www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 5 May 2017, Page No. 21336-21342
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i5.25

SVM classifier based CAD system for Lung Cancer Detection

Apoorva Mahale[1], Chetan Rawool[2], Dinesh Tolani[3], Deepesh Bathija[4], Prof. Kajal Jewani[5]
[1,2,3,4]Dept. of Computer Engineering, VESIT, Mumbai, India
[5]Asst. Prof., Dept. of Computer Engineering, VESIT, Mumbai, India

Abstract

This paper discusses the formulation of a Computer Aided Detection (CAD) system for Lung cancer detection by using an interdisciplinary approach based on the techniques of Image Processing and Machine Learning. This paper is an extension of image processing using lung cancer detection and produces the results of feature extraction and feature selection after segmentation. Here the proposed model is developed using SVM algorithm for feature selection and classification. The system accepts Lung CT(Computed Tomography) images as input. This present work proposes a method to detect the cancerous cells effectively from the CT scan and images. Modified Fuzzy Possibilistic C Means (MFPCM) has been used for segmentation and Gabor filter has been used for De-noising the medical images. Simulation results are obtained for the cancer detection system using the MATLAB software.

Keywords: **CAD, CT, Gabor filter, MATLAB, MFPCM, SVM**

I. INTRODUCTION

Statistics have it that in the United States, lung cancer strikes 225,000 people every year, accounting for $12 billion in health care costs. Early detection is critical in providing patients with the best shot at survival and recovery.

Lung cancer is considered to be the major cause of cancer deaths worldwide, symptoms appear only at advanced stages causing the mortality rate to be the largest in all other types of cancer, due to this detection is difficult in it's early stages. More people die of lung cancer than any other types of cancer such as: breast, colon, and prostate cancers. There is prominent proof signifying that the early detection of lung cancer will decrease the mortality rate. The recent estimates provided by the World Health Organization shows that around 7.6 million deaths occur worldwide due to lung cancer per year. Moreover, mortality due to cancer is supposed to continue rising, to become around 17 million worldwide in 2030.

The early detection of lung cancer in the primary stage is a challenging problem, because of complicated structure of the cancer cells, where most of the cells are overlapping with each other. It is a computational procedure that classifies the images into groups based on similarities. The manual analysis of samples is very time consuming, inaccurate and requires well trained person to avoid diagnostic errors. The quantitative procedure is very helpful for earlier detection of lung cancer. Experimental analysis will be made with dataset to assess the performance of the various categories of SVM classifiers. The performance is based on effective classification by the classifier. The main aim of our proposed CAD systems is to increase the accuracy and decrease the time of diagnoses.

The conventional approach is to implement a multistage CAD system capable of revealing the presence or absence of nodules to the radiologist. A critical stage in this system is the detection of ROIs (regions of interest) that could quite possibly be nodules, for the sake of reducing the scope of the problem. Image segmentation is the process of partitioning regions in an image into meaningful segments belonging to different objects. During image analysis, it is very often much more convenient to isolate objects of interest before performing the actual analysis, in the form of either cutouts or highlights; using distinguishable borders. Image segmentation usually precedes other analysis related processes. Since most applications necessitate the segmentation of specific objects of interest, it is quite common for image segmentation techniques to include certain forms of object recognition too. In medical image analysis, medical experts are usually only interested in certain organs visible in the image, which is chiefly why image segmentation is required in almost all medical image related applications.

II. PROPOSED SYSTEM

Broadly, our proposed system (*Fig.1*) has 3 major processes: image pre-processing, feature extraction and finally the classification process. The CADe system accepts the CT scan image of lungs as an input. In the real-world scenario, the CT scan image is expected to contain noise and hence needs processing to facilitate extraction of lung features so that classification may be performed successfully based on these features. The first step of our system is image pre-processing. Image pre-processing includes de-noising i.e. removing unwanted noise from the image. Image features extraction stage plays an integral part in our working in which algorithms and techniques are used to detect and isolate the various desired portions or shapes(features) of an image. Feature extraction is an essential stage that represents the final results to determine the normality or abnormality of an image. These features act as the basis for classification process.

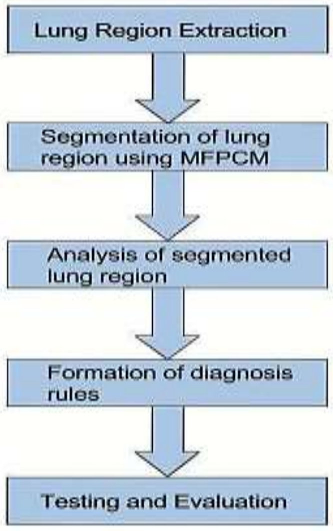

Lung Region Extraction
Segmentation of lung region using MFPCM
Analysis of segmented lung region
Formation of diagnosis rules
Testing and Evaluation

***Fig. 1*: System Block Diagram**

1. Image Acquisition

First step is to acquire the CT scan image of lung cancer patient. The lung CT images are having low noise when compared to X-ray and MRI images; hence they are considered for developing the technique. The main advantage of using computed tomography images is that it affords enhanced clarity with lower distortion. For research work, the CT images are acquired from NIH/NCI Lung Image Database Consortium (LIDC) dataset. DICOM (Digital Imaging and Communications in Medicine) has become a standard for medical Imaging. The acquired images are in raw form. In the acquired images lot of noise is observed. To improve contrast, enhance clarity, remove the background noise, pre-processing of images is indispensable. Hence, various techniques like smoothing and enhancement are used to get image in the required format.

2. Image Pre-processing

Smoothing

It suppresses the noise or other small fluctuations in the image; equivalent to the suppression of high frequencies in the frequency domain. Smoothing also blurs all sharp edges bearing important information about the image. For removing noise from images, median filtering is used. Median filtering is a non-linear operation often used in image processing to reduce salt and pepper noise. In general, the median filter allows a great deal of high spatial frequency detail to pass while remaining very effective at removing noise on images where less than half of the pixels in a smoothing neighborhood have been affected. B=medfilt2(A,[m,n]) performs median filtering of the matrix A in two dimensions. Each output pixel contains the median value in the m x n neighborhood around the corresponding pixel in the image. Medfilt2 pads the image with 0's on the edges, so the median values for points within one -half the width of the neighborhood ([m,n]/2) of the .edges might appear distorted.

Enhancement

Enhancement techniques are used for improving the interpretability or perception of information in image for human viewers, or for providing better input to other automated image processing techniques. Image enhancement can be classified in two main categories, spatial and frequency domain.Bit plane slicing is used here for the purpose of enhancement.

3. Segmentation

The segmentation is performed for determining the cancer nodules in the lung. This phase will help identify the Regions of Interest(ROI) in the lung nodule, that can help identify the cancerous region. Modified Fuzzy Possibilistic C Mean (MFPCM) is used in the proposed technique for segmentation because of better accuracy of MFPCM.

4. Feature extraction

Feature extraction is a crucial step for the CADe system. It uses different methods and algorithms for feature extraction from the segmented image. Based on the extracted features normality and abnormality of the lung are decided. The features extracted include area, perimeter and average intensity. Images on segmentation have only two values; 1 and 0. Nodule part will be represented with value 1. Then area of the nodule can be calculated by finding number of pixel with value 1. Perimeter of the nodule means the number of pixels in the boundary region of the nodule. Average intensity is another feature which is used for the purpose of cancer detection. Choose two threshold values for mean intensity, and then compute the average intensity value for the candidate region. If the average intensity value is between the threshold values then this part is assumed to be cancerous, else non-cancerous. Based on their area, cancerous nodules are identified. A nodule size greater than 25mm is considered an abnormal image, while a nodule size of less than 25mm is taken to be a normal image.

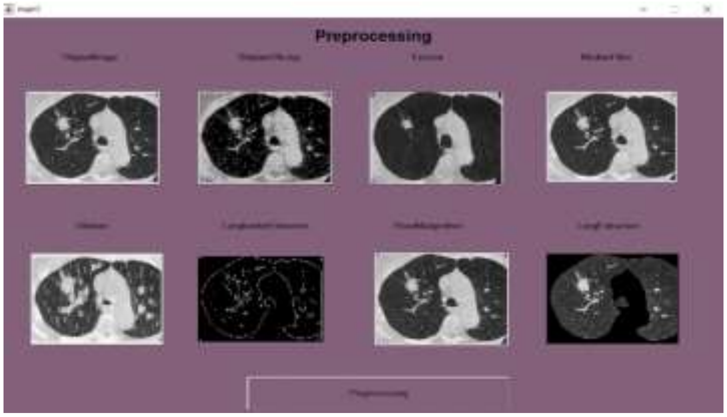
5. Classification

Support vector machines are supervised learning models for analyzing data and recognizing patterns, used for classification. The basic model of SVM takes a set of input data and for each given input, predicts which of two classes forms the input, and hence is a non-probabilistic binary linear classifier.

III. IMPLEMENTATION

main1.m

The first phase of the proposed Computer Aided Diagnosing system (main1.m) is the extraction of lung region from the (CT) scan image. This phase uses the basic image processing methods such as Bit-Plane Slicing, Erosion, Median Filter, Dilation, Outlining, Lung Border Extraction and Flood-Fill algorithms. Usually, the CT chest image not only contains the lung region, it also contains background, Heart, liver and other organs areas. The main aim of this lung region extraction process is to detect the lung region and regions of interest (ROIs) from the CT scan image.



<u>Fig. 2</u>: main1.m

DOI: 10.18535/ijecs/v6i5.25

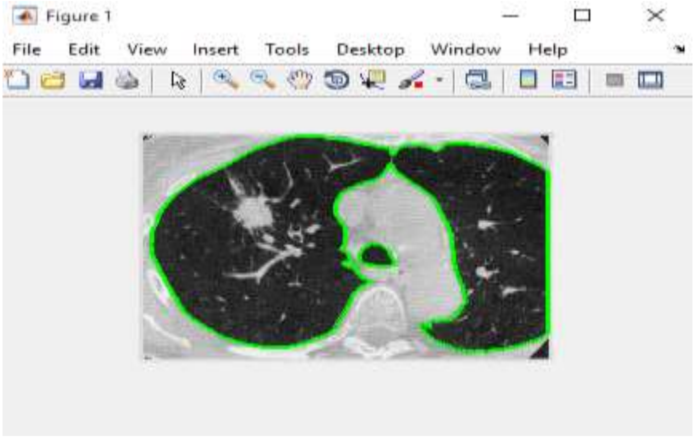


***Fig. 3**: sample input CT scan image*

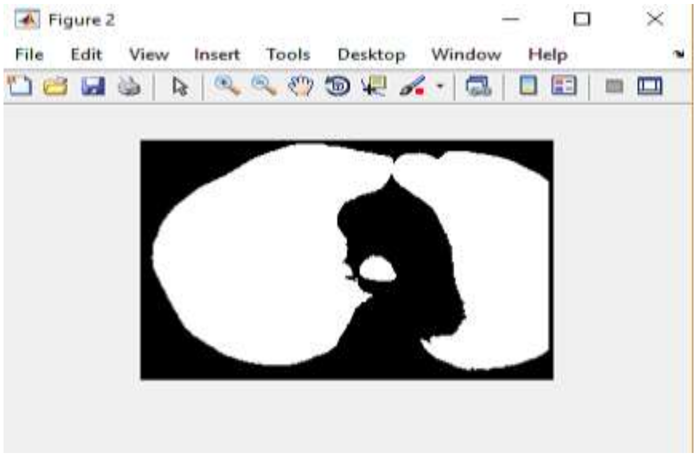


***Fig. 4**: Lung region extracted from sample CT scan image*

main2.m

The second phase of the proposed CAD system is the Segmentation of lung region. The segmentation is performed for determining the cancer nodules in the lung. This phase will help determine the Regions of Interest (ROI) which helps in identifying the cancerous region as shown in below.

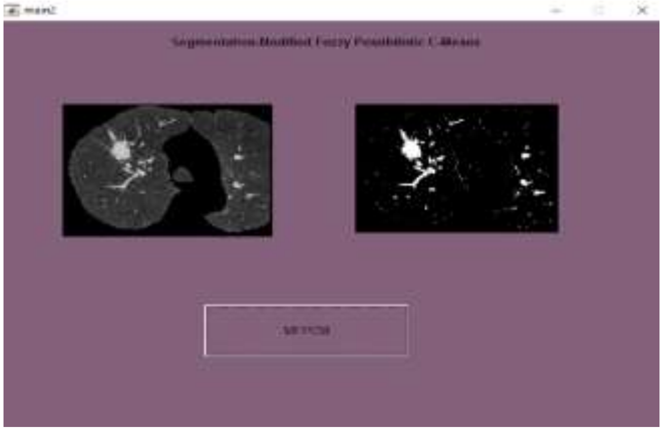


***Fig.5**: main2.m*

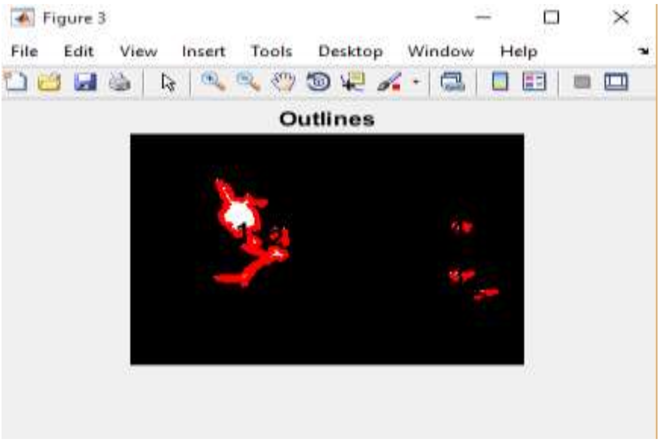


***Fig. 6**: number of cancer nodules in the given sample input CT scan image*

.

main3.m

Once segmentation is performed on the lung region, various features like Gray level co-occurrence matrix and gabor features are extracted from the given input image and the classification of occurrence and nonoccurrence of cancer nodule for the supplied lung image is done using Support Vector Machine.

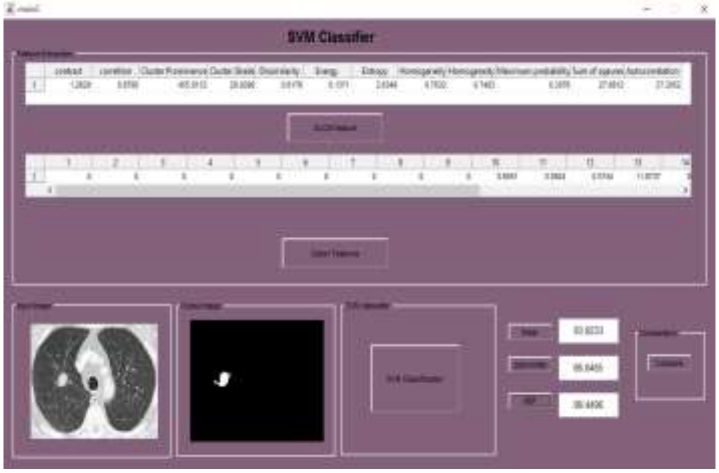
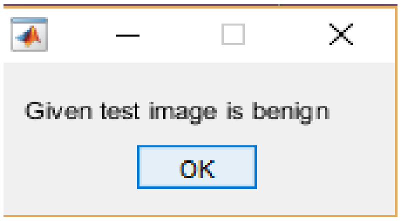


***Fig. 7**: main3.m*

***Fig. 8**: Gabor features extracted from the lung region*

DOI: 10.18535/ijecs/v6i5.25



Given test image is benign

***Fig. 9**: SVM classifier outputs supplied input CT lung image as 'benign'.*

The accuracy results of the proposed CAD system with different SVM kernels is shown below:

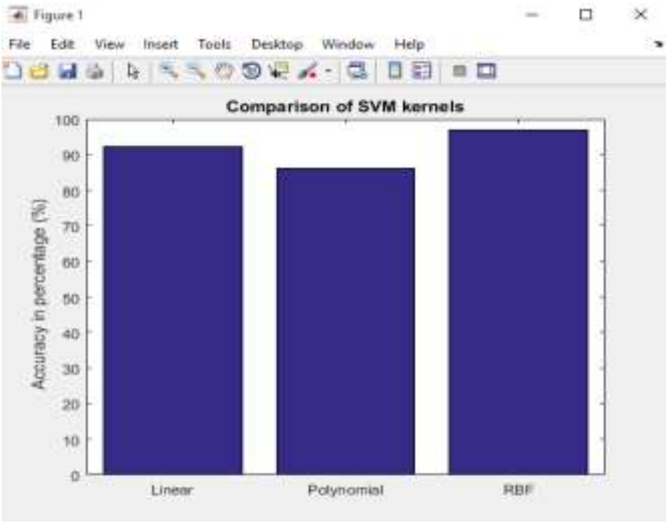


***Fig. 10**: Accuracy Graph*

IV. OBSERVATIONS AND ANALYSIS

The experiments are conducted on the proposed computer-aided diagnosis systems with the help of real time lung images. This experimentation data consists of 30 lung images. Those 30 lung images are passed to the proposed CAD system. The diagnosis rules are then generated from those images and these rules are passed to the classifier for the learning process. After learning, a ung image is passed to the proposed CAD system. Then the proposed system will process through its processing steps and finally it will detect whether the supplied lung image is with cancer or not. we proposed an approach of classification using Support Vector Machine Classifier which has very good working efficiency and produces the accurate results as compare to other classifiers. So that by the SVM classifier we can more accurately and effectively detect the cancer nodule by the analysis of CT images.

Table shows the results obtained by applying different svm kernels to the CT images.

Lung Image	Feature extraction technique	Linear	RBF	Polynomial
1-10 -------- 11-20 -------- 21-30	GLCM	93.023 --------- 91.472 --------- 92.248	98.449 ---------- 96.124 ---------- 96.8992	86.046 ---------- 84.496 ---------- 86.046

Table 1: Accuracy table

The accuracy results of the proposed CAD system with different SVM kernels is shown in Table 1. From the table it is apparent that the accuracy for the usage of radial basis function (RBF) kernel is better as compared to other SVM kernels. As the table indicates, the average accuracy of other kernels for the lung images 1-10 is 89.5349% whereas the usage of RBF yields an accuracy of 98.4496%; which is better than that of the others. For the lung images 11-20, RBF produces an accuracy of 96.124%, whereas, others produce an accuracy of only 89.1473%. The overall accuracy is higher for RBF than for other SVM techniques.

V. CONCLUSION

Early detection of lung cancer in the primary stage is a challenging problem, because of complicated structure of the cancer cells, where most of the cells are overlapping with each other. It is a computational procedure that classifies the images into groups based on similarities. The manual analysis of samples is very time consuming, inaccurate and requires well trained person to avoid diagnostic errors. The quantitative procedure is very helpful for earlier detection of lung cancer.

In many applications, the performance of the machine learning-based systems is comparable to that of experienced radiologists. The application of machine learning may benefit patients either by reducing costs, improving accuracy, or disseminating expertise that is in short supply.

The use of machine learning in radiology is still evolving. As machine learning research progresses, we expect there to be more applications to radiology. Machine learning will be a critical component of advanced software systems for radiology and is likely to have wider and wider application in the near future.

VI. FUTURE SCOPE

Future scope includes enhancing the CAD system to include more features for calculating parameters like sensitivity and specificity. Future work also includes constantly upgrading the accuracy of the classifier by training it on larger and more comprehensive datasets. Furthermore, different machine learning algorithms can be incorporated to compare and understand which technique returns the best results in the context of cancer detection.

VII. ACKNOWLEDGEMENTS

We thank the National Cancer Institute for granting us access to NCI's data NCI's data collected by the National Lung Screening Trial (NLST). We accept complete responsibility for the statements contained herein. NCI does not represent or imply concurrence or endorsement for the findings stated in this paper.

VIII. REFERENCES

[1] Shijun Wang and Ronald M. Summers, " Machine Learning and Radiology " , Published in final edited form as:Med Image Anal. 2012 Jul; 16(5): 933–951.Published online 2012 Feb 23. doi: 10.1016/j.media.2012.02.005.

[2] M.Gomathi, "A Parameter Based Modified Fuzzy Possibilistic C-Means Clustering Algorithm for Lung Image Segmentation", Global Journal of Computer Science and Technology Vol. 10 Issue 4 Ver. 1.0 June 2010.

[3] Swati P. Tidke, Vrishali A. Chakkarwar, "Classification of Lung Tumor Using SVM", International Journal Of Computational Engineering Research (ijceronline.com) Vol. 2 Issue. 5 September 2012

[4] S.Shaik Parveen, C.Kavitha, "Classification of Lung Cancer Nodules using SVM Kernels", International Journal of Computer Applications (0975 – 8887) Volume 95– No.25, June 2014

[5] M.Gomath , Dr.P.Thangaraj, "A Computer Aided Diagnosis System for Detection of Lung Cancer Nodules Using Extreme Learning Machine", International Journal of Engineering Science and Technology,ISSN: 0975-5462,Vol. 2(10), 2010.

[6] Sruthi Ignatious, Robin Joseph, "Computer Aided Lung Cancer Detection System", Global Conference on Communication Technologies (GCCT 2015)

[7] Erkan Emirzade., "A Computer Aided Diagnosis System for Lung Cancer Detection using SVM", a thesis submitted to the Graduate School of Applied Sciences of Near East University, NICOSIA, 2016.

AUTHOR PROFILE

Kajal Jewani is an Assistant Professor at the Department of Computer Science, Vivekanand Education Society's Institute of Technology, Mumbai (India). Her interests include Image Processing and Theory of Computation.

Apoorva Mahale is a final year Computer Science undergraduate student at Vivekanand Education Society's Institute of Technology, Mumbai (India). Her interests include Data Analytics and Web Development.

DOI: 10.18535/ijecs/v6i5.25

Chetan Rawool is a final year Computer Science undergraduate student at Vivekanand Education Society's Institute of Technology, Mumbai (India). His interests include exploring new Open Source tools.

Dinesh Tolani is a final year Computer Science undergraduate student at Vivekanand Education Society's Institute of Technology, Mumbai (India).

Deepesh Bathija is a final year Computer Science undergraduate student at Vivekanand Education Society's Institute of Technology, Mumbai (India).