

Classification of benign and malignant lung nodules using image processing techniques

Moffy Crispin Vas¹, Amita Dessai²

¹Student, Dept. of ETC, Goa Engineering College, Goa, India

²Assistant Professor, Dept. of ETC, Goa Engineering College, Goa, India

Abstract – Cancer is the second leading cause of most number of deaths worldwide after the heart disease, out of which, lung cancer is the leading cause of deaths among all the cancer types. Hence, the lung cancer issue is of global concern and thus this work deals with detection of malignant lung cancer nodules and tries to distinguish it from the benign nodules by processing the Computer tomography (CT) images with the help of Haar wavelet decomposition, Haralick feature extraction followed by artificial neural networks (ANN)

Key Words: Computer tomography, Lung cancer, malignant, Haralick features, ANN

1. INTRODUCTION

Lung cancer contributes to about 19% of the deaths globally. 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. If the cancerous nodules are detected at an early stage the survival rate can shoot up to 50-60%. 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%. A lung nodule is the white spot that appears on the CT image having a size of about 3 cms. It becomes difficult to conclude visually whether the nodule is malignant or benign. Around 40% of the lung nodules are cancerous and have to be detected as early as possible to degrade the mortality rates. It becomes necessary to develop an automated system that will classify the malignant nodules at an early stage with increased accuracy and speed.

2. RELATED WORKS

Khin Mya Mya Tun extracted geometrical features from the CT images and classified the images using feed forward artificial neural networks [1]. S.A.Patil and used x-ray images and preprocessed them with median filters. Segmentation techniques like region growing and morphological operations were used and they also utilized geometrical and first order statistical texture features for classifying the cancer using ANNs [2]. Anita Chaudhary and used three different image enhancement techniques, out of which they reported that Gabor filter gave them best results. Thresholding, watershed segmentation techniques and extracted features such as area, roundness and eccentricity of lung nodules were used for classifying the lung cancer and its stages [3]. Md. Badrul Alam Miah segmented out the left and right lung separately using edge maps. By extracting 33 different features they classified the images using feed forward neural network [4]. Muhammed Anshad gives a comparative survey of all the methods used for automated cancer detection systems. Comparisons are made based on accuracy, advantages and disadvantages of the methods [5]. Amjed S.AlFahoum designed an automated intelligent system for nodule detection and classification of lung cancer in CT images. In this work they made use of morphological operations and utilized geometrical features for classification [6]. Gawade Prathamesh Pratap carried out p-tile thresholding and watershed processing on PET/CT images followed by the use M type morphology to display cancer image if any with the help of MATLAB [7]. Mohsen Keshani used an active contour for lung segmentation and detected ROIs by stochastic 2D features. They further used

3D anatomical features to detect the nodules and eliminate bronchioles and some segmented bronchus. Active contour modeling is used for accurate contour extraction of nodules [8]. Nooshin Hadavi used region growing based thresholding algorithm on lung CT images and used size as a feature of lung nodule. These features were used by cellular learning automata for training thus making it possible to detect the lung cancer [9]. K.Punithavathy used PET/CT images and utilized second order statistical features calculated from GLCMs as an input to a FCM classifier. This detection system achieved an overall accuracy of 85.69%. [10]

3. METHODOLOGY

The methodology adopted in this paper is shown in the form of flow chart diagram in Fig-1 and each step in the flow chart is explained in detail in the subsequent sections.

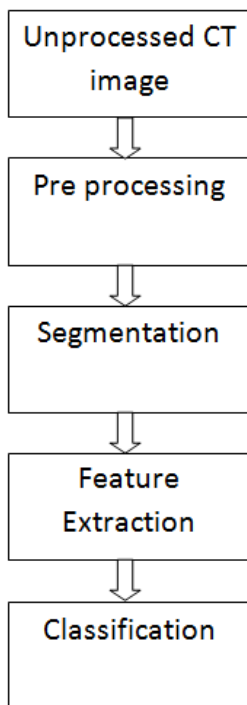


Fig-1: proposed methodology

3.1 IMAGE COLLECTION

The images have been downloaded from The Cancer Imaging Archive (TCIA) database. The images are stored in the DICOM format. Medical images were also obtained from the V. M. Salgaocar hospital (SMRC).

3.2 PRE PROCESSING

The CT images are mostly affected by salt and pepper noise. Median filter has been found to be quite effective for eliminating this impulse noise while preserving the edges. The pattern of neighbours is called the window, which slides pixel by pixel over the entire image. A window of size 3*3 was used in this paper and was found to be successful in the intended use of deleting the salt and pepper noise.

3.3 SEGMENTATION

Segmentation is the process of separating out the region of interest, which is processed further. In this study, morphological operations are used, as these operations account for the form and structure of the image and help in removing the imperfections which arise as a result of distortion by noise and texture, while converting the images to binary. The method used for obtaining the lung region of interest is shown in the form of a schematic diagram in Fig-2.

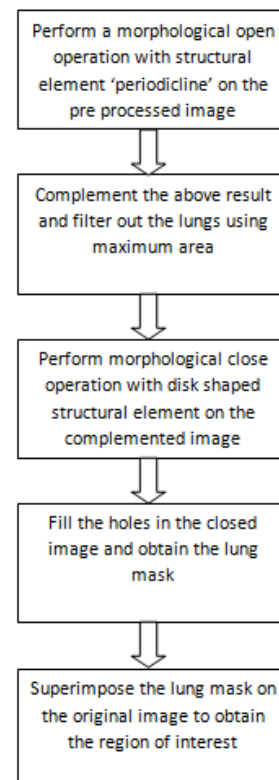


Fig-2: steps involved in segmentation of lungs

3.4 FEATURE EXTRACTION

Since the malignant and the benign CT images are not visually distinguishable, the image is first divided into three different resolutions, Haar wavelet decomposition is applied to each image resolution as it conserves the energy of the signal while compressing it in a compact form. Haar wavelets also have no effects on the edges in the image and are exactly reversible. After using Haar wavelet on the three resolution images only the horizontal, vertical and diagonal decomposed approximations are considered and gray level co-occurrence matrix (GLCMs) are calculated in four directions to extract the Haralick features. Seven textural features namely energy, entropy, contrast, homogeneity, maximum probability, cluster prominence and inverse difference moment normalized were extracted. Textural features were extracted as they provide more detailed information about the medical images.

3.5 CLASSIFICATION

In this study, feed forward neural network was used with 252 inputs, 20 hidden nodes and two output nodes. The back propagation algorithm used, guarantees convergence. Even though ANNs behave as black box systems, they are self-adaptive, flexible and capable of capturing complex and nonlinear characteristics of any physical process with a good accuracy. ANNs can also handle large amounts of datasets easily.

4 RESULTS

The two kinds of images used in this study are shown in Fig-3. A total of 228 downloaded images were used and divided into training and testing set. A total of 170 hospital images were used.



Fig-3.a: An image with Benign nodules



Fig-3.b: An image with malignant nodules

Fig-4 shows the noise reduction effect of median filters with a size of 3*3 pixels.



Fig-4: Preprocessed images



Fig-5 shows the lung mask obtained by morphological close operation with disk structural element of size 2.



Fig-5: Lung masks



Fig-6 shows the effect of super imposing the lung mask on the pre processed image. The region of interest 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.

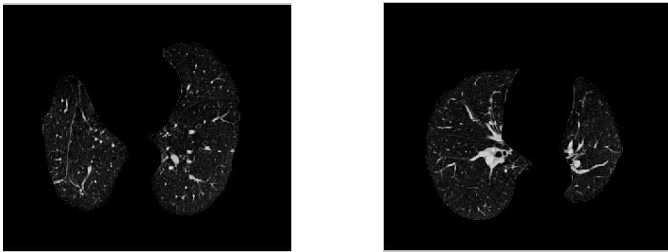


Fig-6: The segmented lungs

Fig-7 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 malignant image features can be differentiated from the benign ones and hence used for neural classification.

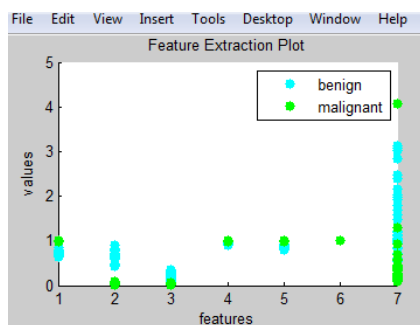


Fig-7: Feature extraction plot

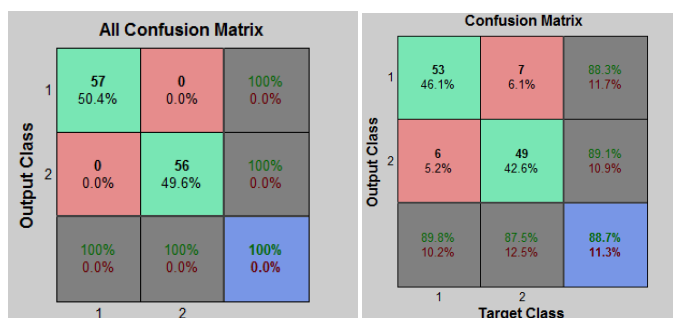


Fig-7: training and testing confusion matrices respectively for downloaded database

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%.

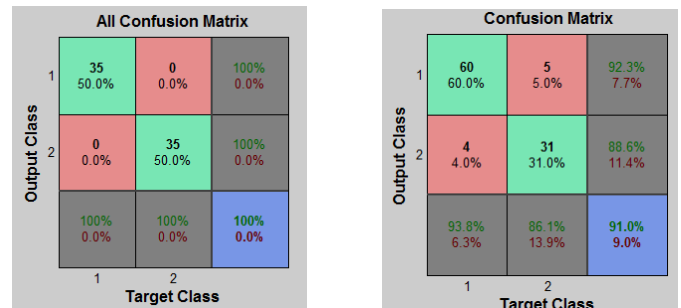


Fig-8: training and testing confusion matrices respectively for hospital images.

For training with 35 benign and 35 malignant images, an accuracy of 100% was achieved and for testing with 64 benign and 36 malignant images, an accuracy of 91% was obtained. Sensitivity of 93.75% and specificity of 86.11% was calculated.

5. 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 other classifiers like SVM or neuro fuzzy classifiers.

ACKNOWLEDGEMENT

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

This project has been a great learning and knowledge enrichment source, right from its beginning to the current state of 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 would also like to thank SMRC hospital, Goa for providing me with the required CT lung images.

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] 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
- 2] 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.
- 3] Anita chaudhary and Sonit Sukhraj Singh "Lung cancer detection on CT images using image processing", computing sciences 2012 international conference, IEEE 2012.
- 4] Badrul Alam Mia and Mohammad Abu Yusuf "Detection of lung cancer from CT image using image processing and neural network" ICEEICT, IEEE May 2015
- 5] Muhammed Anshad and S.S Kumar "Recent methods for the detection of tumor using computer aided diagnosis", ICCICCT, IEEE transactions 2014
- 6] Ajmed S. 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.
- 7] Gawade Prathamesh Pratap and R.P Chauhan "Detection of lung cancer cells using image processing techniques" ICPEICES, IEEE 2016
- 8] Mohsen Keshani, Zohreh Azimifar and Reza Boostani "Lung nodule segmentation using active contour modelling" MVIP, IEEE 2016
- 9] Nooshin Hadavi, Md Jan Nordin and Ali Shojaeipour "Lung cancer diagnosis using CT-scan images based on cellular learning automata", ICCOINS, IEEE 2014
- 10] 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