Automatic Tuberculosis Screening Using Chest Radiographs

Stefan Jaeger*, Alexandros Karargyris, Sema Candemir, Les Folio, Jenifer Siegelman, Fiona Callaghan, Zhiyun Xue, Kannappan Palaniappan, Rahul K. Singh, Sameer Antani, George Thoma, Yi-Xiang Wang, Pu-Xuan Lu, and Clement J. McDonald

Abstract—Tuberculosis is a major health threat in many regions of the world. Opportunistic infections in immunocompromised HIV/AIDS patients and multi-drug-resistant bacterial strains have exacerbated the problem, while diagnosing tuberculosis still remains a challenge. When left undiagnosed and thus untreated, mortality rates of patients with tuberculosis are high. Standard diagnostics still rely on methods developed in the last century. They are slow and often unreliable. In an effort to reduce the burden of the disease, this paper presents our automated approach for detecting tuberculosis in conventional posteroanterior chest radiographs. We first extract the lung region using a graph cut segmentation method. For this lung region, we compute a set of texture and shape features, which enable the X-rays to be classified as normal or abnormal using a binary classifier. We measure the performance of our system on two datasets: a set collected by the tuberculosis control program of our local county's health department in the United States, and a set collected by Shenzhen Hospital, China. The proposed computer-aided diagnostic system for TB screening, which is ready for field deployment, achieves a performance that approaches the performance of human experts. We achieve an area under the ROC curve (AUC) of 87% (78.3% accuracy) for the first set, and an AUC of 90% (84% accuracy) for the second set. For the first set, we compare our system performance with the performance of radiologists. When trying not to miss any positive cases, radiologists achieve an accuracy of about 82% on this set, and their false positive rate is about half of our system's rate.

Index Terms—Computer-aided detection and diagnosis, lung, pattern recognition and classification, segmentation, tuberculosis (TB), X-ray imaging.

I. INTRODUCTION

TUBERCULOSIS (TB) is the second leading cause of death from an infectious disease worldwide, after HIV, with a mortality rate of over 1.2 million people in 2010 [1]. With about one-third of the world's population having latent TB, and an estimated nine million new cases occurring every year, TB is a major global health problem [2]. TB is an infectious disease caused by the bacillus Mycobacterium tuberculosis, which typically affects the lungs. It spreads through the air when people with active TB cough, sneeze, or otherwise expel infectious bacteria. TB is most prevalent in sub-Saharan Africa and Southeast Asia, where widespread poverty and malnutrition reduce resistance to the disease. Moreover, opportunistic infections in immunocompromised HIV/AIDS patients have exacerbated the problem [3]. The increasing appearance of multi-drug resistant TB has further created an urgent need for a cost effective screening technology to monitor progress during treatment.

Several antibiotics exist for treating TB. While mortality rates are high when left untreated, treatment with antibiotics greatly improves the chances of survival. In clinical trials, cure rates over 90% have been documented [1]. Unfortunately, diagnosing TB is still a major challenge. The definitive test for TB is the identification of Mycobacterium tuberculosis in a clinical sputum or pus sample, which is the current gold standard [2], [3]. However, it may take several months to identify this slow-growing organism in the laboratory. Another technique is sputum smear microscopy, in which bacteria in sputum samples are observed under a microscope. This technique was developed more than 100 years ago [1]. In addition, several skin tests based on immune response are available for determining whether an individual has contracted TB. However, skin tests

*S. Jaeger is with the Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, Bethesda, MD 20894 USA (e-mail: stefan.jaeger@nih.gov).

A. Karargyris, S. Candemir, F. Callaghan, Z. Xue, S. Antani, G. Thoma, and C. J. McDonald are with the Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, Bethesda, MD 20894 USA (e-mail: alexandros.karargyris@nih.gov; sema.candemir@nih.gov; fiona.callaghan@nih.gov; zhiyun.xue@nih.gov; sameer.antani@nih.gov; george.thoma@nih.gov; clement.mcdonald@nih.gov).

L. Folio is with Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, MD 20892 USA (e-mail: les.folio@nih.gov).

J. Siegelman is with the Department of Radiology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115 USA (e-mail: jen.siegelman@gmail.com).

K. Palaniappan and R. K. Singh are with the Department of Computer Science, University of Missouri-Columbia, Columbia, MO 65211 USA (e-mail: palaniappank@missouri.edu; rksc82@mail.missouri.edu).

Y.-X. Wang is with the Department of Imaging and Interventional Radiology, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, Hong Kong, China (e-mail: yixiang_wang@cuhk.edu.hk).

P.-X. Lu is with the Department of Radiology, The Shenzhen No. 3 People's Hospital, Guangdong Medical College, Shenzhen 518020, China (e-mail: lupuxuan@126.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

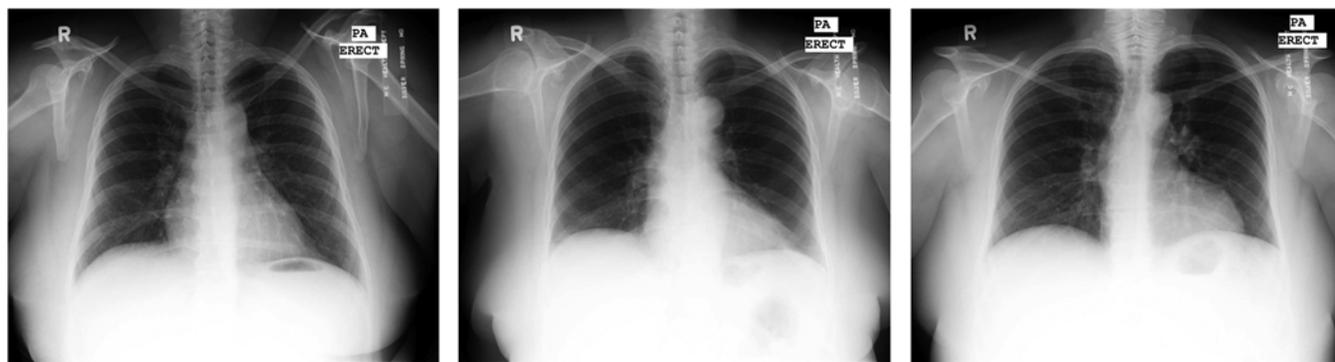Digital Object Identifier 10.1109/TMI.2013.2284099

Fig. 1.   Examples of normal CXRs in the MC dataset.

are not always reliable. The latest development for detection are molecular diagnostic tests that are fast and accurate, and that are highly sensitive and specific. However, further financial support is required for these tests to become commonplace [1]–[3].

In this paper, we present an automated approach for detecting TB manifestations in chest X-rays (CXRs), based on our earlier work in lung segmentation and lung disease classification [4]–[6]. An automated approach to X-ray reading allows mass screening of large populations that could not be managed manually. A posteroanterior radiograph (X-ray) of a patient's chest is a mandatory part of every evaluation for TB [7]. The chest radiograph includes all thoracic anatomy and provides a high yield, given the low cost and single source [8]. Therefore, a reliable screening system for TB detection using radiographs would be a critical step towards more powerful TB diagnostics. The TB detection system presented here is a prototype that we developed for AMPATH (The Academic Model Providing Access to Healthcare) [9]. AMPATH is a partnership between Moi University School of Medicine and Moi Teaching and Referral Hospital, Kenya, and a consortium of U.S. medical schools under the leadership of Indiana University. AMPATH provides drug treatment and health education for HIV/AIDS control in Kenya. HIV and TB co-infections are very common due to the weakened immune system. It is therefore important to detect patients with TB infections, not only to cure the TB infection itself but also to avoid drug incompatibilities. However, the shortage of radiological services in Kenya necessitates both an efficient and inexpensive screening system for TB. Medical personnel with little radiology background need to be able to operate the screening system. The target platform for our automated system are portable X-ray scanners, which allow screening of large parts of the population in rural areas. At-risk individuals identified by our system are then referred to a major hospital for treatment.

Fig. 1 shows examples of normal CXRs without signs of TB. These examples are from our Montgomery County (MC) dataset that we describe in more detail in Section III. Fig. 2 shows positive examples with manifestations of TB, which are from the same dataset. Typical manifestations of TB in chest X-rays are, for example, infiltrations, cavitations, effusions, or miliary patterns. For instance, CXR A and C in Fig. 2 have infiltrates in both lungs. CXR B is a good example of pleural TB, which is indicated by the abnormal shape of the costophrenic angle of the

right lung. In CXR D, we see irregular infiltrates in the left lung with a large area of cavitation. Additionally, there is scarring in the right apical region. CXR E shows peripheral infiltrates in the left lung. Finally, CXR F shows TB scars resulting from an older TB infection. Readers can find more illustrative examples of abnormal CXRs with TB in [8], [10]–[13].

In this paper, we describe how we discriminate between normal and abnormal CXRs with manifestations of TB, using image processing techniques. We structure the paper as follows. Section II discusses related work and shows the state-of-the-art. Section III briefly describes the datasets we use for our experiments. In Section IV, we present our approach with lung segmentation, feature computation, and classification. A presentation of our practical experiments follows in Section V. Finally, a brief summary with the main results concludes the paper. Note that some of the features we use in this paper are identical to the features used in one of our earlier publications [6]. However, the lung boundary detection algorithm in this paper differs from the one used in our earlier publication.

II. RELATED WORK

The advent of digital chest radiography and the possibility of digital image processing has given new impetus to computer-aided screening and diagnosis. Still, despite its omnipresence in medical practice, the standard CXR is a very complex imaging tool. In the last 10 years, several ground-breaking papers have been published on computer-aided diagnosis (CAD) in CXRs. However, there is no doubt that more research is needed to meet the practical performance requirements for deployable diagnostic systems. In a recent survey, van Ginneken *et al.* state that 45 years after the initial work on computer-aided diagnosis in chest radiology, there are still no systems that can accurately read chest radiographs [14]–[16]. Automated nodule detection is becoming one of the more mature applications of decision support/automation for CXR and CT. Several studies have been published evaluating the capability of commercially available CAD systems to detect lung nodules [17]–[19]. The result is that CAD systems can successfully assist radiologists in diagnosing lung cancer [20]. However, nodules represent only one of many manifestations of TB in radiographs.

In recent years, due to the complexity of developing full-fledged CAD systems for X-ray analysis, research has concentrated on developing solutions for specific subproblems [14], [21]. The segmentation of the lung field is a typical task that any
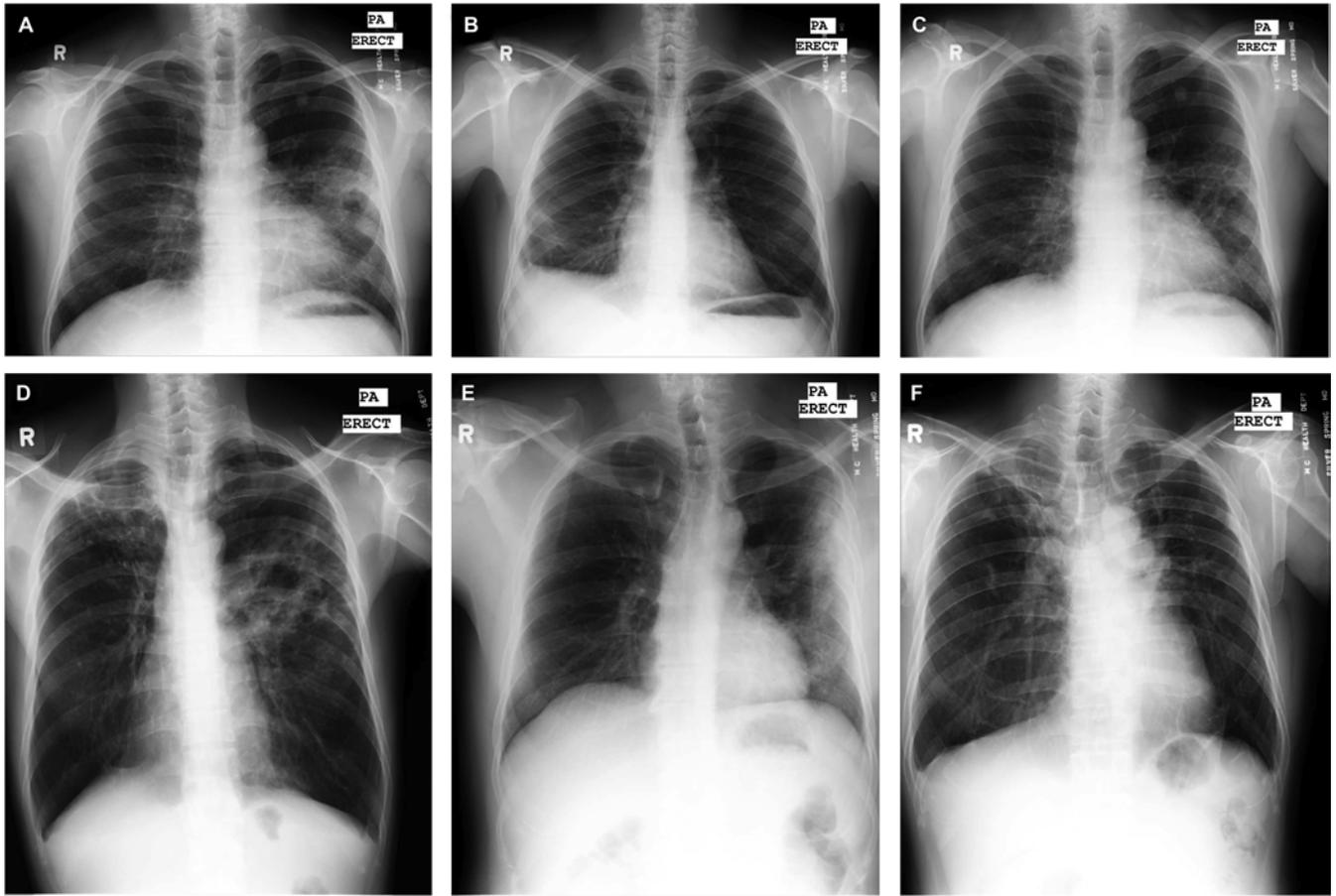
Fig. 2. Examples of abnormal CXRs in the MC dataset. CXR A has a cavitary infiltrate on the left and a subtle infiltrate in the right lower lung. CXR B is an example of pleural TB. Note that the blunted right costophrenic angle indicates a moderate effusion. CXR C has infiltrates in both lungs. CXR D shows irregular infiltrates in the left lung with cavitation and scarring of the right apex. CXR E shows peripheral infiltrates in the left lung. CXR F shows signs of TB, indicated by the retraction of bilateral hila superiorly, which is more pronounced on the right.

CAD system needs to support for a proper evaluation of CXRs. Other segmentations that may be helpful include the segmentation of the ribs, heart, and clavicles [22]. For example, van Ginneken *et al.* compared various techniques for lung segmentation, including active shapes, rule-based methods, pixel classification, and various combinations thereof [22], [23]. Their conclusion was that pixel classification provided very good performance on their test data. Dawoud presented an iterative segmentation approach that combines intensity information with shape priors trained on the publicly available JSRT database (see Section III) [24].

Depending on the lung segmentation, different feature types and ways to aggregate them have been reported in the literature. For example, van Ginneken *et al.* subdivide the lung into overlapping regions of various sizes and extract features from each region [25]. To detect abnormal signs of diffuse textural nature they use the moments of responses to a multiscale filter bank. In addition, they use the difference between corresponding regions in the left and right lung fields as features. A separate training set is constructed for each region and final classification is done by voting and a weighted integration.

Many of the CAD papers dealing with abnormalities in chest radiographs do so without focusing on any specific disease. Only a few CAD systems specializing in TB detection have been published, such as [25]–[28]. For example, Hogeweg *et al.* combined a texture-based abnormality detection system with a clavicle detection stage to suppress false positive responses [26]. In [29], the same group uses a combination of pixel classifiers and active shape models for clavicle segmentation. 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 CXRs can significantly increase a radiologist's performance for nodule detection [30]. A cavity in the upper lung zones is a strong indicator that TB has developed into a highly infectious state [27]. Shen *et al.* therefore developed a hybrid knowledge-based Bayesian approach to detect cavities in these regions automatically [27]. Xu *et al.* approached the same problem with a model-based template matching technique, with image enhancement based on the Hessian matrix [28].

Arzhaeva *et al.* use dissimilarity-based classification to cope with CXRs for which the abnormality is known but the precise location of the disease is unknown [31]. They report classification rates comparable to rates achieved with region classification on CXRs with known disease locations. More information on existing TB screening systems can be found in our recent survey [32].

In addition to X-ray based CAD systems for TB detection, several systems based on other diagnostic means have been reported in the literature. For example, Pangilinan *et al.* presented a stepwise binary classification approach for reduction of false positives in tuberculosis detection from smeared slides [33]. Furthermore, automated systems based on bacteriological examination with new diagnostic tests have been reported recently, such as GeneXpert (Cepheid, Sunnyvale, CA, USA) [34]. Currently, these tests are still expensive. Nevertheless, with costs decreasing over time, these systems may become an option for poorer countries. It is also possible, and indeed very promising, to combine these new systems with X-ray based systems. For the time being, however, these systems are out of the scope of this paper.

III. Data

For our experiments, we use three CXR sets. On the first two sets we train and test our classifiers, and on the third set we train our lung models. The images used in this study were de-identified by the data providers and are exempted from IRB review at their institutions. The data was exempted from IRB review (No. 5357) by the NIH Office of Human Research Protections Programs.

Our first set, the MC set, is a representative subset of a larger CXR repository collected over many years within the tuberculosis control program of the Department of Health and Human Services of Montgomery County (MC), Maryland [6]. The MC set contains 138 posteroanterior CXRs, among which 80 CXRs are normal and 58 CXRs are abnormal with manifestations of TB. All images of the MC set are in 12-bit grayscale, captured with an Eureka stationary X-ray machine (CR). The abnormal CXRs cover a wide range of TB-related abnormalities, including effusions and miliary patterns. For the MC set, we know the ground-truth radiology reports that have been confirmed by clinical tests, patient history, etc.

Our second CXR set, the Shenzhen set, is from Shenzhen No.3 Hospital in Shenzhen, Guangdong providence, China. Shenzhen Hospital is one of the largest hospitals in China for infectious diseases, with a focus both on their prevention and treatment. The CXRs we received from Shenzhen Hospital are from outpatient clinics. They were captured within a one month period, mostly in September 2012, as part of the daily routine at Shenzhen Hospital, using a Philips DR Digital Diagnost system. The set contains 340 normal CXRs and 275 abnormal CXRs with TB. For the Shenzhen set, we have the radiologist readings, which we consider as ground-truth.

We train our lung models on a third set from the Japanese Society of Radiological Technology (JSRT). The JSRT data is the result of a study investigating the detection performance of radiologists for solitary pulmonary nodules [35]. The data was collected from 14 medical centers and comprises 247 CXRs. All CXR images have a size of 2048×2048 pixels and a grayscale color depth of 12 bits. Among the 247 CXRs, 93 CXRs are normal and 154 CXRs are abnormal. Each of the abnormal CXRs contains one pulmonary nodule classified into one of five degrees of subtlety, ranging from extremely subtle to obvious. However, in the JSRT images, the nodules hardly affect the lung shapes. The nodules are either well within the lung boundary or
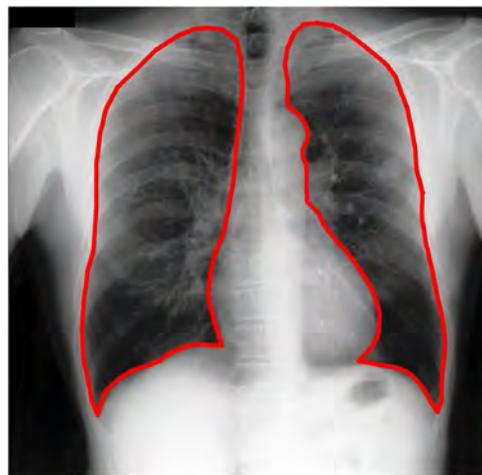

Fig. 3. JSRT CXR with manual ground-truth lung segmentation. Note the nodule overlying the left posterior fifth and sixth ribs.

they are so subtle that the effects on lung shape are minor. We can therefore take advantage of the entire JSRT database to train our shape model for a typical normal lung. To do so, we use the segmentation masks provided by van Ginneken *et al.* [22]. Their SCR dataset (Segmentation in Chest Radiographs) contains the manually generated lung field masks for each CXR in the JSRT database. For example, Fig. 3 shows an abnormal CXR from the JSRT database together with the outline of the left and the right lung as specified in the SCR data. Note that the left mid lung field in Fig. 3 contains a cancer nodule.

IV. Method

This section presents our implemented methods for lung segmentation, feature computation, and classification. Fig. 4 shows the architecture of our system with the different processing steps, which the following sections will discuss in more detail. First, our system segments the lung of the input CXR using a graph cut optimization method in combination with a lung model. For the segmented lung field, our system then computes a set of features as input to a pre-trained binary classifier. Finally, using decision rules and thresholds, the classifier outputs its confidence in classifying the input CXR as a TB positive case, for example.

A. Graph Cut Based Lung Segmentation

We model lung segmentation as an optimization problem that takes properties of lung boundaries, regions, and shapes into account [4], [72]. In general, segmentation in medical images has to cope with poor contrast, acquisition noise due to hardware constraints, and anatomical shape variations. Lung segmentation is no exception in this regard. We therefore incorporate a lung model that represents the average lung shape of selected training masks. We select these masks according to their shape similarity as follows. We first linearly align all training masks to a given input CXR. Then, we compute the vertical and horizontal intensity projections of the histogram equalized images. To measure the similarity between projections of the input CXR and the training CXRs, we use the Bhattacharyya coefficient. We then use the average mask computed on a subset of the most
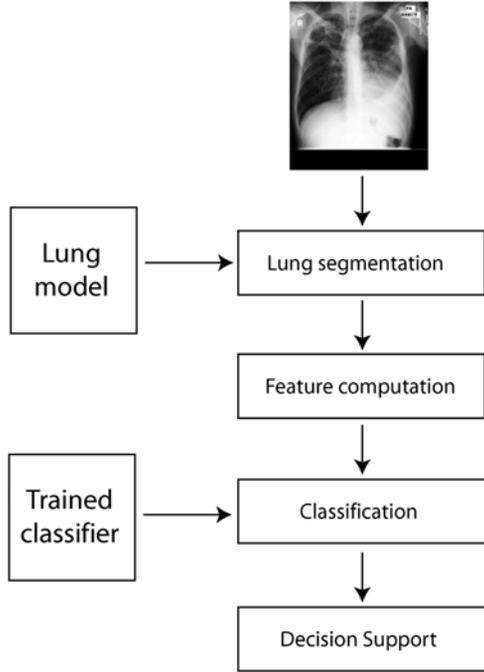
Fig. 4. System overview. The system takes a CXR as input and outputs a confidence value indicating the degree of abnormality for the input CXR.

Fig. 5. CXR and its calculated lung model.

similar training masks as an approximate lung model for the input CXR. In particular, we use a subset containing the five most similar training masks to compute the lung model. This empirical number produced the best results in our experiments. Increasing the subset size to more than five masks will decrease the lung model accuracy because the shapes of the additional masks will typically differ from the shape of the input X-ray.

As training masks, we use the publicly available JSRT set [35] for which ground truth lung masks are available [22]. The pixel intensities of the lung model are the probabilities of the pixels being part of the lung field. Fig. 5 shows a typical lung model we computed. Note that the ground-truth masks do not include the posterior inferior lung region behind the diaphragm. Our approach, and most segmentation approaches in the literature, exclude this region because manifestations of TB are less likely here.

In a second step, we employ a graph cut approach [36] and model the lung boundary detection with an objective function. To formulate the objective function, we define three requirements a lung region has to satisfy: 1) the lung region should be consistent with typical CXR intensities expected in a lung

region, 2) neighboring pixels should have consistent labels, and 3) the lung region needs to be similar to the lung model we computed. Mathematically, we can describe the resulting optimization problem as follows [72]: Let $f = \{f_1, f_2, \ldots, f_p, \ldots, f_N\}$ be a binary vector whose components f_p correspond to foreground (lung region) and background label assignments to pixel $p \in P$, where P is the set of pixels in the CXR, and N is the number of pixels. According to our method, the optimal configuration of f is given by the minimization of the following objective function:

$$E(f) = E_d(f) + E_s(f) + E_m(f) \qquad (1)$$

where E_d, E_s, and E_m represent the region, boundary, and lung model properties of the CXR, respectively. The region term $E_d(f)$ considers image intensities as follows:

$$E_d(f) = \frac{1}{I_{\max}} \left(\sum_{(p,S) \in C} |I_p - I_S| + \sum_{(p,T) \in C} |I_p - I_T| \right) \quad (2)$$

where I_p is the intensity of pixel p and C is the set of edges representing the cut. I_S and I_T are the intensities of foreground and background regions. We learn these intensities on the training masks and represent them using a source (S) and terminal node (T). $I_{\max}$ is the maximum intensity value of the input image. (2) ensures that labels for each pixel are assigned based on the pixel's similarity to the foreground and background intensities.

The boundary constraints between lung border pixels p and q are formulated as follows:

$$E_s(f) = \sum_{(p,q) \in C} \exp \left(-(I_p - I_q)^2 \right). \qquad (3)$$

This term uses the sum of the exponential intensity differences of pixels defining the cut. The sum is minimum when the intensity differences are maximum.

Our average lung model is a 2-D array which contains the probabilities of a pixel p being part of the lung field. Based on this model, we define the lung region requirement as follows:

$$E_m(f) = \sum_{(p,T) \in C} Pr_p + \sum_{(p,S) \in C} (1 - Pr_p) \qquad (4)$$

where Pr_p is the probability of pixel p being part of the lung model. This term describes the probability of pixels labeled as lung belonging to the background, and the probability of pixels labeled as background belonging to the lung, according to the lung model. We want to minimize both probabilities.

Using the three energy terms given above, we minimize the objective function with a fast implementation of min-cut/max-flow algorithm [37]. The minimum-cut is then the optimal foreground/background configuration of f for the input CXR [4]. Note that (4) extends our earlier work in [5], in which we did not use a lung model. Compared to our work in [4], we simplified (2), (3), and (4) so that they describe properties of the cut.

B. Features

To describe normal and abnormal patterns in the segmented lung field, we experimented with two different feature sets. Our

motivation is to use features that can pick up subtle structures in an CXR.

1) Object Detection Inspired Features—Set A: As our first set, we use features that we have successfully applied to microscopy images of cells for which we classified the cell cycle phase based on appearance patterns [38], [39]. It is the same set that we have used in our earlier TB classification work [6]. This set is versatile and can also be applied to object detection applications, for example in [40]–[42].

The first set is a combination of shape, edge, and texture descriptors [6]. For each descriptor, we compute a histogram that shows the distribution of the different descriptor values across the lung field. Each histogram bin is a feature, and all features of all descriptors put together form a feature vector that we input to our classifier. Through empirical experiments, we found that using 32 bins for each histogram gives us good practical results [40], [41]. In particular, we use the following shape and texture descriptors [38], [39].

- Intensity histograms (IH).
- Gradient magnitude histograms (GM).
- Shape descriptor histograms (SD)

$$\mathrm{SD} = \tan^{-1}\left(\frac{\lambda_1}{\lambda_2}\right) \quad (5)$$

  where λ_1 and λ_2 are the eigenvalues of the Hessian matrix, with $\lambda_1 \leq \lambda_2$.
- Curvature descriptor histograms (CD)

$$\mathrm{CD} = \tan^{-1}\left(\frac{\sqrt{\lambda_1^2 + \lambda_2^2}}{1 + I(x,y)}\right) \quad (6)$$

  with $0 \leq CD \leq \pi/2$, where $I(x,y)$ denotes the pixel intensity for pixel (x,y). The normalization with respect to intensity makes this descriptor independent of image brightness.
- Histogram of oriented gradients (HOG) is a descriptor for gradient orientations weighted according to gradient magnitude [43]. The image is divided into small connected regions, and for each region a histogram of gradient directions or edge orientations for the pixels within the region is computed. The combination of these histograms represents the descriptor. HOG has been successfully used in many detection systems [40], [43]–[46].
- Local binary patterns (LBP) is a texture descriptor that codes the intensity differences between neighboring pixels by a histogram of binary patterns [47], [48]. LBP is thus a histogram method in itself. The binary patterns are generated by thresholding the relative intensity between the central pixel and its neighboring pixels. Because of its computational simplicity and efficiency, LBP is successfully used in various computer vision applications [49], often in combination with HOG [38]–[40], [42], [50], [51].

With each descriptor quantized into 32 histogram bins, our overall number of features is thus $6 * 32 = 192$.

The eigenvalues of the Hessian matrix needed for the shape and curvature descriptors in (5) and (6) are computed using a modification of the multiscale approach by Frangi *et al.* [52], [53]. The Hessian describes the second-order surface curvature

properties of the local image intensity surface. This can be seen from the local behavior in the vicinity of a pixel x_o in an image I by means of the second-order Taylor series expansion

$$I(x,s) \approx I(x_0, s) + (x - x_0)^T \nabla I(x_0, s)$$
$$+ \frac{1}{2}(x - x_0)^T \mathcal{H}(x_0, s)(x - x_0) \quad (7)$$

where $\nabla I(x_0, s)$ stands for the gradient vector and $\mathcal{H}(x_0, s)$ is the Hessian matrix, both computed at pixel x_o and scale s. For each scale, we apply Gaussian filters as follows [52]:

$$\frac{\partial}{\partial x} I(x_o, s) = s^\gamma I(x_o) * \frac{\partial}{\partial x} G(x_o, s) \quad (8)$$

where $G(x_o, s)$ is the n-dimensional Gaussian for pixel x_o and scale s, and γ is a weight parameter. With $\gamma = 1$, all scales are weighted equally. The Gaussian $G(x_o, s)$ is given by

$$G(x_o, s) = \frac{1}{\sqrt{(2\pi s^2)}^n} e^{-\frac{\|x_o\|^2}{2s^2}}. \quad (9)$$

Our approach uses the maximum filter response across all scales. The main application in [52] (Frangi *et al.*) was to enhance blood vessels, which have mostly thin elongated shapes, through filtering. On the other hand, our goal is to detect nodular patterns by capturing the spherical or elliptical shapes in the local intensity curvature surface. We therefore use the following structure response SR filter based on the eigenvalues λ_1 and λ_2:

$$\mathrm{SR} = 1 - e^{-\sqrt{k \cdot |\lambda_1 * \lambda_2|}}. \quad (10)$$

A large filter response value of SR indicates the presence of large circular or elongated blobs, and is designed to detect nodular features in CXR images. In the case of very thin linear features, the structural response tends toward zero, with $|\lambda_2| \gg |\lambda_1| \approx 0$. For very large eigenvalues the filter response approaches one. The final eigenvalues that we use to compute the shape descriptor SD and the curvature descriptor CD are the eigenvalues that provide the largest filter response over 10 different Gaussian filter scales, namely $s = 2, 4, 6, \ldots, 20$.

2) CBIR-Based Image Features—Set B: For our second feature set, Set B, we use a group of low-level features motivated by content-based image retrieval (CBIR) [54], [55]. This feature collection includes intensity, edge, texture and shape moment features, which are typically used by CBIR systems. The entire feature vector has 594 dimensions, which is more than three times larger than the feature vector of Set A, and which allows us to evaluate the effect of high-dimensional feature spaces on classification accuracy. We extract most of the features, except for Hu moments and shape features, based on the Lucene image retrieval library, LIRE [56]–[58]. In particular, Feature Set B contains the following features.

- Tamura texture descriptor: The Tamura descriptor is motivated by the human visual perception [59]. The descriptor comprises a set of six features. We only use three of these features, which have the strongest correlation with human perception: contrast, directionality, and coarseness.
- CEDD and FCTH: CEDD (color and edge direction descriptor) [60] and FCTH (fuzzy color and texture histogram) [61] incorporate color and texture information in

one histogram. They differ in the way they capture texture information.

- Hu moments: These moments are widely used in image analysis. They are invariant under image scaling, translation, and rotation [62]. We use the DISCOVIR system (distributed content-based visual information retrieval) to extract Hu moments [63].
- CLD and EHD edge direction features: CLD (color layout descriptor) and EHD (edge histogram descriptor) are MPEG-7 features [64]. CLD captures the spatial layout of the dominant colors on an image grid consisting of 8×8 blocks and is represented using DCT (discrete cosine transform) coefficients. EHD represents the local edge distribution in the image, i.e., the relative frequency of occurrence of five types of edges (vertical, horizontal, 45° diagonal, 135° diagonal, and nondirectional) in the sub-images.
- Primitive length, edge frequency, and autocorrelation: These are well-known texture analysis methods, which use statistical rules to describe the spatial distribution and relation of gray values [65].
- Shape features: We use a collection of shape features provided by the standard MATLAB implementation (regionprops) [66], such as the area or elliptical shape features of local patterns.

C. Classification

To detect abnormal CXRs with TB, we use a support vector machine (SVM), which classifies the computed feature vectors into either normal or abnormal. An SVM in its original form is a supervised nonprobabilistic classifier that generates hyperplanes to separate samples from two different classes in a space with possibly infinite dimension [67], [68]. The unique characteristic of an SVM is that it does so by computing the hyperplane with the largest margin; i.e., the hyperplane with the largest distance to the nearest training data point of any class. Ideally, the feature vectors of abnormal CXRs will have a positive distance to the separating hyperplane, and feature vectors of normal CXRs will have a negative distance. The larger the distance the more confident we are in the class label. We therefore use these distances as confidence values to compute the ROC curves in Section V.

D. System Implementation

While implementation in the field is under the direction of AMPATH, and out of the control of the authors, the current system architecture and status of the project is as follows: AMPATH has finished mounting a portable X-ray machine on a light truck that has been modified to allow radiographic imaging. For example, the truck has been shielded against radiation and has been equipped with an on-board power generation unit and a desk for patient evaluation. The X-ray machine is connected to a portable workstation provided by the manufacturer that acts as a DICOM node, pushing images in a PACS framework. In the initial testing phase, our software runs on a second portable computer that is connected to the workstation of the X-ray machine. The communication module of our software, which we implemented in Java, listens to the DICOM workstation. This module

can automatically receive DICOM files and store them locally. It can also invoke the screening methods described in this paper and output the classification results (normal/abnormal) and their confidence values. Because we coded many of our algorithms, such as segmentation, feature extraction, and classification in MATLAB, we created Java wrappers for these functions and integrated them into our Java code. We also added a straightforward user interface that indicates whether a given X-ray is abnormal.

The truck will start its round trip from Moi University, and all X-rays will be processed on-board by our software. Depending on the availability of long-range wireless connections, the screening results will be transmitted on the fly to the truck's basis at Moi University or saved until the return of the truck.

V. RESULTS

This section presents a practical evaluation of our work. We show lung segmentation examples and we evaluate our features both in combination and individually. We also compare the performance of our proposed TB detection system with the performance of systems reported in the literature, including the performance of human experts.

A. Lung Segmentation Using Graph Cut

Fig. 6 shows three examples of our lung segmentation applied to CXRs from the MC dataset. The leftmost CXR has calcifications in the right upper lung and extensive irregular infiltrates in the left lung with a large area of cavitation. The CXR in the middle of Fig. 6 shows scars in the right upper lung, and the rightmost CXR has scars in the left upper lung and some infiltrates as well. Fig. 6 also shows the outlines of our segmentation masks for all three lungs. We can see that the segmentation masks capture the general shape of the lungs. Due to the use of a lung model, the infiltrates have not impaired the quality of the segmentations, especially in the leftmost CXR. We can see a slight leakage of the segmentation in the apical regions for the second and third CXR. The lower outlines toward the diaphragm could also be tighter in these images.

We compare our segmentation algorithm with the lung boundary detection algorithms in the literature. For the comparison, we use the graph cut implementation of our segmentation described in [5]. As performance measure, we used the overlap measure Ω

$$\Omega = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \qquad (11)$$

where TP is the correctly identified lung area (true positive), FP is the incorrectly identified lung area (false positive), and FN is the missed lung area (false negative). Table I shows the comparison results. We can see that our segmentation method (GC, [4], [5], with best parameter settings) performs reasonably well, though there are better segmentation methods reported in the literature. Our segmentation performance is 4.5% lower than the human performance reported for the JSRT set, which is 94.6%.

We have since significantly improved performance to achieve state-of-the-art results and these will be reported in a companion paper [72].
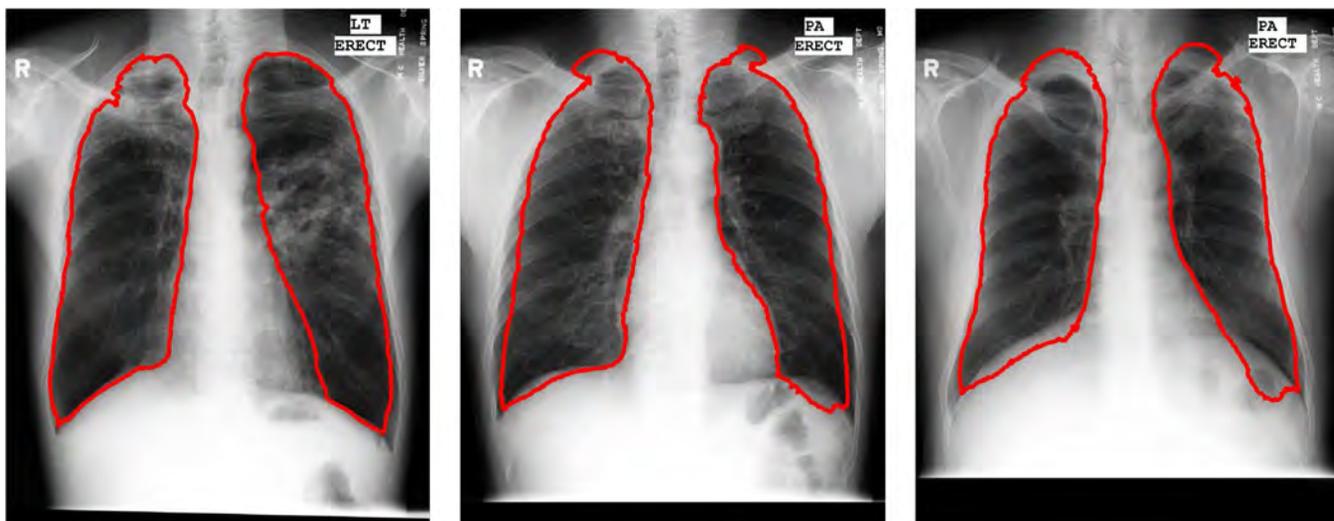
Fig. 6.   Example lung segmentations for MC CXRs. Note the over-segmentation in the apices. The CXR on the left-hand side has irregular infiltrates in the left lung. The CXR in the middle has small noncalcified nodules in the upper lobes. Grouped noncalcified nodules are visible in the CXR on the right-hand side. Note also that we do not include the posterior inferior lung region behind the diaphragm, similar to other lung segmentation methods in the literature.

TABLE I
OVERLAP SCORES ON JSRT DATASET COMPARED TO GOLD STANDARD
SEGMENTATION. GC: GRAPH CUT, PC: PIXEL CLASSIFICATION, MISCP:
MINIMAL INTENSITY AND SHAPE COST PATH, ASMOF: ACTIVE SHAPE
MODEL WITH OPTIMAL FEATURES, ASM: ACTIVE SHAPE MODEL,
AAM: ACTIVE APPEARANCE MODEL

Method	Avg $\pm$ std	Min	Median	Max
Hybrid Voting [22]	0.949 $\pm$ 0.020	0.818	0.953	0.978
Human Observer [22]	*0.946 $\pm$ 0.018*	*0.822*	*0.949*	*0.972*
PC postprocessed [22]	0.945 $\pm$ 0.022	0.823	0.951	0.972
Hybrid ASM-PC [22]	0.934 $\pm$ 0.037	0.706	0.945	0.968
Hybrid AAM-PC [22]	0.933 $\pm$ 0.026	0.762	0.939	0.966
MISCP [69]	0.930 $\pm$ 0.045	-	-	-
ASMOF [70]	0.927 $\pm$ 0.032	0.745	0.936	0.946
ASM-SIFT [71]	0.920 $\pm$ 0.031	0.783	0.928	0.961
ASM [22]	0.903 $\pm$ 0.057	0.601	0.924	0.960
GC [4], [5]	**0.901 $\pm$ 0.054**	**0.541**	**0.911**	**0.969**
ASM [71]	0.870 $\pm$ 0.074	0.608	0.892	0.954
AAM [22]	0.847 $\pm$ 0.095	0.017	0.874	0.956
Mean shape [22]	0.713 $\pm$ 0.075	0.460	0.713	0.891

Fig. 7.   Exhaustive evaluation of all possible feature subsets. Red curve plots ROC performance (AUC) and the black curve is the classifier accuracy.

B. Descriptor Evaluation for Feature Set A

We evaluate the performance of Feature Set A on the MC dataset. For each CXR in the MC dataset, we compute the descriptors in Feature Set A (see Section IV-B) and concatenate them into a single feature vector. We then apply a leave-one-out evaluation scheme, using the SVM-classifier described in Section IV-C. According to the leave-one-out scheme, we classify each feature vector (CXR) in the MC dataset with a classifier trained on the remaining feature vectors (CXRs) of the MC dataset. We thus train as many classifiers as there are CXRs in the MC dataset (138 altogether). To get a better understanding of the performance of individual descriptors and descriptor groups, we perform leave-one-out evaluations for all possible descriptor subsets. Fig. 7 shows the recognition rates we obtain. The x-axis of Fig. 7 represents the different descriptor combinations, where each possible descriptor subset is coded as a 6-digit binary index. Each bit indicates the membership of one of the descriptors mentioned above. The y-axis of Fig. 7 shows the area under the ROC curve (AUC) and the
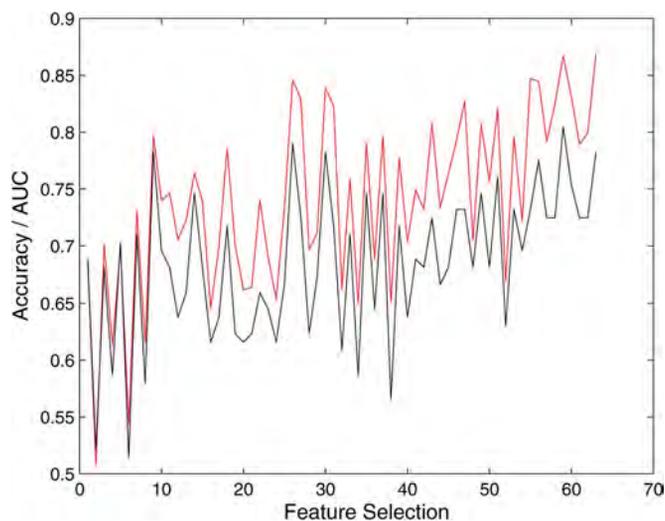
accuracies for each descriptor combination (ACC); see the red and black curve, respectively. To compute the accuracy in Fig. 7, we use the natural decision boundaries of the linear SVM classifier. We thus consider any pattern classified with positive confidence value as abnormal and any pattern with negative confidence as normal. Whenever we report classification results in the following, we will use this standard classification scheme. Accuracy and AUC are highly correlated, with the AUC being higher than the accuracy for most descriptor combinations. The jagged shape of both curves indicates that some descriptors are less likely to increase the performance when added to the descriptor mix. Nevertheless, we see that both the AUC and the accuracy tend to increase with larger descriptor sets. In fact, we achieve the highest AUC value of 86.9%, with an accuracy of 78.3%, when we add all descriptors to the descriptor set. Note that we have removed one feature from the set originally proposed in our earlier publication [6]. The set presented here
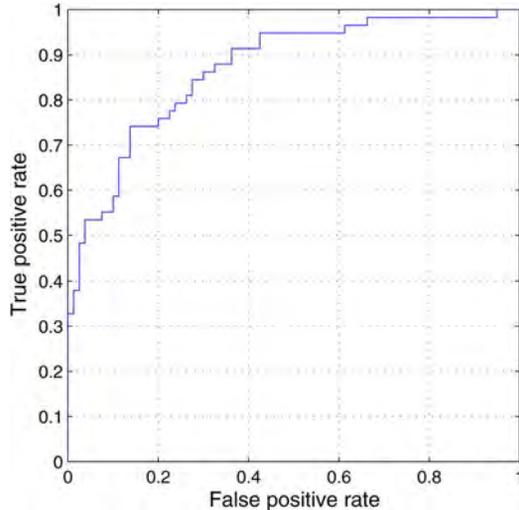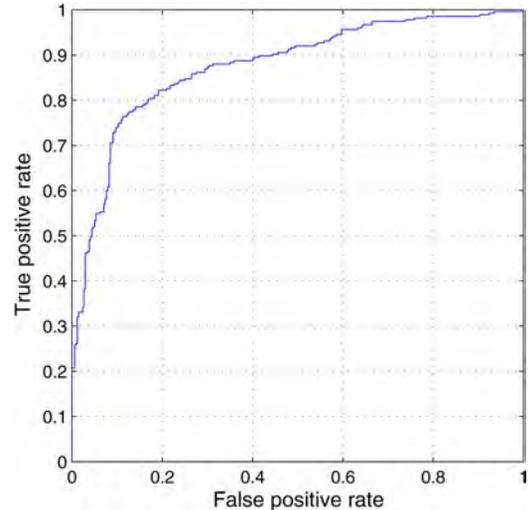
Fig. 8. ROC curve for MC data and Feature Set A.

Fig. 9. ROC curve for Shenzhen data and Feature Set A.

is now optimal in the sense that removing one feature reduces the performance.

C. Machine Performance

We present machine classification results for our two datasets, namely the Montgomery County (MC) dataset from our local TB clinic, USA, and the set from Shenzhen Hospital, China.

1) Montgomery County: Fig. 8 shows the ROC curve that we obtain when using all descriptors of Feature Set A. The ROC curve shows different possible operating points depending on the confidence threshold for the SVM classifier. The y-coordinate indicates the sensitivity (or recall) of our system, while the x-coordinate indicates the corresponding false positive rate, which is one minus the specificity. The area under the ROC curve (AUC) in Fig. 8 is 86.9%, with an overall classification accuracy of 78.3%. According to the ROC curve in Fig. 8, we achieve a sensitivity of about 95% when we accept a false positive rate that is slightly higher than 40%. This means that our specificity is a bit lower than 60% in this case.

2) Shenzhen Hospital: We repeated the same experiments on the set from Shenzhen Hospital. Fig. 9 shows the ROC curve that we computed for this set, again using the full Feature Set A. We see that the ROC curve is slightly better than the ROC curve for the data from our local TB clinic in Fig. 8. In fact, the AUC is approximately 88%. The classification accuracy is also slightly better. We computed an accuracy of about 82.5% for this set, which shows that we can provide consistent performance across different datasets, and for practically relevant data.

We also computed the performance results for our second feature set, Set B. Interestingly, with this feature set, we achieve a similar performance. Using a linear support vector machine, we obtain an accuracy of about 82% and an AUC of 88.5%. Thus, increasing the feature dimensionality does not lead to any improvement in performance.

For comparison purposes, we list the AUC values for our two X-ray sets and our two feature sets again in Table II, using our linear SVM. Table II also contains the result for the second feature set computed on the Montgomery X-ray set, which is lower.

TABLE II
CLASSIFICATION PERFORMANCE (AUC) ON MONTGOMERY COUNTY
AND SHENZHEN CXRS FOR FEATURE SET A AND FEATURE SET B

	Feature Set A	Feature Set B
Montgomery County	86.9	80.1
Shenzhen Hospital	88.0	88.5

TABLE III
CLASSIFICATION PERFORMANCE FOR DIFFERENT CLASSIFIER ARCHITECTURES
ON SHENZHEN CXR SET USING FEATURE SET B

	SVM(L)	SVM(PK)	SVM(RBF)	NN	ADT	LLR
ACC	82.1	76.4	76.4	80.7	82.6	84.1
AUC	88.5	86.0	85.5	88.0	85.6	90.0

For the Shenzhen data and Feature Set B, we experimented with different classification methods to see how the performance varies across different architectures. Table III shows the performance results, accuracy (ACC) and area under the ROC curve (AUC), for the following architectures: support vector machine (SVM) with linear (L), polynomial (PK), and radial basis function kernels (RBF), backpropagation neural network (NN), alternating decision tree (ADT), and linear logistic regression (LLR). The first column of Table III shows the performance for the linear support vector machine that we reported above. It is slightly higher than the rate for the polynomial and radial kernels; in particular the accuracy is higher for the linear machine. We experimented with different C-values for the support vector machines, and found that the standard value ($C = 1$) provides the best performance in our case, with only slight differences in general. Table III shows that the AUC is relatively stable across different architectures, with the linear logistic regression providing a slightly better overall performance.

3) Comparison With Other Systems in the Literature: In the literature, only a few papers have reported performance numbers for full-fledged TB screening systems, for example [25], [26], [31]. Many papers evaluate only part of the detection problem and concentrate on sub-problems, such as cavity detection [27], [28]. Judging by the ROC curves, the performance of our system is comparable to the performance of

some existing systems that address the problem in its entirety. For instance, the AUC value of our system is higher than the AUC values reported for the systems in [25] and [31]. Our AUC value is also slightly higher than the AUC value reported by Hogeweg et al., who use a combination of texture and shape abnormality detectors [26]. However, for a fair comparison of these systems, we would have to evaluate each system on the same dataset. Currently, the training sets in [25], [26], [31] are not publicly available. As yet, there is no publicly available CXR set of sufficient size that would allow training of a TB screening system. For the time being, we have to content ourselves with the fact that some existing systems provide reasonable performances across different datasets. We plan to make both our sets, the MC set as well as the Shenzhen set, available to the research community, so that other researchers can compare their performance. For a more detailed overview of existing TB screening systems, we refer readers to our recent survey in [32].

D. Comparison With Human Performance

In the following, we compare our system performance with human reading performance of two earlier studies reported in the literature. We also conducted our own independent observer study, asking two radiologists to read the MC CXR set (L. Folio, J. Siegelman).

1) Earlier Studies: Van't Hoog et al. investigated the performance of clinical officers in a tuberculosis prevalence survey [73]. Their study shows that clinical officers with sufficient training, rather than medical officers, can achieve an acceptable performance when screening CXRs for *any abnormality*. Van't Hoog et al. therefore recommend training of clinical officers for TB screening programs in regions where medical personnel with radiological expertise is rare. In their study, two experts achieve sensitivities of 81% and 83%, respectively, when screening CXRs of patients with bacteriologically confirmed TB for *any abnormality*. The corresponding specificities are 80% and 74%, respectively. On the other hand, three clinical officers achieve a sensitivity of 95% and a specificity of 73% for the same task. Note that each of the experts has a lower sensitivity than the group of clinical officers. We can compare these numbers with the ROC curve in Fig. 8, which shows the performance of our automatic system. We see that humans still perform better than our automatic system, and also better than other systems reported in the literature. Nevertheless, our system performs reasonably well and its performance, while inferior, is within reach of the clinical officers' performance in the study of Van't Hoog et al. For the same sensitivity provided by the clinical officers (95%), our system achieves a specificity that is about 15% lower than the specificity of the clinical officers.

In another recent study, Maduskar et al. showed that automatic chest radiograph reading for detection of TB has similar performance as clinical officers and certified readers [74]. They collected a dataset of 166 digital CXRs in Zambia, containing 99 positive and 67 negative cases confirmed by sputum cultures. In their observer study, four clinical officers and two certified readers scored all X-rays between zero and hundred. Maduskar et al. compared the human performance with the performance

TABLE IV
RADIOLOGIST AGREEMENT ON MONTGOMERY CXRs

		Radiologist B		
		+	-	
Radiologist A	+	69	15	84
	-	6	48	54
		75	63	138

TABLE V
COMPARISON OF HUMAN CONSENSUS PERFORMANCE WITH GROUND TRUTH OF MONTGOMERY CXRs

		Consensus		
		+	-	
Ground Truth	+	58	0	58
	-	25	55	80
		83	55	138

of their software, which uses the same score range. They computed the areas under the ROC curves and obtained values between 70% (clinical officers) and 72% (software), showing that there is no significant difference between human and machine performance. This result is in accordance with our own study, in which we compared the performance of our system with human performance (see below).

2) Our Study: In our study, we asked two radiologists to provide a second and third reading for our MC CXR set. Both radiologists work at a United States clinical center and hospital, respectively. We made them aware of the purpose of our screening project, that is, to detect tuberculosis in a population from an endemic region which is otherwise to be considered healthy. As a result the recommendations for evaluation of TB screening from the WHO Lime book were considered in the decision making process, in particular the use of *intentional overreading* [1]. To present the CXR data to the radiologists, we adapted our Firefly labeling tool, allowing the radiologists to see the CXRs online and store their readings in a database [75]. Table IV shows the agreement of both radiologists on the MC data. A plus sign indicates CXRs classified as TB positive (abnormal) by one of the radiologists, and the minus sign represents CXRs classified as normal. According to Table IV, both radiologists agree in 84.8% of the cases (95% CI:[77.7,90.3], using exact one-sample test of proportion). The corresponding kappa value is ($\kappa = 0.69$, 95% CI:[0.52,0.86]). This signifies moderate agreement.

After their individual readings, both radiologists convened to come to a consensus decision, reconciling readings for which they have disagreed. For the remaining discrepant cases, they agreed the findings were consistent with TB. Table V shows a comparison of their consensus decision with the labels from our local TB clinic. We consider the latter to be ground-truth labels because they are based on clinical data as well as patient data to which both radiologists had no access. In Table V, the number of false negatives is zero, which means the radiologists succeeded in detecting all TB positive cases. Therefore, the sensitivity (recall) is 100% (95% CI:[93.8,100]). The specificity is 68.8% (95% CI:[57.4,78.7]), so there is a considerable number of false positives, namely 25. Both radiologists agree in 81.9% of the cases with the ground-truth data (95% CI:[74.4,87.9]), which is a relatively low recognition rate due to overreading and trying not to miss any potential positive TB case.

TABLE VI
COMPARISON OF MACHINE PERFORMANCE WITH GROUND
TRUTH OF MONTGOMERY CXRS

		Machine +	Machine -	
Ground Truth	+	43	15	58
	-	15	65	80
		58	80	138

TABLE VII
PERFORMANCE COMPARISON BETWEEN MACHINE AND RADIOLOGIST
CONSENSUS FOR MONTGOMERY CXRS

		Consensus correct	Consensus incorrect	
Machine	correct	89	19	108
	incorrect	24	6	30
		113	25	138

In Table VI, we compare our machine output with the ground-truth data, using again the standard classification scheme that considers patterns classified with positive confidence as abnormal. Here, the agreement with the ground-truth data is slightly lower than in the previous expert consensus table. The machine agrees with the ground-truth data in 78.3% of the cases (95% CI:[70.4,84.8]). This is the same recognition rate we reported in Section V-C for the MC set. Note that the false positives and the false negatives are evenly distributed, with a sensitivity of 74.1% (95% CI:[61.0,84.7]) and a specificity of 81.3% (95% CI:[71.0,89.1]). This is because we have not optimized the true positive rate for our classifier. If we do so, we can see from the ROC curve in Fig. 8, that in order for the sensitivity to be close to 100%, our false positive rate would be slighter higher than 60%. This is about twice as high as the false positive rate for the radiologist's consensus decision in Table V, which is about 31%.

Finally, in Table VII, we compare the correct and incorrect classification results of the two radiologists and the machine. In terms of classification performance, the radiologists are not significantly better than the machine (McNemar test, $p = 0.54$). Note that the number of CXRs for which both the machine and the consensus are wrong is remarkably low. The combined human-machine performance with a significantly lower error rate of 4.3%, compared to the machine-only error rate of 21.7% and the human consensus error of 18.1%, suggests using our system for computer-aided diagnosis and offering a verifying second opinion of radiologist readings. This can help improve human performance because it is unlikely that both the radiologist and the machine classify the same CXR incorrectly.

VI. CONCLUSION

We have developed an automated system that screens CXRs for manifestations of TB. The system is currently set up for practical use in Kenya, where it will be part of a mobile system for TB screening in remote areas. When given a CXR as input, our system first segments the lung region using an optimization method based on graph cut. This method combines intensity information with personalized lung atlas models derived from the training set. We compute a set of shape, edge, and texture features as input to a binary classifier, which then classifies the given input image into either normal or abnormal.

In this paper, we compare two different established feature sets: one set typically used for object recognition and the other used in image retrieval applications. We also experiment with different classifier architectures. Both feature sets and most of the classifier architectures we tested, provide a similar performance. To improve the performance further, we could try to improve the lung segmentation, which provides average performance compared to other systems in the literature. One approach would be to find optimal weights for the terms in the graph cut energy function. Another possibility would be to use more atlas-based lung models for computing the average lung model (see our companion paper [72]). We could also try to partition the lung into different regions, as some of the existing CAD systems do. It is surprising that we achieve a relatively high performance compared to other approaches by using only global features. This may indicate that the combination of local features in the literature is still suboptimal. A final verdict on this issue can only be made once public benchmark data becomes available. Due to the lack of this data for TB detection in CXRs, it is currently difficult to do a fair comparison of the few existing systems that have been reported in the literature. We therefore plan to make both our datasets publicly available. One of these sets, the Shenzhen set, is from a high-incidence area. For both sets, we achieve an AUC of 87% and 90%, respectively. Furthermore, our performance, while still lower than human performance, is reasonably close to the performance of radiologists. In an independent observer study with two radiologists, who were trying not to miss any positive case, our false positive rate is twice as high according to the ROC curve. The likelihood that both the radiologists and the machine reach a wrong conclusion is very low. This shows that it should be possible to reach human performance in the future, or at least have a system that can assist radiologists and public health providers in the screening and decision process. These comparison results have encouraged us to test our system in the field under realistic conditions. In future experiments, we will evaluate our system on larger datasets that we will collect using our portable scanners in Kenya.

APPENDIX

The Montgomery County X-ray set as well as the Shenzhen Hospital X-ray set are available for research purposes upon review of request for data. To submit the request, please visit the following webpage:http://archive.nlm.nih.gov/. Under the "Repositories" tab, a link points to a page with more information on our chest images, including contact information.

ACKNOWLEDGMENT

REFERENCES

[1] World Health Org., Global tuberculosis report 2012.
[2] World Health Org., Global tuberculosis control 2011 2011.
[3] Stop TB Partnership, World Health Org. , The Global Plan to Stop TB 2011–2015 2011.

[4] S. Candemir, S. Jaeger, K. Palaniappan, S. Antani, and G. Thoma, "Graph-cut based automatic lung boundary detection in chest radiographs," in *Proc. IEEE Healthcare Technol. Conf.: Translat. Eng. Health Med.*, 2012, pp. 31–34.

[5] S. Candemir, K. Palaniappan, and Y. Akgul, "Multi-class regularization parameter learning for graph cut image segmentation," in *Proc. Int. Symp. Biomed. Imag.*, 2013, pp. 1473–1476.

[6] S. Jaeger, A. Karargyris, S. Antani, and G. Thoma, "Detecting tuberculosis in radiographs using combined lung masks," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 4978–4981.

[7] C. Leung, "Reexamining the role of radiography in tuberculosis case finding," *Int. J. Tuberculosis Lung Disease*, vol. 15, no. 10, pp. 1279–1279, 2011.

[8] L. R. Folio, *Chest Imaging: An Algorithmic Approach to Learning*. New York: Springer, 2012.

[9] S. Jaeger, S. Antani, and G. Thoma, "Tuberculosis screening of chest radiographs," in *SPIE Newsroom*, 2011.

[10] C. Daley, M. Gotway, and R. Jasmer, "Radiographic manifestations of tuberculosis," in *A Primer for Clinicians*. San Francisco, CA: Curry International Tuberculosis Center, 2009.

[11] J. Burrill, C. Williams, G. Bain, G. Conder, A. Hine, and R. Misra, "Tuberculosis: A radiologic review," *Radiographics*, vol. 27, no. 5, pp. 1255–1273, 2007.

[12] R. Gie, *Diagnostic Atlas of Intrathoracic Tuberculosis in Children*. : International Union Against Tuberculosis and Lung Disease (IUATLD), 2003.

[13] A. Leung, "Pulmonary tuberculosis: The essentials," *Radiology*, vol. 210, no. 2, pp. 307–322, 1999.

[14] B. van Ginneken, L. Hogeweg, and M. Prokop, "Computer-aided diagnosis in chest radiography: Beyond nodules," *Eur. J. Radiol.*, vol. 72, no. 2, pp. 226–230, 2009.

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

[16] 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.

[17] S. Sakai, H. Soeda, N. Takahashi, T. Okafuji, T. Yoshitake, H. Yabuuchi, I. Yoshino, K. Yamamoto, H. Honda, and K. Doi, "Computer-aided nodule detection on digital chest radiography: Validation test on consecutive T1 cases of resectable lung cancer," *J. Digit. Imag.*, vol. 19, no. 4, pp. 376–382, 2006.

[18] J. Shiraishi, H. Abe, F. Li, R. Engelmann, H. MacMahon, and K. Doi, "Computer-aided diagnosis for the detection and classification of lung cancers on chest radiographs: ROC analysis of radiologists' performance," *Acad. Radiol.*, vol. 13, no. 8, pp. 995–1003, 2006.

[19] S. Kakeda, J. Moriya, H. Sato, T. Aoki, H. Watanabe, H. Nakata, N. Oda, S. Katsuragawa, K. Yamamoto, and K. Doi, "Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system," *Am. J. Roentgenol.*, vol. 182, no. 2, pp. 505–510, 2004.

[20] K. Doi, "Current status and future potential of computer-aided diagnosis in medical imaging," *Br. J. Radiol.*, vol. 78, no. 1, pp. 3–19, 2005.

[21] B. Van Ginneken, B. ter Haar Romeny, and M. Viergever, "Computer-aided diagnosis in chest radiography: A survey," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1228–1241, Dec. 2001.

[22] 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," *Med. Image Anal.*, vol. 10, no. 1, pp. 19–40, 2006.

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

[24] A. Dawoud, "Fusing shape information in lung segmentation in chest radiographs," *Image Anal. Recognit.*, pp. 70–78, 2010.

[25] 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," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 139–149, Feb. 2002.

[26] 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," in *Proc. MICCAI*, 2010, pp. 650–657.

[27] R. Shen, I. Cheng, and A. Basu, "A hybrid knowledge-guided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 11, pp. 2646–2656, Nov. 2010.

[28] T. Xu, I. Cheng, and M. Mandal, "Automated cavity detection of infectious pulmonary tuberculosis in chest radiographs," in *Proc. Int. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 5178–5181.

[29] L. Hogeweg, C. I. Sánchez, P. A. de Jong, P. Maduskar, and B. van Ginneken, "Clavicle segmentation in chest radiographs," *Med. Image Anal.*, vol. 16, no. 8, pp. 1490–1502, 2012.

[30] 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.

[31] 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 Recognit.*, vol. 42, no. 9, pp. 1768–1776, 2009.

[32] S. Jaeger, A. Karargyris, S. Candemir, J. Siegelman, L. Folio, S. Antani, and G. Thoma, "Automatic screening for tuberculosis in chest radiographs: A survey," *Quant. Imag. Med. Surg.*, vol. 3, no. 2, pp. 89–99, 2013.

[33] C. Pangilinan, A. Divekar, G. Coetzee, D. Clark, B. Fourie, F. Lure, and S. Kennedy, "Application of stepwise binary decision classification for reduction of false positives in tuberculosis detection from smeared slides," presented at the Int. Conf. Imag. Signal Process. Healthcare Technol., Washington, DC, 2011.

[34] C. Boehme, P. Nabeta, D. Hillemann, M. Nicol, S. Shenai, F. Krapp, J. Allen, R. Tahirli, R. Blakemore, and R. Rustomjee *et al.*, "Rapid molecular detection of tuberculosis and rifampin resistance," *New Eng. J. Med.*, vol. 363, no. 11, pp. 1005–1015, 2010.

[35] 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," *Am. J. Roentgenol.*, vol. 174, no. 1, pp. 71–74, 2000.

[36] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient n-d image segmentation," *Int. J. Comput. Vis.*, vol. 70, no. 1, pp. 109–131, 2006.

[37] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[38] S. Jaeger, C. Casas-Delucchi, M. Cardoso, and K. Palaniappan, "Dual channel colocalization for cell cycle analysis using 3D confocal microscopy," in *Proc. Int. Conf. Pattern Recognit.*, 2010, pp. 2580–2583.

[39] 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 *Proc. Indian Conf. Comput. Vis., Graph., Image Process.*, 2010, pp. 412–418.

[40] K. Palaniappan, F. Bunyak, P. Kumar, I. Ersoy, S. Jaeger, K. Ganguli, A. Haridas, J. Fraser, R. Rao, and G. Seetharaman, "Efficient feature extraction and likelihood fusion for vehicle tracking in low frame rate airborne video," in *Proc. Int. Conf. Inf. Fusion*, 2010, pp. 1–8.

[41] M. Linguraru, S. Wang, F. Shah, R. Gautam, J. Peterson, W. Linehan, and R. Summers, "Computer-aided renal cancer quantification and classification from contrast-enhanced CT via histograms of curvature-related features," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2009, pp. 6679–6682.

[42] R. Pelapur, S. Candemir, F. Bunyak, M. Poostchi, G. Seetharaman, and K. Palaniappan, "Persistent target tracking using likelihood fusion in wide-area and full motion video sequences," in *Proc. Int. Conf. Inf. Fusion*, 2012, pp. 2420–2427.

[43] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. Int. Conf. Comp. Vis. Patt. Recognit.*, 2005, vol. 1, pp. 886–893.

[44] L. Chen, R. Feris, Y. Zhai, L. Brown, and A. Hampapur, "An integrated system for moving object classification in surveillance videos," in *Proc. Int. Conf. Adv. Video Signal Based Surveill.*, 2008, pp. 52–59.

[45] F. Han, Y. Shan, R. Cekander, H. Sawhney, and R. Kumar, "A two-stage approach to people and vehicle detection with HOG-based SVM," in *Performance Metrics Intell. Syst. Workshop*, Gaithersburg, MD, 2006, pp. 133–140.

[46] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *Proc. Int. Conf. Comput. Vis.*, 2009, pp. 32–39.

[47] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

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

[49] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.

[50] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–761, Apr. 2012.

[51] A. Hafiane, G. Seetharaman, K. Palaniappan, and B. Zavidovique, "Rotationally invariant hashing of median binary patterns for texture classification," in *Proc. Int. Conf. Image Anal. Recognit.*, 2008, pp. 619–629.

[52] A. Frangi, W. Niessen, K. Vincken, and M. Viergever, "Multiscale vessel enhancement filtering," in *Proc. MICCAI*, 1998, pp. 130–137.

[53] F. Bunyak, K. Palaniappan, O. Glinskii, V. Glinskii, V. Glinsky, and V. Huxley, "Epifluorescence-based quantitative microvasculature remodeling using geodesic level-sets and shape-based evolution," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2008, pp. 3134–3137.

[54] M. Simpson, D. You, M. Rahman, D. Demner-Fushman, S. Antani, and G. Thoma, "ITI's participation in the ImageCLEF 2012 medical retrieval and classification tasks," in *CLEF 2012 Working Notes*, 2012.

[55] C.-R. Shyu, M. Klaric, G. Scott, A. Barb, C. Davis, and K. Palaniappan, "GeoIRIS: Geospatial information retrieval and indexing system—Content mining, semantics modeling, complex queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 839–852, Apr. 2007.

[56] M. Lux and S. Chatzichristofis, "LIRE: Lucene Image Retrieval: An extensible Java CBIR library," in *Proc. ACM Int. Conf. Multimedia*, 2008, pp. 1085–1088.

[57] M. Lux, "Content-based image retrieval with LIRE," in *Proc. ACM Int. Conf. Multimedia*, 2011, pp. 735–738.

[58] LIRE: An open source library for content-based image retrieval 2013 [Online]. Available: https://code.google.com/p/lire/

[59] P. Howarth and S. Rüger, "Robust texture features for still-image retrieval," *IEE Proc. Vis., Image Signal Process.*, vol. 152, no. 6, pp. 868–874, 2005.

[60] S. Chatzichristofis and Y. Boutalis, "CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval," *Comput. Vis. Syst.*, pp. 312–322, 2008.

[61] S. Chatzichristofis and Y. Boutalis, "FCTH: Fuzzy color and texture histogram—A low level feature for accurate image retrieval," in *Proc. Int. Workshop Image Anal. Multimedia Interactive Services*, 2008, pp. 191–196.

[62] M. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[63] DISCOVIR: Distributed content-based visual information retrieval system 2013 [Online]. Available: http://appsrv.cse.cuhk.edu.hk/ ~miplab/discovir/

[64] M. Lux, "Caliph & Emir: MPEG-7 photo annotation and retrieval," in *Proc. Int. Conf. Multimedia*, 2009, pp. 925–926.

[65] G. Srinivasan and G. Shobha, "Statistical texture analysis," *Proc. World Acad. Sci., Eng. Technol.*, vol. 36, pp. 1264–1269, 2008.

[66] MATLAB. ver. 7.4.0.287 (R2007a), MathWorks, Natick, MA, 2007.

[67] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer Verlag, 2000.

[68] B. Schölkopf, C. Burges, and A. Smola, *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA: MIT Press, 1999.

[69] D. Seghers, D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens, "Minimal shape and intensity cost path segmentation," *IEEE Trans. Med. Imag.*, vol. 26, no. 8, pp. 1115–1129, Aug. 2007.

[70] B. Ginneken, A. Frangi, J. Staal, B. Romeny, and M. Viergever, "Active shape model segmentation with optimal features," *IEEE Trans. Med. Imag.*, vol. 21, no. 8, pp. 924–933, Aug. 2002.

[71] Y. Shi, F. Qi, Z. Xue, L. Chen, K. Ito, H. Matsuo, and D. Shen, "Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics," *IEEE Trans. Med. Imag.*, vol. 27, no. 4, pp. 481–494, Apr. 2008.

[72] S. Candemir, S. Jaeger, K. Palaniappan, J. Musco, R. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with non-rigid registration," *IEEE Trans. Med. Imag.*, to be published.

[73] 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," *Int. J. Tuberculosis Lung Disease*, vol. 15, no. 10, pp. 1308–1314, 2011.

[74] P. Maduskar, L. Hogeweg, H. Ayles, and B. van Ginneken, "Performance evaluation of automatic chest radiograph reading for detection of tuberculosis (TB): A comparative study with clinical officers and certified readers on TB suspects in sub-Saharan Africa," in *Eur. Congr. Radiol.*, 2013.

[75] D. Beard, "Firefly—Web-based interactive tool for the visualization and validation of image processing algorithms," M.S. thesis, Univ. Missouri, Columbia, 2009.