International Conference on Academic Research in Engineering and Management
IETE, Lodhi Road, Delhi, India (ICAREM-17)
30th April 2017, www.conferenceworld.in ISBN: 978-93-86171-43-6

CLASSIFICATION OF CANCEROUS AND NON-CANCEROUS LUNG CANCER NODULES USING IMAGE PROCESSING TECHNIQUES

Moffy Crispin Vas[1], Prof. Amita Dessai[2]

ETC department, Goa College of Engineering (India)

ABSTRACT

*Cancer is the second leading cause of the most number of deaths worldwide, out of which **lung cancer** is the leading cause of deaths among all the cancer types. Computer tomography (**CT**) scan is used by radiologists to detect cancer and its growth, over a period of time. In this Paper, a preprocessing technique known as median filtering is used for noise reduction in lung CT images followed by segmentation for obtaining the lung region of interest (**ROI**) using morphological operation. Textural features are then calculated from gray level co-occurrence matrix (**GLCM**) which is created from the ROI and these features are used an input to an **ANN** classifier for determining whether the lung is cancerous or not.*

Keywords- CT, CLAHE, GLCM, Lung cancer, ROI

I. INTRODUCTION

Lung cancer is the most common cancer in the world and has been the reason for 19% of cancer deaths globally. Most of the deaths are due to the late detection of cancer. A person suffering from lung cancer has an overall 5 years of survival with only 15% assurance in developed countries and 5% in developing countries. Computer tomography scans have been proved useful in detecting lung cancer and hence CT scans have reduced cancer mortality rates by 20% but at the cost of false positive rate of 96%. There are two types of lung cancers namely Non small cell lung cancer which account for 80-85% of all cases and small cell lung cancer which account for 15-20%. Lung cancer has four stages and if the cancer is detected at an early stage (either stage I or II where the cancer is confined to the lungs) the survival rate can increase up to 50-70%. Hence it becomes essential to develop an automated system which will help in early as well as a more accurate detection of lung cancer and thus help in saving more lives.

This paper includes steps namely preprocessing, segmentation, textural feature extraction followed by artificial neural network classification and aims to increase the speed and accuracy and decrease the time involved in cancer detection.

II. RELATED WORKS

Lung cancer detection field is explored by researches to a large extent. Anita chaudhary and Sonit Sukhraj Singh used three image enhancement techniques namely fast Fourier transform, Gabor filter and auto enhancement method. For segmentation, they made use of thresholding and watershed segmentation techniques. With the help of morphological features of lung nodules they classified lung cancer and its stages [1]. Khin Mya Mya Tun used median filter for preprocessing the lung CT images and used Otsus thresholding for segmentation of lungs followed by extraction of geometrical features which were used to train the feed forward artificial neural networks [2]. Md. Badrul Alam Miah applied median filtering to preprocess the lung CT image and segmented out the left and right lung separately using edge maps. They obtained 33 different features and then applied them to a feed forward neural network [3]. S. A. Patil made use of median filter to preprocess the x-ray images. They utilized morphological operations and region growing technique for segmentation. Extracted geometrical features and first order statistical texture features were then applied to ANN for classifying the lung cancer [4].Amjed S.AlFahoum designed an automated intelligent system for nodule detection and classification of lung cancer in CT images[5].Muhammed Anshad and S.S Kumar gives a comparative survey of all the methods used for automated cancer detection systems. Comparisons are made based on advantages, disadvantages and accuracy of the methods [6].K.Punithavathy used PET/CT images and made use of morphological operations for lung segmentation. Second order statistical features were obtained using GLCMs and these features were given as an input to a FCM classifier. This detection system achieved an overall accuracy of 85.69% [7]. Nooshin Hadavi segmented lung CT images using region growing based thresholding algorithm and used size as a feature of lung nodule. These features served as an input for cellular learning automata for training the images [8]. Mohsen Keshani used an active contour for lung segmentation and detected ROIs by stochastic 2D features. To eliminate the segmented bronchus and bronchioles, 3D anatomical features are further used to detect the nodules followed by Active contour modeling [9]. Gawade Prathamesh Pratap segmented PET/CT images by p-tile thresholding and used M type morphology to detect the cancer images with the help of MATLAB [10]. Hence, in this paper, we propose a methodology of using statistical feature extraction and artificial neural network classification for automated detection of lung cancer.

III. METHODOLOGY

The methods used in this study are shown below in Fig.1. Each step shown in the Fig.1. is discussed in detail in the later sections.

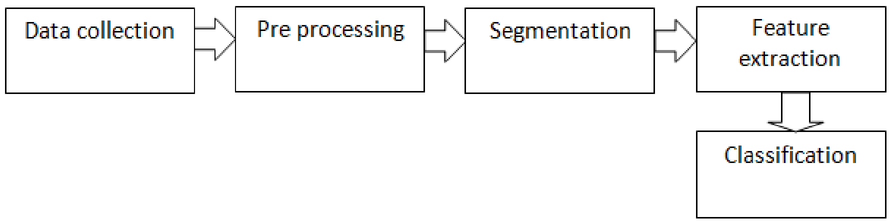


Fig.1. The flowchart of the proposed methodology

3.1 Data collection

The images have been downloaded from The Cancer Imaging Archive (TCIA) database. The images are stored in the DICOM format.

3.2 Preprocessing

CT images are mostly prone to salt and pepper noise. The median filters prove to be the best solution in reducing this impulse noise as it also preserves all the edges in the images.

3.3 Segmentation

For segmentation of the lungs, morphological closing operation was used with a disk structural element of size 2. After complementing the closed image, the holes in the lungs were filled and a lung mask was obtained by separating out the lungs using maximum area. Superimposing the lung mask with the image will give perfectly segmented lungs.

3.4 Feature extraction

After obtaining ROI, the image is divided into three resolutions for further processing. In medical images, texture provides better and more detailed information about the images. Hence, textural features can be calculated from the gray level co-occurrence matrix based on the spatial relationship of pixels .GLCM was calculated in four directions i.e. 0^0, 90^0, 135^0 and 270^0. These textural features tell us how much the property of an image such as spatial structure, contrast and roughness have a certain correlation with the desired output.

3.5 Classification

Neural networks have proved to be an efficient and reliable model for the cancer detection system. A feed forward network with back propagation algorithm provides a flexible and accurate classification method for the lung cancer identification. The network used in this paper consists of 84 input nodes, 20 hidden nodes and two output nodes. Another advantage of ANN is that it can handle large amount of data.

IV.RESULTS

The two kinds of images used in this study are shown in Fig.2. A total of 228 images were used and divided into training and testing set.

Fig.2.a. Raw noncancerous image

Fig.2.b. Raw cancerous image

As can be seen from Fig.2 the salt and pepper noise in the images can be easily noticeable. Fig.3 shows the effect of the median filtering by using a neighborhood of 3*3 pixels.


Fig.3 Preprocessed images **Fig.3 clearly indicates that the images have become smoother and these images will be used as an input for segmentation.**

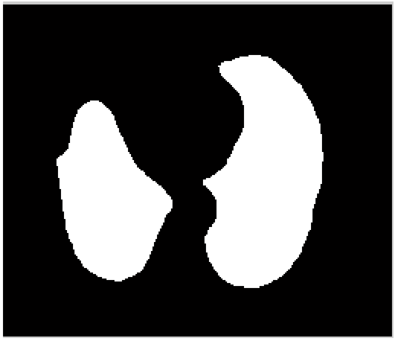
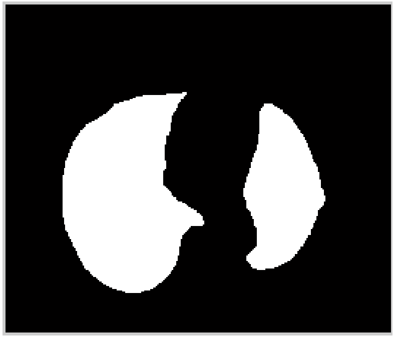



Fig.4. Lung masks

The lung masks as shown in Fig.4 were obtained by performing a closing operation on the preprocessed image with a disc shaped structural element. A disc size of 2 gave best results for accurate segmentation of the lungs.

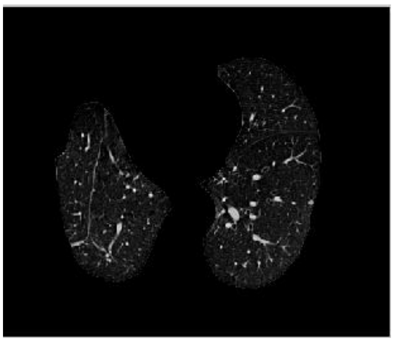
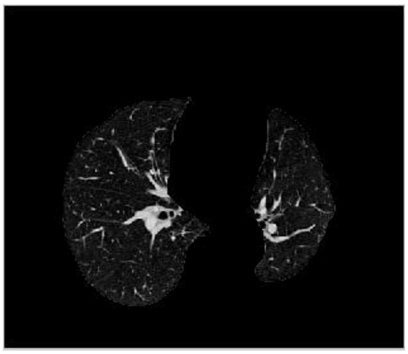



Fig.5. ROI

Fig.5. shows that the lung images have been segmented accurately. The ROI was then divided into overlapping sub images of size 8*8 and seven Haralick features were extracted namely energy, entropy, contrast, homogeneity, maximum probability, cluster prominence and inverse difference moment normalized were obtained from these sub images.

Fig.6. shows the feature extraction plot and the features 1 to 7 are energy, entropy, contrast, homogeneity, maximum probability, cluster prominence and inverse difference moment normalized respectively. As can be seen from the graph, the cancerous image features can be differentiated from the non cancerous ones and hence used for neural classification.

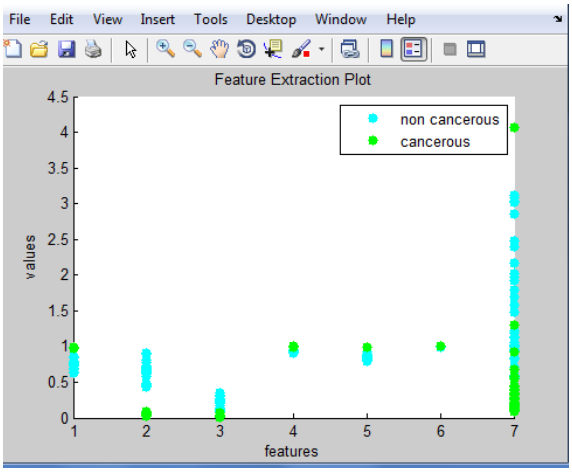
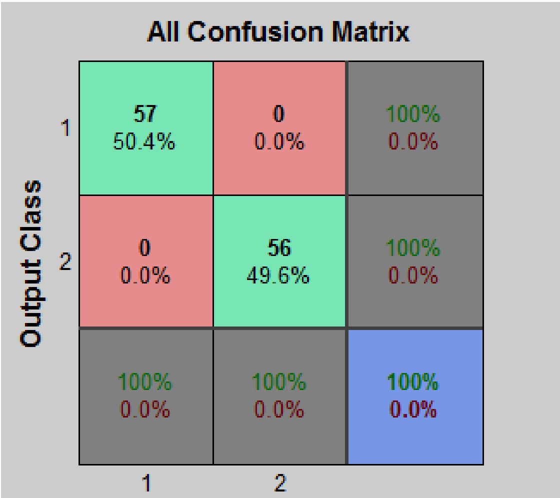


Fig.6.Feature extraction plot



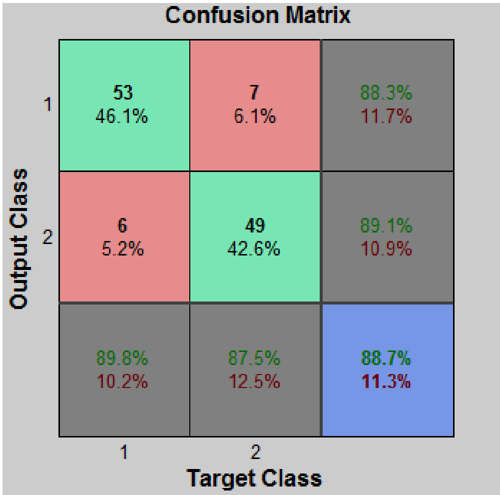


Fig.7.a.training confusion matrix **Fig.7.b.test confusion matrix**

As seen in Fig.7. a total of 215 images used, out of which 113 images were used for training. An accuracy of 100% is achieved for training .For testing 115 images were used, out of which 13 images were misclassified yielding an overall accuracy of 88.7%. The sensitivity thus calculated was found to be 87.5% and specificity was 89.83%.

V. CONCLUSION

This methodology successfully developed an automated lung cancer detection system. Median filtering provided for a severe reduction in image noise and morphological operations led to the accurate segmentation of lungs. The extracted textural features provided a good basis for the neural network classification. The accuracy can be improved by using actual medical images and other classifiers like SVM or neuro fuzzy classifiers.

VI. ACKNOWLEDGEMENT

First of all, I would like to thank my guide Prof. Amita Dessai, for permitting me to carry out my project work under her guidance and for her unconditional and valuable support.

This project has been a great learning and knowledge enrichment source, right from its beginning to the current partial fulfillment. Hence, I take this opportunity to thank Dr.H.G.Virani, HOD of ETC dept. and the other staff members who are always ready to help students with their queries.

I also want to thank Almighty God, my parents and my friends who have always been there, backing and encouraging me to think positive and keep working during instances when this project gave, not so pleasing results.

REFERENCES

[1.] Anita chaudhary and Sonit Sukhraj Singh “Lung cancer detection on CT images using image processing”, computing sciences 2012 international conference, IEEE 2012.

[2.] Khin Mya Mya Tun and Aung Soe Khaing “Feature extraction and classification of lung cancer nodule using image processing techniques” IJERT Vol.3,Issue 3,March 2014

[3.] Badrul Alam Mia and Mohammad Abu Yusuf “Detection of lung cancer from CT image using image processing and neural network” ICEEICT ,IEEE May 2015

[4.] S.A. Patil, M.B. Kuchanur, “Lung Cancer Classification using Image Processing,” International Journal of Engineering and Innovative Technology (IJEIT), 2(3), pp. 37 – 42, 2012.

[5.] AjmedS.Al-Fehoum, Eslam B.Jaber, Mohammed A.Al-Jarrah “Automated detection of lung cancer using statistical and morphological image processing techniques,” Journal of Biomedical Graphics and Computing, 4(2), pp. 33 - 42, 2014.

[6.] Muhammed Anshad and S.S Kumar “Recent methods for the detection of tumor using computer aided diagnosis”, ICCICCT, IEEE transactions 2014

[7.] K.Punithavathy, M.M.Ramya and Sumathi Poobal “Analysis of statistical texture features for automatic lung cancer detection in PET/CT images” 18-20 February, RACE, IEEE 2015

[8.] Nooshin Hadavi, Md Jan Nordin and Ali Shojaeipour “Lung cancer diagnosis using CT-scan images based on cellular learning automata”, ICCOINS, IEEE 2014

[9.] Mohsen Keshani, Zohreh Azimifar and Reza Boostani “Lung nodule segmentation using active contour modelling” MVIP, IEEE 2016

[10.] Gawade Prathamesh Pratap and R.P Chauhan “Detection of lung cancer cells using image processing techniques” ICPEICES, IEEE 2016