



# A new computationally efficient CAD system for pulmonary nodule detection in CT imagery

Temesguen Messay<sup>a,\*</sup>, Russell C. Hardie<sup>a</sup>, Steven K. Rogers<sup>b</sup>

<sup>a</sup> Department of Electrical and Computer Engineering, University of Dayton, 300 College Park, Dayton, OH 45469-0232, United States

<sup>b</sup> Air Force Research Laboratory, AFRL/RW Wright Patterson AFB, OH 45433, United States

## ARTICLE INFO

### Article history:

Received 7 November 2008

Received in revised form 1 February 2010

Accepted 3 February 2010

Available online 19 February 2010

### Keywords:

Computer aided detection (CAD)

Lung nodule

Computed tomography (CT)

LIDC

ANODE09

## ABSTRACT

Early detection of lung nodules is extremely important for the diagnosis and clinical management of lung cancer. In this paper, a novel computer aided detection (CAD) system for the detection of pulmonary nodules in thoracic computed tomography (CT) imagery is presented. The paper describes the architecture of the CAD system and assesses its performance on a publicly available database to serve as a benchmark for future research efforts. Training and tuning of all modules in our CAD system is done using a *separate and independent* dataset provided courtesy of the University of Texas Medical Branch (UTMB). The publicly available testing dataset is that created by the Lung Image Database Consortium (LIDC). The LIDC data used here is comprised of 84 CT scans containing 143 nodules ranging from 3 to 30 mm in effective size that are manually segmented at least by one of the four radiologists. The CAD system uses a fully automated lung segmentation algorithm to define the boundaries of the lung regions. It combines intensity thresholding with morphological processing to detect and segment nodule candidates simultaneously. A set of 245 features is computed for each segmented nodule candidate. A sequential forward selection process is used to determine the optimum subset of features for two distinct classifiers, a Fisher Linear Discriminant (FLD) classifier and a quadratic classifier. A performance comparison between the two classifiers is presented, and based on this, the FLD classifier is selected for the CAD system. With an average of 517.5 nodule candidates per case/scan ( $517.5 \pm 72.9$ ), the proposed front-end detector/segmentor is able to detect 92.8% of all the nodules in the LIDC/testing dataset (based on merged ground truth). The mean overlap between the nodule regions delineated by three or more radiologists and the ones segmented by the proposed segmentation algorithm is approximately 63%. Overall, with a specificity of 3 false positives (FPs) per case/patient on average, the CAD system is able to correctly identify 80.4% of the nodules (115/143) using 40 selected features. A 7-fold cross-validation performance analysis using the LIDC database only shows CAD sensitivity of 82.66% with an average of 3 FPs per CT scan/case.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

According to statistics from the American Cancer Society, lung cancer is the primary cause of cancer related deaths in the United States (Cancer Facts and Figs, 2009). This makes lung cancer a major front in the war against cancer. Early detection of potentially cancerous pulmonary nodules may be a way to improve a patient's chances for survival (Henschke et al., 1999). Computed tomography (CT) is the most commonly used diagnosis technique for detecting small pulmonary nodules because of its sensitivity and its ability to represent/visualize a complete three-dimensional structure of the human thorax (Henschke et al., 1999; Hoffman and McLennan, 1997). However, each scan contains numerous

sectional images that must be evaluated by a radiologist in a potentially fatiguing process. As a result, the performance of a human interpreter can be adversely affected. It is therefore an important task to develop computer aided detection (CAD) systems that can aid/enhance radiologist workflow and potentially reduce false negative findings.

For more than a decade, significant effort has been focused on developing automated systems that detect/recognize suspicious lesions in thoracic CT imagery as well as other types of imagery. For example, in 1998 Okumura et al. (1998) have reported the use of CT screening devices and a filtering technique called "N-Quoit filter". Kanazawa et al. (1998) have described diagnosis rules based on fuzzy clustering. The following year, Hadjiiski et al. (1999) proposed a new hybrid classifier combining an unsupervised and a supervised model to improve classification performance of malignant and benign masses on mammograms (Hadjiiski et al., 1999). In 2001, Armato et al. (1999, 2001) suggested an improved system

\* Corresponding author. Tel.: +1 937 229 3611; fax: +1 937 229 4529.

E-mail addresses: [messayte@notes.udayton.edu](mailto:messayte@notes.udayton.edu) (T. Messay), [rhardie@udayton.edu](mailto:rhardie@udayton.edu) (R.C. Hardie), [srogers5@woh.rr.com](mailto:srogers5@woh.rr.com) (S.K. Rogers).

that combines 2-D and 3-D feature analysis and makes use of an automated classifier based on linear discriminant analysis to distinguish candidates that correspond to actual nodules from false-positive candidates. That same year, Lee et al. proposed an improved template-matching technique based on genetic algorithm template matching for detecting nodules existing within the lung area (Lee et al., 2001). To improve the detection performance of polyps in CT colonography, Nappi and Yoshida (2002) developed new gradient concentration and directional gradient concentration features for recognizing and reducing FP findings generated by their CAD system (Nappi and Yoshida, 2002). Rubin et al. (2005) presented a detection system that uses a surface normal overlap technique followed by a Lantern Transform to form a feature vector that is processed by a simple rule-based classifier to diminish false positive (FP) detections. Wei et al. (2005) consider various classifier approaches including a support vector machine, kernel Fisher discriminant, relevance vector machine, and committee machines (ensemble averaging and AdaBoost) for automated classification of clustered microcalcifications (Wei et al., 2005). Shiraishi et al. (2006) favored using a localized search method based on anatomical classification to detect nodule candidates in 2-D chest radiographs and employed a total of 71 image features for three sequential artificial neural networks in order to reduce the number of false-positive candidates. Recently, Gori et al. have proposed a dot-enhancement filter for nodule candidate selection and a neural classifier for FP reduction (Gori et al., 2007). Hardie et al. (2008) have also recently proposed a CAD system for identifying lung nodules in 2-D chest radiographs that consists of using a weighted mean convergence index detector and an adaptive distance-based threshold algorithm to segment the detected nodule candidates. A set of 114 features is computed for each candidate. This is followed by a classifier to reduce FPs. A Gaussian Bayes linear classifier, a Fisher linear discriminant (FLD) classifier, and a quadratic classifier are compared (Hardie et al., 2008). Many other groups have also recently presented systems and performance studies for detecting nodules (Wiemker et al., 2006; Das et al., 2006; Yuan et al., 2006; Gurung et al., 2006; Opfer and Wiemker, 2007; Sahiner et al., 2007; Buhmann et al., 2007).

In this paper, we introduce a new computationally efficient CAD prototype for the detection of pulmonary nodules in thoracic CT imagery. This work provides a complete description of the entire system and presents a performance analysis by virtue of multiple free-response receiver operating characteristic (FROC) curves. While several other CAD systems have been discussed in the literature (Okumura et al., 1998; Kanazawa et al., 1998; Hadjiiski et al., 1999; Armato et al., 1999, 2001; Lee et al., 2001; Nappi and Yoshida, 2002; Rubin et al., 2005; Wei et al., 2005; Wiemker et al., 2006; Das et al., 2006; Yuan et al., 2006; Gurung et al., 2006; Shiraishi et al., 2006; Gori et al., 2007; Opfer and Wiemker, 2007; Sahiner et al., 2007; Buhmann et al., 2007; Hardie et al., 2008), our approach has several novel aspects. We employ a fully automated algorithm for identifying and segmenting both lungs in the CT scans. Next, an expeditious method that combines intensity thresholding with morphological processing is used to detect and segment nodule candidates *simultaneously*. To the best of our knowledge, the way we combine multiple thresholds and morphological operations and efficiently merge candidates has not been previously reported. For each detected and segmented nodule candidate, 245 features are computed. The final detections of the recommended system are produced by employing a computationally simple FLD classifier (Fisher, 1936; Duda et al., 2000; Hardie et al., 2008; Anderson, 1996; McLachlan, 2004; Martinez and Kak, 2001) using a selected subset of features. Some of the features used here appear to be unique to our system for this application. Other significant contributions of this paper are a comparison of system performance using two different classifiers (FLD and quadratic

classifier, Fukunaga, 1990) and an overall performance assessment as a function of the number of selected features used by the corresponding classifier.

To perform a rigorous validation with our system, we use completely independent training and testing datasets. All modules in the CAD system are first tuned/trained using a dataset provided courtesy of the University of Texas Medical Branch (UTMB). The system is then evaluated using a publicly available database created by the Lung Image Database Consortium (LIDC) (Armato et al., 2004; McNitt-Gray et al., 2007; Reeves et al., 2007; Armato et al., 2007) to provide a realistic performance benchmark for future research efforts. We also present a 7-fold cross-validation performance analysis using just the LIDC collection. These results are included for reference and to serve as an easily reproducible benchmark requiring only the LIDC database. In a final test of the proposed CAD system, we have participated in a study called “Automatic Nodule Detection 2009” (ANODE09) (van Ginneken et al., 2009). With ANODE09, we are able to compare the performance of the proposed CAD system to that of several other CAD systems using a common database and common evaluation protocol (van Ginneken et al., 2009). The ANODE09 results also show the performance of the proposed CAD system for a variety of different types of nodules (van Ginneken et al., 2009).

The remainder of this paper is organized as follows. A thorough description of the materials used is presented in the next section. The architecture of the CAD system and its modules is described in Section 3. This includes the preprocessor, nodule candidate detection and segmentation, features and classifier. Experimental results are shown in Section 4. The results include overall performance results, performance comparisons with five other previously published CAD systems and the ANODE09 study results. In Section 5, we present conclusions and thoughts for future improvements and work.

## 2. Materials

### 2.1. LIDC

In the spirit of (Opfer and Wiemker, 2007; Sahiner et al., 2007), we have evaluated our CAD algorithm using the database created by the LIDC (Armato et al., 2004; McNitt-Gray et al., 2007; Reeves et al., 2007; Armato et al., 2007) that is publicly available in the National Biomedical Imaging Archive (NBIA). The LIDC data have been collected from five different sites in the United States. The LIDC data used in this paper are comprised of *all* 84 CT scans that were available via internet as of May 2008: LIDC case 13614193285030022 through LIDC case 13614193285030147 (note that the case numbers are not consecutive, but are in ascending order). Note also that in this paper, a ‘case’ refers to a CT scan of a single patient and the term ‘case’ will be used synonymously with ‘CT scan’. There are other LIDC cases in the NBIA that were restricted and those will be continuously added to the available data. Four expert chest radiologists drew outlines for nodules 3.00 mm or larger in effective size. The ground truth has been established in a blinded reading session followed by an unblinded one. In the first phase of the review, the four radiologists read and annotated the cases independently in a blinded fashion. The results of the first reading were compiled and presented to the radiologists for a second unblinded reading. Each radiologist reexamined the cases with the benefit of information as to what the others marked. No consensus was enforced during the final decisions about the markings.

There is some dissent among radiologists about many of the nodule findings after the second unblinded reading session (Armato et al., 2007). In our study, we merged the truth lists of all four radiologists so as to include all suspicious structures to

serve as truth against which we score our CAD system. That is, we designate a location to be a nodule if it is identified at least by one of the four radiologists. This has yielded 143 nodules in the range of 3–30 mm in size. Note that we deduce nodule size from LIDC by measuring the maximum diameter in the provided radiologist contour/segmentation of each nodule. Where multiple radiologists provide a segmentation of the same nodule, the maximum diameters are averaged. Using this method, the mean nodule size is  $9.87 \pm 4.53$  mm.

Approximately 26% of the 143 nodules are juxtales nodules. Each CT scan contains between 0 and 11 nodules, has a slice thickness between 1.3 and 3.0 mm, and a reconstruction interval (slice spacing) ranging from 0.75 to 3.0 mm. Each slice has a matrix size of  $512 \times 512$  pixels, and a pixel size ranging from 0.5 to 0.8 mm with a 12-bit gray scale resolution in Hounsfield Units (HU). Note that different scanners from different vendors as well as different scanning protocols have been used for image acquisition. The contents of the LIDC collection in its entirety (i.e., 84 CT scans, including cases without nodules) are used for testing purposes. More information about the LIDC collection can be found in (Armato et al., 2004; Opfer and Wiemker, 2007; Sahiner et al., 2007; Wang et al., 2007; McNitt-Gray et al., 2007; Reeves et al., 2007; Armato et al., 2007).

## 2.2. UTMB

All aspects of CAD algorithm training and tuning has been done using an independent dataset provided by the UTMB. This dataset includes 90 chest CT examinations acquired using a 4-detector LightSpeed QX/i scanner (GE Medical Systems, Milwaukee, WI). Images have been obtained using the HQ setting with 5.00 mm collimation, helical pitch of 0.75/1.0, 120 kVp, and 160–270 mA. Images were then contiguously reconstructed with a slice thickness and slice spacing of 2.5 mm. With the insight of two board-certified radiologists, we identified 606 nodule exemplars with a size range of 3–30 mm. A cue point is placed near the centroid of each nodule. The mean nodule size is estimated to be approximately  $7.6 \pm 3$  mm. Since radiologist segmentations are not available for the UTMB data, we estimated the sizes using segmentations provided by our CAD system's front-end detector/segmentor described in Section 3.2. Each CT scan/case may contain between 0 and 112 annotated nodules. Note that approximately 23% of the 606 annotated nodules are juxtales. The UTMB dataset contains a wide variety of nodule type (e.g., pleura attached, vessel attached and solitary, calcified, regular and irregular nodules). All 90 cases, including cases without annotated nodules, are used for training purposes.

Unlike the cases of the LIDC dataset, that are acquired using different scanners from different vendors as well as different scanning protocols and different reconstruction methods, the UTMB CT scans are acquired using the same scanner and a consistent scanning protocol and reconstruction method. Moreover, although the UTMB training dataset has a greater number of total nodules, the distribution of effective size and the fraction of juxtales nodules are comparable to that of the LIDC testing dataset. Note that the size of the LIDC nodules estimated using the same method as that for UTMB is  $8.5 \pm 4.4$  mm. However, after thoroughly examining the nodules in each dataset, the LIDC database appears to contain a greater number of subtle and irregularly shaped nodules compared to those of the UTMB dataset.

## 3. CAD system

In this section, we describe the proposed CAD system. Fig. 1 shows the top level block diagram of the CAD system. The raw

CT scans are first preprocessed; that includes orienting, down-sampling, performing local contrast enhancement (LCE) and segmenting the lungs. Next, using the down-sampled/oriented CT scans and the generated lung mask, nodule candidates are detected and segmented simultaneously. Features for all of the segmented nodule candidates are then computed based on the information of the generated nodule candidates mask, the down-sampled/oriented slices and the enhanced version of those slices. Finally, a classifier is employed on the computed features (i.e., feature matrix) to reduce the FPs within the nodule candidates mask to produce the final output detections mask.

### 3.1. Preprocessor

#### 3.1.1. Orienting and down-sampling

The first module of the preprocessing stage of the CAD system is designed to consistently orient and down-sample the input CT data. Once properly oriented, the sectional images are down-sampled by an appropriate integer factor to generate CT slices with a comparable slice spacing to that of the training data (i.e., 2.5 mm). The down-sampling process is used not only to provide approximate resolution compatibility between the testing and training data, but to also significantly improve the processing speed and provide noise reduction (Shiraishi et al., 2006). The re-sampling is done by simply averaging the appropriate number of image slices (Gonzalez and Woods, 2007). Each LIDC case is down-sampled by an appropriate integer for that case to get as close as possible to 2.5 mm (without going under). As described in Section 2.1, the slice interval range of the original LIDC dataset is 0.75–3.00 mm. The slice interval range of the re-sampled/down-sampled dataset is 2.00–3.60 mm. The subsequent modules of the preprocessing stage are local contrast enhancement and anatomical segmentation.

#### 3.1.2. Local contrast enhancement

To improve the details and local context of lung nodule candidates, we perform LCE within each down-sampled sectional image of a given case. The proposed LCE process closely follows the one explained in (Schilham et al., 2003; Schilham et al., 2006; Hardie et al., 2008) and is to be performed within each slice image to obtain two different enhanced images. The 2-D LCE operation subtracts from each pixel an estimate of the local mean and then divides this by an estimate of the local standard deviation. This is denoted by

$$o(x,y) = \frac{i(x,y) - \mu(x,y)}{\sigma(x,y)}, \quad (1)$$

where  $i(x,y)$  represents the input image value at coordinates  $x, y$ , and  $o(x,y)$  is the output image. The local mean estimate is given by

$$\mu(x,y) = i(x,y) * h(x,y), \quad (2)$$

and the local standard deviation estimate is given by

$$\sigma(x,y) = \sqrt{i^2(x,y) * h(x,y) - \mu^2(x,y)}, \quad (3)$$

where  $*$  represents 2-D convolution. The choice of the kernel  $h(x,y)$  impacts the type and scale of the objects enhanced by this operation. To enhance both small and medium sized nodules, we generate two LCE images using two different Gaussian kernels with a standard deviation of 16. LCE1 uses an  $11 \times 11$  kernel size (for smaller nodules) and LCE2 uses a  $51 \times 51$  kernel (for medium and large nodules). Fig. 2 shows the two LCE images for two lung nodules of different effective sizes. The local contrast images are used to compute several novel features that have proven to be beneficial to our CAD system.

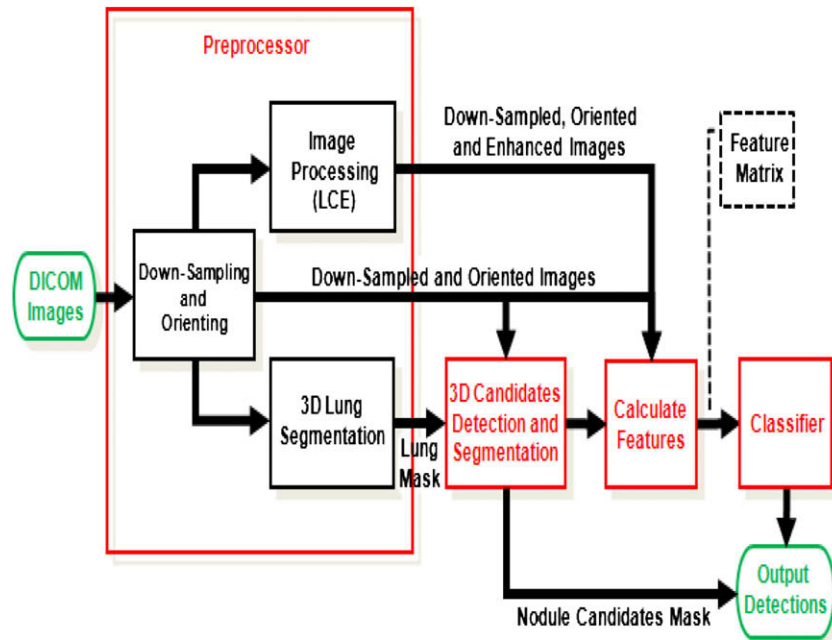


Fig. 1. Top level block diagram of CAD system (LCE: Local Contrast Enhancement).

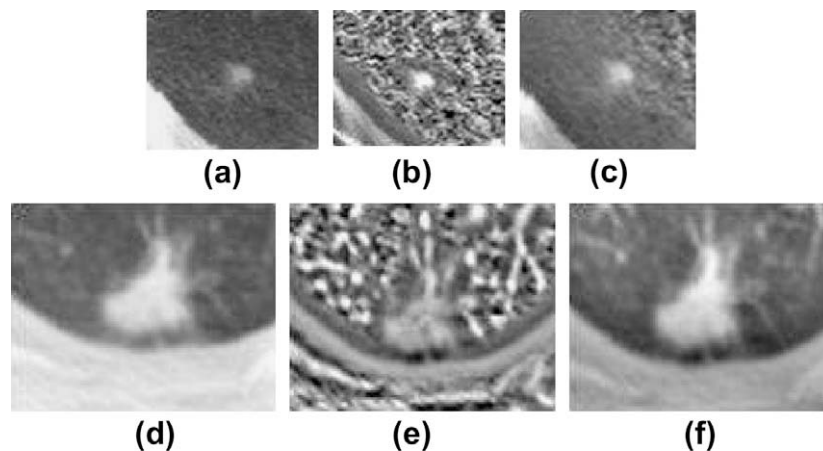


Fig. 2. (a) Example of an original slice image of a small nodule (3.92 mm), (b) and (c) show its first and second LCE images, respectively, (d) is an example of an original slice image of a medium sized nodule (13.94 mm), (e) and (f) are its corresponding LCE images (LCE: Local Contrast Enhancement).

### 3.1.3. 3-D lung segmentation

Another essential preprocessing step is to identify the boundaries of the lungs within the CT images. This is a very crucial process because it restricts the post-processes and subsequent final output detections of the system to the lung regions, preventing FPs outside the lung field. Our system includes a fully automated method for segmenting and identifying both lungs in CT images. Fig. 3 shows a block diagram of the automated lung segmentation algorithm. The general scheme of the lung segmentation algorithm is similar to that described in (Hu et al., 2001; Armato et al., 2001; Leader et al., 2003). It consists of *three* major conventional steps: (1) an extraction of the preliminary lung regions using thresholding, (2) an identification of the lungs followed by a separation of the right and left lungs if needed and (3) an application of a segmentation refinement process (Hu et al., 2001; Armato et al., 2001; Leader et al., 2003). For a down-sampled CT scan, we start off by assuming the presence of two full lungs surrounded by the body and the CT data is in HU.

The first step of the recommended segmentation algorithm is similar to the one explained in (Hu et al., 2001; Armato et al., 2001). Since intensity in CT is proportional to tissue density, it involves the use of intensity thresholding to first isolate the body and then to extract the lung region within the segmentation of the patient's body. Here, unlike (Hu et al., 2001; Armato et al., 2001), fixed threshold values in combination with some morphological and topological analysis are used to produce a preliminary segmentation of the lung region. A threshold of  $-500$  HU, which corresponds to a density between 'water' and 'air', is used to first segment the body. 2-D morphological and topological analysis, similar to (Reinhardt and Higgins, 1998), are thereafter used to remove the CT table (and other non-body) artifacts and to fill the resultant binary image of the body (i.e., a 2-D morphological opening and a 2-D morphological flood fill operation is performed on each slice/section). Fig. 4a depicts an example of a body mask superimposed on the corresponding CT slice image of a healthy person. Then, a threshold of  $-600$  is applied to the CT voxels lo-



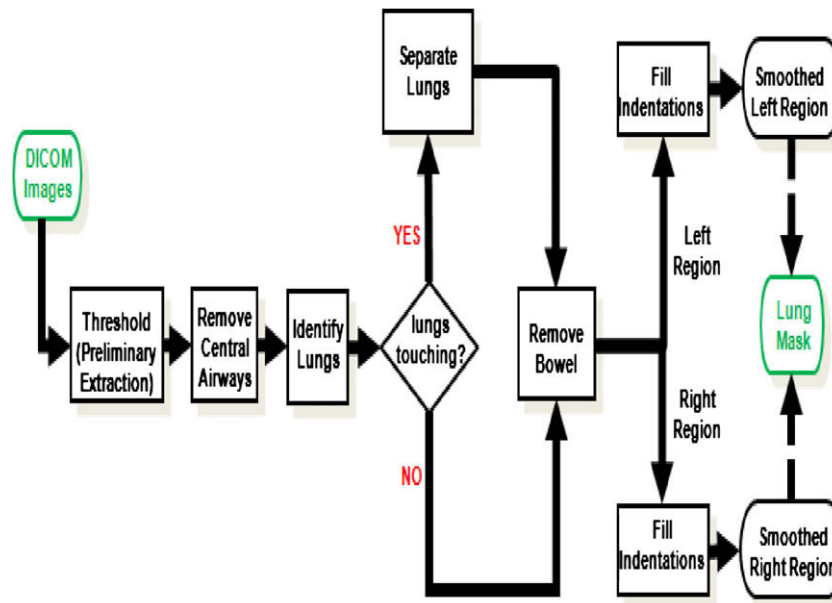


Fig. 3. Top level block diagram of automated lung segmentation scheme.

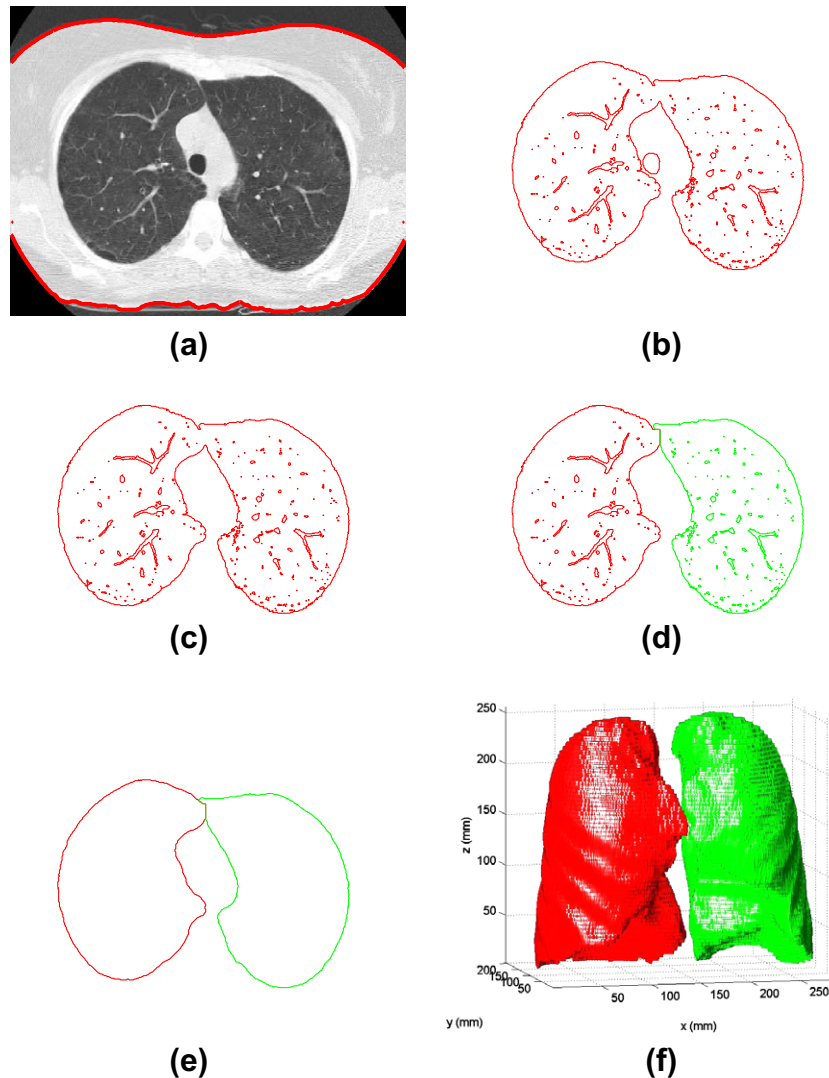
cated inside the body to generate a preliminary lung mask like that illustrated in Fig. 4b. The initial lung mask requires further refinement (e.g., removal of cavities/holes, removal of unwanted components and filling indentations along the perimeter, etc.). Moreover, the back-end processor of our CAD system requires the explicit separation of the right and left regions of the lung for the computation of some geometric features. The remaining two steps are intended to refine the initial lung segmentation so as to address all other related issues to meet the requirements of forthcoming processes of the system.

A pre-requisite to the *second* step is the removal of the central airways from the obtained initial lung mask. Our approach is similar to (Sonka et al., 1994) and calls for a rule-based analysis founded on the knowledge of some anatomical characteristics of the central airways. The distributions over the area and approximate location (with respect to the box bounding both lung regions) of the central airways for all the slices/sections in the training dataset (i.e., UTMB dataset) are examined. The maximum distance in the  $x$ - $y$  plane of any given airway voxel from the center of the lung regions is found not to exceed 41.33 mm. Furthermore, the maximum area of any airway cross-section is found not to exceed 602.35 mm<sup>2</sup>. Accordingly, a search window (100 × 100 mm that encompasses all voxels within 50 mm from the center) centered about the lung region is defined for each slice/section of the lung mask and any object that is sufficiently small (i.e. smaller than 625 mm<sup>2</sup>), bright (i.e., ON-voxels) and not overlapping with the perimeter of the defined central region is removed. After the removal of the central airways, the identification and separation process is initiated. The lungs are identified by employing a 3-D connected components labeling (6-connectivity voxels) and preserving the largest and the second largest (if present) components in the mask volume (Hu et al., 2001). The resultant mask is portrayed in Fig. 4c. As it is in this case, the left and right lungs are fused and hence a left-right lung separation is required.

The lung separation process is applied only to cases that are believed to contain a lung candidate that exceeds 300 mm in maximum width (Simon et al., 1972). A search window (512 × 100 mm) centered about the lung region of each slice of the lung mask is first defined. Our method consists of considering the “non-lung” elements of each slice within the defined search region as a means of generating the anterior and posterior junction

lines. The algorithm identifies 2-D non-lung objects that are large enough or overlapping with the top or/and bottom of the search region to produce a “dividing object”. Then, it initializes the split position by locating the tallest column of such dividing object closest to the middle of the pair. For the actual splitting process, a raster scan technique is brought into play (Leader et al., 2003). The defined region within each section of the lung mask is scanned row by row from top to bottom. During the scan, the breaking column of each row is determined by finding the column position of the voxel belonging to the dividing object closest to the last split point. If there are not any voxels belonging to the dividing object in a specific row, the algorithm uses the last split column position. Fig. 4d illustrates the separated lung mask. Before embarking on the third and final step (i.e., the refinement process), the bowel is removed by discarding attached bowel objects from the patient’s right-side lung. That is accomplished by once again invoking a rule-based analysis similar to the one used to remove the central airways. This time around, the volume of interest (the search region) is the bottom region of the right lung shortly before it ends (i.e., few slices before the last slice/section for the described orientation). Objects that grow in area (or move) from slice to slice by more than a pre-determined amount are removed (Sonka et al., 1994; Brown et al., 1997).

The *third* step improves the overall results by filling the indentations along the perimeter of the segmentation that tend to exclude juxta-pleural nodules and radiodense pulmonary arteries/veins located near or belonging to the mediastinum. A 2-D rolling ball filter is employed within each section of the lung mask to compensate for the indentations along the contour lines of each lung (Armato et al., 2001, 1999). A ball of radius 12.5 mm, which conforms to most of normal anatomical structures (e.g., aorta, mediastinum structures, etc.), is tangentially placed along the outside border of each contour line. Indentations are identified if the ball contacts the perimeter at more than one location. If so, a linear path is interpolated between the two endpoints of the indentation (Armato et al., 2001, 1999). The output mask of the lung segmentation algorithm is finally obtained after a morphological flood filling operation to remove all remaining holes (Reinhardt and Higgins, 1998). Fig. 4e shows a 2-D cross-section of the final output lung mask and Fig. 4f presents a 3-D rendered view of the complete lung segmentation. Note that *all* parameters of the lung segmenta-



**Fig. 4.** Illustration of the steps for the lung segmentation algorithm: (a) example of a body mask of a healthy person (window width, 1500 HU; window level, -600 HU), (b) example of a preliminary lung mask, (c) an image of a lung mask after removal of central airways, (d) a picture of the separated lung regions, (e) and (f) final output of the lung segmentation algorithm.

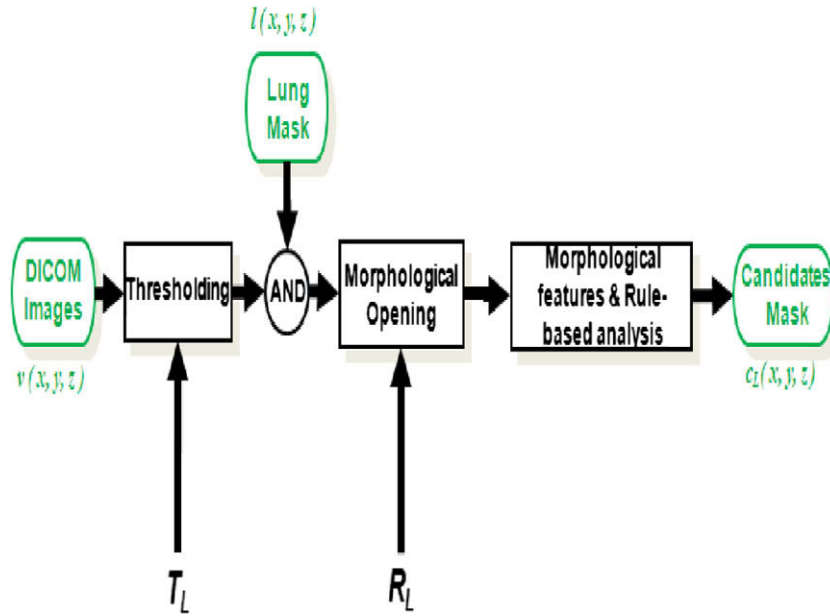
tion algorithm were specified while working with the UTMB data. On average, the proposed lung segmentation algorithm consumed 40–45 s per CT scan on an HP Personal Computer (PC) with a processor speed of 2.6 GHz. Although, the proposed algorithm has provided useful segmentation results for all of the LIDC cases, the proposed automated method may fail in some cases, particularly those with unusual pathology or artifacts. In such cases, manual modification of tuning parameters can usually resolve the problem.

### 3.2. 3-D nodule candidate detection and segmentation

After preprocessing, a nodule candidate detection and segmentation algorithm is employed to locate potential lung nodules. Unlike methods such as the one described in (Hardie et al., 2008) that first detects potential nodule candidates and then employs a separate algorithm to segment those candidates, here, the nodule candidates are detected and segmented simultaneously. Like the method in (Armato et al., 2001), we use multiple gray level thresholding. However, here we pair each threshold operation with a specific morphological opening operation to produce a total of 15 intermediate candidate masks. Each intermediate candidate mask results from one threshold and opening operation. A simple size

and compactness based expert filter is used to remove many unwanted intermediate candidates. Next, we logically OR the 15 intermediate masks to produce the final 3-D nodule candidates mask. We have found that this logical ORing process has a number of benefits. First, we do not need to treat the complex case of having multiple overlapping candidates. Second, this greatly reduces the computational load of the subsequent feature calculations and classifier by reducing the number of total candidates. Finally, the ORing guarantees the largest candidate “footprint” from which to derive features. By carefully choosing and pairing the thresholds and structuring element sizes and by selecting the appropriate expert filter parameters, we have found that this method is capable of producing very good detection sensitivity and relatively good segmentation overlap scores with radiologist segmentations.

To better describe the details of the nodule candidates detection/segmentation algorithm, let us begin by considering the process involved at a single level. Fig. 5 shows a block diagram of the detection and segmentation engine for a single level. Each level is a two-parameter system that produces an intermediate candidate output mask. The two parameters which must be selected are the intensity threshold and the radius of a disk-shaped morphological structuring element. At a given level  $L$ , a threshold value



**Fig. 5.** Block diagram of the detection and segmentation engine for a given single-level  $L$ , a threshold value  $T_L$ , CT voxels  $u(x,y,z)$ , lung field  $l(x,y,z)$  and a circular structuring element of radius value  $R_L$ . An intermediate nodule candidates mask  $c_L(x,y,z)$  is produced.

$T_L$  is first applied to the CT voxels  $u(x,y,z)$ . The obtained 3-D binary data (3-D mask) is then logically ANDed with the lung field  $l(x,y,z)$  to prevent nodule candidate detections outside the defined lung field. A 2-D morphological opening is then performed in each section of the binary data using a circular structuring element of radius value  $R_L$  to remove residual structures such as vessels which may be attached to the detected and segmented nodule candidates (Serra, 1983). A 2-D rather than a 3-D morphological operation is favored here because of the resolution (i.e., down-sampled to fairly thick-slices/sections) and the anisotropic nature of the data. In general, 2-D operations have been found to outperform 3-D operations for relatively thick-slice data and have also shown to greatly reduce computational complexity. The corresponding intermediate nodule candidates mask  $c_L(x,y,z)$  is produced after inspecting the size and compactness of each nodule candidate segmentation. The inspection consists of first evaluating several morphological features. These features include the following:

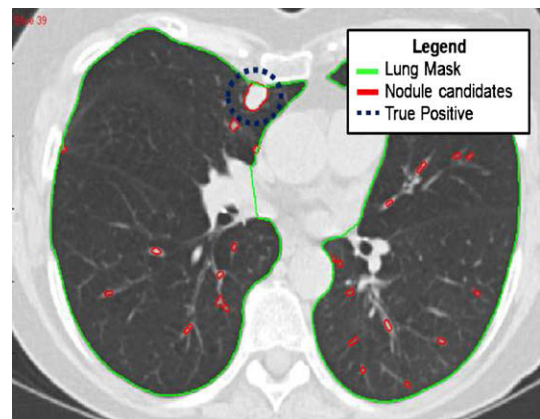
- **Bounding box dimensions:** The width of all 3 dimensions in mm of the 3-D maximum bounding box of the segmentation.
- **Area:** Number of voxels encompassed by the segmentation at the maximum area slice multiplied by the pixel spacing squared.
- **Volume:** Total count of voxels encompassed by the entire segmentation multiplied by both, the pixel spacing squared and the slice spacing.
- **Circularity:** Computed Area divided by  $\pi r^2$ , where  $r$  is the maximum bounding box radius in the maximum area slice.
- **Sphericity:** Computed Volume divided by  $(4/3)\pi r^3$ .

Based on these measurements, a simple rule-based analysis was then used to decide whether to keep or discard the candidate object in question. In order for a segmented object to be admitted as a “good” nodule candidate segmentation, it must meet all of the following basic size and compactness requirements. The computed Area may not exceed the range 4.53–962.11 mm<sup>2</sup> (equivalent to a circular diameter range of 2.4–35 mm). The computed Volume must lie in the range 7.24–2245 mm<sup>3</sup>. The dimensions of the bounding box in the  $x$ - $y$  plane must not exceed 35 mm. The dimension of the bounding box in the  $z$ -axis must be less than 40 mm. Here, the

size criteria are specified so as to identify objects that are too large or too small to be considered. Furthermore, since a lung nodule is assumed to be spherical in 3-D and consequently circular in a 2-D cross-sectional image (Strickland, 2002), a minimum *sphericity* of 0.4 and a minimum *circularity* of 0.15 are required. This concludes the process of a single level.

The fifteen threshold (in HU) and radius (in pixels) combinations used for each level ( $T_L, R_L$ ) are: (−800,1), (−700,1), (−600,1), (−500,1), (−400,1), (−300,1), (−200,1), (−600,2), (−500,2), (−400,2), (−300,2), (−200,2), (−600,3) and (−300,3). The threshold and radii values were determined by examining detection and segmentation performance results for the annotated nodules in the UTMB (training) dataset. The range of thresholds and structuring element sizes are needed to treat the wide range of nodule intensities and varying level of vessel attachment. As mentioned above, we combine the output of the different levels with a logical ORing operation to generate a final nodule candidates output mask (i.e.,  $c(x,y,z)$ ).

A typical output of the proposed detection and segmentation system superimposed on the corresponding CT slice image is presented in Fig. 6. A closer look at some segmented nodules using



**Fig. 6.** Example of a nodule candidates detector output (i.e.,  $c(x,y,z)$ ) superimposed on the corresponding CT slice image.

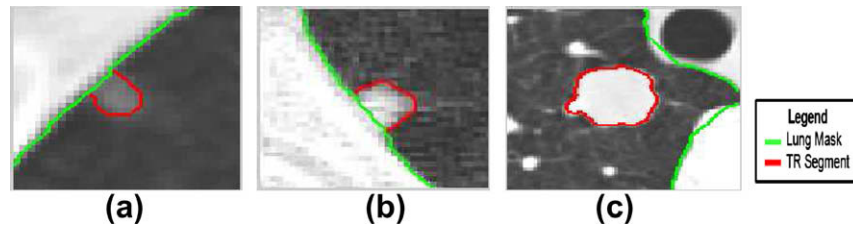


Fig. 7. Examples of some segmented nodules (a) very subtle, (b) relatively obvious, (c) conspicuous (based on 'merged' nodule subtlety rating).

the current system is shown in Fig. 7. We believe that the current approach is a good choice here because it has proven to be efficient and effective in detecting and segmenting nodule candidates. However the various tuning parameters have been selected here based on the nodules in the UTMB data. Consequently, as it will be discussed in Section 4.2, these tuning parameters may not be ideal for some of the irregularly shaped nodules in the LIDC dataset.

### 3.3. 3-D nodule candidate features

For each detected and segmented nodule candidate, 2-D and 3-D features listed in Table 1 are computed. A total of 245 features are evaluated. For each case (i.e., 2-D and 3-D), we have computed the following three types of features: geometric, intensity and gradient features. The geometric features are computed based on the shape and position information provided by the final nodule candidates detection/segmentation mask and the lung mask. All intensity and gradient features are computed for the down-sampled CT data in HU and for the two LCE images using the boundaries defined by the final nodule candidates mask.

All of the 2-D features that are presented in (Hardie et al., 2008) (i.e., geometric, intensity and three types of gradient features) are evaluated at the largest area slice of the nodule candidate segmentation. In addition to the 2-D geometric features presented in (Hardie et al., 2008), we have also included features such as *X-Fraction 2* and *Distance To Center of Current Lung*. The horizontal Euclidean distances from a reference point of the lung field is computed. It is then divided by the width of the box bounding both lungs to produce *X-Fraction 2*. The feature *Distance To Center of Current Lung* is the Euclidean distance between the center of the box bounding both lungs and the nodule candidate at the largest area slice. These two features are intended to incorporate the location of the nodule candidate with respect to the Mediastinum. For 2-D features, the *outside* statistics are computed by enlarging the original segmentation at the maximum area slice. The original segmentation at the largest area slice is dilated using a circular structuring element of radius 5 pixels to obtain the "outside" region.

For the 3-D case, the geometric features warrant explanation. *Elongation* is a feature that describes the ratio between the longest 3-D dimension of the nodule candidate segmentation to the smallest 3-D dimension. *Projection Compactness* is estimated by first projecting the 3-D candidate segmentation onto the  $x$ - $y$  plane. Then, the area of the projected segmentation is divided by the area of its 2-D maximum bounding box to produce its *Projection Compactness*. *Cube Compactness* is computed by dividing the volume of the 3-D maximum bounding box by  $d^3$ , where  $d$  is the longest dimension of the 3-D bounding box. *Square Compactness Mean* and *Square Compactness Standard Deviation* are estimated by computing the square compactness of the segmentation at each relevant CT section. Similarly, *Area Difference* refers to the average change in area across the relevant slices of the segmentation. *Juxtapleural* is a logical feature that indicates if the nodule candidate is touching the lung wall or otherwise. *Fraction Touching Lung* is an extension of

the *juxtapleural* feature and it represents the amount of contact between the two bodies. It is calculated by dividing the number of voxels touching the lung mask with the number of voxels along the entire surface of the segmentation. As one can notice, most of the 3-D geometric features are compactness estimates. This is due to the fact that, as previously mentioned, lung nodules are generally compact ellipsoidal structures with well-formed segmentations (Strickland, 2002).

The 3-D intensity and gradient features are a natural extension of the 2-D features presented in (Hardie et al., 2008). For the 3-D case, we have decided to look at four different 3-D outer regions. The according structuring elements used to dilate the original 3-D segmentation are:

- *Outside 1* is obtained using a spheroidal structuring element that has a radius of 5 pixels in the  $x$ - $y$  plane and that extends by a single slice in each direction of the  $z$ -axis.
- *Outside 2* is generated using a similar spheroidal structuring element as *Outside 1* but with a radius of 10 pixels.
- *Outside 3* is produced by dilating each relevant section of the segmentation using a circular structuring element having a radius of 10 pixels.
- *Outside Z* is produced by using a structuring element that corresponds to a single pixel in the  $x$ - $y$  plane (i.e., radius = 0) but extends by a single slice in each direction of the  $z$ -axis.

Furthermore, for the 3-D case, we have also looked at the intensity statistics of the voxels located *Above* and *Below* of each segmented voxels. The intensity information of these features could potentially allow the system to identify dense tissue (or possibly bone) above or below a candidate. Note that we have computed the same group of statistics (*Inside*, *Outside*, *Contrast* and *Separation*) (Hardie et al., 2008) for all intensity and gradient features (2-D and 3-D features). Those type of statistics, in particular the ones derived with respect to several 3-D outer regions, allow us to better evaluate contrast between the interior of a nodule candidate and its background and thus are believed to provide useful local context for the nodule candidate. For example, if a nodule candidate has strong intensity/gradient statistics (strong *Inside Statistics* such as 'Mean Inside', 'Radial-Gradient Mean Inside', etc.) and if the background of that nodule candidate has also equivalently strong intensity/gradient properties (strong *Outside Statistics* such as 'Mean Outside', 'Radial-Gradient Mean Outside', etc.), this could suggest that the nodule candidate is less likely to be a True-Positive (TP).

Note that before classification, the size of each of the segmented nodule candidate is reinspected. At this stage, the system takes advantage of some of the computed features. Based on the relative size derived from four features (*Size*, *Area*, *Perimeter* and *Volume*), a simple threshold is applied to exclude nodule candidates that are outside the range of the system's nodule size criterion (i.e., 3–30 mm). Note that none of the nodule candidates (within both datasets) that correspond to true nodules (i.e., deemed TPs by the labeling and scoring technique described in Section 4.1)



**Table 1**

List of selected features for the CAD system. The features are computed based on the binary segmentation image (TR), the CT data in HU (CT) and the LCE images (LCE1 and LCE2). Of the full 245 set of computed features, the 60 features selected by the FLD classifier are marked with an (F) and the 29 features selected by the quadratic classifier are marked with a (Q).

	Slice image			
	TR	CT	LCE1	LCE2
<i>2-D geometric features</i>				
Size	F			
Eccentricity	F			
Circularity 1	F			
Circularity 2	F			
X-fraction 2	F			
Distance to center of current lung	F			
<i>3-D geometric features</i>				
Projection compactness	FQ			
Square compactness standard deviation	F			
Area difference	F			
Juxtapleural	Q			
Fraction touching lung	F			
Distance to center of projected lung	Q			
<i>2-D intensity features</i>				
Minimum value inside		F		
Mean inside			F	
Mean outside			F	
Contrast			Q	
Standard deviation inside		F		
Standard deviation outside			Q	
Standard deviation separation			F	
Skew inside		Q		
Kurtosis inside		F		
Moment 2 inside		Q		
Moment 3 inside			Q	
Moment 5 inside		Q		
Moment 6 inside		F		
Moment 7 inside		Q		
<i>3-D intensity features</i>				
Minimum value above		F	FQ	
Maximum value above			F	F
Minimum value below			F	
Mean above			F	
Mean below			F	
Contrast 1		F		F
Contrast Z		F	F	
Mean separation Z		F	F	
Standard deviation inside		F		
Standard deviation above		Q		
Standard deviation below			FQ	F
Standard deviation outside 1			Q	
Standard deviation outside Z		Q		
Standard deviation separation 1		FQ		
Standard deviation separation 3		Q		
Standard deviation separation Z		F		
<i>2-D gradient features</i>				
XY gradient magnitude separation inside			F	F
Radial-deviation mean inside		F	F	
Radial-deviation mean outside				F
Radial-deviation mean separation			F	
Radial-deviation standard deviation inside		F	Q	
Radial-deviation standard deviation separation			F	
Radial-deviation mean perimeter inside				F
Radial-gradient mean inside		F	F	
Radial-gradient standard deviation inside			F	
Radial-gradient standard deviation outside			F	
Radial-gradient mean perimeter inside			Q	
Radial-gradient mean perimeter outside			F	
Radial-gradient standard deviation perimeter inside			Q	
Radial-gradient standard deviation perimeter outside		FQ	Q	F
Radial-gradient standard deviation perimeter separation		Q		
<i>3-D gradient features</i>				
Gradient magnitude standard deviation inside				F
Radial-deviation mean outside 1		FQ	F	
Radial-deviation mean outside 2				Q
Radial-deviation standard deviation inside			F	
Radial-deviation standard deviation outside 1		FQ	F	F
Radial-deviation standard deviation outside 2		Q		
Radial-gradient standard deviation separation 1		F		

Table 1 (continued)

	Slice image			
	TR	CT	LCE1	LCE2
Radial-gradient mean surface inside		F		Q
Radial-gradient mean surface separation 1			Q	

disappear when the nodule size criterion is imposed. A cross-correlation analysis (Papoulis and Pillai, 2002) is thereafter performed by evaluating the correlation coefficient matrix of the feature matrix for all of the remaining nodule candidates. Redundant feature vectors (i.e., linearly dependent: correlated by a factor of 1.00) are discarded. The 245 features computed for each nodule candidate are therefore linearly independent (Rodgers et al., 1984). Note that although some of the 245 features might be highly correlated (but linearly independent), they are believed to contain different and hence relevant information about the nodule candidate at hand that might be useful to the subsequent processes of the system (particularly, as those features interact with one another).

### 3.4. Feature selection and classifiers

A Sequential Forward Selection (SFS) (Fukunaga, 1990) technique is used to determine a subset of the 245 candidate features that is to be used for the classifier in the CAD system. The objective function of a SFS process is based on the area under the FROC curve for a chosen classifier. Starting from an empty set, the feature that results in the highest area under the FROC curve when combined with the features that have already been selected is sequentially added to the set. A FROC curve in our case is defined as the fraction of TP nodule candidates passed versus the average number of FPs per case (Maxion and Roberts, 2004). The curve is generated exclusively with the UTMB training dataset using a 10-fold cross-validation process. The validation process is as follows. First, the training dataset is broken up into 10 groups (by case index) each containing 9 cases (90 cases total). The classifier posterior probabilities for the nodule candidates of each group are then determined using a classifier that is trained on the remaining 9 groups. Similarly, the classifier posterior probabilities for the nodules candidates of the next group are determined using a classifier that is trained on the remaining 9 groups. The particular group is cycled through in this fashion until posterior probabilities are computed for all nodule candidates within the dataset. After assigning a posterior probability value for *all* the nodule candidates in that fashion, it is possible to generate a FROC curve by simply thresholding *all* those posterior probabilities and scoring. Note that a range of threshold values is applied to all posterior probabilities with the intent of generating numerous operating points with approximately uniform spacing for the average number of FPs per case (i.e., horizontal axis). These points form the FROC curve. Note that the plotting function employed displays a straight line between computed operating points (i.e., linear interpolation).

We have considered two distinct classifiers and performed an SFS for each. The two classifiers are a FLD classifier (Fisher, 1936; Duda et al., 2000; Hardie et al., 2008; Anderson, 1996; McLachlan, 2004; Martinez and Kak, 2001) and a quadratic classifier (Fukunaga, 1990). Many of the computed features appear to be very salient. Table 1 lists the features selected using a SFS process for each classifier. Fig. 8 depicts the SFS objective/merit function. The figure shows the area under the according FROC curve from 0 to 10 FPs per CT scan as a function of the number of selected features. This range of FPs has been selected because we believe it includes suitable operating points for clinical purposes. The maximum number

of selected features for each classifier is obtained after examining the slope of the line tangent to the according curve shown in Fig. 8. That is, when the derivative of the SFS objective/merit function with respect to the number of selected features is *sufficiently* small (in our case, less than 0.0024). Accordingly, for the FLD classifier, 60 features are selected and these are marked in Table 1 by an (F). A cross-correlation analysis of all 60 selected features is performed and the maximum correlation coefficient between any two selected features is found not to exceed 0.87 (mean value of  $0.4969 \pm 0.2335$ ). For the quadratic classifier, 29 features are selected and these are marked by a (Q). In turn, the maximum correlation coefficient between any two features selected by the quadratic classifier is found not to exceed 0.77 (mean value of  $0.4841 \pm 0.2194$ ). The number of features indicated here is chosen so as to maximize the objective function of the SFS. However, an elaborate study of the performance of each classifier as a function of the number of the selected features will be discussed in Section 4.3.

The design of the FLD classifier for this application is identical to the one explained in (Hardie et al., 2008). There, it is also reported that such a classifier presented a surprisingly good performance for a similar application due its ability to generalize well, even with a relatively small number of TP training samples (Hardie et al., 2008). For a FLD classifier, the objective function is evaluated by *adding* the class conditional covariance estimates (Fisher, 1936; Duda et al., 2000; Hardie et al., 2008; Anderson, 1996; McLachlan, 2004; Martinez and Kak, 2001). The mixture of the covariance estimates allows the FLD to operate robustly with the relatively few training examples. In contrast, the objective function of a quadratic classifier is evaluated based on both TP and FP class conditional covariance estimates *independently* (Fukunaga, 1990). Thus, the large discrepancy in the number of samples of each class can potentially harm the performance of a quadratic classifier based CAD system. Furthermore, our experimental results presented in

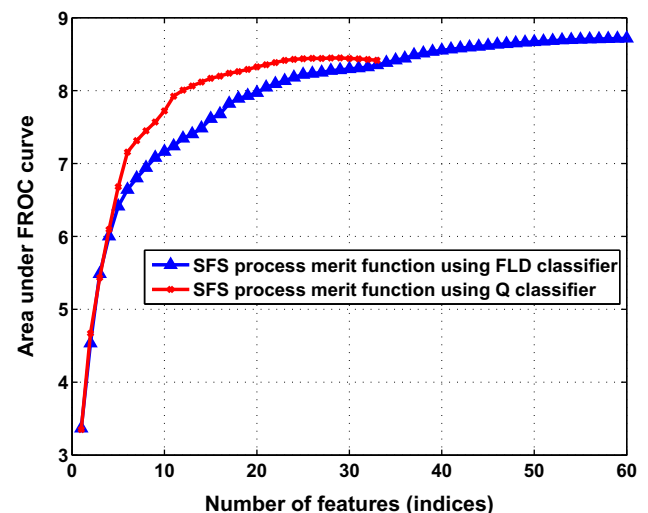


Fig. 8. SFS merit/objective function with increasing numbers of features.

Section 4.3 are in agreement with the results obtained by Hardie et al. (2008) in the sense that a FLD classifier generally outperforms a quadratic classifier. For these reasons, a FLD classifier is recommended for the back-end processor of the CAD system.

#### 4. Experimental results

In this section, we present a number of results to demonstrate the effectiveness of the new CAD system. We first discuss the labeling and scoring scheme we employ. Next, we present a performance analysis of the nodule candidate detector. We then elaborate on the overall performance results of the new CAD system. The performance of the proposed CAD system is then compared to the performance of five other reported detection systems. That is followed by the results from the ANODE09 study. Finally, we present conclusions and thoughts for future improvements and work.

##### 4.1. Labeling and scoring

In this subsection, we explain how the detected candidates of each dataset are labeled using the provided truth information. Recall that the truth information of the training dataset (UTMB) includes cue points near the center of each nodule. For the testing dataset (LIDC), at least one manual segmentation for each nodule in the merged truth is available (with no cue points) (Armato et al., 2004; McNitt-Gray et al., 2007; Reeves et al., 2007; Armato et al., 2007).

For the purposes of training the classifier with the UTMB dataset, we have adopted a simple candidate labeling rule. A candidate is labeled a TP if its segmentation includes a truth cue. It is labeled an FP otherwise. However, since cue points are not provided for the LIDC data, a different labeling method must be employed. Here, if the centroid of a candidate segmentation is contained within any of the radiologist segmentations, the candidate is considered a potential TP. Note that multiple candidate centroids may reside within the provided segmentations of a single nodule. Therefore, only the candidate that overlaps the most with the merged manual segmentation of that specific nodule is labeled a TP and the other potential TPs are ignored for LIDC scoring purposes (they are not counted as TPs or FPs). The merged manual segmentation is obtained by ORing the individual radiologist segmentations. A total of 42 LIDC nodule candidate segmentations are ignored in this fashion. Any candidate with its centroid outside any of the manual segmentations is labeled an FP for LIDC scoring purposes. We believe that the proposed labeling and scoring method for the LIDC data is a rigorous method that makes the best use of the available truth and will ensure that “lucky” candidates that are not associated with a real nodule will not be rewarded. Furthermore, while consistency between labeling for training and testing might be ideal if the available truth information supports that, it need not be identical to fairly evaluate a CAD system. What is paramount is that the labeling/scoring on the testing data be acceptable (since that is how CAD performance is measured).

##### 4.2. Nodule candidate detection and segmentation performance

Here, we examine the performance of the front-end nodule candidates detector. The nodule candidates detector is able to correctly detect 591 out of the total of 606 nodule exemplars in the UTMB/training dataset (sensitivity of 97.52%). When running it on the LIDC/testing dataset, 132 out of the total 143 nodules (based on merged ground truth) are detected (sensitivity of 92.3%). These sensitivities come with an average 900–1200 detections per case depending on the original resolution of the CT scan. That range is

reduced to 450–600 detections per case ( $517.5 \pm 72.9$ ), after rejecting nodule candidates that do not meet the nodule size criterion of the system (i.e., 3–30 mm).

We attribute the reduction in sensitivity of the front-end detector on the LIDC to the presence of some irregularly shaped nodules not well represented in UTMB. We speculate that during the rule-based analysis described in Section 3.2, the intermediate segmentations of those nodules failed to meet the minimum compactness requirements. It is interesting to observe the sensitivity of the front-end detector stratified by nodule subtlety. We have merged the subtlety evaluation of the radiologists by averaging their ratings (out of 5). In our experiment on the LIDC dataset, the front-end detector is able to detect 97.83% of all the conspicuous/obvious nodules (‘merged’ subtlety rating > 3), 96.92% of all the nodules deemed relatively obvious (‘merged’ subtlety rating = 3) and 75% of all the subtle ones (‘merged’ subtlety rating < 3).

The simplicity of the candidate detector and its ability to perform detection and segmentation of candidates simultaneously have strongly contributed to the processing speed of the CAD system. On average, our detection and segmentation method consumed 25–30 s per case on an HP Personal Computer (PC) with a processor speed of 2.6 GHz. This translates to approximately 0.025–0.03 s per nodule candidate. Compared to the processing speed of a segmentation algorithm recently reported by Wang et al. (2007) that consumed on average 13 s for segmenting a nodule using a comparable PC, the proposed detection and segmentation algorithm appears extremely expeditious.

To allow for further evaluation of the proposed detection/segmentation front-end processor, the spatial overlap (intersection divided by union) between the nodule regions provided by the proposed detection/segmentation method and the ones by three or more radiologists is computed. The proposed segmentation method achieved a mean overlap value of approximately 63% (mean overlap value of  $63 \pm 16\%$ ). Note that this result is obtained when testing the detection and segmentation method on 68 nodules identified by three or more radiologists using the original slice resolution of the LIDC dataset (i.e., without any preprocessing such as the down-sampling process) and a 50% consensus level (i.e., PMAP values greater than 50% were considered truth). For context, the reported mean overlap values for the segmentation algorithms presented by Tachibana and Kido (2006), Way et al. (2006) and Wang et al. (2007) with cases of the LIDC are 51%, 58% and 64%, respectively.

##### 4.3. Overall system performance

The overall system performance is quantified using FROC curves. The main result, a two datasets hold-out validation analysis using the recommended system, is presented in Fig. 9. The figure shows CAD sensitivity versus the average number of FPs per case/patient obtained when training a FLD classifier exclusively on the UTMB data (including feature selection) and testing it on the LIDC collection. Scoring is done using the LIDC scoring method described in Section 4.1. In this result, the classifier is using the features labeled (F) in Table 1. The result also includes several other FROC curves as a function of the number of the selected features used by the FLD classifier. Here, we decrement the number of features by 10 from the original pool of 60 and observe the corresponding overall performance result of the system. As shown in Fig. 9, with a specificity of 3 FPs per case/patient, the FLD classifier based CAD system using the first 40 features (i.e., the first 40 features from the original pool of the 60 features identified as the optimum from the independent training set) is able to detect 80.4% (95% confidence interval: 65.4–86.37% obtained using 7 LIDC subsets) of all the nodules in the LIDC collection. On the other hand, at the same specificity, the FLD classifier system based

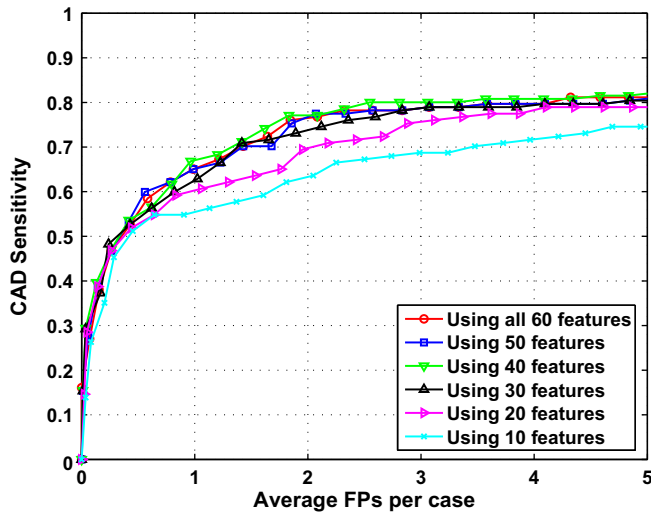


Fig. 9. FROC curves showing CAD performance as a function of the number of selected features used with the FLD classifier. The CAD system was exclusively trained on the UTMB dataset and tested on the LIDC dataset.

system operating with all 60 selected features is able to detect 78.14% (95% confidence interval: 65.33–91.58%) of all those nodules. Note that the performance results of the system for different subsets of the 60 selected features are derived *subsequent* to the SFS (i.e., training) process described in Section 3.4. In other words, the fact that the first 40 selected features yield the best performance result has been deduced *after* testing the system on the LIDC dataset using the desired subsets of selected features and by comparing the corresponding overall performance results. Based on this it appears that the FLD hyperplane decision boundary in the 40 dimensional space generalizes better than the one in 60 dimensional space (significance:  $p = 0.7145$ , i.e., the probability of observing the given result, or one more extreme, by chance if the hypothesis that the two independent samples obtained from 7 LIDC subsets come from distributions with equal means). Note that the maximum correlation coefficient between any two of those first 40 selected features is found not to exceed 0.87 (mean value of  $0.5607 \pm 0.2280$ ).

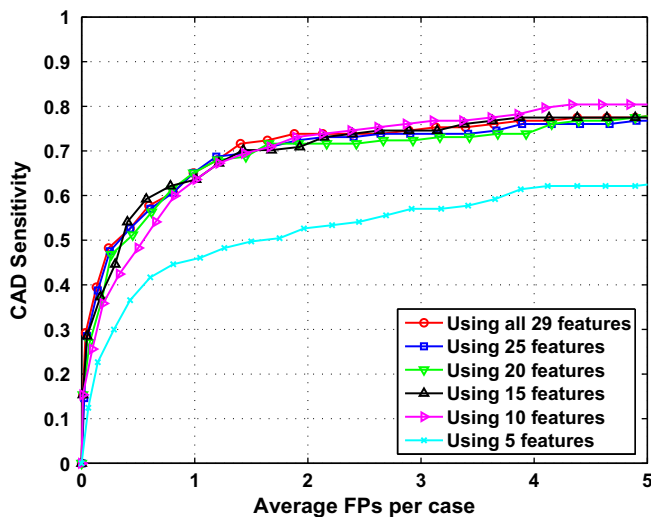


Fig. 10. FROC curves showing CAD performance as a function of the number of selected features used with the quadratic classifier. The CAD system was exclusively trained on the UTMB dataset and tested on the LIDC dataset.

Fig. 10 presents the two dataset hold-out validation performance analysis using a quadratic classifier (exclusively trained on UTMB and tested on LIDC). It shows performance curves with different numbers of selected features (from those listed in Table 1 and labeled (Q)). Fig. 10 shows that, for a specificity in the order of 3 FPs per case/patient, the quadratic classifier based system using only the first 10 features (i.e., the first 10 features from the original pool of the 29 features identified as the optimum from the independent training set) has detected 76.83% (95% confidence interval: 57.7–81.76%) of all the nodules in the LIDC dataset. In contrast, at the same specificity, the quadratic classifier based system using all of the 29 selected features is able to detect 74.89% (95% confidence interval: 62.98–85.49%) of all those nodules. The highest sensitivity at 3 FPs per case is obtained when the quadratic classifier is operating with the first 10 selected features instead of all of the 29 selected features (significance:  $p = 0.5167$ ). Considering the number of samples, it is possible that the quadratic classi-

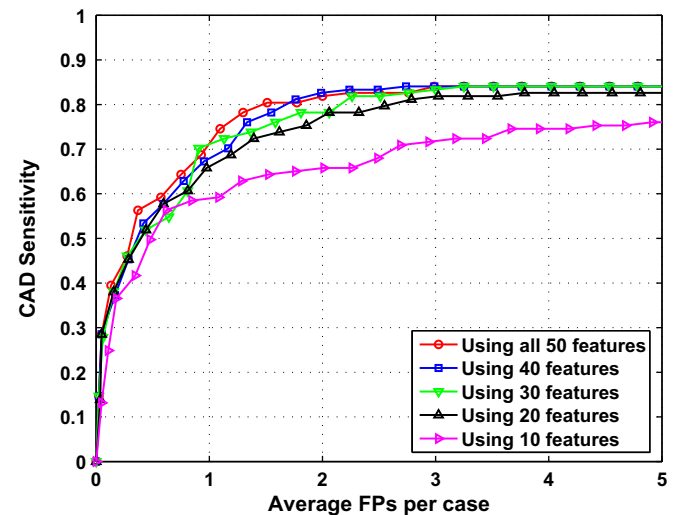


Fig. 11. FROC curves showing CAD performance as a function of the number of selected features used with the FLD classifier for the 7-fold cross-validation case with the LIDC dataset only: trained on LIDC and tested on LIDC.

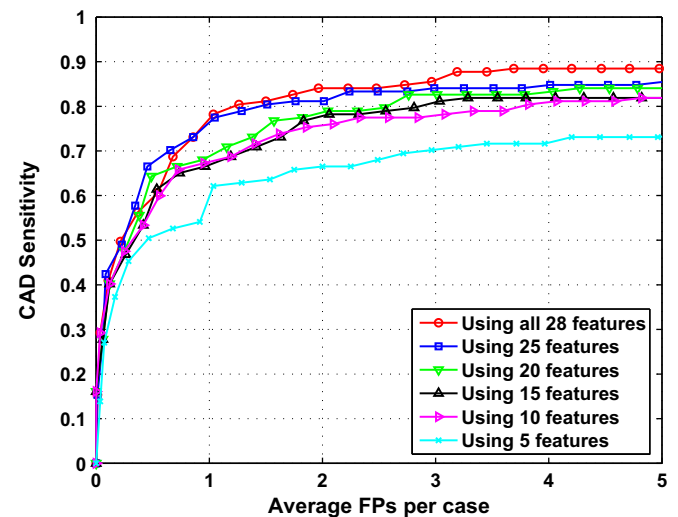


Fig. 12. FROC curves showing CAD performance as a function of the number of selected features used with the quadratic classifier for the 7-fold cross-validation case with the LIDC dataset only: trained on LIDC and tested on LIDC.



fier based system is over-tuned to the training data when using all 29 features.

We also present 7-fold cross-validation performance analysis using the LIDC data only. Figs. 11 and 12 depict 7-fold cross-validation performance curves of the CAD system when using a FLD classifier and a quadratic classifier, respectively. For those results, the LIDC dataset is divided equally into 7 groups, each containing 12 cases (for the total of 84 cases). Here, the FLD classifier is using subsets of 50 features selected using a SFS process and a cross-validation process like that explained in Section 3.4, but here exclusively with the LIDC collection. As it can be seen in Fig. 11, the FLD classifier based system using all 50 selected features shows a sensitivity of 82.66% (95% confidence interval: 66.25–90.72%) with an average of 3 FPs per case/patient. A separate SFS using a quadratic classifier yields to the selection of 28 features. Fig. 12 shows the 7-fold cross-validation results of the quadratic classifier based system, and it can be observed that such a classifier using all 28 selected features is able to detect 84.09% (95% confidence interval: 75.81–95.36%) of all nodules with an average of 3 FPs per case. Since the principle behind the SFS process is to select features so as to maximize the area under the FROC curve, it is not surprising that, for the 7-fold cross-validation results, the performance of the system deteriorates as we reduce the number of selected features used by each classifier. And hence, the best performance result in each case is achieved when using all of the features selected by the according SFS process. However it is interesting to notice that the classifiers were able to generalize fairly well even with few features. For example, the FLD classifier presented comparable performance results when operating with 30 and up of the selected features. In turn, the quadratic classifier based system detected 77.2% of the nodules with an average of 3 FPs using only 10 of the selected features.

One can also compare the performance results of the 7-fold cross-validation analysis (i.e., Figs. 11 and 12) to the performance results of the two datasets hold-out validation analysis (i.e., Figs. 9 and 10) for each classifier and make some interesting observations. However, one must take in consideration the different statistical properties of those two types of performance results, and hence only limited conclusions can be drawn from such comparison. Notwithstanding this, we believe it is important to attempt a relative comparison between the two types of performance results. Note the improved results for the 7-fold cross-validation performance analysis (FLD classifiers significance:  $p = 0.6996$ , quadratic classifiers significance:  $p = 0.0278$ ). This may imply that the UTMB training dataset is not fully representative of the LIDC dataset. As mentioned in Section 2.2, the LIDC database appears to contain a greater number of subtle and irregularly shaped nodules compared to those of the UTMB dataset. The lack of nodule exemplars with similar intensity and morphological characteristic in the training data could explain the improved results for the 7-fold cross-validation performance analysis. However, the two datasets hold-out validation is a more rigorous validation procedure.

In addition to the performance results comparison between the two types of validation analysis, one can also compare the selected features used by each classifier. For example, for the FLD classifiers, one can inspect the discrepancies between the 60 features that are selected by the SFS process exclusively using the UTMB dataset and the 50 features selected by the SFS process using the LIDC data only. In doing so, it is illuminating to note that the two FLD classifiers share some common features. In particular, when examining the first few selected features, it is clear that features such as 'Eccentricity', 'X-Fraction 2' (Geometric features), 'Standard Deviation Separation 1', 'Radial-Deviation Mean Outside 1', 'Radial-Gradient Standard Deviation Perimeter Outside' and 'Minimum Value Below LCE1' (Intensity/Gradient features) are strikingly prominent. Similarly, for the quadratic classifiers, features such as 'Standard

Deviation Separation 1', 'Radial-Deviation Mean Outside 1' and the 7th invariant moment 'Moment 7' (Intensity features) are indeed eminent. Note that 'Standard Deviation Separation 1' and 'Radial-Deviation Mean Outside 1' are actually utilized by the quadratic and FLD classifiers.

Moreover, by comparing the performance results of each classifier for the two datasets hold-out validation analysis (i.e., Figs. 9 and 10), one can also easily deduce that the FLD classifier generally outperforms the quadratic classifier (significance:  $p = 0.3639$ ). We attribute the improved performance of the FLD classifier due to its ability to generalize well, even with a relatively small number of TP training samples (Hardie et al., 2008). These experimental results reinforce our choice of a FLD classifier (*in lieu* of a quadratic classifier) as a back-end processor in our CAD system. Fig. 13 presents performance curves using the favored classifier based system (trained on UTMB) operating on the LIDC collection stratified according to the merged nodule subtlety rating.

A typical output of the CAD system is shown in Fig. 14. In this figure, the corresponding lung mask and detection mask are superimposed on a CT slice image (LIDC case 13614193285030078). This result is obtained when solely training on the UTMB dataset and testing on the LIDC collection using a FLD classifier with 40

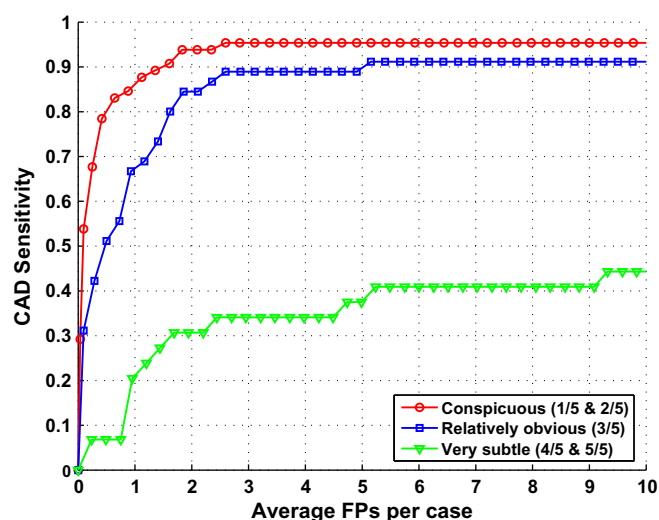


Fig. 13. FROC curves showing CAD performance using the FLD classifier with 40 features operating on the LIDC data stratified by the merged nodule subtlety rating of the four expert radiologists (out of 5).

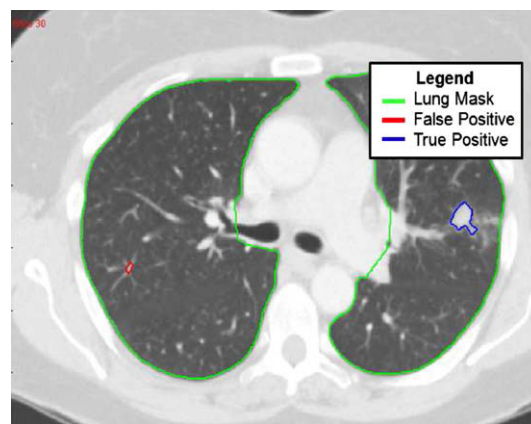


Fig. 14. A typical output of the CAD system superimposed on the according CT slice image.

features operating at a sensitivity of 80.4% accompanied by a false marker rate of 3 per case/patient. The output shows one TP and one FP. As it can be seen in this figure, the cause of the detected FP was the intersection of vessels. After thoroughly inspecting all of the CAD outputs, we also noticed that the lung segmentation algorithm erroneously includes other anatomical structures especially along the clavicle and mediastinum. In such case, a nodule-like object can be segmented and be mistakenly detected by the system. In general, out of the 252 FPs, 52.3% of the FPs were judged to be intersecting vessels, 28.7% of the FPs were judged to be the result of erroneous segmentation of the lung regions, and most other FPs are likely due to atelectasis and scarring. After carefully examining the output results, we also noticed that the majority of the nodules/TPs missed by the CAD system (i.e., classification error) tend to be predominantly small in size and *appear to the authors* to have poorly defined boundaries. Note that the merged subtlety rating of the majority of the missed nodules/TPs is <3 (out of 5). Also note that CAD missed two nodules (i.e., classification error) deemed conspicuous/obvious (merged subtlety rating >3). Those two nodules are irregular in shape with unusual internal intensity characteristics (compared to other conspicuous/obvious nodules) and one of the two is juxtapleural (attached to lung wall near the aorta and other obscuring structures). Consequently, this led to inaccurate segmentations which may be the cause for these misses.

#### 4.4. Performance comparison

To put the performance of the system into context, let us compare our experimental results with reported operating points for previously published CAD systems. It is a difficult task to make definitive comparison between previously published CAD systems due to variability in the datasets (number of cases, scanning protocols, slice thicknesses and spacings), labeling and scoring methods, nodule type and size criterion and validation procedures. It is well known that the performance results of a CAD system can differ significantly depending on those variables. Notwithstanding this, we believe it is still important to attempt a relative comparison. We have identified published results that have used the LIDC dataset to validate detection systems (Opfer and Wiemker, 2007; Sahiner et al., 2007). Opfer and Wiemker (2007) claimed to be the first one to validate a CAD system using the combined LIDC dataset. The set was comprised of 93 cases (slice thickness between 2 and 3 mm) with 127 nodules that were marked at least by one of the four radiologists. Sahiner et al. (2007) evaluated the effect of CAD on radiologists' performance by combining 28 CT scans from the LIDC collection with some other 20 scans yielding to a total

of 48 CT scans containing a total of 73 nodules. Four other radiologists read the CT scans first without and then with CAD, and provided likelihood of nodule ratings for suspicious regions.

Other papers reported CAD performance analysis results using other datasets (Rubin et al., 2005; Gori et al., 2007; Yuan et al., 2006). Rubin et al. (2005) run a CAD system on 20 scans (1.25-mm section thickness and 0.6-mm interval) where two experienced thoracic radiologists in consensus showed 195 non-calcified nodules (reference reading). Gori et al. used a CT dataset that was acquired from the Pisa Centre of the ITALUNG-CT trial (Gori et al., 2007). The system was evaluated *excluding* sub-pleural nodules on 19 CT scans containing 45 internal nodules. We have also looked at the performance results of a commercially available system. In 2006, Yuan et al. (2006) had the opportunity to examine the performance of the ImageCheckerCT LN-1000 (R2 Technology). The model was tested on 150 consecutive patients with 1.25 mm slice original axial images reconstructed with 2.5 mm slice thickness (Yuan et al., 2006). Three hundred thirty-seven regions were classified as true nodules on consensus review.

A comparison of all of the published CAD results for the systems mentioned in this subsection are summarized in Table 2. In this table, each row represents a published benchmark followed by the nodule size criterion and the optimized operating point reported by the according system. Here we show the sensitivities of our CAD system on LIDC using the FLD classifier operating at the same specificity as the reported benchmark system. Note that all training and SFS for our CAD system is done on the independent UTMB dataset and we report results using the first 40 features and using all of the 60 selected features from the SFS process. On the whole, the proposed CAD system seems to be performing well compared to other state-of-the-art systems. The only higher sensitivities reported are for the commercial system, and those results are for 4 mm and larger nodules (and we are operating on 3–30 mm). In addition to the encouraging performance, we believe another important merit of the proposed system is its speed, which is an important factor when considering the use of a CAD system in a clinical setting. On the PC described in Section 4.2, the CAD system consumed approximately 2 min and 15–45 s for a given case. Compared to the CAD system used by Gurung et al. (2006) that was shortest at reconstruction slices of 1.0 mm and required a mean time of 4 min, our system appears to be relatively fast.

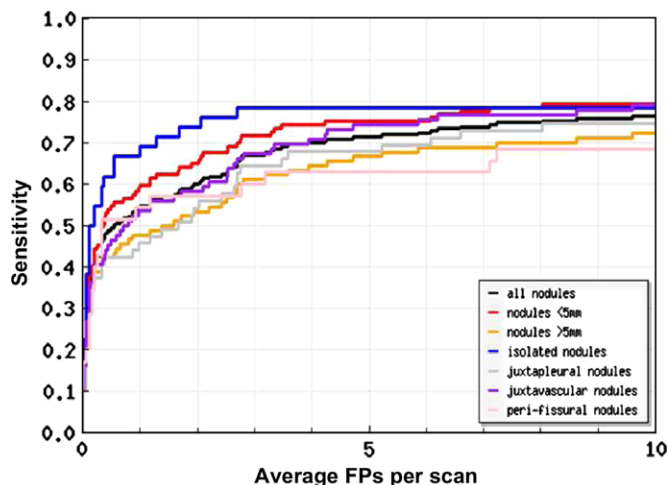
In a final test of the proposed CAD system, we have participated in a study called ANODE09 van Ginneken et al. (2009). This study has been created for the express purpose of comparing CAD systems using a common dataset and common scoring methodology. The ANODE09 testing set is comprised of 50 CT scans without

**Table 2**

CAD system performance comparison. The right most columns show the sensitivity of the proposed CAD system with the described scoring method and operating with the reported average number of FPs per case for the benchmark system listed on the left.

CAD systems	Nodule size criterion used	Average FPs per case	Reported sensitivity (%)	Proposed Sensitivity <sup>*</sup>	
				FLD w/40 features (%)	FLD w/60 features (%)
Opfer and Wiemker (2007)	≥4 mm	4	74	81.54	79.38
Sahiner et al. (2007)	3–36.4 mm	4.9	79	82.81	81.77
Rubin et al. (2005)	≥3 mm	3	76	80.4	78.14
Gori et al. (2007)	≥5 mm	3.8	74.7	81.21	79.38
ImageCheckerCT LN-1000 Yuan et al. (2006)	≥4 mm	3	83.08	80.4	78.14
CAD systems					
	Number of nodules and reported truth “gold” standard method				
Opfer and Wiemker (2007)	127 nodules marked at least by one of the four radiologists				
Sahiner et al. (2007)	73 nodules identified by four other radiologists who provided likelihood of nodule ratings for suspicious regions				
Rubin et al. (2005)	195 non-calcified nodules shown by two experienced thoracic radiologists in consensus				
Gori et al. (2007)	45 internal nodules ( <i>excluding sub-pleural nodules</i> ) annotated by experienced radiologists				
ImageCheckerCT LN-1000 Yuan et al. (2006)	337 nodules classified by consensus review				

<sup>\*</sup> Validated on the LIDC dataset (i.e., 143 nodules identified at least by one of the four radiologists (“gold” standard) Nodule size criterion: 3–30 mm, Slice thickness range: 1.3–3.0 mm, Reconstruction interval range: 0.75–3.0 mm).



**Fig. 15.** FROC curves showing the proposed CAD system performance using a FLD classifier operating with the first 40 selected features. The system is trained exclusively on UTMB and scored on the ANODE09 50 cases based on “relevant” nodules of size 4 mm and up.

annotations that contain 207 various types of “relevant” nodules (van Ginneken et al., 2009). The images are reconstructed with 1.0 mm thickness and 0.7 mm increment (van Ginneken et al., 2009). Participants upload detection coordinates along with values corresponding to the level of suspicion. The scoring is done by the ANODE09 sponsors according to the criteria described in (van Ginneken et al., 2009) based on “relevant” nodules of size 4 mm and up. The results for the CAD system presented here can be found on the ANODE09 website (van Ginneken et al., 2009) under the name “FlyerScan”. Note that the CAD system we used is solely trained on the UTMB dataset using the FLD with 40 features. Our system is optimized for nodules in the size range of 3–30 mm in 2.5 mm slice thickness data, and does not take into account the concept of “relevant” and “irrelevant” findings as specified in the ANODE09 study (van Ginneken et al., 2009). The FROC curves generated courtesy of the ANODE09 study sponsors are shown in Fig. 15. This plot shows how our system performed for a variety of different types of nodules, as well as for all nodules. Note that our CAD system correctly identified 70% of all the “relevant” nodules with a specificity of 4 FPs/case. ANODE09 compares various CAD systems based on the average sensitivity over specificities of 1/8, 1/4, 1/2, 1, 2, 4, and 8 FPs per scan. The sensitivities of the proposed CAD at these false positive rates for all “relevant” nodules and all subsets of “relevant” nodules are tabulated in Table 3. Across all nodules, the proposed system scored 0.552 using this metric. As of December 2009, this was the second best score among 11 participating systems. The next best score is 0.293. The highest scoring system, ISI CAD, was the work of the ANODE09 sponsors (van Ginneken et al., 2009) and that system scored 0.632. It should be noted that that system had the unique advantage of 722 training scans/cases from the same study where the 50 test cases were obtained (not currently available to other participants) (van Ginne-

ken et al., 2009). It is also interesting to note that ISI CAD and FlyerScan together produce better combined results than with any other pair of participating systems. Combining results can be done on the ANODE09 website according to the method described in (van Ginneken et al., 2009).

## 5. Conclusion

In this paper, a new CAD system for the detection of lung nodules in CT imagery is presented. The paper describes the complete design of the CAD system and presents a detailed performance analysis on the publicly available LIDC database. We believe that by training on a diverse and independent dataset, but testing on a publicly available dataset, we were able to thoroughly test the generalizability of the system and to provide a realistic performance benchmark for future research efforts. We have also included 7-fold cross-validation performance results using the testing dataset to provide a reproducible benchmark requiring only the LIDC database.

This paper has explored the performance of the CAD system using different numbers of features selected using an SFS process. The SFS process for the FLD classifier led to the selection of 60 features on the training data. However, the best results on LIDC are obtained when the FLD classifier uses only the first 40 selected features. This illustrates how the SFS process can lead to somewhat over-tuned CAD systems and a conservative approach may be to use only a fraction of the full set of the SFS features from the training set for testing (in this case two-thirds).

We believe that the proposed system offers several useful innovations. The threshold/opening algorithm used for simultaneous detection and segmentation of nodule candidates is attractive because it is conceptually and computationally simple. In Section 4.3 we have compared the FLD classifier performance results to that of a quadratic classifier and found that the FLD generally yields the best results for our data. We believe that the FLD classifier is well suited for this application because it is not only computationally simple, but it also has the ability to generalize well even with relatively few training samples. We also believe that many of the features employed here are novel in this application and represent another contribution of this paper. In particular, we have not seen the use of inside, outside, separation and contrast feature groups used similarly elsewhere. Based on the results presented in Section 4.4, we believe the proposed system performs well compared with other state-of-the-art CAD systems. We believe the combination of the computational simplicity and competitive performance make the proposed system very promising.

This paper has presented a relatively thick-slice analysis due to the resolution of the available independent training (UTMB) data and the high variability in resolution of the testing (LIDC) data. We believe that it is instructive to note from our study that comparatively good CAD results can be obtained with moderate thickness CT scans. However, in view of the expectation that isotropic and high resolution CT data (resolution around 0.7 mm in all directions) will be the standard in the future, an important area of fu-

**Table 3**  
CAD sensitivities over false positive rates for all “relevant” nodules and all subsets of “relevant” nodules in the 50 ANODE09 cases. CAD is solely trained on the UTMB data and uses a FLD classifier operating with the first 40 selected features.

FPs/case	1/8	1/4	1/2	1	2	4	8	Average
Small nodules	0.385	0.444	0.547	0.590	0.650	0.744	0.786	0.592
Large nodules	0.333	0.389	0.422	0.478	0.533	0.644	0.700	0.500
Isolated nodules	0.500	0.548	0.619	0.667	0.738	0.786	0.786	0.663
Vascular nodules	0.360	0.407	0.465	0.535	0.581	0.709	0.767	0.547
Pleural nodules	0.339	0.373	0.424	0.458	0.542	0.678	0.729	0.506
Peri-fissural nodules	0.286	0.400	0.514	0.543	0.571	0.629	0.686	0.518
All nodules	0.362	0.420	0.493	0.541	0.599	0.700	0.749	<b>0.552</b>



ture research would be to further examine the performance of the CAD scheme using isotropic and higher resolution data. In other words, a uniform thin-slice analysis where training and testing the system is accomplished using completely independent, isotropic and high resolution data (excluding the proposed down-sampling process). Note that although the down-sampling process appears justifiable for the LIDC data (i.e., despite the fact that it ensures approximate resolution compatibility, improves the processing speed and reduces noise in data), it implies a loss in resolution (i.e., loss of information). It is known that higher resolution data provides more structural information (i.e., a more accurate anatomical representation). Thus, the use of high resolution data could lead to better segmentations (i.e., accurate lung regions and nodule candidate boundaries), more precise feature computation and subsequent classification (Ernst et al., 2004). Moreover, assuming the availability of thin-slice isotropic data (for training and testing purposes), one could replace many of the 2-D features used in our current system with isotropic 3-D features and get more accurate segmentations and shape information. However, processing such thin-slice data adds to the computational complexity of the CAD system and a possible increase in noise must be considered.

Although the new/proposed CAD system appears promising, further tuning/optimization of the different module parameters could improve the overall performance of the CAD system. A valuable area of future work would be to improve the lung segmentation algorithm. As mentioned in Section 4.3, the lung segmentation erroneously includes some of the anatomical structures and as a result a nodule-like object is segmented and is mistakenly detected by the CAD system. Furthermore, one could also consider improving the segmentation of the detected nodule candidates. While the proposed front-end detector is found to be fast and efficient enough in detecting the majority of the nodules, incorporating a separate nodule candidate segmentation algorithm could aid the performance of the system by providing more accurate candidate boundaries. More anatomical context features may also help to improve the overall system performance by targeting certain classes of observed FPs (e.g., FPs resulting from the intersections of vessels, as mentioned in Section 4.3). Considering that feature selection is a complex and a difficult issue, one could also exploit other feature reduction/extraction techniques (e.g., sequential backward selection, “minimum-Redundancy-Maximum-Relevance” selection, etc.). Another potentially fruitful area of future research might be the use of alternative classifiers as a back-end processor of the system.

## Acknowledgments

The authors would like to thank Dr. Randy Ernest of the University of Texas Medical Branch for providing the CT scans and the corresponding truth to serve here as an independent training data set. We also wish to express our appreciation to the Lung Image Database Consortium and other institutions who contributed to the creation of the LIDC collection used here as an independent testing data set. We would also like to thank the sponsors of ANODE09, and in particular Dr. Bram van Ginneken of the Image Sciences Institute for allowing us to participate in ANODE09 and for processing our results. Finally, we wish to thank the anonymous reviewers for helping to strengthen this paper.

## References

- Anderson, T.W., 1996. R.A. Fisher and multivariate analysis. *Statistical Science* 11 (1), 20–34.
- Armato III, S., Giger, M., Moran, C., Blackburn, J., Doi, K., MacMahon, H., 1999. Computerized detection of pulmonary nodules on CT scans. *RadioGraphics* 19 (5), 1303–1311.

- Armato III, S., Giger, M., MacMahon, H., 2001. Automated detection of lung nodules in CT scans: preliminary results. *Medical Physics* 28, 1552–1561.
- Armato III, S., McLennan, G., McNitt-Gray, M., Meyer, C., Yankelevitz, D., Aberle, D., Henschke, C., Hoffman, E., Kazerooni, E., MacMahon, H., et al., 2004. Lung image database consortium: developing a resource for the medical imaging research community. *Radiology* 232 (3), 739.
- Armato, S., McNitt-Gray, M., Reeves, A., Meyer, C., McLennan, G., Aberle, D., Kazerooni, E., MacMahon, H., van Beek, E., Yankelevitz, D., et al., 2007. The lung image database consortium (LIDC): an evaluation of radiologist variability in the identification of lung nodules on CT scans. *Academic Radiology* 14 (11), 1409–1421.
- Brown, M., McNitt-Gray, M., Mankovich, N., Goldin, J., Hiller, J., Wilson, L., Aberle, D., 1997. Method for segmenting chest CT image data using an anatomical model: preliminary results. *IEEE Transactions on Medical Imaging* 16 (6), 828–839.
- Buhmann, S., Herzog, P., Liang, J., Wolf, M., Salganicoff, M., Kirchhoff, C., Reiser, M., Becker, C., 2007. Clinical evaluation of a computer-aided diagnosis (CAD) prototype for the detection of pulmonary embolism. *Academic Radiology* 14 (6), 651–658.
- “Cancer Facts and Figs. 2009,” The American Cancer Society, 2009.
- Das, M., Muhlenbruch, G., Mahnen, A., Flohr, T., Gundel, L., Stanzel, S., Kraus, T., Gunther, R., Wildberger, J., 2006. Small pulmonary nodules: effect of two computer-aided detection systems on radiologist performance. *Radiology* 241 (2), 564.
- Duda, R., Hart, P., Stork, D., 2000. *Pattern Classification*, second ed. Wiley Interscience.
- Ernst, R., Hardie, R., Gurcan, M., Oto, A., Rogers, S., Hoffmeister, J., 2004. Cad performance analysis for pulmonary nodule detection: comparison of thick- and thin-slice helical CT. *Radiology Society of North America (RSNA)*.
- Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7, 179–188.
- Fukunaga, K., 1990. *Introduction to Statistical Pattern Recognition*. Academic Press, Inc..
- Gonzalez, R., Woods, R., 2007. *Digital Image Processing*. Prentice Hall.
- Gori, I., Fantacci, M., Preite Martinez, A., Retico, A., 2007. An automated system for lung nodule detection in low-dose computed tomography. In: Giger, Maryellen, L., Karssemeijer, Nico (Eds.), *Proceedings of the SPIE on Medical Imaging 2007: Computer-Aided Diagnosis*, vol. 6514, p. 65143R.
- Gurung, J., Maataoui, A., Khan, M., Wetter, A., Harth, M., Jacobi, V., Vogl, T., 2006. Automated detection of lung nodules in multidetector CT: influence of different reconstruction protocols on performance of a software prototype. *Rofu* 178 (1), 71–77.
- Hadjiiski, L., Sahiner, B., Chan, H., Petrick, N., Helvie, M., 1999. Classification of malignant and benign masses based on hybrid ART2LDA approach. *IEEE Transactions on Medical Imaging* 18 (12), 1178–1187.
- Hardie, R., Rogers, S., Wilson, T., Rogers, A., 2008. Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs. *Medical Image Analysis* 12 (3), 240–258.
- Henschke, C.I., McCauley, D.I., Yankelevitz, D.F., Naidich, D.P., McGuinness, G., Miettinen, O.S., Libby, D.M., Pasmantier, M.W., Koizumi, J., Altorki, N.K., Smith, J.P., 1999. Early lung cancer action project: overall design and findings from baseline screening. *Lancet* 354, 99–105.
- Hoffman, E., McLennan, G., 1997. Assessment of the pulmonary structure–function relationship and clinical outcomes measures: quantitative volumetric CT of the lung. *Academic Radiology* 4 (11), 758–776.
- Hu, S., Hoffman, E., Reinhardt, J., et al., 2001. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Transactions on Medical Imaging* 20 (6), 490–498.
- Kanazawa, K., Kawata, Y., Niki, N., Satoh, H., Ohmatsu, H., Kakinuma, R., Kaneko, M., Moriyama, N., Eguchi, K., 1998. Computer-aided diagnosis for pulmonary nodules based on helical CT images. *Computerized Medical Imaging and Graphics* 22 (2), 157–167.
- Leader, J., Zheng, B., Rogers, R., Sciarba, F., Perez, A., Chapman, B., Patel, S., Fuhrman, C., Gur, D., 2003. Automated lung segmentation in X-ray computed tomography development and evaluation of a heuristic threshold-based scheme 1. *Academic Radiology* 10 (11), 1224–1236.
- Lee, Y., Hara, T., Fujita, H., Itoh, S., Ishigaki, T., 2001. Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. *IEEE Transactions on Medical Imaging* 20 (7), 595–604.
- Martinez, A.M., Kak, A.C., 2001. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (2), 228–233.
- Maxion, R., Roberts, R., 2004. *Proper Use of ROC Curves in Intrusion/Anomaly Detection*. University of Newcastle upon Tyne, School of Computing Science, UK. Technical Report Series: CS-TR-871, November 2004.
- McLachlan, G., 2004. *Discriminant Analysis and Statistical Pattern Recognition*. Wiley Interscience.
- McNitt-Gray, M., Armato, S., Meyer, C., Reeves, A., McLennan, G., Pais, R., Freymann, J., Brown, M., Engelmann, R., Bland, P., et al., 2007. The lung image database consortium (LIDC) data collection process for nodule detection and annotation. *Academic Radiology* 14 (12), 1464–1474.
- Nappi, J., Yoshida, H., 2002. Automated detection of polyps with CT colonography: evaluation of volumetric features for reduction of false-positive findings. *Academic Radiology* 9 (4), 386–397.
- Okumura, T., Miwa, T., Kako, J., Yamamoto, S., Matsumoto, M., Tateno, Y., Iinuma, T., Matsumoto, T., 1998. Variable N-Quot filter applied for automatic detection of lung cancer by X-ray CT. *Computer Assisted Radiology and Surgery (CAR98)*, 242–247.



- Opfer, R., Wiemker, R., 2007. Performance analysis for computer-aided lung nodule detection on LIDC data. In: Jiang, Yulei, Sahiner, Berkman (Eds.), *Proceedings of the SPIE – Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment*, vol. 6515, p. 65151C.
- Papoulis, A., Pillai, S.U., 2002. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York.
- Reeves, A., Biancardi, A., Apanasovich, T., Meyer, C., MacMahon, H., van Beek, E., Kazerooni, E., Yankelevitz, D., McNitt-Gray, M., McLennan, G., et al., 2007. The lung image database consortium (LIDC): a comparison of different size metrics for pulmonary nodule measurements. *Academic Radiology* 14 (12), 1475–1485.
- Reinhardt, J., Higgins, W., 1998. Paradigm for shape-based image analysis. *Optical Engineering* 37, 570.
- Rodgers, J., Nicewander, W., Toothaker, L., 1984. Linearly independent, orthogonal, and uncorrelated variables. *The American Statistician* 38 (2), 133–134.
- Rubin, G., Lyo, J., Paik, D., Sherbondy, A., Chow, L., Leung, A., Mindelzun, R., Schraedley-Desmond, P., Zinck, S., Naidich, D., et al., 2005. Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. *Radiology* 234 (1), 274.
- Sahiner, B., Hadjiiski, L., Chan, H., Shi, J., Cascade, P., Kazerooni, E., Zhou, C., Wei, J., Chughtai, A., Poopat, C., et al., 2007. Effect of CAD on radiologists' detection of lung nodules on thoracic CT scans: observer performance study. *Proceedings of SPIE* 6515, 65151D.
- Schilham, A.M.R., van Ginneken, B., Loog, M., 2003. Multi-scale nodule detection in chest radiographs. In: Ellis, R., Peters, T. (Eds.), *Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science Series*, vol. 2878. Springer, pp. 602–609.
- Schilham, A., van Ginneken, B., Loog, M., 2006. A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database. *Medical Image Analysis* 10 (2), 247–258.
- Serra, J., 1983. *Image Analysis and Mathematical Morphology*. Academic Press, Inc., Orlando, FL, USA.
- Shiraishi, J., Li, Q., Suzuki, K., Engelmann, R., Doi, K., 2006. Computer-aided diagnostic scheme for detection of lung nodules on chest radiographs: localized search method based on anatomical classification. *Medical Physics* 33 (7), 2642–2653.
- Simon, G., Reid, L., Tanner, J., Goldstein, H., Benjamin, B., 1972. Growth of radiologically determined heart diameter, lung width, and lung length from 5–19 years, with standards for clinical use. *Archives of Disease in Childhood* 47 (253), 373.
- Sonka, M., Sundaramoorthy, G., Hoffman, E., 1994. Knowledge-based segmentation of intrathoracic airways from multidimensional high-resolution CT images. *Proceedings of SPIE* 2168, 73–85.
- Strickland, R., 2002. *Image-Processing Techniques for Tumor Detection*. CRC Press.
- Tachibana, R., Kido, S., 2006. Automatic segmentation of pulmonary nodules on CT images by use of NCI lung image database consortium. *Progress in Biomedical Optics and Imaging* 7 (30).
- van Ginneken, B., Armato III, S., de Hoop, B., van de Vorst, S., Duindam, T., Niemeijer, M., Murphy, K., Schilham, A., Retico, A., Fantacci, M., et al., 2009. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. Imaging Sciences Institute, University Medical Center Utrecht, The Netherlands, Technical Report. <<http://anode09.isi.uu.nl/>>.
- Wang, J., Engelmann, R., Li, Q., 2007. Segmentation of pulmonary nodules in three-dimensional CT images by use of a spiral-scanning technique. *Medical Physics* 34 (12), 4678–4689.
- Way, T., Hadjiiski, L., Sahiner, B., Chan, H., Cascade, P., Kazerooni, E., Bogot, N., Zhou, C., 2006. Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3 D active contours. *Medical Physics* 33 (7), 2323–2337.
- Wei, L., Yang, Y., Nishikawa, R., Jiang, Y., 2005. A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. *IEEE Transactions on Medical Imaging* 24 (3), 371–380.
- Wiemker, R., Rogalla, P., Opfer, R., Ekin, A., Romano, V., Bülow, T., 2006. Comparative performance analysis for computer aided lung nodule detection and segmentation on ultra-low-dose vs. standard-dose CT. *Proceedings of SPIE* 6146, 614605.
- Yuan, R., Vos, P., Cooperberg, P., 2006. Computer-aided detection in screening CT for pulmonary nodules. *American Journal of Roentgenology* 186 (5), 1280–1287.