International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Design and Development of an Efficient Mining Framework for Pre-Cancerous Lesion Detection in Lung Using Non-Invasive CT Imaging

R.Priyatharshini , Aparajitha.K, Aarthi.B, Dhivya.N

Abstract*: *The ability to analyze and identify meaningful patterns in clinical data must be addressed to provide a better understanding of disease. Currently existing solutions for disease diagnosis systems are costly, time consuming and prone to errors, due to the diversity of medical information sources. Lung Disease Diagnosis individual is based on medical images (Lung CTs) includes Lung segmentation, and the detection of cancerous lesions in the Lung. Segmenting the region of interest from medical imaging is a challenge, since the images are varied, complex and can contain irregular shapes with noisy values.In this context, the segmentation of the Region of Interest from Lung CT and detecting the pre-cancerous lesions is an important research problem that is receiving growing attention. Hence an efficient methodology on ACM based automatic segmentation and precancerous lesion detection is proposed.

Keywords: About Active contour model, Disease diagnosis, Lung Segmentation, Lesion Detection.

I. INTRODUCTION

Imaging modalities such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI),and Positron Emission Tomography (PET) are commonly used to detect the disease patterns from medical images. Lung cancer is a lethal disease which causes death of human beings at an early age. This is due to uncontrolled cell growth in the lung tissues. The diagnostic methods available are less effective for detection of cancer. The ability to analyze and identify meaningful patterns in clinical data must be addressed to provide a better understanding of disease. Mining disease patterns from clinical data is a challenge, given the gray-level ambiguity in the soft tissues of the lung, problems with leakage in clinical images. Segmenting anatomical structures, alongside detecting and quantifying tumors or lesions from medical images, is extremely difficult due to the closeness in the gray-level values of the soft tissues in the lung.

Revised Manuscript Received on November 30, 2019.
***** Correspondence Author
R.Priyatharshini*, Department of Information Technology,Easwari Engineering College, Chennai, India.
Aparajitha.K,Department of Information Technology,Easwari Engineering College, Chennai, India.
Aarthi.B ,Department of Information Technology,Easwari Engineering College, Chennai, India.
Dhivya.N, Department of Information Technology,Easwari Engineering College, Chennai, India.

In this context, the extraction of disease patterns from clinical data is a key research problem that has received growing attention. Lung Disease Diagnosis individual is based on medical images (Lung CTs) includes Lung segmentation, and the detection of cancerous lesions in the Lung. Segmenting the region of interest from medical imaging is a challenge, since the images are varied, complex and can contain irregular shapes with noisy values which will affect the accuracy of segmentation techniques. The detection of pre-cancerous lesions from Lung CT imaging modalities is difficult, with the current atlas-based and region-based methods. The rate of false positives introduced with the existing methods is also high, as a result of variations in the shape of the anatomical structures. Although a lot of work has been reported, it is still not easy to design an efficient and efficient framework because of shape variations and gray-scale ambiguity issues in clinical images. These variations make risk level assessment a challenge. Currently existing solutions for disease diagnosis systems are unable to obtain complete information about disease patterns. Therefore, designing a classifier that can distinguish between cancerous and non-cancerous tissues and quantifying the lesions is still an open problem. In this context, the segmentation of the Region of Interest from Lung CT and detecting the pre-cancerous lesions is an important research problem that is receiving growing attention. Therefore, an automatic precancerous lesion segmentation method with computed tomography (CT) scans has been developed. Due to existence of variation in lesions, it is quite difficult to identify and segment lung tumor automatically with good accuracy. The application of a robust lesion detection and segmentation techniques to segment every individual cell from pathological images to extract the required features is essential for early detection of cancerous lesion from Lung CT Images. In this project, we are planning for non-invasive methodology for pre-cancerous lesion identification.

II. LITERATURE REVIEW

The healthcare domain currently depends on medical images and medical imaging technology like MRIs, CTs, and Ultrasound scans. Doctors use these for the study of anatomical structure and treatment planning. In general, medical images are complex and noisy. Threshold-based segmentation acts as a pre-processing technique, reducing the complexity of medical image analysis, and eliminating noise and unnecessary regions without any loss of crucial information in the image.

Threshold-based methods are developed based on image intensity [9], gradient magnitude or a combination of both. Image intensity-based methods use region growing techniques, while methods based on gradient magnitude are Canny edge and Laplacian techniques. The simple threshold operation does not work well when segmenting images holding many objects, each with a separate gray-level value varying over a band of values and the Image gradient methods are affected by noise and partial volume. Unsupervised techniques such as the K-means and Fuzzy C-means are used for medical image segmentation. Fuzzy clustering is an excellent method for segmenting both CT and MR images. A novel kernelized fuzzy C-means algorithm was developed for medical image segmentation [2]. An unsupervised region-based image segmentation was developed using texture statistics and level set methods for segmenting the medical images [6].

Atlas-based segmentation approaches are the most often used and powerful approaches within the field of medical image segmentation. In this approach, details on anatomy, shape, and size, as well as the features of miscellaneous organs and soft tissues, are compiled in the form of an atlas or a lookup table (LUT). Atlas-based segmentation approaches is the third-generation medical image segmentation strategies [11]. An approach using Neural Networks Arbitration is developed. The diagnosis helps to predict whether heart disease is present or absent in the patient. An Efficient Disease Diagnosis framework must be developed to improve the accuracy of lesion detection for disease diagnosis.

The Active Contour Model (ACM)-based image segmentation technique, including both edge- and region-based approaches, can be most effective. This model is analytical, involving extensive computations to incorporate effective and useful concepts of energy, force, velocity, and curve constraints. As far as current segmentation methods are concerned, particularly the snakes, further parameters of energy could perhaps be studied to provide a better adjustment of the real contour.

A modified Particle Swarm Optimization (PSO) based feature selection for the classification of lung CT images was developed. PSO is employed to implement a feature selection in the wrapper-based methodology, and therefore the K-NN classification function is a fitness function of the PSO for the classification problem. The PSO-based feature selection methodology is applied to the features extracted from the lung CT scan images. It is observed from the literature surveyed that an optimal feature selection algorithm is required to identify the most important features for better class discrimination. The presently available algorithms are unable to accurately discriminate between cancerous tissues and background due to issues with intensity and homogeneity in medical images. It is evident from the literature surveyed that an efficient framework is required to detect pre-cancerous lesions for early detection of lung disease.

.

III. MATERIAL & METHODS

In this section we propose a system for segmentation of lung from Lung CT image and detecting the pre-cancerous lesions from segmented lung. The training data set is preprocessed to remove noise and to improve the visual quality of the image if needed. Preprocessed data set is segmented using ACM-based segmentation from 3-D CT images. Features characterizing precancerous lesions are extracted for detecting cancerous lesions. Detected precancerous lesions helps in early diagnosis of lung cancer. Fig. 1 represents the working methodology of the proposed system.

A. Image Segmentation

An Energy Efficient Active Contour Model (EE_ACM) based region growing technique for segmenting lung from lung CT has been proposed in this work. A methodology has been designed to handle variations in the shape of the lung and tackle gray-scale ambiguity in order to develop a framework for Lung segmentation from a Lung CT. The ACM method follows the deformable model-based approach and comprises the following processes. First, the Lung CT input image is pre-processed by applying morphological operations to enhance the CT image. A shape feature to handle shape variants, as well as image gradient features based on image intensity, is extracted from the input image. Contours are initialized to steer the evolution of active curves to reduce computation time. Objectives can be formulated to segment the lung which best approximates the shape of the object. The contour evolves to neighborhood pixels with minimum energy. This method works across all types of images such as CT images with low or poor contrast and weak edges and is robust enough to adapt to shape variations in the lung. It provides promising results when compared with other techniques. Figure 2 represents the segmentation of lungs during contour formation

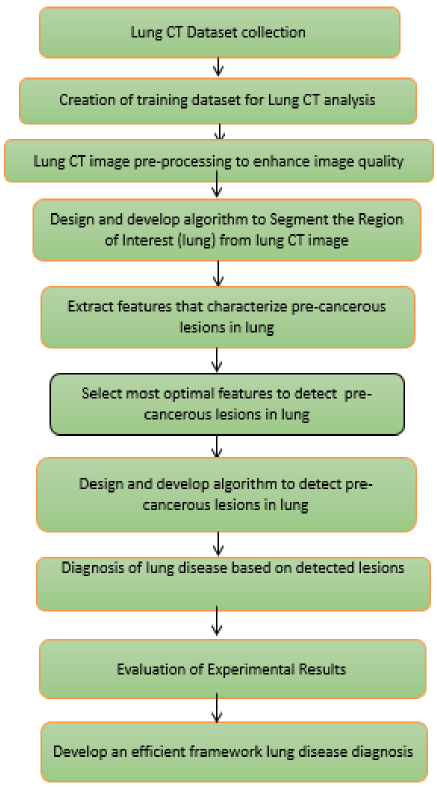

Figure 1. Workflow of Proposed Precancerous Lesion Detection Framework

Energy Efficient Active Contour Model (EE_ACM) Algorithm:

1. A set of snakes (seed) points residing on the image plane are defined
2. By using the external energy and internal energy, next location of those snake points are determined. External energy and internal energy are calculated using equation 1 and 2 respectively.
   Eexternal = argmin *j ((E), K(t)), j <- Wi* (1)
   $E_{internal} = \frac{1}{2}(\alpha(s)|v_s(s)|^2) + \frac{1}{2}(\beta(s)|v_{ss}(s)|^2) = \frac{1}{2}(\alpha(s)\|\frac{dv}{ds}(s)\|^2 + \beta(s)\|\frac{d^2v}{ds^2}(s)\|^2)$ (2)
3. The associated form of those snake points is considered as the contour (The snake points are firstly placed at more distance from the boundary of the object)
4. Then, every point moves towards the optimum coordinates, where the energy utility converges to the minimum.
5. The snake points end on the boundary of the object.

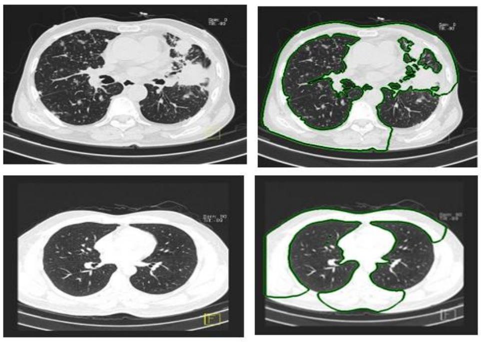

Figure 2. Contour evolution during segmentation.

B. Precancerous Lesion Detection

A method for detecting the pre-cancerous lesions present in Lung CT images are presented to tackle the issues in the literature with optimal set of features. An approach is designed to improve the accuracy of pre-cancerous lesions detection. During online processing, when a user supplies an input image, selective features are extracted from it and used to classify pixels as cancerous or non- cancerous, based on the intensity threshold. Subsequently, features which discriminate between cancerous or non- cancerous pixels are selected using an embedded feature selection method. The selected features are thereafter extracted from the candidate cancerous objects and true cancerous objects are detected. The optimality of the chosen features improves performance, when compared with the existing methods. Table-I lists the characteristic features that are observed in the segmented lungs portion. The basic feature recognized are density (40-50) and size (<=3cm) which characterizes precancerous lesions.

Table-I FEATURES FOR LESION DETECTION

Type of feature	Feature	Characteristics of feature
Density	Hounsfield units	40-50
Size	Area	<=3cm

C. *Disease Diagnosis*

Before lung cancer develops, precancerous lesions are found in the airway which can be categorized as malignant (harmful), benign (harmless) and precancer conditions. These are differentiated using featuring characteristics like 8-20mm (15% malignin), >20mm (75% malignin), 4-7mm (1% malignin), 15HU (benign), Appearance ground-glass opacity, cavities, three-dimensional ratio>1.78 etc. The lesion features obtained enhances prevention of severity in diseases and can be utilized for periodically monitoring the health of individuals.

IV. EXPERIMENTAL SETUP

The proposed system on detection of precancerous lesions is been implemented in MATLAB2015. The results of segmentation of lungs from non-invasive image is been shown in figure 3.

Segmentation using EE_ACM has shown better accuracy when compared to other traditional methods which is been illustrated in performance evaluation. For accuracy purpose we have used several images considering both cancered lung CT and normal CT. Table-II lists the number of datasets collected from various radiologic centers.

TABLE-II LUNG CT DATASETS FOR EXPERIMENTAL EVALUATION

Type of lung CT image	Number of images
Unaffected (normal) lung images	20
Affected (cancerous) lung images	Benign- 9 Malignant- 12 Pre cancer condition- 15

The accuracy of lesion detection form segmented image depends on the datasets given as input. The method we have proposed has given a better performance than compared to the existing systems.

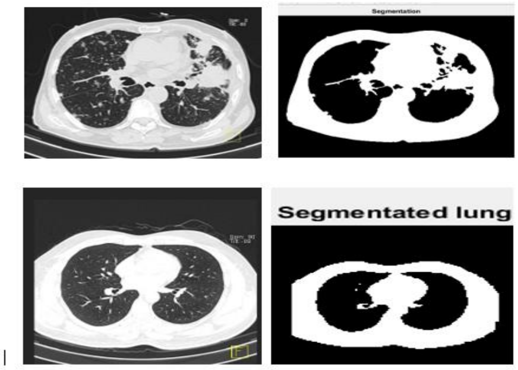

Figure 3. Segmented results of infected lungs and unaffected lungs

V. PERFORMANCE EVALUATION

Similarity measures such as the Jaccard and Dice indices are also used to evaluate the performance of the image segmentation algorithm which is shown in Table -III.

The Dice co-efficient is a similarity measure, largely used in medical image processing, to evaluate the performance of segmentation algorithms with predefined ground truth information or a dataset. The Jaccard index (A, B) and Dice index (A, B) are similarity measures located in the range [0, 1], used to evaluate binary variables.

$$J(A,B)=\frac{|A\cap B|}{|A\cup B|} \qquad (3)$$

$$D(A,B)=\frac{2|A\cap B|}{|A|+|B|} \qquad (4)$$

In this work, locations segmented by means of existing techniques is represented by A and B are utilized to depict regions outlined by the Expert. In these similarity measures, if regions A and B are fully superimposed, the attained outcome is 1, and 0 when both regions are clearly distinct. Figure 4 represents the performance evaluation of segmentation using jaccard and dice indices. The graph shows the performance of the proposed system is better than the existing system in the aspect of the performance measures considered. The approximate percentage of jaccard index is 0.93% and of dice index is 0.89% which indicates the accuracy of the proposed system.

TABLE -III. PERFORMANCE EVALUATION ON BASIS OF JACCARD AND DICE INDICES

	Methods			
Measures	**Threshold based**	**Region based**	**Fuzzy Contour based**	**Proposed EE_ACM**
Jaccard Index	0.59	0.62	0.77	0.93
Dice Index	0.71	0.77	0.84	0.89

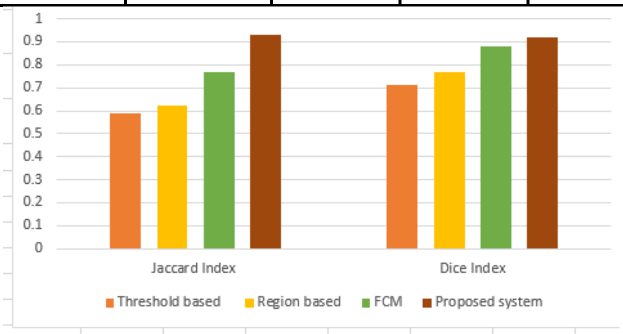

Figure 4. Comparative Performance Analysis of Lung Segmentation using Jaccard and Dice indices.

VI. CONCLUSION

Automatic detection of pre-cancerous lesions from Lung CTs for early detection of lung disease is an active research area in the design of machine vision systems and is used in miscellaneous applications. The Outcome of the proposed research work can be utilized for periodically monitoring the health of Individuals. It will also be used as an aid for medical practitioners to detect early the presence of lung disease in individuals with the following risk factors like having allergies, being overweight, smoking, being exposed frequently to pollutants, family history etc.

ACKNOWLEDGEMENT

This work was supported by DST-FIST Programme No.SR/FST/College-110/2017, Government of India

REFERENCES

1. Damodar Dipika A, Prof. Krunal Panchal, "A Survey of Lung Tumor Detection on CT Images", December 2015, Volume 2, Issue 12, JETIR (ISSN-2349-5162).
2. Dao Qiang Zhanga & Song Can Chena, 'A novel kernelized fuzzy C-means algorithm with application in medical image segmentation',Artificial Intelligence in Medicine Elsevier, vol. 32, no. 1, pp. 37-50
3. Deepthy Janardhanan, Dr.R.Satishkumar, "Lung Cancer Detection Using Active Contour Segmentation Model "International Journal of Innovative Research in Advanced Engineering (IJIRAE)ISSN: 2349-2163 Issue 03, Volume 4 (March 2017) Special Issue .
4. Hongwei Ji, Jiangping He, Xin Yang, Rudi Deklerck, and Jan Cornelis, *Senior Member, IEEE, "*ACM-Based Automatic Liver Segmentation From 3-D CT Images by Combining Multiple Atlases and Improved Mean-Shift Techniques", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 17, NO. 3, MAY 2013.
5. Jun Lai, Ming Ye, "Active Contour Based Lung Field Segmentation", 2009 International Conference on Intelligent Human-Machine Systems and Cybernetics.
6. Karoui,I, Fablet,R, Boucher,J & Augustin, J 2007, 'Unsupervised region-based image segmentation using texture statistics and levelset methods', Proceedngs of WISP IEEE International Symposium on Intelligent Signal Processing, pp. 1-5
7. M Lavanya, P Muthukumaran, "Lung Lesion Detection in CT Scan Images Using the Fuzzy Local Information Cluster Means (FLICM) Automatic Segmentation Algorithm and Back Propagation Network Classification", Article 32, Volume 18, Issue 12, December 2017.
8. Minu George, GopikaS, A Survey on Lung Nodule Detection in CT-Images, International Journal of Scientific & Engineering Research, Volume 6, Issue 12, December-2015 ISSN 2229-5518
9. Pohle, R & Toennies, KD 2001, 'Segmentation of medical images using adaptive region growing', International Society for Optics and Photonics in Medical Imaging, pp. 1337-1346.
10. S. Kalaivani, Ms. T. Sathiya, Dr. B. Sathiyabhama, "ACM based Segmentation of pulmonary lung nodule in CT images", ISSN: 1995-0772 Published BY AENSI Publication EISSN: 1998-1090 2017 May 11(7): pages 592-601.
11. Temesguen Messay a, Russell C. Hardie a, Timothy R. Tuinstra b," Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the Lung Image Database Consortium and Image Database Resource Initiative dataset", Medical Image Analysis 22 (2015) 48–62.
12. Withey, DJ & Koles, ZJ 2007,'Three generations of medical image segmentation: Methods and available software', International Journal of Bioelectromag, vol. 9, no. 2, pp. 67-68.

R.Priyatharshini received her BE degree from University of Madras in 2001 and ME degree from Anna University in 2008. She has completed her PhD degree from Anna University in 2017. She has 16 years of teaching experience and currently working as an associate professor in the Department of Information Technology, Easwari Engineering College, Chennai. She has published more than 20 research papers in International Conferences and Journals. Area of interest towards Data Mining, Medical Image Processing and Big Data Analytics. She is a life time member of Institution of Engineering and Technology.

Aparajitha.K completed her B.Tech degree in Information Technology from Easwari Engineering College (Anna University) in 2019. Her area of interest is towards Data mining and Artificial intelligence.

Aarthi.B completed her B.Tech degree in Information Technology from Easwari Engineering College (Anna University) in 2019. Her area of interest is towards Data mining and Artificial intelligence.

Dhiya.N is currently pursuing her B.Tech degree in Information Technology from Easwari Engineering College (Anna University) and is currently in her final year (2015-2019). Her area of interest is towards Data mining and Artificial intelligence.