

A novel CAD system to automatically detect cancerous lung nodules using wavelet transform and SVM

Ayman AbuBaker¹, Yazeed Ghadi²

¹Department of Electrical Engineering, Applied Science Private University, Jordan

²Department of Electrical Engineering, Al Ain University of Science and Technology, Uni Emirat Arab

Article Info

Article history:

Received Jan 29, 2019

Revised Mar 23, 2020

Accepted Apr 30, 2020

Keywords:

Cancer detection

Computed tomography

DICOM

Wavelet features

Wavelet transform

ABSTRACT

A novel cancerous nodules detection algorithm for computed tomography images (CT-images) is presented in this paper. CT-images are large size images with high resolution. In some cases, number of cancerous lung nodule lesions may missed by the radiologist due to fatigue. A CAD system that is proposed in this paper can help the radiologist in detecting cancerous nodules in CT- images. The proposed algorithm is divided to four stages. In the first stage, an enhancement algorithm is implement to highlight the suspicious regions. Then in the second stage, the region of interest will be detected. The adaptive SVM and wavelet transform techniques are used to reduce the detected false positive regions. This algorithm is evaluated using 60 cases (normal and cancerous cases), and it shows a high sensitivity in detecting the cancerous lung nodules with TP ration 94.5% and with FP ratio 7 cluster/image.

Copyright © 2020 Institute of Advanced Engineering and Science.

All rights reserved.

Corresponding Author:

Ayman AbuBaker,

Department of Electrical Engineering,

Applied Science Private University,

116-11931, Shafa Badran, Amman, Jordan.

Email: a_abubaker@asu.edu.jo

1. INTRODUCTION

Lung cancer is one of the most relevant public health issues in United states, Europe and Middle East [1, 2]. Early Detection and treatment of this types of cancer is require to effectively overcome this burden. As an initial and cheapest method to detect the nodule lung cancer is Chest X-ray. Computed tomography (CT) as a second diagnosis stage is the best imaging modality for the detection of small pulmonary nodules, particularly since the introduction of the helical technology [3, 4]. The CT images are a high resolution images with high amount of data storage. Therefore, researchers tries to help the radiologist to easily process these huge image and automatically detected the potential nodule lung cancers using computer aided diagnosis system (CAD) [5, 6].

Detection nodule lung cancer is one of the most difficult cases for the radiologist specially in CT images since they are closely connected to the surrounding parenchymal tissue [7, 8]. Therefore, visual appearance for cancerous nodules have similar visual characteristics of normal tissues [9]. Therefore, this paper proposed a novel detection and classification method for cancerous cells. Three main stages are used to accurately detect the cancerous nodules in the CT lung images[10].

The paper present a brief description about the lung CT image in section 2. Then literature review section which presented in section 3. Proposed detection and classification algorithm is introduced in Section 4. Finally, discussion and conclusion are presented in Section 5 and Section 6 respectively.

2. DATABASE

The database that are used in this paper in downloaded from cancer imaging archive (TCIA), which is organized into purpose-built collections of subjects [11]. This database has different kind of high resolution images of cancer type and/or anatomical site (lung, brain, etc.) in common. Moreover, the database include a huge dataset of MRI, CT, and X-Ray images that are stored as DICOM file format [12].

3. LITERATURE REVIEW

Many authors proposed different CAD techniques to help the radiologist in diagnoses lung cancer in CT images. One of these techniques is using Laplacian of Gaussian filter as in [13]. Their algorithm is divided to two main stages, transmission of high intensity and LOG Filter. These stages are applied calculate the contrast differences inside and outside region of interest. Snake algorithm is another enhancement and segmentation tool used to find the internal energy in CT image like Vivekanandan D et al [14]. Messay et al. [15] implement local contrast enhancement filter that follow the nodule enhancement method to detect the chest radiographs. Dots, lines, and planes enhancement techniques are used by Li et al. [16]. Such these enhancement techniques can be implemented on specific shapes and suppress other objects. So, as an initial step the CT-Image is blurred then a Gaussian kernel filter (GKF) of a nodule size is implemented to detect the nodule. As final stage, multi-scales GKF are used as to find a match with the nodule size. Yamamoto et al. [17] used a statistical enhancement filter to enhance CT images. They implemented the Quoit Filter that has large ring and disk filters. Differential intensity value procedure is implemented in this algorithm between the internal and external disks in order to find a potential region of interest. Another segmentation algorithm is presented by Maciej Dajnowiec et al. [18]. Initially, an optimum threshold value was calculated based on different image data sets. Threshold algorithm then is implemented to segment lung region from CT image. After that, connecting component labeling (CCL) is applied to detect remaining regions in the image. Therefore, lung nodule can successfully segmented but have lobes with nodules region. On the other hand, CAD system is a powerful tool in detecting cancerous nodules in the CT images. Another CAD system to detect cancerous lung region is proposed by Lee et al. [19]. Authors implement GA to set a target position and also to match the best template image from previous data sets. Four generations every time are established based on grey level of Gaussian distribution. Algorithm sensitivity was 85% in detecting cancerous nodules in the CT-images. Other author like Opfer et al [20] used distance transformation technique to improve the performance of CAD system. The distance transformations of various thresholds and subsequent crest line extraction are used enhance the CAD sensitivity. CAD system performance is slightly improved when using this technique since it produce large number of FP regions. Another CAD system is used by Moreover, Golosio et al. [21] to detect cancerous nodules in CT images. They used multi-adaptive threshold surface triangulation approach in the detection algorithm. The proposed algorithm show good sensitivity in detection cancer nodules but many detected FP are presents. K.Devaki et al. [22] develop a technique to accurately segment lung region from CT- images. The proposed algorithm can efficiently segment the lung regions from CT-images.

4. RESEARCH METHOD

The novel CAD system is used to accurately detect the cancerous nodules in CT- image with minimum number of false positive regions. This technique is mainly divided into four stages. Enhancement the suspicious regions using adaptive average method is the first stage. Detection region of interest algorithm that have the cancerous nodule is implemented in the second stage. In order to reduce the detected false positive regions, four wavelet features are used to reduce the detected FP regions. These wavelet features are used with SVM in order to classify the detected regions as FP and TP regions. The proposed algorithm is shown in Figure 1.

4.1. CT-image enhancement

The fundamental operation needed to assist cancerous nodule in CT-image is contrast enhancement. In many image processing applications, the Laplacian filter is one of the simplest and effective techniques for intensity enhancement that presented in as (1). Laplacian filter improves contrast of the cancerous nodules in CT-image by applying the Laplacian mask of size 9×9 .

$$S(x, y) = f(x, y) + c[\nabla^2 f(x, y)] \quad (1)$$

where $S(x,y)$ is the intensity value for the processed image, $f(x,y)$ the intensity value for the input image and c is consider as one in this paper.

Cancerous nodules appear on digitized CT as small regions of size less than 4 mm [23], with intensity values higher than their surrounding background. So it is not easy to enhance the cancerous nodules regions since surrounding lung tissue makes the abnormality areas almost invisible. Therefore, the modified average filter is implemented to smooth the edges of the processed Laplacian enhancement image [24]. This in case will slightly enhance the Cancerous nodules regions to be easily detected in the next stage.

After extensive analysis of 60 Cancerous CT images, we concluded that all cancerous nodules have grey scale values in the range from 80 to 230. In accordance with these observations, each CT-image is processed using the modified average filter that presented on (2).

$$S_k = \frac{1}{mn} \sum_{j=80}^{230} r_j \quad (2)$$

where S_k is the intensity value for the processed image, r_j the intensity value for the input image and m and n are the mask size. After processing the Laplacian and modified average filter, the cancerous nodule regions become slightly brighter corresponding to the neighbor regions as shown in Figure 2. This will assist in detection the region of interest that will be discussed in the next section.

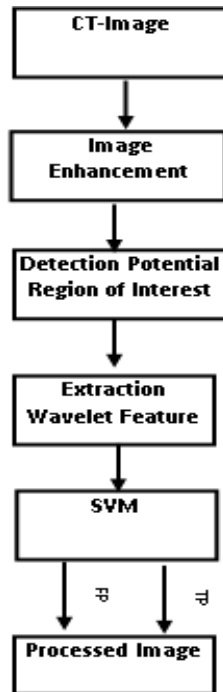


Figure 1. Proposed CAD system

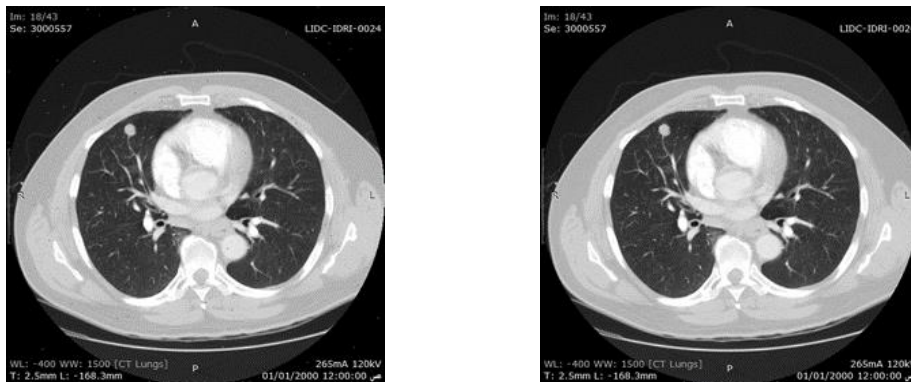


Figure 2. Cancerous nodule enhancement result

4.2. Potential region of interest

In order to detect cancerous nodules region, two concentric circular masks are used as shown in Figure 3. When centered of the cancerous nodule, the inner masked region included the cancerous nodule while the outer masked region included the surrounding region. Based on CT-Image resolution which is $45 \mu\text{m} \times 45 \mu\text{m}$, the inner mask of the filter is determined. Whereas, the outer mask size is computed as trial and error and it was found that mask of size $120 \mu\text{m} \times 120 \mu\text{m}$ is significant to be used in this case [25].



Figure 3. Two concentric circular masks

The two concentric circular masks was tested on 60 CT-images and it found that all cancerous nodules are detected with large number of false positive regions. Detection the PROI cluster is designed based on the fact that the cancerous nodules are brighter than the neighbor pixels. Therefore, in order to select PROI two conditions should be satisfied, average value for the inner mask should be greater than outer mask and the intensity pixel value of the center of the inner mask should be the highest intensity in the mask.

After processing 60 CT-images using PROI algorithm, it was noticed that all cancerous nodules in the CT-images are detected but many detected false positive (FP) regions as shown in Figure 4. This in case will reduce the sensitivity of the proposed CAD system. Therefore, wavelet features will be applied to reduce number of detected FP clusters and increase the sensitivity of this CAD system.

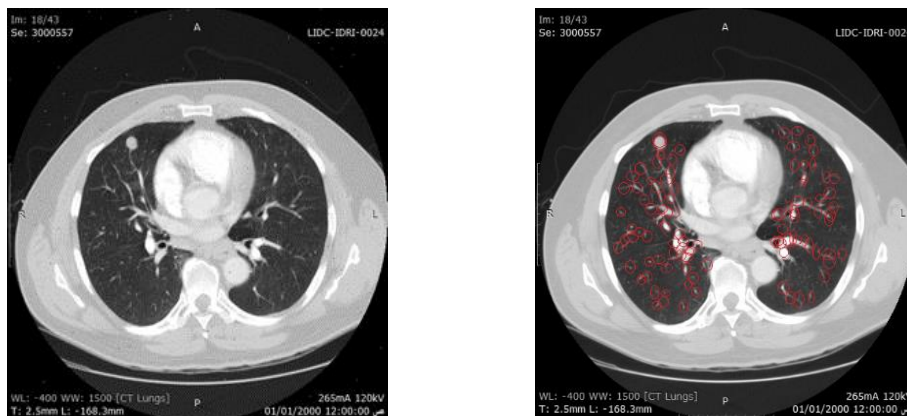


Figure 4. PROI processing result

4.3. Wavelet features

Wavelet Daucechies (BD4) transform is used in this paper to generate the wavelet coefficient that will be used to classify the detected regions to TP and FP regions. ROI of size 13×13 pixels is implemented using wavelet DB4 to generate TP wavelet coefficient features which are minimum_value and maximum_value of the coefficients, average_value of the coefficients, and standard_deviation between the coefficients. These features are extracted from low frequency image at Daucechies (BD4) with level 2 as shown in Figure 5.

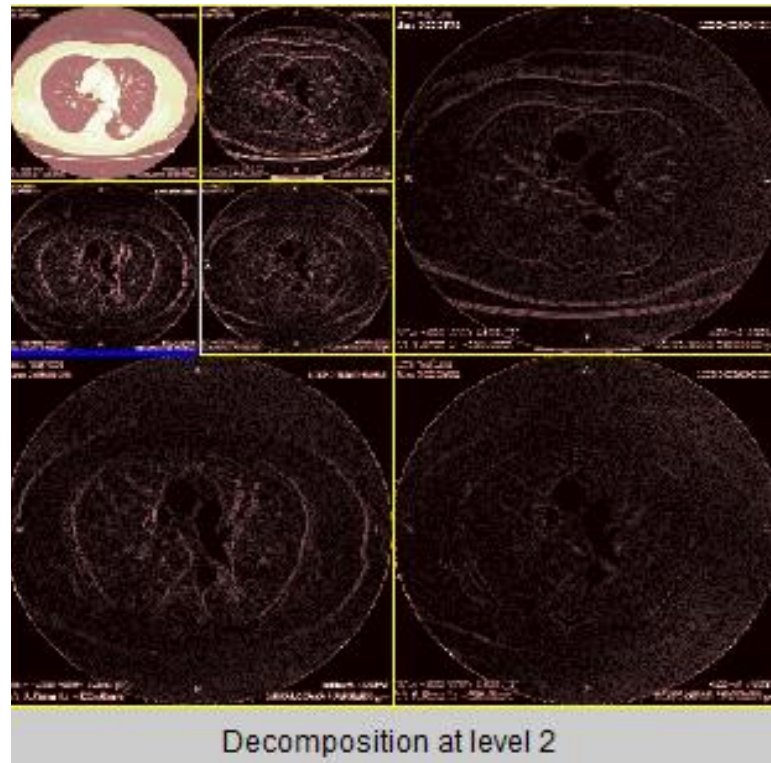


Figure 5. Wavelet decomposition at level 2

4.4. Support vector machine

Classification and regression technique is used in this paper to classify the TP and FP region using support vector machines (SVM). SVM is known to be an excellent tool for binary classification problems, similar to the one here, by seeking the optimal separating hyperplane that provides efficient separation of the data and maximizes the margin. SVM is mainly divided into two main stages: collecting data stage and learning process stage. These stages are presented as follow:

4.4.1. Collecting data stage

In SVM both input and output data should be known. So, the four wavelet features are extracted from 50 normal regions and 60 cancerous nodule region. As a results, we generate an input vectors: four vectors of size 50 for normal cases and four vectors of size 60 for cancerous cases. On the other hand, two outputs nodes are used to classify the cases to normal and abnormal cases based on output vector of 0.1 and 0.9 for normal and abnormal cases respectively.

4.4.2. Learning process stage

The values of the input matrix are arranged as a training vector in a manner similar to the Jackknife technique, where 70% of the inputs data were used for the SVM training phases and the remaining 30% were used for the SVM testing phases. ANOVA kernel technique is considered as a best classification tool comparing with other kernel methods. The ANOVA kernel, which is shown in (3), has two parameters, the gamma (γ) parameter and the exponential degree (d) parameter. These two parameters control the shape of the kernel.

$$K(x, y) = \left[\sum_i \exp(-\gamma(x_i - y_i)) \right]^d \quad (3)$$

Therefore, different experimental results are implemented to find the optimal value to gamma and the degree as shown in Figure 6.

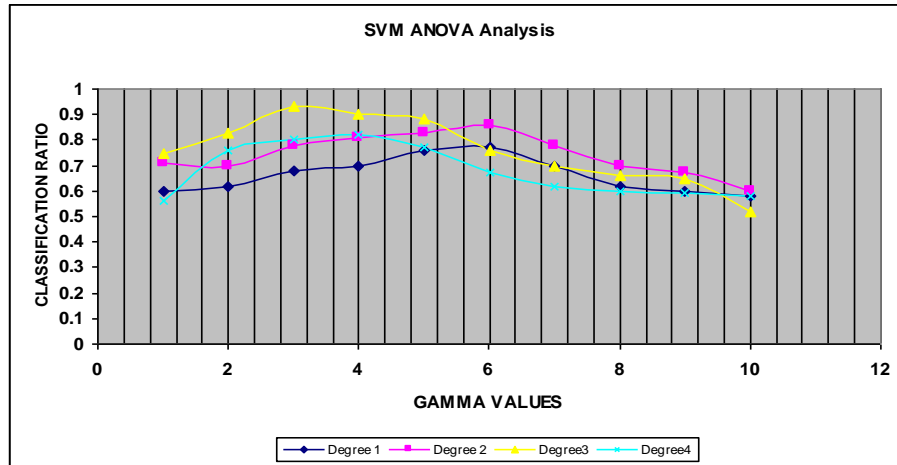


Figure 6. SVM ANOVA analysis curves

5. ALGORITHM EVALUATION

The cancerous nodule enhancement and detection algorithm is applied on 60 CT images. Then, four wavelet coefficient features are generated. The algorithm is subjectively evaluated using three radiologist, where number of detected FP regions are counted per image. Also TP percentage of each image is also recorded. Finally, the average of detected FP region and TP percentage is presented in the Table 1 after processing 60 CT images. From Table 1, it is clearly noticed the proposed CAD system can accurately detect the cancerous nodules with minimum number if detected FP regions. So the TP ratio achieved 94.5% with FP rate 7 clusters/ image. Figure 7 show complete CAD system detection stages.

Table 1. Image quality evaluation

Variable	Average of detected FP region (cluster)	Average of TP (%)
Radiologist 1	8	93
Radiologist 2	6	95
Radiologist 3	9	96

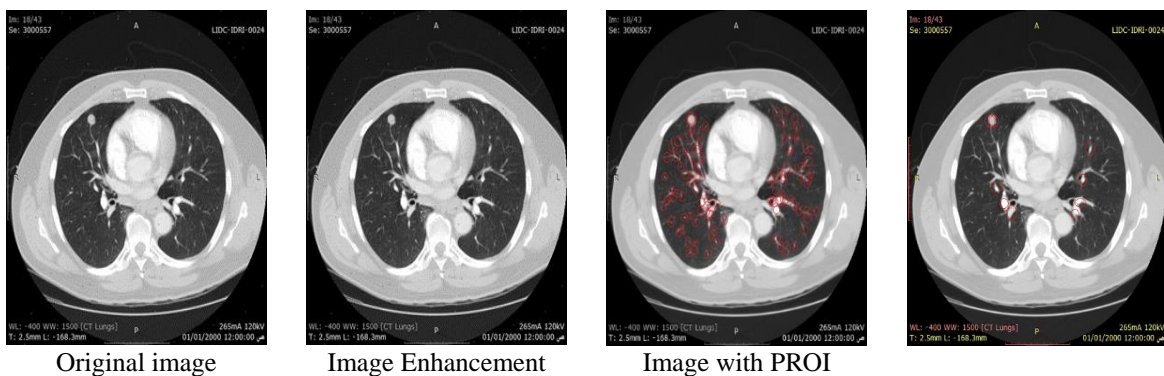


Figure 7. Accurate detection of cancerous nodules

6. CONCLUSION

An adaptive CAD system that is used to accurately detect the cancerous nodules in CT-images is proposed in this paper. The proposed algorithm is divided into four stages. The cancerous nodules are enhanced using the Laplacian filter. Then, the average filter is modified based on the lower and upper grey levels of the cancerous nodules in the CT images. This incase, slightly enhances the cancerous nodules in the mammogram images. In the second phase, the potential region of interest is detected based on the visual appearance of cancerous nodules regions in CT-Images. The processed CT-image has many detected FP

regions, therefore, SVM and wavelet features are used to reduce the detected FP regions. So four wavelet features are generated then SVM is implemented to classify the detected regions to TP and FP regions. As a result, the proposed algorithm is subjectively and objectively tested on 60 CT images and it shows that its algorithm can detect the cancerous nodules with an average rate 94.5% with FP regions of 7 cluster/image.

ACKNOWLEDGEMENTS

The author is grateful to the Applied Science Private University, Amman, Jordan, for the full financial.

REFERENCES

- [1] Shah, S. K, et al., "Computer aided characterization of the solitary pulmonary nodule using volumetric and contrast enhancement features," *Academic radiology*, vol. 12, no. 10, pp. 1310-1319, 2005.
- [2] Lo, S. B, et al., "Computer-aided detection of lung nodules on ct with a computerized pulmonary vessel suppressed function," *American Journal of Roentgenology*, vol. 210, no. 3, pp. 480-488, 2018.
- [3] Li, Q., "Recent progress in computer-aided diagnosis of lung nodules on thin-section CT," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 248-257, 2007.
- [4] Furqan, S. Gulistan, R. and Alejandro F. F., "Computer-aided detection of lung nodules: A review," *Journal of Medical Imaging*, vol. 6, no. 2, pp. 20901-201911, 2019.
- [5] Anita, C., Sonit, S. S., "Lung cancer detection on CT images By using image processing," *International Conference Computing Sciences*, no. 14-15, pp. 142-146, 2012.
- [6] Pu J, et al., "An automated CT based lung nodule detection scheme using geometric analysis of signed distance field," *Medical Physics*, vol. 35, no. 8, pp. 3453-3461, 2008.
- [7] Matsumoto S, et al., "Computer-aided detection of lung nodules on multidetector row computed tomography using three-dimensional analysis of nodule candidates and their surroundings," *Radiation medicine*, vol. 26, no. 9, pp. 562-569, 2008.
- [8] Steven S., et al., "Computer-aided detection improves detection of pulmonary nodules in chest radiographs beyond the support by bone-suppressed images," *Radiology*, vol. 272, no. 1, pp. 252-261, 2014.
- [9] Sousa, J. R., "Methodology for automatic detection of lung nodules in computerized tomography images," *Computer methods and programs in biomedicine*, vol. 98, no. 1, pp. 1-14, 2010.
- [10] Faridoddin, S, and Mojtaba M., "Application of CAD systems for the automatic detection of lung nodules," *Informatics in Medicine Unlocked*, vol. 15, no. 10, pp. 1001-1013, 2019.
- [11] "LIDC lung nodule image database," National cancer imaging archive, [Online], Available: <https://imaging.nci.nih.gov/ncia/>
- [12] Yingru Zhao, "Performance of computer-aided detection of pulmonary nodules in low-dose CT: Comparison with double reading by nodule volume," *European Radiology*, vol. 22, no. 10, pp. 2076-2084, 2012.
- [13] Zhenghao, Shi, et al., "A method for enhancing lung nodules in chest radiographs by use of LoGFilter," *2nd International Congress Image and Signal Processing*, pp. 1-4, 2009.
- [14] Vivekanandan, D. and Sunil, R.R., "A feature extraction model for assessing the growth of lung cancer in computer aided diagnosis," *IEEE-International Conference on Recent Trends in Information Technology*, pp. 953-958, 2011.
- [15] Messay T, Hardie R, and Rogers S., "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," *Medical image analysis*, vol. 14, no. 3, pp. 390-406, 2010.
- [16] Li Q, Li F, and Doi, K., "Computerized detection of lung nodules in thin-section CT images by use of selective enhancement filters and an automated rule-based classifier," *Academic radiology*, vol. 15, no. 2, pp. 165-175, 2008.
- [17] Yamamoto, S, et al., "Qoit Filter-a new filter based on mathematical morphology to extract the isolated shadow, and its application to automatic detection of lung cancer in X-ray CT," *Proc 13th International Conference of Pattern Recognition*, vol. 2, pp. 3-7, 1996.
- [18] Maciej, D., and Javad, A., "Computer simulation for segmentation of lung nodules in CT images," *IEEE International Conference Systems, Man and Cybernetics*, vol. 5, pp. 4491-4496, 2004.
- [19] Lee, Y, et al., "Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique," *IEEE Transactions on medical imaging*, vol. 20, no. 7, pp. 595-604, 2001.
- [20] Opfer, R, and Wiemker, R., "Performance analysis for computer aided lung nodule detection on LIDC data," *Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment*, vol. 6515, pp. 1-9, 2007.
- [21] Golosio B, Masala GL, Piccioli A, Oliva P, and Carpinelli M, "A novel multithreshold method for nodule detection in lung CT," *Medical physics*, vol. 36, no. 8, pp. 3607-3618, 2009.
- [22] K. Devaki and V. Murali Bhaskaran, "Study of computed tomography images of the lungs," *IEEE-International Conference on Recent Trends in Information Technology*, pp. 837-842, 2011.
- [23] Sahiner, B, et al., "Effect of CAD on radiologists' detection of lung nodules on thoracic CT scans: Observer performance study," *Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment*, vol. 6515, pp. 1-7, 2007.
- [24] A. A. Abu Baker, et al., "One Scan Connected Component Labelling Technique," *IEEE International Conference on Signal Processing and Communication*, pp. 1283-1286, 2007.
- [25] Hardie, R. C, Rogers, S. K, Wilson, T, and Rogers, A., "Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs," *Medical Image Analysis*, vol. 12, no. 3, pp. 240-258, 2008.