# Training a Computer Aided Detection System with Simulated Lung Nodules in Chest Radiographs

Peter R. Snoeren, Geert J.S. Litjens, Bram van Ginneken, and Nico Karssemeijer

Radboud University Nijmegen Medical Centre
Nijmegen, The Netherlands
P.Snoeren@rad.umcn.nl

**Abstract.** This paper addresses the hypothesis that artificially implanted lung nodules from computed tomography exams (CT exams) into chest radiographs can improve the performance of a computer aided detection system (CAD system). Twenty-four three-dimensional lung nodules were segmented and projected in five directions, mimicking 120 X-rayed nodules. The CAD system was tested by fivefold cross validation on a publicly available database. The results were evaluated by a free-response receiver operating characteristic analysis (FROC). It was found that the performance of the CAD system trained with simulated nodules comes close to the performance of state of the art CAD systems that are trained with real nodules. The CAD system trained with real nodules did improve by adding simulated nodules, but only when there were few real nodules used for training.

## 1 Introduction

Chest radiography is often the first study performed on symptomatic patients and on asymptomatic patients in screening. While it is a cost-effective imaging tool it has the drawback that images are hard to interpret, even for trained radiologists. After some decades of development computer aided detection (CAD) is becoming a widespread tool in medical diagnosis to assist radiologists. It is obvious that a radiologist in training improves when he has diagnosed more radiographs with feedback, with different kinds of diseases, with abnormalities on different locations, and under different circumstances. That is not very different for CAD systems. A CAD system's performance also improves when training data are richer in content and amount. A CAD system must first of all be well-designed, but it also needs to be trained with many normal and abnormal images before it can be applied successfully. Many papers have been written about the design of CAD systems [1,2,3,4], in most cases silently assuming that training with a large training database is possible. Unfortunately though, large training databases are rare. This paper describes the results of a feasibility study with artificially implanted lung nodules to increase the size of training databases.

| Stratum | Number |
|---|---|
| Obvious | 12 |
| Relatively obvious | 38 |
| Subtle | 48 |
| Very subtle | 24 |
| Extremely subtle | 19 |
| Total | 141 |

**Table 1. Subtlety rating of abnormal JSRT images.**

A small number of three-dimensional lung nodules were segmented in CT volumes. Projection of these volumes in different directions mimics X-rayed CT nodules that can be implanted in chest radiographs. This way a training database can be extended with abnormalities at will. Simulated nodules can for example be placed at locations with few training data, e.g., at locations that are partly obscured by other organs, for separately training the CAD system. They can be systematically varied in size and/or contrast by scaling or placed on the edges of ribs or just in between. To prove the concept, a single contrast and the real projected size of the simulated nodules were used in this paper. Furthermore, the location variation is limited by locations of contralateral real nodules on the same radiograph.

The following describes the segmentation of CT nodules, the construction of simulated nodules, the implanting of these simulated nodules in chest radiographs, and the CAD system that is used to test the hypothesis that implanted nodules can in principle improve CAD systems. An FROC analysis is used to objectively evaluate the CAD system with differently enriched training data.

## 2 Materials and Methods

### 2.1 Databases

**Chest Radiography Database.** The publicly available database of the Scientific Committee of the Japanese Society of Radiological Technology (JSRT, Shiraishi et al. [5]) was used for training and testing the CAD system. The performance of CAD systems on this database often serves as a benchmark for comparison between CAD systems. At present, it is the only publicly available database of chest radiographs of moderately large size with normals and proven abnormals (lung nodules). The JSRT database consists of 247 Posterior-Anterior PA chest radiographs, digitized at a spatial resolution of 175 $\mu m$/pixel and 12 bit gray levels. 154 images contain exactly one pulmonary lung nodule; the other 93 images contain no lung nodules. Additionally, a publicly available database of lung field segmentations of the JSRT-images was used. This database consists of manually segmented unobscured lung fields (only the parts that are unobscured by the diaphragm or heart), heart, and clavicles (van Ginneken et al. [6]). Thirteen images with lung nodules that were not part of the unobscured lung fields

(a) JSRT nodules

| Stratum | Number |
|---|---|
| $< 10\,mm$ | 33 |
| $\geq 10\,mm$ and $< 20\,mm$ | 82 |
| $\geq 20\,mm$ | 26 |
| Total | 141 |

(b) Simulated nodules

| Stratum | Number |
|---|---|
| $< 10\,mm$ | 6 |
| $\geq 10\,mm$ and $< 20\,mm$ | 68 |
| $\geq 20\,mm$ | 43 |
| Total | 117 |

**Table 2. Diameter of Lung Nodules.** The total number of implanted CT nodules is 117. The diameter of simulated nodules is defined by the diameter of a disk with 95% of the total 'mass'.

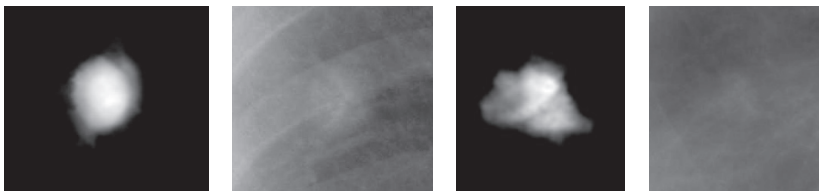
were removed from the dataset. Before features were computed the images were downsampled to $350\,\mu m$/pixel to speed up processing. Table 1 gives the subtlety rating of the abnormal JSRT images and Table 2(a) gives the effective diameters of the nodules in those images.

**Computed Tomography Database.** CT volumes in an anonymized database from a lung cancer screening study were used. To minimize influence of inaccurate segmentation in our simulations, only volumes with nodules that are not in contact with large blood vessels were selected from this database. Twenty-four lung nodules were annotated by a radiologist.

The CT volumes were acquired on a 16-slice Philips Brilliance scanner with a slice thickness of $1\,mm$ and a slice spacing of $0.7\,mm$. The in-plane resolution was between $0.6\,mm$ and $0.8\,mm$.

### 2.2 Simulated Nodules from Computed Tomography

**X-rayed CT Nodules.** Twenty-four CT nodules were segmented using a method by Kostis et al. [7]. This is basically a region growing algorithm with a refinement by morphological operators. It leads to segmented nodule templates that are rescaled to the resolution of the radiographs and placed in a box of air ($-1000\,HU$). Subsequently, nodules were projected along five randomly oriented axes using simple orthogonal ray casting. This leads to 120 two-dimensional CT intensity profiles. Three simulated nodules that looked unrealistic were not used, resulting in 117 simulated nodules. Table 2(b) gives the sizes of the 117 simulated nodules. The last step is a conversion from the projected CT units to radiograph units. This is done by a calibration obtained from studies with both a CT exam and a radiography exam [8]. The calibration function is determined by measuring the intensities in corresponding regions of the radiographs and the projected CTs. The intensities were measured in the heart, the spine, and in the lungfields. A calibration function was constructed by fitting a third-order polynomial to these data. It will not completely undo the effects of all post-processing steps that manifactures apply, like edge detection or unsharp masking, but it was adequate for the purpose of this study.



(a) (b) (c) (d)

**Fig. 1. Examples of Implanted Nodules.** Fig.s 1(a) and 1(c) depict two simulated nodules and Fig.s 1(b) and 1(d) depict parts of chest radiographs with these nodules implanted.

**Implanting X-rayed CT Nodules.** The CAD system is trained with both normal and abnormal images. We implanted exactly one simulated nodule in the abnormal images. The location was chosen in the contralateral lung field of the real nodule. (Remember that each abnormal image in the JSRT database contains exactly one real nodule.) Hence, the total number of available lung nodules for training (real and implanted) doubles. The vertical locations of the real and implanted nodules are chosen approximately the same and also their horizontal distances from the vertebral column are chosen approximately the same. The reason for this careful choice of placement is that we test the hypothesis that adding artificially implanted lung nodules to a training database improves the performance of a CAD system. If locations were carelessly chosen, e.g., by random placement, we might have impaired the CAD performance beforehand.

After projection of the CT nodules there is still one degree of freedom, namely the orientation, which is chosen to be random.

141 simulated lung nodules are sampled from the set of 117 simulated nodules with replacement and assigned to the 141 JSRT chest radiographs with a real lung nodule. For a wide range of pixel values the pixel value is approximately proportional to the thickness of tissue in film-screen radiography. Implanting nodules is therefore nothing more than adding pixel values. Figure 1 depicts two examples of simulated nodules.

### 2.3 CAD system

**Local Direction Features.** Detection is mainly based on two *local direction features* features: the first measures whether the number of local gradients pointing to a common center is statistically larger than the expected number for random gradients; the second measures whether local gradients are rotation symmetric.

If many gradient vectors are directed to a certain location this indicates that a lung nodule may be present. To quantify such events we define the statistic

$$s_{i,j} = \begin{cases} 1 - p_{i,j}, & \text{if “hit”, (pixel } j \text{ is oriented to pixel } i) \\ -p_{i,j} & \text{if “miss”.} \end{cases}$$

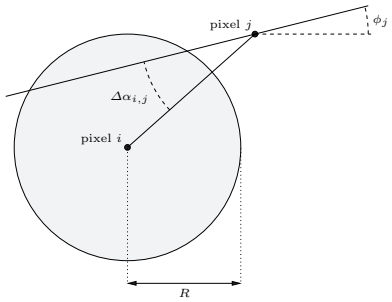



**Fig. 2. Hit/Miss.** A pixel $j$ is oriented to pixel $i$ when the line through $\mathbf{r}_j$ with orientation $\phi_j$ crosses a disk with radius $R$ centered at pixel $i$.

A pixel $j$ is oriented to pixel $i$, i.e., we speak of a "hit", when the line through the location of pixel $j$, denoted by $\mathbf{r}_j$, with orientation $\phi_j \in [0, \pi]$ crosses a disk with radius $R$ centered at $\mathbf{r}_i$ (see Figure 2). This happens when

$$\sin \Delta\alpha_{i,j} \leq \frac{R}{\|\mathbf{r}_j - \mathbf{r}_i\|}, \tag{1}$$

where $\Delta\alpha_{i,j} \in [0, \pi/2]$ is the acute angle between a line with orientation $\phi_j$ and the line through $\mathbf{r}_i$ and $\mathbf{r}_j$. The prior probability that a line through $\mathbf{r}_j$ is oriented to pixel $i$ is given by $p_{i,j}$. Assuming that the prior probability of orientations $\phi_j$ is uniform on $[0, \pi]$, we have

$$p_{i,j} = \begin{cases} \frac{2}{\pi} \arcsin \frac{R}{\|\mathbf{r}_j - \mathbf{r}_i\|}, & \text{if } \|\mathbf{r}_j - \mathbf{r}_i\| \geq R, \\ 1 & \text{otherwise.} \end{cases} \tag{2}$$

A sum is computed by

$$S_i = \sum_{j \in \mathcal{N}_i} s_{i,j}$$

where an annular neighborhood $\mathcal{N}_i$ with outer radius $D$ and inner radius $D_{\min}$ is defined as

$$\mathcal{N}_i = \left\{ \forall\, \mathbf{r}_j \in \mathbb{R}^2 \,|\, R < D_{\min} \leq \|\mathbf{r}_j - \mathbf{r}_i\| \leq D \right\}.$$

The first feature, the *concentration feature*, is defined as

$$g_1 = \frac{S_i}{\operatorname{std}\{S_i\}} = \frac{\sum_{j \in \mathcal{N}_i} s_{i,j}}{\sqrt{\sum_{j \in \mathcal{N}_i} p_{i,j}\,(1 - p_{i,j})}}. \tag{3}$$

The normalization $\operatorname{std}\{S_i\}$ in the denominator is computed by assuming that all orientations at $j \in \mathcal{N}_i$ are uniformly distributed and statistically independent. For random orientations this feature is zero, and only when orientations systematically have a common center at $i$ (or the opposite, not one is oriented to $i$) it deviates significantly from zero.

(a) X-location (b) Y-location (c) Heart border (d) Lung border

**Fig. 3. Location Features within the Unobscured Lungs Fields.**

The second feature takes the symmetricity of orientation patterns into account. Each neighborhood $\mathcal{N}_i$ is divided into $K$ directional bins, i.e., like a pie ($K = 24$ is used). For each bin the statistic $S_i$ is computed as before, but now with $\mathcal{N}_i$ replaced by the region of one piece of the pie. When $S_i > 0$ for all bins, there is evidence that the pattern of orientations is symmetric around pixel $i$. The second feature, the *concentration symmetricity feature*, is defined as

$$g_2 = \frac{n_+ - K_i}{\sqrt{K_i/4}}, \tag{4}$$

where $K_i \leq K$ is the number of bins at pixel $i$, and $n_+ \in [0, K_i]$ is the number of bins with positive $S_i$. The number of bins is variable as pixels close to the breast or image boundary are excluded if they do not contain enough pixels. The denominator is the standard deviation of $n_+$ when all orientations are uniformly distributed and statistically independent, i.e., when $n_+$ is a binomial random variate (with sample size $K = 24$ and probability $p = 1/2$ that bin $i$ has a positive statistic $S_i$).

Note that when the summations above are computed for a certain upper radius $D$, it takes little effort to compute the summations for a somewhat larger upper radius. This means that $g_1$ and $g_2$ can be computed efficiently for all encountered radii $D$ in a rectangular grid of pixels. Instead of using $g_1$ and $g_2$ for fixed upper radius $D$, we use the multi-scale features after computing them as a function of $D$

$$G_1 = \max_D \{g_1(D)\} \tag{5}$$

and

$$G_2 = \max_D \{g_2(D)\}. \tag{6}$$

Before determining the maxima the functions $g_i(D)$ are smoothed.

**Location Features.** Because nodules are not uniformly distributed over the lungs and, more importantly, because the local direction features suffer from the vessels near the heart, four *location features* are added: the relative X- and Y-coordinates within the pixel's unobscured lung field (computed by the cumulative area horizontally or vertically); the relative distance to the heart; and the relative

distance to the border of the unobscured lung fields (see Figure 3). The four location features are normalized between zero and one.

**Classifier.** We used a three-layer feedforward neural network with sigmoid transfer functions and one hidden layer to classify feature vectors. The number of input nodes was equal to the length of the feature vector. The number of nodes in the hidden layer was five and one output node was used. The network is trained by the backpropagation algorithm.

The neural network output was computed on regularly spaced grid locations (the grid spacing was $2.8\,mm$) in all testing images and thresholded local maxima of the resulting output images were further analyzed. A local maximum of an output image was considered a true-positive (TP) when its location was less than $2.5\,cm$ from the center of any annotated lesion, otherwise the local maximum was considered a false-positive (FP). This criterion was proposed by Hardie et al. [4].

In most CAD systems a second stage is employed in which candidate nodules detected in the first stage are segmented and classified with more features. This study is limited to investigating the use of simulated nodules in the initial detection stage.

**Training and Testing.** Lung nodule images in the JSRT database are classified into six strata according to the degrees of subtlety shown, ranging from "obvious" to "extremely subtle" and "no lung nodules (normals)" (taken from Shiraishi et al. [5]). The numbers of images per degree of subtlety are given in Table 1. Stratified fivefold cross-validation was used to partition the dataset into training sets and testing sets.

Let us call the set with testing images $S_{20}$ (20% of all normal and abnormal JSRT images) and the disjoint set with training images $S_{80}$ (80% of all normal and abnormal JSRT images). The pixel classifier is trained with the set $S_{80}$ (with existing and implanted lung nodules) and tested with the disjoint set $S_{20}$ (only with existing lung nodules). This is repeated with circulating sets $S_{20}$ until all images have been tested, i.e., five times. The five sets $S_{80}$ are randomized once and remain the same in all experiments thereafter. The only thing that differs between experiments are the regions of interest (ROI), comprising the set of pixels used for training. When a nodule, irrespectively whether it is a real or a simulated nodule, is not to be used in a certain experiment then it is removed from the ROI. This is done by removing all pixels in a disk with a diameter of $4\,cm$ centered at the nodule's location from the ROI. When $N_{real}$ real nodules and $N_{implanted}$ simulated nodules are to be used for training then all nodules are blanked, except the real nodules in the first $N_{real}$ images and the simulated nodules in the following $N_{implanted}$ images. Of course, the ROIs of the testing sets $S_{20}$ are not manipulated at all. For experiments with a constant of $N_{real} + N_{implanted}$ lung nodules in the training set, the only thing that may be varied is the relative amount of implanted lung nodules in the training sets,

i.e., the blanked nodules; The sets of training and test images remain the same under all conditions.

## 3 Experimental Results

**Location Features and Local Direction Features.** The main experiment is performed with different numbers of real nodules and simulated nodules. For this experiment all six features are included. The results, evaluated by a free-response receiver operating characteristic analysis (FROC; [9]), are given in Figure 4.

In the next section a control experiment is performed to test whether these results can be explained by the way the simulated nodules were implanted.

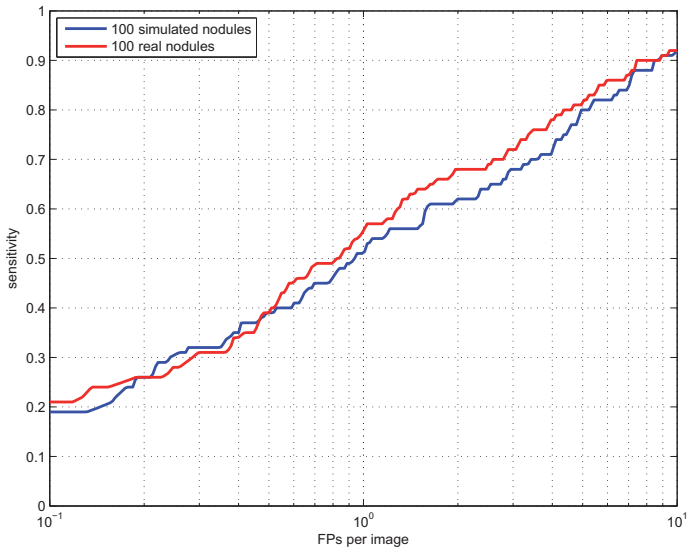
**Location Features.** To show that the previous results cannot be explained by the locations of the implanted lung nodules alone, a control experiment is performed with only location features. The results are given in the FROC curves of Figure 5. Evidently, the curves lie much lower than the curves in Figure 4(a) (e.g., at one FP per image the sensitivity drops from 0.56 to 0.14 for the real nodules). This proves that the previous results cannot be explained by location features alone. In other words, the morphology of the simulated nodules adds something essential.

The slight difference between the two curves can be explained as follows. Although the relative locations of the implanted nodules are the same as the relative locations of the real nodules, the distances to the heart and the distances to the unobscured lung border, also two location features, are not.

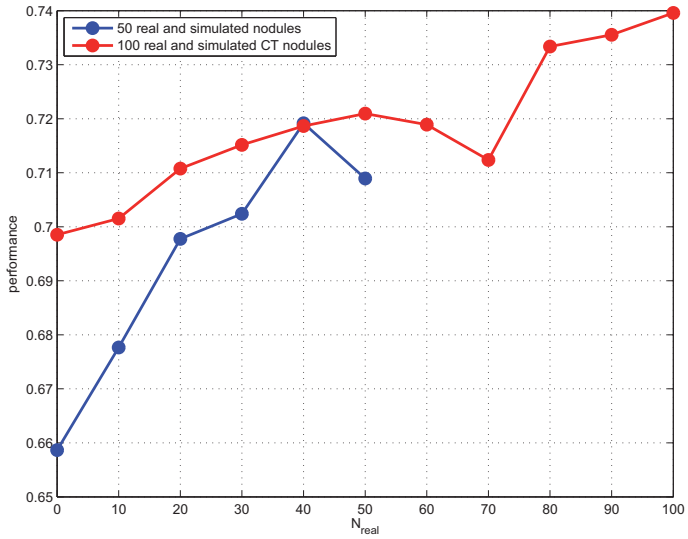
## 4 Discussion

**Conclusion.** This paper presents the preliminary results of a study with implanted lung nodules for the purpose of training CAD systems. The hypothesis that a CAD system improves with artificially added X-rayed CT nodules is confirmed. Even when no real nodules are in the training set, the performance of the CAD system is reasonably high, certainly if one considers that only twenty-four three-dimensional CT nodules form the basis of the implants. The sensitivity of the CAD system trained with 100 simulated nodules was 0.51 at one FP per image. The best results reported to date were obtained by Hardie et al. [4]. With the same criterion for TPs, they found a sensitivity of 0.56 at one FP per image (tenfold cross validation on the JSRT database). The CAD system they used was however trained with more than 100 abnormal images, but more importantly it included an extra classifier for candidate lung nodules, an extra step that was omitted here.

**Future Work.** In an extended CAD scheme the candidate nodules are further processed in a subsequent, second classifier. It will be investigated whether such a CAD schemes will increase the performance (which is already relatively high).




(a) FROC curves




(b) Performance with $N_{real} + N_{implanted} = 50$ and 100

**Fig. 4. FROC analysis of a CAD system with Location Features and Local Direction Features.** Figure 4(a) shows the FROC curves when the CAD system is either trained with either $N_{real} = 100$ real nodules or with $N_{implanted} = 100$ simulated nodules. Figure 4(b) shows aggregated results for other numbers of real and simulated nodules by integrals over the FROC curves with respect of $\log_{10}$ (FPs per image) from 0.1 to 4 FPs per image. Notably, the red points at $N_{real} = 0$ and 100 are the aggregate performances of Figure 4(a).

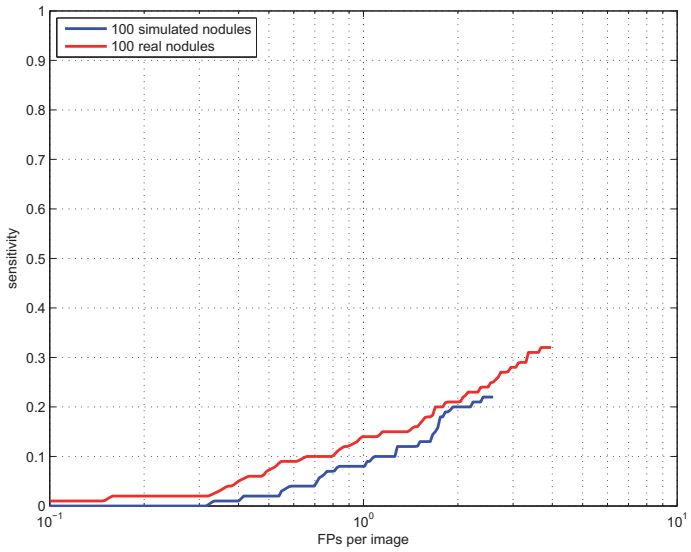



**Fig. 5. FROC analysis of a CAD system with Location Features only.** The CAD system is either trained with $N_{real} = 100$ real nodules or with $N_{implanted} = 100$ simulated nodules.

Probably more than twenty-four CT nodules (the number used in the current study) are required to capture the more subtle properties that nodules may possess.

The locations of implants were realistic locations in this work, because it is our believe that the location of a nodule is critical for its detection. This assumption will be investigated. If it is true that locations of implants are critical for training then a generic, smooth distribution of nodule locations may be constructed from a database of real nodules.

It would be very interesting to see whether CAD results improve when chest radiographs are split in different parts, and each trained with a different classifier. The first thing that comes in mind are the occluded regions, the heart, the clavicles, and the diaphragm, but one can also use different classifiers for half-occluded on non-occluded nodules by the ribs. When implanted nodules cannot be distinguished from real nodules then availability of trained data is not an issue anymore.

## 4 References

1. Schilham, A., Ginneken, B.v., Loog, M.: 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) (2006) 247–258

2. Shiraishi, J., Abe, H., Li, F., Engelmann, R., Macmahon, H., Doi, K.: Computer-aided diagnosis for the detection and classification of lung cancers on chest radiographs: ROC analysis of radiologists' performance. Academic Radiology **13**(8) (2006) 995–1003
3. Freedman, M., Chung, S., Osicka, T., Lure, F., Xu, X.W., Lin, J., Zhao, H., Zhang, R.: Computer-aided detection of lung cancer on chest radiographs: effect of machine CAD false-positive locations on radiologists' behavior. Volume 4684., SPIE (2002) 1311–1319
4. Hardie, R., Rogers, S., Wilson, T., Rogers, A.: Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs. Medical Image Analysis **12**(3) (2008) 240–258
5. Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K., Matsui, M., Fujita, H., Kodera, Y., Doi, K.: Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. American Journal of Roentgenology **174**(1) (2000) 71–74
6. Ginneken, B.v., Stegmann, M., Loog, M.: Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Medical Image Analysis **10**(1) (2006) 19–40
7. Kostis, W., Reeves, A., Yankelevitz, D., Henschke, C.: Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. IEEE Transactions on Medical Imaging **22**(10) (October 2003) 1259–1274
8. Litjens, G., Hogeweg, L., Ginneken, B.v.: Simulation of nodules and interstitial disease in chest radiographs using CT templates. In: Proceedings MICCAI 2010. Lecture Notes in Computer Science, Springer (2010)
9. Miller, H.: The froc curve: A representation of the observer's performance for the method of free response. The Journal of the Acoustical Society of America **46**(6B) (1969) 1473–1476