

Automated Scoring of Chest Radiographs for Tuberculosis Prevalence Surveys: A Combined Approach

Bram van Ginneken¹, Rick H.H.M. Philipsen¹, Laurens Hogeweg¹, Pragnya Maduskar¹, Jaime C. Melendez¹, Clara I. Sánchez¹, Rahmatulai Maane², Beatrice dei Alorse², Umberto d'Alessandro², and Ifedayo M.O. Adetifa²

¹Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands (<http://www.diagnijmegen.nl>)

²Disease Control and Elimination Theme, Medical Research Council Unit-Gambia, Fajara, The Gambia

Abstract. Chest radiography is one of the key techniques for investigating suspected tuberculosis (TB). Computerized reading of chest radiographs (CXRs) is an appealing concept because there is a severe shortage of human experts trained to interpret CXRs in countries with a high prevalence of TB. This paper presents a comprehensive computerized system for the detection of abnormalities in CXRs and evaluates the system on digital data from a TB prevalence survey in The Gambia. The system contains algorithms to normalize the images, segment the lung fields, analyze the shape of the segmented lungs, detect textural abnormalities, measure bluntness of the costophrenic angles and quantify the asymmetry in the lung fields. These subsystems are combined with a Random Forest classifier into an overall score indicating the abnormality of the radiograph. The results approach the performance of an independent human reader.

Keywords: Chest radiography, Tuberculosis, Medical image analysis

1 Introduction

Tuberculosis (TB) is the only disease in the top 10 of global causes of mortality for which a cheap and effective cure is available [1]. In 2011 an estimated 8.7 million new cases and 1.4 million deaths were reported. The incidence of TB is highest in Africa, followed by the Asian countries [2].

The World Health Organization (WHO) has identified 21 global focus countries for TB prevalence surveys for accurate assessment of the disease burden and impact of TB control programmes [3]. The screening algorithm for TB prevalence surveys recommends screening of eligible target population first with a symptom questionnaire and then chest radiography (CXR). Stationary or mobile digital chest radiography units are now being increasingly used for these surveys. Subjects with symptoms indicative of TB, or abnormalities on the radiograph, submit a sputum sample which is analyzed with microscopy, culture

or other tests to determine if that person has TB [4]. It is too time-consuming and costly to analyze sputum of all participants. In this scenario it is vital that the decision whether the chest radiograph is normal or abnormal can be made quickly. Ideally, this decision is made directly after obtaining the image, so that a survey participant with an abnormal radiograph can immediately be asked to submit sputum. To facilitate such a high throughput prevalence screening scenario, which could also be applied in active TB case finding among high risk groups, we aim to develop a computerized system that runs on standard computer hardware, analyzes a chest radiograph automatically, and produces a score that indicates how abnormal the radiograph is within minutes. This score can be thresholded to determine which subjects should undergo further testing. The threshold can be optimized depending on the requirements and capacity of the screening program.

This paper describes a platform and its components that comprise such an automated image analysis solution. We evaluated the performance of the system on data from a pilot phase of the TB prevalence survey of The Gambia, that was carried out during 2012 using the WHO recommended protocol.

A large diversity of pathologic changes in the lungs can be observed on CXRs because of the complex pathophysiology of TB and manifestations also differ with different severity and stages of the disease [5]. A generally applicable computer system therefore should be able to detect the presence of all these patterns. Moreover, the software should be able to process images from different radiography units.

In this work we describe such a system that has the following components: image normalization (to be able to deal with images from different machines);

Novel contributions of this paper are the integration of the various components into a combination classifier and the symmetry analysis system. Furthermore, this is the first paper to evaluate automated detection of TB on a database of digital chest radiographs obtained from TB prevalence survey.

2 Methods

This section describes the components of the automated analysis system.

2.1 Image normalization

Supervised computer analysis systems are typically trained on data from a small number of sources and as a result, they may not perform optimally on data from different sources. For example in this work, training data was provided by a digital slotscanner system and the completely independent test data was obtained from mobile units with full field digital detectors. We therefore applied an energy normalization procedure to both training and test data prior to further processing.

This procedure started by resizing the image to a standardized width of 1024 pixels and scaling the pixel values such that they have zero mean and unit

standard deviation. The image was then divided into frequency bands. Starting with the original image, a blurred version of the original image I , obtained by convolution with a Gaussian kernel with standard deviation σ_1 was subtracted from it which resulted in a high frequency image I_1 . The same procedure was applied to the blurred image now using a larger standard deviation σ_2 , and so on, and in this way 6 energy bands were obtained with $\sigma_i = 2^{i-1}$. From a set of reference radiographs obtained from a single type of digital radiography unit, we determined the average standard deviation (energy) in the central 70% of the image for each energy band. We now scaled the pixel values in each band so as to match these reference energy levels and reconstructed the radiograph by adding the energy bands.

The result of this preprocessing operation is that the intensity ranges and perceived sharpness of the high and medium frequency details is similar in all images. We have recently described this algorithm in more detail [6] and shown that after application of this energy normalization, a lung segmentation algorithm similar to what we describe here (Sect. 2.2) that is trained with data from one source can successfully be applied to data from other sources.

2.2 Lung field segmentation

A lung segmentation is required to limit subsequent analysis to the region inside the lung fields, where radiographic patterns associated with TB are primarily visible. The method we used to segment the lung fields (see [7] for details) is based on pixel classification and postprocessing. For computational efficiency, the normalized images were subsampled to a width of 256 pixels. A multi-scale local jet [8] of second order was computed at scales 1, 2, 4, 8, 16 pixels. In addition, the original pixel value and the x and y position were used as features. From a set of 309 training images, examples of pixels in- and outside manually outlined lung fields were sampled (a random selection of 0.6% for both classes). A k-nearest-neighbor classifier ($k = 15$) was trained and used to assign all pixels in a test image a lung likelihood p . The resulting lung likelihood map was converted to a binary segmentation in a series of steps: Gaussian blurring with $\sigma = 0.7$ pixels, thresholding at $p = 0.5$, selection of the two largest components, morphologically closing with a spherical kernel of radius = 10 pixels, and applying hole filling.

2.3 Lung field shape analysis

When severe abnormalities are present in the image, lung field segmentation may fail. In the worst case scenario, the segmented part would contain only normal lung parenchyma and subsequent texture analysis would not flag the image as being highly abnormal. Therefore, we added an analysis step that evaluates the shape of segmented lung fields for normality.

For each lung, 80 equiangular rays were cast from the center of gravity of the object and the distances to the intersection with the boundary, normalized by the height of the bounding around both lungs, were computed. All distances were stacked in a feature vector. For a set of normal images with correct lung

segmentation, principal component analysis was applied retaining 95% of the variance. For new shape vectors, the Mahalanobis distance after projecting on the principal components was computed as a lung shape score. A large Mahalanobis distance of a test shape to the model indicates a more abnormal shape and is indicative of a grossly abnormal image.

2.4 Costophrenic recess analysis

Pleural fluid in the costophrenic recesses may cause the costophrenic angle, the point where the diaphragm meets the outer rib cage, to appear ‘blunted’. Although such a sign is not always associated with TB, in areas with a high prevalence of TB it is possible that a TB patient exhibits no abnormalities on a CXR except for a blunt costophrenic angle. We therefore developed an algorithm exclusively devoted to detecting the presence of blunt costophrenic (CP) angles caused by pleural effusion on chest radiographs. The CP angle is the angle formed by the hemidiaphragm and the chest wall. We defined the intersection point of both as the CP angle point. We detected the CP angle point automatically from the lung segmentation by locating the foreground pixel of each lung with maximum y location. Patches were extracted around the CP angle point and boundary tracing was performed to detect 10 consecutive pixels along the hemidiaphragm and the chest wall and we computed the CP angle from these. The method was recently evaluated by Maduskar et al. [9]. The two angular measurements (for the left and right costophrenic angle) were used as features indicative of pleural effusion in the costophrenic recesses.

2.5 Texture analysis

Textural abnormalities in CXRs occur often as a result of TB infection and they typically reflect inflammatory changes in the lung parenchyma, but can also be the result of fluid accumulation or fibrotic changes in the pleural space. The texture analysis system applied, has been previously described in Hogeweg et al. [10] and was inspired by Arzhaeva et al. [11] and Van Ginneken et al. [12]. The algorithm operates on normalized images resampled to a width of 1024 pixels. On a grid with a spacing of 8 pixels circular patches with a radius of 32 pixels were defined. Within these patches, the first four moments of the output of the multiscale jet described above were computed as texture features. In addition several spatial features (normalized x and y position and distance to the lung wall and center of gravity of both lungs) were computed. Pixels were classified with a k-nearest-neighbor classifier. Training samples have been obtained from a set of 1,000 digital CXRs of TB suspects collected at a busy urban clinic in Lusaka, Zambia. In these images, all abnormal regions were manually outlined and pixels in such regions count as abnormal examples. Normal samples were obtained from completely normal images.

This texture analysis produced a ‘heat map’ indicating regions with abnormal texture. To aggregate this analysis into an image score, the 95% percentile

from the cumulative distribution of patch likelihoods was computed. This score measures in a robust way the extent and the severity of the affected lung [13].

2.6 Symmetry analysis

Symmetry is an important feature of human anatomy and the absence of symmetry in medical images may indicate the presence of pathological changes. Quantification of image symmetry can be used to improve the automatic analysis of medical images. We have developed a method that computes global symmetry in 2D medical images for which two regions that should be roughly mirror symmetric have been defined. The segmented lung fields serve as those regions.

First, an initial vertical symmetry axis was determined based on a vertical projection of the lung fields and an analysis of the projection profile. With this axis available, the lung fields were converted to a mirror symmetric representation. Now for each position p in the left lung field, a mirrored position p' can be determined in the right lung field, and vice versa. In the neighborhood of p' , an optimally corresponding position p'' was determined by minimizing a cost function d that adds intensity differences in an n by n patch around p and the mirrored patch around p'' and the Euclidean distance between p'' and p' , multiplied by a weighting term m . The optimal value of d was used as a measure of local symmetry, and all values of d were averaged to obtain the global symmetry D .

One major issue with this approach is that a slight rotation of the CXR has a large effect on D . The measure was therefore optimized by a greedy minimization of D for a rotation between -10 and +10 degrees, and for slight horizontal displacements of the vertical symmetry axis. The free parameters n and m have been optimized in pilot experiments where settings were determined that yielded optimal separation between normal and abnormal images using only D as a measure.

An attractive property of symmetry computation is that it is self-normalizing: the contralateral side of the image is the 'control image', and this circumvents the problem of supervised schemes where CXRs of different subjects are compared with inherent variation in image appearance due to anatomical and physiological variation. A limitation of this approach is that cases where the both sides exhibit similar abnormal patterns may achieve a high symmetry score, similar to normal cases. It is hypothesized that such cases, because of their more widespread abnormalities, will usually be detected by the texture analysis.

2.7 Classifier combination

The analyses described above yield for each image 5 scores (lung shape, left and right costophrenic angle, texture and symmetry). These scores were used as features for a Random Forest classifier [14] using 50 decision trees with a maximum tree depth of 7. This classifier was trained with a reference set of 1,000 radiologically normal and abnormal images.

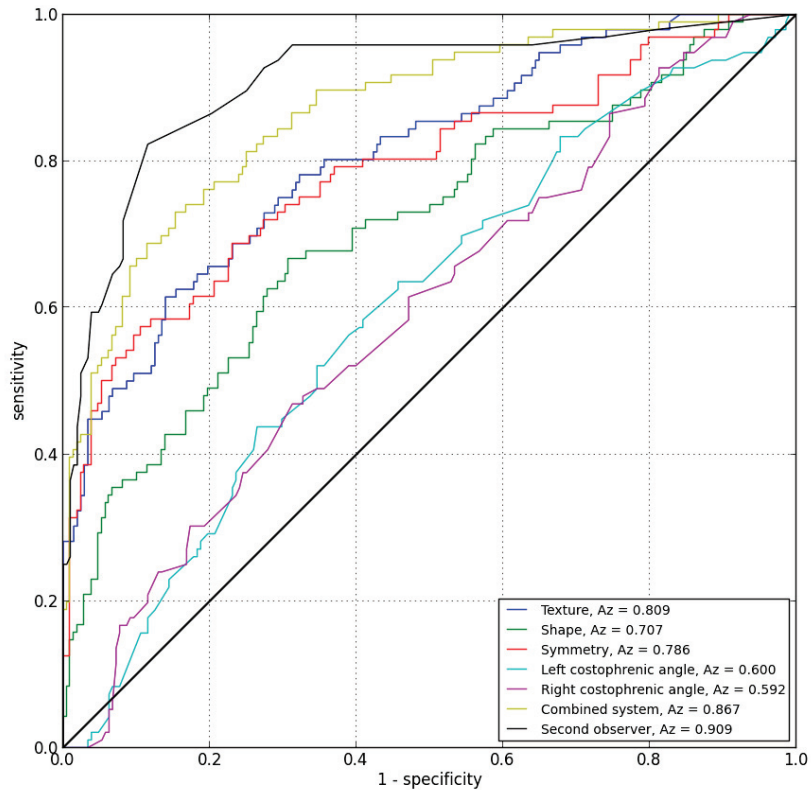


Fig. 1. ROC curves for each of the individual components, the combined system, and an independent second observer.

3 Materials

The training database used to train the texture analysis system, the lung shape analysis system, and the Random Forest classifier consisted of 1,000 images from a digital Odelca-DR system with a slotscan detector (Delft Imaging Systems, The Netherlands). The CXRs were collected in an African country with a high incidence of tuberculosis (TB). The radiographs had image widths in the range of 1500-1800 pixels and a pixel spacing of $256 \times 250 \mu\text{m}$. Each image was read by a chest radiograph and recording system (CRRS) [15] certified "B" reader. Three types of abnormalities were manually outlined: small and large opacities, and consolidations. A subset of this data set was used to train the lung segmentation system.

| | Shape | Left CA | Right CA | Texture | Symmetry | Combination |
|-------------------------|-------|---------|----------|---------|----------|-------------|
| Components individually | 0.707 | 0.600 | 0.592 | 0.809 | 0.786 | 0.867 |
| Combination without | 0.858 | 0.851 | 0.758 | 0.823 | | |

Table 1. Performance in terms of area under the ROC curve for the individual subsystems, the combination, and the combination without individual subsystems, where both costophrenic angle measurements were taken together. CA = costophrenic angle.

The performance of the individual components and the combined system was tested on a set of 304 images from a digital Atomed mobile X-ray system (Delft Imaging Systems, The Netherlands) with a Canon CXDI detector acquired in a prevalence survey in The Gambia. The CXR widths were in the range of 2000-2500 pixels and have a pixel spacing of 160 μm isotropic. The radiological reference standard for each image was set by the same reader. 96 images were read as abnormal. This set was also read by an independent second observer who rated each image on a scale from 0 (normal) to 100 (very abnormal) so that an ROC curve could be constructed.

4 Results

Fig. 1 shows the results of the individual systems and the combined system as ROC curves. The area under the curve, A_z , a global performance measure, is listed in the legend. Additionally, the performance of an independent human observer is provided. Fig. 2 provides illustrative examples of the output of the individual systems. Table 1 lists A_z values of individual systems, the combined system and instantiations of the combined system in which the texture, shape, symmetry and costophrenic angle measurements were removed. In this way it can be seen what the effect of removing one type of analysis was on the combined approach.

5 Discussion and conclusion

From all the individual systems, the measurements of the costophrenic angles were performing the poorest. This is not surprising. Not all abnormal chest radiographs have a left or right blunt costophrenic angle, so these two systems at best detect only a subset of the abnormal cases. Still, inclusion of the costophrenic recess analysis in the combined system demonstrated added value: performance increased from 0.851 to 0.867. It appears there is room for improvement in the blunt costophrenic angle detection system though: the ROC curves start below the chance line, indicating that the images with the most blunt angles are actually normal, and close inspection of these cases revealed that in these cases the costophrenic point has not been correctly detected.

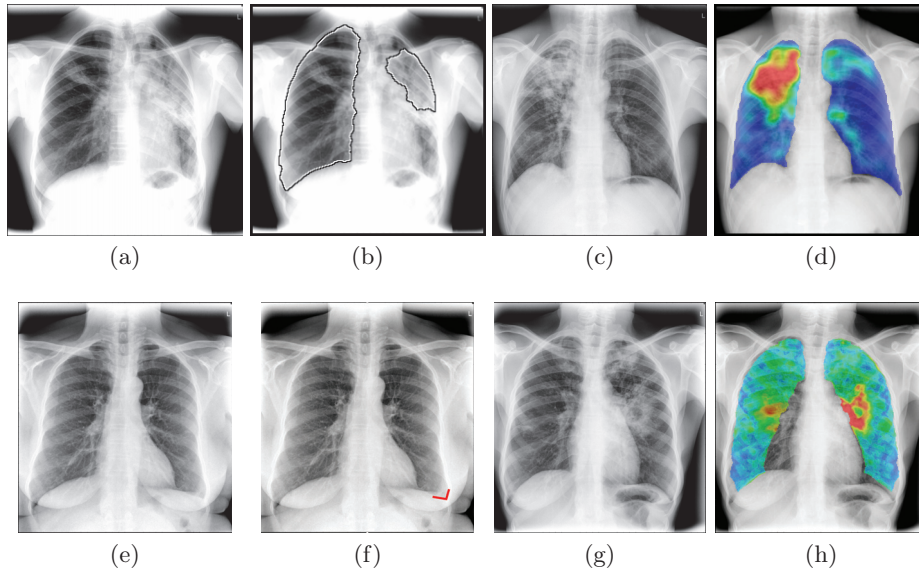


Fig. 2. Illustrative examples of the output of components of the analysis system. (a-b) In this highly abnormal image the lung field segmentation failed. The dense abnormality in the middle part of the left lung was classified as background and the left lung segmentation therefore consisted of multiple parts of which only the largest component was retained. This image therefore received the largest shape score (most abnormal shape) and was therefore considered highly abnormal. (c-d) Another case with clear abnormalities and a cavity in the upper right lung field. These abnormalities were detected by the texture analysis system. (e-f) A case where the only obvious abnormality is a blunt costophrenic angle of the left lung. This was detected by the costophrenic recess analysis (100.5°). (g-h) A case with an abnormal hilar structure that resulted in high asymmetry on both sides. Other more subtly abnormal regions gave a lower (greenish) output of the asymmetry detection.

The subsystem that estimates how abnormal the shape of the segmented lung fields is performed admirably on its own, with an A_z of 0.707. This indicates that the lung segmentation partly fails quite often in the case of abnormal radiographs. Given that the segmentation is based on pixel classification and was largely trained with normal images, this is as may be expected: the lung segmentation system effectively detects normal lung parenchyma. If we leave out the shape system from the total system, performance dropped from 0.867 to 0.858. This implies that occasionally indeed the segmented part of the lungs is completely normal and there is complementary information in the shape score.

The symmetry analysis achieved, on its own, a good performance of 0.786, close to the best performing subsystem, texture analysis, which obtained $A_z = 0.809$. This is remarkable because the symmetry system is not a supervised system trained with other radiographic images, but a completely standalone analysis of a single test image. The texture system was, as expected, the best

single system. But the combined analysis improved substantially upon texture analysis alone.

Observing the ROC curves, we note that performance of the combined system was close to that of a human expert for operating points at either a high specificity or a high sensitivity. At a sensitivity of 50% both the automatic and the human observer had a false positive rate of only a few percent. And both systems reached a sensitivity above 95% with a specificity of 40%. The human observer still achieved better sensitivity in the specificity range between 50 and 90%. An inspection of the abnormal images that the human reader detected additionally to the computer at these specificity levels showed that these images contained relatively subtle abnormalities. Some of these were limited to the hilar region. An additional subsystem focused on the detection of lymphadenopathy will therefore be useful. The architecture of the system presented here naturally supports the inclusion of an additional subsystems.

Besides further improvement of the underlying algorithms, and the development of additional subsystems, future work will also focus on evaluation of the system in an operational setting. Comparison to multiple human observers is needed to reliably assess the potential of computerized reading versus current practice.

This is the first paper to present a comprehensive approach to the detection of abnormalities in chest radiographs acquired to detect signs of tuberculosis. Moreover, it is the first validation of a computer-aided detection system for tuberculosis using data obtained from a TB prevalence program. The promising results indicate that automated reading may be incorporated in TB prevalence systems employing digital chest radiography, which could greatly improve the logistics and cost effectiveness of such programs.

References

1. Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y., Alvarado, M., Anderson, H.R., Anderson, L.M., Andrews, K.G., Atkinson, C., Baddour, L.M., Barker-Collo, S., Bartels, D.H., Bell, M.L., Benjamin, E.J., Bennett, D., Bhalla, K., Bikbov, B., Bin Abdulhak, A., Birbeck, G., Blyth, F., Bolliger, I., Boufous, S., Bucello, C., Burch, M., Burney, P., Carapetis, J., Chen, H., Chou, D., Chugh, S.S., Coffeng, L.E., Colan, S.D., Colquhoun, S., Colson, K.E., Condon, J., Connor, M.D., Cooper, L.T., Corriere, M., Cortinovis, M., de Vaccaro, K.C., Couser, W., Cowie, B.C., Criqui, M.H., Cross, M., Dabhadkar, K.C., Dahodwala, N., De Leo, D., Degenhardt, L., Delossantos, A., Denenberg, J., Des Jarlais, D.C., Dharmaratne, S.D., Dorsey, E.R., Driscoll, T., Duber, H., Ebel, B., Erwin, P.J., Espindola, P., Ezzati, M., Feigin, V., Flaxman, A.D., Forouzanfar, M.H., Fowkes, F.G.R., Franklin, R., Fransen, M., Freeman, M.K., Gabriel, S.E., Gakidou, E., Gaspari, F., Gillum, R.F., Gonzalez-Medina, D., Halasa, Y.A., Haring, D., Harrison, J.E., Havmoeller, R., Hay, R.J., Hoen, B., Hotez, P.J., Hoy, D., Jacobsen, K.H., James, S.L., Jasrasaria, R., Jayaraman, S., Johns, N., Karthikeyan, G., Kassebaum, N., Keren, A., Khoo, J.P., Knowlton, L.M., Kobusingye, O., Koranteng, A., Krishnamurthi, R., Lipnick, M., Lipshultz, S.E., Ohno, S.L., Mabweijano, J., MacIntyre, M.F., Mallinger,

- L., March, L., Marks, G.B., Marks, R., Matsumori, A., Matzopoulos, R., Mayosi, B.M., McAnulty, J.H., McDermott, M.M., McGrath, J., Mensah, G.A., Merriman, T.R., Michaud, C., Miller, M., Miller, T.R., Mock, C., Mocumbi, A.O., Mokdad, A.A., Moran, A., Mulholland, K., Nair, M.N., Naldi, L., Narayan, K.M.V., Nasser, K., Norman, P., O'Donnell, M., Omer, S.B., Ortblad, K., Osborne, R., Ozgediz, D., Pahari, B., Pandian, J.D., Rivero, A.P., Padilla, R.P., Perez-Ruiz, F., Perico, N., Phillips, D., Pierce, K., Pope, 3rd, C.A., Porrini, E., Pourmalek, F., Raju, M., Ranganathan, D., Rehm, J.T., Rein, D.B., Remuzzi, G., Rivara, F.P., Roberts, T., De León, F.R., Rosenfeld, L.C., Rushton, L., Sacco, R.L., Salomon, J.A., Sampson, U., Sanman, E., Schwebel, D.C., Segui-Gomez, M., Shepard, D.S., Singh, D., Singleton, J., Sliwa, K., Smith, E., Steer, A., Taylor, J.A., Thomas, B., Tleyjeh, I.M., Towbin, J.A., Truelsen, T., Undurraga, E.A., Venketasubramanian, N., Vijayakumar, L., Vos, T., Wagner, G.R., Wang, M., Wang, W., Watt, K., Weinstock, M.A., Weintraub, R., Wilkinson, J.D., Woolf, A.D., Wulf, S., Yeh, P.H., Yip, P., Zabetian, A., Zheng, Z.J., Lopez, A.D., Murray, C.J.L., AlMazroa, M.A., Memish, Z.A.: Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. *Lancet* **380** (2012) 2095–2128
2. World Health Organization: Global tuberculosis report 2012 (2012)
 3. World Health Organization: TB impact measurement: Policy and recommendations for how to assess the epidemiological burden of TB and the impact of TB control (2009)
 4. van der Werf, M.J., Enarson, D.A., Borgdorff, M.W.: How to identify tuberculosis cases in a prevalence survey. *International Journal of Tuberculosis and Lung Disease* **12** (2008) 1255–1260
 5. Leung, A.N.: Pulmonary tuberculosis: the essentials. *Radiology* **210** (1999) 307–322
 6. Philipsen, R., Maduskar, P., Hogeweg, L., van Ginneken, B.: Normalization of chest radiographs. In: *Medical Imaging. Volume 8670 of Proceedings of the SPIE.* (2013) 86700G
 7. van Ginneken, B., Stegmann, M.B., Loog, M.: Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Medical Image Analysis* **10** (2006) 19–40
 8. Florack, L.M.J., ter Haar Romeny, B.M., Viergever, M.A., Koenderink, J.J.: The Gaussian scale-space paradigm and the multiscale local jet. *International Journal of Computer Vision* **18** (1996) 61–75
 9. Maduskar, P., Hogeweg, L., Philipsen, R., van Ginneken, B.: Automated localization of costophrenic recesses and costophrenic angle measurement on frontal chest radiographs. In: *Medical Imaging. Volume 8670 of Proceedings of the SPIE.* (2013) 867038
 10. Hogeweg, L., Mol, C., de Jong, P.A., Dawson, R., Ayles, H., van Ginneken, B.: Fusion of local and global detection systems to detect tuberculosis in chest radiographs. In: *Medical Image Computing and Computer-Assisted Intervention. Volume 6363 of Lecture Notes in Computer Science.* (2010) 650–657
 11. Arzhaeva, Y., Tax, D.M.J., van Ginneken, B.: Dissimilarity-based classification in the absence of local ground truth: application to the diagnostic interpretation of chest radiographs. *Pattern Recognition* **42** (2009) 1768–1776
 12. van Ginneken, B., Katsuragawa, S., ter Haar Romeny, B.M., Doi, K., Viergever, M.A.: Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE Transactions on Medical Imaging* **21** (2002) 139–149

13. Loog, M., van Ginneken, B.: Static posterior probability fusion for signal detection: applications in the detection of interstitial diseases in chest radiographs. In: International Conference on Pattern Recognition. (2004) 644–647
14. Breiman, L.: Random forests. *Machine Learning* **45** (2001) 5–32
15. Den Boon, S., Bateman, E.D., Enarson, D.A., Borgdorff, M., Verver, S., Lombard, C.J., Iruen, E., Beyers, N., White, N.W.: Development and evaluation of a new chest radiograph reading and recording system for epidemiological surveys of tuberculosis and lung disease. *International Journal of Tuberculosis and Lung Disease* **9** (2005) 1088–1096