LUNG CANCER TUMOR DETECTION USING IMAGE PROCESSING AND SOFT COMPUTING TECHNIQUES

Shraddha G. Kulkarni[1], Sahebrao B. Bagal[2]

[1] *PG student, Department of Electronic and Telecommunication,*
Late G. N. Sapkal College of Engineering, Anjaneri, Nasik, Maharashtra, (India)
[2] *Head: Department of Electronic and Telecommunication,*
Late G. N. Sapkal College of Engineering, Anjaneri, Nasik, Maharashtra, (India)

ABSTRACT

Lung cancer is very challenging to detect at its early stage with medical examinations. Early detection of lung cancer can help for the commendable decrease in the lung cancer mortality rate. In recent years, the image processing methodologies are widely used in medical image diagnosis, especially in detection of lung cancer tumors. This paper presents lung cancer detection system using image processing and soft computing techniques. The proposed method uses Gaussian filter and anisotropic diffusion filter for pre-processing of CT images. Features such as energy, standard deviation, entropy and mean absolute deviation are calculated for the extracted lung nodule. Comparative study of classifiers is done using SVM and KNN classifier.

Keywords : Anisotropic Diffusion Filter, Gaussian Filter, Image Processing Techniques, KNN, SVM.

I INTRODUCTION

Due to the high occurrence of lung cancer in 21st century, there is need to have some methodology which can detect lung cancer precisely in its early stage. The early detection of lung cancer will not only help in its treatment but also it will contribute to decrease in death rate.

According to the study done by National Lung Screening Trial (NLST), it was observed that CT images are more reliable than any other screening method. "American Cancer Society" conduct survey every year regarding cancer issue and according to its study in 2016 it was observed that the lung cancer is secondly most identified after breast cancer. They also observed the reason behind this is that the only 15% of lung cancer cases are detected at its early stages.

Hence it may conclude that earlier the detection of lung cancer lesser will be the death rates. In this work a new algorithm is developed for detection of lung cancer at its early stage by applying image processing and soft computing techniques. The classification of tumor can be done by implementing four stages those are image acquisition, image pre-processing, feature extraction and finally the classification. For research work all the images used are CT images in DCOM format. The further paper will include literature survey, proposed method, results and discussion, conclusion and lastly the references.

II LITERATURE SURVEY

Ada, Ranjeet Kaur [4] proposed a computational procedure that sort the images into groups according to their similarities. In their work Histogram Equalization is used for preprocessing of the images and feature extraction process and neural network classifier to check the state of a patient in its early stage whether it is normal or abnormal. After that they predict the survival rate of a patient by extracted features. Experimental analysis is made with dataset to evaluate the performance of the different classifiers. The performance is based on the correct and incorrect classification of the classifier. In this paper Neural Network Algorithm is implemented using open source and its performance is compared to other classification algorithms. It shows the best results with highest TP Rate and lowest FP Rate and in case of correctly classification, it gives the 96.04% result as compare to other classifiers.

Authors of [5] described the CT image segmentation using anatomical constraints. For incorporating that they framed a new segmentation cost function based on Bayesian framework. They engaged in investigating substitution method for histogram in intensity cost function determination. The watershed algorithm from mathematical morphology is powerful for segmentation. Nguyen has explained some methods of segmentation techniques which are being classified here [6].

In the paper proposed by Dasu Vaman and Ravi Prasad [7] the image quality and accuracy is the core factors of their research, image quality assessment as well as improvement are depending on the enhancement stage where low pre-processing techniques is used based on Gabor filter within Gaussian rules. Following the segmentation principles, an enhanced region of the object of interest that is used as a basic foundation of feature extraction is obtained. Relying on general features, a normality comparison is made. In this research, the main detected features for accurate images comparison are pixels percentage and mask-labeling.

In paper [8] authors mostly focus on significant improvement in contrast of masses along with the suppression of background tissues is obtained by tuning the parameters of the proposed transformation function in the specified range. The manual analysis of the sputum samples is time consuming, inaccurate and requires intensive trained person to avoid diagnostic errors. The segmentation results will be used as a base for a Computer Aided Diagnosis (CAD) system for early detection of cancer, which improves the chances of survival for the patient. In this paper, authors proposed Gabor filter for enhancement of medical images. It is a very good enhancement tool for medical images.

III PROPOSED METHOD

The block diagram of proposed technique is presented in Fig. 1 All the images used are CT images in DICOM format. The images are enhanced using Gaussian filter and anisotropic diffusion filter. The features such as energy, standard deviation, entropy and mean absolute deviation are extracted and provided to SVM and KNN classifiers for classification.

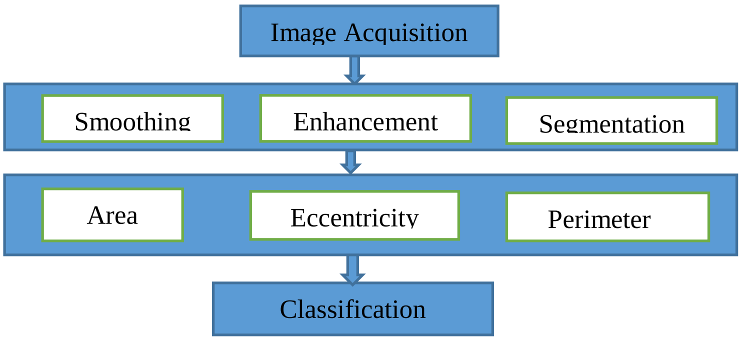

Figure 1: Block schematic of proposed work

3.1 Database collection

In order to perform classification of lung cancer, images from various stages of lung cancer is required. CT images are used due to the variety of advantages they provide as mentioned earlier. The images are strictly in DICOM (Digital Imaging and Communication in Medicine) format in order to maintain medical standardization. These images are obtained from NIH/NCI Lung Image Data Consortium and Early Lung Cancer Action Program (ELCAP). The sample CT image of lung cancerous patient is shown in Fig. 2.

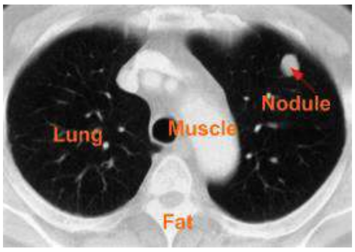

Figure 2: CT scan image of lung cancer patient

3.2 Image Pre-processing

The image pre-processing is the process in which images is subjected thorough various filtering technique in order to give noble results. The image enhancement starts with smoothing.

3.2.1 Image Smoothing

Image smoothing is done in order to remove noise and negligible fluctuations in an image. Here Gaussian filter is used for smoothing. Gaussian smoothing operator is a convolution operator that is used to 'blur' images and remove detail and noise. In this sense it is similar to the mean filter, but it uses a different kernel that represents the shape of a Gaussian ('bell-shaped') hump. This kernel has some special properties which are detailed below;

$$G(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}$$

Where, σ is the standard deviation of the distribution.

We have also assumed that the distribution has a mean of zero (*i.e.* it is centered on the line x=0). The distribution is illustrated in Fig. 3.

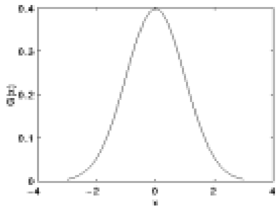

Figure 3: 1-D Gaussian distribution with mean 0 and σ =1

3.2.2 Image Enhancement

The aim of image enhancement is to improve the interpretability or perception of information in images for human viewers. This is completely a subjective process. It accentuates or sharpens image features such as edges, boundaries, or contrast to make a graphic display more helpful for display and analysis. This method uses Anisotropic filtering for enhancement.

Anisotropic filtering is a technique aiming at reducing image noise without removing significant parts of the image content, typically edges, lines or other details that are important for the interpretation of the image. In this we have to define eight masks in eight directions (North, South, East, West, NE, SE, SW and NW).After sliding of each mask in its particular direction only centre pixel is modified preserving the other pixels and hence preserving the edges. The two dimensional network structure is shown in Fig. 4.

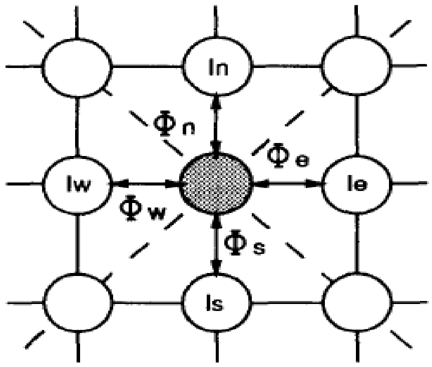

Figure 4: Two Dimensional Network Structure

The diffusion coefficient for anisotropic filtering can be given as,

$$c(\|\nabla I\|) = \frac{1}{1 + (\frac{\|\nabla I\|}{k})^2}$$

3.2.3 Image Segmentation

Image segmentation is the process of partitioning an image into different meaningful regions. Here Watershed algorithm is implemented for the purpose of segmentation. Watershed technique is a powerful method of image segmentation with high success rate.

3.3 Feature Extraction

This is an important stage in order to classify an image. Feature extraction is a process of transforming input data into set of features. The basic geometrical features such as area, perimeter and eccentricity are extracted here.

3.3.1 Area

Area is a scalar quantity. It is actually total of number of pixels identified in that extracted tumor [10].

3.3.2 Perimeter

This is a scalar quantity. It is the total number of boundary pixels found in connected component of tumor [10].

3.3.3 Eccentricity

It mainly decides the circularity of identified tumor. Minimum the eccentricity less circular will be the identified tumor and wise versa [10].

3.4 Classification

After feature extraction we have to classify features into cancerous and non-cancerous types. Hence for classification purpose two classifiers namely SVM and KNN are implemented.

3.4.1 SVM Classifier

The SVM is a supervised machine learning algorithm. With SVM each data item is plotted as point in n-dimensional space where n is number of features. The main objective of SVM is to find the hyperplane that gives largest minimum distance to training example.

Various types of kernels such as Gaussian, Radial basis kernel, Linear, polynomial can be implemented in SVM. Here we have used RBF kernel for better results. Figure 5 shows classification of SVM with different kernel functions.

3.4.2 KNN Classifier

KNN stands for 'K' nearest neighbors. Whenever we have a new point to classify, we find its K nearest neighbors from the training data. The distance method here considered is Euclidean distance method. We can also calculate distance using city block distance, Mahalanobis distance, Minkowski metric, Chebychev distance, Jaccard distance and many others.

The principle behind nearest neighbor methods is to find a predefined number of training samples closest in distance to the new point, and predict the label from these. The number of samples can be a user-defined constant or vary based on the local density of points.

While selecting value of K, it should be large so that error rate is minimized. Larger the value of K, smoother will be the classification. KNN is very simple and intuitive. It can be applied to any distribution and also larger the data size result into good classification.

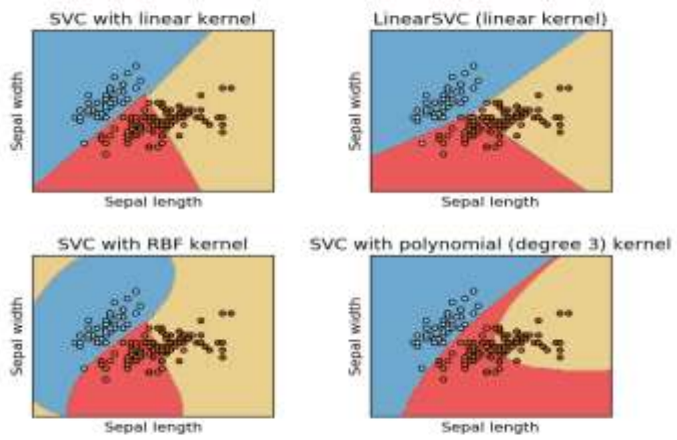

Figure 5: SVM with different kernels

IV RESULTS AND DISCUSSION

CT images in DICOM format are successfully undergone through pre-processing. The four features namely energy, entropy, and standard deviation and mean absolute deviation are extracted. These extracted features are stored in excel sheet and provided further to SVM and KNN classifier for further classification. The proposed system uses database of 90 lung cancer patients.

4.1 Results of Image Enhancement

As image enhancement is purely subjective phenomenon, the desired results may vary person to person. This is a cosmetic process and hence do not add any additional information in original. The image enhancement is done by using MATLAB software and results are as shown in Fig.6.

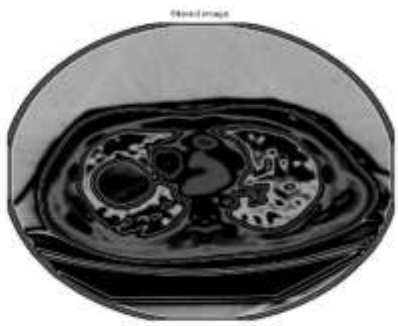

Figure 6: Result of Image Enhancement

4.2 Results of Segmentation

Proper segmentation of image is the most important in image processing here it is satisfactorily achieved by MC watershed algorithm. Fig.7 shows results of segmentation.

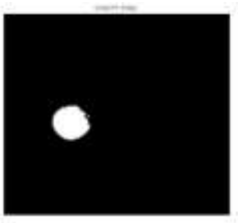

Figure 7: Result of Segmentation

4.3 Results of Feature Extraction

Feature extraction is the most vital part in detecting the stage of lung cancer. The images are processed in binary format hence limited important features are extracted. Table below shows five images with their extracted features.

Table 1: Results of Feature Extraction

SR.No.	Area	Perimeter	Eccentricity
1	1779	154.372	0.620039
2	5532	269.832	0.651564
3	5520	298.705	0.366621
4	13776	464.092	0.777694
5	1779	154.372	0.620039

4.4 Results of Classification

The results of classification using SVM and KNN are shown in graphs. From the Fig. 8 it is clear that KNN shows better performance in classification.

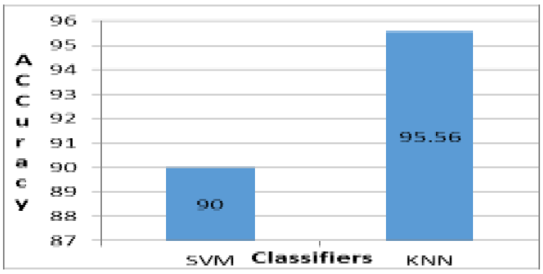

Figure 8: Performance Analysis

V CONCLUSION

Lung cancer detection technique is proposed here using different image processing and soft computing techniques. It primarily uses CT images. Images are preprocessed using Gabor, median and anisotropic diffusion filtering. The image segmentation is done by MC watershed algorithm. Features such as area, parameter and eccentricity are calculated for detection of cancer. Finally SVM and KNN classifiers are implemented for classification of images. From the results it is observed that KNN classifiers provide better results. By using this algorithm detection of cancer is possible with impressive accuracy.

REFERENCES

[1] http://www.ncbi.nlm.nih.gov
[2] http://www.cancer.org
[3] J. Kaur, Nidhi Garg, Daljeet Kaur, A survey of lung cancer detection techniques on CT scan images, *International Journal of Scientific & Engineering Research, Volume 5, Issue 6, June-2014.*
[4] Ada, Rajneet Kaur, Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier, *IJAIEM volume 2, Issue 6, June 2013*
[5] Siqi Chen, D. Michael Lovelock, and Richard J. Radke, Segmenting the prostate and rectum in CT imagery using anatomical constraints, *Medical Image Analysis Vol 15, Issue 1, February 2011*.
[6] Nguyen, H. T., Watersnakes: Energy-Driven Watershed Segmentation, *IEEE Transaction on pattern analysis and Machine Intelligence, Vol 25, No. 3, Pp.330-342,March 2003*.
[7] Dasu Vaman, Lung cancer detection using image processing techniques, *International journal of latest trends in engineering and technology (2013).*
[8] S Vishukumar K. Patela and pavan Shrivastavab, Lung A Cancer Classification Using Image Processing, *International Journal of Engineering and Innovative Technology Volume 2, Issue 3, September 2012.*
[9] Disha, Gagandeep, Identifying Lung Cancer Using Image Processing Techniques, International conference on computational techniques and artificial intelligence ICCTAI, 2011.
[10] Anitha Chaudary, Sonit.S.Singh, Lung Cancer Detection on CT Images by image processing, *IEEE International Conference on Computer Science, 2012.*