

COMPUTERIZED SYSTEM DESIGN FOR THE DETECTION AND DIAGNOSIS OF LUNG NODULES IN CT IMAGES

¹ALI ABDRHMAN UKASHA, ²EMHMED SAAID ALFAKHRY, ³RASIM AMER ALLI,
⁴MARUKAH EDREES FADEL

^{1, 3, 4}Electrical and Electronics Engineering Department,
²Sebha Medial Center-Specialist of Radiology Department
Sebha University, Sebha, Libya

Email: ¹ali.ukasha@sebhau.edu.ly, ²saied10@yahoo.com, ³ras.alsanbani@sebhau.edu.ly,
⁴kookfadel@gmail.com

ABSTRACT

In recent years, image processing techniques have become widely used in many medical fields to improve image quality in early detection and treatment stages, as time is a very important factor to detect and detect abnormal cells in the target image, especially in many cancerous tumors such as lung cancer, breast cancer, and many more. In this paper, a method was proposed to detect and diagnose lung nodes (malignant and benign) from CT images. The detection stage is divided into two parts: - 1) The first part is obtained by the lung image of the CT-SCAN image of the chest area, lung using the vehicle-related learning algorithm (CCL) and a number of morphological processes, 2) The second part involves obtaining the lung nodules using the properties of the anatomical characteristics and then using the characteristics of the shape (geometric characteristics) of the lung nodules. The diagnosis of the nodules (identification of benign and malignant nodules) was based on the characteristics of the edges of the discovered nodules. The MATLAB programming language was used to design the algorithm for the proposed method, and an easy-to-use graphical interface was designed to display the results to the user. Experimental results show that the system accuracy exceeds to the 77.27% for the 22 tested CT scan images.

Keywords: lung nodules detection; thresholding; morphological processes; CCL; nodules features extraction; computer-aided design (CAD); tumor metastasis; CT scan;

1. INTRODUCTION

In this section, lung cancer is the disease of abnormal cell proliferation and its growth randomly to form the tumor (cancer). Lung cancer mortality is the highest among all other types of cancer [1]. Lung cancer is one of the most dangerous cancers in the world, with the lowest rate of survival after diagnosis, nearly 20% of lung nodules cases represent lung cancers. Thus, the identification of malicious lung nodules is necessary for examine and diagnosing the lung cancer [2]. CT scan is the most accurate imaging method for obtaining anatomical information about lung nodules and surrounding structures [3]. In current clinical practice, CT images interpretation is a challenge for radiologists due to the large number of CT scans from a single area, and this is a consequence has a major burden on the doctor in diagnosing all slides in a traditional way. This traditional reading of the images leads to eye strain of the doctor, and thus can lead to the neglect of a malignant lung nodules. Computer aided diagnostic systems (CAD) are useful for radiologists by providing initial examination or second opinion helpful for the classification of lung nodules [4]. CAD systems are described by the quantitative metrology and can analyze a large number of small nodules determined by CT scan [4]. The purpose of this paper is to detect lung nodules using image processing based on a number of formulations to distinguish between them and the surrounding tissues that fall within the lung, especially the vessels that have the same round shape, and then work on the diagnosis on the basis of knowledge of the form of malignant tumor, which has a regular form [1]. This work is a continuation and update of the previous work, where the design of the graphical interface was added automatically [5].

2. METHODOLOGY

There are three main processes used in this research, processing, extraction of features, and finally the classification process. The MATLAB program was used in every process applied throughout the project. The processes included in the lung nodules detection system are shown in Figure 1. The first step is to get a CT image of a patient with lung cancer. CT scans of the lungs have a very low noise ratio compared to X-rays and MRI images. The main advantage of using the resulting images of CT is that it gives better clarity and less distortion. For this research, the CT scan of the lungs shown in Figure 2a was obtained from the Lung Image Database Consortium (LIDC) NIH / NCI [3].

1. First stage:- Extract the lungs from the original CT image

- a) Automatic image resolution: The threshold was obtained using the Otsu method [6] and applied to the input image.

Figure 2b shows the image after thresholding operation for the original image.

- b) Connected-component labeling (CCL): Using the CCL algorithm and calculating the area of all components, and selecting the larger area component that represents the region around the lungs (see Figure 3a).
- c) Get the image of the lung mask: This is done in two steps: Process filling holes, to obtain the outer boundary of the object (see Figure 3b). Form the lung mask by multiplexing the previous image by the resulting image from step (2) (see Figure 4a).

Using the CCL algorithm, then we calculate the area of all components, and select the larger area component that represents the region around the lungs.

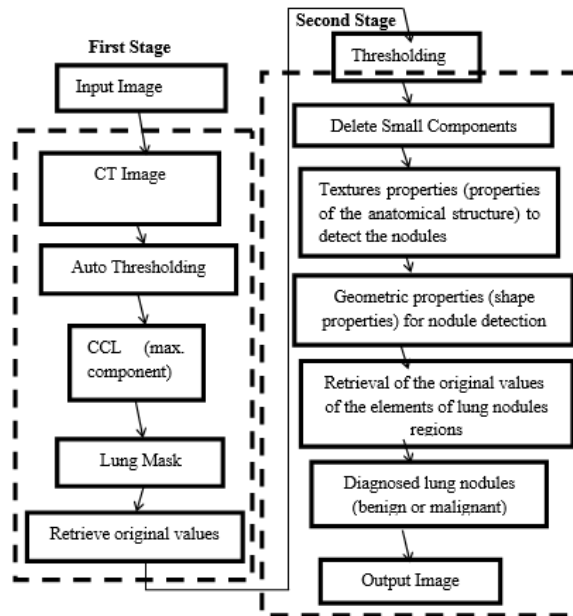


Figure. 1 Block diagram for detection and diagnoses of lung nodules.

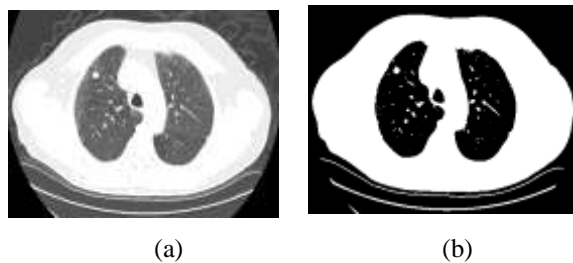


Figure. 2 a) Original CT scan image, and b) CT scan after thresholding.

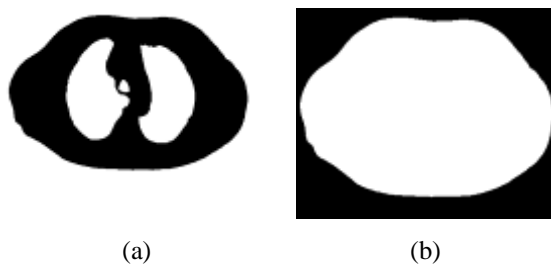


Figure. 3 a) Resulting image from selection of max CCL component, and b) Process of filling holes for the previous image.

d) Dispose of organs near the lungs

The objects in the surrounding of the lungs are deleted using the morphological processes, the morphological opening process followed by the closing process, where the morphology element has a radius of 10 (see Figure 4b). Restore the original values of the lung area

The original values of the lung components are restored according to the thresholding criterion (see Figure 5a).

2. Second stage:- Extract lung nodules from lung image

a) Image thresholding

From the histogram shown in Figure (5b), we notice that the elements of the image fall below 150, so the threshold value is chosen according to the image histogram (see Figure 6a).

b) Delete small components

In this step, components that its area are less than 15 elements are deleted using CCL technology (see Figure 6b).

c) Use textures properties (properties of anatomical composition) to detect lung nodules

This step was applied using the CCL technique and calculating the texture properties of the image elements in each component. These properties are illustrated in Table I. Where:

$m = \sum_{i=0}^{255} z_i * p(z_i)$ is the average of the gray level, z_i is the element value at gray level i , and $p(z_i)$ is the value of the histogram at the gray level i .

The Textures Properties of the lung nodules were found in the ranges listed in Tables I and II. These characteristics of the lung nodules were calculated in all cases used in this paper is shown in Tables III and IV. Then delete all connected components, which do not fulfill the properties of the textures. Figure 7a shows the resulting image of these deletion of components that do not match the properties of the textures.

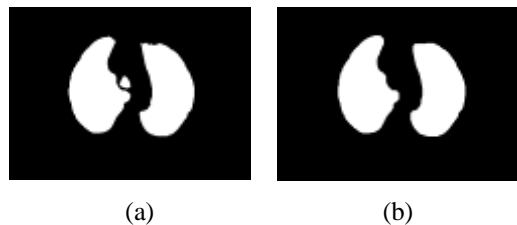


Figure. 4 a) Lung mask image, and b) Resulting image after applied the morphology operation.

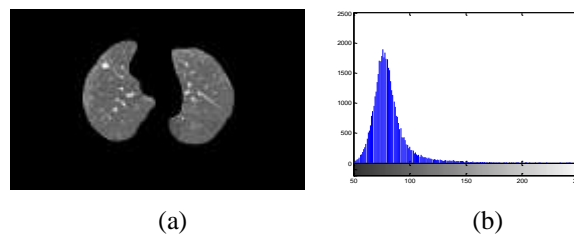


Figure. 5 a) Lungs extracted from the original CT image, and b) Histogram for Lungs image.

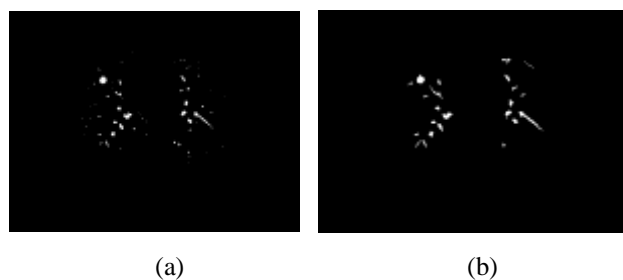


Figure. 6 a) Resulting Lungs image after thresholding, and b) Resulting Lungs image after deleting small components.

Table. 1 Textures properties

Normalized third moment	$\mu_2 = 255 * \sum_{i=0}^{255} (z_i - m)^2 p(z_i)$
Standard deviation	$\sigma = \sqrt{u_2}$
Smoothness	$R = 1 - \frac{1}{1 - \sigma^2}$
Third moment	$\mu_3 = \sum_{i=0}^{255} (z_i - m)^3 p(z_i)$
Similarity measure	$U = \sum_{i=0}^{255} p^2(z_i)$
Entropy	$e = - \sum_{i=0}^{255} p(z_i) * \log_2 p(z_i)$

Table. 2 Textures properties range

Normalized third moment	177-245
Standard deviation	13 - 35
Smoothness	0.0022 - 0.0189
Third moment	-0.82 – 0.240
Similarity measure	0.0189 - 0.22
Entropy	3.5 - 6

Table. 3 Textures properties values for extracted lung nodules

case No	nodule No	normalized third moment	Standard deviation	smoothness
1	1	207.734	22.05	0.0074
	2	205.171	25.33	0.0098
2	1	205.255	24.63	0.0092
3	1	202.96	26.71	0.0109
4	1	190.278	18.09	0.0055
5	1	0.5750	1716.8	0.7886
6	1	222.801	34.1	0.0101
7	1	178.677	20.93	0.0067
8	1	180.171	17.06	0.0045
	2	181.2121	15.49	0.0037
	3	213.764	21.57	0.0071
	4	214.761	27.48	0.0115
	5	199.696	26.91	0.0112

9	1	227.501	33.52	0.0145
	2	201.912	32.91	0.0103
10	1	233.503	28.63	0.0029
	2	235.223	27.24	0.0148
	3	238.912	26.10	0.0139
	4	223.434	33.01	0.0150
	5	227.101	31.22	0.0141
11	1	228.702	31.61	0.0029
12	1	193.219	18.234	0.0051
13	1	1.0982	243.14	0.9655
	2	0.9453	1319.5	0.9485
14	1	0.5712	3474.7	0.7737
	2	0.8825	562.397	0.9184
	3	1.179	182.59	0.9200

Table. 4 Textures properties values for extracted lung nodules

case No	nodule No	Third moment	Similarity	Entropy
1	1	-0.2023	0.0584	5.30
	2	-0.0732	0.0548	4.31
2	1	-0.1642	0.0230	5.51
3	1	-0.0945	0.0302	5.2175
4	1	-0.0668	0.0208	5.7
5	1	0.5774	41.434	1.4321
6	1	-0.501	0.100	3.8
7	1	0.0795	0.0489	4.567
8	1	0.0295	0.0384	4.78
	2	0.0086	0.0451	4.55
	3	-0.1493	0.0268	5.37
	4	-0.1945	0.0422	4.68
	5	-0.0683	0.0435	4.52
9	1	-0.601	0.196	4.9
	2	-0.292	0.068	5.1
10	1	-0.553	0.191	4.7
	2	-0.610	0.152	4.6
	3	-0.301	0.100	4.4
	4	-0.491	0.064	5.1

	5	-0.587	0.023	4.0
11	1	-0.501	0.104	4.4
12	1	-0.0653	0.0296	5.4
13	1	0.6667	15.5930	1.2144
	2	0.7603	36.3255	1.0589
14	1	0.5104	58.947	1.5546
	2	0.7500	23.7149	1.5825
	3	0.7667	13.5126	1.2621

d) Use geometric properties (shape properties) to detect lung nodules

The lung nodules have an external shape that approximates the shape of the circle. The characteristics of the shape can be found in many literature reviews as listed in Table 5 [1].

Then connected components that do not realized these characteristics are deleted as shown in Figure 7b.

e) Restore the original values of the elements of lung nodules regions using thresholding operation (see Figure 8a).

Table. 5 Shape characteristics range

<i>form factor</i>	0.42-1.29
<i>roundness</i>	180-9050
<i>solidity</i>	0.63-1.2
<i>extent</i>	0.5-0.88
<i>comactness</i>	13-98
<i>aspect ratio</i>	1-1.6

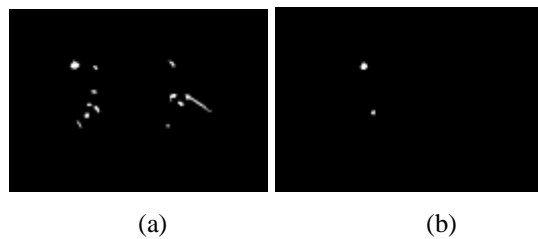


Figure. 7 a) Resulting image after deleting the components which not fullfill the texture properties., and b) Resulting image after deleting the CCL components.

Table. 6 Textures properties values for extracted lung nodules

case No	nodule No	Form factor	roundness	Solidity
1	1	0.5608	1282.9	0.8392
	2	1.1790	185.87	0.9583
2	1	1.0044	1379.3	0.9574
3	1	1.059	602.867	0.9643

4	1	0.5935	1248.8	0.8977
5	1	0.5750	1716.8	0.7886
6	1	0.9487	199.988	0.9565
7	1	0.5700	329.912	0.7949
8	1	1.2093	334.616	0.9722
	2	1.1823	294.360	1
	3	0.8581	835.824	0.8831
	4	1.1033	285.812	0.9375
	5	1.0635	192.950	0.9583
9	1	0.7953	1876.60	0.9274
	2	0.6747	477.209	0.8302
10	1	0.9942	1178	0.9451
	2	0.9584	3257.3	0.9551
	3	0.8448	2528	0.9424
	4	0.7119	1223	0.8706
	5	1.1643	244.40	0.9643
11	1	0.8303	769.073	0.8971
12	1	0.8897	3010.3	0.9463
13	1	1.0982	243.14	0.9655
	2	0.9453	1319.5	0.9485
14	1	0.5712	3474.7	0.7737
	2	0.8825	562.397	0.9184
	3	1.179	182.59	0.9200

Table. 7 textures properties values for extracted lung nodules

case No	nodule No	Extent	Compactness	Aspect ratio
1	1	0.5776	113.26	1.1774
	2	0.7667	13.6334	1.3167
2	1	0.7500	37.1388	1.2374
3	1	0.7500	24.553	1.087

4	1	0.7802	53.383	1.1894
5	1	0.5774	41.434	1.4321
6	1	0.6286	14.1417	1.5995
7	1	0.6458	18.1635	1.4016
8	1	0.8333	18.2924	1.2422
	2	0.6735	17.157	1.1196
	3	0.6800	28.9106	1.0259
	4	0.7143	16.9060	1.4058
	5	0.6571	13.8907	1.3957
9	1	0.6805	43.3199	1.0678
	2	0.5432	21.845	1.1578
10	1	0.7107	34.330	1.0382
	2	0.7556	57.070	1.0275
	3	0.7798	50.279	1.3079
	4	0.6325	34.970	1.4943
	5	0.7500	15.630	1.3930
11	1	0.5545	27.732	1.1499
12	1	0.6779	54.866	1.5403
13	1	0.6667	15.5930	1.2144
	2	0.7603	36.3255	1.0589
14	1	0.5104	58.947	1.5546
	2	0.7500	23.7149	1.5825
	3	0.7667	13.5126	1.2621

f) Automated diagnosis of lung nodules

It is intended to diagnose the lung nodules is the distinction between the malignant and benign nodules. In this study, the diagnosis of the lung nodules is dependent on the form of the nodules. The malignant nodules have an irregular (random) contour, while the benign nodes have a normal contour. Shape properties of the lung nodules can be calculated as:

$$z = \frac{\text{perimeter}}{\text{minor axes length}} \quad \text{Eq. (1)}$$

where z is a classifier that identifies lung nodules.

g) Identification of benign nodules from malignant nodules

After calculating the (z) classifier of the discovered nodules, it was concluded that the malicious nodules fulfills the following condition: $z > 3.5$

The good nodules fulfill the following condition: $z \leq 3.5$

Figure 8b shows the result of diagnosis. The malignant lung nodules were surrounded by red color, and the benign nodules were surrounded by green color. Table VIII contains all the results of lung nodules diagnosis used in this study.

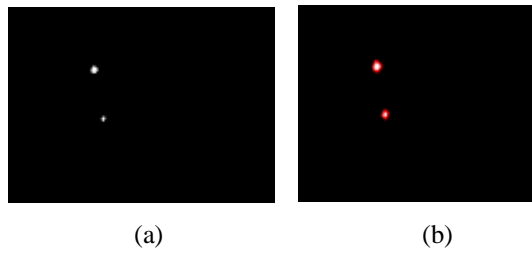
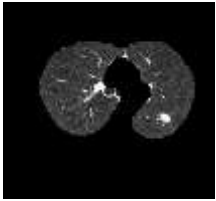
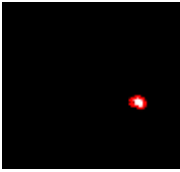

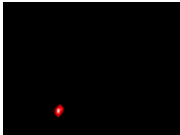
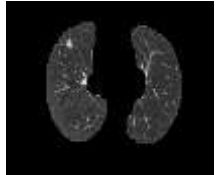

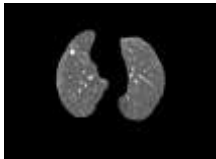
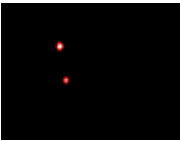

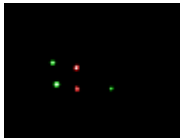
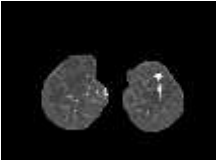
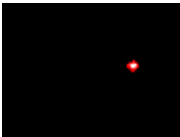
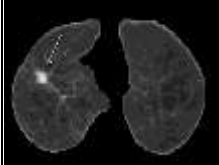
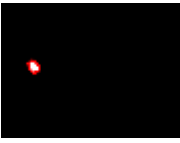
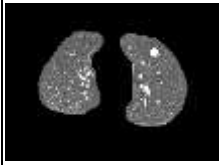
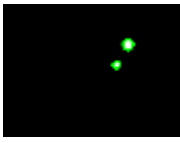
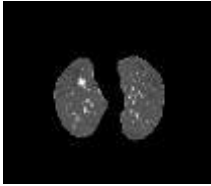
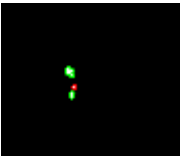
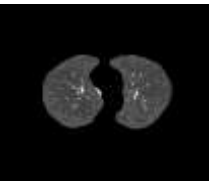
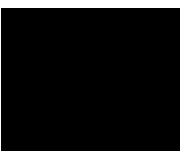


Figure 8 a) Detected Lung nodules, and b) Diagnosed Lung nodules.

Table. 8 Diagnoses result lung nodules in all used images in the study

No	Lungs image	Diagnosis result	explaining of case
1			Infected image (malignant & benign)
2			Infected image (benign)
3			Infected image (benign)
4			Infected image (malignant)

5			Infected image (malignant)
6			Infected image (malignant)
7			Infected image (malignant)
8			Infected image (malignant)
9			Infected image (malignant & benign)
10			Infected image (malignant)

11			Infected image (malignant)
12			Infected image (benign)
13			Infected image (malignant and benign)
14			Intact image

3. GRAPHICAL USER INTERFACE (GUI)

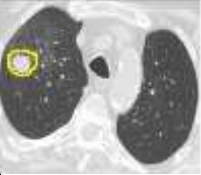
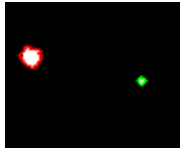

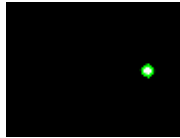

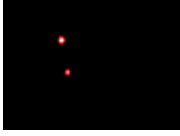


The graphical interface is designed to help the non-expert user in programming languages such as the doctor so that he can use the program code and see the results without dealing with the lines of commands and instructions, using the programming language MATLAB (see Figure 9).

4. DISCUSSION AND EVALUATION OF RESULTS

In this section, the results of the diagnosis of a group of images used in the study obtained using the proposed method will be compared with the diagnosis of the specialist doctor for these images (see Figure 10). These results were evaluated using the Receiver Operating Characteristic (ROC) method. It's a statistical method used to compare the performance of the classification in diagnostic tests (see Table IX). First: Evaluation of the automatically detection of the lung nodules:



Figure. 9 Designed graphical user interface

No	doctor's diagnosis	automatic diagnosis result
1		
2		
3		
4		

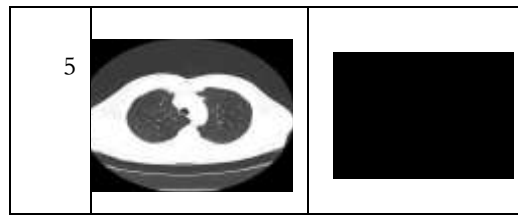


Figure. 10 Designed graphical user interface

Table. 9 Diagnoses result using roc method

Case no	Detection by doctor	<i>TP</i>	<i>FP</i>	<i>FN</i>
1	1	1	1	0
2	1	1	0	0
3	1	1	0	0
4	1	1	0	0
5	1	1	0	0
6	1	1	0	0
7	1	1	0	0
8	1	1	1	0
9	3	2	3	1
10	1	1	0	0
11	1	1	0	0
12	1	1	1	0
13	1	1	2	0
Total	15	14	8	1

Where: *TP* is the nodules that are correctly detected.

FP is the incorrectly detected nodes and is not actually exist in the doctor's detection.

FN The undiscovered nodules and is present in the doctor's detection.

$$(2) \text{ detection ratio} = \frac{\text{number of nodules detected by system}}{\text{number of nodules detected by doctor}}$$

$$= 93.3\% \text{ detection ratio} = \frac{16}{17}$$

Second: Evaluation of the mechanical diagnosis of the pulmonary contract:

$$SEN = \text{Sensitivity} = \frac{TP}{TP + FN} = \frac{9}{9 + 1} = 90\%$$

$$SPE = \text{Specificity} = \frac{TN}{TN + FN} = \frac{8}{8 + 1} = 88.90\%$$

$$ACC = \text{Accuracy} = \frac{TP + TN}{n} = \frac{8 + 9}{22} = 77.27\%$$

Where: *TP* is malignant nodules which are properly diagnosed and are identical to a doctor's diagnosis, *TN* is a benign nodule that has been properly diagnosed which is identical to a doctor's diagnosis, *FP* malignant nodules that are mistakenly diagnosed as a benign contract, and *FN* is benign nodules are incorrectly diagnosed.

N	TP	TN	FP	FN	NC	SEN
22	9	8	1	1	2	90%

SPE		ACC				
88.90%		77.27%				

where: NC is not classified, SEN is sensitivity, SPE is specificity, and ACC is accuracy.

5. CONCLUSIONS

Lung cancer is one of the most serious diseases in the world. Correct diagnosis and early detection of lung cancer can increase the survival rate. Current techniques include x-ray, CT, MRI and PET. Expert doctors rely on experience in diagnosing lung nodules. The nodules are formed in some patients due to inflammatory, bacterial or tumor causes. Treatment includes surgery, chemotherapy, and radiotherapy. These treatments are long, and expensive. This paper is developed and continue to the previous work by Mabrukah Fadel and etc. In this study we tried to help the doctor with automated diagnosis, where doctors depend on the size of nodules and their growth over time as well as the patient's age and whether he is a smoker or not. But this study depends on the form of the nodules to distinguish between benign nodules and malignant nodules. Therefore, the diagnosis resulting from this algorithm can be considered an indication of the doctor, where the detection of at least one nodule will push the doctor to review the image to diagnose it accurately and quickly. Experiments was done for 22 CT scan images and the system accuracy, sensitivity, and specificity values are 77.27%, 90%, and 88.90% respectively.

REFERENCES

1. Levner, I. and H. Zhang, *Classification-driven watershed segmentation*. IEEE transactions on image processing, 2007. **16**(5): p. 1437-1445.
2. Gajdhane, A.V. and L. Deshpande, *Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques*. IOSR Journal of Computer Engineering (IOSR-JCE), 2014. **16**(5): p. 28-35.
3. Chaudhary, A. and S.S. Singh. *Lung cancer detection on CT images by using image processing*. in *2012 International Conference on Computing Sciences*. 2012: IEEE.
4. Van Ginneken, B., B.T.H. Romeny, and M.A. Viergever, *Computer-aided diagnosis in chest radiography: a survey*. IEEE transactions on medical imaging, 2001. **20**(12): p. 1228-1241.
5. Mabrukah Edrees Fadel, R.A.A., and Ali Abdirahman Kashif,. *Computerized System For Lung Nodule Detection in CT Scan Images by using Matlab*. in *3rd International Conference on Automation, Control, Engineering and Computer Science (ACECS'2016)*. 2016. Hammamet, Tunisia.
6. Gonzalez, R.C. and R.E. Woods, *Digital image processing second edition*. Beijing: Publishing House of Electronics Industry, 2002. **455**.

AUTHORS PROFILE