

Fusion of Local and Global Detection Systems to Detect Tuberculosis in Chest Radiographs

Laurens Hogeweg^{1,3}, Christian Mol^{1,3}, Pim A. de Jong², Rodney Dawson⁴,
Helen Ayles⁵, and Bram van Ginneken^{1,3}

¹ Image Sciences Institute, University Medical Center Utrecht, The Netherlands

² Department of Radiology, University Medical Center Utrecht, The Netherlands

³ Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre,
The Netherlands

⁴ University of Cape Town Lung Institute, Cape Town, South Africa

⁵ Department of Infectious and Tropical Diseases,
London School of Hygiene & Tropical Medicine, London, United Kingdom
l.hogeweg@rad.umcn.nl

Abstract. Automatic detection of tuberculosis (TB) on chest radiographs is a difficult problem because of the diverse presentation of the disease. A combination of detection systems for abnormalities and normal anatomy is used to improve detection performance. A textural abnormality detection system operating at the pixel level is combined with a clavicle detection system to suppress false positive responses. The output of a shape abnormality detection system operating at the image level is combined in a next step to further improve performance by reducing false negatives. Strategies for combining systems based on serial and parallel configurations were evaluated using the minimum, maximum, product, and mean probability combination rules. The performance of TB detection increased, as measured using the area under the ROC curve, from 0.67 for the textural abnormality detection system alone to 0.86 when the three systems were combined. The best result was achieved using the sum and product rule in a parallel combination of outputs.

Keywords: chest radiography, tuberculosis, computer aided detection, system combination.

1 Introduction

Tuberculosis (TB) is a major cause of mortality and morbidity worldwide, with 9.3 million new cases and 1.8 million deaths reported in 2007 [10]. Chest radiography is becoming increasingly important in the fight against TB, because in populations with a high prevalence of AIDS existing screening diagnostics such as sputum staining are less reliable. With the increasing availability of digital radiography, computer-aided detection (CAD) systems can be developed that could facilitate mass population screening for TB. This work is part of the larger CAD4TB project aimed at developing such a system.

On chest radiographs (CXR) the presentation of TB is diverse: textural abnormalities are often present, but focal abnormalities (e.g. nodules) and changes of the lung shape may be found as well. So far, only the problem of textural abnormality detection in TB detection has been addressed [1]. If the lungs have different shapes due to disease, texture analysis might miss abnormalities, an example of this is shown in Figure 3d. Another problem is that chest radiographs have a complex appearance caused by the projection of overlapping structures. This increases the difficulty of characterizing normal structures in the lung, such as clavicles, and can result in false positive responses of a textural abnormality detection system (Figure 3b).

These issues provided the motivation for this work, in which multiple detection systems are combined to improve detection performance. It is expected that combining systems detecting different types of abnormalities as well as systems detecting normal structures (segmentation) will lead to an overall improvement of a CAD system. The idea of system or classifier fusion has been extensively studied theoretically (e.g. [6]) and has also been applied to medical imaging (e.g. [3]). To our knowledge, this work is the first to apply such a method to chest radiographs (CXRs). Different strategies for combining subsystems are discussed.

2 Method

2.1 Detection Systems

The detection systems considered in this study are derived from pixel classification of CXRs. Features are calculated for a number of samples (positions) in an image. A classifier is trained using labeled samples from a database of training images. Examples of labels are inside/outside lung field and normal/abnormal structure. Samples in a test image are classified and assigned a probability of belonging to a particular label. Systems for lung field, clavicle, textural abnormality and shape abnormality detection will be shortly described below.

Lung field detection and clavicle detection (CD). A segmentation system for CXRs based on pixel classification to detect the lung fields and clavicles described by van Ginneken et al. [9] was slightly modified. Features in that system are based on Gaussian derivative filtered images calculated at different scales and in different directions. For clavicle detection two extra position features derived from the lung segmentation were added: the distance of a pixel to the border of the lung field segmentation and the distance to the center of gravity of the two detected lung fields.

Textural abnormality detection (TAD). The detection of textural abnormalities is based on texture analysis of small circular image patches (radius = 32 pixels) sampled every 8 pixels from radiographs downsampled to 1024×1024 pixels. Features are based on the moments of intensity distributions of Gaussian derivative filtered images sampled in a patch. This method has recently successfully been used to detect textural abnormalities related to TB in chest

radiographs [1]. The two lung segmentation derived position features described above were added to the texture features. A total of 106 features per patch were extracted. Image patches were sampled inside the segmented lung fields and classified as normal or abnormal using linear discriminant analysis (LDA) preceded by feature reduction using principal component analysis (PCA) (retaining 95% of the variance).

Shape abnormality detection (SAD). Segmentation of the unobscured lung fields can be done accurately in normal images, or in abnormal images with only small abnormalities such as nodules [7]. When large abnormalities close to the lung walls are present, segmentation of the lung fields becomes much more difficult as can be seen in Figure 3d,f. This causes the boundaries of the detected lung fields to be displaced and abnormalities to fall outside the detected lung fields, rendering them undetectable for the TAD system. In such situations the changed shape of the detected lung fields can be used to detect abnormal images.

A simplified version of shape contexts [2] is used to describe shapes. Instead of using multiple points on the shape to construct angle-distance histograms the centroid of the shape is used. Rays are cast in multiple directions from the centroid and the distance to the intersection with the boundary is recorded. This creates a feature vector of length n (the number of directions used) for each shape. The procedure is performed separately for each lung and the feature vectors are concatenated to obtain one describing both lung shapes.

Using a training set of normal and abnormal lung shapes a classifier can be trained to detect abnormal shapes. It was observed from preliminary experiments that abnormal lungs in the training database were very varied in shape and few in number. In a situation where the number of positive examples is small, it is difficult for two-class classifiers to obtain good performance and one-class classifiers might perform better [8]. A one-class classifier based on a PCA model of normal shapes (retaining 90% of the variance) was used to detect abnormal shapes. A large Mahalanobis distance (generalisation of the standard score) of a test shape to the model indicates that the shape is more abnormal. By thresholding this value an image can be classified as normal or abnormal.

2.2 Combining Detections

The three detection systems (TAD, CD and SAD) are combined to improve the performance of the system in discriminating between normal and abnormal images. The texture and clavicle detection output is combined at the pixel level, fused to an image decision and combined with the shape decision to give a final decision (Figure 1). Each step is now described in detail.

Clavicle corrected TAD. Clavicle induced false positive responses in the TAD system are reduced by combining it with the probabilistic output of the clavicle detection system. This gives one clavicle corrected output $p_{tcc} = p_t \cdot (1 - p_c)$ per pixel, where p_t and p_c are the probabilistic outputs of the TAD and CD systems respectively. The combination reduces the probability of abnormality from the TAD system when the probability of the clavicle being located at that position is high.

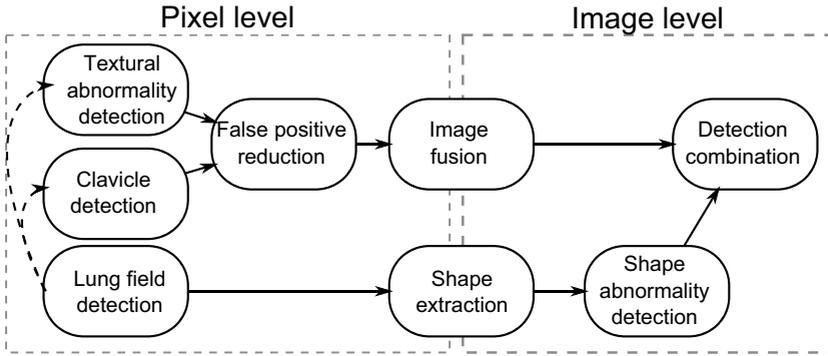


Fig. 1. Combination of detection systems. The detection combination is shown in detail in Figure 2

Image fusion. The texture detection system operates at the pixel level. Before combining it with the shape system the pixel decisions are fused to one image decision, which is performed using the quantile rule. This rule has been shown to have a good performance in texture classification for TB detection [1].

Detection combination. In general the output of a classification system - e.g. LDA, the quantile rule, Mahalanobis distance - is not a real probability estimate. Therefore, different outputs cannot be directly combined and first need to be converted to calibrated probability estimates in the range 0 – 1. In this work the probability of an output being normal is calibrated by taking the percentage of false positives (1 – specificity) produced by the classifier at a decision threshold equal to the output.

System combination (see Figure 2) can be performed in a serial fashion, where the output of one system is the input for the next one, or in a parallel fashion, in which outputs of multiple systems are combined. For serial combination the shape system serves as a filter for the texture system: if the output of the shape system is higher than a certain threshold the image is judged to be abnormal. The threshold is set to a calibrated probability of being abnormal of 0.95. Shape probabilities above this threshold are considered abnormal and set to 1, and 0 otherwise. In parallel combination the population probabilities were directly combined. Widely used combination rules - taking the minimum, maximum, product and mean value - were used to combine outputs.

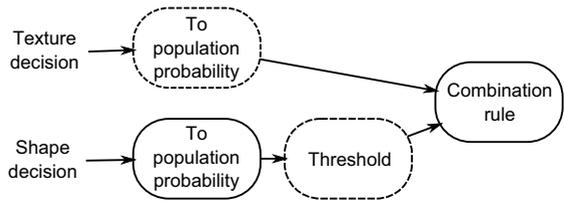


Fig. 2. Detail of detection combination of TAD and SAD system. Optional blocks are indicated by a dashed border.

2.3 Experiments

Data. A database containing 365 digital CXRs (2048×2048 resolution, pixel size 0.25 mm, Delft Imaging Systems, The Netherlands) was used. The training database consisted of 216 images (110 normal, 106 abnormal) for which full manual outlines of abnormalities were created by a chest radiologist in training. The training database contained 15 images with abnormal shapes as judged by the author of this work. The test database consisted of 149 consecutive images from a TB clinic in Africa of which the normal/abnormal decision was made by a local expert (69 normal, 80 abnormal). The database is part of a larger database of CXRs from areas in Africa with high TB incidence that is constructed within the CAD4TB project.

Analysis. Three types of systems were evaluated: (1) TAD, (2) TAD with clavicle correction, and (3) TAD with clavicle correction combined with SAD. The performance of the systems was evaluated using receiver operator characteristic (ROC) analysis and calculation of the area under the ROC curve (A_z). The different combination rules (minimum, maximum, product and mean) for system 3 were also evaluated.

3 Results

3.1 Detection Systems

Figure 3 shows examples of the effect of the clavicle correction and the combination of texture and shape system. Clavicle correction of the TAD system reduces false positive pixel responses (a,b,c) and the SAD system reduces false negatives when the texture inside the detected lung fields is normal (d,e,f).

3.2 Combination

Figure 4 shows a comparison of ROC curves for the different detection system combinations. The TAD system (1) achieved a performance, measured using A_z , of 0.69. When false positive responses were corrected using the CD system the performance of the TAD system increased to 0.75. The combination of outputs from the clavicle corrected TAD system and the SAD system increased the A_z to 0.85. Differences between A_z values were computed according to the method by Hanley et al. [4]: system 1 and 3, and 2 and 3 were significantly different ($p < 0.05$). The difference between 1 and 2 was not significant, but in the low false positive rate region of the ROC curve, which is important for screening, there was a clear improvement in sensitivity.

The effect of different combination strategies of the TAD system and the SAD system is shown in table 1. When outputs are combined in a parallel configuration (no threshold on shape probabilities) the sum and product rule give the best results. For a serial configuration (shape probabilities are thresholded) the sum and maximum rule perform best. In a serial configuration the maximum rule is

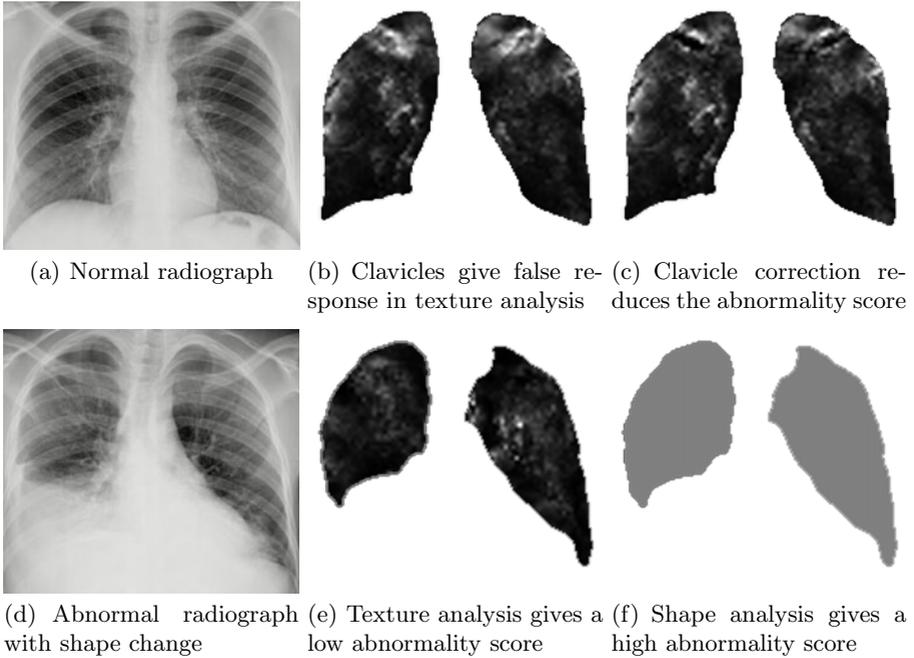


Fig. 3. Examples of results for the clavicle corrected TAD system (a,b,c) and the combination of the clavicle corrected TAD system and the SAD system (d,e,f). Figure b, c and e are maps indicating the probability of being abnormal for each pixel (white = high).

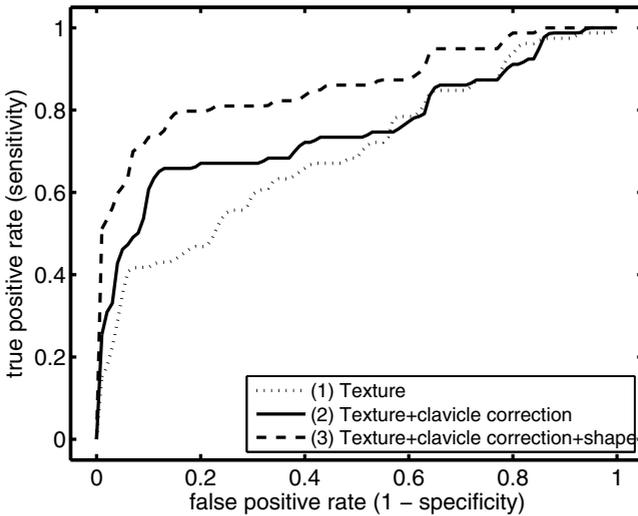


Fig. 4. ROC analysis of TAD system (1) and combined systems (2 and 3)

Table 1. Effect of different combination strategies (values indicate A_z)

Fusion method	Shape threshold		Practical effect of shape threshold
	No (parallel)	Yes (serial)	
Sum	0.86	0.85	Filtering of abnormal shapes
Maximum	0.84	0.85	Filtering of abnormal shapes
Minimum	0.81	0.76	Filtering of normal shapes
Product	0.86	0.76	Filtering of normal shapes

equivalent to preventing images with abnormal shapes to be further analyzed using texture analysis. The (unweighted) sum had the same practical effect. The minimum and product rule, which have the effect of analyzing only images which have abnormal shapes in the TAD system, perform worse than the other strategies.

4 Discussion and Conclusion

This paper described the combination of multiple subsystems to automatically detect tuberculosis in digital chest radiographs. The combination of a lung field, clavicle, textural abnormality and shape abnormality detection system improved detection performance.

The performance improvement of the addition of the shape abnormality detection system demonstrates the potential of combining systems at different levels (pixel and image). Alternatively the lung field segmentation algorithm could be made robust so that it can provide accurate lung field segmentations in the presence of pathology. Although an interesting research topic on its own the problem of segmenting abnormal images has not been solved yet [7]. Normal undiseased anatomy already shows substantial variation and this problem is even larger when pathology is present. Training a system which could solve that problem would require a far larger amount of training data than a system based on normal anatomy would use.

Parallel and serial configurations of the combination system were evaluated using a few well known probability combination rules. Generally, configurations and rules that retained the most information about the systems (sum rule and no thresholding) performed best. Depending on computational or other resource restraints (e.g. the choice of using expensive tests) other configurations might be useful in practical applications.

Instead of using multiple separate detection systems to detect disease in images it might be possible to design one large detection system to handle that task. However, there are a number of reasons to prefer multiple small systems over one large complex system. The generalization ability of a classification system drops when it becomes more complex [5], requiring larger amounts of training data to achieve good performance. From a perspective of system design, multiple highly specialized subsystems are also preferable since these can be more easily tested and evaluated than larger general ones. A practical reason to use multiple

subsystems, such as the clavicle and texture detection system, is that they can be developed parallelly by different research groups.

Within a framework for detection system combination such as presented in this paper other subsystems can be easily added. For example, in the application of TB detection in CXRs, detection of focal lesions and hilar abnormalities could further improve the performance of the whole system. It was demonstrated in this paper that the combination of multiple detection subsystems improves the detection of TB on CXRs.

References

1. Arzhaeva, Y., Hogeweg, L., de Jong, P.A., Viergever, M.A., van Ginneken, B.: Global and local multi-valued dissimilarity-based classification: Application to computer-aided detection of tuberculosis. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 724–731. Springer, Heidelberg (2009)
2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(4), 509–522 (2002)
3. Das, S.K., Chen, S., Deasy, J.O., Zhou, S., Yin, F., Marks, L.B.: Combining multiple models to generate consensus: application to radiation-induced pneumonitis prediction. *Medical Physics* 35(11), 5098–5109 (2008)
4. Hanley, J.A., McNeil, B.J.: A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148(3), 839–843 (1983)
5. Ho, T.K., Basu, M.: Complexity measures of supervised classification problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(3), 289–300 (2002)
6. Kuncheva, L.I.: A theoretical study on six classifier fusion strategies. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(2), 281–286 (2002)
7. Seghers, D., Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P.: Minimal shape and intensity cost path segmentation. *IEEE Transactions on Medical Imaging* 26(8), 1115–1129 (2007)
8. Tax, D.: One-class classification. PhD thesis, Delft University of Technology (2001)
9. 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(1), 19–40 (2006)
10. World Health Organization. WHO report 2009: Global tuberculosis control, epidemiology, strategy, financing (2009)