International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-4, February 2020

CAD System for Lung Cancer and its Stages Detection using Image processing Techniques

Jenif D Souza W S, Jothi S, Chandrasekar A

Abstract: ***The Lung Cancer is a most common cancer which causes of death to people. Early detection of this cancer will increase the survival rate. Usually, cancer detection is done manually by radiologists that had resulted in high rate of False Positive (FP) and False Negative (FN) test results. Currently Computed Tomography (CT) scan is used to scan the lung, which is much efficient than X-ray. In this proposed system a Computer Aided Detection (CADe) system for detecting lung cancer is used. This proposed system uses various image processing techniques to detect the lung cancer and also to classify the stages of lung cancer. Thus the rates of human errors are reduced in this system. As the result, the rate of obtaining False positive and (FP) False Negative (FN) has reduced. In this system, MATLAB have been used to process the image. Region growing algorithm is used to segment the ROI (Region of Interest). The SVM (Support Vector Machine) classifier is used to detect lung cancer and to identify the stages of lung cancer for the segmented ROI region. This proposed system produced 98.5 % accuracy when compared to other existing system.***

Keywords: CAD, Region growing, SVM, Staging, binarization

I. INTRODUCTION

The lung Cancer is one of the most common cancers in the world which causes death in people. The cancer is caused by the multiplication of abnormal cells and growing into tumor. The rate of survival of lung cancer can be is high only if the cancer is detected in the early stage. Currently CT (Computed Tomography) scan is used for diagnosing and detecting of lung cancer. This CT scan images are much efficient and accurate than the X-ray images. The CT scan images of lung can be taken from databases such as "The Cancer Imaging Archive database", NIH/NCI Lung Image Database Consortium (LIDC), IMBA Home (VIA-ELCAP Public Access) database, etc. These databases store large number of scanned medical images for different cancers. Recently the CADe based system is used for detecting the lung cancer. This CADe system uses various techniques for detecting the lung cancer.

Revised Manuscript Received on February 28, 2020.
***** Correspondence Author
Jenif D Souza W S*, Assistant professor, Department of Computer Science and Engineering, St. Joseph's College of Engineering. OMR, Chennai, India, Email: jenifdsouzaggp@gmail.com
Jothi S, Associate professor, Department of Computer Science and Engineering, St. Joseph's College of Engineering. OMR, Chennai, India, Email: jothi14.baskar@gmail.com
Chandrasekar A, Professor, Department of Computer Science and Engineering, St. Joseph's College of Engineering. OMR, Chennai, India, Email: drchandrucse@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

The detection of lung cancer can be done using various image processing stages such as Image acquisition, Pre-processing, segmentation, feature extraction, classification, localization, identification of cancer stages, etc. These modules or stages uses many image processing algorithms to detect cancer in the accurately at the early stage. The SVM classifier is mostly used to classification whether the image is cancerous or not. The staging of lung cancer is done for affected image. The stages of lung cancer are the categorized by the degree at which the cancer is spreads. Generally, there are four stages of lung cancer; from Stage I to Stage IV. The stages of cancer are detected by the size of the tumor in the lung and the location of the lymph node of the lung.

Rest of the paper is organized as: Section 2 gives idea of existing system, methodology followed and its drawbacks. Section 3 provides the research methodology of the proposed system. Section 4 shows the evaluation parameters and results. Finally the paper is concluded with conclusion part in Section 5.

II. LITERATURE SURVEY

In existing system X-ray images are taken for the detection of cancer. Later on, CT scan images are taken as it gives better efficiency. For the CT scan images various image processing techniques are applied because manual prediction had resulted in lots of false positive and false negative test results. The CT scan images are taken, the region of interest (ROI) is identified. Features such as area, contrast, energy, perimeter and eccentricity are extracted for the ROI regions using feature extraction algorithms. These feature values are used for identifying various stages of cancer. In this system a good potential for lung cancer detection is obtained at early stage [1]. In this system a CADe (Computer Aided Detection) system is used for the lung cancer detection. This system can help in assisting radiologists for the diagnosis and detection of lung cancer easily [8].

The images are taken from IMBA Home (VIA-ELCAP Public Access) database. In this work quality and accuracy is taken as the core factor. By using binarization and masking better results are obtained [8]. In this research three image segmentation methods such as Region Growing, Marker Controlled Watershed, and Marker Controlled Watershed with Masking are analyzed for lung cancer. Among these three methods, the Marker Controlled Watershed method with Masking gives us the best performance result for segmentation.

By using the binarization method, the condition of lung such as normal or cancer is determined easily [5]. In some system, the histogram values are taken for segmentation and detection of cancer [2]. In a system multiple parameter algorithm approach is used for identification of image and its stages [4]. A Linear filtering technique is used for automatic reconstruction and decomposition of image. This is done by using global thresholding techniques [6]. In some system the gene analysis of cancer is also done using several medical analysis systems [7].

III. PROPOSED SYSTEM

In the proposed system a CADe based system is used. This system used various image processing algorithm to detect and classify the stages of cancer. The proposed system contains six stages such as Image Acquisition, Preprocessing, Segmentation, Feature Extraction, Classification and Identifying stages. The Fig. 1 shows the proposed system architecture. The flow of this proposed system is shown in the below algorithm:

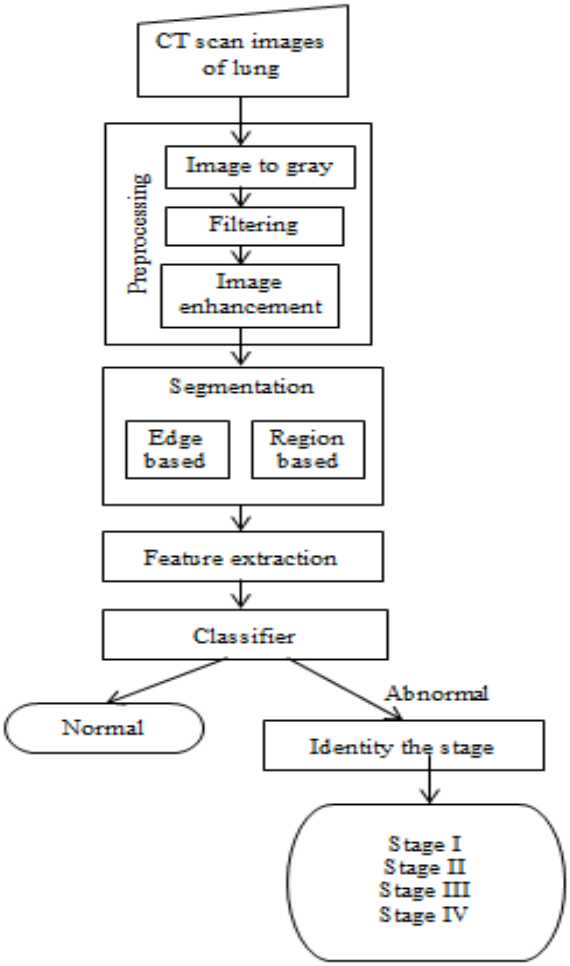

Fig. 1 Proposed system architecture

Algorithm:

Step 1: Acquire CT scan images from the LIDC (NIH/NCI Lung Image Database Consortium) or IMBA Home (VIA-ELCAP Public Access) database.
Step 2: The image is converted into gray scale image.
Step 3: The gray scale image is filtered to remove noise.
Step 4: Image is enhanced.
Step 5: The Segmentation is done to find ROI.
Step 5.1: Region based segmentation is carried out.
Step 5.2: Edge based segmentation is done.
Step 6: Using feature extraction the Features such as Area, Centroid, Perimeter, Diameter and Eccentricity are extracted.
Step 7: The Normal and abnormal images are classified using classifier.
Step 7: Finally, Identification of the stages of cancer is done using classifier.

A. Image Acquisition

First the CT scan image is acquired for the cancer patient. When comparing X-ray image and MRI images, the CT scan image has low noise, better clarity and less distortion. Both normal and abnormal CT images are acquired from LIDC (NIH/NCI Lung Image Database Consortium) or IMBA Home (VIA-ELCAP Public Access) database. This image has lots of noises which can be improved in preprocessing stage. The Fig 2.a shows the lung CT scan input image.

B. Image Preprocessing

The preprocessing is the initial stage which is done in order to improve the quality, contrast, clarity and also to separate the background noise from the image. In this system, preprocessing modules includes three stages such as converting image to grayscale, noise filtering and image enhancement.

Step 1: Image to Gray

The input image (CT scan image) is converted into grayscale format. This is the initial process done in image processing. The Fig2.b. Shows the Gray scale converted image.

Step 2: Filtering

The filtering is done in order to reduce the rate of detection of false positive and false negative nodules. Using the Median filter, the salt and pepper noise of the image is removed. The salt and pepper noise of 0.05 densities is added which is shown in Fig2.c.Then this image is filtered by using Gaussian filter. This filter can reduce the noise and also can blur the image. The Gaussian filtered image is shown in Fig2.d.

Step 3: Image enhancement:

Image enhancement is done in order to increase the quality of image. The filtered image has to enhance by using histogram equalization method. The histogram of the output image matches approximately with a specified histogram. This is done by modifying the values in an intensity image.

This results in gaining higher contrast for areas of lower local contrast. The quality and clarity of image is improved.

The Fig.2.e shows enhanced image for the filtered image.

C. Image Segmentation

The image segmentation is the process of dividing or segmenting or partitioning an image into multiple segments. The Segmentation is done in order to detect the possible abnormal areas present in the image.

Segmentation plays an important role in medical image processing. In the segmentation Region of Interest (ROI) is segmented. The segmentation of image is done by using block based segmentation method such as region based method and Edge or boundary based method. Image segmentation includes following steps. The Fig. 2.f and Fig. 2.g show segmentation from region growing (region based) and prewitt (edge based) respectively.

Step 1: Region based method

In region based method all the pixels of same grey level are taken in one region as in Fig 2.f. The region growing algorithm is used for this method; this algorithm first selects a pixel and goes on increasingly adding all the surrounding pixels of similar grey level. This process repeats until all the pixels are added in that region.

Step 2: Edge based method

The edge based method is used to detect the edges. In edge based method, edges are detected by localization of the lung cancer using prewitt operator as in Fig.2.g. It is an edge detection method which detects the affected region of the breast. The Prewitt method finds edges using the derivative. This method, if the gradient of I is maximum then that edges are returned. It specifies the sensitivity threshold for the Prewitt method. The edges that is not stronger than Threshold has been ignored by EDGE method. If you do not specify Threshold, or if threshold is empty, Edge chooses the value automatically.

D. Feature Extraction

The feature extraction is a process of calculation of some properties or features of the image. The regionprops function is used for this process, in which 21 features are extracted from the image. Among those features area, Centroid, Perimeter, Eccentricity and Diameter are taken in this proposed system. In training phase the results of feature extraction are stored in database. These features are then taken for classification process.

E. Classification

The Classification is the process of classifying or sorting out the images as normal or abnormal images. The SVM (Support Vector Machine) algorithm classifies and shows the whether the image as affected or not affected. The SVM uses functions such as train, struct and classify to train the data, store information and classify the data respectively. The SVM is a supervised learning technique that analyzes data and recognizes patterns for classification. For every image training and testing is done.

Step 1: SVM Training

The SVMTRAIN(TRAINING, Y) trains a SVM classifier on data taken from two groups. Each element of Y specifies the corresponding row of training belongs to. The TRAINING and Y must have the same number of rows. This is how training is done.

Step 2: SVM Classify

SVMSTRUCT contains information about the trained classifier and the support vector, which is used by SVMCLASSIFY for classification of normal and abnormal images.

F. Identifying stages

The degree which the cancer spreads is categorized in this stage. The SVM takes the feature extraction values and finds the stages of lung cancer. In this phase, the cancer staging is categorized into four:

Stage I:

This is the initial stage in which the cancer is founded only in the lung and has not spread to the lymph nodes or other parts.

Stage II:

In this stage, the lung cancer is present in lung and has spread to nearby lymph nodes.

Stage III:

In this stage, the cancer has spread in the lung, chest and lymph nodes.

Stage IV:

This advanced stage in which the cancer has spread from the chest to other parts of the body.

Thus the stages are categorized. By the feature extraction values, the staging of lung cancer is identified such as diameter, area and centroid. Finally result will be shown in the form of message box.

IV. EXPERIMENAL RESULTS

The experimental result of proposed system is shown in Fig.2. This Figure include input image, gray converted image, enhanced image and noise filtered image as a result of preprocessing. The Fig.2.a shows the input image which is converted into gray image as in Fig.2.b. Salt and pepper noise is added for gray image as shown in Fig.2.c. The Fig.2.d shows the noise filtered image by applying Gaussian filter. The Fig.2.e shows the enhanced image by using histogram equalization. After preprocessing, segmentation is done. In segmentation ROI is segmented by region based method and edge based method using region growing algorithm as shown in Fig.2.f and Prewitt algorithm as shown in Fig.2.g respectively. After segmentation, the Feature extraction is done in which various features were extracted. Classifications of images are done after feature extraction, in this phase abnormal images and normal images are classified. Finally by using SVM classifier staging of lung cancer is done for the abnormal image.

Fig. 2.a Input Image

Fig. 2.b Gray image

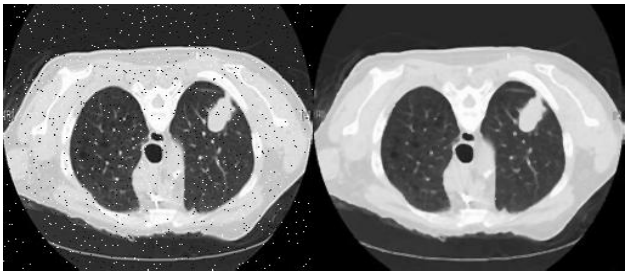

Fig. 2.c salt and pepper Noise filtered image

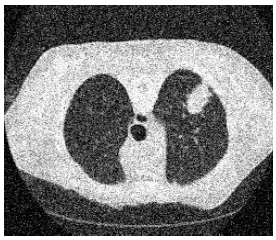

Fig. 2.d Noise filtered using Gaussian filter

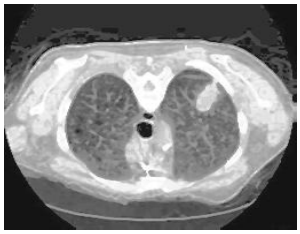

Fig. 2.e Enhanced image

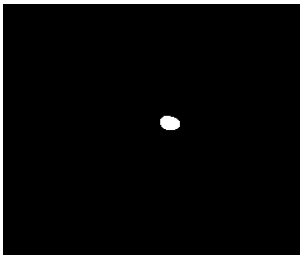

Fig. 2.f Region based segmentation

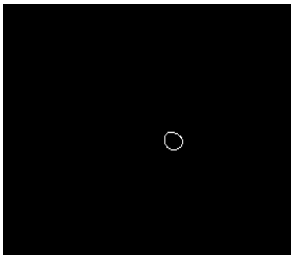

Fig. 2.g Edge based segmentation

Fig 2. Experimental Results

In this proposed system the accuracy is measured from the rate of TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative). The equation (1) shows the formula for finding accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TN+FP+FN+TP} \quad (1)$$

The proposed system produces an accuracy of 98.5% when compared to the existing systems which is calculated using the equation (1).

V. CONCLUSION

In this proposed system the rates of obtaining human errors are reduced because of using CADe based system. This proposed system produced 98.5 % accuracy when compared to other existing system. The accuracy is obtained by measuring the rate of TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) values. In this system, Region growing algorithm is used to segment the ROI (Region of Interest) which produced a better result than the existing system. The SVM (Support Vector Machine) classifier is used to detect lung cancer and to identify the stages of cancer for the ROI segment. This system helps in detecting the stages of cancer easily, accurately and in lesser time.

REFERENCES

1. Prof. Deshpande L.M, Mr.Vijay A.Gajdhane, , “Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques”, “IOSR Journal of Computer Engineering (IOSR-JCE)”, p-ISSN: 2278-8727, e-ISSN: 2278-0661, Volume 16, Issue 5, Sep – Oct. 2014, PP 28-35.
2. Jenif D Souza W.S, S. Jothi, A. Chandrasekar. "Automated Attendance Marking and Management System by Facial Recognition Using Histogram", 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019.
3. Mokhled S. AL. Tarawneh, “Lung Cancer Detection Using Image Processing Techniques”, “Leonardo Electronic Journal of Practices and Technologies”, ISSN 1583-1078, Issue 20, January-June 2012, p. 147-158.
4. Balamurali R., and A. Chandrasekar (2019), ‘Multiple parameter algorithm approach for adult image identification’, Cluster Computing-The Journal of Networks, Software Tools and Applications. Vol.22, Supplement 5,pp. 11909–11917 . DOI: 10.1007/s10586-017-1510-3.
5. Alhadi Bustamam, Bariqi Abdillah, and Devvi Sarwinda, “Image processing based detection of lung cancer on CT scan images”,“The Asian Mathematical Conference 2016”, Conf. Series: Journal of Physics: Conf. Series 893 (2017), DOI :10.1088/1742-6596/893/1/ 012063.
6. Prabhu Chakkaravarthy A and Chandrasekar, A (2019), ‘A linear filtering on automatic decomposition and reconstruction of dermoscopy images using global thresholding’, International Journal of Innovative Technology and Exploring Engineering. Vol.8, Issue.10, pp.4257-4263.DOI: 10.35940/ijitee.J9959.0881019.
7. Nithya P and Chandrasekar A (2019), ‘NBN gene analysis and its impact on breast cancer’ Journal of Medical Systems. Vol.43, Issue.270, DOI: 10.1007/s10916-019-1328-z.
8. Pramit Chatterjee, Rishi Gupta, S. Kalaivani, Shikhar Juyal, “Lung Cancer Detection Using Digital Image Processing and Artificial Neural Networks”, “International Conference on Electronics, Communication and Aerospace Technology ICECA 2017”, 978-1-5090-5686-6/17.

AUTHORS PROFILE

Jenif D Souza W S is Assistant Professor of the Department of CSE at St. Joseph’s College of Engineering, Chennai, Tamil Nadu. She has 3 years of teaching experience on graduate and post-graduate level. She has published many papers in various national and international journals and Conferences. Her area of interest includes Network Security, Image Processing, Machine Learning and Deep Learning.

S. Jothi is Associate Professor of the Department of CSE at St. Joseph’s College of Engineering, Chennai, Tamil Nadu. She has 14 years of teaching experience. She has published 15 technical papers in various national and international journals. Her area of interest includes Wireless Sensor Networks, Mobile ad-hoc Networks, Big Data Analysis and Image Processing.

A. Chandrasekar is Professor of Department of CSE at St. Joseph’s College of Engineering, Chennai, Tamil Nadu. He has 22 years teaching experience. He has guided more than 24 Research Scholars. He has published over 125 research articles in refereed International and National journals. His areas of interest are Network Security, Cloud Security, Data mining, Artificial Intelligence and Big Data Analysis.