Asian Journal of Information Technology 15 (21): 4189-4194, 2016
ISSN: 1682-3915

Lung Cancer Detection by Automatic Region Growing with Morphological Masking and Neural Network Classifier

[1]T. Manikandan and [2]N. Bharathi
[1]Department of ECE, Rajalakshmi Engineering College, 602105 Tamil Nadu, India
[2]Department of EEE, Velammal Engineering College, 600066 Tamil Nadu, India

Abstract: An effective lung cancer detection scheme is developed in the present research for segmenting the suspected nodules from the consecutive slices of Computer Tomography (CT) images using Automatic Region Growing (ARG) with morphological masking. The seed values are chosen automatically for the region growing technique. After the initial segmentation of the suspected lung nodules by the region growing technique, features like eccentricity, area and convex area were computed to eliminate unwanted line like structures and small tiny non-cancerous nodules, (>3 mm in size). Further, the centroid shift for the remaining suspected nodules in consecutive slices were calculated to reduce the false positives (vessels), as the centroid shift do not vary much for cancerous nodules and calcifications (calcium deposition on the lungs). Finally, the texture features like homogeneity, auto-correlation and contrast were calculated to remove the calcifications. The extracted features like centroid shift, homogeneity, auto-correlation and contrast were utilized to train and test the neural network classifier. The study was carried out on a total of 106 patients CT scan images (56 cancerous patients and 50 non-cancerous subjects) retrospectively collected from the Bharat Scans, Chennai, which had 60 malignant (cancerous) nodules and 238 benign (non-cancerous) nodules. Out of these nodules, 56 true malignant nodules and 204 true benign nodules were classified correctly by the neural network classifier and remaining was misclassified. Present work produced good sensitivity and specificity and accuracy of 93, 86 and 87%, respectively. The False Positive (FP) per patient is reported as 0.32.

Key words: Lung cancer, lung nodule, region growing, morphological masking, neural network

INTRODUCTION

Lung cancer is foremost cause of cancer-related deaths in world-wide (Siegel *et al*., 2012). Diagnosis of lung cancer in its early stage is the only way to control the death rate (Lo *et al*., 1995). However, confirming the disease in its early stage is critical and physician request the patients to undergo several Computed Tomography (CT) images at regular intervals (Kumar and Ganesh, 2013). Though the CT imaging accurately captures the lung images, physician still finds it difficult to diagnose the cancerous nodules. This is due to the continuous cross-sectional images produced by the CT machine and necessary to analyze the every cross-section. This leads for additional effort put by the radiologist to diagnosing the disease and therefore high possibility of making mistake. The advancement of Computer-Aided Diagnostic (CAD) system may assist the radiologist and physician to precisely analyze the CT images to potentially improving the cancerous nodule detection (Camarlinghi *et al*., 2012; Messay *et al*., 2010).

Generally, nodules are the indicator of lung cancer and are tiny mass inside the lung. Lung nodule are categorized into juxta-pleural, well-circumscribed, pleural tail and vascularised based on their location inside the lung (Kostis *et al*., 2003). Most of the nodules are non cancerous but about 40% of the lung nodules only cancerous (Kuruvilla and Gunavathi, 2014; Sousa *et al*., 2010). Therefore, it is the real challenge for the researchers to quantify and to describe the lung nodules. Hence efficient technique is required to segment the lung nodules from the CT scan images.

Commonly lung nodule detection and classification in CT images involves four important steps: lung region segmentation, nodule candidate detection, false positive reduction and classification. To extract the lung region from the CT scan images several methods were found in the literatures. The methods such as multiple thresholding, optimal thresholding and global thresholding were successfully implemented for lung region segmentation (Suzuki *et al*., 2003; Ye *et al*., 2009; Suarez-Cuenca *et al*., 2009). However, in these studies the threshold values are calculated manually by considering the pixel intensity of the CT scan images. Also morphological processing was performed on CT images to remove fat, bone and background noise of lung

Corresponding Author: T. Manikandan, Department of ECE, Rajalakshmi Engineering College, 602105 Tamil Nadu, India

parenchyma in threshold based segmentation. Since the threshold value is chose based on the CT technology and X-ray dose, it vary from one CT machine to another CT machine. Hence, this technique will not segment the lung region universally.

After the segmentation of the lung volume, the nodule candidates are detected for finding out the cancerous nodules. Various methods were successfully implemented in the literatures for the candidate nodule detection. It includes filtering based methods, active contour methods, shape-based methods and morphological approaches (Kass *et al*., 1988; Li *et al*., 2008; Pu *et al*., 2011; Paik *et al*. 2004; Kubota *et al*., 2011). The limitations of these methods were detection of more false positive nodules. The various thresholding techniques were proposed for the nodule candidate detection, however all the methods are based on the intensity values (Armato *et al*., 2001; Zhao *et al*., 2003; Leader *et al*., 2003). These methods needs human intervention to select the threshold value as it is not universal for all the CT machines. Template based matching were found in literatures to segment the nodule candidates detection (Jo *et al*., 2014; Lee *et al*., 2001). As the orientation of the image changes from one CT image to another CT image template matching segmentation fails. Hence this method is not reliable. The work published by Ozekes *et al*. (2008) using genetic and fuzzy based model produced 100% sensitivity with false positive per patient of 13.4. Complex hybrid model implemented by Lu *et al*. (2015) for nodule detection yielded 3.9 false positives/patient.

In general, all the above discussed two Dimensional (2-D) methods showed more false positives. To overcome the more false positives in 2-D studies, three dimensional (3-D) algorithms were proposed. Texture 3-D model for the lung nodule detection implemented by Demir and Camurcu (2015) produced 2.45 false positive/patient. Alilou *et al*. (2014) demonstrated 3-D structural visualization model yielded false positive of 3.13/patient. An automated pulmonary nodule detection scheme using hierarchical 3-D block classification proposed by Wook-Jin and Tae-Sun produced 6.76 false positives/patient (Dehmeshki *et al*., 2007). Later, they proposed an automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor produced 2.27 false positives/patient (Choi and Choi, 2012). Still there is a scope for the reduction of false positive nodules.

The extracted features from the segmented nodule candidate were given to the input of classifiers for lung nodules classification. The classifier classifies them as either malignant nodule or benign nodule based on the extracted features. Most of the papers in literatures used geometric features, gray level features and gradient features for classification of the nodule candidate (Choi and Choi, 2012, 2013). The linear discriminant analysis classifier and rule-based classifier have been most widely used in the literatures (Messay *et al*., 2010). Machine learning classification methods like genetic algorithm, support vector machine and neural network were also found in the literatures (Dehmeshki *et al*., 2007; Golosio *et al*, 2009; Riccardi *et al*., 2011). The available CAD systems in the literates have its own strengths and weakness for lung nodule classification.

In this study, we proposed an automatic region growing and morphological masking for segmenting the lung regions and initial suspected lung nodules. The structural (shape/size) features like area, eccentricity and convex area were computed for the initial suspected nodules to eliminate tiny clusters and line like structures and retain the final suspected nodules. After the segmentation of final suspected nodules, the 3-D centroid shift was carried out on the consecutive CT scan images to remove the vessels. Finally, texture features like homogeneity, auto-correlation and contrast are computed to eliminate the calcifications. Thus, the present work reduced the false positives to greater extent. This step makes our work more efficient than the available existing works in this area.

MATERIALS AND METHODS

Database: A total of 106 patients (56 malignant cases and 50 non-cancerous cases) CT scan (General Electric, New York, USA) images are retrospectively collected from the Bharat Scans (a wing of Bharat Education and Research Foundation), affiliated to Tamil Nadu Dr. M.G.R Medical University, Chennai, India. The mean age of patient is considered as 62.8 year (youngest patient is 37 years and oldest patient is 81 year). The low dose CT images are taken at a kilo voltage peak of 120 kV and current rating of 356 mA. The approved document was signed by each patient as a means of voluntary written consent.

The institutional ethical committee of Bharat Scans approved the practice adopted for this study. Two CT chest radiologists were allotted by the department of radiology of Bharat Scans to select 106 patient's cases which had a total of 60 malignant nodules and 238 malignant nodules from their data base of size >3 mm (which are used as ground truth images). Of these 106 subjects (Mean±SD age = 62.92±5.7 year), 78 were males (Mean±SD age = 65.28±4.4 year) and 28 were females (Mean±SD age = 58.24±7.1 year), for whom the analysis was carried out. Segmentation of final suspected nodules from the lungs by ARG and morphological masking The detailed steps involved in Automatic Region Growing (ARG) with morphological masking for segmentation of lung region and suspected nodules are as follows:

Step 1: Read the input CT scan image.

Step 2: Choose the first pixel (background pixel) in the chest CT lung scan image automatically as initial seed point for region grow technique which creates the mask for lung region segmentation. This step segment the background region from the rib and lung lobe region. This seed point is common for the CT scan images taken from any CT machine.

Step 3: All the pixels which has the value '0' in the mask are replaced by the input CT image pixels and all the pixels which has the value '1' are replaced by '0'. Thus the lung region is segmented from the input CT image.

Step 4: Choose the first non-zero pixel from an image segmented in step 3 as the next seed point (corresponds to a border of the lung region) to segment the lung lobe portion from ribs and fat tissues. This seed point is also common for the CT scan images taken from any CT machine.

Step 5: All the pixels which has the value '0' in the mask are replaced by the input CT image pixels and all the pixels which has the value '1' are replaced by '0'. Thus the lung parenchymal region is segmented from the input CT image.

Step 6: Gray threshold is computed on the resultant image by Otsu's method which segments the initial suspected binary lung nodules from the lung parenchymal region as shown in Fig. 1a-d. The initial segmented nodules may be of malignant or benign.

Step 7: Most of the lines like irregular structures, small tiny non cancerous clusters have area <10 pixels in size, eccentricity >1 and convex area <220. Therefore apply the area, eccentricity and convex area features to keep only the final suspected nodules.

Step 8: Convert the final suspected binary nodule into the gray nodules. These final suspected gray nodules may be of malignant or benign (vessels and calcifications).

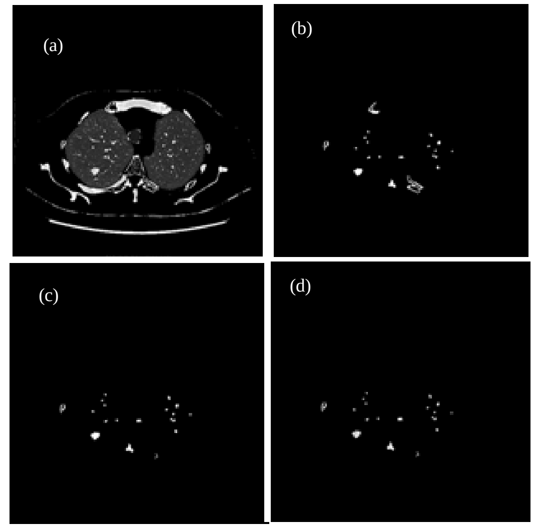

Fig. 1: Steps involved in segmentation of the suspected lung nodules; a) Segmented lung parenchymal region; b) Segmented initial suspected lung nodules before applying area and eccentricity criteria; c) Segmented final suspected lung nodules; d) Segmented final suspected gray nodules

Centroid analysis and texture features for the elimination of vessels and calcifications: The final suspected nodules after applying area, eccentricity and convex area may be of vessels, calcifications and cancerous. Therefore the vessels and calcifications are to be removed which are not nodules. The centroid shift in the successive slices remains unchanged for the cancerous nodules and shapes do vary. This peculiar property helps the physicians to identify the malignant nodules in clinical practice. Hence change in centroid position is computed to avoid the vessels. After the removal of vessels the remaining suspected nodules may be of calcification or malignant nodules.

Calcifications will have uniform auto-correlation, homogeneity properties and high contrast value. However, the real malignant nodules will have non-uniform auto-correlation, homogeneity properties and low contrast values when compared to the calcifications. Hence auto-correlation, homogeneity and contrast values are extracted for the final suspected nodules to detect the true malignant nodules. The extracted features were utilized to train and test the neural network classifier to make the final decision.

Neural network classifier: Neural network consists of artificial neurons which replaces the functions of neurons in the human brain. It consists of three layers. They are input layer, hidden layer and output layer. The extracted lung nodule features (centroid shift, homogeneity, auto-correlation and contrast) were fed to the input layer and then it is propagated to output layer through the hidden layer.

The feed forward-back propagation algorithm was employed in the neural network classifier. During the training process, the weights between input layer and hidden layer and hidden layer and output layer were adjusted to minimize the mean square error between neural network output value and target value. Once the training has got over, it will have the capability of making the decision about the lung nodule as either malignant or benign during the testing phase.

RESULTS AND DISCUSSION

Input CT scan slice of a cancerous patient is shown in Fig. 2a. The mask for lung region segmentation after choosing the first pixel from the C T scan image as the seed point is shown in Fig. 2b. The segmented lung region after applying the mask is shown in Fig. 2c. Mask for the segmented lung lobe portion is given in Fig. 2d by choosing the first non-zero pixel from the Fig. 2c. Fig.1a shows the segmented lung parenchymal region by replacing all the pixels which has the value '0' in the mask by the input CT image pixels and all the pixels which has the value '1' are replaced by '0'. The initial suspected lung nodules after applying the Otusu's threshold are shown in Fig. 1b. For the image shown in Fig. 2a, the total initial suspected nodules after applying the region growing with morphological masking are computed as 560.

The entire segmented initial suspected nodules from the CT slices are not malignant. It may be of tiny clusters, line like structures, vessels and calcifications also. Hence further analysis carried out to find the final suspected nodules. After applying the area, eccentricity and convex area features (which removes tiny clusters and line like structures) the initial suspected nodules are greatly reduced to 19 which is shown in Fig. 1c. The gray suspected final nodules are shown in Fig. 1d. All these final suspected nodules are still not cancerous. It may be of vessels and calcifications. Therefore, further analysis must be carried out to detect the true malignant nodules.

The extracted centroid shift from the consecutive CT slices, homogeneity, contrast and auto-correlation features were utilized to train and test the neural network classifier. A total of 106 patient cases obtained from Bharat scan centre, Royapettah, Chennai are analyzed the occurrence of malignant and benign nodules (used as ground truth nodules). Two chest radiologists allotted by the radiological department of the scan centre showed 60 true malignant nodules and 238 benign nodules. Out of these, 60% of the total nodules (36 malignant nodules and 143 benign nodules) are used for training and remaining 40% of the nodules (24 malignant nodules and 95 benign nodules) for testing.

The proposed algorithm using region growing with morphological masking and neural network classifier detected 56 true malignant nodules and 204 true benign nodules. Therefore, True Positive (TP), False Negative (FN), True Negative (TN) and False Positives (FP) are 56, 4, 204 and 34 respectively. The sensitivity, specificity and accuracy of this work are computed as 93, 86 and 87% respectively. The FP per patient is reported as 0.32. The comparative analysis of the proposed work with the existing works for the lung nodule classification is given in Table 1.

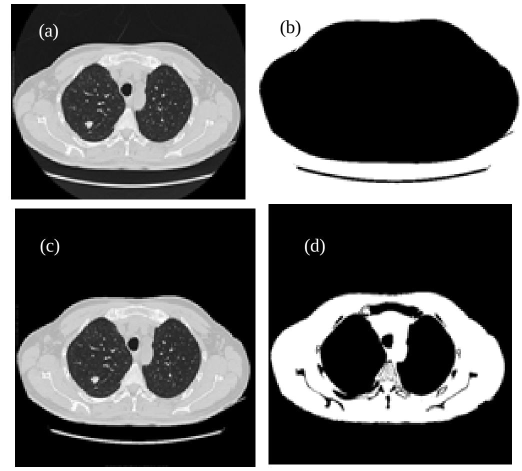

Fig. 2: Steps involved in seed selection for lung lobe segmentation; a) Input CT image; b) Mask for background region segmentation; c) Segmented lung region; d) Mask for lung lobe segmentation

Table 1: Comparative analysis of the proposed work with existing works

CAD system	Sensitivity (%)	False positive/ patient	Nodule size criteria used (mm)
Zhao *et al.* (2003)	84	5.00	2-7
Opfer and Wiemeker (2007)	74	4.00	>4
Dehmeshki *et al.* (2007)	90	15.50	3-20
Ozekes *et al.* (2008)	100	13.40	>5
Golosio *et al.* (2009)	79	4.00	>4
Suarez-Cuenca *et al.* (2009)	80	7.70	>4
Messay *et al.* (2010)	83	3.00	3-30
Sousa *et al.* (2010)	85	0.42	> 3
Choi and Choi (2013)	98	6.76	3-30
Camarlinghi *et al.* (2012)	80	3.00	>3
Choi and Choi (2012)	96	2.27	>3
Alilou *et al.* (2014)	80	3.90	>4
Kuruvilla and Gunavathi (2014)	82	0.50	> 3
Lu *et al.* (2015)	85	3.13	>5
Demir and and Camurcu (2015)	98	2.47	NA
Proposed frame work	93	0.32	>3

CONCLUSION

Lung cancer is the very dangerous among all other types of cancers in the world. It needs to be detected in its earlier stages to increase the survival rate of the patients. The prediction of lung cancer is tedious work for the radiologists and physicians as they need to analysis all the slices of CT scan images. This may lead to the large human error. To overcome this error, physicians and radiologists in the hospitals may use CAD system as the second opinion for detecting the lung cancer. The implemented CAD system using the proposed algorithm for the lung cancer detection showed good sensitivity and selectivity of 93 and 87%, respectively. The accuracy of this work is calculated as 87% with greatly reduced FP per patient (0.32 FP/Patient). Thus, the proposed CAD system using automatic region growing with morphological masking and neural network classifier could be used as a second opinion for the physicians and radiologists to make the final decision about the lung cancer.

ACKNOWLEDGEMENTS

This research was conducted at Bharat Scans, Royapettah, Chennai. The Institutional ethical committee of the Bharat Education and Research Foundation approved the protocol used for this study (Ref: IEC-BERF/Approval Lr./Date: 4-6-2014). The authors would like to thank the authorities of Bharat Scans for providing necessary facilitative infrastructure to complete this research.

REFERENCES

Alilou, M., V. Kovalev, E. Snezhko and V. Taimouri, 2014. A comprehensive framework for automatic detection of pulmonary nodules in lung CT images. Image Anal. Stereology, 33: 13-27.

Armato, S.G., M.L. Giger and H. MacMahon, 2001. Automated detection of lung nodules in CT scans: Preliminary results. Med. Phys., 28: 1552-1561.

Camarlinghi, N., I. Gori, A. Retico, R. Bellotti and P. Bosco *et al*., 2012. Combination of computer-aided detection algorithms for automatic lung nodule identification. Int. J. Comput. Assisted Radiol. Surgery, 7: 455-464.

Choi, W.J. and T.S. Choi, 2012. Genetic programming-based feature transform and classification for theautomatic detection of pulmonary nodules on computed tomography images. Inform. Sci., 212: 57-78.

Choi, W.J. and T.S. Choi, 2013. Automated pulmonary nodule detection system in computed tomography images: A hierarchical block classification approach. Entropy, 15: 507-523.

Dehmeshki, J., X. Ye, X. Lin, M. Valdivieso and H. Amin, 2007. Automated detection of lung nodules in CT images using shape-based genetic algorithm. Comput. Med. Imaging Graoh., 31: 408-417.

Demir, O. and Y.A. Camurcu, 2015. Computer-aided detection of lung nodules using outer surface features. Bio. Med. Mater. Eng., 26: S1213-S1222.

Golosio, B., G.L. Masala, A. Piccioli, P. Oliva and M. Carpinelli et al., 2009. A novel multithreshold method for nodule detection in lung CT. Med. Phys., 36: 3607-3618.

Jo, H.H., H. Hong and J.M. Goo, 2014. Pulmonary nodule registration in serial CT scans using global rib matching and nodule template matching. Comput. Biol. Med., 45: 87-97.

Kass, M., A. Witkin and D. Terzopoulos, 1988. Snakes: Active contour models. Int. J. Comput. Vision, 1: 321-331.

Kostis, W.J., A.P. Reeves, D.F. Yankelevitz and C.I. Henschke, 2003. Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. IEEE Trans. Med. Imaging, 22: 1259-1274.

Kubota, T., A.K. Jerebko, M. Dewan, M. Salganicoff and A. Krishnan, 2011. Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models. Med. Image Anal., 15: 133-154.

Kumar, T.S. and E.N. Ganesh, 2013. Proposed technique for accurate detection-segmentation of lung nodules using spline wavelet techniques. Int. J. Biomed. Sci, 9: 9-17.

Kuruvilla, J. and K. Gunavathi, 2014. Lung cancer classification using neural networks for CT images. Comput. Methods Programs Biomed., 1131: 202-209.

Leader, J.K., B. Zheng, R.M. Rogers, F.C. Sciurba and A. Perez *et al*., 2003. Automated lung segmentation in X-ray computed tomography: Development and evaluation of a heuristic threshold-based scheme1. Acad. Radiol., 10: 1224-1236.

Lee, Y., T. Hara, H. Fujita, S. Itoh and T. Ishigaki, 2001. Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. IEEE Trans. Med. Imaging, 20: 595-604.

Li, Q., F. Li and K. Doi, 2008. Computerized detection of lung nodules in thin-section CT images by use of selective enhancement filters and an automated rule-based classifier. Acad. Radiol., 15: 165-175.

Lo, S.C., S.L. Lou, J.S. Lin, M.T. Freedman and M.V. Chien *et al*., 1995. Artificial convolution neural network techniques and applications for lung nodule detection. IEEE Transa. Med. Imaging, 14: 711-718.

Lu, L., Y. Tan, L.H. Schwartz and B. Zhao, 2015. Hybrid detection of lung nodules on CT scan images. Med. Phys., 42: 5042-5054.

Messay, T., R.C. Hardie and S.K. Rogers, 2010. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Med. Image Anal., 14: 390-406.

Opfer, R. and R. Wiemker, 2007. Performance analysis for computer-aided lung nodule detection on LIDC data. Proceeding of the International Conference on Medical Imaging 2007: Image Perception, Observer Performance and Technology Assessment, March 08-08, 2007, SPIE, San Diego, California, pp: 65151C-65151C.

Ozekes, S., O. Osman and O.N. Ucan, 2008. Nodule detection in a lung region that's segmented with using genetic cellular neural networks and 3D template matching with fuzzy rule based thresholding. Korean J. Radiol., 9: 1-9.

Paik, D.S., C.F. Beaulieu, G.D. Rubin, B. Acar and R.B. Jeffrey et al., 2004. Surface normal overlap: A computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT. IEEE Trans. Med. Imaging, 23: 661-675.

Pu, J., D.S. Paik, X. Meng, J. Roos and G.D. Rubin, 2011. Shape break-and-repair strategy and its application to automated medical image segmentation. IEEE Trans. Visual. Comput. Graphics, 17: 115-124.

Riccardi, A., T.S. Petkov, G. Ferri, M. Masotti and R. Campanini, 2011. Computer-aided detection of lung nodules via 3D fast radial transform, scale space representation and Zernike MIP classification. Med. Phys., 38: 1962-1971.

Siegel, R., D. Naishadham and A. Jemal, 2012. Cancer statistics, 2012. CA: Cancer J. Clinicians, 62: 10-29.

Sousa, J.R.F.D.S., A.C. Silva, A.C.D. Paiva and R.A. Nunes, 2010. Methodology for automatic detection of lung nodules in computerized tomography images. Comput. Methods Programs Biomed., 98: 1-14.Sousa, J.R.F.D.S., A.C. Silva, A.C.D. Paiva and R.A. Nunes, 2010. Methodology for automatic detection of lung nodules in computerized tomography images. Comput. Methods Programs Biomed., 98: 1-14.

Suarez-Cuenca, J.J., P.G. Tahoces, M. Souto, M.J. Lado, M. Remy-Jardin, J. Remy and J.J. Vidal, 2009. Application of the iris filter for automatic detection of pulmonary nodules on computed tomographyimages. Comput. Biol. Med., 39: 921-933.

Suzuki, K., III.S.G. Armato, F. Li, S. Sone and K. Doi, 2003. Massive training artificial neural network for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography. Med. Phys., 30: 1602-1617.

Ye, X., X. Lin, J. Dehmeshki, G. Slabaugh and G. Beddoe, 2009. Shape-based computer-aided detection of lung nodules in thoracic CT images. IEEE Trans. Biomed. Eng., 56: 1810-1820.

Zhao, B., G. Gamsu, M.S. Ginsbers, L. Jiang and L.H. Schwartz, 2003. Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm. J. Applied Clin. Med. Phys., 4: 248-260.