**A Report to the Board of Scientific Counselors
April 2015**

Automated Detection of Lung Diseases in Chest X-Rays

Sameer Antani, PhD
Communications Engineering Branch

Sema Candemir, PhD, Postdoctoral Fellow
Stefan Jaeger, PhD, Research Fellow
Alexandros Karargyris, PhD, Research Fellow
Santosh KC, PhD, Postdoctoral Fellow
Szilárd Vajda, PhD, Postdoctoral Fellow
Zhiyun Xue, PhD, Staff Research Engineer, Medical Science & Computing, Inc.

Advisors:
George Thoma, PhD, Chief, CEB/LHC/NLM
Clement McDonald, MD, Director, LHC/NLM

Collaborating Radiologists:
Marc Kohli, MD, Indiana University School of Medicine
Les Folio, DO, MPH, NIH, Clinical Center
Jenifer Siegelman, MD, Brigham and Women's Hospital

TABLE OF CONTENTS

1   Introduction and Background

Tuberculosis (TB) is an infectious disease caused by the bacillus *M. Tuberculosis*. According to the World Health Organization (WHO) *Global Tuberculosis Report 2014* [1], TB remains one of the world's deadliest communicable diseases. In 2013, an estimated 9.0 million people developed TB, of these 1.1 million (13%) were HIV-positive. According to the report, 1.5 million died from the disease, 360 000 of whom were HIV-positive. The rates of TB mortality are slowly declining each year and it is estimated that 37 million lives were saved between 2000 and 2013 through effective diagnosis and treatment. However, TB is now present in all regions of the world, increasingly as *drug resistant* variants.
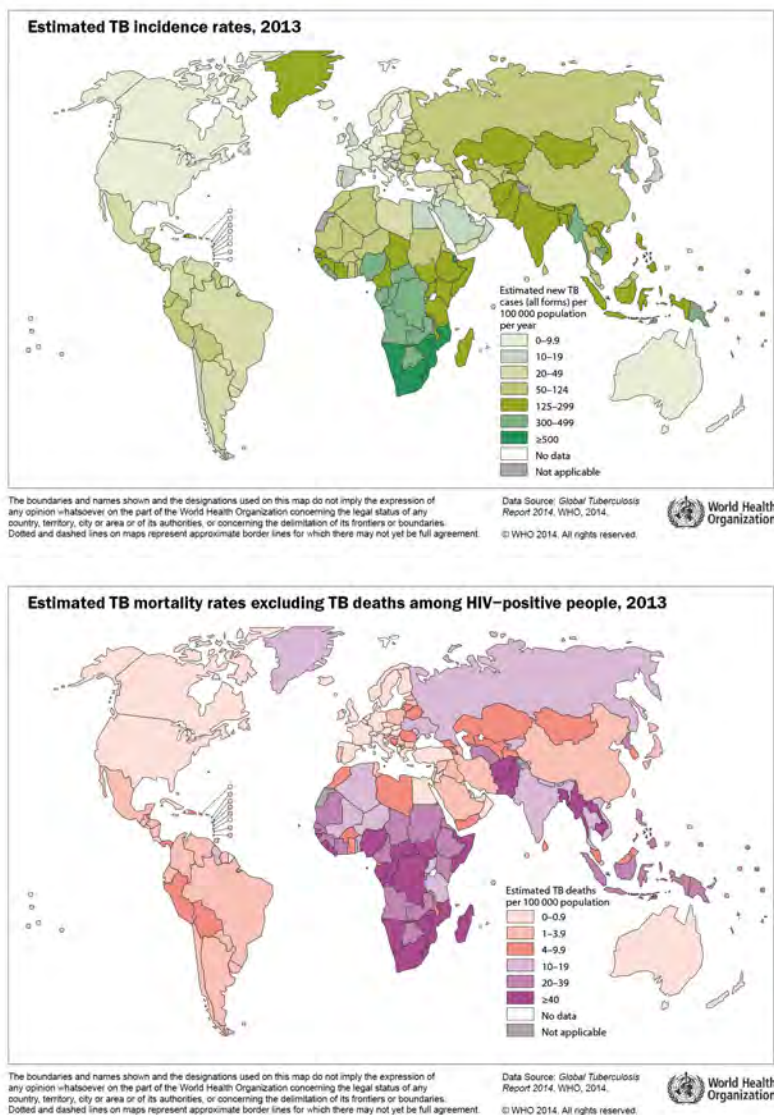
Figure 1. WHO maps (c. 2013) showing (Top) global TB incidence rates; (Bottom) TB mortality among HIV positive cases.

TB commonly affects the lung, but can also affect other organs such as bones and other soft tissues. It spreads through the air when people with active TB cough, sneeze, or otherwise expel infectious bacteria. According to the WHO, the largest impact of the increase in TB burden relates to inadequate control in developing nations, particularly those with high endemic rates of HIV infection where opportunistic

coinfections in immunocompromised HIV/AIDS patients have exacerbated the problem [2, 3]. TB is most prevalent in sub-Saharan Africa and Southeast Asia, where, in addition to HIV, widespread poverty and malnutrition reduce resistance to the disease. The African Region accounts for about four out of every five HIV-positive TB cases and TB deaths among people who were HIV positive [1].

Several skin tests based on immune response are available for determining whether an individual has been exposed to TB. However, skin tests are not an indicator of active disease, and are also affected by vaccinations. A definitive test for diagnosing TB is the identification of the bacteria in a clinical sputum or pus sample, which is the current gold standard [1, 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. Recent advances in automated diagnostic tests are based on DNA analysis[1]. The test is fast and accurate, but is expensive making it unsuitable for population screening. As a result, taking postero-anterior (PA) chest X-rays (CXRs) continues to remain as an inexpensive and mandatory part of every evaluation for TB [4, 5].

In this report, we present a reliable automated CXR image-based screening system for detecting pulmonary diseases, with a special initial focus on TB [6, 7]. The system is being tested through our collaboration with AMPATH (Academic Model for Prevention and Treatment of HIV/AIDS) [7]. AMPATH is a partnership between Moi University School of Medicine and Moi Teaching and Referral Hospital, Eldoret, 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. AMPATH currently treats nearly 200,000 patients for HIV. Large numbers of HIV patients need to be X-rayed and screened for active TB to ensure proper treatment of their infection(s) and avoid drug incompatibilities. However, the shortage of radiological services in Kenya necessitates an automated and inexpensive screening system. The system should be automatic and provide adequate clinical decision support to regional medical officers with little radiology background. The target imaging source for our automated system are portable X-ray scanners mounted on trucks designed for radiological imaging to access population in rural areas with poor road access. At-risk individuals identified by our system would be referred for treatment to regional clinics, or the hospital in Eldoret.

1.1   Radiographic impressions of TB

Patients exposed to TB can be characterized clinically into two groups: (1) those with active TB disease; and, (2) those who have been exposed but do not have active disease. This latter group may harbor small colonies of active bacilli sheltered within their bodies, particularly in their lungs, and are thus at risk for activation of TB, if their physical health state is compromised. This happens, for example, when immune surveillance and control systems fail in populations with immunocompromised states such as HIV/AIDS.

Major radiographic manifestations of active pulmonary TB [8] include the following, some of which are shown in Figure 2:
  • *Air space consolidation*: lobar opacity, often reported as pneumonia or pneumonitis;
  • *Miliary pattern:* Fine granular sandy or seed-like appearance throughout the entirety of both lungs, reported as diffuse bilateral infiltrates, sometimes (correctly) referred to as micronodular [9];
  • *Cavity formation:* a finding with a detectible radio-dense rim, which may be continuous or discontinuous, is differentiated from a mass as it has some central complete or relative radiolucency;
  • *Bronchiectasis* or enlargement of airways can appear as tubular rings or cylinders of irregular diameter extending radially from the lung hila, with or without central radiolucency;

---

[1] http://www.who.int/mediacentre/news/releases/2010/tb_test_20101208/en/

- *Lymph node enlargement* or *adenopathy:* smoothly marginated well-defined mass in close anatomic proximity to the mediastinum. These are silhouetted against the margins of normal structures such as the predictable pulmonary arteries and veins, which may facilitate their detection. Also a specific feature of an enlarged lymph node is frequently seen in patients with TB [10];
- *Thickened Pleura:* a finding that can be easily detected through careful survey of the periphery of the lungs just adjacent to ribs superiorly and laterally, and along the diaphragm inferiorly;
- *Pleural effusions:* loss of visualization (indistinctness) of the lateral costophrenic and medial cardiophrenic sulci on PA radiographs.
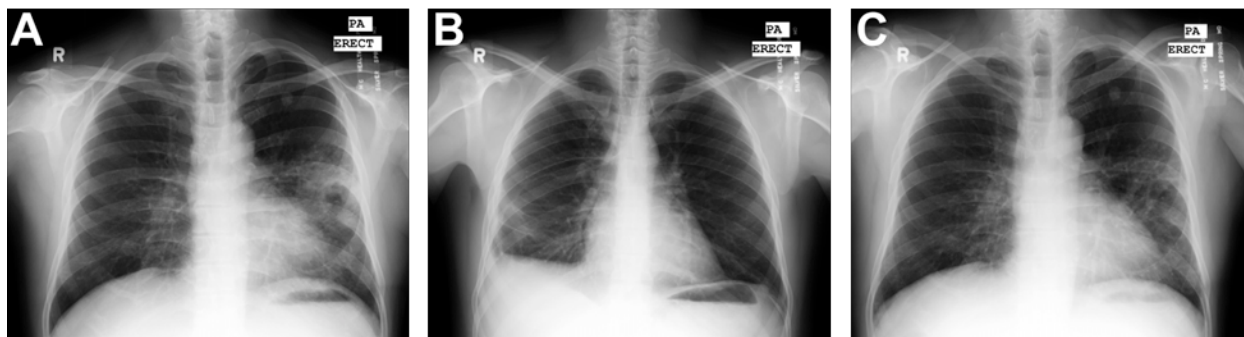

Figure 2. Example CXRs with manifestations of TB. CXR A and C in have infiltrates in both lungs. CXR B is a good example of pleural TB, which is indicated by the abnormal shape of the costophrenic angle of the right lung.

In addition to these findings of active disease, classic features of TB exposure include calcified granulomata and apical pleural thickening. Many of these characteristic findings occur in TB, but also occur in many other pulmonary diseases such as pneumonia.

1.2   Radiographic screening

According to the "WHO Lime book" [2], CXRs can serve both as a screening and measurement tool to identify everything that is or could be related to a patient having pulmonary diseases including manifestations of TB. In a mobile screening program all participants who either have CXRs "suspicious for TB" or whose CXRs are "minimally abnormal" should be "*intentionally over-read*" by the interpreter [2]. An embedded computerized program would allow this to occur during the primary encounter, limiting the resources required to identify the patient whose CXR has been screened and interpreted as normal. The potential for immediate, on-the-spot interpretation of "normal", which is entirely negative for TB or any pulmonary disease possibly related to TB; or "abnormal", which is positive for TB or any other disease that could, even in the rarest of circumstances, have a statistical association with TB, is of potential value in the mobile screening program. TB screening in high prevalence HIV/AIDS regions demands *intentional over-reading*. The relative frequency of typical TB-associated imaging features is different in patients with HIV. In such screening programs, each patient has a reasonably high likelihood of having disease, or pre-test probability of having disease. Pre-test probability is considered as an input variable by human readers, and may bias their interpretation of images toward "intentionally over-reading". In the screening context, a low false negative rate is desirable and a modestly high false positive rate is tolerated.

1.3   Related Work

Despite its near omnipresence in medical practice, digital chest radiography and computer aided screening and diagnosis is still a challenge. 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 [11-13]. Automated nodule detection is becoming one of the more mature applications

of decision support/automation for CXR and CT [14-16]. The result is that computer-aided diagnosis (CAD) systems can successfully assist radiologists in diagnosing lung cancer [17].

The segmentation of the lung field is an essential task that any 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 [18, 19]. For example, van Ginneken et al compared various techniques for lung segmentation, including active shapes, rule-based methods, pixel classification, and various combinations thereof [19, 20]. Their conclusion was that pixel level 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 3) [21]. Depending on the lung segmentation, different feature types and ways to aggregate them have been reported in the literature [18, 19].

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 reported, such as [22–25]. Further, TB screening is not just limited to normal/abnormal classification, but also to spot the location in CXRs where the abnormality lies [20]. For example, van Ginneken et al subdivide the lung into overlapping regions of various sizes and extract features from each region [22]. Hogeweg et al combined a texture-based abnormality detection system with a clavicle detection stage to suppress false positive responses [23]. A cavity in the upper lung zones is a strong indicator that TB has developed into a highly infectious state [24]. Shen et al developed a hybrid knowledge-based Bayesian approach to detect cavities in these regions automatically [24]. Xu et al approached the same problem with a model-based template matching technique, with image enhancement based on the Hessian matrix [25]. In [26], Hogeweg et al use 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 could significantly increase a radiologist's performance for nodule detection [27]. 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 [28]. 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 [29].

2 Project Objectives

This project exploits the convergence of imaging research and system development at the NLM and NIH policy objectives in global health. The following are project objectives:

1. Advance the state-of-the-art in automated CXR image analysis. Automatically detect presence of pulmonary diseases including TB and other relevant disease in digital CXRs, leading to suitable discrimination for screening, as well as compute a measure of confidence in its determination.
2. Develop deployable screening software such that it can aid field clinical officers in decision making at the point of care, and for radiologists to organize their workload.
3. Recognizing the severity of HIV/TB co-infections and the shortage of radiological services in western Kenya, deploy developed software on to a self-powered mobile X-ray truck that AMPATH uses in rural areas. Their staff take chest x-rays of the population and employ the NLM-developed software to screen for the presence of TB and other diseases.

3   Project Significance

The project makes a significant impact in the following ways:

- Advancing the NIH Director's global health initiative by making a significant impact in the development and application of low cost disease detection technologies in resource-challenged regions.
- Developing screening technology for TB, a major global health challenge identified by the WHO as the second leading cause of mortality from infectious disease. HIV and TB coinfections result in treatment complications and spread of the disease.
- Advancing the science in image analysis for automatically detecting pulmonary diseases from digital CXR images.

4   Datasets

Algorithmic development requires large image datasets for training and testing. We reached out to several sources – and the sets acquired are described below. The datasets were de-identified by the data providers and were exempted from IRB review at their institutions. The NIH Office of Human Research Protections Programs reviewed the collection of image data (review No. 5357).

JSRT set: Set compiled by the Japanese Society of Radiological Technology (JSRT) [30]. The set contains 247 CXRs, among which 154 have lung nodules (100 malignant cases, 54 benign cases), and 93 have no nodules. All images have a size of 2048 pixels x 2048 pixels with a 12-bit gray-scale depth. The pixel spacing in vertical and horizontal directions is 0.175mm. The JSRT set is publicly available and has gold standard lung masks [31] for performance evaluation.

Montgomery set: Set from the Department of Health and Human Services, Montgomery County, Maryland, USA. The set contains CXRs collected under Montgomery County`s Tuberculosis control program. The dataset consists of 138 images, 80 of which are normal, and 58 images are abnormal with manifestations of tuberculosis. The CXR images are captured with a Eureka stationary X-ray machine in 12-bit gray-scale. The images in the dataset have a resolution of 4020 pixels x 4892 pixels. The pixel spacing in vertical and horizontal directions is 0.0875mm.

Shenzhen set: The Shenzhen dataset was collected in collaboration with Shenzhen No.3 People`s Hospital, Guangdong Medical College, Shenzhen, China. The CXR images are from outpatient clinics, and were acquired as part of the daily hospital routine, using a Philips DR digital diagnostic system. The set contains 662 images, of which 326 are normal cases and 336 are cases with manifestations of TB. This set includes some pediatric images. As a result, the image sizes in the dataset vary: image width: minimum is 1130 pixels and maximum is 3001 pixels; for the image height, the minimum is 948 pixels and the maximum is 3001 pixels.

Indiana set: The Indiana Set comprises 3,996 de-identified narrative radiology reports and associated CXR images in posterior –anterior (PA) and lateral exposures. The images were acquired from two large hospital systems within the Indiana Network for Patient Care database. The chest x-ray reports that were pulled mention tuberculosis evaluation ordered from any setting, and then a random subset of all outpatient chest x-ray reports were selected to achieve the target dataset. There is only one study per patient. The set has images at various sizes that range from a minimum dimension of 1024 pixels and a maximum dimension of 4248 pixels.

India set: Contains 397 chest X-rays from a private clinic in India. These have a 12-bit gray-scale depth. The images have following different resolutions: 2446 pixels x 2010 pixels, 1772 pixels x 1430 pixels and 2010 pixels x 1572 pixels. The pixel spacing in vertical and horizontal directions is 0.175mm. It contains more challenging X-ray images due to rotated torsi scans, stronger edges within the rib-cage, denser shade effects at apices, wide variations of lung sizes, and large lung deformations.

We have made the Montgomery set and Shenzhen Chest X-ray datasets and associated data publicly available[2] for research purposes [32].

5   Methods and Procedures

The system architecture is designed as a set of cascaded modules, with the flexibility to implement alternate image analysis pathways followed by late stage decision fusion. As currently developed, every image is analyzed for automatic lung region localization. Image features are extracted from within the localized lung boundary, leading to 2-class normal/abnormal decision for the input CXR image. We are also studying alternate techniques for detecting abnormalities in the CXR without localizing the exact lung boundary. The method also uses edge detection to find spurious contours that could be indicative of disease. Initial results suggest that the method is fast and quite powerful in detecting certain kinds of pathologies.

5.1   Lung Region Localization

A key first step in a computer-aided CXR screening system is the automatic detection of the lung regions. We have developed a novel automated lung boundary detection algorithm for CXR images [6]. It consists of three stages. An overview of the method is shown in Figure 3 below.
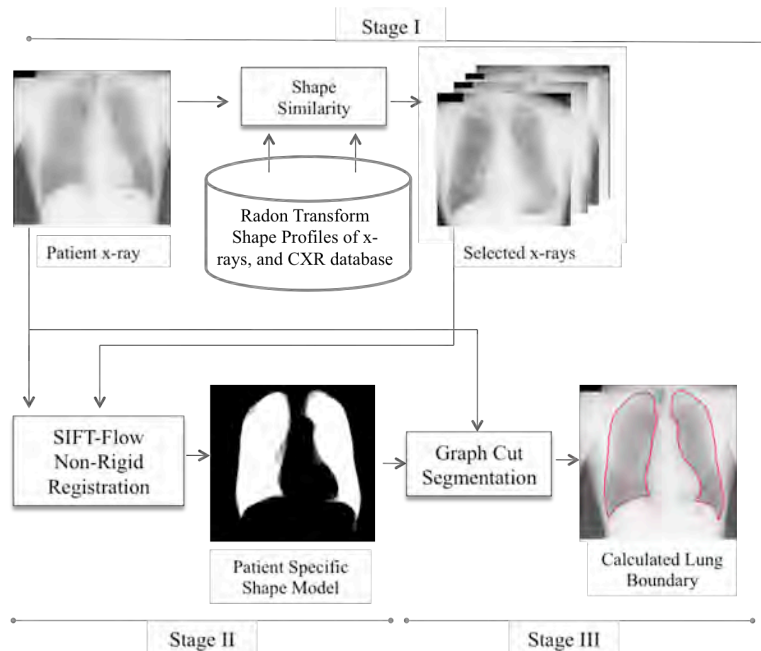

Figure 3. An overview of our lung segmentation algorithm: (Stage-I) finding the similar lung CXRs from an atlas; (Stage-II) warping selected images to patient CXR; and, (Stage-III) lung boundary detection using a graph-cuts optimization approach.

Stage 1: First we use a content-based image retrieval (CBIR) method to identify a small set of similar appearing CXR images from an expert-annotated set, that we shall call the "atlas set" hereafter. Horizontal and vertical projection profiles [33] are computed for all CXR images in the atlas set. Then, we measure the similarity of each projection profile between the atlas set and the patient chest X-ray using the average Bhattacharyya coefficient [34].

Stage 2: In order to create the lung model, we register the selected set of CXR images that have similar appearance but may have different lung outlines. The transformation mapping is done using the SIFT-flow algorithm [35]. The algorithm first models the local gradient information of the observed image using scale invariant feature transforms (SIFT) [36]. Next, a minimization algorithm calculates the SIFT-flow, the transformation mapping between each selected atlas image and the patient image. The mapping is used to register and warp the selected atlas CXRs, making them geometrically aligned to the patient image. The lung model for the patient X-ray is then composed as the mean of the warped lung masks from the registered atlas images. The model is a probabilistic shape prior in which each pixel value is the probability of the pixel being part of the lung field in the patient image.

Stage 3: As a refining stage, we perform image segmentation using the graph cut algorithm [37] and model the segmentation process with an objective function. The max-flow/min-cut algorithm [38] minimizes the objective function to find a global minimum that corresponds to foreground (within-lung) and background (outside-lung) labeling of the pixels.
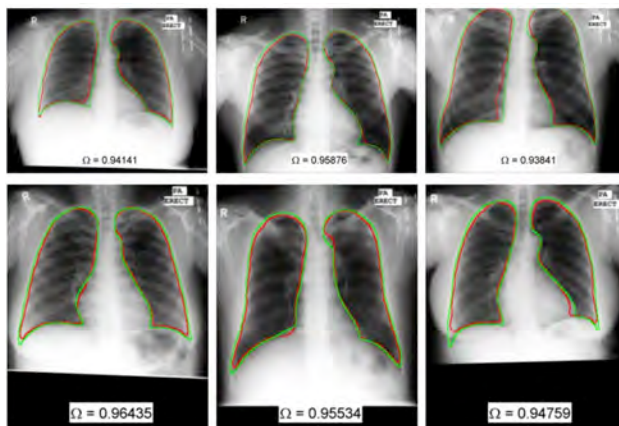

Figure 4. Sample segmentation results on Montgomery dataset. Green and red contours indicate the gold standard and automatic segmentation results, respectively.

Evaluation: A radiologist manually generated "gold standard" segmentations for the atlas chest X-ray images. The process was aided using an interactive boundary marking tool [39], developed in prior NLM research and reported to the Board. The radiologist then corrected these outlines using FireFly [40], a web-based labeling tool developed at the University of Missouri. The method was evaluated on three datasets (JSRT, Montgomery, and India) described, previously, in Section 3 above. We also compared the system performance with the systems in the literature. The Jaccard Index[3] (which measures overlap agreement) resulted in average accuracy of 95.4% on the public JSRT database, which bests all prior published results. A similar degree of accuracy of 94.1% and 91.7% on Montgomery and India datasets, respectively, demonstrates its robustness to image variety.

5.2   Image Features

To describe normal and abnormal patterns in the segmented lung field, we experimented with three different feature sets identified in the research literature for object detection and image retrieval; based on the research experience of the team, we expected these features to be meaningful for detecting pulmonary anomalies.

1) Object Detection Inspired Features—Set A: This is a versatile feature set that is a combination of shape, edge, and texture descriptors [7]. For each descriptor, a histogram is computed within the localized lung field. The CXR feature vector is concatenated across all computed feature histograms. In particular,

---

[3] Jaccard Index (Wikipedia): http://en.wikipedia.org/wiki/Jaccard_index

we use the following shape and texture descriptors in Set A [41, 42]: Intensity Histograms (IH), Gradient Magnitude histograms (GM), Shape Descriptor histogram (SD), Curvature Descriptor histograms (CD), Histogram of Oriented Gradients (HOG) [43], and the Local Binary Patterns (LBP) texture descriptor [44, 46]. A modified multiscale approach by Frangi et al [46, 47] is used to compute the eigenvalues of the Hessian matrix needed for the shape and curvature descriptors. The Hessian describes the second-order surface curvature properties of the local image intensity surface. Normalization with respect to intensity makes these descriptors independent of image brightness. We empirically determined that using 32 bins for each histogram provides good results.

2) CBIR-Based Image Features—Set B: This feature set contains low-level features motivated by content-based image retrieval (CBIR) research at NLM [48]. The feature vector has 594 dimensions and includes intensity, edge, texture, and shape moment features. The set of features include the following features from the Lucene image retrieval library[4] (LIRE) [49]:   three features – contrast, directionality, and coarseness – (out of six) from the Tamura texture descriptor [50] that have the strongest correlation with human perception; CEDD (color and edge direction descriptor) [51], FCTH (fuzzy color and texture histogram) [52], Hu invariant moments[5] [53], MPEG-7 features [54] – CLD (color layout descriptor) and EHD (edge histogram descriptor), Gabor texture feature [55], and texture features, specifically, primitive length, edge frequency, and autocorrelation [56].

3) Shape Features—Set C: In this set, a collection of shape features provided by the standard MATLAB[6] implementation (`regionprops`) is used to describe properties of a lung region. Of these, we use: size, orientation, eccentricity, extent, centroid, and bounding box.

5.3   Performance Evaluation

A Receiver Operating Characteristic (ROC) curve plots the true positive rate versus the false positive rate. We used area under the ROC curve (AUC) and Accuracy (ACC) to measure performance of our classifier. A larger AUC value generally means a larger sensitivity and/or specificity for the same operating point [57]. The ROC curves show different possible operating points depending on the confidence threshold set for the classifier. The Y-axis indicates the sensitivity (or recall) of a system, and the X-axis indicates the corresponding false positive rate, which is (1 – specificity). ACC is computed as $ACC = (TP + TN)/Total$, where *TP* is the number of true positives and *TN* is the number of true negatives, and *Total* is the number of images.

We trained a support vector machine (SVM) [58, 59] classifier model on feature vectors computed for the Montgomery and Shenzhen data to discriminate between normal and abnormal CXR images. A unique characteristic of an SVM is that it classifies by computing the hyperplane with the largest margin between two classes; 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 sign of the distance corresponds to the "side" of the hyperplane that the feature vector lies in, in hyperspace.) The larger the distance the more confident we are of the class label. In our model we use this confidence measure to provide feedback to the operator. This enables confidence-based thresholding to find the optimal operating point of the classifier and the computation of the ROC curve. The optimal operating point of the classifier is the point on the ROC curve that provides the best ratio of sensitivity and specificity, given a cost function

---

[4] LIRE: https://code.google.com/p/lire/

[5] DISCOVIR (for Hu moments): http://appsrv.cse.cuhk.edu.hk/ ~miplab/discovir/

[6] MATLAB: http://www.mathworks.com

that describes the operational costs of missing an abnormal case in the field. Mathematically, this point is given by the following expression:

$$S = (|Normal| \cdot C(Abnormal \mid Normal)) / (|Abnormal| \cdot C(Normal \mid Abnormal))$$

Here $|Abnormal|$ is the number of abnormal (TB positive) cases in the training set, $|Normal|$ is the number of normal (TB negative) cases in the training set, $C(Abnormal \mid Normal)$ is the false positive cost, and, conversely, $C(Normal \mid Abnormal)$ is the false negative cost. The parameter S defines the slope of a line starting in the upper left corner of the ROC curve. When the false negative cost is much higher than the false positive cost, which is typically the case when screening large populations for TB, then S will be small. Consequently, the optimal operating point will reflect a higher sensitivity at the expense of specificity.
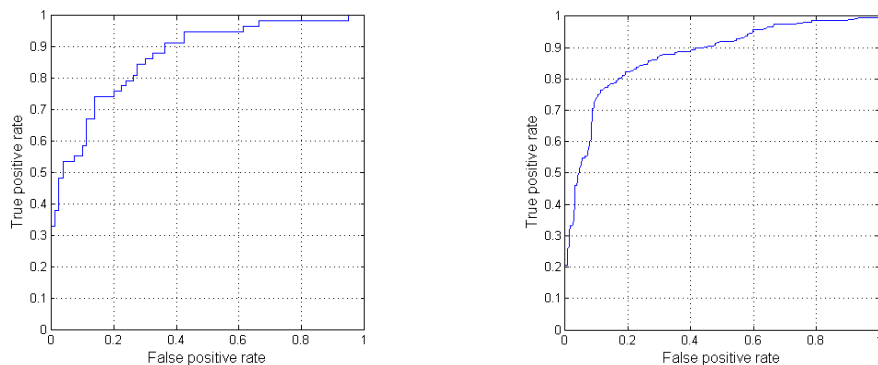

Figure 5. ROC curves for Montgomery dataset (Left) and Shenzhen dataset (Right) using SVM classifier on feature sets A and B

In [7] we reported results on using all features (Set A, B) using an SVM classifier for both the Montgomery and the Shenzhen datasets. The best performance was achieved for the Shenzhen set with Linear Logistic Regression, with an AUC of 0.9 and an accuracy (ACC) of 84.1%. The paper also reported the performance of an SVM with a linear kernel. It was slightly lower at an AUC of 0.885 and ACC of 82.1%. Figure 5 shows ROC curves computed for the datasets, using a linear SVM as classifier that has been trained and tested with features in Set A using a leave-one-out evaluation scheme. For the Montgomery dataset, the ROC curve, which is shown on the left-hand side of Figure 5, has an AUC of 0.86. For the Shenzhen dataset, the classifier performs a little better, as indicated by the slightly higher AUC value of 0.88 for the ROC shown on the right-hand side of the figure.

Next, we wanted to identify features that contribute the most to the classification result, and if an alternate classifier would be able to model the data better. A neural network classifier [60] was implemented due to the high dimensionality of the feature vector and relative ease of parallelization. The network also generates complex hyperplanes to separate samples from different classes by adjusting its internal parameters *(weights)* by minimizing the classification error at each node for different input training samples. All the features were extracted from the segmented lung region only. Several different experiments were conducted using both datasets. However, we report results for only the Shenzhen set because of its size and balance.

First we compute the effect of each feature set on the results.
Table 1 shows results using individual feature sets A, B, C, and the combined set, respectively, validating the superiority of the features in Set A.

Table 1. Accuracy an AUC measures using different features for the Shenzhen dataset.

Set A (%)		Set B (%)		Set C (%)		Combined Set A,B,C (%)	
ACC%	AUC	ACC%	AUC	ACC%	AUC	ACC%	AUC
95.57	0.99	81.06	0.90	70.40	0.77	92.00	0.97

Next, an experiment was designed to find the most effective combination of 18 individual features from the combined set. We performed an exhaustive feature space search by grouping different features in all possible combinations. The combination study resulted in 2^n-1 different possible feature groups for each test with *n features*. A parallel implementation of the evaluation was implemented using GPUs (graphical processing unit) for accelerated computation. The results in Table 2 show the best feature combinations. In the table the bolded features indicate those that are common for achieving a maximum AUC and ACC score. Recall that the accuracy is a point on the ROC curve, and even if we have the maximum AUC, we may not have the maximum ACC on that curve!

Table 2: Optimization results for the Shenzhen dataset.

Optimization criteria	Score	Score	Features
Max (ACC) %	83.73	97.03	**Gabor**, **HOG**, **LBP,** Auto Correlation
Max (AUC)	0.91	0.99	**Gabor**, **HOG**, **LBP,** FCTH, GM, CD

In
Table 1, the best scores for the Shenzhen were ACC of 95.57% and AUC of 0.99. After optimization the ACC increased to 97.03%, while the AUC of remained unchanged. The results demonstrate the importance of the feature selection and combination.

Table 3: False positive rates for decreasing recall values.

Dataset	Sensitivity		
	0.99	**0.95**	**0.90**
Montgomery	0.276	0.276	0.265
Shenzhen	0.0668	0.0230	0.0056

Figure 6. ROC curve showing performance of the Neural Network classifier on Shenzhen dataset

Table 3 reports false positive rate results at recall (sensitivity) of 0.90, 0.95 and 0.99. Note that for the Shenzhen data, the sensitivity is continues to remain very high even when more false positives are forcibly introduced. This result provides further support for feature selection. As shown in Figure 6, the ROC curve using the Neural Network classifier improves to near perfection (AUC=0.993, ACC=97.05%) showing a large improvement over SVM classifier with linear kernel (AUC=0.885, ACC=82.1%).
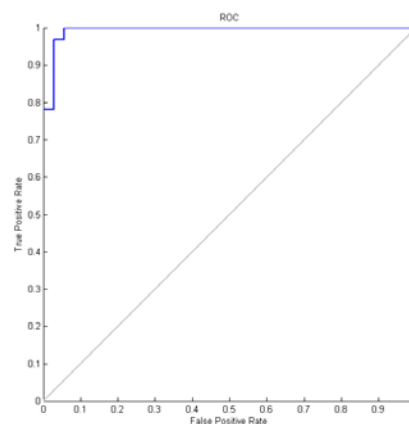
Note on current definition of "Abnormals": This performance evaluation is discussed as a 2-class classification between abnormals and normals. However, the abnormal datasets used in this are exclusively those positively identified as having active TB. Broadening the abnormal class to include other pulmonary or chest-cavity abnormalities, such as cardiomegaly, atelectasis, nodules, and others, remains a goal and is described in our next steps (Section 7).

Impact of Image Size on System Performance: To examine the scale invariance of the features, we also tested the performance of the SVM classifier using the Shenzhen dataset with different image scales. The

results indicate the features are not scale invariant, but are robust within a small range of image sizes near those used for training. It is better to use the model that was trained by the images that have similar scales as the images the tool is going to be applied on. Further, lower image sizes often results in elimination of image subtleties indicative of certain pulmonary diseases.

5.3.1 Comparing the Computer and the Screening Radiologist's Performance

To compare the performance of the system with the performance of radiologists, we asked two experienced radiologists to provide a second and third reading for the Montgomery CXRs. Both radiologists were working at a U.S. hospital at the time of our study. We asked them to follow the recommendations outlined in the WHO Lime book for diagnosing TB in endemic regions of the world [1]. In particular, we asked them to perform *intentional overreading* to ensure a high sensitivity so that missing a positive TB case becomes less likely. We adapted the Firefly labeling tool from the University of Missouri [40] to aid the radiologists in annotating the CXRs. Table 4 shows the agreement of both radiologists on the data, where R1 denotes the readings of the first radiologist and R2 denotes the readings of the second radiologist. A plus sign indicates CXRs that either the first or the second radiologist classified as TB positive, while a minus sign indicates CXRs that are, similarly, classified as normal. Both radiologists agree in 84.8% of the cases (95% CI:[77.7,90.3], using exact one-sample test of proportion). The kappa value is $\kappa = 0.69$ (95% CI:[0.52,0.86]), which signifies moderate agreement. After performing their individual readings, the two radiologists convened and came to a consensus decision for each CXR for which they have disagreed. For remaining cases, they agreed that their findings were consistent with TB.

Table 4. Radiologist agreement on Montgomery CXRs

	R2 +	R2 -	
R1 +	69	15	84
R1 -	6	48	54
	75	63	138

Table 5. Comparison of radiologist consensus performance with ground truth for Montgomery CXRs

	R +	R -	
GT +	58	0	58
GT -	25	55	80
	83	55	138

Table 5 shows a comparison of their consensus decision with ground-truth labels from Montgomery County's local TB clinic that are based on clinical data as well as patient data to which the radiologists had no access. Here the terms R+ and R- denote the X-rays that are TB positive or negative, respectively, according to the consensus decision of the radiologists. The rows, GT+ and GT-, represent the X-rays that are TB positive or negative as confirmed by clinical data or other patient information. The number of false negatives is zero in Table 5, which is the result of overreading. The radiologists therefore succeeded in detecting all TB positive cases, achieving a sensitivity (recall) of 100% (95% CI:[93.8,100]). The specificity is 68.8% (95% CI:[57.4,78.7]), which shows that overreading also results in a considerable number of false positives, namely 25 X-rays. 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 6. Comparison of machine performance with ground truth for Montgomery CXRs

	M +	M -	
GT +	43	15	58
GT -	15	65	80
	58	80	138

Table 7. Performance comparison between machine and radiologist consensus decisions for Montgomery CXRs

	R-correct	R-wrong	
M-correct	89	19	108
M-wrong	24	6	30
	113	25	138

Table 6 compares the agreement of machine classification with the ground-truth data. The columns M+ and M- represent the X-rays that are identified by the machine as TB positive or negative, respectively. For the machine output, the agreement with the ground-truth data is slightly lower compared to the radiologist consensus decisions shown in Table 5. The machine agrees with the ground-truth data in 78.3% of the cases (95% CI:[70.4,84.8]). It achieves a sensitivity of 74.1% (95% CI:[61.0,84.7]) and a specificity of 81.3% (95% CI:[71.0,89.1]), with the same number of false positives and false negatives, as shown in Table 6. This is because we have not optimized the sensitivity of the classifier used in Table 6. If we do so, the classifier's ROC curve for the Montgomery X-rays (Figure 1) indicates that in order for the sensitivity to be close to 100%, the classifier will produce a false positive rate slightly higher than 60%. This is about twice the false positive rate for the radiologists' consensus decision in Table 5, which is about 31%.

Finally, Table 7 compares the correct and incorrect consensus decisions of the two radiologists with the classification decision of the machine. The rows M-correct and M-wrong denote the correct and incorrect machine classifications, and the columns R-correct and R-wrong denote the correct and incorrect consensus decisions of the radiologists. In terms of classification performance, the radiologists are not significantly better than the machine (McNemar test, $p = 0.54$). The number of X-rays for which both the machine classification and the consensus decision of the radiologists are wrong is remarkably low. In fact, if an X-ray is considered correctly classified when either the machine or the radiologist consensus is correct, then this combined human-machine performance has an error rate of 4.3%. This is significantly lower than the machine-only error rate of 21.7% and the human consensus error rate of 18.1%. Therefore, using the system for making computer-aided diagnostic decisions is a promising approach. The system could provide a verifying second opinion of radiologist readings, which can help improve the performance of radiologists because it is unlikely that both the radiologist and the machine classify the same X-ray incorrectly.

5.4 Quality Control

This section describes our efforts to ensure that the images have minimal acquisition errors, and that our system implementation is robust.

5.4.1 A Novel Method for Image Rotation Correction

To acquire an upright CXR, a radiographer needs to align the patient so that the imaged region is orthogonal to the central axis of the x-ray beam cone. Deviation from this procedure often results in a rotated CXR. This is likely to occur in images acquired using portable machines in non-hospital settings or under more challenging outdoor conditions due to hardware limitations of the screening setup [29, 61]. Typically, rotation is estimated by analyzing the relationship of the medial heads of the clavicles to adjacent appearing vertebral spinous processes in the upper thorax region [7, 62-64]. However, this requires the accurate automatic detection of these anatomical regions; since a small spatial error in the location of these regions can adversely affect the computation of the amount of rotation in the image. State-of-the-art automatic methods use generalized line histogram techniques to compute a principal rib-orientation measure. While this may be adequate for human reading, the same may not hold for computerized reading of CXRs. In our method [65], we compute principal rib- orientation using a set of filter parameters (*kernels*) that attempt to account for all possible orientations $[0, 2\pi]$. As shown in Figure 7, we convolve the edge image of the lung in the CXR with filters that use these kernels. We then transform the resulting sets of lines into a 1D histogram, where the maximum magnitude is the principal rib-orientation.
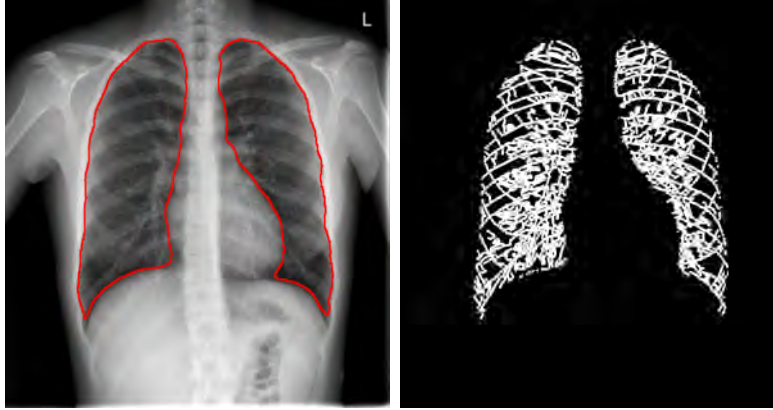
Figure 7. Segmented lung region in CXR and final edge map image with superimposed results from kernel line filters applied

For greater robustness, we apply the method to each lung separately. The CXR is considered upright if the difference in orientation angle for each lung is zero or within acceptable tolerance of zero. Our method is also robust to missing ribs. The method was tested on the Montgomery and Indiana datasets and resulted in a maximum accuracy of 90% with a tight 2° tolerance.
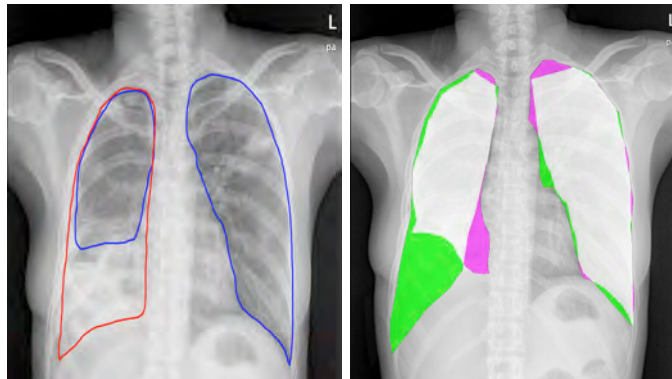
5.4.2 Lung Segmentation and Classification

Figure 8. (Left) Consolidation in right lung results in smaller lung boundary (blue line) compared to manual markup (red line). (Right) Two radiolgists' markup (green and magenta) and overlap (white) superimposed on a CXR

Our lung segmentation algorithm uses an atlas-based approach to dynamically fit similar appearing annotated model CXRs to the patient CXR by iteratively registering air cavity spaces. Therefore, the resulting binary mask has a very strong preference to outline the air cavity space ("dark region"). Figure 8 (left) shows this effect on a CXR diagnosed with pleural effusion: the green line outlines the output of the segmentation algorithm and the red line outlines the actual anatomy of the lung field. We pose the question: does this limitation have an adverse effect on classification because features are extracted only within the air cavity space whereas an abnormality, such as consolidation, may lie outside the segmented regions? Further, it is difficult to determine a priori when the segmentation algorithm parameters could be relaxed to include lower lobe opacities, by how much, and if the modified segmentation result would be adequate. To determine the need for parameter modification, we selected 48 abnormal CXRs with and without lung opacities from the Shenzhen dataset. Two experienced radiologists manually annotated the lung outlines. The outlines were overlapped to compute inter-observer agreement (Figure 8, right). The results revealed that while the radiologists had excellent inter-observer agreement on normal lung fields ($(\mu, \sigma) = (86\%, 13.6)$), they scored much lower agreement on abnormal lung fields ($(\mu, \sigma) = (73\%, 18.1)$). To measure the impact on overall classification, we modified the segmentation algorithm parameters such

that its output closely matched the annotated lung boundaries. The resulting analysis showed that adjusting segmentation parameters did not meaningfully improve the overall performance. However, we observed a 2.4% increase in AUC for TB classification when we incorporated shape differences between left and right lung fields as features into our existing feature set.

5.5 Lung classification without region segmentation

The previous method, outlined above, is based on lung segmentation followed by image feature extraction and classification. While it is powerful in its capabilities, there may be alternatives for quickly identifying abnormal CXRs. We note that radiographic presentations of pulmonary disease in CXR images such as, cavitation, consolidation, infiltrates, and small broadly distributed nodules, alter the appearance of the image in the form of corrupted and or deformed rib edge profiles. Figure 9 shows examples from both normal and abnormal CXRs and their corresponding rib profiles that we call *edge maps*. In this method, we measure changes in the rib profiles within the chest cavity to identify abnormal CXR images. We use a single feature: pyramid histogram of orientation gradients (PHOG) [66] using edge map with the goal of improving quality and speed of abnormality detection. A key advantage of the proposed method is that it does not require precise lung contours and an approximate bounding box around the lungs is sufficient. The Montgomery county and Shenzhen datasets were used in the evaluation. We used a neural network based multi-layer perceptron with one hidden layer for classification. The PHOG feature is parameterized by *levels* and *orientation bins*.
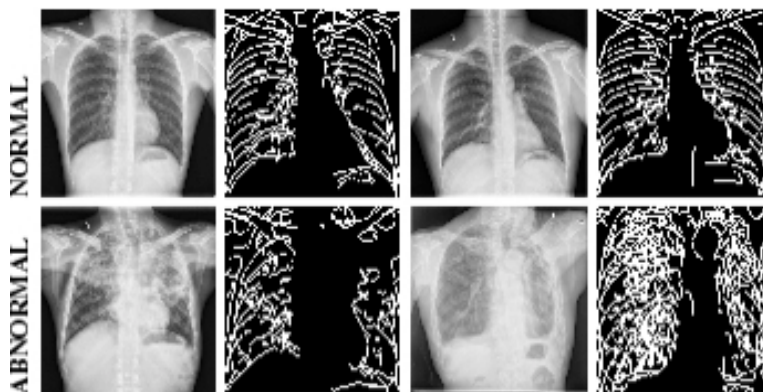

Figure 9. Edge maps from normal (top row) and abnormal (bottom row)

Results in Figure 10 show that we achieved the best performance on the Shenzhen data set, using level 2 with 15 orientation bins, where we reached a maximum accuracy (ACC) of 85.92% and area under the curve (AUC) of 0.91. These results nominally outperform the current system [7], and the method is approximately 25 times faster. An additional advantage of the system is that it is robust against opacities due to consolidation or pleural effusion. It does not at present have the capability of isolating pulmonary abnormalities; however, it can be quite powerful as an efficient and reliable staging module. We are considering including this capability as one processing stage in the deployed system (Section 6).
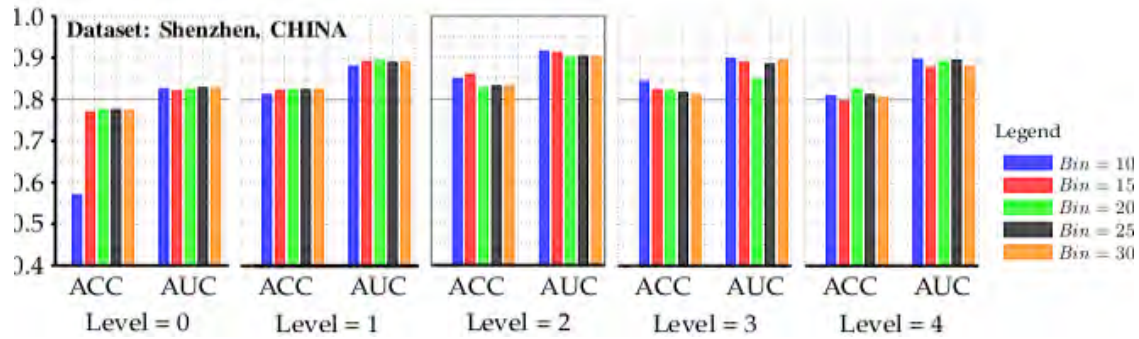
Figure 10. Accuracy (ACC) in % and Area under the curve (AUC) at different number of bins and levels for the Shenzhen dataset. The best performance is at Level = 2 and 15 orientation

6   Screening Application Design and Development

Figure 11 AMPATH "Karibuni" or mobile X-ray truck

Figure 12. Truck running on generator power and with X-ray machine being conditioned by Dr. Marc Kohli (Indiana University School of Medicine.)

A primary goal of this research is to make the algorithms realizable on a small-scale low-powered system that can be deployed in low resource settings. The driving example (pun intended) for us is deploying the system on a generator-powered truck in rural Kenya, shown in Figure 11 and Figure 12 The truck houses a portable X-ray machine, cassette reader, and laptop, the screening system and a WiFi capable router that are powered with a 7.5KW diesel generator. Figure 13 shows the portable X-ray machine inside the truck

(left), the cassette reader (left, on the shelf), and (right) the image acquisition laptop, the screening system (small black box) and router. The schematic of the workflow shown in Figure 14 has the following steps. First, the radiographer prepares the patient for CXR imaging following the presentation of the prescription. The prescription indicates the reason why the patient is being imaged. Basic patient information, the prescription, and rural imaging site information are recorded in the image acquisition software. After the imaging, the exposed cassette is inserted into the reader and the image is imported into the image acquisition software as a patient exam. The radiographer verifies appropriateness of the image quality. Next, he straightens the image, and marks the patient's left side on the image. The image exam is then accepted and saved on the local PACS [7].

Figure 13. (Left) Radiographer Peter Otunga preparing a patient for X-ray imaging with X-ray cassette reader on shelf. (Right) On-board image acquisition laptop showing screening system UI, NLM Screening system (black box on the right), and router.
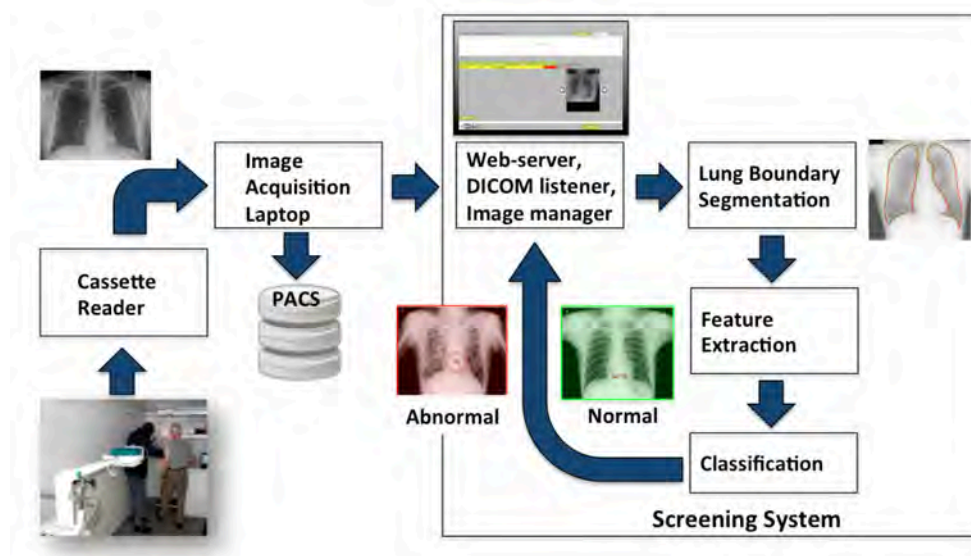

Figure 14. Schematic showing the image acquisition and screening workflow. The NLM Screening System Software modules appear in the box on the right side of the figure.

NLM Screening System Hardware Design: Key hardware considerations for the screening software systems included: (1) low power consumption; (2) avoid use of a computer screen; (3) low cost; and (4)

---

[7] PACS: Picture Archiving and Communications System is a special system for storing radiological images.

concealable (for theft prevention). We selected an Intel NUC PC and configured it with 8 GB RAM and 256 GB SSD storage. The unit, shown in Figure 13 (right)**Error! Reference source not found.**, is approximately 4.3" x 4.3" x 1.3" in physical dimensions and has a peak power consumption of 50W. It supports wired Gigabit Ethernet LAN, WiFi (802.11n), and Bluetooth® communications.

NLM Screening Tool Software Design: Images that are acquired and stored in the PACS database are in DICOM[8] format. The DICOM software has the capability to "send" the images to any other software on the network that identifies itself as a DICOM NODE. Such a node acts as a *listener* and adheres to the specifications of the DICOM protocol. The radiographer (or assistant) submits the patient's CXR to the listener, which submits the image to the processing system, and, on completion, records the classification results, and displays them in a Web browser.

The server side of the screening software contains several important components, shown in Figure 14. They include the DICOM listener, Web server, lung segmentation module, and image classification module. The user interface is implemented using AJAX on the client side, which communicates with a Java servlet on the server side. The list of processed images is stored locally and shown in the history table. The classification algorithms are written in MATLAB that is wrapped in server software written in Java. The modules were tested for robustness and efficiency before they were included in the final system. The system completes its classification and display of results in less than 20 seconds.

7 Summary and Next Steps

The report describes ongoing research in improving the state of the art in automated computer-aided diagnosis algorithms for analyzing posteroanterior CXRs for cardiopulmonary diseases. We have initially focused on TB since opportunistic coinfections with immunocompromised HIV/AIDS patients have exacerbated the TB problem in sub-Saharan Africa. Our collaborators at AMPATH are running one of the largest HIV treatment programs in Kenya. With the assistance of NLM they have designed and put in operation trucks with portable X-ray units to increase their reach in rural areas and curb the spread of the disease. The system was deployed in February 2015, and is processing an average of 40 patients per week. Our algorithms could be significantly improved with data from the truck in Kenya, and we are working through the necessary IRB process.

Table 8. Example pediatric CXRs acquired at various ages showing the variety in lung shapes and imaged regions
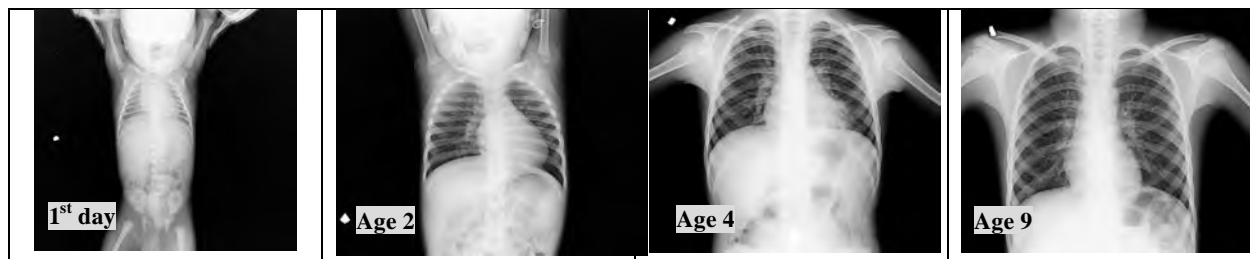
Next Steps: The following enhancements are being studied as next steps:

1. **Pediatric CXR analysis:** Our progress in adult CXR screening has piqued the interest of pediatric TB researchers at the National Institute of Allergic and Infectious Diseases (NIAID). There is an interest in

---

[8] DICOM: Digital Imaging and COmunications in Medicine is the international standard for medical images and related information (http://dicom.nema.org)

testing our system with images from their ongoing research in South Africa. As a first step, we evaluated our model-based lung segmentation algorithm on pediatric images and recorded reduced accuracy. As shown in Table 8, the lung shape changes during the pediatric developmental stages. For an infant the lungs are smaller, have a triangular shape, and, the cardiac contour is relatively large. The horizontal diameter of the heart may approach 60% of thoracic horizontal diameter [67]. Also, for infants and children much of the body is also often imaged which can distract the segmentation module. But, merely adding pediatric models to the atlas did not improve results. We conducted an image analysis study, which revealed three lung shape-based pediatric age groupings of 0–2, 2-10, and 11-18 years. Modifying the algorithm according to the schematic in Figure 15, where age information is used to select the lung models, resulted in up to 230% improvement in segmentation accuracy measured using the Dice similarity measure[9] [68].
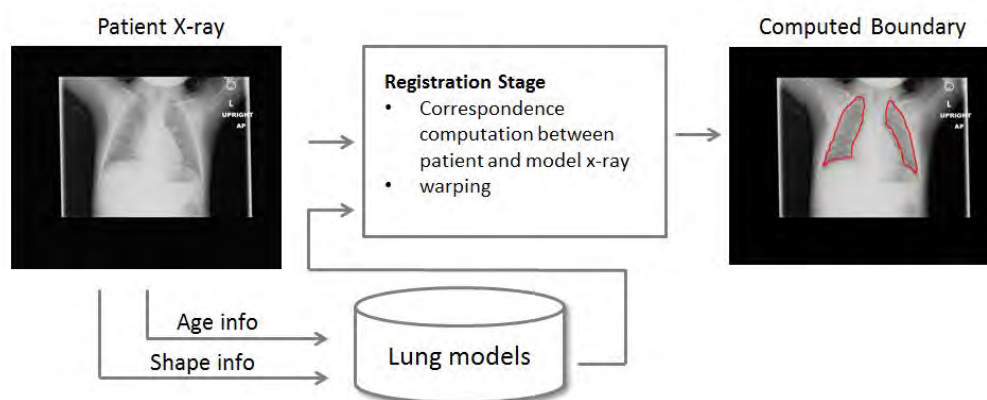

Figure 15. Modified lung region segmentation method for adult and pediatric CXR

2. **Different pulmonary diseases:** So far, all of our studies in this project have used datasets where the abnormal cases are explicitly active TB. In addition to screening for TB prevalence in HIV positive populations (~2.5% in Kenya [1]), it is our goal to also detect other cardiopulmonary diseases. Initial tests on the Indiana dataset, which has no cases of active TB, using the Neural Network classifier, described in Section 5.3, results in comparatively unimpressive scores. Even after feature selection the maximum ACC is 68.83% with an AUC of 0.74. When the sensitivity is set at 0.99, the false positive rate (FPR) is 0.86 (compared to Shenzhen TB dataset FPR of 0.0668). These results suggest that advanced algorithms need to be developed for detecting different cardiopulmonary diseases in CXRs.

---

[9] Dice Index: A statistic used for comparing the similarity of two samples. http://en.wikipedia.org/wiki/Sørensen–Dice_coefficient
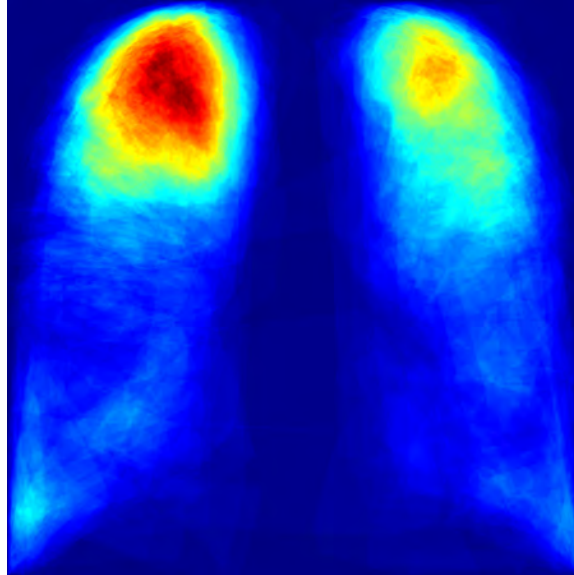
Figure 16. Heat map showing locations of TB infection marked on Shenzhen dataset images.

3. **Pulmonary disease localization:** When we deployed the software in Kenya, we observed that the radiographer on the truck informs the clinical officer at the rural clinic (who may not be trained in radiology) of noticeable abnormality in a CXR. A radiologist reads the image only days later; this is a significant delay that could result in AMPATH (or, similar organization in other regions where the system may be deployed) "losing" the patient. The screening system currently provides an immediate decision, but it does not provide information about the location of the abnormality. It does provide its confidence for the overall classification result. We plan on developing a heat-map interface that will suggest location information on the abnormality and the confidence of the classifier in its decision. Warmer colors indicate greater confidence while cooler colors indicate lower confidence. An initial rendering of the heat map is shown in Figure 16. It shows location of disease in the Shenzhen dataset.

Figure 17. Example AMPATH radiology order for a
"baseline" reading.

4. **Using radiology order information for improving classifier decision:** When patients are referred for imaging, the ordering physician specifies a reason that is recorded in the DICOM image header. An example AMPATH imaging prescription, shown in Figure 17, typically contains the patient name, birth year, gender, and reason for imaging. We believe that using the clinical findings (or imaging

19

order) to adjust the classifier threshold may result in improved specificity without loss of sensitivity measures.

In conclusion, this project is a unique combination of imaging research, machine learning, and practical deployment challenges of a computer-aided screening system in resource constrained regions. Initial results suggest that this project is making good on meeting NIH global health goals and NLM's imaging research and technology exploration objectives, while making significant impact in addressing the diagnostic needs of this critical health challenge.

References

1. World Health Org., Global Tuberculosis Report 2014.
2. World Health Organization. Tuberculosis Prevalence Surveys: A Handbook. 2011.
3. Stop TB Partnership, World Health Organization. The Global Plan to Stop TB 2011–2015.
4. Leung C. Reexamining the role of radiography in tuberculosis case finding. Int. J. Tuberculosis Lung Disease, 15(10):1279, 2011.
5. Folio, L., Chest Imaging: An Algorithmic Approach to Learning, Springer, 2012
6. Candemir, S., Jaeger, S., Palaniappan, K., Musco, J. P., Singh, R. K., Xue, Z., Karargyris A., Antani S., Thoma G., McDonald, C. J. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. Medical Imaging, IEEE Transactions on, 33(2), 577-590, 2014.
7. Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan FM, Xue Z, Palaniappan K, Singh RK, Antani S, Thoma G, Wang Y, Lu P, McDonald C. Automatic tuberculosis screening using chest radiographs. IEEE Trans. Med. Imaging, 33(2):233–45, 2014.
8. Daley C, Gotway M, Jasmer R. Radiographic manifestations of tuberculosis. A Primer for Clinicians. San Francisco, CA: Curry International Tuberculosis Center, 2009.
9. Folio LR. Robochest Web Teaching Tool. Available online: www.robochest.com (last accessed 10 April 2013)
10. Lee KS, Song KS, Lim TH, et al. Adult-onset pulmonary tuberculosis: findings on chest radiographs and CT scans. AJR Am J Roentgenol. 160:753-8, 1993.
11. van Ginneken B, Hogeweg L, Prokop M. Computer-aided diagnosis in chest radiography: Beyond nodules. Eur. J. Radiol., 72(2):226–30, 2009.
12. Lodwick GS. Computer-aided diagnosis in radiology: A research plan. Investigative Radiology, 1(1):72, 1966.
13. Lodwick G, Keats T, Dorst J. The coding of Roentgen images for computer analysis as applied to lung cancer. Radiology, 81(2):185, 1963.
14. Sakai S, Soeda H, Takahashi N, Okafuji T, Yoshitake T, Yabuuchi H, Yoshino I, Yamamoto K, Honda H, Doi K. Computer-aided nodule detection on digital chest radiography: Validation test on consecutive T1 cases of resectable lung cancer. J. Digital Imaging, 19(4):376–82, 2006.
15. Shiraishi J, Abe H, Li F, Engelmann R, MacMahon H, Doi K. Computer-aided diagnosis for the detection and classification of lung cancers on chest radiographs: ROC analysis of radiologists' performance. Acad. Radiol., 13(8):995–1003, 2006.
16. Kakeda S, Moriya J, Sato H, Aoki T, Watanabe H, Nakata H, Oda N, Katsuragawa S, Yamamoto K, Doi K. Improved detection of lung nodules on chest radiographs using a commercial computer- aided diagnosis system. Am. J. Roentgenol., 182(2):505-10, 2004.
17. Doi K. Current status and future potential of computer-aided diagnosis inmedical imaging. Br. J. Radiol. 78(1):3-19, 2005.
18. van Ginneken B, ter Haar Romeny B, Viergever M. Computeraided diagnosis in chest radiography: A survey. IEEE Trans. Med. Imag., 20(12):1228–41, 2001.
19. van Ginneken B, Stegmann M,Loog M. Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database. Med. Image Anal., 10(1):19–40, 2006.
20. van Ginneken B, ter Haar Romeny B. Automatic segmentation of lung fields in chest radiographs. Med. Phys., 27(10):2445-55, 2000.
21. Dawoud A. Fusing shape information in lung segmentation in chest radiographs. Image Anal. Recognit., pp. 70–8, 2010.
22. van Ginneken B, Katsuragawa S, ter Haar Romeny B, Doi K, Viergever M. Automatic detection of abnormalities in chest radiographs using local texture analysis. IEEE Trans. Med. Imag., 21(2):139–49, 2002.
23. Hogeweg L, Mol C, de Jong P, Dawson R, Ayles H, van Ginneken B. Fusion of local and global detection systems to detect tuberculosis in chest radiographs. Proc. MICCAI, pp. 650–57, 2010.

24. Shen R, Cheng I, Basu A. A hybrid knowledge-guided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs. IEEE Trans. Biomed. Eng. 57(11): 2646–56, 2010.

25. Xu T, Cheng I, Mandal M. Automated cavity detection of infectious pulmonary tuberculosis in chest radiographs. Int. IEEE Eng. Med. Biol. Soc., pp. 5178–81, 2011.

26. Hogeweg L, Sánchez CI, de Jong PA, Maduskar P, van Ginneken B. Clavicle segmentation in chest radiographs. Med. Image Anal. 16(8):1490–1502, 2012.

27. Freedman M, Lo S, Seibel J, Bromley C. Lung nodules: Improved detection with software that suppresses the rib and clavicle on chest radiographs. Radiology, 260(1): 265–73, 2011.

28. Arzhaeva Y, Tax D, van Ginneken B, "Dissimilarity-based classification in the absence of local ground truth: Application to the diagnostic interpretation of chest radiographs. Pattern Recognition, 42(9):1768–76, 2009.

29. Jaeger S, Karargyris A, Candemir S, Siegelman J, Folio L, Antani S, Thoma G. Automatic screening for tuberculosis in chest radiographs: a survey. Quant Imaging Med Surg. 2013; 3(2):89-99.

30. Shiraishi J, Katsuragawa S, Ikezoe J, Matsumoto T, Kobayashi T, Komatsu K, Matsui M, Fujita H, Kodera Y, Doi K. Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists detection of pulmonary nodules. Am. J. Roentgenol., 174:71–4, 2000.

31. van Ginneken B, Stegmann M, Loog M. Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database. Med. Image Anal., 10(1):19-40, 2006.

32. Jaeger S, Candemir S, Antani S, Wáng YX, Lu PX, Thoma G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant Imaging Med Surg 2014;4(6):475-477.

33. Jain R, Kasturi R, Schunck BG. Machine Vision. McGraw-Hill, Inc., ISBN 0-07-032018-7, 1995

34. Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distributions. Bull. Calcutta Math. Soc. 35:99–109, 1943.

35. Liu C, Yuen J, Torralba A. SIFT flow: Dense correspondence across different scenes and its applications. IEEE Trans. Pattern Anal. Mach. Intell., 33(5): 978-94, 2011.

36. Lowe D. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis., 60(2):91-110, 2004.

37. Boykov Y, Funka-Lea G. Graph cuts and efficient n-d image segmentation. Int. J. Comput. Vis., 70:109–131, 2006.

38. Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell., 23(11):1222–39, 2001.

39. Gururajan A, Sari-Sarraf H, Hequet E. Interactive texture segmentation via IT-SNAPS. IEEE Southwest Symp. Image Anal. Interpret., 2010.

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

41. Jaeger S, Casas-Delucchi C, Cardoso M, Palaniappan K. Dual channel colocalization for cell cycle analysis using 3D confocal microscopy. Proc. Int. Conf. Pattern Recognit., pp. 2580–3, 2010.

42. Jaeger S, Casas-Delucchi C, Cardoso M, Palaniappan K. Classification of cell cycle phases in 3D confocal microscopy using PCNA and chromocenter features. Proc. Indian Conf. Comput. Vis., Graph., Image Process., pp. 412–418, 2010

43. Dalal N, Triggs B. Histograms of oriented gradients for human detection. Proc. Int. Conf. Comp. Vis. Patt. Recognit., 1:886–93, 2005

44. Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell., 24(7): 971–87, 2002.

45. Ojala TPietikäinen M, Harwood D. A comparative study of texture measures with classification based on feature distributions," Pattern Recognit., 29:51–9, 1996.

46. Frangi A, Niessen W, Vincken K, Viergever M. Multiscale vessel enhancement filtering. Proc. MICCAI, pp. 130–7, 1998.

47. Bunyak F, Palaniappan K, Glinskii O, Glinskii V, Glinsky V, Huxley V. Epifluorescence-based quantitative microvasculature remodeling using geodesic level-sets and shape-based evolution. Proc. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 3134–7, 2008.

48. Simpson MS, You D, Rahman MM, Xue Z, Demner-Fushman D, Antani S, Thoma G. Literature-based biomedical image classification and retrieval. Comput Med Imaging Graph. 2015 Jan;39:3-13. doi: 10.1016/j.compmedimag.2014.06.006.
49. Lux M. Content-based image retrieval with LIRE. Proc. ACM Int. Conf. Multimedia, pp. 735–8, 2011.
50. Howarth P, Rüger S. Robust texture features for still-image retrieval. IEE Proc. Vis., Image Signal Process. 152(6):868–74, 2005.
51. Chatzichristofis S, Boutalis Y. CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. Comput. Vis. Syst., pp. 312–22, 2008.
52. Chatzichristofis S, Boutalis Y. FCTH: Fuzzy color and texture histogram—A low level feature for accurate image retrieval. Proc. Int. Workshop Image Anal. Multimedia Interactive Services, pp. 191–6, 2008.
53. Hu M. Visual pattern recognition by moment invariants. IRE Trans. Inf. Theory, 8(2) pp. 179–87, 1962.
54. Lux M. Caliph & Emir: MPEG-7 photo annotation and retrieval. Proc. Int. Conf. Multimedia, pp. 925–6, 2009.
55. Manjunath BS, Ma WY. Texture features for browsing and retrieval of image data. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI - Special issue on Digital Libraries), 18(8):837-42, 1996.
56. Srinivasan G, Shobha G, Statistical texture analysis. Proc. World Acad. Sci., Eng. Technol., 36:1264–9, 2008.
57. Fawcett T, An introduction to ROC analysis. Pattern Recognition Letters 27(8): 861-74, 2006.
58. Vapnik V. The Nature of Statistical Learning Theory. NewYork: Springer Verlag, 2000.
59. Schölkopf B, Burges C, Smola A. Advances in Kernel Methods: Support Vector Learning. Cambridge, MA: MIT Press, 1999.
60. Jordan MI, Bishop CM: Neural Networks. Computing Handbook, 3rd ed. 2014
61. Bongard F, Sue D, Vintch, J. Current Diagnosis and Treatment Critical Care, McGraw-Hill Education, 2008
62. Corne J, Pointon, K. Chest X-Ray Made Easy, Elsevier Health Sciences, 2010
63. Puddy E, Hill C. Interpretation of the chest radiograph, Continuing Education in Anaesthesia, Critical Care & Pain 7(3):71–5, 2007
64. Herring W. Learning Radiology: Recognizing the Basics, Elsevier Health Sciences, 2011
65. KC S, Candemir S, Jaeger S, Folio L, Karargyris A, Antani SK, Thoma GR. Automatically detecting rotation in chest radiographs using principal rib-orientation measure for quality control. Int. J. Patt. Recogn. Artif. Intell, 29(2), 2015. http://www.worldscientific.com/doi/abs/10.1142/S0218001415570013
66. Zisserman BA, Munoz X, "Representing shape with a spatial pyramid kernel," in ACM Int. Conf. Image and Video Retrieval, 401–8, 2007.
67. Arthur R. Interpretation of the paediatric chest X-ray. Paediatric respiratory reviews 1.1, 41-50, 2000.
68. Candemir S, Antani SK, Jaeger SR, Thoma GR. Lung boundary detection in pediatric chest x-rays. PACS and Imaging Informatics: Next Generation and Innovations, SPIE Medical Imaging, Vol. 9418, 2015.