Automated detection and quantification of micronodules in thoracic CT scans to identify subjects at risk for silicosis

C. Jacobs[1], S.H.T. Opdam[2], E.M. van Rikxoort[1], O.M. Mets[3], J. Rooyackers[4], P.A. de Jong[3], M. Prokop[1], B. van Ginneken[1]

[1]Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, The Netherlands
[2]Department of Biomedical Engineering, Eindhoven University of Technology, The Netherlands
[3]Department of Radiology, University Medical Centre Utrecht, The Netherlands
[4]Netherlands Expertise Centre for Occupational Respiratory Disorders, Utrecht University, The Netherlands

ABSTRACT

Silica dust-exposed individuals are at high risk of developing silicosis, a fatal and incurable lung disease. The presence of disseminated micronodules on thoracic CT is the radiological hallmark of silicosis but locating micronodules, to identify subjects at risk, is tedious for human observers. We present a computer-aided detection scheme to automatically find micronodules and quantify micronodule load. The system used lung segmentation, template matching, and a supervised classification scheme. The system achieved a promising sensitivity of 84% at an average of 8.4 false positive marks per scan. In an independent data set of 54 CT scans in which we defined four risk categories, the CAD system automatically classified 83% of subjects correctly, and obtained a weighted kappa of 0.76.

1. PURPOSE

Silicosis is an incurable lung disease and one of the oldest known occupational diseases: The ancient Greeks and Romans were already aware that breathing dust may cause respiratory illness. Chronic silicosis is radiologically characterized by widespread, well-defined solid pulmonary micronodules, measuring 3mm or less.[1] In an advanced state, this may be seen on chest radiographs, but micronodules are much better visible, and probably in a much earlier stage already detectable on computed tomography (CT) scans. It is difficult to diagnose silicosis unequivocally, and therefore a combination of a history of dust exposure, radiological manifestations and exclusion of other diseases is used.[2] Even then, silicosis is difficult to recognize, especially in early phases of the disease where symptoms are often absent, or resemble those of chronic obstructive pulmonary disease. Early detection is however, very important because there is no cure and ending exposure is the only way to avoid progression.

In a recent study, Mets et al. compared CT scans of two groups of 54 subjects.[3] The study group consisted of subjects at high risk for silicosis extracted from a database of construction workers and miners. The control group consisted of heavy

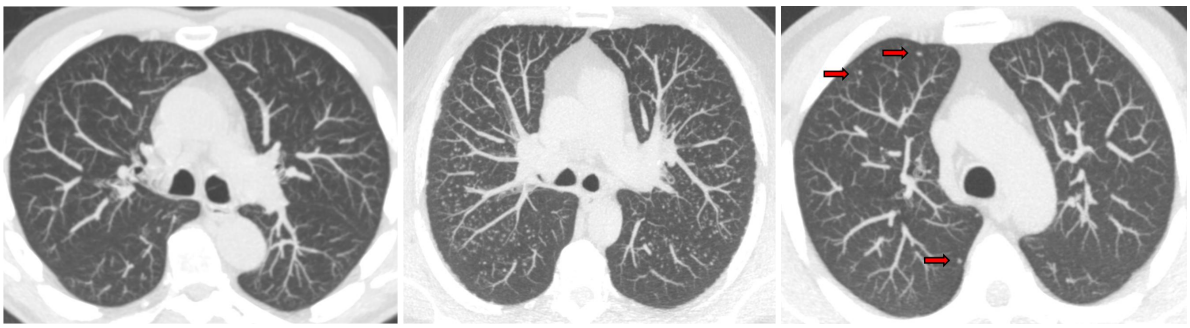

Figure 1. Three axial sections which illustrate the detection task for micronodules. All images are axial maximum intensity projections of 10 mm. The left scan is a normal subject without micronodules. The second scan shows disseminated micronodules, compatible with radiological features of manifest silicosis. In the right image, a subject which had 14 micronodules is depicted. The arrows indicate the three detected micronodules in this particular section of the scan. Evidently, the detection of a small amount (> 13) of micronodules to detect early stage silicosis is a tedious task.

Medical Imaging 2014: Computer-Aided Diagnosis, edited by Stephen Aylward, Lubomir M. Hadjiiski,
Proc. of SPIE Vol. 9035, 90351I · © 2014 SPIE · CCC code: 1605-7422/14/$18 · doi: 10.1117/12.2043536

smokers from a lung cancer screening program. The groups were matched for age, sex and smoking behavior (research showed that the vast majority of construction workers with high silica exposure are also heavy smokers). Two radiologists visually, in a blinded fashion, scored how many micronodules were present in each scan. The authors found that in almost all scans micronodules were present (median 4 in both groups) and that in the control group the 95th percentile was 13 micronodules, which they regarded as the upper limit of normal (ULN). Subjects with more than 13 micronodules should be considered at high risk for silicosis. In the group of construction workers 12 subjects (22%, significantly more than 5%), had $>$13 micronodules, while only two subjects had disseminated micronodules, one of them 144, compatible with the radiological definition of silicosis. Mets et al. hypothesized that the presence of moderate numbers of micronodules is an early sign of silicosis.

Detecting micronodules is, however, a tedious task. Such small nodules are easy to miss because their size is often similar to cross-sections of small vessels. Even searching for larger nodules ($>$4mm), as should be done to detect possible early lung cancer, takes expert human observers between 10 and 15 minutes per scan. Detecting subjects at high risk for silicosis requires finding those patients which have more than 13 micronodules, which is even more challenging. The difficulty of this task is depicted in Fig.1. Counting micronodules seems therefore unfeasible in clinical practice unless the task can be automated. This was the purpose of the current study, and we had the data of Mets et al. and additional scans of subjects at high risk of developing silicosis at our disposal. To our knowledge, this is the first paper on automated detection of micronodules, even though the problem of detecting larger nodules has been widely studied.[4–9]

2. METHODS

2.1 Data

The database with CT scans of construction workers at high risk for silicosis, described in the publication by Mets et al.,[3] was used to select cases for the development set. In that publication, the study group contained 54 out of the 159 subjects in the database: only the subjects which matched for age and smoking behavior with the control group were included. We selected 15 scans from the remaining subjects of the database as a development data set. On these 15 scans, a radiology resident annotated micronodules for training, 85 in total. Note that the radiology resident did not extensively search all 15 scans for micronodules, but created a subset of micronodules for training of the CAD algorithm.

For evaluation of the CAD system, all scans of the study group of 54 dust-exposed subjects were used. Two observers counted the number of micronodules in all 54 scans. The average of the scores of the two observers was used. The median number of micronodules on these scans was 4 with a range from 0 to 144.

All subjects received low-dose chest CT examination with either 16x0.75-mm collimation (Mx8000 IDT or Brilliance 16P; Philips Medical Systems, Cleveland, OH) or 128x0.625-mm collimation (Brilliance iCT; Philips Medical Systems). Axial slices with a slice thickness of 1.0 mm were reconstructed with an interval of 0.7 mm using a moderately soft reconstruction kernel.

2.2 Candidate detection

The candidate detection stage consisted of four stages. First, the lung fields were automatically extracted using a method proposed by van Rikxoort et al.[10] Second, the scan was cropped around the lung segmentation and resampled to an isotropic resolution of 0.7 mm^3. Third, all voxels above a threshold t were selected. Fourth, similar to Lee et al.,[11] we performed template matching on selected voxels. A 3D Gaussian blob template with scale σ was used. Normalized cross-correlation coefficient (CCC) was used as a similarity measure. Finally, local maxima detection in a 26-neighborhood was applied to acquire the final candidate locations. During pilot experiments it was found that $t = -700$HU and $\sigma = 0.35$mm yielded optimal results on the training set.

2.3 False positive reduction

A set of 30 features was used here to further describe the candidates and to reduce false positive responses. They are listed in Table 1. Features were calculated from a region around a candidate. Region 1 ($R1$) is a spherical region with a diameter of 3 mm with the candidate as center. This region has approximately the size of a micronodule. Region 2 ($R2$) is a spherical

region around the candidate with a diameter of 8 mm, from which $R1$ has been excluded. The features can be subdivided into 4 classes: intensity features, gradient features, correlation features and histogram features.

Features 1,2,3 and 5-8 were defined with respect to the scan intensity. Since micronodules have a high density, they most likely have a higher attenuation on the scan than the surroundings, even when located close to blood vessels or other structures. Features 3 and 5-8 took into consideration the intensity in a small region around the candidate, while features 1 and 2 were also dependent on the statistics of the wider surroundings. Features 4 and 13-16 were obtained from the magnitude of the gradient vectors. The gradient vector was calculated using a Sobel kernel and measures the gradient in scan intensity and its direction. The magnitude of this vector is the total change of intensity in the region and is a scalar. This scalar was used in feature calculations. In a round object on a CT scan, the gradient vectors on the edge all radiate from the center. When the candidate represents a blood vessel, a common source of false positives, gradients will all point in the same direction, and their magnitudes will show a different pattern than that of a round object. Features 9-12 were calculated from the CCC of the candidate and its surroundings. In addition to the maximum CCC value, statistics of the surroundings are also of interest. Features 17-30 were calculated from the histogram of local scan intensity values in R1. The histograms have a bin width of 10 HU. Histogram features give a more detailed description of the intensity distribution around candidates, which can also be used to discriminate between micronodules and other objects.

Candidates were classified with a k-nearest neighbor classifier with k set to the square root of the number of training samples, which gave $k = 149$.

2.4 Optimizing the training data set using an active learning approach

During preliminary evaluation of the CAD algorithm, it became evident that there was a substantial amount of micronodules not annotated on our 15 training cases. Therefore, we used an active learning approach to improve the annotations on the training data set. In this approach, we use the CAD algorithm itself to update the annotations on the training data set. We start by performing a leave-one-out cross-validation experiment. Next, we perform FROC analysis to evaluate the performance of the system. Subsequently, we inspect the 100 most suspicious false positives and mark the false positives which are actually micronodules. These micronodules are then added to the reference standard on the training set and we repeat the same procedure again. Note that the training in the cross-validation now uses these new annotations and the performance of the system improves. This process is iterated until no micronodules are found among the 100 most suspicious false positives. Note that we are using the CAD algorithm to update the annotations on the training set and as a consequence, although leave-one-out cross-validation is applied, a biased estimate of the performance of our system on the training set will be obtained. An additional evaluation of our system will be performed on independent data, the study group used by Mets et al.

#	Description
1	Ratio: Max intensity $R1$/Max intensity $R2$
2	Ratio: Mean intensity $R1$/Mean intensity $R2$
3-4	Intensity and gradient magnitude of candidate voxel
5-8	Max, min, mean, std. dev. of intensities
9-12	Max, min, mean, std. dev. of CC
13-16	Max, min, mean, std. dev. of gradient vector magnitudes
17	Histogram entropy
18	Histogram mean position
19	Histogram std. dev.
20-21	Histogram mean height/nonzero mean height
22-23	Histogram max peak position/nonzero max peak position
24-25	Histogram max peak height/nonzero max peak height
26-30	Histogram quantile 5, 25, 50, 75, 95

Table 1. Features used for false positive reduction. All features are computed over region $R1$, except for features 1 and 2.

3. RESULTS

Four iterations of the active learning approach were performed to update the annotations of the training data set and the number of added micronodule annotations are shown in Table 2.

Iteration	# micronodules among 100 most suspicious false positives	Total # of micronodule annotations
1	63 / 100	85
2	37 / 100	148
3	9 / 100	185
4	0 / 100	194

Table 2. Results from optimizing the training data set using the active learning approach.

Leave-one-nodule-out cross validation was used to assess the performance of the micronodule CAD system on the final development set with 194 micronodule annotations. Free-response operating characteristic (FROC) analysis was applied to measure the performance of the CAD system. The FROC curve is depicted in Fig. 2 and shows a good performance with a sensitivity of 84% at an average of 8.4 false positives (FPs) per scan.

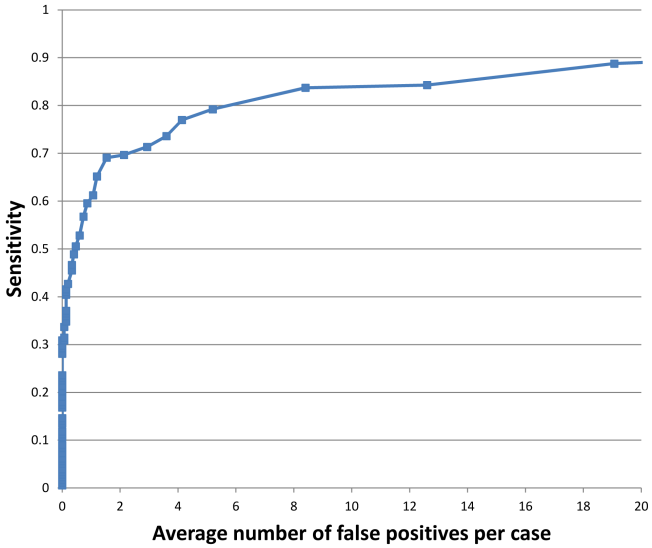

Figure 2. FROC curve of the micronodule CAD system, obtained using leave-one-nodule-out cross validation.

An experiment was conducted to test whether the CAD system can detect subjects with disseminated micronodules, compatible with radiological features of manifest silicosis. The micronodule counts by the radiologists suggested that 2 out of the 54 subjects in the study group had radiological patterns of manifest silicosis. We hypothesize that a micronodule CAD system set to operate at a high sensitivity is able to pick out cases with high micronodule counts. To do so, the CAD algorithm was set to operate at an operating point of 84% sensitivity and an average of 8.4 FPs/scan. The CAD system processed all 54 cases and we evaluated the potential of the CAD system running fully automatic, e.g. without visual checking of the CAD marks by a radiologist.

On the test set of 54 cases, the CAD system generated 637 CAD marks (11.8 CAD marks per scan), with a median of 3 (range 0-303). The two cases with a radiological pattern of manifest silicosis had the highest number of CAD marks, 303 and 45 CAD marks, respectively. If a threshold of 40 CAD marks per scan would be used, the two cases with manifest silicosis could be detected in this database without any false positive cases. Fig. 3 displays two axial sections of these two cases in which many micronodules are present.

We ordered the cases into four groups, (1) low risk: cases with less or equal than 4 micronodules, (2) intermediate risk: cases containing between 5 and 13 micronodules, (3) high risk: cases containing between 13 and 40 micronodules and (4)

manifest silicosis: cases with more than 40 micronodules. Note that cases in group 3 are at higher risk for silicosis and above the UPN of the control group.[3] The confusion matrix which compares the reference standard set by the observers with the CAD system is displayed in Table. 3. The weighted kappa between the risk assessment by human observers and the CAD system was 0.76.

		CAD Prediction			
		Low Risk	Intermediate Risk	High Risk	Manifest Silicosis
Observers	Low Risk	28	4	2	0
	Intermediate Risk	1	9	2	0
	High Risk	0	0	6	0
	Manifest Silicosis	0	0	0	2

Table 3. Confusion matrix for the four risk categories.

4. DISCUSSION AND FUTURE WORK

In this study, we developed an automated detection system for micronodules that has good performance and can be used to automatically classify scans into low, intermediate, high risk, and manifest silicosis. The confusion matrix in Table. 3 shows that both cases of manifest silicosis were correctly detected without any false positive detections, and that only 2 out of 54 subjects were not assigned to either the correct risk group or a neighboring risk group. Manual annotation of micronodules by human observers is exhausting, labor intensive and prone to errors. A CAD tool may be able to give a more precise quantification which can be important for the purpose of monitoring disease progression. Therefore, an efficient CAD system for micronodules is of clinical significance. Future work includes experimenting with other classifiers, optimization and expansion of the feature set, and validation on a larger cohort.

5. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

This is the first paper to present an automated system for the detection and quantification of micronodules from thoracic CT scans. This is an effective automatic risk assessment tool for subjects at high risk of developing silicosis.

6. CONCLUSIONS

An automated system for the detection and quantification of micronodules from thoracic CT scans has been presented. The system was validated by comparing with a counting by humans in a high risk screening cohort and shows excellent performance. This paves the way for automated risk assessment for silicosis, which could be useful to screen high risk subjects such as construction workers or miners, and could also be included in lung cancer screening programs.

REFERENCES

[1] Hansell, D. M., Bankier, A. A., MacMahon, H., McLoud, T. C., Müller, N. L., and Remy, J., "Fleischner society: glossary of terms for thoracic imaging," *Radiology* **246**, 697–722 (2008).

[2] Rees, D. and Murray, J., "Silica, silicosis and tuberculosis," *International Journal of Tuberculosis and Lung Disease* **11**, 474–484 (2007).

[3] Mets, O. M., Rooyackers, J., van Amelsvoort-van de Vorst, S., Mali, W. P., de Jong, P. A., and Prokop, M., "Increased micronodule counts are more common in occupationally silica dust-exposed smokers than in control smokers," *Journal of Occupational and Environmental Medicine* (2012).

[4] 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., Belloti, R., Carlo, F. D., Megna, R., Tangaro, S., Bolanos, L., Cerello, P., Cheran, S. C., Torres, E. L., and Prokop, M., "Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study," *Medical Image Analysis* **14**, 707–722 (2010).

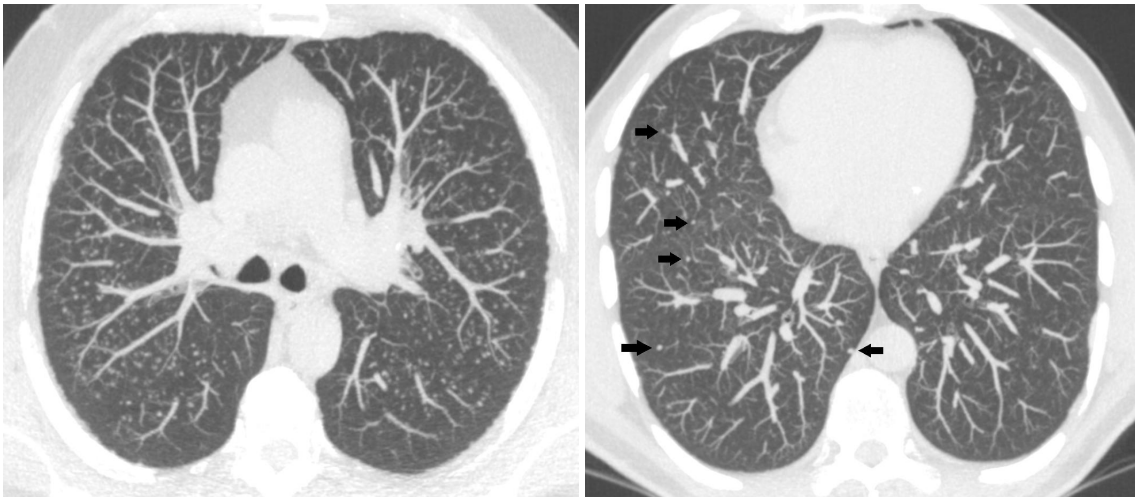

Figure 3. Two axial sections of the two scans with the most micronodules according to both CAD system and human observers. These two scans were the only ones compatible with radiological features of manifest silicosis. Both images are axial maximum intensity projections of 10 mm. In the right image, arrows indicate the detected micronodules (in the left image there were too many detections to indicate).

[5] Murphy, K., van Ginneken, B., Schilham, A. M. R., de Hoop, B. J., Gietema, H. A., and Prokop, M., "A large scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification," *Medical Image Analysis* **13**, 757–770 (2009).

[6] Messay, T., Hardie, R. C., and Rogers, S. K., "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," *Medical Image Analysis* **14**, 390–406 (2010).

[7] Jacobs, C., van Rikxoort, E., Twellmann, T., Scholten, E., de Jong, P., Kuhnigk, J., Oudkerk, M., de Koning, H., Prokop, M., Schaefer-Prokop, C., and van Ginneken, B., "Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images," *Medical Image Analysis* **18**, 374–384 (2014).

[8] 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., and Golosio, B., "A CAD system for nodule detection in low-dose lung CTs based on region growing and a new active contour model," *Medical Physics* **34**, 4901–4910 (2007).

[9] Retico, A., Delogu, P., Fantacci, M. E., Gori, I., and Martinez, A. P., "Lung nodule detection in low-dose and thin-slice computed tomography," *Computers in Biology and Medicine* **38**, 525–534 (2008).

[10] van Rikxoort, E. M., de Hoop, B., Viergever, M. A., Prokop, M., and van Ginneken, B., "Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection," *Medical Physics* **36**, 2934–2947 (2009).

[11] Lee, Y., Hara, T., Fujita, H., Itoh, S., and Ishigaki, T., "Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique," *IEEE Transactions on Medical Imaging* **20**, 595–604 (2001).