Grenze International Journal of Computer Theory and Engineering , Special issue

Identification of Efficient Patterns for the Detection of Lung Cancer

K Guna Shekar[1], Pancham Bopanna D[2], C Jayvardhan Chetty[3], Basavaraj V Hiremath[4] and S C Prasannakumar[5]
[1-3]Department of Medical Electronics, Ramaiah Institute of Technology, Bengaluru, India.
[4]Assistant Professor, Department of Medical Electronics, Ramaiah Institute of Technology, Bengaluru, India.
[5]Professor, Dept. of Electronics and Instrumentation Engineering, R.V College of Engineering, Bengaluru, India.
pancham271195@gmail.com_bvhiremath@msrit.edu_prasannakumar@rvce.edu.in

***Abstract*—Lung cancer is one of the most occurring cancers among both men and women [1]. Detection of cancer at the early stage is the only method to improve the survival rate. CT images of lung are helpful in the diagnosis of lung cancer. Doctor analyses the CT image and predicts the presence of cancer nodule. This manual detection can also result in false detection. So in order to overcome this problem a computerized method for cancer detection is needed. Image processing technique can be used for this purpose. By using the following image processing techniques lung cancer detection can be developed further. Lung cancer detection system has three steps to detect the presence of cancer nodule in lung. Pre-processing stage, segmentation and feature extraction stage. Pre-processing step includes image enhancement using Gaussian filter and image segmentation. Enhanced CT image of lung is then fed to segmentation phase. From the segmented output, features are extracted. By using these extracted features the lung is differentiated as normal lung or cancerous lung. And further the cancerous tumor can be differentiated as Benign and Malignant.**

***Index Terms*— Lung cancer, MATLAB, Pre processing, segmentation, GLCM.**

I. INTRODUCTION

Cancer is one of the most serious health problems in the world. But among them, the number one cause of death in the world among men and women is Lung Cancer. The disease has a high risk of mortality. Not many patients diagnosed with lung cancer live long enough to fight it. The survival rate for lung cancer is very small. Survival rate can be estimated directly related to its detection time and rate at which it is growing. The earlier it is detected, chances of successful treatment is more. There are 2 main types of lung cancer: about 80% to 85% of lung cancers are non-small cell lung cancer (NSCLC), about 10% to 15% are small cell lung cancer (SCLC) [2]. The overall incidence of lung cancer is higher in men than women. An estimated 85% of lung Cancer cases in males and 75% in females are caused by cigarette smoking [3]. But it is not true that cancer and smoking go hand in hand. Globocan estimate of lung cancer in India would indicate that incidence of lung cancer in India is 70,275 (for all ages and both genders) with an age standardized incidence rate being 6.9 per 100,000 of our population [4]. Although there are different methods for cancer treatment such as surgery, radiation therapy and chemotherapy, but the survival rate is still very low. Early detection of lung cancer can increase the survival rate of a person. In this paper we are going to discuss the various image processing techniques used for the detection of lung cancer.

Grenze ID: 01.GIJCTE.3.4.79

II. LITERATURE REVIEW

Lung Cancer Detection from CT Image using Image Processing Techniques by Ajil M V et al. [1] we came to understand that in the enhancing technique, using log of gabor filter gave the best output. While in segmentation weibull was preferred. Also by plotting a histogram the image can be segmented into different regions by selecting the minimum and maximum grey levels. From the ROI obtained, they have extracted features like area, perimeter and average intensity. They have also considered nodule size as a factor, considering it to be normal if it was less than 25mm. clustering techniques were used. Lung Cancer Tumour Detection Using Image Processing and Soft Computing Techniques by Shraddha G. Kulkarni et al. [5] we came to understand that the image was smoothened using a Gaussian filter, also with an anisotropic filter. Marker controlled watershed techniques was used. Features such as area, perimeter and eccentricity were fed into SVM and KNN classifiers. And better accuracy was obtained using KNN classifier. Watershed segmentation was especially useful in separating touching objects. Lung cancer detection using medical images through image processing by Neelima Singh et al. [6] they have used 6 CT images and 15 MRI scan images. Enhancement of these images was done using a gabor filter and edge detection was done using canny edge detector. Super pixel segmentation was used to extract a lung nodule. From the extracted lung nodule, shape, area, size and calcification features were extracted. These extracted features were used in the classification process using the Pearsons and Spearman algorithms. Cancer Cells Detection Using Digital Image Processing Methods Bhagyashri G. Patil et al. [7] the image was enhanced using gabor filter the followed by segmenting the image using Otsu's method of thresholding also with marker controlled watershed segmentation. Binarisation approach was used in extracting the features. Accuracy using marker controlled watershed segmentation was observed to be better. Identification of Lung Cancer Cell using Watershed Segmentation on CT Images by S.Logeshkumar et al. [8] images in JPEG format of the size 512 x 512 were used. The enhancement of these images was done using a gabor filter, followed by segmenting the ROI using thresholding and watershed techniques. Features such as area, perimeter, average intensity and roundness were extracted. Upon comparison of the results obtained it was seen that watershed segmentation gave a better result.

III. METHODOLOGY

Lung cancer CT images are acquired from a private hospital and internet sources (ELCAP) are used. The images are in 2 dimensions. These images are of the sixe 512 x 512 size and are stored in JPEG format. The processes are then carried out using MATLAB. Image pre-processing, image-segmentation and feature extraction are the three main steps. The process is as shown in fig. 1.

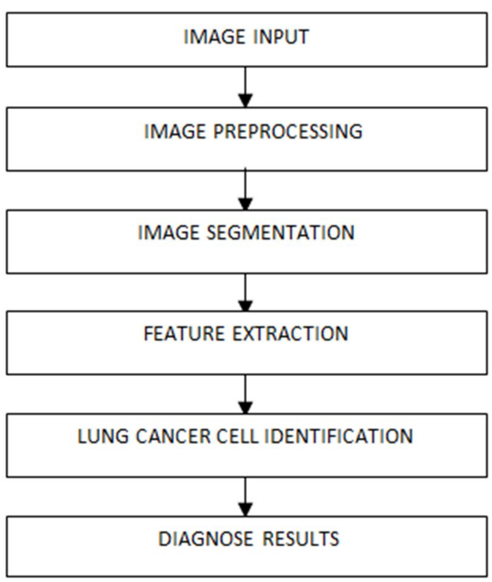


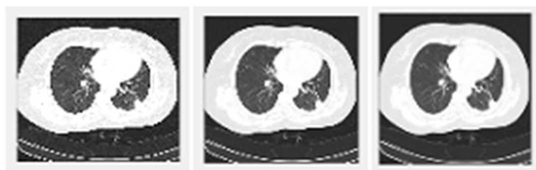
Fig. 1: Flow diagram of the methodology used

A. Image pre-processing

Once a CT image is loaded into the algorithm, there is a lot of noise that needs to be filtered and hence we use Gaussian smoothening technique. Using the following formula.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)}$$

The Gaussian matrix is 11 x 11. The sigma is kept to be at 1.5. Different sigma values were compared before choosing this value. It can be seen in fig. 2 that the image output for the sigma value of 1.5 is better.



(a) (b) (c)

Fig. 2: outputs after Gaussian smoothening with different sigma values (a) 0.5 (b) 1 (c) 1.5

B. Image segmentation

The image after pre processing has a grey scale level from 0-255. Separation of the object and the background has to be done now. Thresholding is used to do so, a histogram is plotted and the minimum and maximum values of grey level values are taken and the centre of it is calculated. This value becomes the center of that and is taken to be the thresholding value. All values below the thresholding value is made 0 and the remaining is left as it is. The output is as shown in fig. 3.

Fig. 3: Output image after thresholding

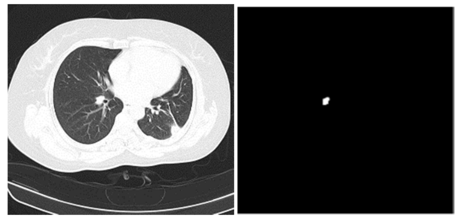
Once the image is thresholded the edges needs to be detected as to separate the nodule from the remaining region. For this purpose a sobel filter is applied. Thos helps in getting all the edges in image. The output is as shown in fig.4.

Fig. 4: Output image after sobel edge detection

Once the edges are detected, the regions are filled using the region filling operator.

C. Feature extraction

The nodule is extracted using the segmentation technique by removing all the regions that are attached to the nodule. The segmented nodule is then multiplied to the original image to obtain the grey scale of the nodule region and the output appears as shown in fig. 5.



(a) (b)

Fig. 5: (a) original image; (b) segmented nodule

Once the nodule is extracted we can use different region operators to get the area and perimeter, also Gray-Level Co-Occurrence Matrix (GLCM) to find the energy, contrast and correlation.

Lung cancer cell identification
The obtained values are then compared to check for the significant difference in energy and roundness of benign and malignant cells.

IV. RESULTS

The nodule from the original image is segmented and features such as energy, area and perimeters are extracted.
The roundness can be calculated from the area and perimeter that has been obtained from the boundary operator using the following formula.

$$\frac{4(Area)}{\pi(Length_{FeMax})^2}$$

Where $Length_{FeMax}$ is the perimeter.

The obtained nodule is as shown in fig. 6.

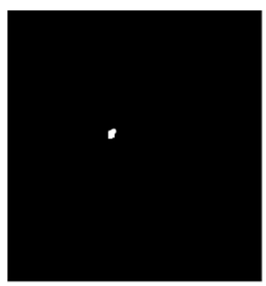


Fig. 6: output image

The energy and roundness of the obtained nodules are as shown in table 1.

TABLE I: ENERGY AND ROUNDNESS

ENERGY	ROUNDNESS
Benign cells	
0.0251	0.0123
0.0892	0.0519
0.0181	0.0567
0.0308	0.0814
0.0190	0.0973
Malignant cells	
0.1148	0.2979
0.3125	0.2387
0.2083	0.1414
0.3125	0.2387
0.1562	0.1994

V. Conclusion

The lung that is cancerous and non cancerous can be differentiated looking at the output image in the results. From the features extracted the cancerous lung tumor can also be differentiated into benign and malignant. Significant differences in values are seen between benign and malignant cells. Benign cells have lower energy and are less circular as compared to the malignant cells. Hence these two features can be used to find the type of cancerous cells.

Furthur Advancements

Further using these features and also SVM and KNN classifiers, many other features can be extracted and plotted using scatter plot. A non linear decision region can be made and this will make the classification of benign and malignant cells more precise.

Acknowledgement

We would like to thank the management and the radiology department of Ramaiah Memorial Hospital for providing us the database. Also, we would like to thank Mr. Israel (Radiology department, Ramaiah Institute of Technology, Bengaluru), for helping us with the database and the technical knowledge that was required. Last but not the least, we would like to thank our parents, friends and all our well-wishers for their constant support.

References

[1] Ajil M V, Sreeram S, "Lung Cancer Detection from CT Image using Image Processing Techniques." International Journal of Advance Research in Computer Science and Management Studies Volume 3, Issue 5, May 2015, pp 249-254.
[2] About Non-Small Cell Lung Cancer, (May 16, 2016.), American cancer society.https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/what-is-non-small-cell-lung-cancer.html
[3] Singapore Cancer Society, http://www.singaporecancersociety.org.sg
[4] National center for biotechnology information. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4991145/
[5] Kulkarni, Shraddha G., and Sahebrao B. Bagal. "LUNG CANCER TUMOR DETECTION USING IMAGE PROCESSING AND SOFT COMPUTING TECHNIQUES.". International journal of science technology and management, Volume No. 5, Issue No. 5, May 2016, pp 451-458.
[6] Singh, Neelima, and A. Asuntha. "Lung cancer detection using medical images through image Processing." Journal of Chemical and Pharmaceutical Sciences 9.3 (2016).
[7] Patil, Bhagyashri G., and Sanjeev N. Jain. "Cancer cells detection using digital image processing methods." International Journal of Latest Trends in Engineering and Technology (2014). pp 45-49.
[8] Kumar, S. Logesh, et al. "Identification of lung cancer cell using watershed segmentation on CT images." Indian Journal of Science and Technology 9.1 (2016), pp 1-4.