QUANTITATIVE LUNG NODULE ANALYSIS SYSTEM (NAS) FROM CHEST CT

Amal Farag[1], PhD, Salwa Elshazly[1], Asem Ali[2], PhD, Islam Alkabbany[2], Albert Seow[2], MD and Aly Farag[2], PhD

[1]Kentucky Imaging Technologies, [2]University of Louisville

ABSTRACT

A modular nodule analysis system (NAS) for early lung nodules detection and classification is presented. Elastic active appearance models (AAM) are used to create realistic templates for nodule detection by template matching. A novel approach is presented to automatically annotate nodules in the design of the AAM templates. On sample nodules from ELCAP and LIDC studies, the AAM nodule templates provided 95% sensitivity of detection vs 80% for geometric templates, at 95% specificity rate. The paper presents an SVM classifier in a two-tier cascade framework for nodule classification. The classification results using Gabor features showed overall better results for identifying non-nodules, malignant and benign nodules with total average Area Under the Receiver Operating Characteristics (AUC-ROC) curves of 0.99 and average f1-score of 0.975. Results lay the foundation for a fully model based nodule analysis system for early detection of lung cancer at Stage 1a.

Index Terms— AAM models, nodule annotation, nodule detection, nodule segmentation, nodule classification.

1. INTRODUCTION

Lung cancer is devastating in terms of cost and human tolls; it is the second most common cancer and the leading cause of cancer deaths in both men and women in the United States (US). In 2014 (the most recent year numbers available)—215,951 people in the United States were diagnosed with lung cancer, including 113,326 men and 102,625 women; 155,526 people died from lung cancer, including 84,859 men and 70,667 women [1]. Globally, lung cancer is the most common malignancy; an estimated 1.8 million diagnosed cases in 2012 (most updated figures) and 1.6 million deaths occurring that same year [2]. Screening studies in the US and worldwide have been conducted in the past three decades, concluding that Low Dose CT Scanning (LDCT) is more efficient than Chest radiography/Chest X-ray (CXR) in early detection and diminished mortality, even with the risk of radiation associated with CT (e.g., ELCAP: Henschke et al., 1999 [3]; LIDC, Armato et al., 2004 [4]; NLST, Aberle et al., 2011 [5]; and the French Screening Study: Blanchon et al., 2007 [6]). These studies provided data for developing computer-assisted diagnosis systems (CAD) for lung nodule detection.

Computationally, sensitivity of nodule detection and classification have varied significantly. For example, in the past two years, Golan el al., 2016 [7], used a deep Convolutional Neural Network (CNN), reported sensitivity rate on the LIDC study of 71.2% with 10 false positives per scan. Setio et al., 2016 [8] reported: on 888 scans of the LIDC-IDRI dataset, Multi-View CNN reaches detection sensitivities of 85.4% and 90.1% at 1 and 4 false positives per scan, respectively. The most recent work (Setio, et al., 2017) [9] lists progress on automatic detection of lung nodules, using the 888 scans from the LIDC study, where leading solutions employed CNN, and combination of solutions achieved sensitivity over 95% at fewer than 1.0 false positives per scan.

Despite the enormous progress made, CAD based on LDCT suffer from two main problems: large false positive rates in nodule detection and large uncertainty with classification into benign or malignant; e.g., Rubin, 2015 [10]. Resolving these issues and proper integration with PACS viewers will lead to workflow that facilitates both efficiency and effectiveness of interpretation, and widespread acceptance of CAD.

Typical CAD systems consist of four steps (Fig.1): a) Filtering for removal of artifacts in the CT scans; b) Segmentation of lung regions from surrounding anatomy, without affecting nodules on the pleural surface; c) Detection of lung nodules by discriminating their characteristics from surrounding tissues; and d) Classification of nodules into benign or malignant.

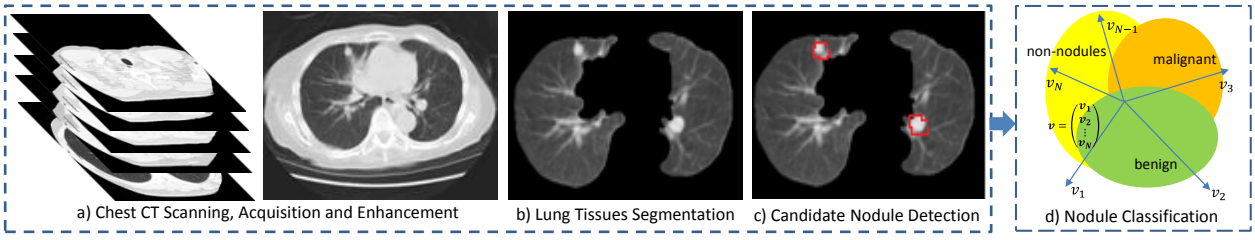


Fig. 1: Typical components of CAD systems for nodule analysis.

Robust methods are in place for the first two components in Fig. 1, Filtering and Lung Segmentation; for example, our homegrown methods (filtering based on Anisotropic Diffusion, and lung segmentation using statistical approach [11, 12]) are able to extract the lung tissues with minimal affect to nodules larger than 5 mm, including those on the outside pleural surface. *Reducing false positives in detection and improving the estimation of malignancy in Stage 1a, are the bottlenecks of CAD systems; and the focus of this work.*

The contributions of this paper over our related work (e.g., [12, 22]) are twofold: 1) we introduce a novel approach to annotate large nodule ensembles into the creation of active appearance model (AAM) templates, which significantly improved the detection using template matching, in terms of the sensitivity and specificity. 2) We improve the nodule classification using SVM using optimized multi-feature feature model, and investigate CNN-based nodule classification.

2. APPROACH

2.1 Deformable AAM models for lung nodules

In combined AAM models of *shape* $\mathbf{S}(\mathrm{x})$ and *appearance* $\mathbf{A}(\mathrm{x})$ of lung nodules, as a particular location x in the nodule's spatial support, a single set of parameters $\mathrm{c} = (c_1, c_2, \dots, c_L)^T$ is used; that is:

$$\mathbf{S}(\mathrm{x}) = \boldsymbol{S}_0(\mathrm{x}) + \sum_{\mathrm{i}=1}^{\mathrm{L}} c_i \, \mathrm{S_i}(\mathrm{x}) \qquad (1.a)$$

$$\mathbf{A}(\mathrm{x}) = \mathbf{A_0}(\mathrm{x}) + \sum_{\mathrm{i}=1}^{\mathrm{L}} c_i \mathrm{A_i}(\mathrm{x}) \qquad (1.b)$$

where $\boldsymbol{S}_0(\cdot)$ and $\mathbf{A_0}(\cdot)$ are *average* shapes and appearances, respectively. These parameters are identifiable using pre-annotated nodules and dimensionality reduction by Principle Component Analysis (PCA); e.g., [13].

Fig. 2 shows the average and first five Eigen nodules using an ensemble of size 24 per nodule category from the ELCAP study. Adding Eigen nodules to $\mathbf{A_0}(\cdot)$ would provide more resemblance to actual nodule topologies; their impact on nodule detection has not been previously studied. In the case of matched filter approach, templates (resemble the impulse responses) that possess characteristics to the desired signals (lung nodules in our case) which will provide better detection performance in terms of sensitivity and specificity.

However, the size and topology of nodules at Stage 1a pose difficulties in training the AAM models. Indeed, small-size nodules ($< 1\,cm$ in diameter) are harder to outline by the radiologists; e.g., the outlines by the investigators in the LIDC study – Armato et al., 2011 [14] – focused on the nodules' heads, which if used to construct the AAM models won't provide the desired specificity in nodule detection. Farag et al., 2013 [12] devised an empirical approach to annotate small-size nodules using a magnification of the ROI, then using curvatures measures to identify the contours' critical points. From these critical points, co-registration is performed by estimating the rotation, translation and scale.

After co-registration of the nodule ensemble, the PCA is conducted to provide the Eigen nodules (Fig. 2). Fig. 3 shows critical points on the contours of four small-size nodule categories.

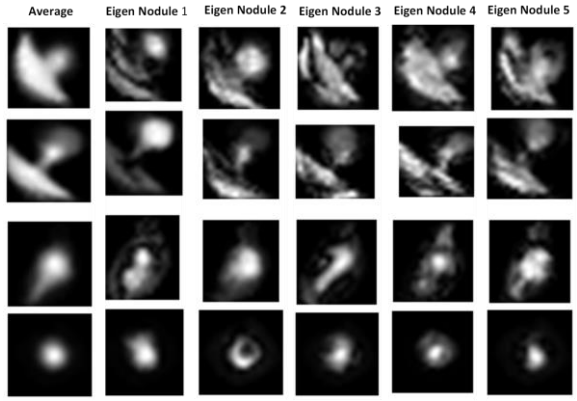


Fig. 2: Average and five Eigen nodules.

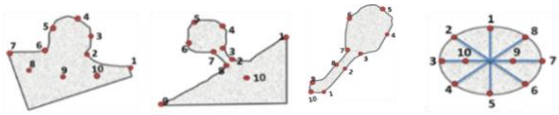


Fig. 3: Critical points for annotation (left to right): juxta-pleural, pleural-tail, vascularized and well-circumscribed.

2.2 Novel Nodule Annotation

We automated the nodule annotation and alignment by three basic steps, given a larger ensemble of nodules in Stage 1a: a) use the manual approach on an adequate set of nodules under each nodule category (e.g., well-circumscribed, vascular, juxta-pleural, pleural tail) – e.g., using 24 nodules per category; b) Construct the AAM model using these nodules, per category, using the approach of Farag, 2012 [11]; and c) co-register the rest of the nodules in the ensemble per category with the average nodule, then perform the PCA on the larger co-registered nodules.

Table 1 lists the main steps of the proposed algorithm. We note that nodules' annotation for AAM modeling is performed *offline*; hence, the elaborate and time-demanding tasks in this step will not affect detection and classification,

Table 1: Adaptive Nodule Modeling Algorithm

i) Generate nodule models using 24 manually annotated nodules per categories.
ii) Extract feature points for each nodule image per type using high and low curvature regions on the surface contours.
iii) For each nodule contour, the maximum or minimal curvature points are used as a biomarker for registering the samples.
iv) A modified Iterative Closest Point (ICP) [15] is used to compute the transformation matrix to register the samples to a commons reference; it exploits the curvature information in the matching process. Other methods may also be used (e.g., Abdelmunim et al., 2013 [16])
v) Construct the updated AAM model using Farag's approach [11] for small objects.

which are conducted *online*. We will examine the impact of the ensemble size on the quality of the AMM nodule template for the detection vs geometric templates.

2.3 Nodule Detection by Template Matching

Using the updated templates for the vascular ($\widetilde{\boldsymbol{A}}_{0V}$), well-circumscribed ($\widetilde{\boldsymbol{A}}_{0W}$), juxta-pleural ($\widetilde{\boldsymbol{A}}_{0J}$) and pleural-tail ($\widetilde{\boldsymbol{A}}_{0P}$), template matching is performed using the normalized cross-correlation (NCC) as the similarity measure (Farag, 2012 [11] showed that NCC was optimal for detection of small deformable objects over 10 other popular similarity measures). Within the body of the lungs (blue box regions in Fig. 4(b)) we may only use templates $\widetilde{\boldsymbol{A}}_{0V}$ and $\widetilde{\boldsymbol{A}}_{0W}$; while near the pleural surface we may only use $\widetilde{\boldsymbol{A}}_{0J}$ and $\widetilde{\boldsymbol{A}}_{0P}$ with equal weight and an *Exclusive OR* thresholding for acceptance as a *nodule candidate*. We also explored several search approaches for accuracy, speed and weight of templates per lung region.

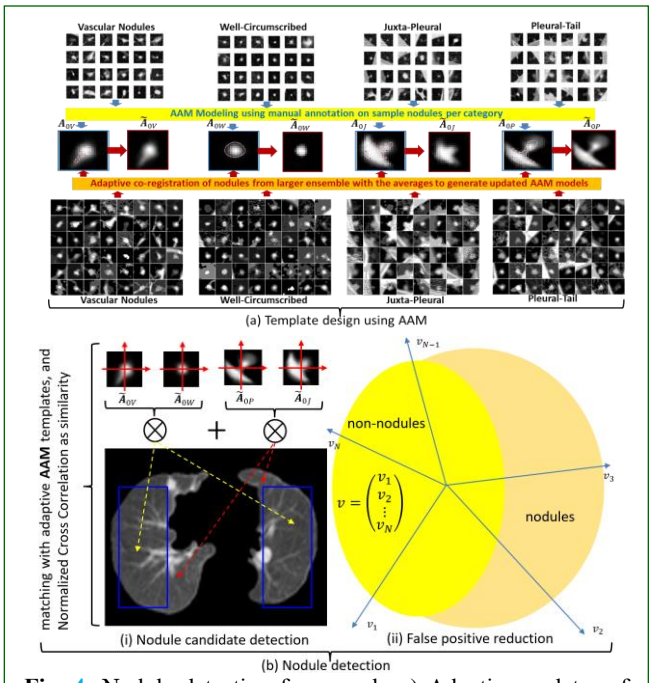
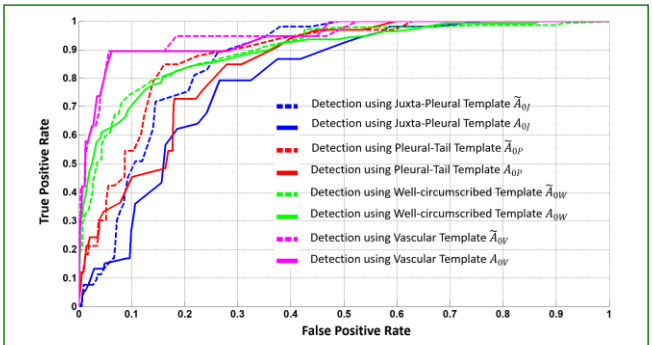
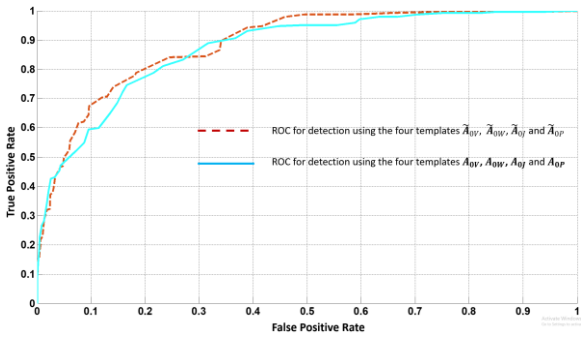


Fig. 4: Nodule detection framework. a) Adaptive updates of AAM nodule models. Manually annotated ensemble is used to generate average $\boldsymbol{A_0}$ nodule model per category. Larger ensembles are co-registered with $\boldsymbol{A_0}$ to generate updated AAM models $\widetilde{\boldsymbol{A}}_{\boldsymbol{0}}$ for detection. b) Detection performed by template matching followed by false positive reduction.

Experiment 1: Testing was performed on the LIDC using 1018 helical thoracic CT scans from 1010 different patients [22]; and ii) The ELCAP [17] has 397 nodules, 291 identified and categorized nodules are used in the detection process. On LIDC and ELCAP, we used only the average (mean) template model generated from the AAM approach vs parametric nodule template [11]; circular and semi-circular of radius 10 pixels. For the AAM templates, we used 24 nodules per nodule type to design the nodule models and the rest to test



(a) ROC for detection by updated AAM templates.



(b) ROC for detection by average AAM templates.

Fig. 5: Adaptive effect of AAM templates on detection.

the detection. Fig. 5 shows the ROC of 1-specificity vs. sensitivity, which has been reported in our earlier work [12].

Experiment 2: Testing was performed on the LIDC only using 1018 helical thoracic CT scans from 1010 different patients. We used 24 annotated nodules per category to generate the averages $\boldsymbol{A_{0V}}, \boldsymbol{A_{0W}}, \boldsymbol{A_{0J}}$ and $\boldsymbol{A_{0P}}$ and, average of 75 unannotated nodules per category, to generate $\widetilde{\boldsymbol{A}}_{0V}$, $\widetilde{\boldsymbol{A}}_{0W}$, $\widetilde{\boldsymbol{A}}_{0J}$ and $\widetilde{\boldsymbol{A}}_{0P}$, using the algorithm in Table 1. We then applied the template matching using normalized cross-correlation as the similarity measure. Fig. 5(a) shows the detection curves (without false positive reduction) per nodule category for the two AAM nodule models. Fig. 5(b) shows the average performance using combinations of the templates (as illustrated in Fig. 4(b)). These results when compared to results obtain using parametric templates [11], provide credence to the hypothesis that improved AAM models provide better performance in nodule detection, and it is worth further investigation for creating a data-driven CAD system.

2.4 False Positive Reduction

False positive reduction in detection can generally be achieved by a classification step that follows candidate nodule detection (e.g., [8], [18] and [19]). Various criterions have been introduced for false positive reduction which include: 1) distance from the thoracic wall, 2) nodule area/volume and mean diameter, and 3) mean Hounsfield Units (HU) in the nodule's head. In this paper, reduction of the false positive rates is obtained through the following: a)

improving the AAM templates – above experiment showed impact on detection; and b) adding features to discern nodules (e.g., size, location with respect to the thorax, and HU in the nodules' head region). The final outcome of the detection is a cropped region (nodule's spatial support) on which we apply nodule classification into benign/malignant. Fig. 6 shows a sample of detected and segmented nodules.

2.5 Nodule Classification

Our most recent work [22] examined features for nodule classification (benign vs malignant). Features based on i)

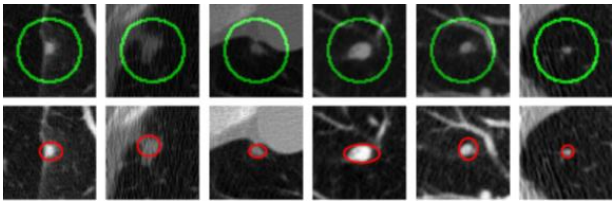


Fig. 6: Segmented nodules by level set method. Green: Initial contour; Red: segmented detected nodule.

Gabor filter [23, 24], ii) Multi-resolution Local Binary Pattern (LBP) texture features [25, 26] and iii) signed distance fused with LBP which generates a combinational shape and texture feature, are utilized to provide feature descriptors of malignant and benign nodules and non-nodule regions of interest. Support Vector Machines (SVM) classifier in serial and two-tier cascade frameworks are optimized and analyzed for optimal classification results of nodules (Fig. 7(i)-(ii)). On 1191 nodule and non-nodule samples from the LIDC database, where samples were annotated into one of three categories; Benign (B), Malignant (M) or non-nodule (N). The data distribution is as follows: 723 benign and 223 malignant nodules between $3mm \leq n \leq 10mm$ and 245 non-nodules. We obtained the following conclusions: The classification results from the two-tier cascade SVM using Gabor features showed overall better results for identifying non-nodules, malignant and benign nodules with average Area Under the Receiver Operating Characteristics (AUC-ROC) curves of 0.99 and average F_1 score of 0.975 over the two tiers.

Recently, Convolutional Neural Network (CNN) has been presented as an end-to-end framework that performs both feature extraction and classifier training. Fig. 7(iii), illustrates the proposed CNN architecture for nodule/non-nodule classification. In this experiment, we used the LIDC database, same nodule ensemble as in the SVM experiment, and cross validation method to evaluate the approaches. The number of samples was inadequate (i.e. smaller than required) for training a deep learning model, thus, data samples were augmented to avoid overfitting, the proposed classifier provided an F_1 score of 0.93.

3. DISCUSSION AND CONCLUSION

The focus of this paper has been on nodule detection and classification. The AAM approach to design elastic templates was enhanced by designing an automatic annotation framework using biomarkers on the nodule contours that highlighted the major features that were then used for registration. A simultaneous improvement in sensitivity and specificity of nodule detection was observed from using the new templates. The level set methods enable accurate segmentation of the detected nodules. A fully model-based mechanism to detect, authenticate (reduce false positives) and segment/crop the nodules for the last step in the CAD system, classification, is presented. Classification using cascaded SVM and statistical methods provided significant results in identifying non-nodules, malignant and benign nodules with average AUC-ROC curves of 0.99 and average F_1 score of 0.975. Our current efforts is focused on deploying the NAS system mainly as model-based in order to enable the creation of a large discriminative nodule database, such that further enhancements and deployments of machine learning methodologies can be tested for detection and classification.

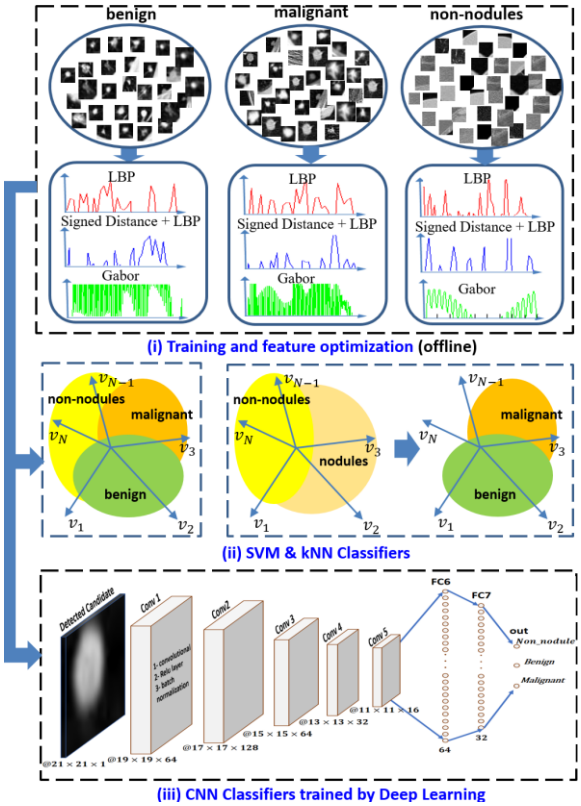


Fig. 7: Framework for nodule classification using two approaches: SVM + kNN and CNN.

4. REFERENCES

[1] Centers for Disease Control and Prevention. National Center for Health Statistics. CDC WONDER On-line Database, compiled from Compressed Mortality File 1999-2014 Series 20 No. 2T, 2016.

[2] World Health Organization. International Agency for Research on Cancer. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. Lung Cancer.

[3] Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999; 354:99-105.

[4] Armato, G. McLennan, G., McNitt-Gray, MF, Meyer, CR, Yankelevitz, D, Aberle, DR, Henschke, CI, Hoffman, EA, Kazerooni, EA, MacMahon, H, Reeves, AP, Croft, BY and Clarke, LP "Lung image database consortium: developing a resource for the medical imaging research community," Radiology, vol. 232, no. 3, pp. 739–748, 2004.

[5] National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011 Aug 4; 365(5):395-409.

[6] Blanchon, et al., Baseline results of the Depiscan study: A French randomized pilot trial of lung cancer screening comparing low dose CT scan (LDCT) and chest X-ray (CXR), Lung Cancer, Volume 58, Issue 1, October 2007, Pages 50–58

[7] R. Golan, C. Jacob and J. Denzinger, "Lung nodule detection in CT images using deep convolutional neural networks," 2016 Int. Joint Conf. on Neural Networks (IJCNN), Vancouver, BC, 2016, pp. 243-250.

[8] Arnaud A. Setio, et al.; "Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks". IEEE Tran. on Med.Imaging. 2016, Volume: 35, Issue 5, Pages: 1160 – 1169.

[9] Arnaud A. Setio, et al "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge". Medical Image Analysis 42 (2017) 1–13.

[10] Rubin, GD. Lung Nodule and Cancer Detection in CT Screening. Journal of thoracic imaging 30 (2), 2015: 130-138

[11] Amal A. Farag, "Modeling Small Objects under Uncertainties: Novel Algorithms and Applications," – PhD Dissertation, University of Louisville, Department of Electrical and Computer Engineering, 2012.

[12] Amal Farag, James Graham and Aly Farag, "Deformable Models for Random Small-size objects: Case of Lung Nodules in CT Tomography." International Conference on Acoustics, Speech, and Signal Processing (ICASSP-13). Vancouver, Canada, May 26-31, 2013.

[13] T.F. Cootes, G.J. Edwards and C.J. Taylor. "Active Appearance Models", IEEE PAMI, Vol.23, No.6, pp.681-685, 2001.

[14] Armato SG, et al., Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys. 2011 Feb; 38 (2):915-31.

[15] Besl, PJ, McKay, ND "A Method for Registration of 3-D Shapes". IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol.14 (2): 239–256, 1992.

[16] H. Abdelmunim, Amal Farag, and Aly A. Farag, "Shape Representation and Registration in Vector Implicit Spaces: Adopting a Closed Form Solution in the Optimization Process," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 3, pp. 763-768, March 2013.

[17] ELCAP public lung image database: http://www.via.cornell.edu/databases/lungdb.html

[18] Amal Farag, Mostafa Farag, James Graham, Salwa Elshzaly, Mohamed El-Mogy and Aly Farag, "Modeling of Lung Nodules from LDCT of the Human Chest: Algorithms and Evaluation for CAD Systems." Advances of Shape Analysis in Medical Image Analysis. Editors: Li and Tavares. Pp. 259-290. Elsevier 2014.

[19] Hosseini R, Qanadli SD, Barman S, Mazinani M, Ellis T, Dehmeshki J, An Automatic Approach for Learning and Tuning Gaussian Interval Type-2 Fuzzy Membership Functions Applied to Lung CAD Classification System, IEEE Transactions on Fuzzy Systems, April 2012, 224-234.

[20] Amal Farag, Hossam Abdelmunim, James Graham and Aly Farag, "A Novel Approach for Lung Nodules Segmentation in Chest CT using Level Sets," – IEEE Transactions on Image Processing 22(12): 5202-5213, December 2013.

[21] H. Abdelmunim, Amal Farag, and Aly A. Farag, "Shape Representation and Registration in Vector Implicit Spaces: Adopting a Closed Form Solution in the Optimization Process," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 3, pp. 763-768, March 2013.

[22] Amal Farag, Asem Ali, Salwa Elshazly and Aly Farag, "Feature fusion for lung nodule classification". Int. J Comput Assist Radiol Surg. 2017 Jun 16. doi: 10.1007/s11548-017-1626-1.

[23] Tai Sing Lee, "Image Representation Using 2D Gabor Wavelets," IEEE Trans. On Pattern Analysis and Machine Intelligence, VOL. 18, NO. 10, pp. 1-13, OCTOBER 1996.

[24] C. Liu, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 26, no. 5, pp. 572–581, 2004.

[25] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, 2002, pp. 971-987.

[26] Farag AA, Biomedical Image Analysis: Statistical and Variational Approaches. Cambridge University Press, 2014.