



## Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images



Colin Jacobs<sup>a,b,\*</sup>, Eva M. van Rikxoort<sup>a,b</sup>, Thorsten Twellmann<sup>c</sup>, Ernst Th. Scholten<sup>d,e</sup>, Pim A. de Jong<sup>d</sup>, Jan-Martin Kuhnigk<sup>b</sup>, Matthijs Oudkerk<sup>f</sup>, Harry J. de Koning<sup>g</sup>, Mathias Prokop<sup>h</sup>, Cornelia Schaefer-Prokop<sup>i,a</sup>, Bram van Ginneken<sup>a,b</sup>

<sup>a</sup> Diagnostic Image Analysis Group, Department of Radiology, Radboud University Medical Center, Nijmegen, The Netherlands

<sup>b</sup> Fraunhofer MEVIS, Bremen, Germany

<sup>c</sup> MeVis Medical Solutions AG, Bremen, Germany

<sup>d</sup> Department of Radiology, Utrecht University Medical Center, Utrecht, The Netherlands

<sup>e</sup> Department of Radiology, Haarlemmer Kennemer Gasthuis, Haarlem, The Netherlands

<sup>f</sup> Department of Radiology, University Medical Center Groningen, Groningen, The Netherlands

<sup>g</sup> Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands

<sup>h</sup> Department of Radiology, Radboud University Medical Center, Nijmegen, The Netherlands

<sup>i</sup> Meander Medical Centre, Amersfoort, The Netherlands

### ARTICLE INFO

#### Article history:

Received 11 December 2012

Received in revised form 22 November 2013

Accepted 2 December 2013

Available online 17 December 2013

#### Keywords:

Computer aided detection (CAD)

Computed tomography (CT)

Lung nodule

Subsolid nodule

Lung cancer

### ABSTRACT

Subsolid pulmonary nodules occur less often than solid pulmonary nodules, but show a much higher malignancy rate. Therefore, accurate detection of this type of pulmonary nodules is crucial. In this work, a computer-aided detection (CAD) system for subsolid nodules in computed tomography images is presented and evaluated on a large data set from a multi-center lung cancer screening trial. The paper describes the different components of the CAD system and presents experiments to optimize the performance of the proposed CAD system. A rich set of 128 features is defined for subsolid nodule candidates. In addition to previously used intensity, shape and texture features, a novel set of context features is introduced. Experiments show that these features significantly improve the classification performance. Optimization and training of the CAD system is performed on a large training set from one site of a lung cancer screening trial. Performance analysis on an independent test from another site of the trial shows that the proposed system reaches a sensitivity of 80% at an average of only 1.0 false positive detections per scan. A retrospective analysis of the output of the CAD system by an experienced thoracic radiologist shows that the CAD system is able to find subsolid nodules which were not contained in the screening database.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

Lung cancer is the most deadly cancer in both men and women. The American Cancer Society estimates that lung cancer will account for 28% of all cancer-related deaths in the United States in 2012 (American Cancer Society, 2012). This can be largely attributed to the fact that at present, the 5-year survival rate for all stages combined is only 16% (American Cancer Society, 2012). The 5-year survival rate is 52% for cases detected when the disease is still localized, but only 15% of lung cancers are diagnosed at this early stage (American Cancer Society, 2012). Therefore, early

detection of lung cancer, in which it is still treatable, is of major importance to reduce lung cancer mortality.

In the last decade, many screening trials have been initiated to investigate the potential of early detection of lung cancer with low-dose chest computed tomography (CT). Lung cancer in an early stage manifests itself as a pulmonary nodule. Thin-slice helical chest CT scans have a sub-millimeter resolution at which very pulmonary nodules can be detected and therefore, this exam is able to show early stage lung cancer (Henschke et al., 1999). In November 2010, the National Lung Screening Trial (NLST) published a 20% lung cancer mortality reduction in their study group which received 3 annual rounds of low-dose CT screening in comparison to their control group, which received 3 annual rounds of chest X-ray screening (Aberle et al., 2011). This was the first study which showed clear scientific evidence that screening for lung cancer reduces lung cancer mortality.

\* Corresponding author at: Diagnostic Image Analysis Group, Department of Radiology, Radboud University Medical Center, Nijmegen, The Netherlands. Tel.: +31 24 3668112.

E-mail address: [colin.jacobs@radboudumc.nl](mailto:colin.jacobs@radboudumc.nl) (C. Jacobs).

Pulmonary nodules are described as round opacities, well or poorly defined, measuring up to 3 cm in diameter (Hansell et al., 2008). Pulmonary nodules can be differentiated into subsolid and solid nodules (Hansell et al., 2008). Solid nodules have homogeneous soft-tissue attenuation on CT scans. Subsolid nodules can be further differentiated into non-solid nodules (synonym: ground glass nodules) and part-solid nodules (synonym: semi-solid nodules). Non-solid nodules manifest as focal areas of hazy increased attenuation that do not obliterate the bronchial or vascular margins. Areas of hazy increased attenuation are called ground glass opacity and therefore, these nodules are also referred to as ground glass nodules. Part-solid or semi-solid nodules contain both ground glass and solid components. In the Early Lung Cancer Action Project (ELCAP), 81% of all positive findings at baseline were solid nodules and 19% were subsolid nodules, which indicates that subsolid nodules are less common in screening (Henschke et al., 2002). However, this study published a significantly larger malignancy rate of 34% for subsolid nodules, compared to 11% for solid nodules (Henschke et al., 2002). Consequently, about half of the lung cancers found in this study originated from subsolid nodules. Therefore, early detection of subsolid nodules is of major importance.

Computer-aided detection (CAD) of lung nodules in chest CT scans is an extensively researched area in medical research. Although many systems have been proposed for detection of solid nodules (e.g. Tan et al., 2011; Messay et al., 2010; Murphy et al., 2009; Sousa et al., 2009; Ye et al., 2009; Li et al., 2008; Retico et al., 2008; Bellotti et al., 2007; Dehmeshki et al., 2007; Enquobahrie et al., 2007; Marten and Engelke, 2007), only few studies have focused on detection of subsolid nodules (Jacobs et al., 2011; Tao et al., 2009; Ye et al., 2007; Zhou et al., 2006; Kim et al., 2005). In a study by Beigelman-Aubry et al. (2009), it was shown that both a CAD system designed for solid nodules, and radiologists have difficulties in detecting subsolid nodules. Therefore, dedicated CAD algorithms for subsolid nodules are needed.

A few papers have been published on computerized detection of subsolid nodules. Kim et al. (2005) described a slice-based CAD system using texture and intensity features. The system classified regions of interest (ROI) on manually chosen slices from CT examinations of 14 patients into ground glass opacity (GGO) or non-GGO. Performance was measured using receiver operating characteristic (ROC) analysis on all ROIs and showed an area under the ROC curve of 0.92. Zhou et al. (2006) developed an automatic scheme for both detection and segmentation of subsolid nodules based on vessel suppression, intensity and texture analysis. They reported good performance but the test data set contained only 10 subsolid nodules. Ye et al. (2007) presented a voxel-based method with rule-based filtering that was tested on 50 CT examinations with 52 subsolid nodules. They reported a high sensitivity of 92.3% but also a high false positive (FP) rate of 12.7 per scan. Tao et al. (2009) developed a multi-level detection scheme with classification at voxel-level and object-level. They focused on small volumes of interest (VOIs) generated by a candidate detector algorithm which was not otherwise specified. The method was tested on a set of 1100 VOIs including 100 positive ones, from 153 healthy and 51 diseased patients. Results were provided for VOIs only, and neither the FP rate per scan nor the total number of VOIs per scan were reported. Finally, we published a preliminary version of our subsolid nodule CAD system which we trained and evaluated on a data set of 140 scans from one site of a large lung cancer screening trial. In this study, we reported promising results as we reached 73% sensitivity on the independent test set at an average of 1.0 FP/scan (Jacobs et al., 2011).

In this work, a novel automatic computer-aided detection system for subsolid nodules is presented and evaluated on a large database of a multi-center lung cancer screening trial.

The first algorithms for subsolid nodule detection have been evaluated on rather small amounts of data ranging from 10 to 50 thoracic CT scans containing between 10 and 52 subsolid nodules in total (Kim et al., 2005; Zhou et al., 2006; Ye et al., 2007). The most recent study used 1100 subvolumes from CT scans from around 200 subjects of which 100 subvolumes contained subsolid nodules (Tao et al., 2009). In this study, we collected cases from two sites of a large lung cancer screening trial and included all CT examinations in which subsolid nodules were annotated, leading to a larger database than presented by prior studies. Data collection is described in detail in Section 2.

The different components of the proposed CAD system, including candidate detection, feature calculation and classification, are described in Section 3. In comparison to the publication by Tao et al. (2009), which is the most recent publication on subsolid nodule detection, we describe a full CAD system including the candidate detection step, which was not specified in the publication by Tao et al. (2009). Different classes of features have been utilized in previous publications. The first work by Kim et al. (2005) and Zhou et al., 2006 only used texture and intensity features. In the publications by Ye et al. (2007) and Tao et al. (2009), shape features are added to the feature set. In this work, we add another class of features by including context features which are calculated at the image-level, describing the relation of an area of ground glass opacity to its surroundings such as the lung, airways, vessels and other nodule candidates. We performed an experiment to show that these type of features benefit the classification performance. Experiments to optimize the configuration of the CAD system are also explained in Section 3. We experiment with many different classifiers to find the optimal classifier for this classification task. This is an important optimization of the CAD system which has also been mentioned as future work by prior publications (Tao et al., 2009).

Section 4 outlines the results of the CAD system on the independent test set. This section also reports the performance of a solid nodule CAD system on our database with many subsolid nodules and how the combination with a subsolid nodule CAD would benefit the detection performance for detection of subsolid nodules. At the best of our knowledge, we believe this has not yet been done in any publication to date.

In order to get a good insight into the performance of the CAD system and its potential to be applied in a clinical setting, an experienced thoracic radiologist reviewed the CAD marks of the CAD system on the independent test set and the results of this investigation are presented in Section 5. Finally, we discuss the performance and limitations of the proposed CAD system and opportunities for future work in Section 6 and conclude in Section 7.

## 2. Materials

Data for this study was collected from the NELSON trial, a large multi-center lung cancer screening trial, organized in the Netherlands and Belgium (van Klaveren et al., 2009). NELSON is an ongoing randomized control trial established to test if screening for lung cancer by low-dose CT in high-risk (ex-)smokers will lead to a 25% mortality reduction in lung cancer mortality.

In total, the trial includes 7557 participants who receive multiple rounds of screening with low-dose CT. All pulmonary nodules which were found during visual reviewing of the CT scans were recorded in the trial database. Among other characteristics, the screening radiologists indicated the location, diameter and nodule type of all detected pulmonary nodules; calcified, solid, part-solid or non-solid (van Klaveren et al., 2009). This database served as reference standard for our CAD analysis. We collected all thin-slice,

low-dose CT examinations from two sites of the NELSON trial in which at least one subsolid (part-solid or non-solid) nodule was annotated. In total, the data set from which the scans were selected consisted of around 20,000 scans from around 4500 subjects. All subsolid nodule annotations with a diameter smaller than 5 mm were discarded because current clinical guidelines state that these nodules do not require follow-up CT (Naidich et al., 2013; Godoy and Naidich, 2009). It has to be noted that in this work, solid nodule annotations were not included in the analysis.

The first site was the University Medical Center in Utrecht, The Netherlands. Screening CT examinations were made with a 16-detector row CT scanner (MX8000 IDT or Brilliance 16; Philips Medical Systems, Cleveland, Ohio) using a moderately soft reconstruction kernel (B; Philips Medical Systems). The second site was the Haarlemmer Kennemer Gasthuis in Haarlem, The Netherlands. At this site, CT examinations were made using a Somatom Sensation 16 (Siemens Medical Solutions) and reconstructed using a B30f kernel. All CT examinations at both sites were acquired in helical mode with  $16 \times 0.75$  mm collimation. Exposure settings were 30 mAs at 120 kVp for patients weighing less than 80 kg and 30 mAs at 140 kVp for those weighing more than 80 kg. Axial images of 1.0-mm thickness were reconstructed at a 0.7-mm increment with a  $512 \times 512$  matrix. In-plane voxel sizes varied from 0.53 mm to 0.89 mm.

After removal of the annotations of the subsolid nodules smaller than 5 mm, 209 scans from 103 patients were collected from the first site. In these patients, 122 subsolid nodules (63 part-solid, 59 non-solid) were found. Note that a pulmonary nodule can be annotated in multiple scans because follow-up examinations of a patient were included. Consequently, 225 annotations in 209 scans were found. This data set was solely used for training and optimization of the CAD system.

At the second site, 109 scans from 56 patients were collected after removal of the small subsolid nodules. In these patients, 60 subsolid nodules (32 part-solid, 28 non-solid) were found, which led to 114 annotations in the 109 scans. This data set, acquired with a different type of scanner from a different manufacturer, was used for independent evaluation of the CAD system.

The effective diameter of the nodules in both data sets varied between 5 and 34 mm, with a median of 10.7 mm.

### 3. Methods

This section describes the different steps of the CAD system. Prior to the detection pipeline, previously published lung, airway and vessel segmentation algorithms were applied (van Rikxoort et al., 2009; van Ginneken et al., 2008; Dongen and Ginneken, 2010). An initial detection stage including nodule segmentation generates a set of candidates. A rich set of features was defined for subsolid nodule candidates. Previous publications on subsolid nodule CAD have used intensity, shape and texture features (Kim et al., 2005; Zhou et al., 2006; Ye et al., 2007; Tao et al., 2009). In this paper, we add another class of features by including context features. We performed classification experiments excluding and including context features to show the additional value of these features. We experimented with different classification schemes and different classifiers to investigate their influence on the classification performance. Previous publications already expressed the optimization of the classification as future work (Tao et al., 2009). In this paper, we present an extensive and structured evaluation to select the best classification scheme. Based on the results of these experiments, the optimal configuration of the CAD system is chosen and this final system was evaluated on the independent test set. Finally, we combined our subsolid CAD system with a previously published solid nodule CAD system (Murphy et al., 2009).

#### 3.1. Candidate detection

##### 3.1.1. Coarse candidate detection

The candidate detection procedure is started by applying a double-threshold density mask within the lung regions to obtain a mask of voxels with attenuation values commonly observed in ground glass opacities. A range between  $-750$  and  $-300$  Hounsfield units (HU) is used, similar to previous studies (Heitmann et al., 1997; Kauczor et al., 2000). Partial volume effects at the edges of the lungs, vessels and airways can also give rise to attenuation values in the defined range. To remove these voxels, a morphological erosion operation using a spherical structuring element with a diameter of 3 voxels is applied. After this step, a connected component analysis is performed to cluster all voxels into candidates. Since subsolid nodules with a diameter smaller than 5 mm do not require follow-up CT (Godoy and Naidich, 2009), all candidates which have a volume smaller than  $34 \text{ mm}^3$  (corresponding to the volume of an ideal sphere with a diameter of 4 mm are removed). Subsequently, a morphological dilation operation with the same structuring element is applied to undo the shrinking induced by the erosion operation. Finally, the volume and center of mass of all candidates are computed and candidates for which the centers of mass are within 5 mm of each other are merged. This merging procedure is applied to ensure that a nodule is covered by only one candidate.

##### 3.1.2. Nodule segmentation

The candidate detection procedure described above generates clustered regions, but these are not an accurate segmentation of the subsolid nodules. Therefore, a previously published robust pulmonary nodule segmentation algorithm is used to acquire accurate nodule segmentations (Kuhnigk et al., 2006). This algorithm works on a cubic volume of interest and applies an efficient combination of morphological operations to acquire a robust nodule segmentation. Automatic chest wall removal and separation from attached vasculature are incorporated into the algorithm. This method has been extensively evaluated for solid nodules and showed excellent results (Kuhnigk et al., 2006). We slightly adjusted the segmentation algorithm to work for subsolid nodules. The algorithm described in the paper by (Kuhnigk et al., 2006) used a global lower threshold of  $-450$  HU. In order to get good segmentation results for subsolid nodules (both part-solid and non-solid), the lower threshold is changed to  $-750$  HU. The volume-of-interest for the segmentation algorithm is created around the center of mass of the candidate and the size of the VOI is set to 1.5 times the equivalent diameter of the initial candidate. In this way, accurate segmentations for all candidates are created and this forms the final set of candidates which go into the classification process.

#### 3.2. Features

A rich set of features to describe the candidates is computed which can be subdivided into four categories: intensity, texture, shape features and context features.

##### 3.2.1. Intensity features

Intensity features are calculated on four different sets of voxels:

- *segmentation*, voxels inside the candidate segmentation,
- *boundingBox*, voxels inside a bounding box defined around the candidate segmentation,
- *surrounding3*, voxels inside the surrounding of the candidate segmentation, created by dilating the candidate segmentation with a rectangular structuring element of size  $3 \times 3 \times 3$  voxels,

- *surrounding5*, voxels inside the surrounding of the candidate segmentation, created by dilating the candidate segmentation with a rectangular structuring element of size  $5 \times 5 \times 5$  voxels.

An example of these regions is depicted in Fig. 1. These four regions are defined to extract features from the intensity profiles of the inner and the surrounding of a candidate. For each set of voxels, a normalized histogram is computed using a bin size of 50 HU. The bin size has been empirically determined such that the histograms are not too sparse, but still contain the necessary information to describe the underlying intensity distribution. For each normalized histogram, the following statistics are computed: entropy, mean, height of mean bin, mode (position of maximum peak), height of mode bin and value of the bins of the quantiles 5%, 25%, 50%, 75% and 95%. Furthermore, the standard deviation, minimum, maximum value and the first 7 Hu moments (Hu, 1962) are computed for voxel set *segmentation*. The Hu moments are translation, scale and rotation invariant and used to describe the underlying intensity profile. Finally, the maximum vesselness (Frangi et al., 1998) over multiple scales (1.0, 1.77, 3.16, 5.62 and 10.0 voxels) is computed and the minimum, maximum, mean and standard deviation of the maximum vesselness in voxel set *segmentation* are used as features. Partial volume effects can create areas of ground glass opacity close to vessels or on vessel walls and by including these features, we capture information whether the candidate is in vicinity to or at a vessel wall.

In total, 54 intensity features are collected.

### 3.2.2. Texture features

For texture analysis, local binary patterns (LBP) and 2D Haar wavelets are used (Ojala et al., 2002). Both are commonly used texture descriptors to describe localized spatial texture information and have been used for parenchymal texture analysis in CT images (Sørensen et al., 2008). These features for example help to exclude false positive candidates in regions of homogeneous ground glass opacity caused by motion artifacts. A VOI is created from the bounding box around the candidate segmentation and this volume is resampled (Lanczos resampling,  $a = 3$ ) to two cubic volumes of interest of  $16 \times 16 \times 16$  and  $32 \times 32 \times 32$  voxels, respectively. Subsequently, 2D local binary patterns using a neighborhood of  $3 \times 3$  ( $P=8$ ,  $R=1$ ) are computed for every slice of the resampled volumes. Then, a normalized histogram with bin size 1 is computed of the LBP output for each volume and the same histogram statistics as mentioned in paragraph 3.2.1, except the quantiles, are used as texture descriptors.

Furthermore, 2D Haar wavelets are applied on the  $32 \times 32 \times 32$  resampled volume. Each slice of the resampled volume is decomposed into four bands. Four volumes are constructed from the four bands from every slice. Three normalized histograms are computed from the three volumes built from the high-frequency bands. The volume built from the low-frequency band is not used. Again, the same histogram statistics are used as texture descriptors.

This leads to 40 texture features in total.

### 3.2.3. Shape features

The third group of features consists of shape features, which are computed from the candidate segmentation. Segmentation on other structures than nodules can create odd shapes and therefore, shape is an important feature to discriminate true positive from false positive samples. First, the following features are calculated: *sphericity*, *compactness1*, *compactness2* and *guessRadius*. In order to calculate the *sphericity*, a sphere  $S$  is defined at the center of mass of the candidate region with the same volume as the candidate segmentation. Then, *sphericity* is defined as the ratio between the volume of the voxels of the candidate segmentation within sphere  $S$  and the total volume of sphere  $S$ . Then, in order to calculate *compactness1*, *compactness2* and *guessRadius*, the bounding box around the candidate segmentation is used and the dimensions are named *dimx*, *dimy* and *dimz*. To calculate *compactness1*, the number of voxels of the candidate cluster is divided by the total number of voxels within the bounding box. *Compactness2* is calculated by dividing the number of voxels in the candidate cluster by the number of voxels in a cube for which the size is defined by the largest dimension of the bounding box ( $\max(\text{dimx}, \text{dimy}, \text{dimz})$ ). The feature *guessRadius* is calculated by dividing the volume of the bounding box by 6. In case of a perfect spherical nodule, this will produce the exact radius of the sphere. Secondly, the number of voxels and the cluster size in  $\text{mm}^3$  are computed to describe the size of the candidate. These two features are almost identical, but the cluster size takes the resolution of the CT scan into account. Finally, the same set of 7 invariant Hu moments are computed from the candidate mask voxels to describe its shape. Note that in contrast to the previous calculation of Hu moments for the intensity features, the voxels are in this case set to 1 inside the segmentation and 0 outside the segmentation.

In total, 13 shape features are computed.

### 3.2.4. Context features

Finally, a novel group of context features is defined, which describe the location of the candidate region in respect to the lung boundary, the airway tree, the vessels and other subsolid nodule candidates. The location of the candidate with respect to the lungs, vessels and airways is important for multiple reasons. For example, larger areas of ground glass opacity can be seen at the gravity dependent portions of the lung (base of the lung when CT is performed supine) due to microatelectasis. This will result in a candidate which has an elongated shape along the boundary of the lung. A combination of shape and context features is able to capture this. Another example is airways filled with mucus, which can manifest in the intensity range of ground glass opacities. These candidates will show an overlap with the airway segmentation and this can be used to classify them as false positive. Furthermore, the relation of candidates to other candidates is relevant contextual information. For example, a small candidate which is surrounded by many other candidates is more likely to be originating from an area of microatelectasis than to be a subsolid nodule.

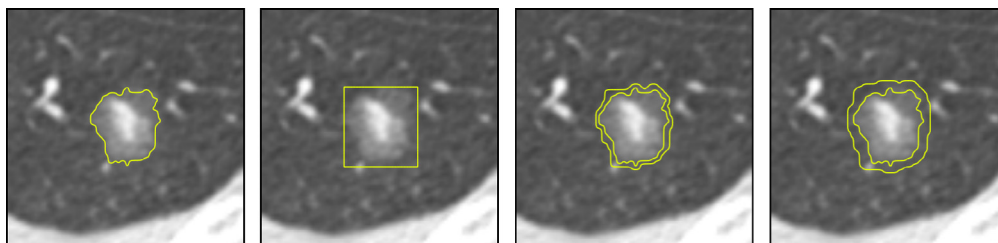


Fig. 1. Four different sets of voxels on which features are calculated. Left: *segmentation*, Middle left: *boundingBox*, Middle right: *surrounding3*, Right: *surrounding5*.



First, two distance transforms are calculated within the lung regions; the first using the lung segmentation and the second using the airway tree. The distance to the lung boundary and distance to the closest airway is extracted from the distance transforms for all voxels inside the candidate segmentation. The mean, standard deviation, minimum and maximum distance to the lung boundary and airways are computed and used as context features.

Secondly, a bounding box is defined around the lungs and this is used to compute relative position features; relative  $X$ ,  $Y$  and  $Z$  position, and distance to left bottom corner of the bounding box are computed. Furthermore, the distance to the center of mass of both lungs is calculated.

Thirdly, the absolute and relative airway and vessel overlap are computed. To calculate this, we count the number of voxels within voxel set *boundingBox* which are part of the airway segmentation or the vessel segmentation. The exact number of voxels is the absolute overlap and the relative overlap is calculated as the number of voxels inside the segmentation divided by the total number of voxels in *boundingBox*.

Finally, the relation of a candidate with respect to other candidates is described. First, we use the total amount of candidates within the scan as a feature. This provides information about the number of ground glass areas in the lung. Secondly, we calculate the number of candidates within a distance of 30 mm and 50 mm of the candidate and the distance to the closest other candidate.

In total, this sums up to 21 context features.

### 3.3. Classification

In this section, the experiments to optimize the classification performance are described. Furthermore, we describe the evaluation of the CAD system on the independent test set. During evaluation of the system, a nodule is marked as detected when the center of mass of the candidate is within a distance  $R$  of the center of the nodule. In order to ensure that the CAD mark is displayed within the nodule on the CT scan, we set  $R$  to be the radius of the nodule size. This radius is half of the diameter which is reported by the radiologist during reading of the CT in the screening.

#### 3.3.1. Optimization of the classification scheme

In order to select the best classification scheme, several classification experiments are conducted. These experiments are performed in 10-fold cross-validation on the training set. Since patients can have multiple scans of the same pulmonary nodule, the folds are created by splitting at a patient level to prevent bias. Candidates are classified into two classes: nodule or false-positive (FP), and the final performance of the CAD system is evaluated using free-response operating characteristic (FROC) analysis.

Classification of candidates is tested using a single versus a two-stage classification scheme. The one-stage classification scheme computes the complete set of features for all candidates and uses one supervised classifier to classify all candidates into the two classes. In contrast, the two-stage classification scheme utilizes only five features in the first stage to perform a first-stage classification. This first-stage classification is aimed at removing as many false-positive candidates as possible. Then, the complete set of features is only calculated for all remaining candidates. This two-stage approach has two advantages. Firstly, the computation time of the CAD system for the two-stage classification scheme will be shorter since the complete set of features does not have to be calculated for all candidates. Secondly, the first-stage classification could make the data set more balanced, which could be beneficial for the classifiers tested in the second stage. Both approaches are tested to

evaluate which classification scheme is optimal in terms of classification performance.

The first stage classification of the two-stage classification scheme is performed using a Linear Discriminant Classifier (LDC) (Fukunaga, 1990) because of its simplicity and speed. The optimal set of five features for this first-stage classification is determined using three approaches. In the first two approaches, a sequential forward floating selection (SFFS) procedure (Pudil et al., 1994) is used to select 5 features. The SFFS procedure uses a random 50% of the training fold as training data and the other 50% as testing data. In the first approach, accuracy is used as optimization criterium for the SFFS procedure.

In the second approach, the partial area under the FROC curve between 0 and 3 FP/scan is used as optimization criterium. Finally, in the third approach, the Fisher's linear discriminant ratio (Jobson, 1992) is calculated for all features and the five features with the highest ratio are selected. Note that this approach does not consider feature combinations. Three different LDC classifiers are trained using these three different sets of five features and this produces a likelihood for all candidates for each classifier. The likelihood threshold for stage one for each classifier is determined by sorting all likelihoods of the positive samples and selecting the lowest likelihood. Consequently, no true positives are removed in the training set. All candidates with a likelihood below this threshold are removed. In this way, we obtain the set of five features which removes the most false positive candidates without removing true positives.

For the one-stage classification scheme and the second stage of the two-stage classification scheme, a  $k$ -nearest neighbor classifier (kNN) (Cover and Hart, 1967), random forest classifier (RF) (Breiman, 2001), GentleBoost classifier (GB) (Friedman et al., 2000), nearest mean classifier (NM) (Fukunaga, 1990), support vector machine using radial basis function kernel (SVM-RBF) (Vapnik, 1995), and LDC are tested in order to find the optimal classifier for this classification task. Parameters of the different classifiers were also optimized in cross-validation on the training set. In the kNN classifier,  $K$  was set to the square root of the number of positive samples. The random forest classifier was trained with 100 trees with a maximum tree depth of 20. For the GentleBoost classifiers, regression stumps were used as weak classifiers and 250 weak classifiers were used to train the classifier. Next to these classifiers, a combination of 10 GentleBoost classifiers, referred to as GB10, is used. This classifier consists of 10 GentleBoost classifiers, which are all separately trained using a random 75% of the training set. The final output of the GB10 classifier is the median of the 10 different classifier probabilities. In pilot experiments, this classifier produced improved performance compared to a single GentleBoost classifier. For the GB10, we also used 250 regression stumps in each separate GB classifier. The  $C$  and gamma parameter of the SVM-RBF classifier were optimized in an inner 5-fold cross-validation loop within the training fold of the 10-fold cross-validation loop. As an optimization criterium, the partial area under the FROC curve between 0 and 3 FPs/scan was used. All features are normalized to zero mean and unit variance.

#### 3.3.2. Evaluation of benefit of the context features

We hypothesize that the presented context features contribute to a significantly better classification performance. We conducted an experiment on the training set using the final system with and without the context features to test this hypothesis. The bootstrap method is used to test statistical significance. Scans were sampled with replacement from the cross-validation set 5000 times. Every bootstrap sample had the same number of scans as the original data set. Classification performance is measured by the partial area under the FROC curve between 0 and 8 FPs/scan.

### 3.3.3. Evaluation of the optimal classification scheme on independent test set

The optimal classification scheme, one-stage or two-stage classification, and the optimal classifier was chosen based on FROC analysis of the cross-validation results on the training set. Then, the optimal classification scheme was trained using the complete training set and tested on the independent test set to evaluate the performance of the CAD system. Note that this test set has not been used in any way during the optimization of the classification scheme.

### 3.4. Combination with solid nodule CAD

In clinical practice, the subsolid nodule CAD system will operate in combination with a solid nodule CAD system. Although solid nodule CAD algorithms are not optimized and trained for detection of subsolid nodules, they may still detect a fraction of all subsolid nodules. In particular, they may be sensitive to detecting the solid core of part-solid nodules. Therefore, the combination of a solid nodule CAD and the proposed subsolid nodule CAD may increase the overall detection sensitivity of subsolid nodules. To evaluate this, a previously published nodule CAD system (Murphy et al., 2009) is applied to all cases in the test set. This CAD system reached an excellent sensitivity in a large comparative study of nodule CAD algorithms, the ANODE09 study (van Ginneken et al., 2010). In this work, the CAD system was set to operate at an average of 4 false positives per scan. Note that this operating point has been determined on an independent data set so the system is not guaranteed to generate precisely this false positive rate on our test set.

## 4. Results

### 4.1. Candidate detection

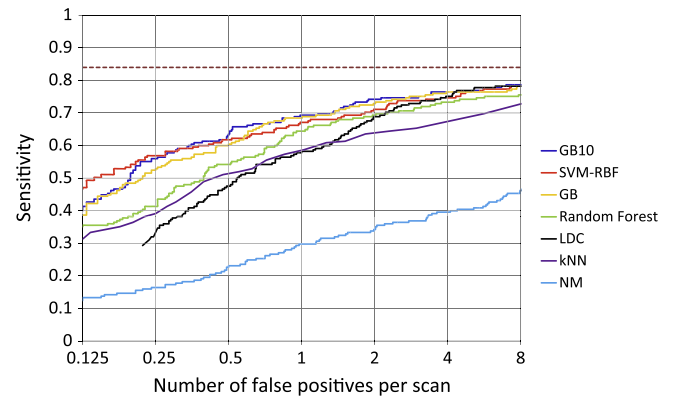
The candidate detection step generated  $237 \pm 267$  candidate regions per scan in the training set and  $109 \pm 127$  candidate regions per scan in the test set. In the training set, the candidate detection sensitivity was 84% for all subsolid nodules, where the sensitivity for part-solid nodules and non-solid nodules separately was 81% and 87%, respectively. In the test set, the sensitivity was 88% for all subsolid nodules, and 85% and 90% for part-solid and non-solid nodules, respectively.

### 4.2. Classification

#### 4.2.1. Optimization of the classification scheme

The FROC curves of the different classifiers in the single stage classification scheme on the full training set in 10-fold cross-validation are depicted in Fig. 2. This figure shows that the GB10 classifier performs best and reaches 69% sensitivity at 1 FP/scan and 74% sensitivity at 2 FPs/scan. Note that the candidate detection sensitivity is 84%, which means that the classification sensitivity cannot be higher than this value. This is indicated by the dotted line in Fig. 2.

The goal of the first stage classification is to reduce the amount of FPs in the set of candidate regions. Using the three approaches explained in Section 3.3.1, three different sets of five features for stage 1 classification are constructed and their performance is tested in 10-fold cross-validation. Table 1 shows the amount of samples which are removed when a posterior probability threshold  $T$  is used which removes no true positives in the training set. This table shows that the feature set based on Fisher's linear discriminant ratio removes the most samples and therefore, this feature set is selected for the first stage of CAD system.



**Fig. 2.** FROC curves of the different classifiers in a single stage classification scheme tested in 10-fold cross-validation on the full training set. The horizontal axis is logarithmic. The dotted line indicates the maximal sensitivity which can be reached due to the candidate detection sensitivity. GB10: combination of 10 GentleBoost classifiers, as explained in Section 3.3.1, GB: GentleBoost classifier, SVM-RBF: support vector machine with radial basis function kernel, LDC: linear discriminant classifier, kNN: k-nearest neighbor classifier, NM: nearest mean classifier.

**Table 1**

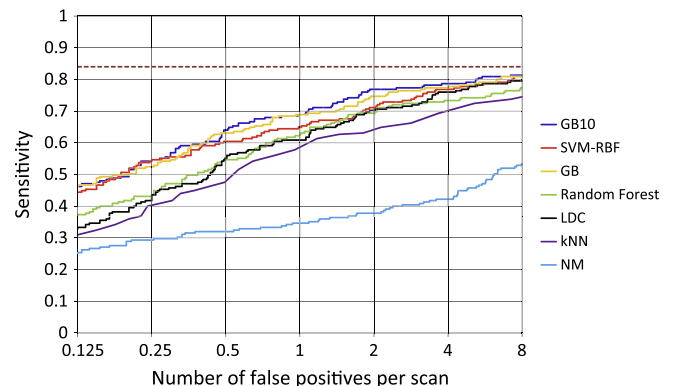
Performance of the first-stage LDC classifier for different feature groups selected by: (a) sequential feed-forward selection (SFFS) using accuracy as optimization criterion, (b) SFFS using partial area under the FROC curve between 0 and 3 FPs/scan as optimization criterion, (c) Fisher's linear discriminant ratio (FLDR). Right column shows the reduction ratio: the percentage of remaining candidates after removal of candidates below the likelihood threshold.

Feature set	Reduction ratio (%)
SFFS – accuracy	68
SFFS – partial area under FROC curve	79
Fisher linear discriminant ratio	59

The performance of the different classifiers in the second stage is depicted in Fig. 3. Comparable to Fig. 2, the GB10 classifier also performs best in this second-stage classification. Again, note that the candidate detection sensitivity is 84%, which means that the classification sensitivity cannot be higher than this value.

### 4.2.2. Evaluation of the optimal classification scheme on independent test set

The performance of the best classifier for the single stage classification scheme, GB10, is compared to the best configuration



**Fig. 3.** FROC curves of the different classifiers in a two-stage classification scheme tested in 10-fold cross-validation on the full training set after first stage classification. The horizontal axis is logarithmic. The dotted line indicates the maximal sensitivity which can be reached due to the candidate detection sensitivity. GB10: combination of 10 GentleBoost classifiers, as explained in Section 3.3.1, GB: GentleBoost classifier, SVM-RBF: Support vector machine with radial basis function kernel, LDC: linear discriminant classifier, kNN: k-nearest neighbor classifier, NM: nearest mean classifier.

for the two-stage classification scheme; LDC with the FLDR features and the GB10 for the second stage. There is no significant difference between the two classification schemes.

Computation time of the system with different classification schemes was measured on a Dell laptop using 1 core with a processor speed of 2.70 GHz. Using the one-stage classification scheme, the average computation time per case was 129 s. Using the two-stage classification scheme, the average computation time per case was 122 s. These results exclude the computation time of the lung, airway and vessel segmentation. Although the computation time difference is small, the two-stage classification scheme is selected as the final classification scheme. The complete configuration of the CAD system is depicted in Fig. 4.

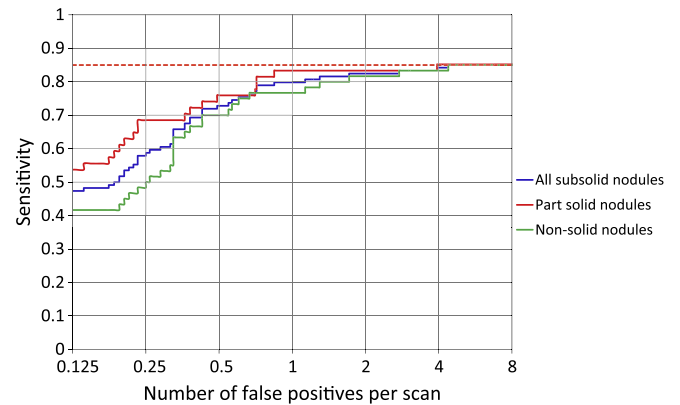
Finally, Fig. 5 shows the FROC curve of the performance of the final configuration of the CAD system on the test set. This system has a two-stage classification scheme using a LDC and GB10 classifier, is trained on the complete training set and tested on the test set. Furthermore, FROC curves for part-solid nodules only and for non-solid nodules only are displayed. This figure shows that the final configuration of the CAD system reaches 80% sensitivity at 1 FP/scan and 83% sensitivity at 2 FPs/scan. Note that the candidate detection sensitivity on the test set was 88% (85% for part-solid nodules and 90% for non-solid nodules). After the first stage classification, three true-positives (TP) were removed from the test set. These were all non-solid nodules. Consequently, the maximal classification sensitivity in this graph is 85% (97 out of 114) for all subsolid nodules, 85% (46 out of 54) for part-solid nodules and 85% (51 out of 60) for non-solid nodules. Note that the CAD system is able to detect all subsolid nodules when the system is set to operate at 4 FPs/scan.

#### 4.2.3. Evaluation of benefit of the context features

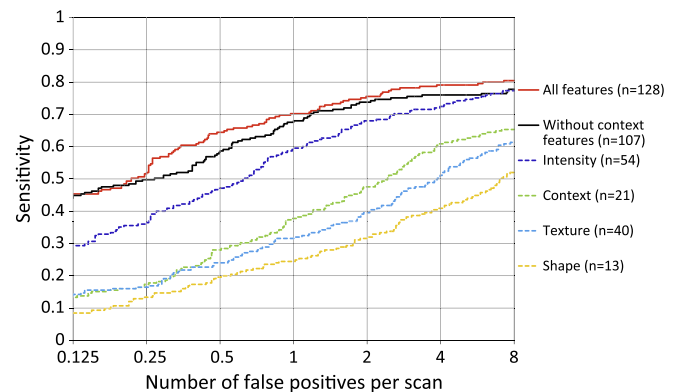
Fig. 6 shows the FROC curves of the performance of the CAD system on the training set with and without context features. In addition, to get an impression of the value of each feature group, we also added the FROC curves of the system when only one of the feature groups is used. The performance of the CAD system, measured by the partial area under the FROC curve between 0 and 8 FPs/case, is significantly increased when the context features are added to the system ( $p = 0.001$ ). The FROC curves of the single feature groups also confirm the value of the context features as they are the most important feature group after the intensity features.

#### 4.3. Combination with solid nodule CAD

The complete test set has been processed using the solid nodule CAD system, which was set to operate at an operating point of 4 false positives per scan. At this operating point, the solid nodule CAD detected of 55% of all subsolid nodules in the test set. The

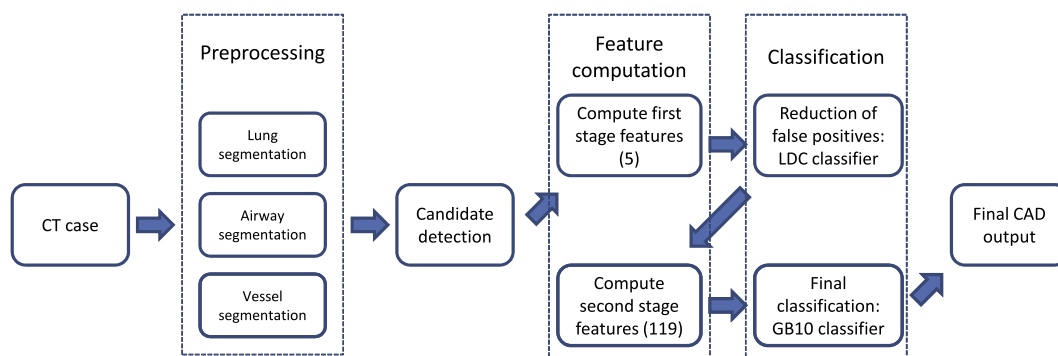


**Fig. 5.** FROC curves of the CAD system on the independent set for all subsolid nodules, part-solid nodules only and non-solid nodules only. The horizontal axis is logarithmic. The dotted lines indicate the maximal sensitivity which can be reached due to the candidate detection sensitivity for the different nodule groups.



**Fig. 6.** FROC curves of the CAD system with and without context features. In addition, the FROC curves when the CAD system is trained with only one single feature group are depicted. The FROC curves are obtained by a 10-fold cross-validation on the full training set. The horizontal axis is logarithmic.

detection rate for part-solid nodules and non-solid nodules was 71% and 42%, respectively. Furthermore, the solid nodule CAD detected 9 (6 part-solid and 3 non-solid) of the 23 subsolid nodules which were missed by the subsolid CAD system when operating at 1.0 FP/scan. Consequently, adding the solid nodule CAD system to the subsolid nodule CAD system increased the detection sensitivity for subsolid nodules from 80% to 88%.



**Fig. 4.** Schematic diagram of proposed CAD system.

## 5. Observer study

In this section, we describe an analysis of the output of the CAD system using an independent experienced chest radiologist as reference. Prior publications have shown that databases used for evaluation of CAD have its limitations (van Ginneken et al., 2010; Armato et al., 2011). Although the process of obtaining a reliable reference standard is a widely known problem when evaluating CAD systems, it is important to investigate the output of a CAD system and evaluate whether additional lung lesions are detected (Peldschus et al., 2005). The annotations of the database used for this study were obtained without the support of CAD and therefore, it is likely that CAD identified lesions not annotated in the screening trial. It remains open whether these nodules have been overlooked by the radiologists during screening or whether they were deemed not suspicious enough to be annotated, for example in case of small subsolid nodules. In order to see if CAD find additional lesions and to get an idea of the nature of false positives which were generated by the CAD system, all false positives of the CAD system when set to operate at an average of 1 FP/scan are selected for evaluation by the radiologist.

Prior research has shown that there is a lot of debate among radiologists as to what constitutes a pulmonary nodule. This is clearly illustrated by the study of the Lung Image Database Consortium (LIDC) (Armato et al., 2011). The database of this study is publicly available and contains 1018 thoracic CT scans. In this study, four experienced thoracic radiologists first independently reviewed the CT cases and recorded the nodules they found; the initial blinded-read phase. They categorized the nodules into three categories: “nodule  $\geq 3$  mm”, “nodule  $< 3$  mm”, “non-nodule  $\geq 3$  mm”. Then, in a second session, the radiologist could see the anonymized annotations of the other three radiologists and gave a final opinion about the nodules in the CT case. After both sessions, 2669 lesions were marked as “nodule  $\geq 3$  mm”. Of these, only 34.8% were marked by all radiologists and 29.1% were marked by a single radiologist only. Since the investigators of LIDC did not force consensus, they were able to show that radiologists disagree considerably about what constitutes a pulmonary nodule. In order to test whether there are debatable subsolid nodules within the annotated nodules of the test set, the independent radiologist was also confronted blindly with all subsolid nodules from the reference standard.

In total, the radiologist was thus confronted with 223 marks in random order; 109 false positives and 114 nodule annotations from the reference standard. An in-house developed, dedicated reading workstation was used in which the radiologist could easily navigate from mark to mark and inspect findings in all orthogonal planes simultaneously. All reading facilities of a usual reading workstation were present. Note that the radiologist read these marks in a blinded fashion. Based on experience from prior experiments, the following categories were defined: subsolid nodule, subsolid nodule  $< 5$  mm, solid nodule, scar, acute inflammation, interstitial lung disease (ILD), motion/pulse artifacts and other FP. The last category contains everything not belonging to the first groups and the radiologist could provide a comment with the mark. In addition, when a subsolid nodule was marked, the radiologist was also asked to indicate whether it was a part-solid or non-solid nodule. The results from this observer study are presented in Table 2.

## 6. Discussion

A fully automatic CAD system for automatic detection of subsolid nodules was described and extensively evaluated. The data set used in this work was collected from two sites of a large multi-cen-

ter lung cancer screening trial and was considerably larger than the data sets used by previous publications (Zhou et al., 2006; Ye et al., 2007; Tao et al., 2009). CAD systems trained and tested on data from the same scanner may show better performance than the actual performance of the CAD system on data from different scanners. To circumvent this problem, we decided to test the CAD system on data from a different site, which was acquired with a different scanner.

The CAD system is initiated with a lung, airway and vessel segmentation algorithm. These algorithms were successful on all scans from the training set and failed on only one scan of the test set. Consequently, no marks were generated on this scan in the test set and lesions on this scan were missed. Evaluation and improvement of these algorithms is beyond the scope of this paper.

The candidate detection procedure is started in the next step in which a robust and accurate nodule segmentation algorithm is integrated. As a basis for a robust performance of the CAD system, a sensitive candidate detection step is essential. As the results in Section 4.1 show, the sensitivity of the candidate detection step is high; 84% on the training set and 88% on the independent test set. We believe this is a sufficiently high sensitivity for the candidate detection step.

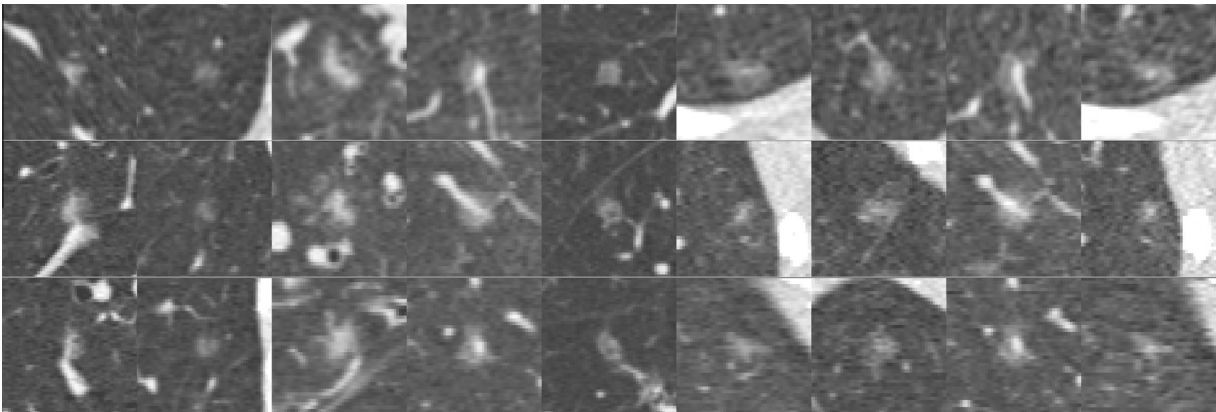
Subsequently, a rich set of features is defined for the candidate regions. In addition to intensity, shape and texture features, we added a novel group of context features. Contextual information is commonly used in classification problems in many fields. However, despite the large number of publications on CAD, contextual information is barely used in this area, with few exceptions such as the work of Sánchez et al. (2012), Song et al. (2012) and Hupse and Karssemeijer (2009). Hupse and Karssemeijer (2009) and Song et al. (2012) used specific contextual features for detection of masses in mammograms and detection of tumors and lymph nodes in thoracic images, respectively. Sánchez et al. (2012) presented a general framework for including contextual information and showed that this significantly improved the classification of two CAD applications: identification of exudates and drusen in 2D retinal images and coronary calcifications in 3D computed tomography scans. The contributions of our paper in terms of contextual classification are twofold. Firstly, we introduced novel context features dedicated to subsolid nodules and their relation to vasculature, pleural surface and the bronchial tree. We show that these features significantly increase the classification performance ( $p = 0.001$ ). Secondly, compared to Sánchez et al. (2012) we introduced a new class of context features which take the context of the complete image into account instead of only the context of individual candidates. The rationale behind this new class of features is that in a scan with many subsolid nodule candidates, it is more likely that the high number of candidates is caused by other factors such as interstitial lung disease, low inspiration level or poor image quality than that there are actually many subsolid nodules present within the scan.

In order to get the best performance out of all features, two different classification schemes and several different classifiers have been tested. To prevent a positively biased performance on the test set, the complete optimization of the CAD system is performed in 10-fold cross-validation on the training data. The FROC curves in Fig. 2 and Fig. 3 show that classification of candidate regions using all features achieves the best performance when the GB10 classifier is used. Therefore, this classifier was used in the final configuration of the CAD system. There was no clear difference in performance between the two classification schemes. Since the two stage classification scheme reduces the computation time of the CAD system, this is used in the final system. Although the reduction in computation time is only minor at this point, the benefit of the two stage approach might increase when the feature set is extended.

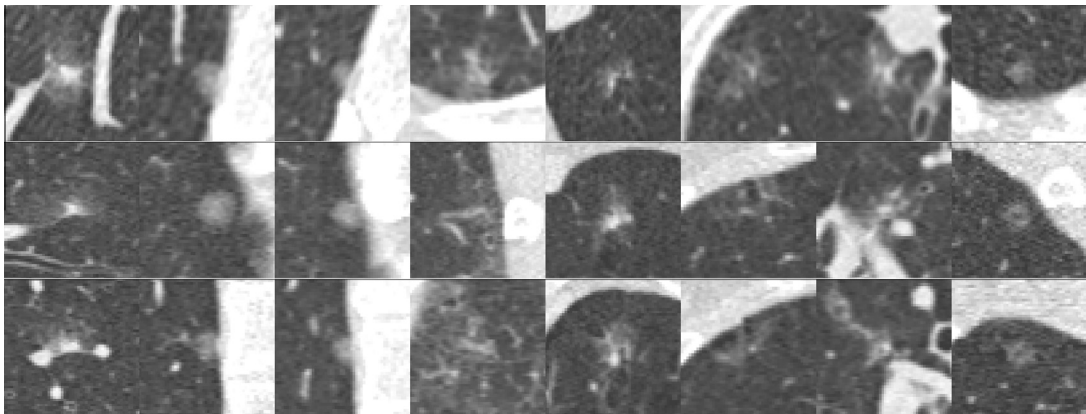


**Table 2**  
Analysis of the true positives, false negatives and false positives of the CAD system at 1FP/scan by an experienced thoracic radiologist. All marks are categorized into one of the 8 categories, represented by 8 columns in the table.

Mark type	Subsolid nodule	Subsolid nodule < 5 mm	Solid nodule	Inflammation	Scar	Interstitial lung disease	Motion/pulse artifacts	Other FP	Total number
True positive	75	0	2	0	6	0	0	8	91
False negative	10	0	4	0	2	0	0	7	23
False positive	44	6	1	0	5	2	4	47	109



**Fig. 7.** Subsolid nodules detected by CAD and confirmed by radiologist, but not present in reference standard. Every column shows one nodule. The top image displays an axial image, the middle image displays a sagittal image and the bottom image a coronal view. Images show a field of view of  $40 \times 40$  mm in which the nodule is centered.

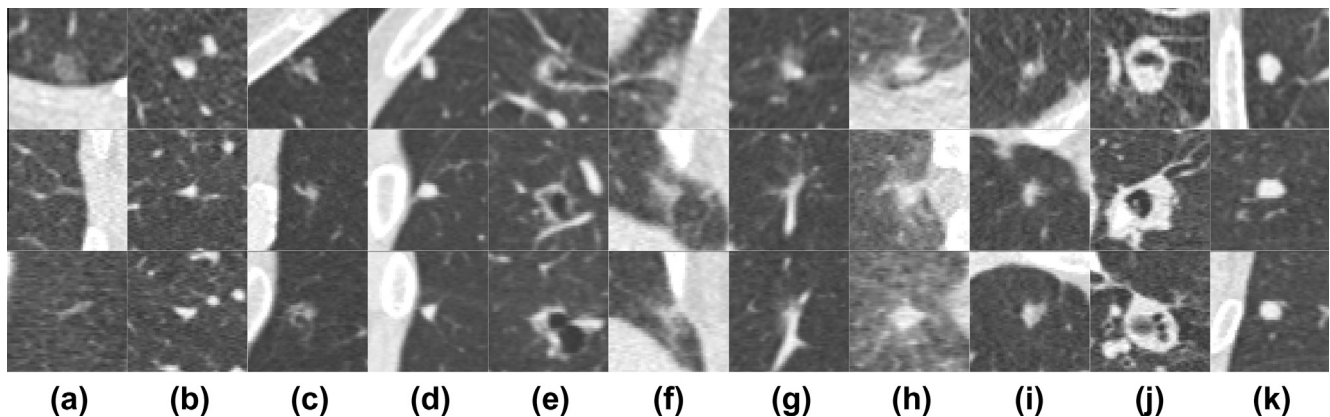


**Fig. 8.** Subsolid nodules missed by subsolid nodule CAD when set to operate at 1 FP/scan. Every column shows one nodule. The top image displays an axial image, the middle image displays a sagittal image and the bottom image a coronal view. Images show a field of view of  $40 \times 40$  mm in which the nodule is centered. Also note that the second and third nodule is the same lesion but at different time points.

The lack of performance difference between the one stage and two stage classification scheme using the GB classifier is an interesting observation which shows that the GB classifier intrinsically is able to handle unbalanced data sets. Boosting classifiers increase the weights of misclassified examples in each iteration step of the training and therefore, many obvious negative samples will soon get a low weight which effectively balances the data set. Therefore, no large improvement can be found for the GB classifier when comparing the one-stage versus two-stage classification. On the contrary, a substantial improvement in the two-stage classification situation can be seen for the LDC and NM classifier, which are known to suffer from unbalanced data sets. In the LDC for example, the estimate of the common covariance matrix of the two classes is a weighted mean of the two sample matrices. Therefore, it will be dominated by the variation of the prevalent class. Consequently, a

substantial bias may be present if the assumption of a common covariance matrix does not hold. CAD researchers typically invest considerable amount of time in optimizing the classification scheme and one aspect of this is to start experimenting with one stage versus multiple stage classification schemes. Next to that, different classifiers are usually tested. We show that the possible performance increase of a two-stage classification is dependent on the classifier, since certain classifiers are able to handle unbalanced data better than others. This should be taken into account while developing CAD algorithms.

In clinical practice, the subsolid nodule CAD system will operate in combination with a solid nodule CAD system. Most published and commercial nodule CAD systems generate between 2 and 4 false positives per scan. In order to control the false-positive rate of the combined system, good sensitivity at a low false-positive



**Fig. 9.** Subsolid nodules in the reference standard which are not classified as subsolid nodules by the independent radiologist. Every column shows one nodule. The top image displays an axial image, the middle image displays a sagittal image and the bottom image a coronal view. Images show a field of view of  $40 \times 40$  mm in which the nodule is centered. The categories which were chosen by the experienced radiologist for these nodules were: (a) other FP: plate atelectasis, (b) other FP: perifissural nodule (PFO), (c) other FP: small complex lesion next to bullae, (d) other FP: PFO, (e) other FP: small complex lesion next to bullae, (f) scar, (g) scar, (h–k) solid nodule

rate is desired. The FROC analysis in Fig. 5 shows that the proposed subsolid CAD system reaches a sensitivity of 80% on our independent test set at a low false-positive rate of 1 FP/scan. Given that the system will be combined with a solid nodule CAD, we believe that this is a suitable operating point of our subsolid nodule CAD system. As Fig. 5 also shows, the performance for part-solid nodules is better than for non-solid nodules. The most likely reason for this is that the solid component in the part-solid nodule is well described by our feature set and this benefits the classification process.

Furthermore, we explored whether a solid nodule CAD system would be able to detect the missed nodules of the subsolid CAD system. At an average of 1 FP/scan, the subsolid nodule CAD reaches a 80% sensitivity which means that it misses 23 of the 114 annotations in the test set. As described in Section 4.3, the solid nodule CAD detected 9 of the 23 missed subsolid nodules when set to operate at 4 FPs/scan. Consequently, the sensitivity would increase from 80% to 88% which represents an important improvement. To get an idea which type of nodules we still miss, 7 of the 18 missed subsolid nodules are depicted in Fig. 8.

Finally, an experienced thoracic radiologist scored all findings in the reference standard as well as all CAD marks when the CAD is set to operate at an average of 1 FP/scan. Table 2 shows that 44 from the 109 (40%) false positives CAD marks were retrospectively marked as subsolid nodules. These nodules were initially not annotated during the screening. These results show that a subsolid nodule CAD system may detect subsolid nodules overseen by radiologists. In Fig. 7, 9 of these false-positive CAD marks which were retrospectively classified as subsolid nodule are shown. Furthermore, Table 2 also shows that from the 23 missed subsolid nodules, a considerable amount (13) were not marked as subsolid nodule by the radiologist. This indicates that these nodules are debatable. In Fig. 9, examples of these 13 debatable nodules are shown.

In future work, we intend to further explore how to optimally combine the proposed subsolid nodule CAD system with a solid nodule CAD and evaluate this on a large data set containing both solid and subsolid nodules. Then, robust detection of the complete spectrum of pulmonary nodules is at hand.

## 7. Conclusions

In this work, a fully automatic CAD system for detection of subsolid nodules was presented. The data set used in this work was collected from a large multi-center lung cancer screening trial

and is considerably larger than the data sets used in previous studies. A novel set of context features is introduced which describe the relation of a nodule candidate to the lung boundary, airways, vessels and other nodule candidates. Using experiments, we have shown that these features significantly improve the classification performance of the CAD system. The CAD system reached 80% sensitivity on the independent set at an average of 1.0 false positive per scan. We believe this is an appropriate operating point because in clinical practice, a subsolid nodule CAD will be used in combination with a solid nodule CAD, which usually generates between 2 and 4 FPs/scan. When the subsolid nodule CAD system was combined with a previously published solid nodule CAD, the sensitivity for subsolid nodules increased to 88% on the independent test set. An extensive evaluation of the CAD output using an experienced thoracic radiologist showed that a substantial fraction of the false positives of the system when operating at 1.0 false positive per scan were actually considered to be subsolid nodules which were missing in the reference standard.

## Acknowledgements

We thank the investigators of the NELSON trial for providing data for this study. Funding for this research was provided by MeVis Medical Solutions AG.

## References

- Aberle, D.R., Adams, A.M., Berg, C.D., Black, W.C., Clapp, J.D., Fagerstrom, R.M., Gareen, I.F., Gatsonis, C., Marcus, P.M., Sicks, J.D., 2011. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* 365, 395–409.
- American Cancer Society, 2012. Cancer Facts and Figs. 2012.
- Armato, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., Kazerooni, E.A., MacMahon, H., Beek, E.J.R.V., Yankelevitz, D., Biancardi, A.M., Bland, P.H., Brown, M.S., Engelmann, R.M., Laderach, G.E., Max, D., Pais, R.C., Qing, D.P.Y., Roberts, R.Y., Smith, A.R., Starkey, A., Batrah, P., Caligiuri, P., Farooqi, A., Gladish, G.W., Jude, C.M., Munden, R.F., Petkovska, I., Quint, L.E., Schwartz, L.H., Sundaram, B., Dodd, L.E., Fenimore, C., Gur, D., Petrick, N., Freymann, J., Kirby, J., Hughes, B., Castele, A.V., Gupte, S., Sallamm, M., Heath, M.D., Kuhn, M.H., Dharaiya, E., Burns, R., Fryd, D.S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B.Y., 2011. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med. Phys.* 38, 915–931.
- Beigelman-Aubry, C., Hill, C., Boulanger, X., Brun, A., Leclercq, D., Golmard, J., Grenier, P., Lucidarme, O., 2009. Evaluation of a computer aided detection system for lung nodules with ground glass opacity component on multidetector-row CT. *J. Radiol.* 90, 1843–1849.
- Bellotti, R., Carlo, F.D., Gargano, G., Tangaro, S., Cascio, D., Catanzariti, E., Cerello, P., Cheran, S.C., Delogu, P., Mitri, I.D., Fulcheri, C., Grosso, D., Retico, A., Squarcia, S., Tommasi, E., Golosio, B., 2007. A CAD system for nodule detection in low-dose

- lung CTs based on region growing and a new active contour model. *Med. Phys.* 34, 4901–4910.
- Breiman, L., 2001. Random forests. *Machine Learn.* 45, 5–32.
- Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 21–27.
- Dehmshki, J., Ye, X., Lin, X., Valdivieso, M., Amin, H., 2007. Automated detection of lung nodules in CT images using shape-based genetic algorithm. *Comput. Med. Imag. Graph.* 31, 408–417.
- van Dongen, E., van Ginneken, B., 2010. Automatic segmentation of pulmonary vasculature in thoracic CT scans with local thresholding and airway wall removal. In: *IEEE Int. Symp. Biomedical Imaging*, pp. 668–671.
- Enquobahrie, A.A., Reeves, A.P., Yankelevitz, D.F., Henschke, C.I., 2007. Automated detection of small pulmonary nodules in whole lung CT scans. *Acad. Radiol.* 14, 579–593.
- Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998. Multiscale vessel enhancement filtering. In: *Med. Image Comput. Comput. Assist. Interv.*, pp. 130–137.
- Friedman, J., Hastie, T., Tibshirani, R., 2000. Special invited paper. additive logistic regression: a statistical view of boosting. *Ann. Stat.* 28, 337–374.
- Fukunaga, K., 1990. *Introduction to Statistical Pattern Recognition*, second ed. Academic Press.
- van Ginneken, B., Armato, S.G., de Hoop, B., van de Vorst, S., Duindam, T., Niemeijer, M., Murphy, K., Schilham, A.M.R., Retico, A., Fantacci, M.E., Camarlinghi, N., Bagagli, F., Gori, I., Hara, T., Fujita, H., Gargano, G., Bellotti, R., Carlo, F.D., Megna, R., Tangaro, S., Bolanos, L., Cerello, P., Cheran, S.C., Torres, E.L., Prokop, M., 2010. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. *Med. Image Anal.* 14, 707–722.
- van Ginneken, B., Baggerman, W., van Rikxoort, E.M., 2008. Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In: *Med. Image Comput. Comput. Assist. Interv.*, pp. 219–226.
- Godoy, M.C.B., Naidich, D.P., 2009. Subsolid pulmonary nodules and the spectrum of peripheral adenocarcinomas of the lung: recommended interim guidelines for assessment and management. *Radiology* 253, 606–622.
- Hansell, D.M., Bankier, A.A., MacMahon, H., McCloud, T.C., Müller, N.L., Remy, J., 2008. Fleischner society: glossary of terms for thoracic imaging. *Radiology* 246, 697–722.
- Heitmann, K.R., Kauczor, H., Mildenberger, P., Uthmann, T., Perl, J., Thelen, M., 1997. Automatic detection of ground glass opacities on lung HRCT using multiple neural networks. *Eur. Radiol.* 7, 1463–1472.
- 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.
- Henschke, C.I., Yankelevitz, D.F., Mirtcheva, R., McGuinness, G., McCauley, D., Miettinen, O.S., 2002. CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. *AJR Am. J. Roentgenol.* 178, 1053–1057.
- Hu, M.K., 1962. Visual pattern recognition by moment invariants, computer methods in image analysis. *IRE Trans. Inform. Theory* 8, 179–187.
- Hupse, R., Karssemeijer, N., 2009. The use of contextual information for computer aided detection of masses in mammograms. In: *Medical Imaging*, p. 72600Q.
- Jacobs, C., Sánchez, C.I., Saur, S.C., Twellmann, T., de Jong, P.A., van Ginneken, B., 2011. Computer-aided detection of ground glass nodules in thoracic CT images using shape, intensity and context features. In: *Med. Image Comput. Comput. Assist. Interv.*, pp. 207–214.
- Jobson, J.D., 1992. *Applied Multivariate Data Analysis*. Springer-Verlag.
- Kauczor, H.U., Heitmann, K., Heussel, C.P., Marwede, D., Uthmann, T., Thelen, M., 2000. Automatic detection and quantification of ground-glass opacities on high-resolution CT using multiple neural networks: comparison with a density mask. *AJR Am. J. Roentgenol.* 175, 1329–1334.
- Kim, K.G., Goo, J.M., Kim, J.H., Lee, H.J., Min, B.G., Bae, K.T., Im, J.C., 2005. Computer-aided diagnosis of localized ground-glass opacity in the lung at CT: initial experience. *Radiology* 237, 657–661.
- van Klaveren, R.J., Oudkerk, M., Prokop, M., Scholten, E.T., Nackaerts, K., Vernhout, R., van Iersel, C.A., van den Bergh, K.A.M., van 't Westeinde, S., van der Aalst, C., Thunnissen, E., Xu, D.M., Wang, Y., Zhao, Y., Gietema, H.A., de Hoop, B., Groen, H.J.M., de Bock, G.H., van Ooijen, P., Weenink, C., Verschakelen, J., Lammers, J.J., Timens, Willebrand, D., Vink, A., Mali, W., de Koning, H.J., 2009. Management of lung nodules detected by volume CT scanning. *N. Engl. J. Med.* 361, 2221–2229.
- Kuhnigk, J.M., Dicken, V., Bornemann, L., Bakai, A., Wormanns, D., Krass, S., Peitgen, H.O., 2006. Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. *IEEE Trans. Med. Imag.* 25, 417–434.
- Li, Q., Li, F., Doi, K., 2008. Computerized detection of lung nodules in thin-section CT images by use of selective enhancement filters and an automated rule-based classifier. *Acad. Radiol.* 15, 165–175.
- Marten, K., Engelke, C., 2007. Computer-aided detection and automated CT volumetry of pulmonary nodules. *Eur. Radiol.* 17, 888–901.
- Messay, T., Hardie, R.C., Rogers, S.K., 2010. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Med. Image Anal.* 14, 390–406.
- Murphy, K., van Ginneken, B., Schilham, A.M.R., de Hoop, B.J., Gietema, H.A., Prokop, M., 2009. A large scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification. *Med. Image Anal.* 13, 757–770.
- Naidich, D.P., Bankier, A.A., MacMahon, H., Schaefer-Prokop, C.M., Pistolesi, M., Goo, J.M., Macchiarini, P., Crapo, J.D., Herold, C.J., Austin, J.H., Travis, W.D., 2013. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the fleischner society. *Radiology* 266, 304–317.
- Ojala, T., Pietikainen, M., Maenpää, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Machine Intell.* 24, 971–987.
- Peldschus, K., Herzog, P., Wood, S.A., Cheema, J.I., Costello, P., Schoepf, U.J., 2005. Computer-aided diagnosis as a second reader: spectrum of findings in CT studies of the chest interpreted as normal. *Chest* 128, 1517–1523.
- Pudil, P., Novovicova, J., Kittler, J., 1994. Floating search methods in feature selection. *Pattern Recognit. Lett.* 15, 1119–1125.
- Retico, A., Delogu, P., Fantacci, M.E., Gori, I., Martinez, A.P., 2008. Lung nodule detection in low-dose and thin-slice computed tomography. *Comput. Biol. Med.* 38, 525–534.
- van Rikxoort, E.M., de Hoop, B., Viergever, M.A., Prokop, M., van Ginneken, B., 2009. Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Med. Phys.* 36, 2934–2947.
- Sánchez, C.I., Niemeijer, M., Isgum, I., Dumitrescu, A.V., Suttrop-Schulten, M.S.A., Abràmoff, M.D., van Ginneken, B., 2012. Contextual computer-aided detection: Improving bright lesion detection in retinal images and coronary calcification identification in CT scans. *Med. Image Anal.* 16, 50–62.
- Song, Y., Cai, W., Kim, J., Feng, D., 2012. A multi-stage discriminative model for tumor and lymph node detection in thoracic images. *IEEE Trans. Med. Imag.* 31, 1061–1075.
- Sörensen, L., Shaker, S.B., de Bruijne, M., 2008. Texture classification in lung CT using local binary patterns. In: *Med. Image. Comput. Comput. Assist. Interv.*, pp. 934–941.
- Sousa, J.R., Silva, A.C., de Paiva, A.C., Nunes, R.A., 2009. Methodology for automatic detection of lung nodules in computerized tomography images. *Comput. Methods Programs Biomed.* 98, 1–14.
- Tan, M., Deklerck, R., Jansen, B., Bister, M., Cornelis, J., 2011. A novel computer-aided lung nodule detection system for CT images. *Med. Phys.* 38, 5630–5645.
- Tao, Y., Lu, L., Dewan, M., Chen, A.Y., Corso, J., Xuan, J., Salganicoff, M., Krishnan, A., 2009. Multi-level ground glass nodule detection and segmentation in CT lung images. In: *Med. Image Comput. Comput. Assist. Interv.*, pp. 715–723.
- Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.
- Ye, X., Lin, X., Beddoe, G., Dehmshki, J., 2007. Efficient computer-aided detection of ground-glass opacity nodules in thoracic CT images. In: *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, pp. 4449–4452.
- Ye, X., Lin, X., Dehmshki, J., Slabaugh, G., Beddoe, G., 2009. Shape based computer-aided detection of lung nodules in thoracic CT images. *IEEE Trans. Biomed. Eng.* 56, 1810–1820.
- Zhou, J., Chang, S., Metaxas, D.N., Zhao, B., Ginsberg, M.S., Schwartz, L.H., 2006. An automatic method for ground glass opacity nodule detection and segmentation from CT studies. In: *IEEE EMBS*, pp. 3062–3065.