RESEARCH ARTICLE

Active Contour Based Segmentation and Classification for Pleura Diseases Based on Otsu's Thresholding and Support Vector Machine (SVM)

M Malathi[1]*, P Sinthia[1], K Jalaldeen[2]

Abstract

Objective: Generally, lung cancer is the abnormal growth of cells that originates in one or both lungs. Finding the pulmonary nodule helps in the diagnosis of lung cancer in early stage and also increase the lifetime of the individual. Accurate segmentation of normal and abnormal portion in segmentation is challenging task in computer-aided diagnostics**. Methods:** The article proposes an innovative method to spot the cancer portion using Otsu's segmentation algorithm. It is followed by a Support Vector Machine (SVM) classifier to classify the abnormal portion of the lung image. **Results:** The suggested methods use the Otsu's thresholding and active contour based segmentation techniques to locate the affected lung nodule of CT images. The segmentation is followed by an SVM classifier in order to categorize the affected portion is normal or abnormal. The proposed method is suitable to provide good and accurate segmentation and classification results for complex images. **Conclusion:** The comparative analysis between the two segmentation methods along with SVM classifier was performed. A classification process based on active contour and SVM techniques provides better than Otsu's segmentation for complex lung images.

Keywords: Lung- SVM classifier- Otsu's thresholding-Computer tomography- Active contour segmentation

Introduction

The lungs are the one of the primary organs of the respiratory system in the human body. A pair of spongy, air filled lungs is located on either side of the chest. The main function of the respiratory system is to remove oxygen from the atmosphere and transfer it into the blood stream. The trachea helps to conduct the inhaled air into the lungs via its tubular branches, called bronchi. It is further divided into smaller and smaller branches (bronchioles) finally it looks like microscopic air sacs called alveoli. Oxygen from the atmosphere is absorbed into the blood and carbon dioxide voyages from the blood to the alveoli. Lung cancer is one of the dangerous diseases, which occurs due to abnormal growth of the cells either in one or in both the lungs. It affects the proper supplying of oxygen from the lung to the human body via blood. The main reasons of lung cancer are smoking, toxic particles in the air, ageing etc., Early detection of lung cancer helps the patient to improve the life time. It is a very difficult task to locate the correct position of lung nodules. Hence the researchers developed a computer aided detection called lung nodule segmentation using Otsu's theresholding along with SVM classifier. Many medical imaging techniques are available to capture the lung image like chest X-ray, CT imaging techniques etc.

The accuracy of the segmentation is improved with the help of CT images. Many segmentation methods are available, among all these Otsu's thresholding based segmentation used and it is followed by SVM classifier. The exact boundaries of the lung cancer cells are accurately segmented and classified using active contour segmentation along with SVM classifier. CT imaging has the ability to produce 3D images of the chest, which provide greater resolution of nodules and tumour pathology. Digital X ray is also the imaging technique to detect the nodule size of 1 to 2 cm. The chest CT is one of the best equipment to find the abnormalities found in other imaging technique and also it helps to find the reason for unexplained cough, shortness of breath, chest pain, fever and other chest symptoms. CT is non-invasive, fast, painless and accurate diagnostic methods. Especially chest CT helps to identify the small nodules in the early stage.

Tao Xu et al., (2011) discussed the existence of a cavity in the upper lung zones which is the most important symptom of highly infectious tuberculosis (TB). The author has proposed a fully automatic TB cavity detection method. The method performs cavity segmentation and

*[1]Department of Electronics and Instrumentation, Saveetha Engineering College, Chennai, [2]Department of Electronics and Instrumentation, Sethu Institute of Technology, Madurai, India. *For Correspondence: malathiyoga08@gmail.com*

feature extraction by combining 2D Gaussian model based template (GTM) and hessian matrix based image enhancement. The method provides high accuracy and easy extraction of features, but it has lack of principled approaches for disparate features.

Previous Studies

Medical image segmentation is the imperative area in the field of research. The following studies discuss the diverse segmentation methods in abnormal lung nodule dissection for CT lung images.

Xuechen et al., (2018) proposed a new method for lung nodule detection based on the chest X-ray radiograph. The article used a solitary feature lung nodule detection method. It was a texture related features lung nodule detection method. The texture related features are extracted using stationary wavelet transform and convergence index filter. Further the extracted features are used by adaboost algorithm to produce a white nodule likeness map. Finally the SVM classifier used for the classification of lung nodules based on the extracted features. It requires less computation time with other techniques.

Muzzamil et al., (2016) the author proposed a method to segment the lung nodules from CT image. The method uses intensity thresholding to segment lung region from the thorax CT image. The threshold value is selected from the histogram of CT image for better segmentation output. Further it uses K means clustering for segmentation of potential nodules, shape and size. Next it was followed by using SVM classifier for further classification of large nodules and the sensitivity of the system was also calculated.

Alex Martins et al., (2014) developed a computer aided detection system to automate the segmentation process. The established automatic detection method is used to segment small lung nodules using image processing and pattern recognition techniques. Further, it uses region growing techniques to segment the pulmonary paraenchyma. The internal structure of the lungs is obtained by Gaussian mixture and hessian matrix. The texture is obtained by tsallis's and Shannon's entropy. Support vector machine is one of the neural networks which is used to classify the affected region as normal nodules or non-nodules.

Brahim et al., (2018) the author proposed lung CT image segmentation using deep neural networks. U- Net architecture in deep learning is used in order to perform segmentation. The method provides correct segmentation result for the lung nodule with the dice co-efficient of 0.9502

Suren et al., (2017) the author proposed the segmentation of lung CT images. It uses watershed segmentation to extract the lung cancer portion from the CT images. After completing the segmentation of lung cancer nodule, the different features like area, perimeter, mean, intensity, eccentricity, and centroid are extracted from the lung nodule. The features are trained by support vector machine to classify the nodules as benign or malignant.

Hao et al., (2015) implemented a computer aided detection method in order to assist the radiologist in early identification of lung cancer nodule from lung CT images. The lung nodule from the chest volume is extracted using the hierarchical vector quantization scheme. Comparing with other conventional method the proposed method provides accurate lung segmentation from the chest volume. After the segmentation process SVM classifier helps to classify the segmentation portion as normal or abnormal.

Hongyang et al., (2017) developed one of the computer aided detection technique to detect lung nodule from lung CT images. The proposed method uses convolutional neural network to reduce false positive of nodule detection. For the segmentation of pulmonary border, local thresholding segmentation is used.

Candemir et al., (2014) discussed digital X- ray screening system developed by the national library medicine in order to focus on early detection of tuberculosis. Automatic detection of the lung regions is difficult for computer aided diagnosis. The lung boundaries are detected using non-rigid registration driven, robust lung segmentation using image retrieval. The method has to be performed in the following three stages. The first step is identifying the training images using content based image retrieval. The second step is using a SITFT - flow model create the initial patient specific anatomical model. Third step performs the extraction of lung boundaries using graph cut.

Anita et al., (2012) discussed lung cancer as the most affecting disease. Early detection of lung cancer helps to improve the survival rate. For detection of diseases CT imaging methods are used. The author used MATLAB and image processing procedures like preprocessing, segmentation and feature extraction in order to classify lung nodules.

Hence, from this review, it is clearly recognized that Image processing tools helps for automatic segmentation of lung nodules accurately. MATLAB supports to implement the segmentation algorithm and classify the lung nodule into normal or abnormal.

Materials and Methods

The proposed work uses Otsu's thresholding and active contour based segmentation for separating lung nodule from CT chest image. Further the segmented image is classified using SVM classifier and the performance of the both segmentation method is compared in terms of PSNR, time, iterations, Mean Square error. The block diagram of the proposed work is given in the following Figure 1.

Image Pre-Processing

For computer aided segmentation preprocessing is one of the most important and challenging tasks, because of the fact that the same tissue type occurs in different scale of intensities. This is mainly due to atmospheric noise and sensor regularities. Hence, for accurate segmentation, image preprocessing is basic necessary step for both MRI, CT images. The proposed research work uses the CT images from SCIAS medical repository, in which more number of normal and abnormal Lung CT images are taken. The preprocessing is performed using median
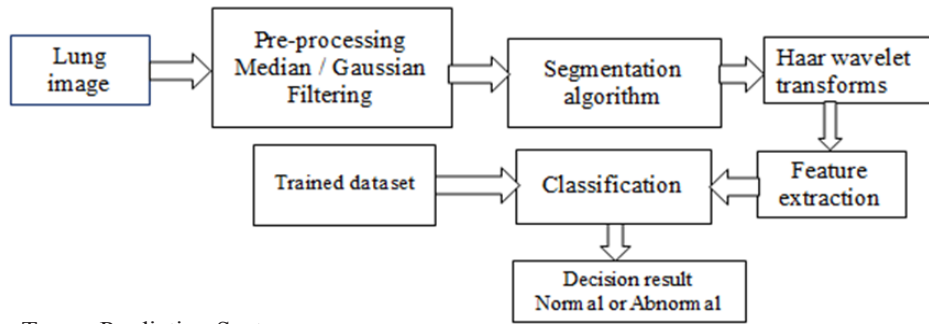
Figure 1. Lung Tumor Prediction System

filter or Gaussian filter. Most of the segmentation process uses median filter to remove noise by preserving the edges of an image. Like the median filter, Gaussian filtering also removes the noise without affecting edges. The CT normal and abnormal lung images for our research work is taken from SCIAS medical repository.

Otsu's Thresholding Based Segmentation

Yang et al., (2012) discussed Segmentation as one of the basic process in many images, video and medical applications. The gray level for the object completely differs from the gray level of the pixels in the background. Hence thresholding is one of the effective tools to distinct those background objects from the background. The gray level plays an important role to differentiate background and objects of an image. Hence the gray level is called as detectors. Thresholding is one of the simple segmentation methods to extract the region of interest. According to the gray level of an image, it is divided into two groups. Further choose the threshold value between these two groups of gray level. There are many types of thresholding based segmentation like global, local, multilevel, Otsu's thresholding based segmentation.

The intensity distribution between the objects and foreground are very dissimilar, then we use global thresholding for segmentation of an object. A single threshold value is used to differentiate object and foreground of an image. When the image has many dissimilar parts, global threshold does not provide good results. Hence we use multilevel thresholding, which contains more than one threshold value. From this analysis, it is identified that it is essential to find optimal threshold value to segment the affected region of interest from an image. Hence the Otsu's threshold method is used to segment the object from an image. The main aim of

this Otsu's method is to obtain optimum threshold value. It can be calculated by grouping of pixels into two classes C1 and C2 and it has bimodal histograms. The method also reduces the intra class variance by selecting a suitable threshold value.

The intra-class variance can be represented for each cluster using weighted equation:

$$\sigma^2 w(t) = q_1(t)\,\sigma_1^2(t) + q_2(t)\sigma_2^2(t)$$

Where the weight q_i is the probability for each class. It can be calculated as follows:

$$q_1(t) = \sum_{i=1}^{t} P_i \quad q_2(t) = \sum_{i=t+1}^{l} P_i$$

$$\mu_1(t) = \sum_{i=1}^{t} \frac{iP(i)}{q_1(t)}$$

$$\mu_2(t) = \sum_{i=t+1}^{l} \frac{i\,P(i)}{q_2(t)}$$

At last the individual class variance is given by

$$\sigma_1^2(t) = \sum_{i=1}^{t}[i - \mu_1(t)]^2 \frac{P_i}{q_1(t)}$$

$$\sigma_2^2(t) = \sum_{i=l+1}^{t}[i - \mu_2(t)]^2 \frac{P_i}{q_2(t)}$$

After calculation of variance, the process is stopped here. This algorithm is repeated for all threshold values of t within the class variance.

$$\sigma_b^2(t) = \sigma - \sigma_w^2(t) = q_1(t) - q_2(t)[\mu_1(t) - \mu_2(t)]^2$$

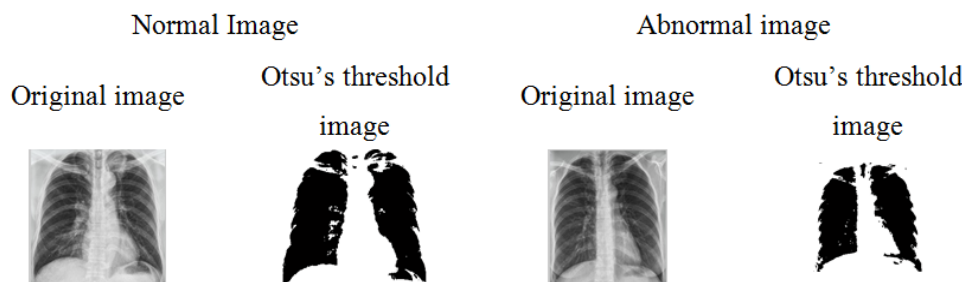The method is used to reduce the intra-class variance and maximize the between-class variance.

Figure 2. Final Segmented Nodules and Their Segmented Lungs (a) Result of Otsu's Threshold Segmentation of Normal and Abnormal Image

Table 1. Features and Parameters for Affected Images

IMAGES	Otsu Thresholding by Svm					Active Contour Method				
	Variance	MEAN	SD	PSNR	MSE	Variance	MEAN	SD	PSNR	MSE
IMAGE 01	3,309.5	131.7256	57.52	3.22	178	536	123.77	47.36	2.99	145
IMAGE 02	345	126.658	58.62	3.11	168	224	125.55	50.26	2.02	143
IMAGE 03	2,250	132.759	54.89	3.45	163	4380	133.54	44.23	2.01	135
IMAGE 04	36,200	129.68	62.59	4.89	158	34,800	123.86	47.23	2.97	148
IMAGE 05	36,200	129.68	62.59	2.99	167	64,500	123.59	60.25	2.04	156
IMAGE 06	2,650	130.25	52.69	2.78	190	5,350	126.75	49.54	2.99	161
IMAGE 07	346	131.26	50.8	3.9	189	4,230	130.62	47.26	3.01	156
IMAGE 08	3,285	121.6256	45.52	3.44	182	4,530	120.26	42.36	3.02	145
IMAGE 09	255	123.548	56.62	3.89	178	323	122.65	53.48	2.97	135
IMAGE 10	3,350	112.761	46.89	4.01	158	2,320	111.56	43.36	2.87	149

After performing otsu's threshold segmentation the output is as shown in Figure 2.

Feature Extraction

Transforming of the input data into a set of features is defined as feature extraction. It can be calculated by arithmetical measurement and numerical calculations, and it helps to differentiate the region of interest from the background. Edges play an important role for segmentation, because it helps to detect the boundaries of the lung CT image. The statistical features are based on the gray scale intensities of an image. The features are namely mean, median, standard deviation. The features are extracted by using Haar wavelet transform in our proposed work.

Mean: The ratio of the sum of all pixel value of images in a total number of pixels in an image is called as mean

$$\mu = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}P(i,j)$$

P (i, j)is the pixel value at that point (i, j).

Standard deviation: The probability distribution of a certain object is defined by standard deviation. It also gives the measures of inhomogeneity. The high value of standard deviation refers the image has high intensity and high contrast edges of an image.

$$\Sigma = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(P(I,j)-\mu)^2}$$

Variance: Variance is the square root of median

$$var = \sqrt{SD}$$

SD-standard deviation

Mean square error: It is the measure of signal fidelity or image fidelity. It states the amount of resemblance between two images and it provides the numerical score.

$$MSE = \frac{1}{MN}\sum_{i=0}^{m-1}\sum_{j-0}^{n-1}[I(i,j)-k(i,j)]^2$$

PSNR: It is the ratio used to evaluate the amount of reconstruction of processed image.

$$PSNR = 10\log_{10}\frac{MAX^2}{MSE}$$
$$= 20\log_{10}\frac{MAX}{\sqrt{MSE}}$$
$$= 20\log\_10[(MAX)-10\log_{10}[MSE]$$

The value varies from 0 to ∞. The PSNR measured in dB. If the PSNR value is 100%, then the processed image is reconstructed accurately.

Table 2. Features and Parameters for Normal Images

IMAGES	Otsu's Thresholding Along Svm					Active Contour Method Along With Svm				
	Variance	MEAN	SD	PSNR	MSE	Variance	MEAN	SD	PSNR	MSE
IMAGE 01	3,309.5	31.7256	57.52	123	1.08	536	123.77	47.36	90	0.11
IMAGE 02	345	126.658	58.62	128	1.98	224	125.55	50.26	87	0.67
IMAGE 03	2,250	132.759	54.89	124	1.01	4,380	133.54	44.23	67	1.01
IMAGE 04	36,200	129.68	62.59	125	1.78	34,800	123.86	47.23	45	0.99
IMAGE 05	36,200	129.68	62.59	127	2.01	64,500	123.59	60.25	34	0.04
IMAGE 06	2,650	130.25	52.69	125	1.64	5,350	126.75	49.54	89	0.9
IMAGE 07	346	131.26	50.8	129	2.02	4,230	130.62	47.26	92	0.1
IMAGE 08	3,285	21.6256	45.52	128	1.78	4,530	120.26	42.36	39	0.2
IMAGE 09	255	123.548	56.62	130	1.56	323	122.65	53.48	89	0.3
IMAGE 10	3,350	112.761	46.89	132	1.99	2,320	111.56	43.36	67	0.4

Figure 3. Various Haar Wavelet Decomposition Levels for Normal and Affected CT Lung images

SVM Classifier

An SVM classifier was introduced by vapnik. Support vector machine is one of the discriminative classifier and it is previously called as separating hyper plane. SVM is one of the supervised learning. It is a useful tool for data classification, function approximation, image recognition. Lee et al., (2009) studied the performance of Bayesian classifier artificial neural Network and SVM. Among all the methods SVM provides the best result for classification of an image. Many types of linear and nonlinear architectures were developed such as linear discriminant analysis, multi-layer perceptron, K nearest neighbor, least square minimum distance(LMSD) to classify an abnormal portion of an image. Among all those nonlinear classifier, SVM provides the best classification result. SVM classification involves two basic processes, namely training and testing.

Training of SVM classifier is performed by feeding of known data to SVM, along with the known decision value, hence it provides a finite training set. It helps SVM classifier to classify the unknown data accurately.

For example, consider a pattern classifier, for which hyper plane helps to separate two classes of patterns based on given example $\{(x_i, y_i)\}$ Where i= 1 to n

Where I is the vector in the input space $I=R^k$

Y(i) represents the class index which takes the value either 1 or 0.

For this machine learning, SVM classifier using class boundary, the hyper plane that maximizes the margin in the given training data. Support vectors are said to be the training samples along the hyper plane near the class boundary. The margin is the distance between the support vectors and the class boundary hyper planes. SVM helps to find an optimal separating hyper plane which gives a maximum margin between two data's. A transform which non-linearly maps the data into high dimensional space, which is not possible to perform linear separation in the original space.
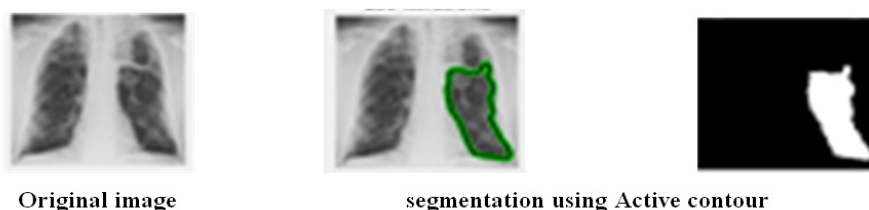

Figure 4. Results of the Proposed Method (a) Original Image (b) Active Contour Result

Consider a training set

D={ (xj,yi)}

$f(x,(w,b))=sign(wx+b)$

The output is represented by $y_i \in \{-1,+1\}$

Initially, every inputs is mapped into higher dimension feature space F by $Z=\Phi(x)$ via nonlinear mapping

$\Phi: R^n \rightarrow F$

The hyper plane that optimally separate the data on F

Minimize $\|w\|^2 + c\sum_{i=1}^{m} \in_i$

Subject to $y_i(W^1 Z_i+B) \geq -\in \in \geq 0 \forall I$

C represents regularization parameters

It helps to calculate the trade off between maximum margin and minimum classification error.

There are three types of kernel function available, namely linear kernel function, polynomial kernel function and radial basis function

Active Contour

Active contour is introduced by Kass et al in 2D space and it can be improved for 3- D by Terzopoulos et al. The method is based on moving deformable contours under the influence of forces, which helps to track boundaries and motions accurately. Based on the way in which the curves are represented there are two types of deformable models. Selin et al., (2017) discussed parametric active contour the curves are represented clearly during deformation. But in level set contours the curves are defined in implicit level. Deformable model provides the information which is the combination of geometry, physics and approximation theory. The shape of the object is obtained from geometrical information, physics represents in which way the shape varies over space and time. With the help of approximation theory to make the parameter suitable for the curve.

Step 1: Locate the Active contour or snake near to the region of interest.

Step 2: Due to the internal and external forces created within the image the snake is moved near to the target by an iterative process.

Step 3: The energy function is estimated for internal and external forces, which is capable of finding the accuracy of Region of interest.

Step 4: The main motive of this method is to minimize the energy function. The internal forces help to smooth the data set and external forces push the contours near to the region of interest.

The various features like variance, mean, standard deviation, PSNR, Mean square error for normal and abnormal lung CT images are extracted and the same has been listed in the has been listed in the tables 1 and 2.

The proposed method utilizes Otsu's and active contour segmentation for separating affected region from CT images. Further the features are extracted using Haar wavelet transform. Next the segmented image is classified using SVM classifier.

Performance Analysis

The performance of the proposed classifier is measured using following measures.

Accuracy:

It is a measure to evaluate the overall usefulness of the classifier. Accuracy can be model can be defined by following expression

Accuracy= TP+TN / TP+FP+TN+FN

Sensitivity:

The ability of the classifier to identify the positive class patterns. It is defined by the following equation

Sensitivity= TP/TP+TN

Specificity:

It helps to spot the negative class patterns. It can be estimated by using the following equation

Specificity= TN/TN+FP

Discussion

The performance analysis of proposed method is measured with the help of parameters like PSNR, MSE. Peak signal to Noise ratio is one of the important parameter which helps to measure the image quality between the input and output image. Similarly Mean square error is another parameter, which helps to measure cumulative squared error between the input and output image. Hence the PSNR, MSE value is calculated for the Otsu's and active contour based segmentation along with the SVM classifier. The MSE, PSNR value for active contour based segmentation along with SVM classifier provides the best result for complex multimodal CT lung image.

Active contour is one of the best image segmentation methods. Because it has the ability to generate sub regions with continuous boundaries. Level set theory helps to give more flexibility in the implementation of active contours. Hence the proposed active contour segmentation based SVM classifier provides good result to extract the boundaries for complex multimodal medical images exactly. It can be substantiated by the relating the value of PSNR (Peak signal to Noise ratio) and MSE (Mean Square Error).

In conclusion, Pleural Diseases persists as a great public health problem. Increase in the overall age of the population and the rise of drug-resistant have reinforced the need for rapid diagnostic improvements and new modalities to detect it and to be drug-resistant, as well as to improve disease control. An automatic method is presented to detect abnormalities in frontal chest radiographs which are aggregated into an overall abnormality score. The method is aimed at finding abnormal signs of a diffuse textural nature such as parameters encountered in mass chest screening against pleural diseases. The scheme starts with filtering the chest radiographs using median filter then the features of the filtered image will be extracted using Haar wavelet transform (HWT) and histogram of oriented gradients then the automatic segmentation of the lung fields will be done by using Otsu's thresholding algorithm and Active Contour Method. Finally the extracted features are compared with the trained datasets using support vector machine and the result will be shown whether then tuberculosis is affected or not. All the above process were completed and lung images were successfully classified the lung image is as normal or abnormal by using support vector machine.

References

Alex MS, Antonio Ode CF, Aristófanes CS, et al (2014). Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM. *Eng Appl Artif Intell*, **36**, 27–39.

Anita C, Sonit SS (2012). In proceedings of International conference on computing sciences' December 24, 'Lung Cancer Detection on CT Images by Using Image Processing', Phagwara, India.

Brahim AS, Abdelhamid ELH, Aicha M (2018). In proceedings of First International Conference On Intelligent Computing in Data Sciences', 'Lung CT Image Segmentation Using Deep Neural Networks' Morocco.

Candemir S, Jaeger S, Palaniappan, K, et al (2014). Lung segmentation in chest radiographs using anatomical atlases with non-rigid registration. *IEEE Trans Med Imaging*, **33**, 577-90.

Hao H, Lihong L, Fangfang H, et al (2015). Fast and adaptive detection of pulmonary nodules in thoracic CT images using a hierarchical vector quantization scheme. *IEEE J Biomed Health Inform*, **19**, 648-59.

Hongyang J HM, Wei Q, Mengdi G, Yan L (2017). An automatic detection system of lung nodule based on multi-group patch-based deep learning network. *IEEE J Biomed Health Inform*, **20**, 2168-94

Lee Y, Seo JB, Lee JG, et al (2009). Performance testing of several classifiers for differentiating obstructive lung diseases based on texture analysis at high-resolution computerized tomography (HRCT). *Comput Methods Programs Biomed*, **93**, 206-15.

Muzzamil J, Moazzam J, Muhammad ZUR, Syed I, Ali S (2016). A novel approach to CAD system for the detection of lung nodules in CT images. *Comput Methods Programs Biomed*, **12**, 125–39.

Selin U, Buse U (2017). In proceedings of 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, August 24-25, 'Lung tumor segmentation algorithm', Budapest, Hungary.

Suren M, Prasad PWC, Abeer A, Singhb AK, Elchouemi AK (2017). In proceedings of 6th International Conference on Smart Computing and Communications', December 7-8, 'Lung Cancer Detection using CT Scan Images', Kurukshetra, India.

Tao Xu, Irene C, Mrinal M (2011). In proceedings of International conference on engineering in medicine and biology', august 30- sepetember 3, 'Automated cavity detection of infectious pulmonary tuberculosis in chest radiographs' Boston, MA, USA.

Xuechen L , Linlin S, Suhuai L (2018) A solitary feature-based lung nodule detection approach for chest X-ray radiographs. *IEEE J Biomed Health Inform*, **22**, 516-24.

Yang S, Weidong C, Jinman K, David DF (2012). A multistage discriminative model for tumor and lymph node detection in thoracic images. *IEEE Trans Med Imaging*, **31**, 1061-75.