Detection of Lesion of Lung Cancer in CT Scan Images using Watershed Segmentation

Nur Najihah Sofia Mohd Marzuki, Siti Noraini Sulaiman*, Iza Sazanita Isa, Noor Khairiah A.Karim, Nooritawati Md Tahir

***Abstract*— This paper presents a lesion of lung cancer detection exist in CT scan images using watershed segmentation. Lung cancer is a disease of uncontrolled cell growth in tissues of the lung. It seems to be the common cause of death among people throughout the world. Therefore, diagnosis of lung cancer at early stage can help doctors to treat patients and keep them alive. The main aim of this research is to establish an image processing method for segmentation of lung cancer from CT scan images by using image processing techniques. In order to achieve the main aims, the work is divided into two parts, first is obtaining lung region from CT scan images and second is detecting the lesion of lung cancer. Hence, this paper will present the outcome of the second part. The lung lobes and nodules or lesion in CT image are segmented using two techniques which are convolution watershed and modified watershed within two stage approaches. Firstly, the image will undergo threshold, clustering and image filtering as well as enhancement process to get better and clearer lung area image. This stage is known as pre-processing stage where this stage is needed to improve the quality of images as the images normally comes with some unwanted information that may obscure some feature that are important. In this stage, the distortion or noise will be removed in order to enhance the important features in the images. Next, is the most important stage in this research which is the segmentation stage since convolution and modified watershed is considered as an essential step in medical image analysis, classification and detecting the lesion of lung cancer. Then, the performance of segmentation process is measured whether or not it improves the accuracy, recall, precision and F- score parameters. The comparative analysis is made on 50 set of images. The outcome of this research is very helpful for the doctor to determine the type of treatment should be provided to the patient. and to diagnose the lung cancer and nodule on the image.**

***Keywords*—Computed Tomography (CT) scan, Image Segmentation, Lung nodule, Watershed Segmentation**

This manuscript is submitted on 25th September 2020 and accepted on 18th December 2020.N.N.S.M Marzuki is a Postgraduate student, Faculty of Electrical Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang, Kampus Permatang Pauh, Permatang Pauh, Pulau Pinang, Malaysia (e-mail:najihahsofia93@gmail.com).

S.N.Sulaiman, Faculty of Electrical Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang, Kampus Permatang Pauh, Permatang Pauh, Pulau Pinang, Malaysia (*e-mail:sitinoraini@uitm.edu.my).

I. S. Isa, Faculty of Electrical Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang, Kampus Permatang Pauh, Permatang Pauh, Pulau Pinang, Malaysia (e-mail:izasazanita@uitm.edu.my).

N.K.A.Karim, Regenerative Medicine Cluster, Advanced Medical and Dental Institute Universiti Sains Malaysia, Malaysia (e-mail:drkhairiah@usm.my).

N.M. Tahir, Faculty of Electrical Engineering, Universiti Teknologi MARA, Selangor, Malaysia (e-mail: noori425@uitm.edu.my).

I. INTRODUCTION

LUNG cancer is one of the most serious disease among human in the world such as in the U.S and worldwide [1]. Based on the studies from [2], almost 8.2 million and 1.59 million deaths in 2012 and 2014 are due to lung cancer. In Indonesia, lung cancer becomes the major cause of the deaths for men in the year of 2014 [3]. The number one cause of deaths among men in Malaysia is lung cancer based on the World Health Organization report that released in April 2011[4]. There are two major types of lung cancer, Small Cell Lung Cancer (SCLC) and Non- Small Cell Lung Cancer (NSCLC) [5].

Lungs are the organ that located in the chest in the rib cage. The lungs are the pair of spongy, air-filled organs located on either side of the chest. The right lung has three lobes and left lung has two lobes [6]. According to the Khin and Aung [7], the types of lung cancer disease can be divided into four stages. In stage I, the cancer is confined to the lung. The cancer is confined to the chest on stage II and III, but with larger and more invasive tumors classified as stage III. For stage IV, cancer spreads from the chest to other parts of the body.

CT imaging is the imaging procedure for the lung used to detect pulmonary disease such as lung cancer, tumour, and cystic fibrosis. The diagnosis and treatment of lung cancer are dependent on the type of lung cancer. CT scan become the most reliable method for early detection of cancer as has produced good detection of classification with lower cost, takes shorter time to conduct the procedure and widespread availability [8]. Compare to MRI and X-ray, CT scan are more useful due to its advantage in terms of the low noise and better clarity.

Currently, computer aided diagnosis (CAD) is one of a essential part of diagnosis procedure for early detection. Whereby, CT image of the lung is processed to detect the occurrence of cancer nodule. Occasionally, a radiologist uses CAD systems for pre-processing of the image to assist in discovering the most possible locations for lung nodules. The first step is suppressing the background structure in lungs like ribs, bronchi, and blood vessels. Later, the nodules are categorized based on size, contrast, and shapes by using preprocessing algorithm such as Median filter, Average filter, and Wiener filter. Watershed algorithm is use in segmentation process to divide the lung lobes from CT image and suspicious regions classification.

Watershed segmentation is one of the methods that commonly used in image segmentation, features extraction and surface visualization [9]. The basic concept of Watershed segmentation is based on visualizing an image in three dimensions, which are two spatial coordinates versus gray levels. There are three types of points to be studied in watershed segmentation which are points that have its place to regional minima, points at which the drop of water is positioned and the points at which the water is uniformly likely to befall [10]. S. Makaju et al. [11] applied watershed segmentation in their research to detect lung cancer and the outcome of their research is quite convincing with the accuracy of 92%.

To date, there has been many system developed and on going research conducted for lung cancer detection. However, some systems do not have satisfactory accuracy of detection and some systems need to be improved to achieve highest accuracy tending to 100%. Therefore this study is conducted to improve the gaps left from the previous researchers. The objective of this study is to obtain the lung region region from CT scan images and for this paper, the emphasize is in detecting the lesion of lung cancer. This paper proposed modified watershed algorithm for the image segmentation of the lesion of lung cancer. The rest of the paper is organized as follows. In Section II the methodology of this project is explained. Section III is the result and discussion whereby Section IV is the conclusion of this paper.

II. Methodology

This section discusses in detail the entire process of the proposed methodolgy. In general, the method comprises four (4) parts which are data/image acquisition, image pre-processing, image segmentation, detection of lung cancer and performance evaluation. Figure 1 shows the block diagram of the overall methodology. The following sub-sections will discuss each part of the methodology in detail:

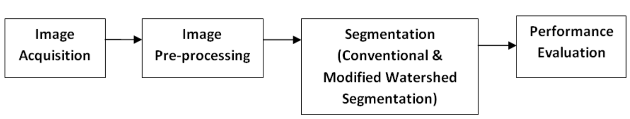


Fig. 1. Block Diagram Of The Overall Methodology

A. *Image acquisition*

Image acquisition is the first part of preparing raw images to be processed for any computer vision system. For this research work, CT images has been acquired from Advanced Medical and Dental Institute (AMDI), Universiti Sains Malaysia database and from The Cancer Imaging Archive (TCIA Images) in the form of DICOM (Digital Image Communications in Medicine). The data were collected restrospectively from 50 subjects. The images were sorted and loaded into a MATLAB 2017b software.

For the purpose of presentation the result in this paper, 5 images are used as shown in Figure 2.

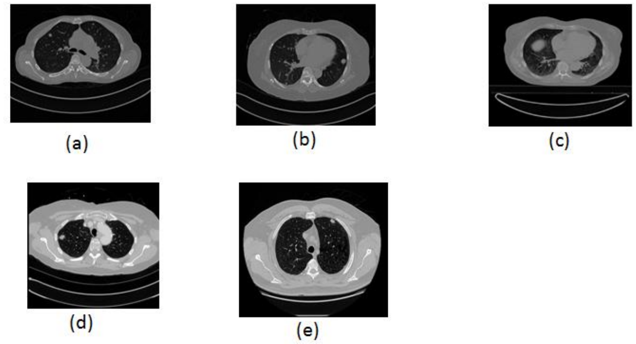


Fig. 2. The original image of lung region

B. *Image pre-processing*

The aim for the pre-processing stage is to improve the contrast and clarity of the images as well as to separate the lung region from its background. There are many techniques such as smoothing, and a few enhancement techniques including the image segmentation technique are applied to the image in order to improve the quality of image in this pre-processing stage. The details explanation on the implementation of this stage was discussed in [12]

C. *Segmentation*

The main purpose of the segmentation stage is to simplify the representation of an image into more significant and easier understanding image [13]. Watershed segmentation is implemented in this image segmentation stage for the purpose of detecting the lesion of cancer from the CT scan image. In this paper, we focus on the lesion segmentation. Nodule or lesion is a mass of tissue located in the lung. It appears as round, white shadows on CT image [14]. If the mass of 25mm or larger, it can cause cancer. Therefore, it is important to segment lung nodules. Conventional and modified Watershed sementation is used to segment the lung lobes and lung nodules.

1) Conventional Watershed Segmentation:

To get a smooth region, Marker- Controlled Watershed segmentation is used in this project. The process of level set segmentation is used to obtain the extracted nodule image by allowing the image undergoes the thresholding, clustering and enhancement processes as well as the circular Hough transform to find the circle region (the lesion region). Figure 3 shows the block diagram of the conventional watershed method.

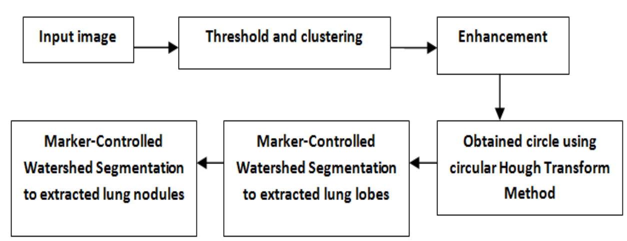


Fig. 3. Block diagram of the conventional watershed method

2) Modified Watershed Segmentation

In this research a modified watershed segmentation is proposed since the preferred result cannot be obtained directly and effectively by the conventional method. The level set segmentation process which consist of the thresholding, clustering and enhancement, as well as the discovery of circle using circular Hough transform process from the previous stage remain the same in order to provide a preliminary segmentation of the lung nodule. These level set segmentation process and watershed process complement each other. In order to get an effective, better smooth and sufficiently robust segmented region, the convolution of the Marker-controlled watershed with the watershed ridge lines is performed. In this work some modification is applied to the watershed ridge lines process to suit with the lung CT scan image environment. Figure 4 shows the block diagram of the modified watershed method.

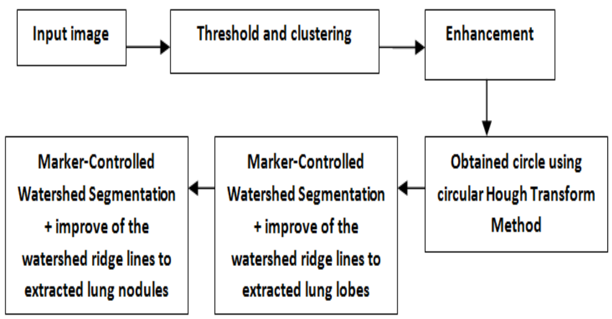


Fig. 4. Block diagram of the modified watershed method

D. *Performance Evaluation*

The performance evaluation was divided into two part which are qualitative and quantitative. The qualitative assessment is done by observation using human eyes which is then validated by radiologist, while quantitative assessment is calculated by referring to the accuracy, precision, recall and F-score test. The formula for each performance indicator were presented below [15]:

$$Recall = \frac{TP}{TP+FN} \tag{1}$$

$$Precision = \frac{TP}{(TP+FN)} \tag{2}$$

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{3}$$

$$Fscore = \frac{2\,(Precision \times Recall)}{(Precision + Recall)} \tag{4}$$

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

Recall also known as sensitivity; is the fraction of the total amount of relevant instances that were actually retrieved. Precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances. It measures the closeness of agrement between predicted output of the system and actual result is the meaning of accuracy. The result of F-score is obtained from the calculated precision and recall where according to [16], F-score is a measure of a test's accuracy.

III. RESULTS AND DISCUSSION

A. *Qualitative*

This section will discuss in detail the result in qualitative for obtaining the lung region and detecting the lesion of lung cancer. For the purpose of result presentation and discussion in this paper, five (5) images of lung with different shape of lung area are presented, tested for the proposed modified watershed segmentation and compare it with the conventional watershed. As mentioned, the primary aim of this research is to establish an image processing method for segmentation of lung cancer from CT scan images by using image processing techniques. In order to achieve that, the work is divided into two parts, first is obtaining lung region from CT scan images and second is detecting the lesion of lung cancer. The outcome of the first part are presented in [12].

Figure 5 shows the resultant image of lung lobes of while Figure 6 shows the nodule of lung cancer obtained by conventional watershed method and proposed modified watershed method.

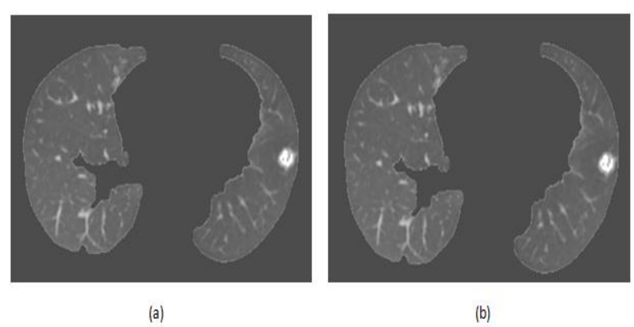


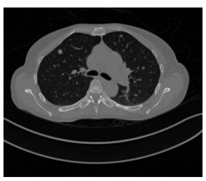
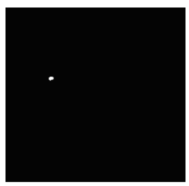
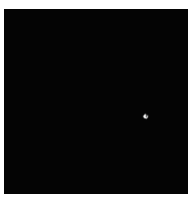
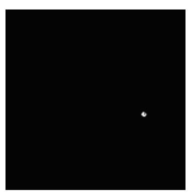
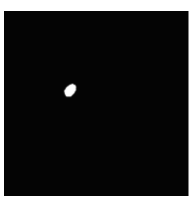
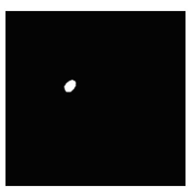
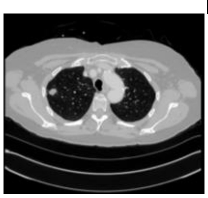
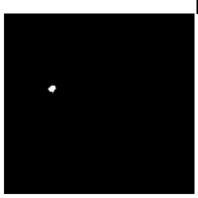
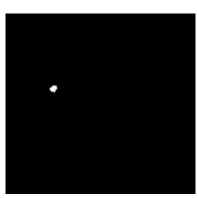
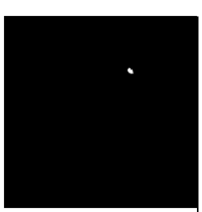
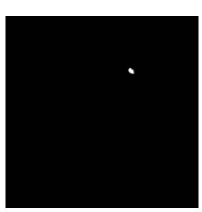
Fig. 5. The resultant image lung lobes of (a) Conventional watershed (b) improve watershed method.

Meanwhile, Table I shows the outcome of the second part that is detecting the lesion of lung cancer. It can be seen that both methods able to detect the lung lobes and nodule but the proposed modified watershed method gives the better, clearer and more enhanced lesion region. It can be seen at the edge of the lesion is smoother and the region between the lung cancer area and the lesion part is more enhanced .

(a) (b)

Fig. 6. The resultant image the nodule of lung cancer of (a) Conventional watershed (b) improve watershed method.

TABLE I
RESULT OF DETECTION OF LESION OF LUNG CANCER

Original Image	Conventional watershed	Proposed Modified watershed
 Image 1		
 Image 2		
 Image 3		
 Image 4		
 Image 5		

In other word, the proposed modified watershed method, provide a better comprehensive image than the conventional watershed. This can be observe from the improve of the watershed ridge lines, the smoother edge and better enhanced lesion region of the resultant image obtained by the proposed watershed method, as compared to the conventional watershed.

B. Quantitative

This section presented a quantitative result. The performance is measured based on statistical parameters of precision, recall, accuracy and F-score is presented in Table II. The average performance of detecting the lesion of lung cancer is as presented in Table III. Conventional Watershed segmentation gave an average value of precision 99.54%, recall 96.13%, accuracy 99.41% and 97.80% for F-score. The proposed modified Watershed segmentation gave a slightly higher performance which are precision 99.80%, recall 98.41%, accuracy 99.93% and 99.09% for F-score

TABLE II
PERFORMANCE OF DETECTING THE LESION OF LUNG CANCER

Segmentation	Precision	Recall	Accuracy	F-score
Image 1	100	93.10	99.99	96.43
Image 2	99.98	100	99.98	99.99
Image 3	99.86	100	99.86	99.93
Image 4	99.16	100	99.89	99.57
Image 5	100	99.11	99.92	99.55

TABLE III
AVERAGE PERFORMANCE OF DETECTING THE LESION OF LUNG CANCER

Segmentation	Precision	Recall	Accuracy	F-score
Conventional Watershed	99.54	96.13	99.41	97.80
Proposed Modified Watershed	99.80	98.44	99.93	99.09

Based on both analyses, the proposed modified Watershed segmentation method gave the better percentage in all the tested parameters. The proposed modified Watershed segmentation also shows better contrast of the lung lesion compared to conventional watershed. This method works by determine the area of the lung and then determind the lung lesion that exist in the lung area.

IV. CONCLUSION

This paper proposed a modified watershed segmentation method for detecting lung nodule of lung cancer from CT scan images. This experiment adds to a growing corpus of research showing that it is possible to detect lesion that exist in CT Lung images via a proper image processing technique. The proposed method gives a better performance as compared to the conventional watershed with the accuracy of 99.93%. We have demonstrated the effectiveness of the method by a few examples from CT scan images

ACKNOWLEDGMENT

This research work is financially supported by Ministry of Higher Education Grant Scheme (FRGS) " Formulation of A New Watershed-based Segmentation Method for Features Extraction of Lung Cancer nodule in CT-scan Images" (Ref: FRGS/1/2019/TK04/UITM/02/15). Ethics approval is obtained from Universiti Sains Malaysia (USM/JEPeM/19040231). The author are grateful to Universiti Teknologi MARA, Cawangan Pulau Pinang, Kampus Permatang Pauh, Pulau Pinang, Research Group Advanced Rehabilitation Engineering in Diognostic and Monitoring (AREDiM) and Advanced Control System and Computing Research Group (ACSCRG) for their contribution that help in finishing this research.

REFERENCES

[1] P. Su *et al.*, “A Fast CT and CT-Fluoroscopy Registration Algorithm With Respiratory Motion Compensation for Image-Guided Lung Intervention,” *IEEE Trans. Biomedial Eng.*, vol. 60, no. 7, pp. 2034–2041, 2013.

[2] P. Yugander, “Liver Tumor Segmentation in Noisy CT Images using Distance Regularized Level Set Evolution based on Fuzzy C-Means Clustering,” *2017 2nd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol.*, pp. 1530–1534, 2017.

[3] Y. F. Riti and H. A. Nugroho, “Feature Extraction for Lesion Margin Characteristic Classification from CT Scan Lungs Image,” *1st Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng.*, pp. 54–58, 2016.

[4] M. Assefa, I. Faye, A. S. Malik, and M. Shoaib, “Lung Nodule Detection Using Multi-Resolution Analysis,” *Int. Conf. Complex Med. Eng.*, pp. 457–461, 2013.

[5] K. Varalakshmi, “Classification of Lung Cancer Nodules using a Hybrid Approach,” *J. Emerg. Trends Comput. Inf. Sci.*, vol. 4, no. 1, pp. 63–68, 2013.

[6] S. B. J. Nitheesha S, “Lung Cancer Recognition in CT Image Using Watershed Segmentation,” *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.*, vol. 5, no. 4, pp. 3405–3409, 2016.

[7] K. Mya, M. Tun, and A. S. Khaing, “Feature Extraction and Classification of Lung Cancer Nodule using Image Processing Technique,” *Int. J. Eng. Res. Technol.*, vol. 3, no. 3, pp. 2204–2211, 2014.

[8] V. Rani and J. Jawhar, “Emerging Trends in Lung Cancer Detection Scheme- a Review,” *Int. J. Res. Anal. Rev.*, vol. 5, no. 3, 2018.

[9] X. Zhang, S. Member, J. Tian, D. Xiang, X. Li, and K. Deng, “Interactive Liver Tumor Segmentation from CT Scans Using Support Vector Classification with Watershed,” pp. 6005–6008, 2011.

[10] S. Ignatious and R. Joseph, “Computer Aided Lung Cancer Detection System,” *2015 Glob. Conf. Commun. Technol.*, no. Gcct, pp. 555–558, 2015.

[11] S. Makaju, P. W. C. Prasad, A. Alsadoon, A. K. Singh, and A. Elchouemi, “Lung Cancer Detection using CT Scan Images,” *Procedia Comput. Sci.*, vol. 125, no. 2009, pp. 107–114, 2018.

[12] N. Najihah *et al.*, “Demarcation of Lung Lobes in CT Scan Images for Lung Cancer Detection using Watershed Segmentation,” pp. 1–5, 2017.

[13] P. R. Katre, “Detection of Lung Cancer Stages using Image Processing and Data Classification Techniques,” *2017 2nd Int. Conf. Converg. Technol.*, pp. 402–404, 2017.

[14] A. S. and Sridhar.S2, “SEGMENTATION OF LUNG LOBES AND NODULES IN CT IMAGES,” *Signal Image Process. An Int. Journal(SIPIJ)*, vol. 1, no. 1, pp. 1–12, 2010.

[15] I. S. Isa, “An Automated Multimodal White Matter Hyperintensities Identification In MRI Brain Images Using Image Processingitle,” 2017.

[16] I. Segmentation, W. Algorithm, K. R. Shrimali, and I. Technology, “ImageSegmentationLung,” pp. 1–6, 2018.

Nur Najihah Sofia Mohd Marzuki completed her Bachelor of degree in Electrical Engineering at Faculty of Electrical Engineering Universiti Teknologi Mara, Pulau Pinang in the year of 2014. Her main interest in research are image processing technique.

Siti Noraini Sulaiman received PhD degree in image processing from School of Electrical and Electronics Engineering, Universiti Sains Malaysia in 2012. She is currently an Associate Professor at Faculty of Electrical Engineering, Universiti Teknologi MARA, Malaysia. She is now attached as the chair of Advanced Rehabilitation Engineering in Diagnostic and Monitoring (AREDiM) and the deputy chair of Advanced Control System and Computing Research Group (ACSCRG), Faculty of Electrical Engineering, Universiti Teknologi MARA, Penang Branch Campus in Permatang Pauh, Penang, Malaysia. Her main research interest is biomedical engineering focusing on image filtering, image clustering and algorithm for image processing

Iza Sazanita Isa received M.Sc. degree in Electronic System Design Engineering from the School of Electrical and Electronics Engineering, Universiti Sains Malaysia, Malaysia, in 2009. She received her Ph.D. degree from Universiti Teknologi MARA, Malaysia. Her research focus is on the image processing of medical imaging which is more specifically is the segmentation of the MRI images and deep learning

Noor Khairiah A. Karim received the Bachelor Degree in Medicine, Bachelor Degree in Surgery (MBBS) and Master of Radiology (MRad) from the University of Malaya, Malaysia. She then obtained her Fellowship in Cardiac Imaging from the National Heart Center Singapore. She is currently a Senior Medical Lecturer of the Regenerative Medicine Cluster, and a Consultant Radiologist at the Advanced Medical and Dental Institute, Universiti Sains Malaysia. Her current research areas include medical image processing and analysis with special interest in cardiac, breast and brain imaging.

Nooritawati Md Tahir received her PhD in Electrical, Electronic & System Engineering from Universiti Kebangsaan Malaysia. She is currently a Professor at the Faculty of Electrical Engineering, Universiti Teknologi MARA, Malaysia. Her research interests include pattern recognition, computational intelligence and artificial intelligence