(IJRST) 2016, Vol. No. 6, Issue No. III, Jul-Sep **e-ISSN: 2249-0604, p-ISSN: 2454-180X**

LUNG NODULE DETECTION USING HYBRID CLASSIFIER

***R.Poornima,**V.Sivasankaran**
**PG Scholar, Communication Systems,*
Arunai college of Engineering, Tiruvannamalai, Tamilnadu
***Assistant Professor, Department of ECE,*
Arunai college of Engineering, Tiruvannamalai, Tamilnadu

ABSTRACT

An abnormality in lung nodule leads to lung cancer which demands an early detection of lung nodule. The study reveals that automatic detection technique of lung nodule with convincing results and increases the speed of analysis. Since the nodules are attached to blood vessels, detection of lung nodule is a challenging task. To deal with this issue MR8 (Maximum response 8) filter bank based approach is used before preprocessing and eight maximum responses are obtained. From that response texture, intensity and gradient features are extracted using LBP (Local Binary Pattern), HOG (Histogram of Oriented Gradient), SIFT (scale invariant feature transform) descriptor respectively. Further the performances of features are analyzed by hybrid classifier. Hybrid classifier approach is to embed SVM (Support Vector Machine) with ID3 (Iterative Dichotomiser).

Keywords- lung nodule, feature extraction, SVM+ID3 classifier

1. INTRODUCTION

Lung cancer is one of the most injurious forms of cancer, which is the leading cause of cancer death in many regions of the world. Early detection of lung nodule is essential in reducing life victim. Detection of lung cancer at early stage is not an easy task. Survival rates in lung cancer vary significantly by stage; overall, less than 15% of newly diagnosed patients will survive for 5 years. If patients are diagnosed at the earliest stage, survival rates approach 70% [1]. Cigarette smoking is the most critical reason for lung cancer other factors such as environment pollution, certain chemicals like asbestos, silica, and diesel exhaust cause lung cancer and excessive alcohol may also be consign to lung cancer.

The lung nodule detection scheme consists of four steps. They are preprocessing, segmentation, feature extraction and classification. In an existing system, an overlapping nodule identification procedure is designed to help the classification, but this task mainly focused on identifying the nodules located in the intersections among different types. K means clustering and manual analysis method is used for segmentation where the results are not accurate and consumes more time. In short time it is not possible to detect multiple images for cancer detection. Medical images contain a noise which can lead to inaccuracies classification. Extract features from those maximum response filters at eight directions to cover an entire image. Plot histogram for each filter output and concatenate into single histogram. By

concatenating accuracy level of extracting features are increased. Extracted features are stored as a trained set. For classification hybrid classifier is used which embeds support vector machine with iterative dichotomiser. Extracted features are compared with the trained set of images to detect whether the nodule is present or not and analyze the performance of classifier. Datasets are collected from publicly available early lung cancer action program (ELCAP) [2] for experiments. The ELCAP database contains 50 sets of low-dose CT lung scans with 379 unduplicated lung nodules expound at the centroid.

2. RELATED WORKS

While many studies reported that the detection and classification of lung nodules. [3] Pulmonary nodules are detected using template-matching approaches; smaller nodules only appear in one slice where larger nodules appears in continuous slices. Since many false positives (FPs) were detected by template-matching methods. [4] embeds the high dimensional multiview features borrowed from multimodal neuroimaging data into a low dimensional feature space and conserve the optimal local embeddings among different views. [5] used local density algorithm for detection where radiation is high, failed to detect non solid nodules and low sensitivity. [6] discussed about nodule with conventional imaging techniques and differentiate benign from malignant nodules. [7] discussed about Computerized scheme using a snake algorithm, grayscale intensity distribution fed to classifiers for classification. Segment the affected field by placing a reference points around the target lung field and polygon is created based on these points. Feature descriptors like scale invariant feature transform, local binary pattern and speeded up robust features are used in [8] at the output of detection. Nodules are classified using a k-NN (k-Nearest neighbour) algorithm with Euclidean distance. [9] discussed benign and malignant nodules using different SVM. Radiologists can miss up to 30% of lung nodules in chest radiographs as a result of background structure of the lungs which hide the nodules. [10] stated that the lung regions are segmented because of the deficient of the similar intensity as a result of lung segmentation done.

3. SYSTEM MODEL

The functioning of preprocessing is to enhance the image in ways that increase the chances for success of other processes by denoising or by filtering. Feature extraction in image processing is a technique of redefining a large set of unnecessary data into a set of features of reduced dimension. Transforming the input data into the set of features is called feature extraction. Based on our visual analysis features are intensity, texture, and gradient [11]. SIFT descriptor depicting the overall intensity, texture, and gradient information. MR8+LBP descriptor representing a texture feature incorporating MR8 filters before calculating LBP histograms. A Multiorientation HOG descriptor, describing the gradients and accommodating rotation variance in a multi coordinate system.

The combination of MR8 filters and LBP feature is designed to provide richer texture description of patches by incorporating multiscale and rotation-invariant properties. LBP is a powerful feature for texture based image classification[12]. Local binary pattern is a texture

operator which labels the pixel of an image by thresholding and neighborhood of each pixel. Therefore, we incorporate the MR filter set before computing LBP histogram. HOG is being widely used and can also improve performance considerably when coupled with LBP[12][13]. However, unlike SIFT and MR8+LBP descriptors, the raw HOG descriptor cannot handle rotation-invariant problems. Therefore, designed a multi orientation HOG descriptor to provide further an advanced gradient description in addition to that from SIFT. The Scale Invariant Feature Transform process generates a 128-length vector for each key point. Since SIFT is invariant to image translation, scaling, rotation and illumination changes. SIFT is robust and efficiency is close to real time performance.

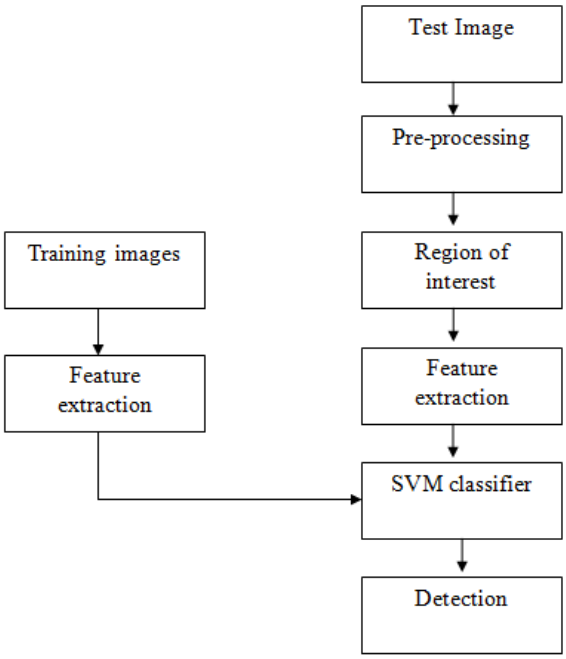

Fig 1. Flow diagram

Hybrid classification is a combination of Support Vector Machine (SVM) with Iterative Dichotomiser3(ID3). Support vector machine is a classifier defined by separating hyperplanes that resolve data and recognize patterns for classification [14]. SVM classifies data points by finding the hyperplane that separates all data points of one class from those of other class. Best hyper plane is the one that represents the largest separation or margin between the two classes. Margin states the maximal width of the slab parallel to hyperplane that has no interior data points. ID3 classifier consists of rules for classifying data using attributes. Tree consists of decision node and leaf node. A decision node has two or more branches, each representing values of tested attributes. A leaf node attribute produces a homogenous result which doesn't required additional classification testing. Information gain is used to select the most useful attribution for classification. Split nodes with head node as A. For each value of A, create new child node or subset. Split training sets into child node or subsets. If subset is pure, stop the splitting process else split the remaining child nodes. Advantages of using ID3 is whole data set is searched to create tree and only needs to test enough attributes until all data are classified.

4. RESULT

CT scan image are in the file format of DICOM (Digital Imaging and Communications in Medicine).

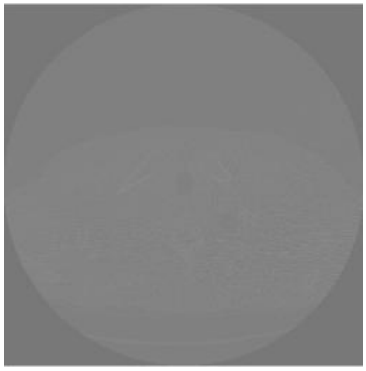

Fig 2. Input image

Fig 3. Filter bank preprocessed image

Apply Maximum response filter to an input image in order to extract wanted region from lung field. It obtains 38 filter output and chooses maximum response of 8 filter output. Here with eight filter responses from input image gives 8×36-dimension histograms.

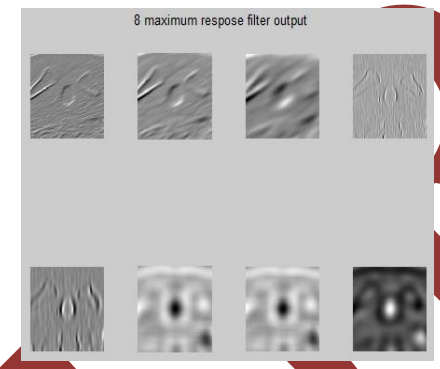

Fig 4. Maximum response filter output

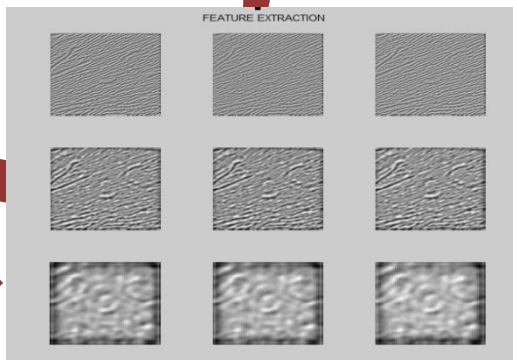

Fig 5. Feature Extraction

For each output of maximum response extract features at eight directions so that it covers an entire image. Totally 64 images of features are extracted and compute histogram for each feature output. Concatenate each histogram into single histogram. All histograms are connected to obtain the final MR8+LBP descriptor.

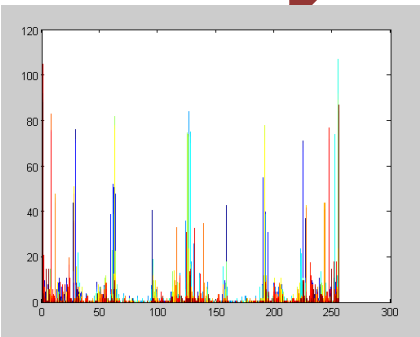

Fig 6. LBP+HOG Histogram

Numerical values of histogram are given as input to classifier section where hybrid classifier is used. Iterative Dichotomiser (ID3) classifier filters unwanted features and increases accuracy and support vector machine classifies the presence of lung nodule.

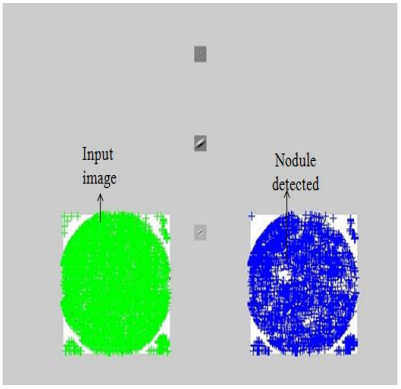

Fig 7. Nodule detection

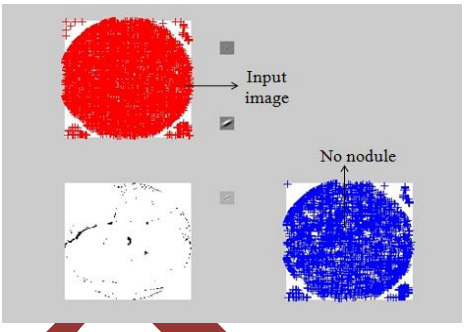

Fig 8. Nodule detection failure

5. PERFORMANCE MEASURES

Sensitivity defines true positive values and specificity defines true negative values. (TP)True positive values indicate correctly identified abnormal nodes. (TN) True negative value indicates correctly rejected as normal node. (FP) False positive value indicates incorrectly identified as abnormal node. (FN) False negative value indicates incorrectly rejected as normal node. These values are composed by confusion matrix. From these values sensitivity, specificity and accuracy can be measured.

$$sensitivity = \frac{TP}{TP+FN} = 97.3\% \quad (1)$$

$$specificity = \frac{TN}{TN+FP} = 92\% \quad (2)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = 96.7\% \quad (3)$$

6. CONCLUSION

The performance analysis of fully automated CT scan lung nodule classification system with maximum feature description based on different range of texture and gradient informations. Filtering is done in preprocessing stage and obtained an eight maximum response of filters. Texture, gradient and intensity features are extracted at eight directions using Local Binary Pattern, Histogram of Gradient, Scalar Invariant Feature Transform. Using Support Vector Machine along with ID3 classification is done. The hybrid classifier achieved an average accuracy of 96.7% .

REFERENCES

[1] N.R. Wardwell, P.P Massion, Novel strategies for the early detention and prevention of lung cancer. Seminars In Oncology,vol.3 ,2005. pp. 259–268.

[2] "ELCAP public lung image database," [Online]. Available:http://www.via.cornell.edu/databases/lungdb.html

[3] Yongbum Lee, Takeshi Hara, Hiroshi Fujita, Shigeki Itoh, and Takeo Ishigaki,"Automated detection of pulmonary nodules in Helical CT images based on an improved template matching technique", IEEE transactions on medical imaging, vol. 20, no. 7, july 2001

[4] sidong liu, lelin zhang, weidong cai, yang song, zhiyong wang, Lingfeng wen, david dagan feng," A supervised multiview spectral embedding method for Neuroimaging classification", IEEE 978-1-4799-2341-0/13/$31.00 ©2013

[5] Binsheng zhao, Gordon gamsu, michelle S.Ginsberg,"Automated Automated detection of small lung nodules on CT utilizing a local density maximum algorithm", Journal of applied clinical medical physics, volume 4, number 3, summer 2003.

[6] 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.

[7] Hong Tan & U. Rajendra Acharya & Collin Tan,"Computer-Assisted Diagnosis Of Tuberculosis: A First Order Statistical Approach To Chest Radiograph," Journal Of Medical Systems · July 2011 Impact Factor: 2.21 · DOI: 10.1007/s10916-011-9751-9

[8] 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. Int. Symp. Biomed. Imag., 2011, pp. 169–172.

[9] M.Gomathi, Dr.P.Thangaraj, "An Effective Classification Of Benign And Malignant Nodules Using Support Vector Machine," Journal of Global Research in Computer Science , Volume 3, No. 7, July 2012

[10] Fan Zhang,Min-Zhao Lee,Heng Huang,Shimin Shan, "Lung Nodule Classification With Multilevel Patch-Based Context Analysis," IEEE transactions on biomedical engineering,vol.61,No.4,APRIL 2014

[11] Yonghong Shi, Feihu Qi, Zhong Xue, Liya Chen, Kyoko Ito, Hidenori Matsuo, and Dinggang She "Segmenting Lung Fields In Serial Chest Radiographs Using Both Population-Based And Patient-Specific Shape Statistics ," IEEE Transactions On Medical Imaging, Vol. 27, No. 4, April 2008

[12] Fan Zhan, Yang Song, Min-Zhao Lee, Yun Zhou, Heng Huang, Shimin Shan, Michael J Fulham, and Dagan D. Feng,," Lung Nodule Classification With Multilevel Patch-Based Context Analysis", IEEE Transactions On Biomedical Engineering, Vol. 61, No. 4, April 2014

[13] Qiang Zhu, Shai Avidan, Mei-Chen Yeh, and Kwang-Ting Cheng," Fast Human Detection Using a Cascade of Histograms of Oriented Gradients", IEEE 0-7695-2646-2/06 $20.00 (c) 2006

[14] Bhavanishankar .K and Dr. M.V.Sudhamani," techniques for detection of solitary Pulmonary nodules in human lung and Their classifications -a survey", International Journal on Cybernetics & Informatics (IJCI) Vol. 4, No. 1, February 2015