34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Detecting Tuberculosis in Radiographs Using Combined Lung Masks

Stefan Jaeger, Alexandros Karargyris, Sameer Antani, and George Thoma

Abstract— **Tuberculosis (TB) is a major health threat in many regions of the world, while diagnosing tuberculosis still remains a challenge. Mortality rates of patients with undiagnosed TB are high. Modern diagnostic techniques are often too slow or too expensive for highly-populated developing countries that bear the brunt of the disease. In an effort to reduce the burden of the disease, this paper presents an automated approach for detecting TB on conventional posteroanterior chest radiographs. The idea is to provide developing countries, which have limited access to radiological services and radiological expertise, with an inexpensive detection system that allows screening of large parts of the population in rural areas. In this paper, we present results produced by our TB screening system. We combine a lung shape model, a segmentation mask, and a simple intensity model to achieve a better segmentation mask for the lung. With the improved masks, we achieve an area under the ROC curve of more than 83%, measured on data compiled within a tuberculosis control program.**

I. INTRODUCTION

Tuberculosis (TB) is the second leading cause of death from an infectious disease worldwide, after HIV [1], killing at least 1.2 million people in 2010. With about one-third of the world's population infected with TB, and an estimated nine million new cases occurring every year, TB is a major global health problem. TB is caused by the bacillus Mycobacterium tuberculosis. It typically affects the lungs but can also affect sites outside the lungs. It spreads through the air when people with active TB cough, sneeze, or otherwise expel infectious bacteria. While mortality rates are high when left untreated, drug treatment with antibiotics drastically improves the chances of survival. Unfortunately, diagnosing TB is still a major challenge nowadays, as diagnostic tests are often too expensive or too slow.

Radiographs are an important part of any medical evaluation for TB [2], among microbiological smears, cultures, and skin tests. A reliable screening system for TB detection on radiographs is therefore a big step toward more powerful TB diagnostics. Generally speaking, TB-related abnormalities in chest x-rays are diffuse, and discriminating between normal anatomical structures and abnormal patterns is a hard problem. Manifestations of TB in chest x-rays are, for example, effusions, nodules, and miliary patterns [3].

In collaboration with AMPATH (The Academic Model Providing Access to Healthcare), which is a partnership between Moi University School of Medicine in Kenya and Indiana University, we plan to screen parts of Kenya's population for TB. Due to the large population and the lack of adequate radiological services in rural Kenya, an automated and computer-aided screening system is essential. Ideally, the system should allow medical personnel with little or no radiological background to screen large numbers of patients, so that at-risk individuals can be referred for further evaluation and treatment.

All authors are with the National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA, email: {stefan.jaeger,sameer.antani,alexandros.karargyris,george.thoma}@nih.gov

This paper extends our previous work [4]. Now, we evaluate our system on real-world data. We also use a more sophisticated lung segmentation by combining multiple segmentation masks. Using image processing techniques, we describe how we differentiate between normal and pathologic chest x-rays, allowing us to screen for the presence of tuberculosis and other lung diseases. We structure the paper as follows: Section II briefly summarizes related work. Section III describes the data sets we use for our experiments. Then, Section IV describes our approach, including lung segmentation, feature computation, and classification. Finally, Section V presents our results, followed by the conclusions.

II. RELATED WORK

Despite the increasing number of publications dealing with computer-aided diagnosis (CAD), only a few systems have been published for TB detection in chest x-rays [5], [6], [7], [8], [9]. In a recent survey, van Ginneken et al. state that forty-five years after the initial work on computer-aided diagnosis in chest radiology, there are still no systems that can accurately read chest radiographs [10], [11], [12]. There is little doubt that more research is needed to meet practical performance requirements. For example, Hogeweg et al. combined a textural abnormality detection system with a clavicle detection system to suppress false positive responses [6]. Note that the clavicle region is a notoriously difficult region for TB detection because the clavicles can obscure manifestations of TB in the apex of the lung. Freedman et al. showed in a recent study that an automatic suppression of ribs and clavicles in chest x-rays can significantly increase a radiologist's performance for nodule detection [13]. A cavity in the upper lung zones is a strong indicator that TB has developed into a highly infectious state [7]. Shen et al. therefore developed a hybrid knowledge-based Bayesian approach to detect cavities in these regions automatically [7]. Xu et al. approach the same problem with a model-based template matching approach and with image enhancement based on the Hessian matrix [8].

For lung segmentation, several different approaches have been proposed [14]. For example, van Ginneken et al. compared various techniques for lung segmentation, including active shapes, rule-based methods, pixel classification, and various combinations thereof [15], [16]. Their result was that pixel classification provided very good performance on their

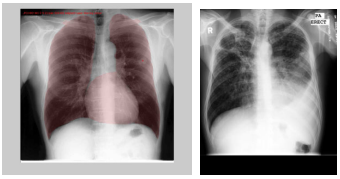

Fig. 1. X-rays from the JSRT database with segmentations for lung, heart, and clavicles (left) and from the MC database (right).

Fig. 2. Average lung mask (left) and another example x-ray from the MC data set (right).

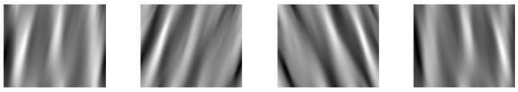

Fig. 3. Four random outputs (energy representations) of the Log Gabor filtering at largest scale and at orientations of $80°$, $70°$, $120°$, $110°$ degrees (from left to right).

test data. Another approach by Dawoud involves an iterative segmentation method that combines intensity information with shape priors [17].

III. DATA

We use two data sets of frontal chest x-rays for our experiments: a set compiled by the Japanese Society of Radiological Technology (JSRT) and a set from our local Department of Health and Human Services, Montgomery County, Maryland (MC).

The JSRT data set comprises 247 chest x-rays, among which 154 x-rays are abnormal and 93 x-rays are normal [18]. Each of the 154 abnormal x-rays features exactly one pulmonary nodule. All x-ray images collected have a size of 2048x2048 pixels and a gray-scale color depth of 12 bit. To augment the JSRT data, Ginneken et al. created the SCR database (Segmentation in Chest Radiographs) [15]. For each image in the JSRT database, the SCR database contains the corresponding manual segmentations for the lung field, the heart, and the clavicles. The left-hand side of Figure 1 shows a typical x-ray from the JSRT database, with segmentations for the lung field, the heart, and the clavicles superimposed. We use the JSRT data to train our lung model (see Section IV-A).

The MC data set is a small representative subset of a large x-ray repository collected over many years under the Montgomery County's Tuberculosis Control Program. Compared to the JSRT data set, which only contains nodules, the MC set contains a wide range of abnormalities most of which are TB-related. Each abnormal x-ray of the MC data set comes with a record that contains the reading of an expert radiologist. The MC data set contains 138 x-rays, among which 80 x-rays are normal and 58 x-rays are abnormal with manifestations of TB. All images are in 12-bit grayscale and their dimensions are either 1005x1223 or 1223x1005. The right-hand side of Figure 1 shows an abnormal example x-ray taken from the MC data set. According to its corresponding record, this x-ray shows extensive bilateral infiltrates, with a large cavity in the right upper lobe and a moderate pleural effusion on the left.

IV. APPROACH

For a given input x-ray, we first extract the lung field for which we then extract a set of features for classification. These steps are now discussed in more detail.

A. Lung Segmentation

We use an average of three different masks for lung segmentation: the intensity mask, the lung model mask, and the Log Gabor mask. The intensity mask is simply the complement of an x-ray. The intensity mask thus highlights the dark parts of an image, which are more likely to be part of the lung. The lung model mask is a probabilistic lung shape model that we compute using the JSRT data set. The left-hand side of Figure 2 displays the lung shape model as a gray-scale image. The pixel intensities in this image are the probabilities of the pixels being part of the lung field. We then use a bilinear alignment of the lung shape model to map the model to a given input x-ray.

The more complex Log Gabor mask is based on Log Gabor Wavelets. It is a true binary mask and presents a lung segmentation in its own right. Log Gabor filters have been introduced as an improvement over Gabor filters to obtain a larger spectral information while maintaining maximum spatial localization. Log Gabor filters can be constructed with arbitrary bandwidth, and the bandwidth can be optimized to produce a filter with minimal spatial extent [19]. In [20], Field defines Log Gabor filter as:

$$G(w) = e^{-\frac{\log\left(\frac{w}{w_0}\right)^2}{2\log\left(\frac{k}{w_0}\right)^2}}, \quad (1)$$

where w_0 is the filter's center frequency. In our previous paper, we demonstrated how, after some minor tuning of Log Gabor Wavelets, we can extract ribs from a chest x-ray [4]. Here, we use Log Gabor wavelets especially for the lung segmentation. In particular, we perform lung segmentation in three steps: In the first step, we apply a filter bank of 2 scales and 18 orientations to the input x-ray. Thus, the image is scanned from $0°$ to $180°$ in $10°$ steps. The right-hand side of Figure 2 shows another x-ray from the MC data set. Figure 3 shows outputs for various orientations from the convolution of our filter bank with this input image. In Figure 4, on the left-hand side, we see the sum of the energy outputs of the Log Gabor filter bank for the x-ray on the right-hand side of Figure 2. In the second step, after the Log Gabor filtering, we perform morphological operations

Fig. 4. Energy output of the Log Gabor filtering (left) and final Log Gabor mask (right).

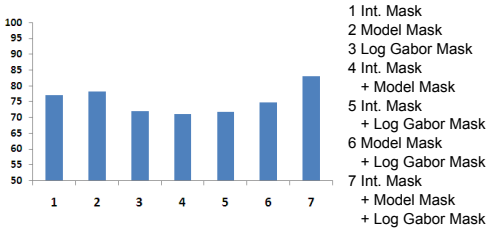

Fig. 5. AUC values for different segmentation combinations.

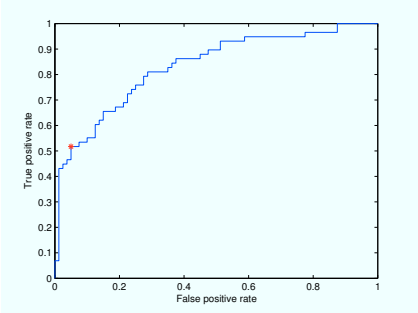

Fig. 6. ROC curve for the system with combined segmentation.

(closing) on the thresholded, inverted energy image. Finally, in Step 3, we search for regions (connected components) in the center of the image, as we know that the lung fields are two large regions close to the center of a chest x-ray. The right-hand side of Figure 4 shows the final Log Gabor mask for the input image on the right-hand side of Figure 2.

The final, overall lung segmentation mask is now the average of the intensity mask, the lung model mask, and the Log Gabor mask. To obtain a binary mask, we apply a threshold of 0.5; every pixel with an intensity higher than 0.5 is considered part of the lung.

B. *Classification*

To measure normal and abnormal patterns in the segmented lung field, we use a set of established shape- and texture descriptors. We then use histograms to represent the distribution of each descriptor across the lung field. The value of each histogram bin is a feature, and all features together form a feature vector that we input to our classifier. Through empirical experiments, we found that using 32 bins for each descriptor gives us good practical results. We therefore use this number of bins for the experiments reported here. In particular, we use the following features, which we describe in more detail in [21]: intensity histograms, gradient magnitude histograms, a shape descriptor, a curvature descriptor, the angle between the x-axis and the largest eigenvector, histograms of oriented gradients, and local binary patterns [22]. To compute the eigenvalues of the Hessian matrix needed for the shape and curvature descriptors, we use the multi-scale approach by Frangi et al. in [23], [24]. With each descriptor quantized into 32 histogram bins, our overall number of features is thus $7 * 32 = 224$.

Using these features, we train a linear support vector machine and classify a given input image into either normal or abnormal. We also compute the distance of the input pattern to the separating hyperplane of the SVM classifier to obtain a confidence value. We can use this confidence value to reject classifications that are borderline, which is important for practical applications. We also use these values for evaluation purposes, in particular to compute ROC curves (see the next section).

V. RESULTS

We evaluate the overall performance of our system on the MC data set. For each x-ray in the MC data set, we compute the features listed in Section IV-B and concatenate them into a single feature vector. We then apply a leave-one-out evaluation scheme, using the SVM-classifier. According to the leave-one-out scheme, we classify each feature vector (x-ray) in the MC data set with a classifier trained on the remaining feature vectors (x-rays) of the MC data set. We thus train as many classifiers as there are x-rays in the MC data set (138 altogether). Furthermore, we perform leave-one-out evaluations for all possible combinations of the segmentation masks mentioned above. Figure 5 shows the AUC (area under the ROC curve) for each combination. The y-axis of Figure 5 shows the AUC value, while the x-axis lists the different combinations. We see that both the intensity mask and the model mask produce an AUC value higher than 75% (see the first two columns in Figure 5). Using only the Log Gabor mask leads to a lower AUC (third column). A combination of the intensity mask with the model mask results in the lowest overall AUC value (Column 4). We also see that the Log Gabor mask reduces the performance when used in combination with either the intensity mask or the model mask (Column 5 and 6). However, the best performance is achieved by combining all masks together. In this case the AUC value is 83.12% and the overall accuracy of the system is almost 75%. Thus, the Log Gabor mask contains information complementary to the other masks, while the intensity mask and the model mask interfere with each other. Figure 6 shows the ROC curve for our system when we combine all segmentation masks. We see that for very low false positive rates of less than 10% (high specificity), we would miss about 50% of the true positive cases (recall). While this number seems high it is consistent with the few results that have been reported in the literature for different data sets and for different methods. Most papers evaluate only part of the detection problem, such as lung segmentation or cavity detection [8], [7]. Judging by the ROC curves, the performance of our system is comparable with the performance of the system reported in [5], [25]. Only

Hogeweg et al. recently reported a higher AUC value in [6]. They used a much larger training set for their experiments, though. For a fair comparison of all systems, we would have to evaluate each system on the same data set. Unfortunately, neither the training sets in [5], [25], [6] nor our own MC data set is publicly available.

VI. CONCLUSIONS

We are developing an automated system that screens chest x-rays for manifestations of TB and other lung diseases. For a given input chest x-ray, we first segment the lung field using a combination of an intensity mask, a statistical lung model mask, and a Log Gabor mask. We then extract a set of features for shapes, curvatures, and textures from the segmented lung field. Using the extracted features, we train a support vector machine that distinguishes between normal and abnormal x-rays. When evaluated on typical data from a TB control program, our system provides an AUC of 83.12%. This performance is comparable with other systems reported in the literature.

A feature of our approach is that we combine different masks to achieve a superior lung segmentation. It is also remarkable that we can apply a lung model trained on a Japanese data set to a very different data set (MC). We do not evaluate our segmentations by comparing them pixel-wise with a ground-truth segmentation, like most other researchers have done. Instead, we evaluate the performance of our segmentations directly on our MC data set, which we think is more practical.

Future steps to improve the performance, and increase the recall of our screening system, include improving the lung segmentation even further and using additional data that we are going to collect on-site in Kenya. For instance, we could use a dynamic lung model alignment instead of a static one, or combine even more segmentation masks. We will also consider partitioning the extracted lung field into regions and using local features for classification.

ACKNOWLEDGMENT

This research is supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC). We thank Dr. Sonia Qasba, Medical Director of Montgomery County's TB Control program, for providing us with x-rays and medical advice.

REFERENCES

[1] W. Report, *Global Tuberculosis Control 2011*. World Health Organization, 2011.

[2] C. Leung, "Reexamining the role of radiography in tuberculosis case finding," *The International Journal of Tuberculosis and Lung Disease*, vol. 15, no. 10, pp. 1279–1279, 2011.

[3] C. Daley, M. Gotway, and R. Jasmer, "Radiographic manifestations of tuberculosis," *A primer for Clinicians. San Francisco: Francis J. Curry National Tuberculosis Center*, 2009.

[4] A. Karargyris, S. Antani, and G. Thoma, "Segmenting anatomy in chest x-rays for tuberculosis screening," in *33rd Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS)*, 2011, pp. 7779–7782.

[5] B. van Ginneken, S. Katsuragawa, B. ter Haar Romeny, K. Doi, and M. Viergever, "Automatic detection of abnormalities in chest radiographs using local texture analysis," *Medical Imaging, IEEE Transactions on*, vol. 21, no. 2, pp. 139–149, 2002.

[6] L. Hogeweg, C. Mol, P. de Jong, R. Dawson, H. Ayles, and B. van Ginneken, "Fusion of local and global detection systems to detect tuberculosis in chest radiographs," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pp. 650–657, 2010.

[7] R. Shen, I. Cheng, and A. Basu, "A hybrid knowledge-guided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs," *Biomedical Engineering, IEEE Transactions on*, vol. 57, no. 11, pp. 2646–2656, 2010.

[8] T. Xu, I. Cheng, and M. Mandal, "Automated cavity detection of infectious pulmonary tuberculosis in chest radiographs," in *33rd Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS)*, 2011, pp. 5178–5181.

[9] A. Hoog, H. Meme, H. van Deutekom, A. Mithika, C. Olunga, F. Onyino, and M. Borgdorff, "High sensitivity of chest radiograph reading by clinical officers in a tuberculosis prevalence survey," *The International Journal of Tuberculosis and Lung Disease*, vol. 15, no. 10, pp. 1308–1314, 2011.

[10] B. van Ginneken, L. Hogeweg, and M. Prokop, "Computer-aided diagnosis in chest radiography: Beyond nodules," *European Journal of Radiology*, vol. 72, no. 2, pp. 226–230, 2009.

[11] G. Lodwick, "Computer-aided diagnosis in radiology: A research plan," *Investigative Radiology*, vol. 1, no. 1, p. 72, 1966.

[12] G. Lodwick, T. Keats, and J. Dorst, "The coding of roentgen images for computer analysis as applied to lung cancer," *Radiology*, vol. 81, no. 2, p. 185, 1963.

[13] M. Freedman, S. Lo, J. Seibel, and C. Bromley, "Lung nodules: Improved detection with software that suppresses the rib and clavicle on chest radiographs," *Radiology*, vol. 260, no. 1, pp. 265–273, 2011.

[14] B. Van Ginneken, B. ter Haar Romeny, and M. Viergever, "Computer-aided diagnosis in chest radiography: a survey," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 12, pp. 1228–1241, 2001.

[15] B. Van Ginneken, M. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, vol. 10, no. 1, pp. 19–40, 2006.

[16] B. van Ginneken and B. ter Haar Romeny, "Automatic segmentation of lung fields in chest radiographs," *Medical Physics*, vol. 27, no. 10, pp. 2445–2455, 2000.

[17] A. Dawoud, "Fusing shape information in lung segmentation in chest radiographs," *Image Analysis and Recognition*, pp. 70–78, 2010.

[18] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule," *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, 2000.

[19] "What are Log-Gabor filters and why are they good?" last viewed in March 2012, http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/PhaseCongruency/Docs/convexpl.html.

[20] D. Field *et al.*, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[21] S. Jaeger, C. Casas-Delucchi, M. Cardoso, and K. Palaniappan, "Classification of cell cycle phases in 3D confocal microscopy using PCNA and chromocenter features," in *7th Indian Conference on Computer Vision, Graphics, and Image Processing*, Chennai (India), 2010, pp. 412–418.

[22] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, pp. 51–59, 1996.

[23] A. Frangi, W. Niessen, K. Vincken, and M. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 1998, pp. 130–137.

[24] R. Kerekes, S. Gleason, N. Trivedi, and D. Solecki, "Automated 3-D tracking of centrosomes in sequences of confocal image stacks," in *Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, vol. 1, 2009, pp. 6994–6997.

[25] Y. Arzhaeva, D. Tax, and B. Van Ginneken, "Dissimilarity-based classification in the absence of local ground truth: Application to the diagnostic interpretation of chest radiographs," *Pattern Recognition*, vol. 42, no. 9, pp. 1768–1776, 2009.