Segmentation Based Methodology for Lung Cancer Detection

Archa Unnikrishnan [1], Remya Mol R [2], Suresh Kumar N [3]
[1]*Mtech Scholar*, Dept. of CSE, College of Engineering and Management, Punnapra, Kerala, India
[2] *Senior Lecturer*, Dept. of IT, College of Engineering and Management, Punnapra, Kerala, India
[3] *Senior Lecturer*, Dept. of CSE, College of Engineering and Management, Punnapra, Kerala, India
(E-mail: archaunni9mtech@gmail.com)

Abstract- The biomedical field plays an important role in diagnosing lung cancer. Lung tumor is a contagious disease due to the growth of abnormal cells in the lung tissue. There are a variety of imaging strategies can be used such as radiography, CT, PET, and MRI. In this paper, we propose an efficient detection of lung cancer from a CT scan image. Our proposed strategy comprises the following three phases: preprocessing, Segmentation, and Classification. The input is the Lung CT scan image. In the first phase, Preprocessing by Grey Scale Imaging, Noise Removal (Median Filter), and Histogram Equalization. For the segmentation, the Content-Adaptive Superpixel algorithm is used. The segmented images are categorized using the SVM classifier.

Keywords— *Lung Cancer, Segmentation, Computer Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), Content-adaptation, Preprocessing, Support Vector Machine (SVM).*

I. INTRODUCTION

Digital Image Processing (DIP) is a computerized PC that uses a particular computer vision algorithm to process digital images. It is a subcategory or field of advanced signal handling. The image is converted into a digital form, and it performs operations. So that get an enhanced image and information is extracted from it. DIP provides a wide variety of algorithms for input data. The basic steps required for doing this include preprocessing, segmentation, feature extraction, classification, and testing. Image Processing can be utilized in different fields such as the biomedical field, agricultural field, robotics, and computer vision, and so on.

The medical field plays an important role in digital image processing. Lung cancer is a deadly type of cancer, and in its early stages, it is difficult to detect. Lung cancer is a disease in which irregular cells expand and develop into a tumor [1]. The discovery of lung malignant growth is the most intriguing examination zone of specialist's in beginning times. Small cell lung cancer (SCLC) [2] and non-small cell lung cancer (NSCLC) [3] are separate forms of lung cancer. The SCLC includes Small cell carcinoma and Combined Small cell carcinoma. The three main types of NSCLC in Figure 1[4] are adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Other forms of tumors are located inside the lung. They're considerably less popular. Types include cancers of the salivary glands, lung sarcoma, and lung lymphoma.

Nowadays, a variety of imaging techniques used by doctors for diagnosing lung tumors. The most efficient imaging strategies can be used such as radiography, CT scan [5], PET scan [6], and MRI scan [7]. In this paper, CT scan Lung images can be used. CT examine picture is considered as extraordinary compared to other strategies that give point by point description about lung nodules. It has low noise and less distortion and gives clarity over various images. CT is the most delicate imaging procedure for recognizing lung modules.

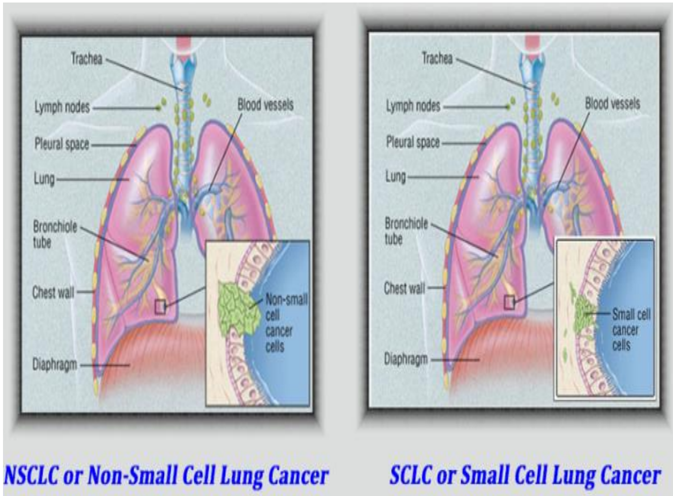

Fig. 1: Types of Lung Cancer

[https://www.google.com/search?q=different+type+of]

This paper explores the idea of automatic lung tumor segmentation. There are different segmentation algorithms [8] that were developed for lung cancer detection, but these algorithms result in less accurate segmentation results. Image segmentation can be of different types such as edge-based, texture-based, thresholding based, and region based. Superpixel is a collection of pixels associated with identical colors or grays. Superpixel segmentation targets gathering pixels in an image into nuclear districts whose limits adjust well to the characteristic item limits. It includes a new feature representation that embraces color, contour, texture, and spatial features for superpixel segmentation. Instead, we implement a clustering-based discriminability measure to iteratively determine the value of different features. We suggest a novel content-adaptive superpixel (CAS) segmentation algorithm [9] combining the feature representation and the discriminability measure. CAS will change the weights of different features automatically and iteratively to match different image instance properties. The approach adopts a modern feature representation that uses color, contour, texture, and spatial features to provide a robust characterization of the input image. A feature discriminability test was proposed to determine the segmentation capability of each category of features. A content-adaptive clustering algorithm is further designed to accomplish the task of superpixel segmentation, based on the function representation and discriminability scale. In particular, we change the weights of different characteristics iteratively and adaptively based on their discrimination on the current pixel partition for different image instances, resulting in a more reliable and rational distance measure to discriminate pixels on natural images. After successful segmentation, a machine learning classifier named Support Vector Machine (SVM) [10] to detect lung tumors is used.

The proposed model consists of the following steps as Collection of the dataset, preprocessing, segmentation, and classification. Each step is described in further sections. This paper is organized as follows: Section II presents related works in connection with lung cancer segmentation. Section III describes the methodology. Section IV gives details regarding the conclusion and future work.

II. Related works

One of the superpixel generating algorithms is the Simple linear iterative clustering (SLIC) algorithm. This algorithm creates superpixels based on their color similarity and proximity in the image plane, by clustering pixels. In paper [11] segmentation method for lung nodule image sequences based on superpixels and density-based spatial clustering of applications with noise, the technique is proposed (DBSCAN). First, for preprocessing, our approach uses three-dimensional computed tomography image features of the average strength projection combined with multi-scale dot enhancement. Hexagonal clustering and morphologically optimized linear iterative clustering (HMSLIC) are then proposed to obtain superpixel blocks for the sequence image over-segmentation. The adaptive weight coefficient is then designed to calculate the distance between superpixels needed for accurate positioning of the lung nodules and the subsequent clustering of the starting blocks. Besides, an effective clustering threshold is obtained by fitting the distance and detecting the shift in slope. An algorithm for rapid grouping of DBSCAN superpixel sequences, optimized by the strategy of grouping lung nodes and the adaptive threshold, is then used to obtain sequences of masks from lung nodes. Ultimately, image sequences of the lung nodule are obtained.

In paper [12] Segmentation Method for Lung Parenchyma Image Sequences Based on Superpixels and a Self-Generating Neural Forest is used. It mainly includes the sequential linear iterative clustering algorithm (GSLIC) to obtain superpixels, the clustering of superpixels with a self-generating neural forest (SGNF), and the segmentation of lung parenchyma image sequences. Then, the SGNF, optimized by a genetic algorithm (GA), is used for superpixel clustering. In paper [13], PSO, Genetic Optimization, and SVM Algorithm used for the Lung Cancer Detection technique are proposed. This paper is an extension of image processing using the identification of lung cancer and provides the findings after the segmentation of feature extraction and feature selection. The developed framework recognizes any medical image as an input within the three choices that consist of MRI, CT, and Ultrasound. After the image has been preprocessed, the canny filter is used for Edge detection. Superpixel segmentation was used to segment and Gabor filter was used to de-noise the medical images.

III. Proposed Methodology

The proposed paper focuses on identifying lung cancer from the CT image that detects lung cancer region automatically. It composed of three stages: Preprocessing, CAS segmentation, SVM classifier. Figure 2 shows a flowchart of the proposed architecture.

A. Dataset

The database that was used in this study includes 50 CT scan images of the lung (25 healthy, 25 cancerous). Lung images were taken by a digital CT scanner in a specialist medical imaging center and the dimension of images was 412×412.

B. Preprocessing

Preprocessing is a method for extracting useful information from an image through various operations [14]. Image preprocessing methods are useful for strength improvement, noise deletion, enhancement of an image. The preprocessing steps are Grey Scale Imaging [15], Noise removal, and Adaptive Histogram Equalization. The Preprocessing includes the following steps:

1) Grey Scale Imaging: The input image is the Lung CT image. Images in their default structure are in the Red Green Blue (RGB) structure. These data are complex to process and this is why we turn it into a grayscale image. Grayscale images are logically a matrix with each pixel represented as a level between 0 and 255. The lowest level is the darkest color available while the highest level is the brightest hue.

2) Noise Removal: CT images are having some form of noise. To eliminate unwanted noise from the original image using a median filtering technique. Median filtering [16] is designed to filter out noise that has a corrupted image. They are focused on a mathematical approach. Typical filters are designed to respond at the desired frequency. Median filtering is a nonlinear method frequently used in the processing of photographs to minimize "salt and pepper" noise [17]. The goal is to simultaneously minimize noise and maintain edges. Median filter is more efficient than convolution. It is a nonlinear optical filtering technique that is often used to remove noise.

3) Histogram Equalization: The histogram of an image is a plot between the tones displayed in the image and the number of pixels sharing the respective tones. A histogram that is large at lower levels shows a dark image while an over-exposed image shows the opposite. The method of smoothing out an image's histogram is called histogram equalization [18]. Histogram Equalization results in a smoother and more uniform image compared with the original image. It includes sharper borders and the points highlighted in the image.

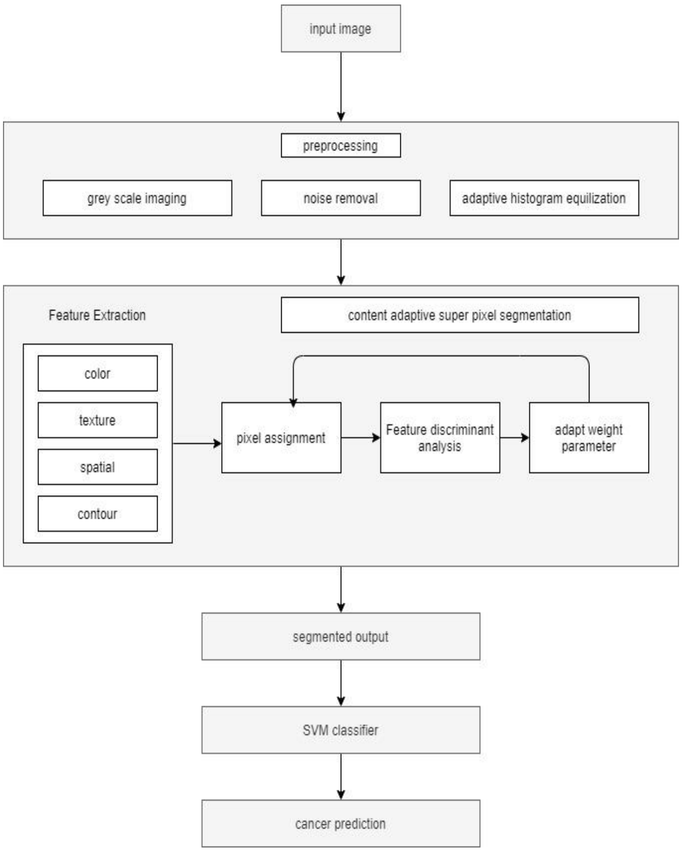

Fig .2: Proposed Architecture

B. Content-adaptive superpixel segmentation

The existing superpixel algorithms have various disadvantages, such as an insufficient representation of attributes, and do not provide discriminability measurements. To solve those problems Content adaptive superpixel (CAS) is proposed. The CAS algorithm is done in the five-dimensional space where [lab] is the color vector of pixels in CIELAB color space and XY is the location of pixels utilized the local characteristics in images To use the Euclidean distance in this 5D space we need to normalize the spatial distances since the maximum possible distance between two colors in the CIELAB space is limited while the spatial distance in the XY plane depends on the size of the image. Therefore, a new distance measure that considers superpixel size was implemented to cluster pixels in this 5D space.

The CAS System the color, space, contours, and texture features of each pixel are extracted from an input image. CAS requires multiple iterations to achieve the final result for segmentation. CAS algorithm as follows.

Algorithm 1: Content- Adaptive Superpixel Segmentation

Step 1: Extracting color, spatial, contour, and texture features.
Step 2: Evaluate different features by using discriminability measures.
Step 3: Adjust the weights of different features on the current pixel.
Step 4: Repeat step 5 to step 9 until several clustering is equal to the specified number of superpixels.
Step 5: Randomly select k cluster center.
Step 6: Calculate the distance between each pixel to each center.
Step 7: Assign the pixel to the nearest center, then generate partition of the image instance.
Step 8: Update the cluster center.
Step 9: For each Feature, Reset the weights.
Step 10: Repeat the process until the clustering process stops.

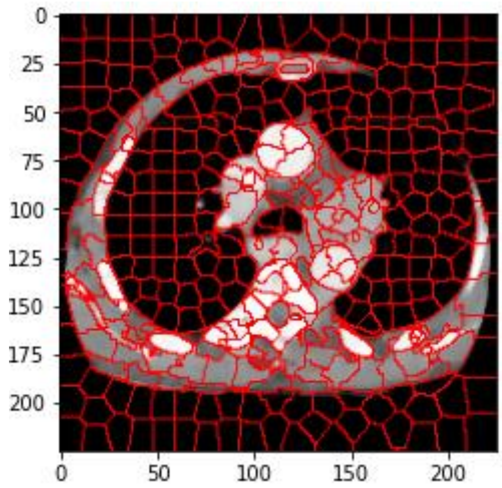

Fig .3: CAS segmentation

C.SVM classifier

A support vector machine (SVM) is a supervised model of machine learning, which uses classification algorithms for classification problems in two classes. In 1995 Vapnik introduced SVM for the classification problem, but in 1997 Vapnik introduced SVM for regression in the AT&T labs. [16].There are mainly two types of classifiers linear and non-linear. Linear classifier means a hyperplane dividing the data into groups. Non-linear classifiers may be used to classify a non-linear dataset. There are a lot of hyperplanes that could classify the data. The best hyperplane is one that shows the biggest difference, or margin, between the two groups. Therefore, select the hyperplane to maximize the distance from it to the nearest data point on each side. If such a hyperplane exists, it is known as the hyperplane with the maximum margin and the linear classifier.

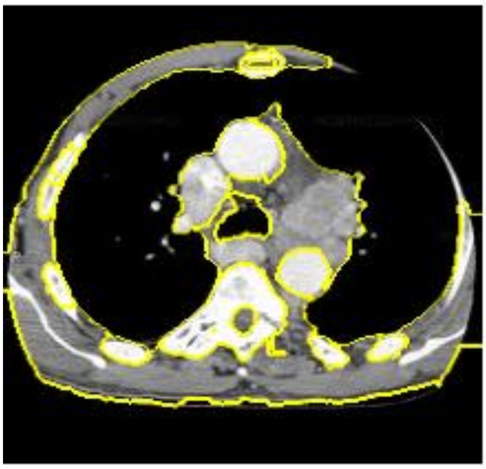

Fig .4: Segmented output

Non-linear classifier renders nonlinearly separable samples per kernel function to high-dimensional feature space and constructs the optimal classification hyperplane in high-dimensional space. SVM uses a kernel function to map the provided information to an individual position. With very complex boundaries, separation can be made. Kernel functions of various types including polynomial, quadratic and Multi-Layer Perceptron (MLP), and so on.

IV. CONCLUSION AND FUTURE WORKS

This paper presented a fully automated method for the segmentation and detection of lung tumors. The proposed model consists of 3 stages: preprocessing, segmentation, and classification. The existing superpixel algorithms have various disadvantages, so we suggest a novel content-adaptive superpixel (CAS) segmentation algorithm combining the feature representation and the discriminability measure. The proposed CAS beats state-of-the-art methods and has low computational costs. The classifier SVM is used to determine whether it is cancerous or not. In the future, our method can be used for detecting different lung diseases.

REFERENCES

[1] Preeti Joon, Aman Jatain and Dr. Shalini Bajaj Bhaskar, "Lung Cancer Detection Using Image Processing Techniques: Review, "IJESC, Volume 7, Issue No: 4, April 2017.

[2] Hui William Chen, Colin D. Medley, Kwame Sefah, Dihua Shangguan, Zhiwen Tang, Ling Meng, Josh E. Smith, and Weihong Tan, "Molecular Recognition of Small-Cell Lung Cancer Cells Using Aptamers",2008.

[3] Holger Hof, Marc Muenter, Dieter Oetzel, Angelika Hoess, Juergen Debus and Klaus Herfarth, MD1 "Stereotactic Single-Dose Radiotherapy (Radiosurgery) of Early Stage Nonsmall-Cell Lung Cancer (NSCLC)", May, 2007.

[4] Image url :"https:/ /www.google.com/ search? q=different + type +of+lung+cancer+images".

[5] Stefano Diciotti, Giulia Picozzi, Massimo Falchini, Mario Mascalchi, Natale Villari, and Guido Valli "3-D Segmentation Algorithm of Small Lung Nodules in Spiral CT Images"IEEE, January 2008.

[6] V. Trillet, D. Revel, V. Combaret, M. Favrot, R. Loire, A. Tabib, J. Pages, P. Jacquemet, A. Bonmartin, J.F. Mornexl, J.F. Cordierl, F. Pinet and M. AmieJ Brunel "Bone marrow metastases in small cell lung cancer: detection with magnetic resonance imaging and monoclonal antibodies,"1989.

[7] Xinhua Qa, Xiaolu Huang, Weili Yan, Lianming Wu and Kerong Dai, "A meta analysis of FDG-PET-CT, FDG-PET-MRI,and bone scintigraphy for diagnosisi of bone metastases in patients with lung cancer",ELSEVIER,2012.

[8] Aarohi Trivedi and Prof.Himani Trivedi "A Comparative Study On preprocessing, Segmentation and Classification Techniques for Cancer Detection,"IJERA, Vol. 8, Issue 10, Oct 2018.

[9] Xiaolin Xiao, Yicong Zhou, and Yue-Jiao Gong, Member, IEEE," Content-Adaptive Superpixel Segmentation," IEEE 2018.

[10] S.Shaik Parveen and C.Kavitha, "classification of lung cancer nodules using SVM kernals", Volume 95–No.25, June 2014.

[11] Wei Zhang, Xiaolong Zhang, Juanjuan Zhao, Yan Qiang, Qi Tian and, Xiaoxian Tang, "A segmentation method for lung nodule image sequences based on superpixels and density-based spatial clustering of applications with noise,"Plos On,September 7,2017.

[12] XiaoleiLiao, JuanjuanZhao, ChengJiao, LeiLei, YanQiang and Qia ngCui, "An Efficient Approach of Image Segmentation for Skin Cancer Detection," August 2016.

[13] A. Asuntha, Neelima Singh and Dr.Andy Srinivasan," PSO, Genetic Optimization and SVM Algorithm used for Lung Cancer Detection," International Journal of Advanced Computer Science and Applications, 2016.

[14] G. Vijaya, A. Suhasini and R. Priya "Automatic Detection of lung cancer in CT images," IJRET, Volume: 03, May-2014.

[15] Md. Badrul Alam Miah and Mohammad Abu Yousuf "Detection of Lung Cancer from CT Image Using Image Processing and Neural Network,"2004.

[16] P.B.Sangamithraa and S.Govindaraju, "Lung Tumor Detection and Classification using EK-Mean Clustering," IEEE, 2016.

[17] Ashwini Kumar Saini, H.S. Bhadauria and Annapurna Singh "A Survey of noise removal methodologies for lung cancer diagnosis", IEEE, 2016.

[18] S.Kalaivani, Pramit Chatterjee, Shikhar Juyal and Rishi Gupta "Lung Cancer Detection Using Digital Image Processing and Artificial Neural Networks", ICECA 2017.

Archa Unnikrishnan is M.Tech scholar in Dept. of CSE, College of Engineering and Management Punnapra affiliated to APJ Abdul Kalam Technological University. She has completed her bachelor degree from Sri Vellappally Natesan College of Engineering in 2017. Her area of interest includes Digital image processing and Machine learning. Email id: archaunni9mtech@gmail.com

Mrs. Remyamol R is working as Assistant Professor in the Department of Information Technology, College of Engineering and Management, Punnapra. Her area of interest Digital image processing.
Email id: remyaremesanm@gmail.com

Suresh Kumar N is working as Assistant professor in the Department of Computer Science Engineering at College of Engineering and Management, Punnapra. His area of interest includes Internet of Things, Digital Image Processing and Machine Learning.
Email id:cnsuresh2000@gmail.com