

Early diagnosis of Lung Cancer with Probability of Malignancy Calculation and Automatic Segmentation of Lung CT scan Images

Dr. Samuel Manoharan,

Professor,
Department of Electronics,
Bharathiyar College of Engineering and Technology,
Thiruvettakudy, Karaikal, India
Email id: jsamuel@bcetedu.in

Prof. Sathish,

Department of EEE,
Eritrea Institute of Technology,
Eritrea
Email id: sathesh4you@gmail.com

Abstract: Computer aided detection system was developed to identify the pulmonary nodules to diagnose the cancer cells. Main aim of this research enables an automated image analysis and malignancy calculation through data and CPU infrastructure. Our proposed algorithm has improvement filter to enhance the imported images and for nodule selection and neural classifier for false reduction. The proposed model is experimented in both internal and external nodules and the obtained results are shown as response characteristics curves.

Keywords: lung cancer, computed tomography, segmentation.

1. Introduction

In recent years, there has been a rapid increase in death rate due to lung cancer. It is a challenge for the radiologists to predict the pulmonary nodule growth earlier [4]. This causes major problem in many of the cases, since the CT scan is high in cost most of the rural patients are getting affected.

Calculation of malignancy probability is predicting the lung cancer early. Lung cancer is one of the deadliest diseases among dangerous diseases in the world. Despite significant research efforts and advances in the understanding of tumor biology, there was no reduction of the mortality over the last decades. More researches are going on about early diagnosis of lung cancer recently [6] [7].

Many conservative methods are focusing to select the nodule by enhancement filter only. This article algorithm is constructed neural network classification for reduction of false positive identification including of enhancement filter [8][11][12][19].

The motivation behind this paper is to reduce the challenges of the radiologists and to detect the early growth rate of pulmonary nodule by calculating probability of malignancy and segmented of lung nodule.

2. Related work

A vast survey has been made in lung nodule analysis and using image processing techniques like radiography [3], computed tomography [1] and MRI. Segmentation method utilizes the tissue intensity properties for the lung images. Various researchers are in progress under image processing identifying the discrimination and spatial resolutions [10] of the tissues. Advantages of the tissue segmentation are done by many statistical methods. Expectation maximization algorithm segments MR images for better visualization. Some researcher's uses adaptive segmentation [14] for obtaining better gain for the MR images. The issues and the gain field are calculated using the segmentation process. Three step segmentation processes is used in analysis of lung nodule such as ground glass opacity, region of interest and preprocessing for better results.

Research works [5] [9] proposes a multi joint segmentation solution for fitting the mean based on linear space theory. Pulmonary modules are geometrical in structure and pulmonary nodule in the CT is used to obtain the details of location, approximations, and region volumes. The issues found in the research highlights in intensity distribution based on the robustness. Threshold segmentation [18] and histogram analysis [16] some other approaches used in image segmentation. The segmented lungs are refined using component connect labeling and circular morphological factors. A three dimensional adaptive fuzzy based segmentation [15] and the segmented contour are smoothed using Gaussian smoothening filter. In this an MR image is process for contrast enhancement and this research work is needed to obtain the issues present in the MR imaging system.

Cellular automata technique is based on graph theory which provides solution for shortest path problem and the spatial smoothness can be accomplished. Neural network based computer aided diagnosis system are also in trend for obtaining the best results for the low resolution images. This type of researches uses hidden layer and output layer for extracting the tumor portion from the image. The detection of SNAs is performed by hidden layer present in s ANN. optimization techniques are also available for obtaining the best results in image classification such optimization techniques are not implemented in CT images. Based on the observation from various articles this proposed model is developed and experimental results are compared with conventional methods.

3. Proposed Method

The preprocessing algorithm has written and it is followed by the proposed algorithm for the segmentation.

Step 1 (Preprocessing)

Erode and then dilate the original image to get the Opening image from the input images. In order to get the enhanced image, superimpose the original and top hat image and subtract the bottom hat image from the output image.

Step 2: Input lung image (512x512 gray scale images)

2(a) Read and show (P_9) (read the preprocessed output image)

2(b) Convert $P_9 \Rightarrow P_{10}$, where P_{10} (0s & 1s) and P_9 (0s & 255s),

Now P_{10} (0s & 1s) $\Rightarrow P_{11}$ (1s & 0s) (convert the gray scale image into binary Image)

Step 3: Distance transform and Superimpose

$$3(a) \quad \text{Take } P_{12} (1s \ \& \ 0s) \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \Rightarrow P_{13} (0s \ \& \ 1s) \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$3(b) \quad P_{14} \Rightarrow \sim (P_{13})$$

$$3(c) \quad P_{15} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Leftarrow P_{14} \begin{pmatrix} 2 \\ 1 & 0 & 4 \\ 3 \end{pmatrix}$$

$$3(d) \quad P_{16} \Leftarrow P_9 + f(P_{15}); \text{ where } f(P_{15}) \Rightarrow \sim P_{15}$$

Step 4: Gradient and Superimpose

$$4(a) \quad P_{17} \Leftarrow \text{sobel } (P_{16})$$

$$4(b) \quad P_{17} \begin{pmatrix} 2 \\ 7 & 0 & 9 \\ 8 \end{pmatrix} \Rightarrow P_{18} \begin{pmatrix} 0 \\ 0 & 0 & 0 \\ 0 \end{pmatrix}$$

Correctness:

$$\text{Assign } P_{18} = \begin{pmatrix} 20 & 20 \\ 20 & 0 & 20 \\ 20 \end{pmatrix}$$

$$4(c) \quad \text{To get } P_{19}: \text{ consider } P_{18} > a; \text{ where } a \Rightarrow 255, \text{ so } P_{19} \Leftarrow P_{18} > 255$$

$$4(d) \quad P_{20} \Leftarrow P_{17} + P_{18}$$

Step 5: Marker control

$$5(a) \quad \text{Extended minima for the image } (P_{20}, h) \text{ where } (h: \text{ non-negative scalar})$$

$$P_{21} \Leftarrow \begin{pmatrix} \max & \max & \max & \max & \max \\ \max & \text{const} & \text{const} & \text{const} & \max \\ \max & \text{const} & \text{const} & \text{const} & \max \\ \max & \text{const} & \text{const} & \text{const} & \max \\ \max & \max & \max & \max & \max \end{pmatrix}$$



26-neighbourhood connected for 3-D image. Do iterations $P_{21}(1), P_{21}(2), P_{21}(3), \dots, P_{21}(n)$ until n reaches max.

$$5(b) \quad P_{22} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \leq P_{21} \begin{pmatrix} 9 & 9 & 9 & 9 & 9 \\ 9 & 3 & 5 & 2 & 9 \\ 9 & 1 & 2 & 8 & 9 \\ 9 & 7 & 6 & 4 & 9 \\ 9 & 9 & 9 & 9 & 9 \end{pmatrix}$$

Step 6: Segmentation

$$6(a) \quad P_{22} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow P_{13} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

6(b) P_{14} segmented output image $\leq P_{13}$

4. Theoretical Analyses

Theoretically, the comparison of different methods of segmentation demerits and merits includes Graph cut segmentation, Level set segmentation, Region splitting segmentation, Thresholding + region growing, Thresholding + morphology Thresholding + ball algorithm with the Proposed algorithm is shown in table 1 below [13], [14], [15], [16], [17]. The early detection of the nodule is not fulfilled in all the methods, so the proposed algorithm has been derived to overcome this drawback.

Table 1: comparison of methods

Methods	Advantages/ disadvantages
Graph cut segmentation	Original segmented image can be detected. Discontinuities between voxels are the disadvantage.
Level set segmentation	Vector method not a scalar; represent the evolving region and shapes. Contraction and expansion is the limitation.
Region splitting segmentation	Instead of using fixed parameters, dynamic methods are used. Method failed for some patients.



Thresholding + region growing	To extract the exact cavity region with accuracy. Results are not satisfied, does not work when there is overlapping of intensities.
Thresholding + morphology	Segmenting the thorax area of the image. The limitation is the selection of size of the ball.
Thresholding + ball algorithm	Segment the thorax area of the image. Segmentation includes unnecessary areas is the drawback.
Proposed algorithm	To extract the nodules from lung images, segmentation is accurate, and early detection is feasible.

5. Performance Analysis

Due to measure the performance of the system, the probability of malignancy has been calculated based on the clinical and radiographic property parameters. These property parameters are explained enough from the past research [18]. In order to calculate the probability of malignancy using the above mentioned parameters, Likelihood ratio and Posterior, Prior odds are used in the Baye's theorem [20], [21].

Posterior odds

Odds are method of expressing probability which is determined after consideration of the final output of the process [22]. Particularly an odds ratio is practiced with case based studies, predict the particular event happening in one group in relation to its frequency of chances in another group.

$$\text{odds}_{\text{posterior}} = \text{odds}_{\text{previous}} \times \text{likelihood ratio}$$

Posterior odd tests consider the following conditions;

- 1) Disease prevalence
- 2) The patient pool
- 3) Specific patient risk factors

Probability ratio

This process is contained sensitivity and specificity of the process. These parameters can give the prediction of how long a test result will change the chances of having a disease. It named likelihood ratio is classified into positive LR and negative LR as follows,

- 1) positive LR = $\frac{\text{sensitivity}}{1-\text{specificity}}$
- 2) negative LR = $\frac{1-\text{sensitivity}}{\text{specificity}}$



Mathematically, the sensitivity and specificity is defined as,

$$\text{sensitivity} = \frac{\text{Number of true positives}}{\text{total number of true positive and false negative}}$$

$$\text{specificity} = \frac{\text{Number of true negatives}}{\text{total number of true negative and false positive}}$$

Finally, this sensitivity provides the probability of positive test that the patient is having infection. Specificity provides the probability of a negative test that the patient is good [23].

6. Results and Discussion

Case 1

Table 2 clearly shows that the probability of malignancy decreases rapidly and reaches the minimum probability within three months. The age of patient considering in this research is 50 years old, non-smoker and free from hemoptysis.

Table 2: Probability calculation of a non-smoker having constant nodule size

P(A) in %	Age	Smoking	Hemoptysis	Previous malignancy	Size	Location	Edge	Growth rate	month	P(B) in %
50	50-59	No	Absent	Absent	1cm	Upper/middle	smooth	Benign	1	11
11	50-59	No	Absent	Absent	1cm	Upper/middle	smooth	Benign	2	1
1	50-59	No	Absent	Absent	1cm	Upper/middle	smooth	Benign	3	0

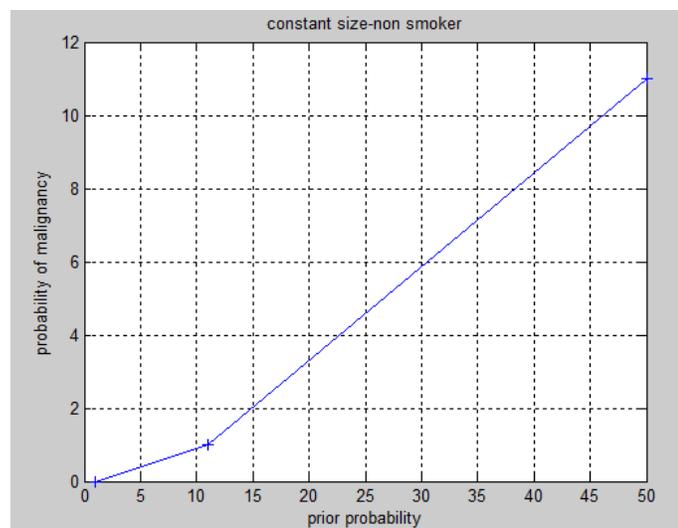


Fig 1: Prior probability vs probability of malignancy-nonsmoker

Fig 1 is plotted for a non-smoker between prior probability and probability of malignancy. It is clear that a non-smoker has rapid decrease in probability of malignancy with respect to the decrease in prior probability.

Table 3 clearly shows that the probability of malignancy increases gradually and reaches the maximum probability within ten months. The patient considered here is 50 years old, smoker and free from hemoptysis.

Table 3: Probability calculation of smoker having constant nodule size

P(A) in %	Age	Smoking	Hemoptysis	Previous malignancy	Size	Location	Edge	Growth rate	month	P(B) in %
50	50-59	yes	Absent	Absent	1cm	Upper/middle	smooth	Benign	1	61
61	50-59	yes	Absent	Absent	1cm	Upper/middle	smooth	Benign	2	71
71	50-59	yes	Absent	Absent	1cm	Upper/middle	smooth	Benign	3	80

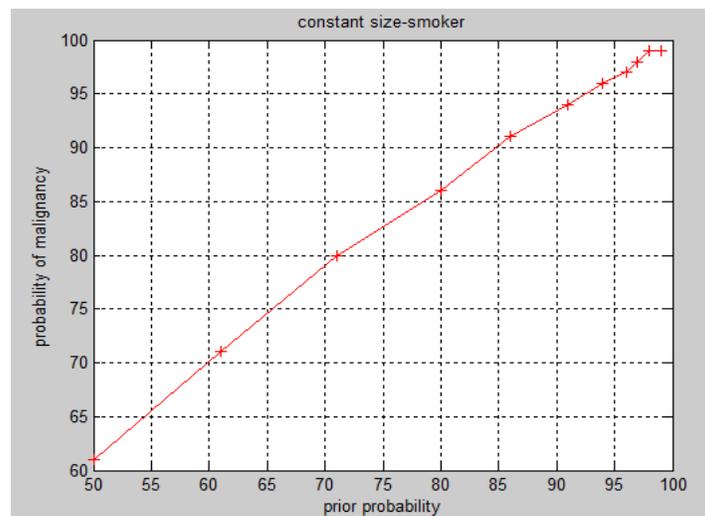


Fig 2: Prior probability vs probability of malignancy-smoker

Fig 2 is plotted for a smoker between prior probability and probability of malignancy. It is clear that a smoker has gradual increase in probability of malignancy respecting the increase in prior probability.

Case 2

Table 4 clearly shows that the probability of malignancy decreases rapidly and reaches the minimum probability within three months. The patient considering here is 50 years old, non-smoker and free from hemoptysis.



Table 4: Probability calculation of a non-smoker having constant nodule size

P(A) in %	Age	Smoking	Hemoptysis	Previous malignancy	Size	Location	Edge	Growth rate	month	P(B) in %
50	50-59	No	Absent	Absent	1cm	Upper/middle	smooth	Benign	1	11
11	50-59	No	Absent	Absent	1cm	Upper/middle	smooth	Benign	2	1
1	50-59	No	Absent	Absent	1cm	Upper/middle	smooth	Benign	3	0

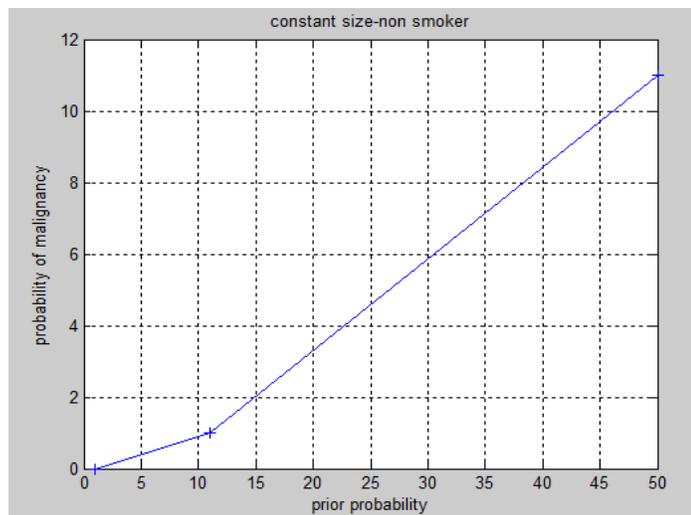


Fig 3: Prior probability vs probability of malignancy-nonsmoker

Table 5: Probability calculation of smoker having variable nodule size

P(A) in %	Age	Smoking	Hemoptysis	Previous malignancy	Size	Location	Edge	Growth rate	month	P(B) in %
50	50-59	yes	Absent	Absent	1cm	Upper/middle	smooth	Benign	1	61
71	50-59	yes	Absent	Absent	2cm	Upper/middle	smooth	Benign	3	85
85	50-59	Yes	Absent	Absent	3cm	Upper/middle	smooth	Benign	4	98

Table 5 clearly shows that the probability of malignancy increases and reaches the maximum probability within five months.

Fig 4 is plotted for a smoker between prior probability and probability of malignancy. It is clear that a non-smoker has gradual increase in probability of malignancy with respect to the decrease in prior probability.

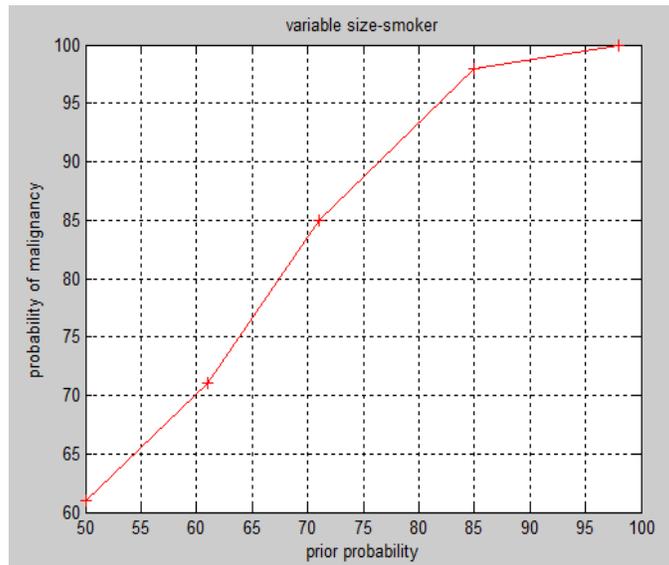


Fig 4: Prior probability vs probability of malignancy-smoker.

Fig 5 is plotted for a smoker between prior size and probability of malignancy. It is clear that a non-smoker has linear increase in probability of malignancy with respect to the decrease in prior probability.

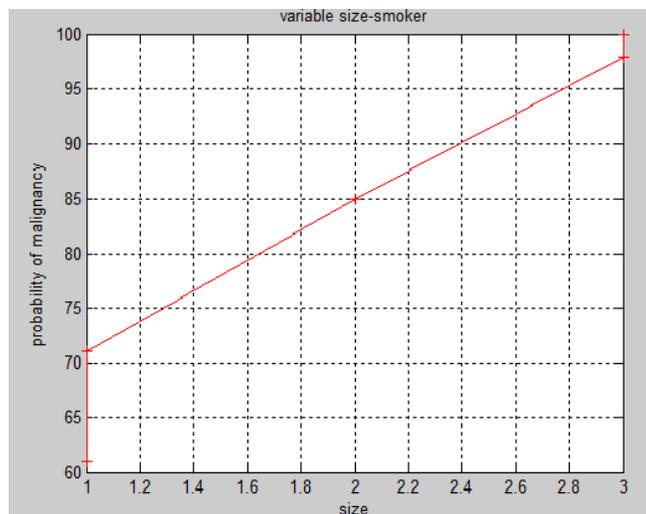


Fig 5: size vs probability of malignancy-smoker

Output figures

The output of the enhanced and segmented images is shown below. It is clear that the nodule in the input image is slightly growing in the sequence of images. The enhanced image is the output of preprocessing in order to increase the visual and the output segmented image clearly segments the nodule part in all the cases.

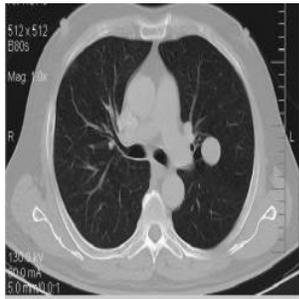


Fig 6



Fig 7



Fig 8

Fig (6) original image of size 1cm (smoker) (7) original image of size 2cm (smoker) (8) original image of size 3cm (smoker)

The original lung CT scan image of the patient with increasing nodule size of 1cm, 2cm, and 3cm of consequent images is shown in fig 6, 7, 8 respectively. The images records that the nodule growth is minimum.

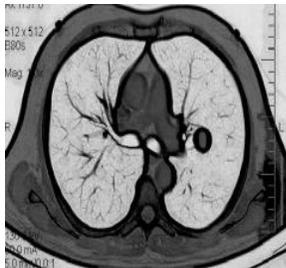


Fig 9

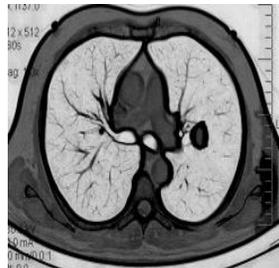


Fig 10

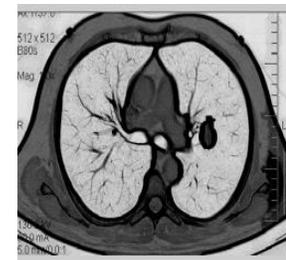


Fig 11

Fig (9) enhanced image of size 1cm (smoker) (10) enhanced image of size 2cm (smoker) (11) enhanced image of size 3cm (smoker)

The enhanced lung CT scan image of the patient with increasing nodule size of 1cm, 2cm, and 3cm of consequent images is shown in fig 9, 10, 11 respectively. This enhanced image is to increase the contrast of the image by preprocessing algorithm as mentioned in step 1. The images are now clearly borders the nodule.



Fig 12

Benign nodule



Fig 13

Malignant nodule

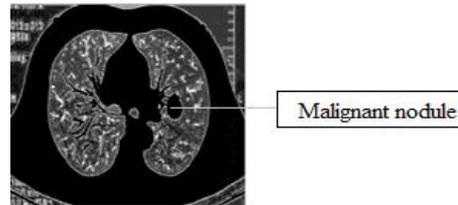


Fig 14

Fig (12) segmented image of size 1cm (smoker) (13) segmented image of size 2cm (smoker) (14) segmented image of size 3cm (smoker)

The segmented output of lung CT scan image of the patient with increasing nodule size of 1cm, 2cm, and 3cm of consequent images is shown in fig 12, 13, 14 respectively. This segmented image has been obtained by the proposed algorithm; the benign and malignant nodules are marked and shown above.

7. Conclusion

From Case 1: (a) The probability of malignancy decreases for the non-smoking patient when the nodule size remains constant. (b) The probability of malignancy increases gradually for the smoking patient even though the nodule size remains constant. So, it is quite confusing because the probability of malignancy seems it doesn't depend on size of the nodule. Further cases have been considered. Case 2: (a) The probability of malignancy decreases for the non-smoking patient when the nodule size remains constant. (b) The probability of malignancy increases rapidly for the smoking patient when the nodule size increases. It is saying that the size of the nodule will directly proportional to probability of malignancy.

Since the size plays the major role in the probability of malignancy. Growth rate of nodule must be detected as soon. This proposed algorithm has been implemented for segmentation and early detection of the nodule. This research will focus deep learning methods as a future enhancements in order to develop the efficient and accuracy of the process. The development of algorithm mainly focuses for better improvement in speed of the process.

Acknowledgement

We would like to thank the Consultant Diagnostic Radiologist & Sonologist Dr. V. Prabakarakumar MBBS, DMRD of Best Scan Centre Coimbatore and Shanmugha priya for his great support to complete this paper. We want to extend our thank the Dr. Yassir Babiker Hamdan, Sudan who comments have greatly improved the paper.

References

1. Lee,S,L,A., Kouzani,A,Z., Hu,E, J.: Automated Detection of Lung Nodules in Computed Tomography Images: a review. In: Machine Vision and Applications. vol. 23, pp. 151-163. Springer (2012)
2. Ezhil E, Nithila.,S.S. Kumar,S, S.: Segmentation of Lung Nodule in CT Data using Active Contour Model and Fuzzy C-mean Clustering. In: Alexandria Engineering Journal. vol. 55, pp. 2583-2588. Elsevier(2016)
3. Nascimento, L,B., De Paiva, A, C., Silva, A, C.: Lung nodules classification in CT images using Shannon and Simpson diversity indices and SVM. In: Machine Learning and Data Mining in Pattern Recognition. pp. 454-466. (2012)
4. Mehdi Aliloub.,Vassili Kovalev., Eduard Snezhko.,Vahid Taimouri .:A Comprehensive Framework for Automatic Detection of Pulmonary Nodules in Lung CT Images. In: Image Anal Stereol. vol. 33, pp. 13-27. (2014)
5. Wook-Jin Choi.,Tae-Sun Choi .: Automated Pulmonary Nodule Detection System in Computed Tomography Images: A Hierarchical Block Classification Approach. In: Entropy. vol. 15, pp. 507-523. (2013)

6. Ezhil E, Nithila.,S.S. Kumar,S, S.: Automatic detection of solitary pulmonary nodules using swarm intelligence optimized neural networks on CT images. In: Engineering Science and Technology. vol. 20, pp. 1192–1202. (2017)
7. Mehrdad Moghbel.,Syamsiah Mashohor.,Rozi Mahmud.,Iqbal Bin Saripan. M. :Automatic Liver Tumor Segmentation on Computed Tomography for Patient Treatment Planning and Monitoring. In: EXCLI Journal. vol.15, pp. 406-423. (2016)
8. Alex Martins Santos., Antonio Oseasde Carvalho Filho., Aristofanes Correa Silva., Anselmo Cardoso de Paiva., Rodolfo Acatauassu Nunes., Marcelo Gattass. : Automatic Detection of Small Lung Nodules in 3DCT Data using Gaussian Mixture Models, Tsallis Entropy and SVM. In: Engineering Applications of Artificial Intelligence. vol.36, pp. 27–39. Elsevier (2014)
9. Juan Juan Zhao., Guo Hua Ji.: Cavitory Nodule Segmentation in Computed Tomography Images Based on Self-Generating Neural Networks and Particle Swarm Optimisation. Int. J. Bio-Inspired Computation. vol. 7, pp. 62-67. (2015)
10. Goncalves L., Novo,J., Cunha,A., Campilho. : Evaluation of the Degree of Malignancy of Lung Nodules in Computed Tomography Images. In: International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP), pp. 74-80. (2017)
11. Pablo G, Cavalcanti., Shahram Shirani., Jacob Scharcanski., Crystal Fong., Jane Meng., Jane Castelli., David Koff. : Lung Nodule Segmentation in Chest Computed Tomography using a Novel Background Estimation Method. In: Quant Imaging Med Surg. Vol.6, pp. 16-24. (2016)
12. Temesguen Messay., Russell C, Hardie., Timothy R, Tuinstra. : 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. In: Medical Image Analysis. vol.22, pp. 48–62. Elsevier (2015)
13. Amal A, Farag., James H, Graham. :A Novel Approach for Lung Nodules Segmentation in Chest CT Using Level Sets. In: IEEE Transactions on Image Processing, vol. 22, pp. 5202-5213. (2013)
14. Qiu Shi.,Wendesheng.,Cui Ying.,Feng Jun. : Lung Nodules Detection in CT Images Using Gestalt-Based Algorithm. In: Chinese Journal of Electronics vol.25, pp.711-718. (2016)
15. Stefano Diciotti., Simone Lombardo., Massimo Falchini., Giulia Picozzi., Mario Mascialchi. : Automated Segmentation and Refinement of Small Lung Nodules in CT Scans by Local Shape Analysis. In: IEEE Transactions on Biomedical Engineering, vol. 58, pp.3418-3428. (2011)
16. Shaik Parveen,S., Kavitha, C. : Segmentation of CT Lung Nodules using FCM with Firefly Search Algorithm. In: IEEE 2nd International Conference on Innovations in Information Embedded and Communication Systems ICIIECS'15. (2015)
17. Hang Zhang., Xi Wang., Parisa Memarmoshrefi.,Dieter Hogrefe. : A Survey of Ant Colony Optimization Based Routing Protocols for Mobile Ad Hoc Networks. In: IEEE Access. vol.5, pp. 24139-24161. (2017)
18. Meng Ma.,Chuang Sun.,Xuefeng Chen. : Discriminative Deep Belief Networks with Ant Colony Optimization for Health Status Assessment of Machine. In:IEEE Transactions on Instrumentation and Measurement, vol.66, pp.3115-3125.(2017)
19. Feifei Zheng., Aaron C, Zecchin., Jeffery P, Newman., Holger R, Maier., Graeme C, Dandy. : An Adaptive Convergence-Trajectory Controlled Ant Colony Optimization Algorithm with Application to Water Distribution System Design Problems. In: IEEE Transactions on Evolutionary Computation. vol.21, pp.773 -791.(2017)
20. Michalis Mavrouniotis., Felipe M, Müller., Shengxiang Yang. : Ant Colony Optimization with Local Search for Dynamic Traveling Salesman Problems. In:IEEE Transactions on Cybernetics. vol.47, pp. 1743-1756. (2017)
21. Yongjun Sun., Wenxin Dong., Yahuan Chen. : An Improved Routing Algorithm Based on Ant Colony Optimization in Wireless Sensor Networks. In: IEEE Communications Letters. vol.21, pp. 1317 - 1320. (2017)
22. Shun-Hung Tsai., Yu-Wen Chen. : A Novel Fuzzy Identification Method Based on Ant Colony Optimization Algorithm. In: IEEE Access. vol.4, pp. 3747 - 3756. (2016)
23. Xu Huang., Raul Fernandez-Rojas., Keng-Liang Ou. : Signal Processing of Brain Activity with Ant Colony Optimization and Wavelet Analysis using Near Infrared Spectroscopy. In: IEEE Sixth International Conference on Communications and Electronics. Pp. 306 - 311. (2016)

