Jurnal Teknologi

Full Paper

GEOMETRICAL FEATURE OF LUNG LESION IDENTIFICATION USING COMPUTED TOMOGRAPHY SCAN IMAGES

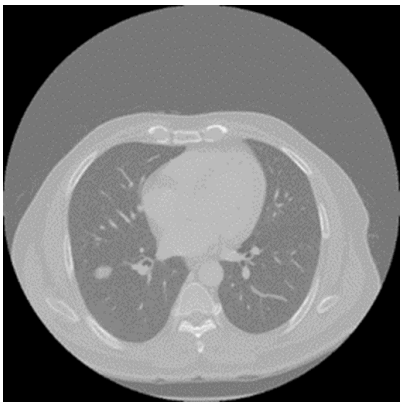
Mohd Firdaus Abdullah[a]*, Siti Noraini Sulaiman[a,b], Muhammad Khusairi Osman[a], Noor Khairiah A. Karim[c], Samsul Setumin[a], Iza Sazanita Isa[a], Adi Izhar Che Ani[a]

[a]Centre for Electrical Engineering Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, 13500 Permatang Pauh, Pulau Pinang Malaysia
[b]Integrative Pharmacogenomics Institute (iPROMISE), UiTM Puncak Alam Campus, Bandar Puncak Alam, Puncak Alam, Selangor, 42300, Malaysia
[c]Advanced Medical and Dental Institute, Universiti Sains Malaysia, Bertam 13200 Kepala Batas, Pulau Pinang, Malaysia

Article history
Received
2 July 2022
Received in revised form
26 December 2022
Accepted
26 December 2022
Published Online
23 February 2023

*Corresponding author
f.abdullah@uitm.edu.my

Graphical abstract



Abstract

Lung lesion identification is an important aspect of an early lung cancer diagnosis. Early identification of lung cancer may assist physicians in treating patients. This paper uses computed tomography scan images to present a lung lesion identification geometrical feature. From the previous studies, lung segmentation is particularly challenging because differences in pulmonary inflation with an elastic chest wall can result in significant variability in volumes and margins when attempting to automate lung segmentation. Besides, the features used to describe a lung lesion focus on image features which are geometric, appearance, texture, and others. This study develops an image processing technique that uses image segmentation algorithms to segment lung lesions in computed tomography images. The suggested approach includes the following stages, which require image processing techniques: data collection, image segmentation, and performance evaluation. The computed tomography scan images were collected from Advanced Medical and Dental Institute (AMDI), Universiti Sains Malaysia database. As a contribution to biomedical engineering, this study has successfully calculated the performance of the image processing method for lung segmentation, which gets an average accuracy of 99.38%, recall is 99.45%, and F-score is 99.6. The lung lesion segmentation approach based on the object's size could help investigate image abnormality for medical analysis. From the study, 80% of the total lesion identification using the proposed method was correctly predicted when compared with the radiologist's lesion mark. The experiment results found clear support for the next stage of this research.

Keywords: Computed tomography, image processing, image segmentation, lung cancer, lung lesion

Abstrak

Pengenalpastian nodul paru-paru adalah aspek penting dalam diagnosis kanser paru-paru pada peringkat awal. Pengenalpastian awal kanser paru-paru boleh membantu pakar perubatan dalam merawat pesakit. Projek ini membentangkan ciri geometri pengecaman nodul paru-paru menggunakan

imej imbasan tomografi yang dikira. Daripada kajian terdahulu, segmentasi paru-paru amat mencabar kerana perbezaan dalam inflasi paru-paru dengan dinding dada yang anjal boleh mengakibatkan kebolehubahan yang besar dalam volum dan margin apabila cuba mengautomasikan segmentasi paru-paru. Selain itu, ciri-ciri yang digunakan untuk menggambarkan nodul paru-paru memfokuskan kepada imej yang berbentuk geometri, rupa, tekstur dan lain-lain. Kajian ini menggunakan teknik pemprosesan imej yang menggunakan algoritma segmentasi imej untuk membezakan nodul paru-paru dalam imej tomografi yang dikira. Kaedah yang dicadangkan adalah, teknik pemprosesan imej: pengumpulan data, pembahagian imej dan penilaian prestasi. Imej imbasan tomografi itu dikumpul daripada Institut Perubatan dan Pergigian Termaju (AMDI), pangkalan data Universiti Sains Malaysia. Sebagai sumbangan kepada kejuruteraan bioperubatan, kajian ini telah berjaya mengira prestasi pemprosesan imej untuk segmentasi paru-paru, yang mendapat purata ketepatan ialah 99.38%, *recall* ialah 99.45%, dan akhir sekali F-skor ialah 99.6. Daripada analisis, 80% daripada jumlah pengecaman nodul menggunakan kaedah yang dicadangkan telah diramalkan dengan betul jika dibandingkan nodul yang di tanda oleh ahli radiologi. Keputusan eksperimen amat sesuai untuk ke peringkat seterusnya dalam penyelidikan ini.

Kata kunci: Imbasan tomografi, pemprosean imej, segmentasi imej, kanser paru-paru, nodul paru-paru

© 2023 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

Lung cancer is a disease that occurs in the lungs and could spread to other body parts. This lung cancer could lead to death and has become the second most common cancer worldwide for both gender [1]. Lung cancers start evolving in the lungs and might spread to lymph nodes or organs within the frame, including the brain [2]. Since lung cancer tends to spread or metastasize at an early stage after it forms, it is a very life risky cancer and one of the most challenging cancers to treat. According to the world health institution, approximately 8.8 million people die yearly from cancer, and cancer mortality is expected to keep increasing to 18 million in 2030. The mortality rate of lung cancer causing death is 27% of all deaths due to cancer [3]. Cancer occurs when the cell starts spreading and growing abnormally. The uncontrolled division of the cell could form a mass which is a tumour [4]. The tumour could occur in any part of the body; if it appears in the lung, it will be called lung cancer.

Early detection can improve treatment effectiveness and enhance patient survival [5]. Before any treatment is used, the doctor will diagnose the lung cancer patient to determine the lesion's location, shape, and size by using imaging modalities. The imaging modalities include CT scan [5-7] and MRI [8,9]. However, a CT scan is preferred because it is easy to use and provides accurate classification and foreign mass location [6]. Besides, the use of MRI is expensive, time-consuming diagnostics, and less available than CT scan [11].

Clinical image processing and analysis are essential, particularly in lung applications. The use of image processing in clinical study or for diagnostic purposes has grown in popularity among medical experts due to the benefits of non-invasive techniques and the ability to analyze complex organs such as arteries, veins, bronchi, and bronchioles in the lung region. Through the innovation of scientific computing and the advancement of imaging technology, various efforts have been made to automate the human role in lung lesion identification [7-9].

Generally, a wealth of well-known publications have looked into developing computer-aided diagnosis (CAD) systems for lung cancer using different image modalities [10]. It is a highly effective method for assisting radiologists in improving diagnostic accuracy [11]. Image filtering, segmentation, nodule detection, and classification are part of a typical CAD system. Researchers have recently shown a growing interest in medical image processing techniques due to the requirement to construct an advanced clinical evaluation and diagnosis system. The procedure of image segmentation extracts the lung region from an image [12]. It differentiates between items in the foreground and background and becomes vital because it enhances the look by detecting lesions with higher precision and accuracy [13]. Despite its extensive clinical success, segmentation presents several challenges, including detecting and distinguishing the anatomical areas of the lung on a CT scan image. It may include applying human or automatic segmentation techniques to the entirety of the area. When different studies attempt to create lesion morphology, they encounter issues such as poor delimitation of the borders associated with the physical structure of the lungs and hazardous noise that varies with the intensity of each image.

Lesions identification in the lung region is the next step of lung cancer detection in CAD systems. Using image processing techniques, [14] implements feature extraction and classification of lung cancer

nodules. Extraction only considered the following characteristics: area, perimeter, and eccentricity. Besides the features used to describe a lung lesion from [15] focus on image features which are geometric, appearance, texture and others. Without referring to the intensity information, geometric images describe the geometric nature of a lung lesion. A lesions volume, area, perimeter, diameter, surface area, and aspect ratio are all geometric features used to characterize it [15].

The ground truth photos were manually tagged by the radiologist as a reference. These precise detection results may aid radiologists in automatically analysing and diagnosing patients' diseases, such that early lung lesion test results may aid patients in therapy by minimising risk or confirming the absence of trouble [16-17]. In line with this, this study seeks to discover the roundness-based characteristics of lesion and non-lesion conditions.

2.0 METHODOLOGY

The method is carried out on a computer with embedded software Matlab R2019b tested using a laptop with GeForceRTX 2070 GPU, 8.00GB RAM and NVIDIA GeForce MX230. Figure 1 depicts the general scheme of the automated technique. The lungs are then extracted from the remainder of the image based on user input. In the proposed approach, the segmentation of the lungs is automatic and requires no user intervention. The threshold is chosen automatically by the segmentation technique. Based on this threshold, the image's outer portion is detected and deleted, and the remaining region is processed further to extract the lungs. The subsequent step is the detection of nodules inside the lung segments. The interiors of the nodule, bronchus, and blood vessels were separated from the parenchyma region using geometric features. The details of each process are explained in the subsection below.

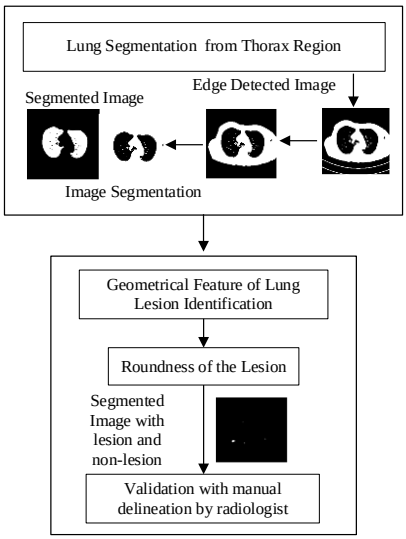


Figure 1 An Overview of Geometrical Feature of Lung Lesion Identification

2.1 Data Acquisition

CT images are collected at the Patient Contributed Image Repository in Digital Imaging and Communications in Medicine (DICOM) pixel size. DICOM header data is extracted and analyzed using the MATLAB function to reads the image data from the compliant DICOM file. This study received institutional ethics approval on 8th August 2019 from the Human Research Ethics Committee of USM (JEPeM) under the School of Medical Sciences, USM, IPPT, Bertam, Pulau Pinang [USM/JEPeM/19040231]. To comply with the used ethical regulations, the data is de-identified and anonymized before being transferred to the public domain.

2.2 Lung Segmentation

The process of grouping pixels into several homogeneous regions is known as image segmentation. This task was essential and one of the most critical requirements in medical image analysis. It is one of the fundamental techniques for understanding images. The medical staff (radiologists) face the problem of manually screening hundreds of CT scan images. It will produce a late diagnosis, resulting in undesirable catastrophic outcomes, missed early treatments, and increased health care costs.

Image pre-processing techniques such as image enhancement and filtering techniques that are beneficial to detect/identify the features/characteris tics of CT scan images are used to improve image quality at this stage. Once the image quality is improved, the subsequent identification stage will be much easier. Most image processing algorithms require pre-processing to enhance image quality. It is a method to perform some operations on an image to get an improved image or extract some useful information from it. It is a type of signal processing in which input is an image, and output may be an image or characteristics/features associated with that image. The image will first be input for this study to start the process, as in Figure 2. Then, thresholding is the most basic technique of image segmentation.

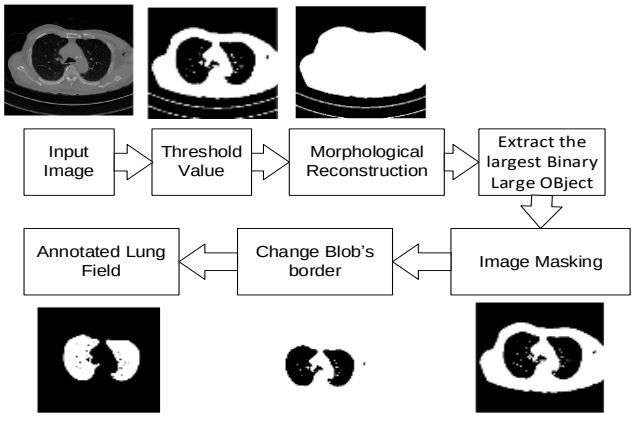


Figure 2 Block Diagram For Image Processing Technique

Threshold value, *thresh* is used to separate lung CT images from their background by assigning an intensity value for each pixel. Each pixel is classified as an object or background point. The threshold value, *thresh* is determined by using Global thresholding Otsu's method on a center windowed image, I_{CW} of size W_W X W_L; where, W_W =W/3 and W_L=L/3. Image threshold will automatically determine different threshold values for one image using different algorithms. The black and white image, $I_{BW}(x,y)$ is obtained by thresholding the intensities of $I_G(x,y)$ based on the *thresh* value as in

$$I_{BW}(x,y) = \begin{cases} 1 & if \quad I_G(x,y) > thresh \\ 0 & if \quad I_G(x,y) \leq thresh \end{cases} \quad (1)$$

Where *thresh* is the threshold value determined using Otsu's thresholding method. The result of thresholding is a binary image, where pixels with an intensity value of 1 correspond to objects, whereas pixels with 0 corresponding to the background. Morphological reconstruction may be conceptualized as successive dilations of one image (the marker image) until the contour of the marker image fits under a second image (the mask image). The peaks in the marker image "spread out," or dilate, during morphological reconstruction. The following procedure fills all the holes with

$$I_{BWF}(x,y)_k = (I_{BW}(x,y)_{k-1} \oplus B) \cap A^C) \quad (2)$$

Where $I_{BW}(x,y)$ is the black and white image, $I_{BWF}(x,y)$ is the black and white image with fills holes, *B* is the symmetric structuring element, and *A* is a set containing one or more connected components. The algorithm is executed iteratively and terminates at iteration step k if $I_{BWF}(x,y)_k = (I_{BW}(x,y))_{k-1}$. The set $I_{BWF}(x,y)_k$ then contains all the filled holes. The set union of $I_{BWF}(x,y)_k$ and *A* contains all the filled holes and their boundaries. The intersection at each step with A^C limits the result to inside the region of interest.

After that, the larges Binary Large OBject (BLOB) in the image will be extracted by computing the areas of all blobs, sorting the computed areas in descending, and selecting the first blob (with the most significant area) only. Masking is an image processing technique in which a small 'image piece' is defined and used to alter a bigger image [16] as follows:

$$I_{M1}(x,y) = I_{DP}(x,y) \circ I_{BB}(x,y) \quad (3)$$

Lastly, the lung region image segmentation will be the final result by applying image inversion to the $I_{INV}(x,y)$, represented as $I'_{INV}(x,y)$. The area must determine each blob using how many connected neighbours and eliminate all blobs with a size less than β pixels. Empirically β=1500 pixels.

2.3 Lung Lesion Segmentation

In this stage, the potential lesions will be identified after the lung segmentation stage from the recorded image sets from the patients. The main challenge is when the lesion signs are too small or close to the blood vessel. This issue requires more computational effort in image segmentation. The overall organs such as arteries, veins, bronchi, and bronchioles in the lung region will be identified based on their roundness to differentiate the lesion and others. Based on previous studies, no differences between genders were found regarding roundness. A roundness of 1 will indicates perfectly round [9]. An integrated analysis of CT roundness may improve the understanding of evaluating these lesions in clinical practice [7].

The geometrical properties of lung lesions are significant in distinguishing between benign and malignant lung lesions; for example, the larger the size of a nodule, the more likely it is to be malignant[11]. The thresholding technique calculates such geometric features from the suspicious binary region (SR). According to the literature [11], the growth of the malignant part (nodule here) is usually circular; thus, the roundness of the nodule is calculated using a simple equation:

Roundness= (4π ×area)/〖Perimeter〗^2 (4)

This metric value, also known as the roundness, circularity, or irregularity index (I), is equal to one for circles and less than one for all other shapes. It has been assumed here that the greater the circularity of the object, the greater the likelihood of that object being a nodule[9-11].

Circular shapes in the detected collection of structures were recognised using the size-invariant round/near-round shape detection technique proposed in [7]. As seen in Figure 3, the bronchioles and ship are cylindrical, while the lesions are spherical. The suggested strategy allows us to extract nodules of any shape, making our methodology template-independent in contrast to previous nodule template-based methodologies. To remove prospective lesion locations, a region-growing algorithm uses the centres of the potential nodule locations as seed points.

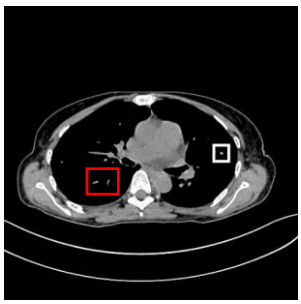


Figure Error! No text of specified style in document.. The region square in white is a lesion and the region squared in red is a non-lesion

2.4 Performance Evaluation

The calculated evaluation performance refers to the accuracy, recall, and F-score test. The accuracy measurement was calculated based on the actual or known value. The total number of lung nodule pixels detected is divided by the number of lung lesion and background pixels seen to calculate recall. The accuracy of an algorithm is more precisely reflected by the F-score, which is the harmonic mean of precision and recall. Only with strong recall and precision can one achieve a high F-score. The formula for each test is provided below [18].

$$\text{Accuracy} = (Tp+Tn)/(Tp+Tn+Fp+Fn) \quad (5)$$

$$\text{Recall} = Tn/(Tp+Fn) \quad (6)$$

$$\text{F-Score} = (2\text{XPrecisionXRecall})/(\text{Precision}+\text{Recall}) \quad (7)$$

Where

Tp= true positive,
Tn= true negative
Fp= false positive
Fn=false negative

True Negative, TN is the number of background pixels and True Positive, TP is the number of lung nodule pixels correctly classified. False Positive, FP is an incorrect classification of the number of lung nodule pixels and False Negative, FN is the incorrect classification of the number of background pixel.

3.0 RESULTS AND DISCUSSION

The primary objective of this research is to develop automated lung lesions segmentation for different CT sequences. The system to be constructed must process the images to detect signal abnormalities and extract the features of the lesion.

This chapter is divided into three main parts. The first part of this chapter presents the data collections, The second part for lung segmentation results. Finally this chapter presents the results of the lung lesion segmentation stages, including the roundness analysis for lesion segmentation.

3.1 Data Collections

A total 20 lung CT images used in this study. According to earlier research, there were no gender variations in roundness. The evaluation of these lesions in clinical practice may be improved by an integrated examination of CT roundness [7]. The CT image dataset used in this study was obtained from the Imaging Department at AMDI, USM, Kepala Batas, Pulau Pinang.

3.2 Analysis of Lung Segmentation

The threshold value is used to distinguish lung CT images from the background by assigning an intensity value to each pixel. Each pixel is classified as either an object point or a background point. The results revealed an optimized threshold value for correctly separating lung CT images from the background automatically was 0.2706. The range level was specified in between 0 and 1, regardless of the class of the input image. Several factors contributed to perfect lung segmentation in the image pre-processing. Higher image sharpness and uniform image had the most significant influence to the process. Figure 4 illustrates the pre-processed images with the automatic threshold value 0.2706.

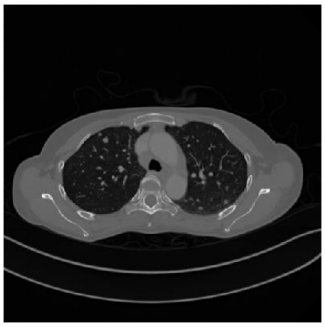


Figure 4 Result of Threshold Value in CT Image Preprocessing

A total of 20 lung imaging samples were collected for evaluation. Each blob's area was calculated using its 4/8 linked neighbours, CC, and P. All blobs with sizes smaller than 1500 pixels were eliminated. The overall effectiveness of the image processing technique based on Accuracy, Recall, and F-score is displayed in Figures 4 to 5.

According to the analysis, even when the values of CC and P change, the result does not significantly alter, demonstrating that the image processing technique is the best way to segment the image for lung cancer segmentation. Because CC = 4 and CC = 8 use the same scanning approach, which is simply scanning two-dimensional connectivities images, the result displays the average score for all 4 values for CC and P with no increase or substantial decrease. Figure 8 displays 4- and 8-connected connections. According to this explanation, it is 4-connected because pixels are connected if their edges contact. If two pixels are on and connected along the horizontal or vertical, they are considered to be a single object. The edges or corners of pixels that are 8-connected are what connect them. If two pixels are connected in a horizontal, vertical, or diagonal direction and are both on, they are considered to be a single object. Because of this, the outcomes have not changed significantly. The items' total number of pixels is P. The findings won't change no matter how many P there are because only the label matrix's equation based on the P-value will be altered.

Because 8 is the standard connection for 2-D images, this study uses that value to achieve the required outcomes. If two pixels' corners or edges

come into contact when they are 8-connected, as in Figure 5, they are linked. Two neighbouring pixels are regarded as belonging to the same object if they are both on and connected in a horizontal, vertical, or diagonal direction.

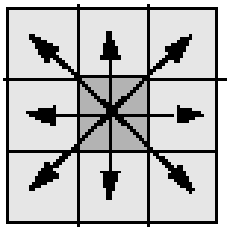


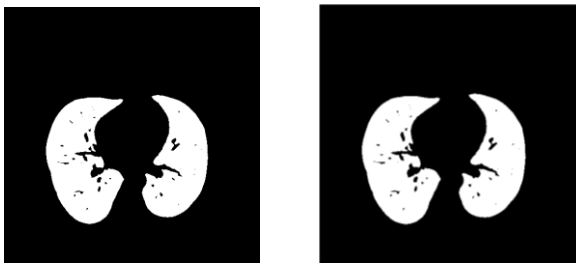
Figure 5 8-connected

Image segmentation is the process of grouping pixels into several homogeneous regions. This task was necessary and one of the essential requirements in analysing medical images. It is one of the basic techniques that help in image understanding. The method was evaluated by calculating the Accuracy, Recall and F-score. Table 1 shows the overall performance value of CC = 8. P=1500 for 20 images used lung segmentation. Based on the output, the performance results for lung segmentation produced an average value of accuracy of 99.38%, recall at 99.45% and F-score at 99.6%.

Table 1 Average performance for lung segmentation

Image	Accuracy	Recall	F-score
1	99.67	99.73	99.84
2	99.57	99.64	99.79
3	99.51	99.60	99.76
4	99.34	99.46	99.67
5	99.27	99.40	99.63
6	99.24	99.37	99.62
7	99.27	99.40	99.63
8	99.29	99.41	99.64
9	99.33	99.45	99.67
10	99.25	99.38	99.62
11	99.28	99.41	99.64
12	99.18	99.32	99.59
13	99.14	99.29	99.57
14	99.50	99.29	98.92
15	99.46	99.22	98.81
16	99.37	99.47	99.68
17	99.44	99.53	99.72
18	99.46	99.55	99.73
19	99.47	99.55	99.73
20	99.48	99.56	99.74
Average	99.38	99.45	99.60

Figure 6 shows the ground truth image before and after the image processing technique has been implemented. To construct a ground truth image, the manual segmentation technique used image segmentation in the MATLAB programme. This manual segmentation was accomplished by employing a graph cut in the generate mask tools to crop the lung image's shape without altering or modifying the format. As a result, it is simpler to see the details of the lung image. To further improve the accuracy of the hand segmentation, the background of the lung was removed by designating the undesirable areas.



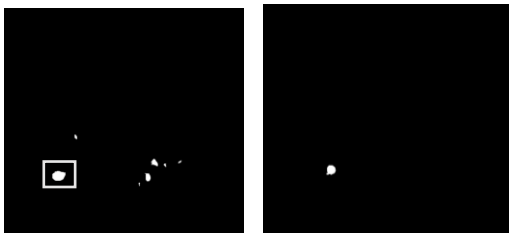
(a) (b)

Figure 6 (a)Ground truth image. (b) The result after applying image processing technique

3.3 Analysis of Lung Lesion Segmentation

The lung lesion segmentation approach based on the object's size was capable of helping in investigating image abnormality for medical analysis. For lung lesion segmentation, it was more challenging due to the complexity of the lung anatomy and structure. Therefore, the output of this study will be used for the following method for the lung lesion segmentation system. The lesions differ from other structures present in the lung in many aspects. One key difference is their shape, and we exploit this property to isolate the nodules from the non-nodules structures, i.e vessels and bronchi. The size-invariant round/ near round shape is used to identify the circular shapes in the detected structures [7].

In contrast to existing nodule template-based techniques, the proposed strategy enables the extraction of lesions of any shape, making this methodology independent of any nodule template. Based on Figure 7(a), the region in a square is a lesion and other non-lesions. While Figure 7(b) shows the ground truth for manually segmentized lesion.



(a) (b)

Figure 7 (a)the region in square is a lesion and others non lesion. (b) ground truth image for lung lesion segmentation

Table 2 shows the average performance results for lung lesions segmentation. The performance results for lung lesions segmentation produced an average accuracy value of 99.92%, recall at 42.79% and F-score at 56.45%. The accuracy test determines how well a diagnostic test identifies and rules out a

specific condition. The percentage of the actual lung region correctly segmented using this method is called recall. The F-score is the harmonic mean of precision and recall, reflecting an algorithm's accuracy more accurately. Table 2 shows that recall and F-score values are lower when compared to accuracy due to the True Positive (TP) value is low. This is because after segmentation there will be non-lesion part as in Figure 7(a) that can effect the TP value when compared to the ground truth image as in Figure 7(b).

Table 2 Average performance for lung lesion segmentation

Image	Accuracy	Recall	F-score
1	99.97	57.74	73.2
2	99.98	83.56	91.04
3	99.92	64.88	78.70
4	99.86	54.29	70.37
5	99.96	45.03	62.10
6	99.95	40.52	57.67
7	99.87	47.05	63.99
8	99.96	45.03	62.10
9	99.90	92.93	96.33
10	99.96	32.28	48.91
11	99.95	8.27	15.27
12	99.91	20.74	34.36
13	99.91	21.80	35.81
14	99.89	16.26	27.97
15	99.93	33.75	50.46
16	99.94	42.85	60.00
17	99.94	39.26	56.38
18	99.87	18.07	30.62
19	99.87	45.60	57.50
20	99.89	45.80	56.30
Average	99.92	42.79	56.45

A roundness value indicates a circular item near to one [9]. For those categorised as test samples of an artificial neural network for upcoming study, the form values of the test samples are comparable to one in a range of +/- 5%, indicating the possibility of the object being a lung nodule is relatively high. Based on Table 3, out of 20 images, 16 or 80% of the total lesion identification using the proposed method was correctly predicted when compared with the lesion mark by the radiologist. This is due to the lesion indications are too small or too close to a blood vessel, it is challenging to diagnose. For image segmentation, this problem calls for more excellent computational work. However, the findings demonstrate that, compared to the image the radiologist had annotated, the proposed method could detect a lesion based on roundness value.

Table 3 Roundness value for lesion and non-lesion

		Lesion and Non-lesion Detection	
Image	Roundness Value	Using Geometrical Features	Marked by Radiologist
1	1.15	Non-Lesion	Non-Lesion
2	1.13	Non-Lesion	Non-Lesion
3	1.08	Non-Lesion	Lesion
4	1.12	Non-Lesion	Non-Lesion
5	0.89	Non-Lesion	Non-Lesion
6	1.03	Lesion	Lesion
7	0.98	Lesion	Lesion
8	1	Lesion	Non-Lesion
9	0.96	Lesion	Non-Lesion
10	1.04	Lesion	Lesion
11	0.92	Non-Lesion	Non-Lesion
12	0.99	Lesion	Non-Lesion
13	1.07	Non-Lesion	Non-Lesion
14	0.95	Lesion	Lesion
15	1.01	Non-Lesion	Non-Lesion
16	1.05	Lesion	Lesion
17	0.89	Non-Lesion	Non-Lesion
18	0.98	Lesion	Lesion
19	0.89	Non-Lesion	Non-Lesion
20	0.94	Non-Lesion	Non-Lesion

4.0 CONCLUSION

In conclusion, the results of the proposed procedure and method for lung segmentation in CT images were presented and discussed in detail. The results were organised into three major parts; image preprocessing procedure and methodology for automated lung identification, the developed software to perform the automated lung lesion identification and performance comparison of the proposed method with ground truth. The procedure has been developed to facilitate radiologists segmenting the lung region for lung cancer identification using this method. As a contribution to biomedical engineering, this study has successfully calculated the performance of the image processing method for lung segmentation, which gets an average of the accuracy of 99.38%, recall is 99.45%, and F-score is 99.6. This result shows that all performance values are above 99%, which is excellent for the image segmentation performance standard. The findings were used for lung lesion segmentation.

The results were compared to the manual lesion segmentation that the radiologist marked to ensure that the proposed automated lung lesion was reliable. From the analysis, 80% of the total lesion identification using the proposed method was correctly predicted when compared with the radiologist's lesion mark. The experiment results found clear support for the next stage of this research. With the excellent results obtained in this research, future

work will apply the outcome, i.e. using the optical flow method to detect the lesions for lung cancer identification purposes. This may be considered a promising aspect in realizing an intelligent, fast, and accurate method for lung cancer identification.

Acknowledgement

This research work was financially supported by the Ministry of Higher Education Grant Scheme (FRGS) "A new Motion-based Optical flow Cancer Detection method for Lung Nodule Segmentation on Lung CT-Scan Images" (Ref: FRGS/1/2021/TK0/UITM/02/58)) and ethics from Universiti Sains Malaysia (USM/JEPeM/19040231). The authors would like to express their gratitude to members of the Machine Learning Research Group (MLRG), Integrative Pharmacogenomics Institute (iPROMISE), and Centre for Electrical Engineering Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang for their assistance and guidance during the fieldwork. Finally, the authors are grateful to Universiti Teknologi MARA, Cawangan Pulau Pinang for their immense administrative and financial support.

References

[1] P. Chalasani and S. Rajesh. 2020. Lung CT Image Classification using Deep Neural Networks for Lung Cancer Detection. *Int. J. Eng. Adv. Technol*. 9(3): 3998-4002. Doi: 10.35940/ijeat.c6409.029320.

[2] S. Bhatia, Y. Sinha, and L. Goel. 2019. Lung Cancer Detection: A Deep Learning Approach. *Adv. Intell. Syst. Comput*. 817: 699-705. Doi: 10.1007/978-981-13-1595-4_55.

[3] W. D. Travis *et al.* 2015. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances since the 2004 Classification. *J. Thorac. Oncol.* 10(9): 1243-1260. Doi: 10.1097/JTO.0000000000000630.

[4] M. Vas and A. Dessai. 2018. Lung Cancer Detection System using Lung CT Image Processing. *2017 Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2017*. 1-5. Doi: 10.1109/ICCUBEA.2017.8463851.

[5] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes. 2004. Texture Analysis of Medical Images. *Clinical Radiology*. Doi: 10.1016/j.crad.2004.07.008.

[6] H. Shaziya, K. Shyamala, and R. Zaheer. 2018. Automatic Lung Segmentation on Thoracic CT Scans Using U-Net Convolutional Network. *Proc. 2018 IEEE Int. Conf. Commun. Signal Process. ICCSP 2018*. 643-647. Doi: 10.1109/ICCSP.2018.8524484.

[7] B. H. Heidinger *et al.* 2017. Lung Adenocarcinoma Manifesting as Pure Ground-Glass Nodules: Correlating CT Size, Volume, Density, and Roundness with Histopathologic Invasion and Size. *J. Thorac. Oncol*. 12(8): 1288-1298. Doi: 10.1016/j.jtho.2017.05.017.

[8] T. J. N. Hiltermann *et al.* 2012. Circulating Tumor Cells in Small-cell Lung Cancer: A Predictive and Prognostic Factor. *Ann. Oncol*. 23(11): 2937-2942. Doi: 10.1093/annonc/mds138.

[9] S. T. Ligthart *et al.* 2013. Circulating Tumor Cells Count and Morphological Features in Breast, Colorectal and Prostate Cancer. *PLoS One*. 8(6): 1-11. Doi: 10.1371/journal.pone.0067148.

[10] K. V. Rani and J. J. S. 2018. Emerging Trends in Lung Cancer Detection Scheme-A Review. *International Journal of Research and Analytical Reviews*. 5(3): 530-542.

[11] S. A. Patil, V. R. Udupi, C. D. Kane, A. I. Wasif, J. V. Desai, and A. N. Jadhav. 2009. Geometrical and Texture Features Estimation of Lung Cancer and TB Images using Chest X-ray Database. *2nd Int. Conf. Biomed. Pharm. Eng. ICBPE 2009 - Conf. Proc*. Doi: 10.1109/ICBPE.2009.5384113.

[12] J. Dhalia Sweetlin, H. K. Nehemiah, and A. Kannan. 2018. Computer Aided Diagnosis of Pulmonary Hamartoma from CT Scan Images using Ant Colony Optimization based Feature Selection. *Alexandria Eng. J*. 57(3): 1557-1567. Doi: 10.1016/j.aej.2017.04.014.

[13] F. Bianconi, I. Palumbo, A. Spanu, S. Nuvoli, M. L. Fravolini, and B. Palumbo. 2020. PET/CT Radiomics in Lung Cancer: An Overview. *Appl. Sci.* 10(5): 1-11. Doi: 10.3390/app10051718.

[14] K. M. M. Tun and A. S. Khaing. 2014. Feature Extraction and Classification of Lung Cancer Nodule using Image Processing Techniques. *Int. J. Eng. Res. Technol.* 3(3): 2204-2210.

[15] N. D. 2020. Detection of Lung Cancer using SVM Classifier. *Int. J. Emerg. Trends Eng. Res*. 85: 21772180. Doi: 10.30534/ijeter/2020/113852020.

[16] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari. 2018. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. *arXiv:1802.06955*.

[17] Y. F. Riti, H. A. Nugroho, S. Wibirama, B. Windarta, and L. Choridah. 2016. Feature Extraction for Lesion Margin Characteristic Classification from CT Scan Lungs Image. *Proceedings - 2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2016*. Doi: 10.1109/ICITISEE.2016.7803047.

[18] A. Baratloo, M. Hosseini, A. Negida, and G. El Ashal. 2015. Part 1: Simple Definition and Calculation of Accuracy. *Sensitivity and Specificity*. 3: 48-49.