A Hierarchical Decision Engine for Computer Aided Lung Nodule Detection from CT Images.

Conference Paper · January 2010

Source: DBLP

Omer Muhammet Soysal
Southeastern Louisiana University

Jianhua Chen
Louisiana State University

Helmut Schneider
Louisiana State University

A Hierarchical Decision Engine for Computer Aided Lung Nodule Detection from CT Images

Omer M Soysal[1*], Jianhua Chen[2], Steven Bujenovic[3], Helmut Schneider[1]
[1]Department of Information Systems and Decision Sciences, Louisiana State University, omsoysal@lsu.edu, hschnei@lsu.edu
[2]Department of Computer Science, Louisiana State University, jianhua@csc.lsu.edu
[3]Lake P.E.T. Imaging Center, Our Lady of the Lake, l.bujenovic@gmail.com
Baton Rouge, USA

Abstract

We present a novel hierarchical modular decision engine for lung nodule detection from CT images implemented by Artificial Neural Networks. The proposed Computer Aided Detection (CAD) technique encompasses several desirable properties such as mimicking physicians by means of geometric multi-perspective analysis, computational efficiency, and most importantly archiving high performance in detection accuracy. One advantage of this decision engine is that it supports the combination of spatial-level and feature-level information analysis in an efficient way. Our methodology overcomes some of the limitations of current lung nodule detection techniques by combining geometric multi-perspective analysis with global and local feature analysis. The proposed modular decision engine design is flexible to modifications in the decision modules; the engine structure can adopt the modifications without re-designing the entire system. The engine can easily accommodate multi-learning scheme and parallel implementation so that each information type can be processed (in parallel) by the most adequate learning technique of its own.

Key Words: Image processing in medicine and biological sciences, Pattern classification and recognition, Computer-based medical systems.

1 INTRODUCTION

Lung cancer is one of the most lethal cancer types. It is reported in (Partain et al. 2005) that it accounts for 32% and 25% of cancer deaths among men and women (Dodd et al. 2004) and causing 150,000 deaths a year (Li et al. 2003) in the United States, respectively. It is proved that early detection of this cancer type may increase the chance of survival (Sluimer et al. 2006, Dood04]. It has been observed that radiologists may overlook some nodules due to heavy load of number of images and fatigue. It is also proven that a computer aided detection and diagnosis (CADD) system can provide a 'second opinion' for radiologists to increase their interpretation performance (Doi 2005).

As seen from the facts stated above, invention of a novel methodology that can assist to detect a cancer would be very crucial. In many areas of medicine, it has been proven for a long time that a CADD system is one of the important tools to help early detection of malicious disease. Main goal of a CADD system is to give a second opinion to enhance the performance of radiologist while increasing the number of true positive detections and decreasing the number of missed nodules. A CADD system is mainly used to assist human interpreters to identify and characterize abnormalities automatically (Ko and Naidich 2004] "making lesions easier to detect and classify and potentially identify at an earlier stage" (Krupinski 2004). Although automatic nodule detection systems may yield false outputs, (Ko and Naidich 2004] showed that image interpretation time is significantly improved with CADD. Another advantage of using a CADD system is to reduce the workload of radiologist and consequently decreasing error as well (Armato et al. 2004). As stated in (Rubin et al. 2000), with the advanced technology of CT scanning, 'data explosion' is a big challenge when we consider that a radiologist has to analyze more than 300 image sections per case. Explosion of image data affects human judgment due to physiological exhaustion like fatigue or distraction (Armato et al. 2000).

Nodule detection techniques can be categorized with respect to their spatial analysis dimension such as 2D or 3D detection and their analysis granularity that can be global or local. It is obvious that high-dimensional analysis is superior to lower ones as regard to detection accuracy; whereas it demands more computational resources. Global analysis approaches, which require typically less computational resources, allow us to characterize the whole volume of interest (VOI) with higher-level abstraction. However, the higher-level abstract features may leave out useful details from the image, thus weakening discriminative power. In contrast, local analysis approaches can provide more descriptive features while leading to the curse of dimensionality that may require advanced processing techniques. Moreover, efficiently integrating the numerous local analysis results into a global decision is a challenging task.

As an important fact that human intelligence is the best nodule detector and yet there is no machine intelligence that can compete with this human detection capability. Whereas human visual recognition power is limited by eye vision capacity, on which the machine can be superior. Therefore, a comprehensive nodule detection technique should be able to combine higher spatial dimensional analysis with global and local feature level analysis while demanding less computational resources and high performance detection accuracy, and be able to mimic human intelligence. In addition, the detection system should be flexible so that different classification approaches can be used together and modification of its units should not require redesigning the entire system.

According to our extensive literature review (e.g., Giger et al. 1994, Kanazawa et al. 1996, Kawata et al. 2000, Lee et al. 2001, Hu et al. 2001, Suzuki et al. 2003, Takiza03, Arma04, Paik et al. 2003, Zhang et al. 2005, Way et al. 2006) about pulmonary nodule detection from 3D CT images, the current approaches analyze a VOI for suspicious objects by 1) searching through transverse plane or 2) performing segmentation. The first group approaches aim either to find the most descriptive section or to model the VOI based on some spatial assumptions before extracting characteristic features; the second group approaches perform, first, a 3D segmentation and then extract features for pattern recognition. On the other hand, they perform feature level analysis in a non-categorical way; so the size of a feature vector has to face curse of dimensionality. Another drawback of the current approaches, which employ 3D analysis, is that they consider entire VOI to produce some global feature vectors, which have less capability of characterizing complicated lung volumes.

In this study, we introduce a novel hierarchical modular decision engine for nodule detection. The central idea is to use a hierarchy of classifiers (predictors) for the task of classifying a volume of interest. The classifiers at the lowest layer of the hierarchy focus on series of (spatial) slices of the 3D image, and classify these slices independently according to features of certain type. The outputs of classifiers at layer *N* form inputs to the classifier at the next layer to produce some intermediate predictions. These predictions can be at spatial- or feature-level to be explained in the sequel. We have implemented a prototype system using the Inter-slice inter-plane (ISIP) model (to be described in Section 3).

The rest of the chapter is organized as follows. In section 2, the data used is described. In section 3, the hierarchical nodule detection method and the implemented ISIP model are presented. Experimental results are reported and analyzed in Section 4 on applying ISIP to a collection of CT images. Section 5 is a brief conclusion.

2 MATERIALS

The nodule and non-nodule image series are provided by National Cancer Institute from National Cancer Imaging Archive (NCIA) and by Lake P.E.T. Imaging Center (LakePet), respectively. The nodules were marked by four different radiologists and their location is saved as an XML file. The non-nodule imjects (image objects) are extracted semi-automatically. In this research, we used 80 nodule and non-nodule imjects in total with 40 imjects from each class. Nodule samples were collected from 20 patients' anonymous CT scans which are provided by NCIA and non-nodule samples were extracted from a healthy person's image series, which were provided by (LakePet). Image series has a resolution of 512 by 512 resolution and slice thickness from 2 to 5 mm.

3 THE HIERARCHICAL ENGINE AND ISIP MODEL

The problem addressed in this section is to classify (as nodule or not) a VOI from 3D CT lung image. The proposed decision engine model is a hierarchical engine which has several tiers. The engine analyzes a volumetric data based on selected features by obtaining some intermediate decisions from the first layer to the final layer. The intuition behind the proposed model is improving accuracy of a prediction through hierarchically connected information levels. Our experimental results show that this multi-level approach has a capability of obtaining the most improved decision at the final level. The engine is composed of several layers which contain several modules of a predictor and a regrouping processor. Each module at the same model layer is totally independent from each other while they are integrated to some modules at the successive layers. This independent processing makes it possible to implement the decision engine in a parallel processing environment.

3.1 The Hierarchical Modular Decision Engine

The decision engine analyzes a volumetric data at different information levels. These levels include spatial levels as slice, plane, and volume and feature levels as feature-class and feature-type. Spatial information levels are obtained by partitioning 3D space into three orthogonal planes[♣] — namely transverse (*x*-*y*), coronal (*x*-*z*), and sagittal (*y*-*z*) — and the data in the direction of each plane is sub-divided into 2D slice sequences which are composed by the most primitive unit (voxel). At the slice level, a prediction is obtained independently for each slice. At the plane level, the prediction is obtained by combining predictions for all slices of the plane. Similarly, the volume level synthesization utilizes the plane level outputs to reach a decision about the entire 3D volume.

The features used for the VOI classification task are partitioned into two feature classes: geometric class and photometric class. Within each class, the features are further partitioned into various feature types. For example, the features in the geometric class are grouped into 'spectral feature type' and 'circularity' feature type". Feature-class (or

[♣] As a convention, the term 'slice' alone refers to 'spatial slice' on a plane and the term 'plane' alone refers to 'spatial plane' in 3D Cartesian coordinate system.

feature-type) information level synthesization is performed using prediction results of other information levels by employing a specific feature-class (or feature-type) predictor.

In a non-modular predictor, all of the volumetric data (and all features) are utilized together to obtain a decision. In contrast, the modular approach partitions the raw data into several spatial and feature levels and combines a hierarchy of predictions at these levels in an efficient way. Therefore, the proposed modular engine model mimics a human expert who analyzes a VOI from different projections; further, the engine, after obtaining machine coded features from the raw data, enhances the decision-making process with multi-level feature analysis which is beyond the limit of a human intelligence.

3.2 The ISIP Model

The inter-slice inter-plane decision (ISIP) model analyzes a volume based on each feature-type separately. For a VOI and a specific feature type, the decision engine will first perform spatial and feature-level prediction (from slice to plane, and from plane to entire volume) using the specific feature type. Namely we are first looking for the answers to questions like "What is class of the object in the VOI according to the spectral feature? As the model name ISIP refers, analysis and synthesis is performed by first combining results from different slices, and then combining results from the 3 spatial planes. The individual feature-type predictions for the whole volume are then integrated for each feature-class separately before reaching the final classification.

The ISIP engine starts analyzing a VOI from the slices of a plane to obtain some predictions at the feature-type level. Then, a synthesization occurs at the plane level followed by the volume level synthesization; the overall decision is obtained by the module M_O. Figure 1 illustrates the ISIP structure of the modular decision engine. The model layers are explained in the sequel.

Layer 1. Feature-type Analysis of the VOI: This layer is composed of feature-type modules $M_{F\bullet}$. The first prediction about each slice based on a specific feature-type is obtained at this layer. The decision engine seeks for an answer to a question like "to what extent does the imject in the slice belong to the nodule class according to each individual feature-type?".

Layer 2. Plane and Volume Level Synthesization: This layer is composed of plane and volume level modules $M_{P\bullet}$ and $M_{V\bullet}$, respectively . An analysis is conducted separately on the VOI based on each individual feature type; first, a prediction is obtained at the plane level, and then output of the 3 plane predictors are used to obtain a volume level prediction. The decision engine seeks for an answer to a question like "to what extent does the imject in the VOI belong to the nodule class according to a shape feature when we look first through planes individually then altogether?". On the other hand, instead of using two types of module (namely plane and volume level modules), plane level prediction can be skipped and only volume level synthesization can be performed; this is depicted as dashed line box ($M_{P\bullet}$) in Figure 1.

Layer 3. Feature-Class Level Synthesization: This layer is composed of feature-class level modules $M_{FC\bullet}$. A prediction about the VOI is obtained based on all feature types of the feature-class. The decision engine seeks for an answer to a question like "to what extent does the imject in the VOI belong to the nodule class according to predictions considering all geometric features?".

Layer 4. Final Decision: This layer has a module M_O which produces the final decision about the VOI. The module has a threshold operator after the predictor unit to obtain the final classification result. The decision engine seeks for an answer to a question like "to what extent does the imject in the VOI belong to the nodule class considering geometric and photometric features altogether?"

Note that under the general framework of the proposed hierarchical model, one can easily design decision engines with different orders of analysis/synthesis. For example, a model can first perform feature-level and then spatial-level analysis/synthesis to obtain the final decision. In addition, one could use different machine learning methods to train the classifiers at different layers.

4 EXPERIMENTAL RESULTS

4.1 Experimental Set-up

We have implemented the ISIP model with back-propagation neural networks as classifiers at each decision node in the hierarchy. We tested our model with six experiment groups; each group has several sub-groups. In this section, we analyze results of experiment group A.

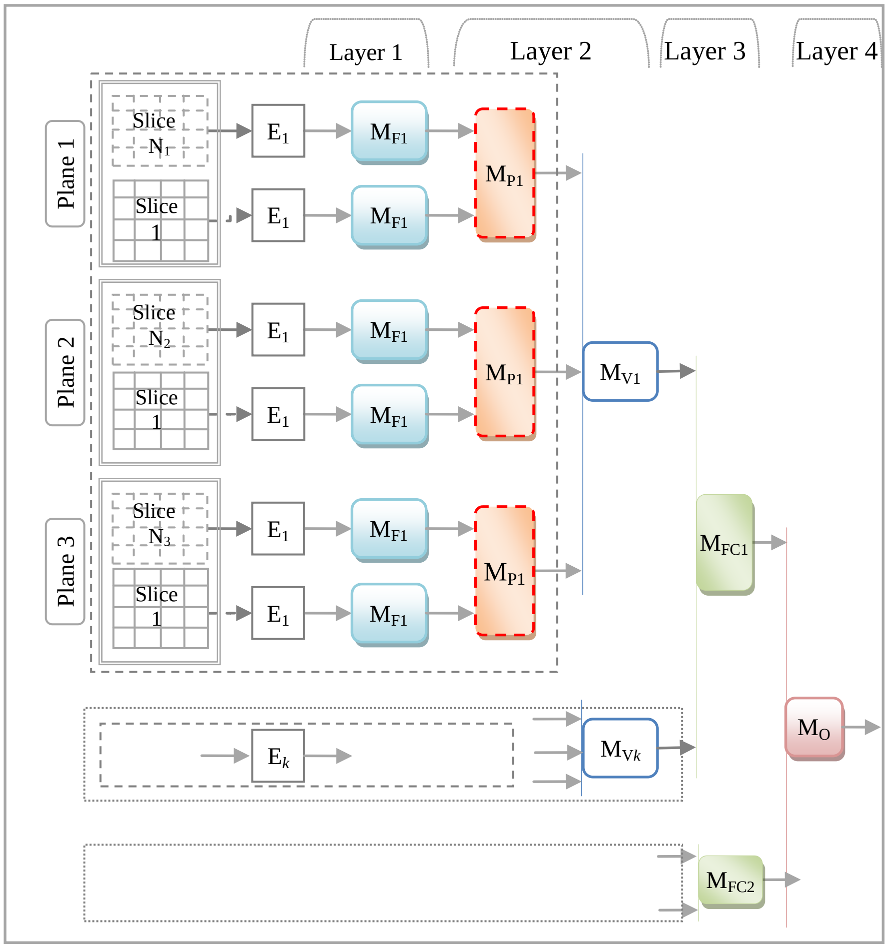

Figure 1 The ISIP conceptual model structure (M_{Fk}: Feature-type level module, $M_{P\bullet}$: Plane level module, $M_{V\bullet}$: Volume level module, $M_{FC\bullet}$: Feature-class level module, M_O: Overall classification module, E_k: Feature Extractor of a feature type k)

In the experiment group-A, we applied the Levenberg-Marquet learning algorithm using batch processing approach. This algorithm trains a back-propagation neural network by adaption of the parameter momentum μ. We refer the reader to (Hagan et al. 1994) for details of the learning algorithm. The algorithm starts with a small stability constant μ (like 0.001), and μ is increased by mu_inc (like multiplying by 5) if the performance function (sum of squared errors) does not produce a smaller value; on the other hand, if the function yields a smaller value, then μ is divided by mu_dec for a faster convergence. We conducted 72 experiments with varying training parameters: mu_dec = {0.05, 0.1, 0.5}, mu_inc = {5, 10}, mu_max = {100, 1000, 10e10}, epochs = {10, 10, 500, 1000}, and keeping other constants which are mu = 0.001, time = 60 min max, goal = 0, min_grad = 1e-10, max_fail = 3. The network structure is composed of 1 hidden layer. The numbers of nodes in the hidden layer of each predictor and feature types which are used in the Decision Engine are given in Table 1; the output layer of each ANN has 1 node. Our experiments aim at assessing the performance of the ISIP model under various choices of ANN training parameters. The network is trained and tested by randomly selected objects. The whole data set is divided in training, validation and testing having the equal number of objects from nodule and non-nodule sets. The size of the training, validation, and testing data set are 48 (60% of 80 imjects), 16 (20%), and 16 (20%), respectively.

Table 1 Number of hidden nodes in each ANN predictor (ML: Model Layer)

Feature Class	Feature Type	ML 1	ML 2	ML 3	ML 4	ML 5
Geometric	Spectral	2	1	2	2	2
	Circularity	9	1	2		
Photometric	Co-occurrence	11	1	2	2	
	Run-length	12	1	2		

4.2. Feature Vector Data Set

We developed a tool to preprocess the image slices from 3 planes and to extract feature vectors. In this study, we used two feature classes each having two feature types, namely geometric and photometric and two object class, namely nodule and non-Nodule. Two types of shape features, namely spectral feature and circularity-regional shape features, are used. The length of the spectral shape feature vector is 50. The circularity-regional feature vector is composed of 5 circularity measures including elongation, eccentricity, and circularity by radial distance, circularity by CPR which is obtained using CPR of the border, circularity by spectral feature (Soysal and Chen 2007), and 3 regional measures including normalized Euler's number, compactness, and convexity. Photometric feature vectors employed are co-occurrence and run-length feature vectors. The co-occurrence vector is composed of energy, entropy, contrast, the inverse difference moment, the correlation among the gray levels. The run-length feature vector is composed of gray-level non-uniformity, short-run emphasis, long-run emphasis, run percentage, run-length non-uniformity.

4.3 Performance Analysis

In this sub-section we analyze nodule and non-nodule detection performance of the decision engine based on assessment measures, namely sensitivity, specificity, and area under the ROC curve. Sensitivity measures the ratio of true positive samples (nodules) detected out of all positive samples (Kraemer et al. 2005). Specificity measures the ratio of true negative samples (non-nodules) out of all negative samples. The area under the ROC curve *Az* is another measure of classification performance in terms of sensitivity and complement of specificity (Lasko et al. 2005). The ROC curve is obtained by applying some threshold values to the classifier output and then calculating corresponding true positive rates (sensitivity) and false positive rates (1 - specificity) (Metz 1986).

Figure 2 depicts a summary of classification performances at the last layer of the Decision Engine for the experiments group-A; in the following, all analysis refers to Figure 2 unless indicated otherwise.

Goodness of classification for nodule imjects: Max-bar reads that the Decision Engine classifies all nodules perfectly for some parameter sets (with Sensitivity = 100%). We can infer from the median and mean bars that most of the parameter sets yield a classification performance above the mean according to sensitivity median of 88%. This indicates that, in classification of nodule imjects, the Decision Engine is not affected too much by the parameters selected for a sample set. The MIN curve shows that some of the parameter sets cannot yield a good classification.

Tendency to nodule and non-nodule (type I and II error analysis): 'Type I' error is defined as the percentage of true negative examples classified as 'positive'; thus type I error = (1-specificity). Similarly, 'type II' error refers to the percentage of true positive examples classified as 'negative', i.e., type II error = (1-sensitivity). The comparison of medians for sensitivity (88%) against specificity (75%) yields that the Decision Engine has a slight tendency towards the nodule prediction in most of the parameter sets.

Goodness of classification for non-nodule imjects: According to the max-bar, the Decision Engine classifies all nodules perfectly for some parameter sets (with specificity = 100%). The median and mean bars read that most of the parameter sets yield a classification performance above the mean according to specificity (75% and 70%, respectively). This indicates that, in classification of non-nodule imjects, the Decision Engine is not affected too much by the parameters observed.

Analysis of performance variations: Variation among sensitivity, among specificity, and among Az scores is 19%, 30%, and 22%, respectively. This means that the engine is more sensitive as regard to non-nodule detection compared to nodule detection. We can also conclude that the engine is less sensitive to the parameter differences in measuring the type II error compared to type I error.

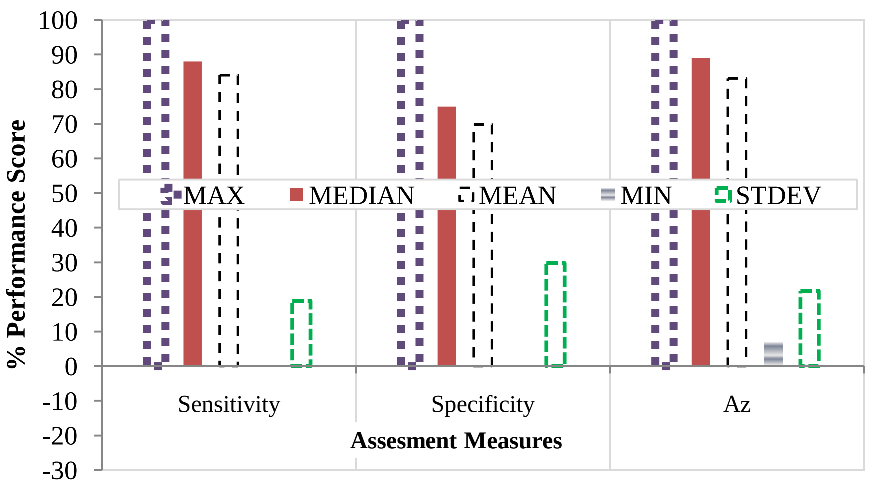

Figure 2 Overall classification performance of the Decision Engine against parameter variations (summary of all experiments for the final layer)

The parameter set which yields a better classification performance: Parameter sets of some selected experiments are shown in Table 2. This parameter set achieved 100% performance scores for sensitivity, specificity, and Az. The correlation among the experiment parameters (mu_dec, mu_inc, and mu_max) are -0.16, -0.01, and 0.01, respectively; this shows that the elapsed time is not sensitive to these parameter variations.

Table 2 Parameters of some selected experiments from group-A

	Epochs	mu_dec	mu_inc	mu_max	Elapsed Time (sec.)
Exp-42	500	0.05	10	1E+11	19
Exp-26	50	0.1	5	1000	20
Exp-33	50	0.5	5	1E+11	11
Exp-55	1000	0.05	5	100	13
Exp-69	1000	0.5	5	1E+11	12

Nodule detection performance for the experiment 42 (number of epochs = 500, mu_dec = 0.05, mu_inc = 10, mu_max = 10^{10}, and elapsed time = 19 sec) is depicted in Figure 3. As seen from the figure, the classification performance improves from the first layer of the Decision Engine to the final layer, in general.

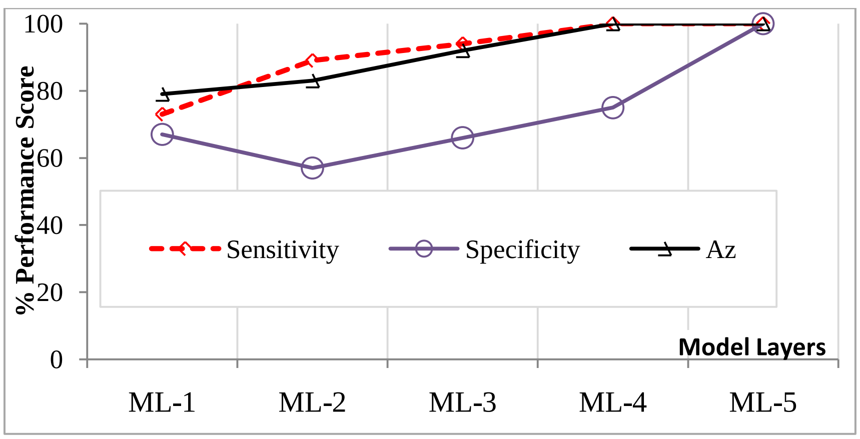

Figure 3 Average change of classification performance through Decision Engine Layers

4 CONCLUSIONS

In this study, we presented a hierarchical modular decision engine for lung nodule detection from 3D CT images. The decision engine consists of a collection of decision components organized in a hierarchical way such that a higher-level decision component synthesize the predictions at the lower-level through either spatial/feature-level synthesization. One advantage of this decision engine is that it supports the combination of spatial/feature-level information analysis in an efficient way. We also reported experimental results of a prototype implementation of the ISIP model on real-world lung CT image data. The ISIP model achieved very good performance assessed by sensitivity, specificity and area under the ROC curve. The empirical study is limited by the size of the available data set. Further experiments with larger data set would be desirable. Nonetheless, the studies so far support the effectiveness of the hierarchical decision model.

References

Armato SG, Giger ML, Macmahon H et al (2000) Analysis Of A Thre Dimensional Lung Nodule Detection Method For Thoracic CT Scans, Proceedings of SPIE, Medical Imaging, Image Processing – 3979:103-109

Armato SG, McLennan G, McNitt-Gray MF (2004) Lung Image Database Consortium: Developing A Resource For The Medical Imaging Research Community, Radiology, 232:739-748

Dodd LE, Wagner RF, Armato SG et al (2004) Assessment methodologies and statistical issues for computer-aided diagnosis of lung nodules in computed tomography: contemporary research topics relevant to the lung image database consortium, Academic Radiology, 11:462– 475

Doi K (2005) Current status and future potential of computer aided diagnosis in medical imaging, British Journal of Radiology, 78:S3-s19

Giger ML, Bae KT, Macmahon H (1994) Computerized Detection Of Pulmonary Nodules In Computed-Tomography Images, Investigative Radiology, 29(4):459-465

Hagan MT, Demuth HB, Beale M (1995) Neural Network Design, PWS Publishing

Hu S, Hoffman EA, Reinhardt JM (2001) Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images, IEEE Trans. Medical Imaging, 20(6):490–498

Ko JP and Naidich DP (2004) Computer-Aided Diagnosis and The Evaluation of Lung Disease, Journal Of Thoracic Imaging, 19(3): 136-155

Kanazawa K, Kubo M, Niki N (1996) Computer Aided Screening System for Lung Cancer Based on Helical CT Images, Lecture Notes In Computer Science, Proceedings of the 4th International Conference on Visualization in Biomedical Computing, 1131:223 - 228

Kawata Y, Niki N, Omatsu H et al (2000) Quantitative analysis of internal texture for classification of pulmonary nodules in three-dimensional thoracic images, Proceedings of SPIE, Medical Imaging Image Processing, 3979:863-871

Kraemer HC, Lowe KK, and Kupfer DJ (2005) How to understand what research tells us about risk, Oxford University Press, New York

Krupinski EA (2004) Computer-aided Detection in Clinical Environment: Benefits and Challenges for Radiologists, Radiology 231(1):7-9

LakePET Lake P.E.T. Imaging Center , Baton Rouge, LA, USA http://www.ololrmc.com/body.cfm?xyzpdqabc=0&id=11&action=list

Lasko TA, Bhagwat JG, Zou KH, and Ohno-Machado L (2005) The use of receiver operating characteristic curves in biomedical informatics, Journal of Biomedical Informatics, 38:404–415

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

Li Q, Li F, Shiraishi J et al (2003) Investigation of new psychophysical measures for evaluation of similar images on thoracic computed tomography for distinction between benign and malignant nodules, Medical Physics, 30(10):2584–2593

Metz CE (1986) ROC methodology in radiographic imaging, Invest Radiol, 21:720–733

NCIA http://ncia.nci.nih.gov

Paik DS, Beaulieu CF, Rubin G D et al (2004) Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT, IEEE Transactions on Medical Imaging, 23(6):661-675

Partain CL, Chan HP, Gelovani JG et al (2005); Biomedical Imaging Research Opportunities Workshop II: Report and Recommendations, Radiology, 236(2):389 – 403

Rubin GD (2000) Data explosion: the challenge of multidetector-row CT, European Journal Of Radiology 36(2):74-80

Sluimer I, Schilham A, Prokop M et al (2006) Computer analysis of computed tomography scans of the lung: a survey, IEEE Transactions on Medical Imaging, 25(4):385–405

Soysal ÖM and Chen J (2007) Matching Polygons Using Canonical String representation, IPCV’07 in WORLDCOMP’07, pp. 407-413.

Suzuki K, Armato SG, Li F et al (2003) Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography, Medical Physics, 30(7):1602-1617

Way TW, Hadjiiski LM, Sahiner B et al (2006) Computer-aided diagnosis of pulmonary nodules on CT scans: Segmentation and classification using 3D active contours, Medical Physics, 33(7): 2323-2337

Zhang X, McLennan G, Hoffman EA et al (2005) Automated Detection of Small-Size Pulmonary Nodules Based on Helical CT Images, Lecture Notes in Computer Science, Processing in Medical Imaging: 19th International Conference, 3565:664-76