

# A Novel Approach for Lung Lesion Segmentation Using Optimization Technique

<sup>1</sup>C.Venkatesh, <sup>2</sup>Polaiah Bojja

<sup>1</sup>Research Scholar, <sup>2</sup>Professor

<sup>1,2</sup> Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh., India

Email: <sup>1</sup>venky.cc@gmail.com, <sup>2</sup>paulraj.bojja@gmail.com

Received: 05<sup>th</sup> January 2019, Accepted: 29<sup>th</sup> January 2019, Published: 28<sup>th</sup> February 2019

## Abstract

Cancer is the most prevalent high-flown disease in all countries. In all cancer types, lung cancer is the most mortal disease. Millions of people were die every year because of lung cancer. Early recognition of disease plays a protuberant role in cancer patients. Accurate prophecy of this disease swells up the survival rate. But accurate detection of lung cancer is very critical with the existing systems and also time consuming. To conquer this difficulty the hybrid method is proposed. Accessibility of present technology has proved the way to explore the genes and its alliances in a variety of ailments like lung cancer. In this paper, a hybrid approach is proposed where genetic optimization algorithm is used for detection of cancer in CT images along with SVM classification combined with novel feature selection technique. This method assists the doctors to discern the lung nodules perfectly at early stages.

## Keywords

*Optimization; Alliances; Classification; Cancer; Nodules*

## Introduction

Cancer is the most dangerous disease concerning with abnormal growth of cells. These cells grow up uncontrollably and form as malignant tumors. These tumors spread and affect the organs within the body. Out of all the cancer diseases lung Cancer is utmost perilous diseases that cause a great number of fatalities in both men and women and also it has been recognized as a major health problem in worldwide[1]. According to ACS (American cancer Society), in the year 2018 approximately 14% are lung cancer out of all cancer types [2]. In 2014, lung cancer causes 1.6 million deaths annually and estimates that by 2030 it will attain just about 10 million deaths per year [3]. Early diagnosis of this disease can avert the spread of cancer, fetches better treatment and increases the endurance rate. However, the computed tomography (CT) imaging modality is used to screen the lung cancer. CT scan forms several slices (detailed images) of lung [3]. Moreover, to diagnose lung cancer still manual methods are used in clinical practices which are not as much as effective for radiologists and pathologists.

Therefore, an effective cancer detection system based on image processing is of great interest for evaluating larger dimensional medical data related to disease. Image processing technology is a dynamic diagnostic tool for medical purposes and has been developed very much to investigate the fine details. So to detect lung cancer the image processing technology and efficient optimization techniques are essential.

In this paper a hybrid strategy is proposed to identify lung lesion effectively and exact. The proposed method employs genetic algorithm as a segmentation method that identifies genes (lung nodule) precisely, support vector machine for classification to classify whether the detected part is malignant or not and LBP feature extraction is used to the extract the features of lung cancer. In this work also compared the results of proposed system with DWT technique.

## Literature Review

In 2019 Ananya Choudhury et.al [4] proposed a multi-objective algorithm using genetic for lung cancer segmentation. In this method multi-level thresholding is performed by genetic algorithm and for classification connected component technique is used. In this work they have calculated classification accuracy in terms of true positive and false negative but not segmentation accuracy.

In 2017 Ammar Odeh et.al [5] proposed a method early prediction system for lung cancer detection using genetic. In this work the authors obtained 84% accuracy.

In 2017 Kamil Dimililer et.al [6] proposed a strategy for lung lesion detection using DWT. In this method DWT (Haar) is used for segmentation and extracted 4 output different images which are represented in vertical, horizontal, diagonal and approximation. Finally all the images are processed through techniques like erosion and subtraction to extract the cancer area. By this method authors obtained 89% accuracy.

In 2016 Mukesh Chandra Arya et.al [7] proposed a method to detect mass tissues in Chest X-ray images using DWT. By this method authors obtained 86% accuracy.

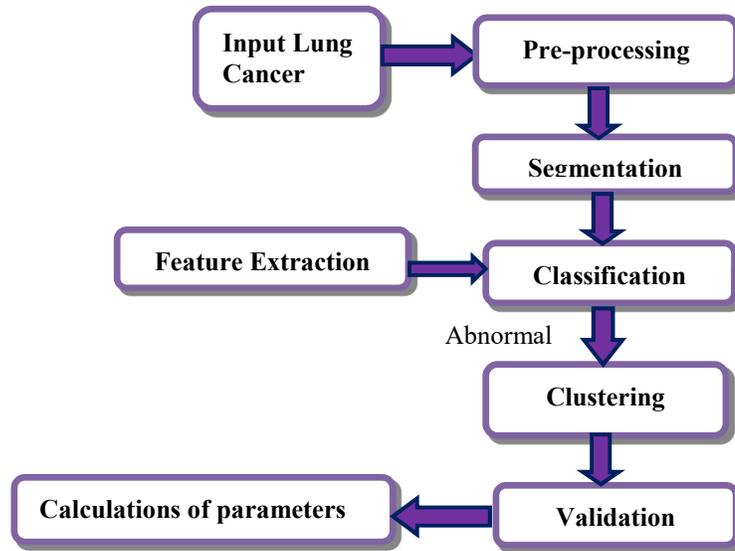
In above two methods [6][7] authors used DWT as segmentation technique. With this technique they obtained less than 90% accuracy. The main drawback is that DWT does not extract fine edges correctly because of poor directionality also it does not provide complete information in all directions.

In 2016 manasee kurkure et.al [8] proposed a strategy for detection and classification of Lung cancer. In this method Naïve Bayesian classifier is used for classification and Genetic Candidate Group search along with canny edge detector is used for segmentation. By this method authors obtained 82% accuracy. In this method Canny edge detector is used for segmentation. It devours more computation time due to its having complex computations also canny detector does not perform a good operation for rotational regularity.

In all the methods which authors have proposed get less accuracy because of limitations present such as poor directionality, less speed-up processing, more computation time and complex computations. So in this work a dynamic method is proposed for detection of lung cancer in CT images where genetic algorithm is used for segmentation and LBP (Local Binary Pattern) for feature extraction to reduce all the drawbacks of the existing methods mentioned above.

### Methodology

In this method the lung cancer is detected and forecast from CT images using dynamic particle swarm optimization method. The processing techniques of proposed method are shown in Fig.1.



**Fig. 1. Proposed Model**

At first the CT image of lung cancer is read from the data base. Usually the acquired image contains low noise and if the noise is removed directly then there is a chance of losing clarity so the included noise is removed by using processing technique. Median filter which is a nonlinear digital filter where it is used to reduce the noise .The enhanced image is then processed through the segmentation by DWT/Genetic algorithm. Then segmented image is processed through the feature extraction technique is known as LBP. The extracted image is then given to the classifier known as SVM which classify whether the lung nodule is benign or malignant. Finally the tumor part is performed clustering operation by using K-means clustering method and different attributes are calculated.

#### A. Image Acquisition

Image acquisition is a process of acquiring a digital image from data base. Generally the images are acquired by different types of scanners like MRI and CT. CT image is acquired from CT scanner. Computed Tomography is an imaging procedure that generates cross sectional images signifying in each pixel [9]. This scan is a Non-Invasive and painless diagnostic tool. It also referred as CAT (computerized axial tomography)

#### B. Preprocessing

Pre-processing is the operation used on images to improve the image data. It reduces undesired distortions and enhances the image relevant features for further processing. Pre-processing techniques can be categorized based on the size of the pixel neighborhood. These methods are realized in image enhancement.

Enhancement operations works on neighborhood image pixels and sub image of corresponding values. Enhancement improves the image quality by reducing the distortion and noise. In most of the bio-medical applications the median filter is utilized to lessen the noise.

#### Median Filtering

The acquired CT images have impulse and salt & Pepper noise. With these noises the fine details are obscured. The median filter reduces impulsive, salt & pepper noise by conserving the boundaries in an image [10].

### C. Segmentation

It extracts objects and boundary features. A challenging problem in medical images is to segment regions with boundary between object (Regions of Interest) and background [11]. To rectify this problem numerous segmentation methods have been proposed by researchers.

### D. Classification

It refers to the undertaking of extricating data classes from a multiband raster picture. The subsequent raster from picture classification can be utilized to make topical maps. Contingent upon the cooperation between the expert and the PC amid classification, there are two sorts of classification: supervised and unsupervised[]. In this work supervised classification i.e. SVM is used. It utilizes the spectral marks got from training samples to characterize a picture.

### E. Clustering

Clustering is a prevalent approach to isolate a huge amount of information from a specific number of gatherings [13]. In the clustering of k-means, the huge amount of information is divided in to k number of individual information and joins such individual information in to disjoint group [14]. K-means involves two discrete phases. In phase 1 it directs centroid of k and in the phase 2 it receives individual point from the cluster which is nearer to the centroid. There are diverse stratagem to describe parting of the closest centroid amongst the most exploited techniques is Euclidean separation. The clustering is performed with the following equations

$$D = \|m(p, q) - C_n\| \quad (1)$$

$$C_n = \frac{1}{n} \sum_{p \in c_n} \sum_{q \in c_n} m(p, q) \quad (2)$$

Where D is Euclidean distance  
n is number of clusters

### Segmentation by DWT & Genetic Algorithm

#### F. DWT:

DWT is mostly utilized here to extricate highlights from a CT by progressive high pass and low pass separating on different scales. The wavelet coefficients are acquired. The segmentation process involves decaying the flag and to get the detailed coefficients and estimation utilizing DWT into N levels with the assistance of separating and obliteration method [15]. Usually in CT images the edges typically gather in the high recurrence parts of the image. Henceforth on the off chance that we can separate high frequencies from the image then get its edge data.

The forward discrete transforms is expressed as:

$$W_f(x, y) = \int_{-\infty}^{\infty} a(t) \varphi_{x,y}(t) dt \quad (3)$$

The inverse discrete transforms is expressed as:

$$a(t) = \frac{1}{D} \int_0^{\infty} \int_{-\infty}^{\infty} w_f(x, y) \varphi_{x,y}(t) dy \frac{dx}{x^2} \quad (4)$$

$$\text{Where } D = \int_{-\infty}^{\infty} \frac{|\varphi\omega|^2}{\omega} \quad (5)$$

#### G. Genetic Algorithm

GA is one type of optimization approach which uses probabilistic methods, motivated by the biological development process. It utilizes ideas of "Natural Selection" and "genetic Inheritance". This algorithm was originally developed by John Holland (1975).GA is especially appropriate for difficult issues where little is thought about the basic inquiry space. A genetic calculation keeps up a populace of competitor answers for the current issue and influences it to advance by iteratively applying a lot of stochastic administrators. It solves Optimization problem with feasible solutions by stochastic operators and these operations iteratively applied on a group of feasible solution.

GA algorithm has numerous targets associated with it. It is related with various stages such as

- The pixels comparisons to chromosomes.
- To perform segmentation and extraction of features

The valid fitness function is calculated by summing all the individual objects in the group sum. For certain hurdles, it is difficult to delineate the fitness function.GA operates based on the characteristics with Selection, mutation, crossover and recombination.

#### Algorithm

- Generate initial population
- If the candidate is not best then
- Generate new population using descendant functions.
- Calculate each candidate fitness function.
- Return to the best solution (candidate).

#### Classification by SVM

SVM is regulated learning model with related learning calculations that dissect information and perceive designs which are utilized for arrangement. The fundamental SVM takes a lot of information from each given point which predicts two classes of shapes the input, making it a non-probabilistic paired straight classifier. From given arrangement of preparing precedents, each

set apart as having a place with one of two classifications, a SVM preparing calculation manufactures a model that doles out new precedents into one classification or the other [17]. In the proposed strategy we are utilizing linear classifier. Best hyper plane is the one that speaks to the biggest partition or edge between the two classes. So we pick the hyper plane with the goal that the separation from it to the closest information point on each side is amplified. On the off chance that such a hyper plane exists, it is known as the most extreme edge hyper plane and the straight classifier it characterizes is known as a greatest classifier. SVM magnifies the edge around the isolating hyper plane. The optimal capacity is completely indicated by a subset of preparing tests. The highest (maximum) margin is formulated by following equations:

$$f(D_i) = \text{sign}(V^T D_i + b) \tag{6}$$

$$F(D_i) = y_i (V^T D_i + b) \tag{7}$$

Where  $F(D_i)$ =Functional margin of  $D_i$

$V$ : Normal vector of decision hyper plane

$D_i$ : Data point  $i$

Useful edge of dataset is double the base practical edge for any point. The factor of 2 originates from estimating the entire width of the edge. Accept that all information is in any event remove 1 from the hyper plane, at that point the accompanying two limitations pursue for a preparation set  $\{(V_i, x_i)\}$

$$V^T D_i + p \geq 1 \quad \text{if } x_i = 1 \tag{8}$$

$$V^T D_i + p \leq -1 \quad \text{if } x_i = -1 \tag{9}$$

For support vectors, the inequality becomes equality. Then, distance and margin from the hyper plane is

$$d = x \frac{V^T D + b}{\|V\|} \tag{10}$$

$$m = \frac{2}{\|V\|} \tag{11}$$

**Future Extraction by LBP**

The fundamental LBP operator was presented by Ojala et al. [18].It was based on the suspicion that the surface has locally two integral viewpoints, a pattern and its quality. In this work, the LBP was proposed as a two-level variant of the surface unit to portray the local textural patterns. The unique adaptation of the local binary pattern administrator works in a  $3 \times 3$  pixel square of an image. The pixels in this square are threshold by its middle pixel value which was multiplied by forces of two and after that summed to get a name for the inside pixel.

Let the black and white image  $i(m, n)$  and let  $Bc$  and  $Bq$  denote the intensity value of an arbitrary pixel and neighborhood sampling point  $Q$  with radius  $D$  respectively then by using the following equations the corresponding pattern is calculated

$$Bc = i(m, n) \tag{12}$$

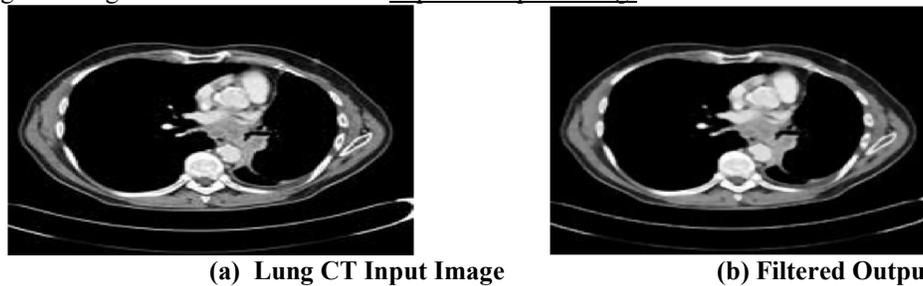
$$Bp = i(x_q, y_q), q = 0 \dots \dots \dots Q - 1 \tag{13}$$

$$m_p = m + R \cos(2\pi p/Q) \tag{14}$$

$$n_p = n - R \sin(2\pi p/Q) \tag{15}$$

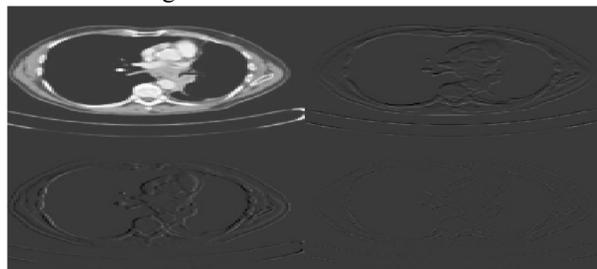
**Results & Discussions**

In this work, CT images of lung cancer are attained from <https://radiopaedia.org/> data base.



**Fig. 2: Input and Filtered Image**

All the acquired images consist of some sort of salt & pepper noise. To confiscate noise and improve the quality, the input image is processed through median filtered as shown in Fig.1.



**Fig. 3: Segmented Image**

To detect the cancer in CT image, the output of the median filter processed through Segmentation. Segmentation technique extracts the smooth edges as shown in Fig.2.

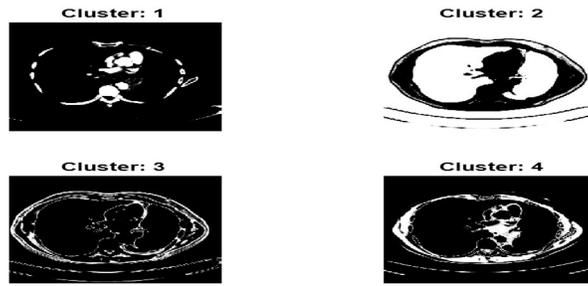


Fig. 4: Grey Level Metrics evaluation of Input Image -1

After obtain smooth edges from segmentation it performs clustering which manages finding a structure in an accumulation of unlabeled information as shown in fig.4.

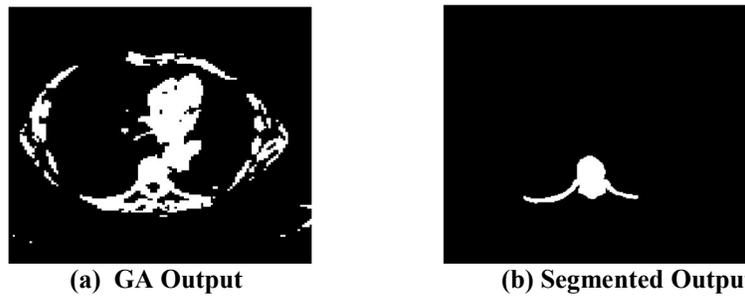


Fig. 5.Genetic and Segmented Output

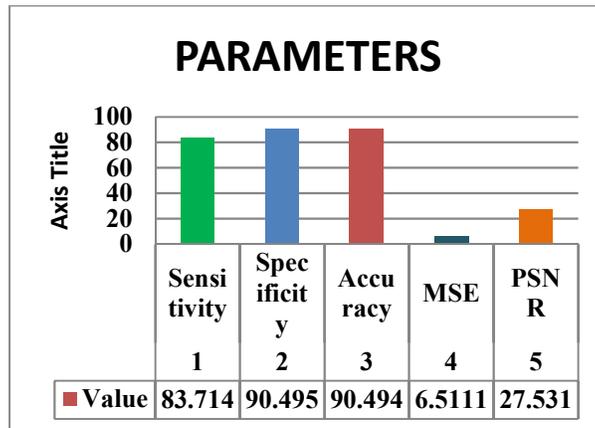
To extract the ROI the cluster output processed through genetic algorithm to identify exact cancer location. The GA output and the cancer segmented output is as shown in fig5.Finally the validation is performed and the corresponding parameters are determined. The parameters as shown in Table.1.

S.NO	Parameter	Value
1	Sensitivity	83.7143
2	Specificity	90.4950
3	Accuracy	90.4937
4	MSE	6.5111
5	PSNR	27.5311

TABLE I Obtained Parameters

S.NO	Method	Accuracy
1	Ammar Odeh et.al [5]	84%
2	Kamil Dimililer et.al [6]	89%
3	Mukesh Chandra Arya et.al [7]	86%
4	Manasee kurkure [8]	82%
5	Proposed	90%

TABLE II Comparison Results



From the Table 1 and Table 2 the accuracy of the proposed is increased compared to the existing methods.

**Conclusion**

In this paper, a novel strategy is proposed to recognize the cancer extant in CT images of Lung. This strategy utilizes Genetic algorithm and obtained 90% accuracy. Thus the proposed framework performs fine and is vigorous against anatomical varieties

of the lungs. This work demonstrated a productive and exact outcome for dividing the lung tumor in CT therapeutic images and it will help radio-significance as second assessment for the better analysis of lung malignant growth.

### Acknowledgement

The First author express sincere thanks to management of K L Deemed to be University, Vaddeswaram, Guntur, Andhra Pradesh, where he is a research scholar and Annamacharya Institute of Technology and Sciences, Rajampet, A.P. for providing good research facilities.

### References

- [1] <https://www.cancer.org/cancer/lung-cancer/prevention-and-early-detection.html>
- [2] Md Rashidul Hasan, Muntasir Al Kabir, “Lung Cancer Detection and Classification based on Image Processing and Statistical Learning”, Research gate, Pg.No:1-6
- [3] Joey Mark Diaz, Raymond Christopher Pinon, Geoffrey Solano, “Lung Cancer Classification Using Genetic Algorithm to Optimize Prediction Models”, IEEE 5th International Conference on Information, Intelligence, Systems and Applications, 2014, Pg.No:1-6
- [4] Ananya Choudhury, Rajamenakshi, R. Subramanian and Gaur Sunder, “A Novel Approach for Tumor Segmentation for Lung Cancer Using Multi-objective Genetic Algorithm and Connected Component Analysis” Proceedings of the 2nd International Conference on Data Engineering and Communication Technology, Advances in Intelligent Systems and Computing Springer Nature Singapore, 2019 Pg.No:367-376
- [5] Ammar Odeh, Ibrahim Al Atoum, Abraham Bustanji, “Novel Genetic Algorithm for Early Prediction and Detection of Lung Cancer” Journal of Cancer Treatment and Research, Volume :5, Issue:2, 2017, Pg.No: 15-18
- [6] Kamil Dimililer, Ali Hesri, Yoney Kirsal Ever, “Lung Lesion Segmentation Using Gaussian Filter and Discrete Wavelet Transform”, ITM Web of Conferences, volume:11, Issue:01018, 2017, Pg.No:1-10
- [7] Mukesh Chandra Arya, Dr. Bhumika Gupta, “Detect Mass Tissue in Lung Images Using Discrete Wavelet Transformation”, IEEE international Conference on Information Processing (IICIP), 2016, Pg.No:1-10
- [8] Manasee Kurkure, Anuradha Thakare, “Lung Cancer Detection using Genetic Approach, IEEE International Conference on Computing Communication Control and automation (ICCCBEA), 2016, Pg.No:1-5
- [9] C.Venkatesh, Polaiiah Bojja, “An Investigation Of Diverse Optimization Techniques On Medical Imagery For Detection Of Perilous Diseases” Fronteiras: Journal of Social, Technological and Environmental Science, volume:6, Issue:2, May-August. 2017, Pg.No: 249-255, ISSN 2238-8869 249
- [10] C.Venkatesh, Polaiiah Bojja, “An Exploration of Optimization Techniques for Detection of Lung Cancer in CT Images”, International Journal of Pure and Applied Mathematics, Volume 117 Issue No 18, 2017, Pg.No: 379-384
- [11] C.Venkatesh, K. Bhagyalakshmi, L. Sivayamini, “Detection of Diverse Tumefactions in Medial images by Various Cumulation Methods” International Research Journal of Engineering and Technology (IRJET), Volume: 04, Issue: 08, Aug 2017, Pg.No:1195-1200,
- [12] S. Perumal, T. Velmurugan, Lung cancer detection and classification on CT scan images using enhanced artificial bee colony optimization”, International Journal of Engineering & Technology, volume:7, Issue:2.26, 2018, Pg.No:74-79
- [13] Nameirakpam Dhanachandra\*, Khumanthem Manglem and Yambem Jina Chanu, “Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm”, Procedia Computer Science, Science Direct, Elsevier, Issue: 54, 2015, Pg.No: 764 – 771
- [14] Shehroz S. Khan and Amir Ahmad, Cluster Centre Initialization Algorithm for K-means Cluster, In *Pattern Recognition Letters*, 2014, pg.No.: 1293–1302
- [15] Nuhman ul Haq, Khizar Hayat, Syed Hamad Sherazi, William Puech, “Segmentation Through DWT and Adaptive Morphological Closing,” 9th European Signal Processing Conference, 2011, Pg.No:31-35
- [16] P. Bhuvaneshwari, Dr. A. Brintha Therese, “Detection of Cancer in Lung With K-NN Classification Using Genetic Algorithm”, Procedia Materials Science, Elsevier, Issue:10, 2015, Pg.No: 433 – 440
- [17] Chunhong Lu & Zhaomin Zhu & Xiaofeng Gu, “An Intelligent System for Lung Cancer Diagnosis Using a New Genetic Algorithm Based Feature Selection Method”, Systems-Level Quality Improvement, Springer, J Med Syst, 2011, Pg.No: 38-97
- [18] T. Ojala, M. Pietikainen and D. Harwood, “A comparative study of texture measures with classification based on feature distributions” Pattern Recognition vol. 29, 1996.