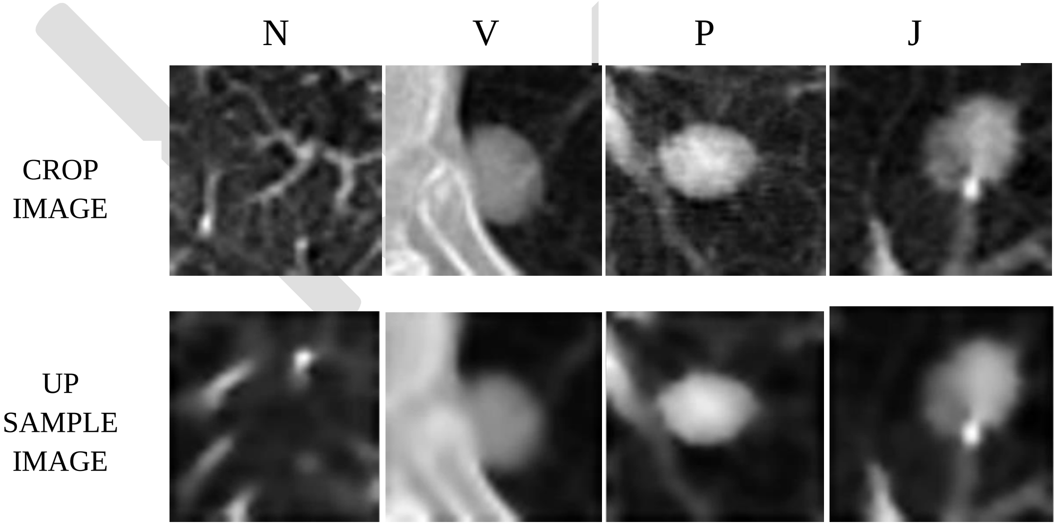
$\mathrm{P}_{\text{SVM}}(tpt|\mathrm{I})$- probability estimate from SVM

NODULE TYPE	PROBABILITY
Vascularized	**0.10-0.19**
Juxta pleural	**0.20-0.25**
Well-circumscribed	**0.25-0.30**
Pleural tail	**0.30-0.35**

Table 1:probability estimations

PERFORMANCE ANALYSIS AND RESULT

Patch based context analysis method is developed for diagnosis and classification of candidate nodules after applying training and testing process. The lung tumour diagnosis is an important criterion in medical field. In this project, we detect and segment the tumour area from the lung LDCT image. The segmented lung tumour can be classified using SVM and LSA classifier. Then the lung tumours are classified as benign or malignant. The performance analysis is carried out in terms of sensitivity, specificity, positive predictive value, negative predictive value and Accuracy. We used publically available early lung cancer action program ELCAP) data base for experiments. The ELCAP database contains 50 sets of low-dose CT lung scans with 379 unduplicated lung nodules annotated at the centroid. The average accuracy achieved is 89% for malignant tumour region in accordance with ground truth images.



ISSN 2091-2730

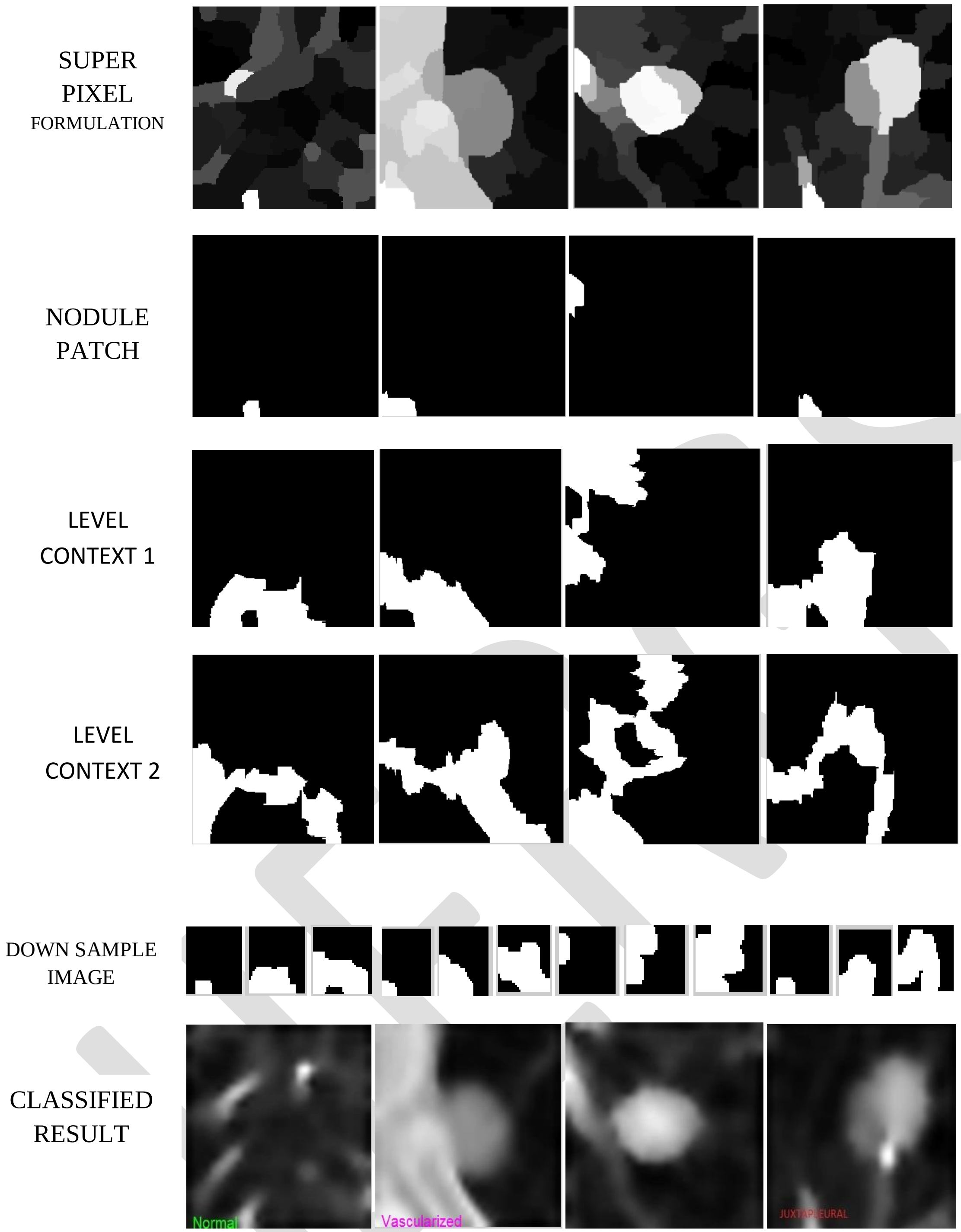


Fig 4: classified result

PERFORMANCE CHARECTERISTICS

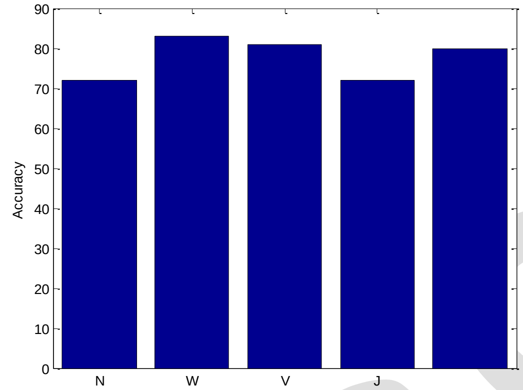


Fig 5: performance characteristics of lung nodule classification system

ACKNOWLEDGEMENTS

The authors wish to thank Dr.T.Krishna Swamy chairman of Arunachala College of engineering for women for stimulating discussions, Dr.S.Joseph Jawahar, principal of Arunachala College of engineering, for his support and encouragement and Mr.E.P.kannan HOD of ECE department for his valuable technical supports.

CONCLUSIONS AND FUTURE RESEARCH

Lung cancer is the most risky and prevailing in the human race according to the stage of detection of the lung cancer nodules in the LDCT images. The process of discovery of disease plays a very vital and essential task to avoid crucial stages and to condense its percentage spreading in the humanity. To achieve further perfect outcomes, the three stages are covered by Image patch division, Features Extraction and nodule classification. Finally, a supervised classifier was designed through combining level nodule probability and level context probability, using image processing tools in MATLAB software.

Future directions are geared towards improving the accuracy rate by removing the SVM and Plsa classifier with decision tree classifier and use the t-test for feature selection. This will improve the efficiency rate and accuracy.

REFERENCE:

[1] J. J. Erasmus, J. E. Connolly, H. P. McAdams, and V. L. Roggli, "Solitary pulmonary nodules: Part I. morphologic evaluation for differentiation of benign and malignant lesions," Radiographics, vol. 20, no. 1, pp. 43–58, 2000.

[2] D. Wu, L. Lu, J. Bi, Y. Shinagawa, K. Boyer, A. Krishnan, andM. Salganicoff, "Stratified learning of local anatomical context for lung nodules in CT images," in Proc. CVPR, 2010, pp. 2791–2798.

[3] A. A. Farag, "A variational approach for small-size lung nodule segmentation," in Proc. ISBI, 2013, pp. 81–84.

[4] A. Farag, S. Elhabian, J. Graham, A. Farag, and R. Falk, "Toward precise pulmonary nodule descriptors for nodule type classification," in MICCAI LNCS, vol. 13, no. 3, 2010, pp. 626–633.

[5] A. Farag, A. Ali, J. Graham, S. Elshazly, and R. Falk, "Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose CT scans of the chest," in Proc. ISBI, 2011, pp. 169–172.

[6] Azian Azamimi Abdullah, Syamimi Mardiah Shaharum, "Lung Cancer Cell Classification Method using Artificial Neural Network", Information Engineering Letters, ISSN:21604114, Volume 2, Number 1, March 2012.

[7] Penedo, M.G., Carreira, M.J., Mosquera, A. and Cabello, D., "Computer-Aided Diagnosis: A Neural-Network-Based Approach to Lung Nodule Detection", IEEE Transactions on Medical Imaging, Pp: 872 – 880, 1998.

[8]M.F.McNittGray , N.Wyckoff,J.W.Sayre,J.G.Goldin,D.R.Aberle,"the effects of co-occurance matrixbased texture parameters on the classification of solitary pulmonary nodules imaged on computed tomography"original research article computerized medical imaging and graphics, vilume 23, issue 6, December 1999

[9] Fan Zhang, Yang Song, Weidong Cai, Min-Zhao Lee, Yun Zhou, Heng Huang, Shimin Shan, Michael J Fulham, and Dagan Feng "Lung Nodule Classification with Multi-Level Patch-based Context Analysis" VOL. 61, NO. 4, APRIL 2014

[10]T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," Machine learning, vol. 42, no. 1-2, pp. 177–196, 2001.

[11] J. W. Hwang and H. S. Lee, "Adaptive image interpolation based on local gradient features," IEEE Signal Processing Letters, vol. 11, no. 3, pp. 359–362, 2004.

[12] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," in Proc. ACM MM, 2010, pp. 1469–1472.

[13] Y. Xu, S. Huang, H. Ji, and C. Fermller, "Scale-space texture description on SIFT-like textons," Computer Vision and Image Understanding, vol. 116, no. 9, pp. 999–1013, 2012.

[14] D. G. Lowe, "Object recognition from local scale-invariant features," in Proc. ICCV, vol. 2, 1999, pp. 1150–1157.

[15] Y. Song, W. Cai, and D. Feng, "Hierarchical spatial matching for medical image retrieval," in Proc. ACM Multimedia 2011 Workshop on MMAR, 2011, pp. 1–6.

[16] M. Varma and A. Zisserman, "Classifying images of materials: Achieving viewpoint and illumination independence," in Proc. ECCV, 2002, pp. 255–271.

[17] Y. Song, W. Cai, Y. Zhou, and D. Feng, "Feature-based image patch approximation for lung tissue classification," IEEE Trans. Medical Imaging, vol. 32, no. 4, pp. 797–808, 2013