

Analysis of Lung Nodule and Classification based on SVM Classifier and Fuzzy C-Means Clustering for Effective Segmentation

Smriti kushwaha

Department of Electronics and
Communication Engineering
B.S.Abdur Rahman University
Chennai, India
Smritikanpur79.sk@gmail.com

P.Maya

Department of Electronics and
Communication Engineering
B.S.Abdur Rahman University
Chennai, India
maya@bsauniv.ac.in

Abstract— Lung cancer is the most type of cancer among all the cancer with less survival rate. It is very difficult to analyze the cancer at its early stage. In this paper, using computed tomography CT images. we proposed a classification method for the four types of lung nodules, i.e., well-circumscribed, vascular-ized, juxta-pleural and pleural-tail using SVM classifier and fuzzy C-means clustering analysis for segmentation and feature extraction with the help of DWT and MWT transform and LBP for the detection of cancerous cells.

Index Terms—Classification- SVM-Support vector machine, FCM- fuzzy C-means clustering analysis, DWT-discrete wavelet transform, MWT-multi wavelet transform , LBP-Local binary pattern.

I. INTRODUCTION

Lung cancer is one of the major cause of cancer deaths in humans now a day. Approximately 20% of cases with lung nodules related to lung cancer. Therefore, the recognition of the malignant lung nodules is necessary for the screening and diagnosis through CT scan or PET scan. Lung nodules are small solitary masses [5] in the human lung, and are mostly oval, round shapes. Whereas, they can be distorted through surrounding anatomical structures, as blood vessels and the adjacent pleura. Intra-parenchyma lung nodules are usually to be malignant than that connected with the surrounding structures, and therefore lung nodules are divided into four types according to their position in the lung. Recently, the classification from Diciotti *et al.* [9]:

- a) **Well-circumscribed (W)**: The nodule which are located centrally in the lung without a connection to vasculature in the lung surface.
- b) **Vascularized(V)**: Nodule which is centrally located in the lung but connected to neighboring vessels.
- c) **Juxta-pleural (J)**: A huge portion of the nodule connected to pleural surface in the lung.

- d) **Pleural-tail (P)**: Nodule which is near the pleural surface connected through a thin tail.

CT Computed tomography is the best and accurate imaging to provide anatomical information in the lung nodules with the surrounding structures. Currently clinical practice, Although, manipulation of CT images is challenging for the radiologists because of the large number of cases. The manual reading is to be error-prone and the manipulator can miss the nodules and also a cancer. Computer-aided diagnosis (CAD) systems can be helpful for radiologists for offering screening and second views to classify nodules in the lung [10].

CADs are helpful and capable of analyzing and interpret, the large number small lung nodules identified by CT scans. Therefore, positron emission therapy (PET) scan is proven a better screening than CT because with this small vasculature information also can be collected which will be strong evidence in representing the affecting lung area with cancer.

Here are some sample images are shown in Fig.1 a, b, c, d with the nodule that is seems to be encircled in red.

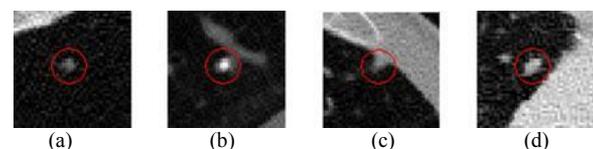


Figure 1. CT images with the four types of nodules, shown from left to right, well-circumscribed, vascularized, juxta-pleural, and pleural-tail.

online at <http://ieeexplore.ieee.org>.

A. Previous Work

In previous work many studies have reported, the detection and segmentation of lung nodules, as there is limited data in lung nodule classification. Farag *et al.* reported in some of the studies in the classification problem [3]. We suggested however, improved performance could be obtained by better feature design and a more advanced classifier. In a recent work [8], F.zhang has designed an overlapping nodule identification procedure to help the classification, but this work only focused on recognition of the nodules located in the intersection in

different types. Prior work mainly describes in patch division method and latent semantic analysis, algorithm (quick shift). A patch based approach, which is mainly based on partitioning the input image into an order less collection of the smaller patches, is usually used to make the bag of feature model. Compared with an overall description of the image, patch-based methods can captured the local details represent the heterogeneous structures

B. Outline of the proposed method

The proposed method involves with SVM classifier, fuzzy c-means clustering analysis, local binary pattern, and comparison of accuracy between discrete wavelet transform, and multi wavelet transform.

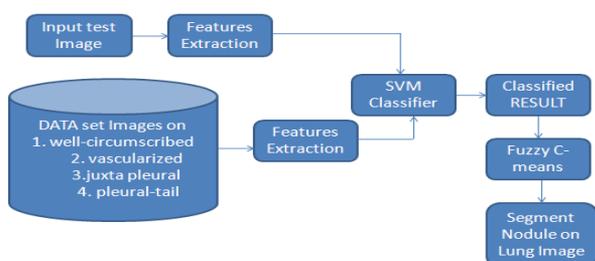


Figure 2. Block diagram of proposed method

II. DISCRETE WAVELET TRANSFORM

DWT has employed in order to keep the high frequency components of the image. Fig.3 DWT separates the image into different sub-band, namely LL, LH, HL, HH. It shown in fig. 3 a high frequency sub-band contains the edge information of the input image and low frequency sub-band contains the clear information of the image. Wavelet transform [1] has a limitation of singularity point detection.

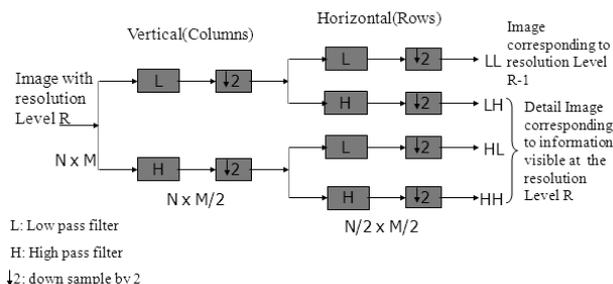


Figure 3. DWT low and high frequency division

Haar wavelet The first DWT was invented through Hungarian mathematician Alfred Haar, it is basically used for equalization of image matrix for providing the rounded value instead of floating. Haar wavelet considered to pair the input values,

storing the difference and passing sum of those values. Input represented by list of 2^n numbers. Whereas pairing the sum for getting the next scale which provide $2^n - 1$ difference and final sum. Fig.4 shows DWT with Haar wavelet.



Figure 4. DWT lung image with four sub-band LL,LH,HL,HH

III. MULTI WAVELET TRANSFORM

The multi wavelet transform [4] originates from the generalization of scalar wavelets. In MWT Instead of one scaling function or one wavelet, there are multiple scaling functions and wavelet are used shown in fig. 5. This will lead to more degree of freedom in constructing wavelet.

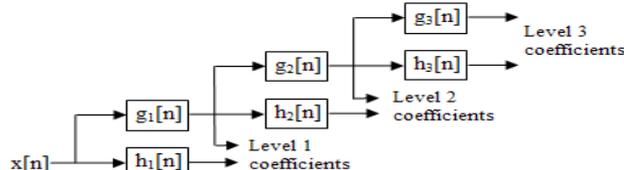


Figure 5. MWT low and high frequency division

The proposed method, multi wavelet is obtained at replacing scalars with matrices, and scalar functions with vector functions and single matrices with the block of matrices. Fig.6 showing sub-band decomposition in MWT with Haar wavelet for equalization of matrix values.



Figure 6. MWT lung image with multiple sub-band up to 3-level LH3, HL3

IV. LOCAL BINARY PATTERN

The local binary pattern (LBP) operator [6] is as a gray level invariant texture measure in a local neighbourhood representing with 0 and 1 binary values. Fig.7 shows LBP for selected portion of lung nodule. The real LBP operator labels the pixel of an image by thresholding the 3X3 neighbourhood

of each pixel and concatenating the results binomially to form a number.

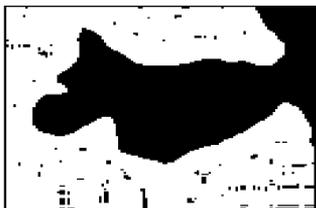


Figure 7. Local binary pattern, separating white and black region selected portion of lung nodul

V. SUPPORT VECTOR MACHINE

SVM is a classifier that maximize the *margin* around the separate hyper plane true values and false values. The classifying function is completely represented by a subset of training samples, *the support vectors*. SVMs is one of the *quadratic programming* problem in which test image feature will be classify shortest path of *similarity measurement*. SVM [2] has proved to be a highly effective classifier. However, for lung nodule image classification, SVM could be error prone due to the overlapping feature spaces of the nodules [8].

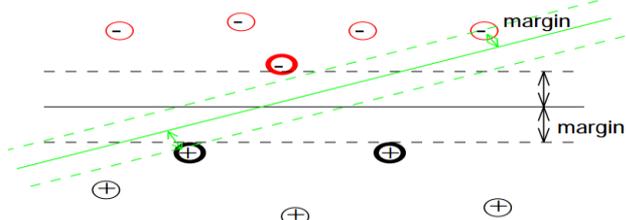


Figure 8. SVM hyper plane with minimum and maximum margin.

With the existing data set the proposed method trained the values in SVM then the algorithm applied. The target vectors match with the input test image, then stage will be classified.

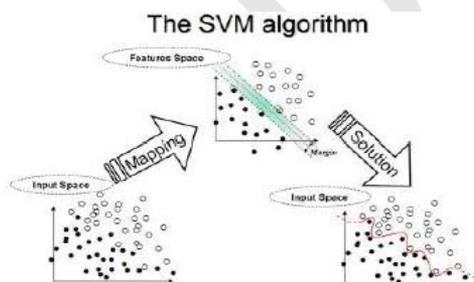


Figure 9. SVM showing true positive and false negative

VI. FUZZY C-MEANS CLUSTERING ANALYSIS

The fuzzy c-means clustering refers [11] as the partitioning an image into different segments. The aim is to simply change the representation of an image into which is something meaningful

to understand and easy to analyze The proposed method algorithm KWFCM that is kernel weighted fuzzy c-means clustering used for segmentation shown in fig.10. This algorithm is an unsupervised algorithm which classifies the input data points into multiple segments based on their inherent distance from each classes.FCM used spatial function to update membership function to get accurate region of the affected area analyzing local neighborhood information.

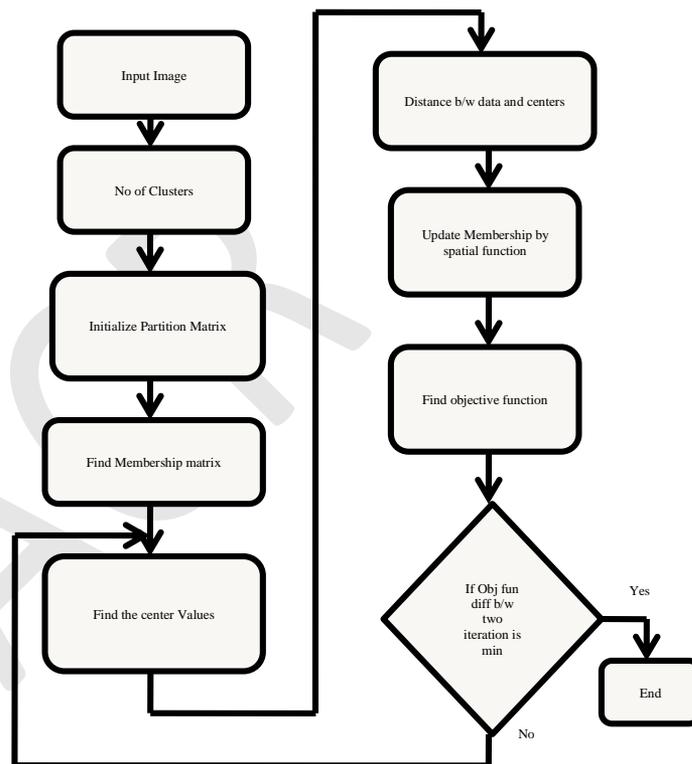


Figure 10. Fuzzy c-means algorithm flow

Morphological process is applied on partitioned or segmented image for more smoothening the lung cancer part alone. Therefore, erosion and dilation process will be using to modify the effected region by removing the unwanted pixels from outer region of cancer of cancer part. This morphological operation is performing on images based on shapes, which will give clear shape of nodule area. Erosion will remove the thickness border by adding '-1' with 0 and dilation will add the thickness border by adding '1' with 0, Therefore 1's are called neighborhood pixels.

VII. FEATURE EXTRACTION

The image feature depend upon transformation and visualization to represent the robust image condition in the affected region in Harlick features. Feature extraction here 200 values that is Energy, Entropy, Contrast, Correlation,

Homogeneity. In DWT visual perception is present whereas, in MWT clear visualization is not present but information will be surely present.

iii. Vascularized lung nodule

TABLE I. Extraction of 200 featured values using DWT

0.42273	0.34092	0.56722	0.27413	0.41315	0.39512	0.39908	0.37056	0.44041	0.36589	0.39556	0.58932	0.74428	0.32805	0.66607	0.49312	0.34943	0.60653	0.58104	0.35078
0.20663	0.62125	0.09603	1.16775	0.14216	0.08551	0.20952	0.21198	0.18898	0.21198	0.22386	0.09652	0.16991	0.26569	0.10562	0.19316	0.35087	0.33372	0.25129	0.23554
0.94826	0.93821	0.9434	0.95191	0.9549	0.93339	0.93026	0.96009	0.95529	0.95980	0.97094	0.93357	0.97169	0.91240	0.97699	0.94424	0.95014	0.95611	0.94078	0.93577
0.91356	0.85777	0.95379	0.80788	0.93325	0.95807	0.90732	0.90995	0.93301	0.90565	0.91327	0.95322	0.95685	0.89378	0.95783	0.93903	0.87941	0.92205	0.92535	0.89923
2.3259	3.00147	1.58534	3.80735	2.19217	1.45192	2.27086	2.54176	1.97761	2.5393	2.54256	1.50828	1.19207	2.57175	1.4264	2.1802	2.84936	1.88493	1.87146	0.89923
0.39185	0.33649	0.39202	0.22953	0.44465	0.35812	0.28857	0.34381	0.44027	0.36467	0.45042	0.50843	0.37389	0.30188	0.72298	0.22712	0.73887	0.89881	0.8045	0.76526
0.85611	1.86036	0.26206	2.79552	0.2727	0.31029	0.98634	0.44027	0.36467	0.45042	0.50843	0.37389	0.30188	0.72298	0.22712	0.73887	0.89881	0.8045	0.76526	0.53433
0.88101	0.80806	0.91316	0.81425	0.93555	0.90407	0.7786	0.89601	0.89415	0.84629	0.9262	0.90125	0.91769	0.84908	0.94403	0.85629	0.90134	0.89192	0.88032	0.88442
0.84436	0.79234	0.88771	0.73592	0.89148	0.87008	0.80455	0.85758	0.89561	0.85954	0.86644	0.85677	0.91349	0.82099	0.92716	0.82636	0.80055	0.89084	0.89547	0.84467
2.84	3.20861	2.38655	4.14263	2.20448	2.54765	3.14987	2.76803	2.00737	2.48951	2.58804	2.65987	1.54401	3.01881	1.61428	2.74509	3.59426	2.04142	1.86167	2.75331

Above table I is for DWT features. Therefore, for MWT features can be extracted similarly with MATLAB tool.

VIII. EXPERIMENTAL RESULTS

The proposed method compared the results of DWT and MWT. It shown in fig.11, 12, 13, 14, 15, 16, 17.

1. DISCRETE WAVELET TRANSFORM

i. Juxta pleural lung nodule

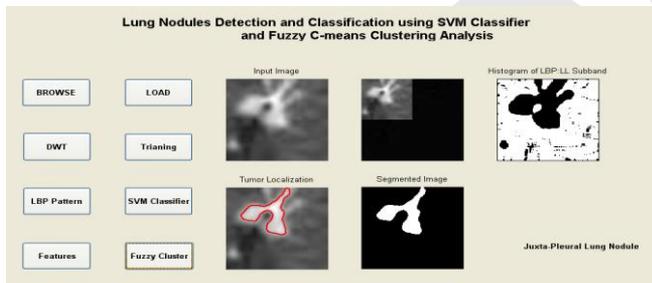


Figure 11-segmented image red color region affected

ii. Pleural tail lung nodule

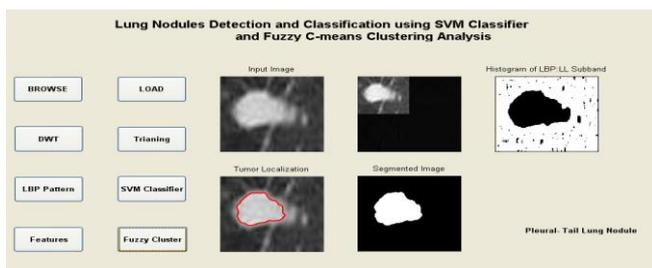


Figure12-segmented image red color region affected

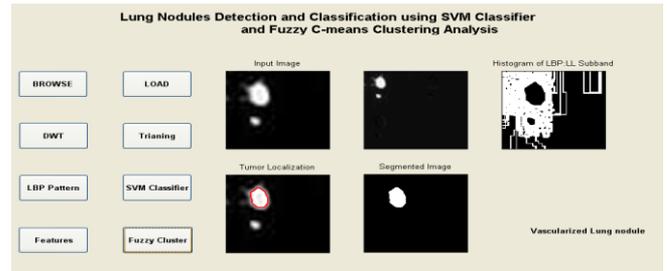


Figure 13- segmented image red color region affected

iv. Well circumscribed lung nodule

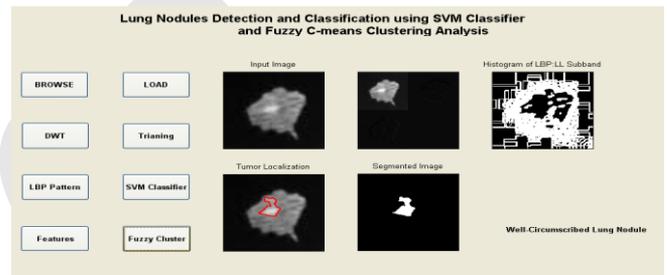


Figure 14-segmented image red color region affected

2. MULTI WAVELET TRANSFORM

v. Multi wavelet transform, juxta pleural lung nodule

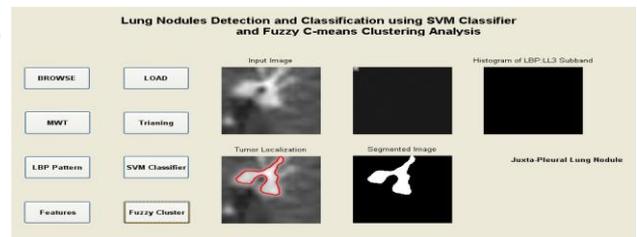


Figure 15-segmented image red color region affected

vi. Multi wavelet transform, pleural tail lung nodule

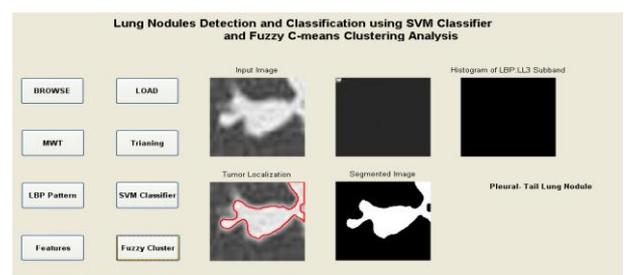


Figure16-segmented image red color region affected area

vii. Vascularized lung nodule

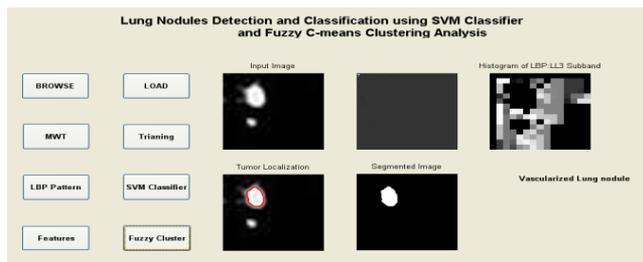


Figure 17- segmented image red color region affected

viii. Well circumscribed lung nodule

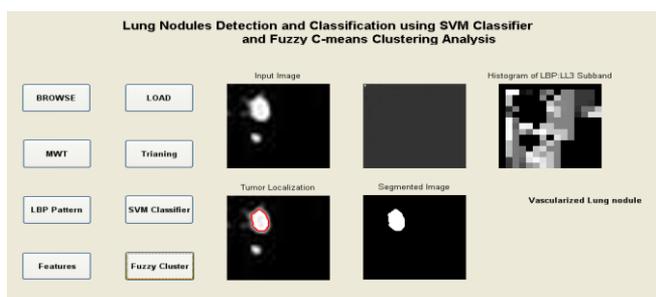


Figure 18- segmented image red color region affected

The fig.19 shows the accuracy comparison chart between discrete wavelet transform (DWT) and multi wavelet transform (MWT) based algorithm.

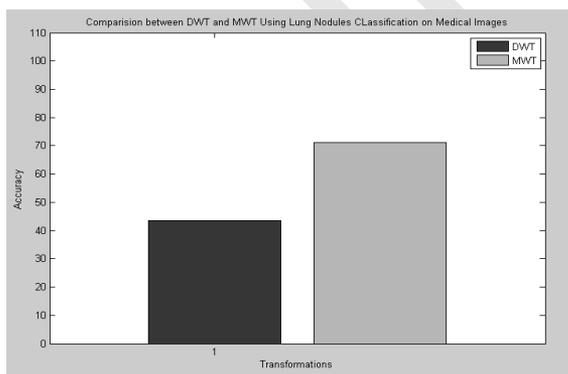


Figure 19. Comparisons between DWT and MWT

TABLE II. Comparison of accuracy

TRANSFORM	ACCURACY
Discrete wavelet	45%
Multi wavelet	75%

IX. CONCLUSION

This paper is presented a supervised classification on lung nodules. The support vector machine had classified stage in abnormal images. The accuracy was improved by multi wavelet and compared with discrete wavelet transform. The fuzzy c-means clustering for segmentation of detected part of the image. The four main categories of lung nodules well-circumscribed, vascular zed, juxta-pleural, and pleural-tail were the objects were differentiated.

REFERENCE

- [1] M.M. Eltokhy, I Faye, B.B. Samir, "A comparison of wavelet and curve-let for breast cancer diagnosis in digital mammogram". *Computers in Biology and Medicine*, vol.40,2010. pp.384–391.
- [2] J. Yao, A. Dwyer, R. M. Summers, and D. J. Mollura, "Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification," *Acad. Radiol.*, vol. 18, no. 3, pp. 306–314, 2011.
- [3] A. Farag, A. Ali, J. Graham, S. Elshazly, and R. Falk, "Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose CT scans of the chest," in *Proc. Int. Symp. Biomed. Imag.*, 2011, pp. 169–172
- [4] H. Wang, X.H. Guo, Z.W. Jia, H.K. Li, Z.G. Liang, K.C. Li, Q. He. "Multilevel binomial logistic prediction model for malignant pulmonary nodules based on texture features of CT image". *European Journal of Radiology*. Online: doi:10.1016/j.ejrad.2009.01.024.
- [5] J. J. Erasmus, J. E. Connolly, H. P. McAdams, and V. L. Roggli, "Solitary pulmonary nodules: Part I. morphologic evaluation for differentiation of benign and malignant lesions," *Radiographics*, vol. 20, no. 1, pp. 43–58, 2000.
- [6] L. Srensen, S. B. Shaker and M. De Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 559–569, Feb. 2010.
- [7] F. Zhang, Y. Song, W. Cai, Y. Zhou, M. Fulham, S. Eberl, S. Shan, and D. Feng, "A ranking-based lung nodule image classification method using unlabeled image knowledge," in *Proc. Int. Symp. Biomed. Imag.*, 2014.
- [8] F. Zhang, W. Cai, Y. Song, M.-Z. Lee, S. Shan, and D. Feng, "Overlapping node discovery for improving classification of lung nodules," in *Proc. Eur. Mol. Biol. Conf.*, 2013, pp. 5461–5464
- [9] S. Diciotti, G. Picozzi, M. Falchini, M. Mascalchi, N. Villari, and G. Valli, "3-D segmentation algorithm of small lung nodules in spiral CT images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 1, pp. 7–19, Jan. 2008.
- [10] L. Fan, C. L. Novak, J. Qian, G. Kohl, and D. Naidich, "Automatic detection of lung nodules from multislice low-dose CT images," in *Proc. SPIE, Med. Imag.*, 2001, vol. 4322, pp. 18281835.
- [11] A. A. Farag, "A variational approach for small-size lung nodule segmentation," in *Proc. Int. Symp. Biomed. Imag.*, 2013, pp. 81–84. online at <http://ieeexplore.ieee.org>.