International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X;pISSN:2321-2241 Volume: 2; Issue: 5; May -2014

Lung Cancer detection in early stage using Modern Clustering Technique, White Pixel Method and Efficient classifier

B.N.RamaChandra[1] and Ramesh Nayak[2]

[1] *Second Year Student, M.Tech, Dept of CSE, Canara Engineering College, Mangalore, India*

[2] *Asst. Professor, Dept of ISE, Canara Engineering College, Mangalore, India*

Email : [1]manubn55@gmail.com and [2]rameshnayak.spi@gmail.com

Abstract : Lung Cancer has been the second most deadliest cause for the death of human beings currently and its dark effect has been prevailed all around the globe. In earlier days manual methods were used to detect the lung nodules which were not very effective and as a result the survival rates were low. So introduction of computer systems would enhance the survival rates to a high significant level. And early detection of lung cancer can even elevate the survival chances of patients up to 70% from 14% which is a very notable rate. In this proposed work the system is divided into 5 different phases which work towards the early, successful detection and classification of lung cancer in patients. The 5 modules being – Enhancement phase, Segmentation phase, White Pixel Counting phase, Feature Extraction phase and the Classification phase. In the Enhancement phase a CT image of the chest is given as the input to the proposed system. Various simple yet effective image enhancement algorithms are used to get a lung region without noise, distortion, blurring etc. Next the enhanced image is given as input to the segmentation phase where a modern segmentation algorithm is applied to get a segmented image with ROI. Next, White Pixel Counting is done with the segmented images as input. Segmented image is analysed to extract various types of features to determine cancerous nodules. Then in the next phase Classification is done for the occurrence and non-occurrence of Cancer based on both the White Pixel count and by SVM Classifier separately. Both the results are compared against each other for their accuracy.Key Words: CT images, Feature extraction, Diagnosis rules, Lung Nodules, Support Vector Machine, Segmentation.

INTRODUCTION

Lung Cancer is the recent trending problem in medical sciences because of its severity in being the cause for high death rate in human beings and also with the pitfalls in its detection. In India, each year nearly 30,000 cases are found with people hit with Lung Cancer. But many a times the survival rate is quite low because the disease is detected in the last stages. There are four stages of lung cancer and if it detected in the earliest only then the survival chances are seen at a good rate. But instead if the lung cancer is detected in its fourth stage then the survival chance is almost nil. So it can be said that there are three deciding factors which can elevate the survival rate of lung cancer hit patients. Early detection, proper detection and proper classification can increase the survival rate of a patient from 14%-70% which is a very good one indeed. But there are certain difficulties in the detection of lung cancer. First of all the CT images of the chest area may contain many other unwanted regions such as the clothes, bed sheets, other bodily parts etc. which are technically termed as artefacts. If suitable algorithms are not used for the image enhancement then in the first stage itself we may lose many useful data in regions where there are chances of lung nodules being present. So a suitable pre-processing need to be done to the CT images to get a clear and enhanced Lung region with noise free, distortion free, blur free images which help in effective detection of lung nodules. Next in order to get the region of interest (ROI) in the lung region we need to go for segmentation of the lung region from the enhanced image. Usually if segmentation is not done properly then there are more chances of losing probable lung nodule areas which will hamper the accuracy as a result. So an effective algorithm for this purpose is chosen and it is the modified version of fuzzy clustering algorithm called Modified Fuzzy Clustering Algorithm. This will efficiently do the segmentation task thus retaining almost all lung nodule regions. After segmentation White pixel counting of the segmented image is done by using Sobel function. If the white Pixel Count is more than a fixed threshold value then the image is regarded as cancerous else it is considered as non-cancerous. Next GLCM and Gabor filters are applied to the segmented image for the purpose of feature extraction. At the last phase, based on the various types of extracted features a training based software called Support Vector Machine classifies these images as either Malignant or Normal.

PROPOSED WORK

Proposed work has five modules. They are as follows –

1. Image Enhancing phase
2. Segmentation phase
3. White Pixel Counting
4. Feature extraction and Analytic Standards
5. Image Classification

Module Description

Module 1 – Image enhancing module

After a patient undergoes CT scanning of his chest we get a CT scanned image which contain many artifacts. Artifacts may be patient's other bodily parts which are in the vicinity of the chest, his clothes, bed spreads etc. Since we are in search of lung nodules the interest is confined to only the lung region. Also CT images will be full of noise, distortion, blurring effects etc. These can be removed in the image enhancement process by applying simple algorithms such as bit-plane slicing, erosion, median-filtering, lung border extraction and flood fill. The output of this stage would be a noise free, blur free, distortion free enhanced lung region.

Module 2 – Segmentation of lung region

An enhanced lung region is the input for this module where segmentation is done with the assistance of an efficient algorithm called Modern Fuzzy Clustering algorithm. Suitable threshold value is fixed and the said algorithm is applied to the enhanced lung region which does the segmentation suitably by ignoring probable areas where lung nodules might not be present and retaining regions which has more probability of the presence of lung nodules. After segmentation is done we get a region of lung where lung nodules are present.

***Module 3*- White Pixel Counting**

Using Sobel operators the white pixel count shall be taken from the segmented images. A threshold value is set for the white pixel count. If the count exceeds the threshold the CT image can be said as cancerous and if does not exceed the set threshold value then it can be concluded that the given CT image is non-cancerous.

Feature Extraction and analytical standards

Feature extraction is done using the GLCM feature extractor where it extracts 12 set of features from the segmented lung region based on which cancerous nodules can be found out. Also Gabor filter is applied to the segmented image so that it helps in efficient classification of CT images.

Image classification

The output of stage 4 would be a set of images where various lung nodules are identified with variety of features. Images are classified as either cancerous or non-cancerous as according to the features extracted of the lung nodules with the assistance of training based software called Support Vector Machine.

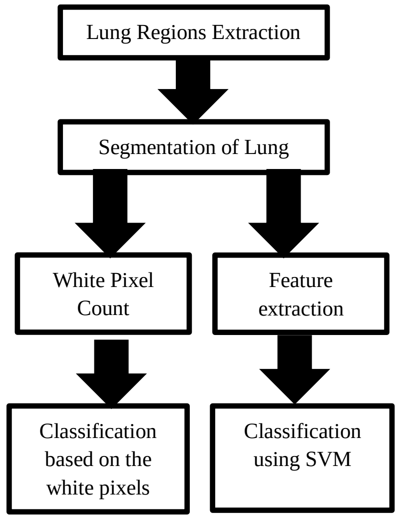

Figure 2 shows the overall architecture of the proposed methodology

RESULTS

Computer Tomography images were collected from the Radiology department of a reputed Cancer hospital. Both cancerous and non-cancerous images were collected where the format of image is in JPEG and all are binary images. These CT images were given as input to the proposed CAD system. Suitable pre-processing was done firstly which removed all noise, blurring effects, distortions etc from the obtained CT images.

Next segmentation was done using an efficient algorithm called Modern Fuzzy Possibility C-Mean algorithm, which yielded a very good segmented image of the given input CT image. Next Sobel function was applied to the segmented image inorder to get the white pixel count of the same. A threshold value was set here and images were classified as accordingly as either cancerous or non-cancerous. Later on the other hand, feature extractions were done from the segmented image. The GLCM and Gabor filters were applied to extract nearly 12 features which are the key factors which decide the presence and growth of lung nodules. These features are fed to a training based software called Support Vector Machine. SVM suitably classified as according to the features extracted whether Normal or Malignant images.

The following experimental results were obtained. Also a comparison has been made between the outputs obtained from the White Pixel

Counting from Sobel function with that of output obtained from the Support Vector Machine.

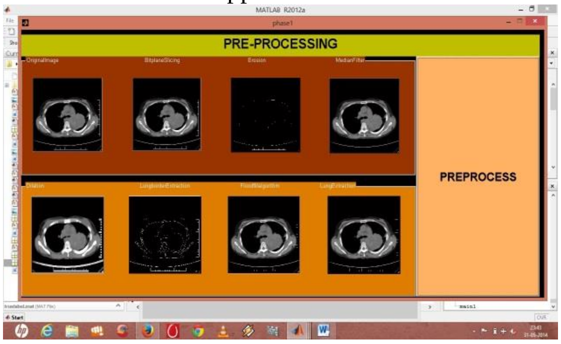

Fig 2 – Phase – 1-Pre-processing of CT Image

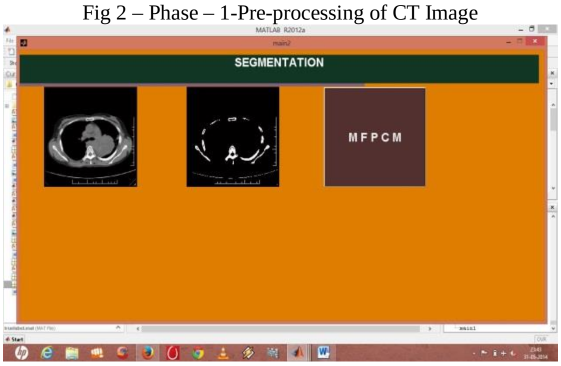

Fig 3 – Phase – 2 - Pre-processed Lung image is Segmented

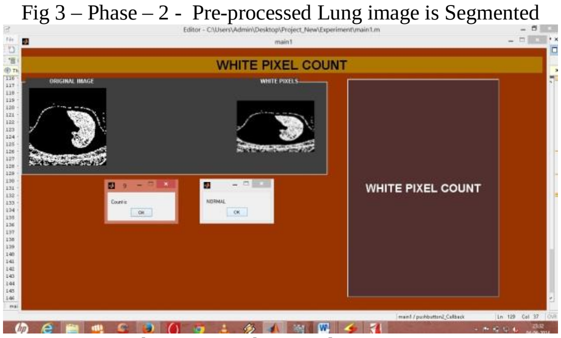

Fig 4– Phase – 3 – White Pixel Count-NORMAL

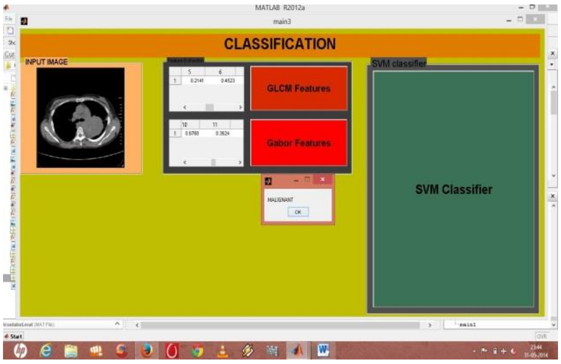

Fig 5 – Last Phase - Image classified as Cancerous by SVM

Comparison between White Pixel Counting and SVM Classification

A threshold of 20 was set for White Pixel Counting which indicates that White Pixel Count below 20 is non-cancerous and those of above 20 are Cancerous. A comparison was made between White Pixel Counting method and the SVM classifier.

Lung Images	Sobel function	SVM Classifier	Accuracy
1-50	88	98.2	SVM
51-100	90	99	SVM
101-150	92	98	SVM
151-200	93	99	SVM

Table – 1 – Comparison between White Pixel Count and SVM Classified Results

CONCLUSION

This work proposes a computer diagnostic system equipped with modern, simple and effective algorithms and a classifier to detect lung nodule in its earliest stage, thus elevating patient's survival rate to a very high significant level. Also this methodology is very economical since it involves normal affordable computers and reduces the manual labour very much. Even patient's data can be kept confidential by the technique of watermarking.

This method can also be extended by testing and working on other body parts by setting suitable threshold values. Thus other than Lung Cancer other diseases can be found efficiently.

FUTURE ENHANCEMENTS

- Lung Cancer has 4 stages. Detecting it in the earliest stage will elevate the survival of the patient. So this project can be enhanced to detect the Lung Cancer in the stage 1 itself.
- This concept can be extended to other body parts where cancerous tissues have certain physical form.
- Along with Lung Cancer other diseases can also be detected.
- Classification of images can be done as according to factors like patient's age, gender etc.

REFERENCES

[1] S.G. Armato, W.F. Sensakovic, "Automated lung segmentation for thoracic CT: impact on computer-aided diagnosis", Academic Radiology 11 (2004) 1011–1021.

[2] S. Hu, E.A. Hoffman, J.M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images", IEEE Transactions on Medical Imaging 20 (2001) 490–498.

[3] B. Zheng, J.K. L., G.S. Maitz, B.E. Chapman, C.R. Fuhrman, R.M. Rogers, F.C. Sciurba, A. Perez, P. Thompson, W.F. Good, D. Gur III, "A simple method for automated lung segmentation in X-ray CT images", SPIE 5032 (2003) 1455–1463, doi:10.1117/12.480290.

[4] J.K. Leader, B. Zheng, R.M. Rogers, F.C. Sciurba, A. Perez, B.E. Chapman, S. Patel, C.R. Fuhrman, D. Gur, "Automated lung segmentation in X-ray computed tomography: development and evaluation of a heuristic threshold-based scheme", Academic Radiology 10 (11) (2003) 1224–1236.