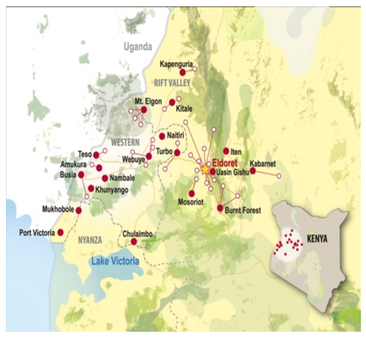
A System for Automated Screening for Tuberculosis using Digital Chest X-rays for Resource-Constrained Regions

Zhiyun Xue, Stefan Jaeger, Alexandros Karargyris, Sema Candemir, Sameer Antani, Rodney Long, George Thoma, and Clement McDonald
Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, National Institutes of Health (NIH)

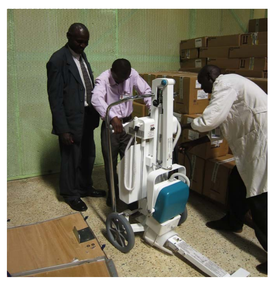
Background

Tuberculosis (TB) is an infectious disease caused by various strains of mycobacteria. It is widespread; about one-third of the world's population has latent TB.

The National Library of Medicine (NLM) is collaborating with AMPATH (the Academic Model Providing Access to Healthcare), an organization that runs the largest AIDS treatment program in the world, to develop a system that screens patients in resource-constrained rural Kenya for evidence of pulmonary tuberculosis in chest x-rays.



NLM has provided AMPATH with lightweight digital x-ray units for use by their staff to screen the population for the presence of disease. These units are already on site in Eldoret, in western Kenya, and are being readied for deployment in a mobile x-ray truck.



Computer-Aided Screening System

The computer-aided screening system is being designed to automatically analyze the patient chest CXR after it is acquired. It comprises a mobile CXR unit and reader mounted on a truck, a laptop computer, and a software system. The CXR unit is connected to a portable DICOM workstation, which pushes images in a PACS framework. Our software runs on a laptop computer that is connected to and listens to the DICOM workstation. The software system can automatically receive DICOM files and store them locally. It can also invoke the screening method, which has three major components: lung segmentation, feature extraction, and classification. First, lung regions are identified automatically using graph-cut segmentation methods. Then, various texture and shape features are calculated from the extracted lung regions. Lastly, a classifier categorizes features to determine whether the input chest x-ray is normal or abnormal. We tested several types of features and classifiers.

Features

1) Object Detection Features - Set A
- Intensity histograms
- Gradient Magnitude Histograms
- Shape descriptor histograms
- Curvature descriptor histograms
- Histogram of oriented gradients
- Local binary patterns

2) CBIR-based Image Features - Set B
- Tamura texture descriptor
- Color and edge directivity descriptor
- Fuzzy color and texture histograms
- Hu moments
- Color layout descriptor
- Edge histogram descriptor
- Primitive length
- Edge frequency
- Autocorrelation
- Standard shape features

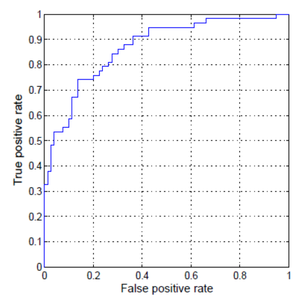
Results

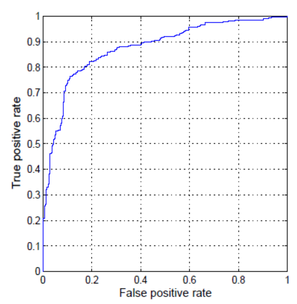
System performance has been evaluated on two TB datasets with promising results.

Datasets

Dataset 1: a representative subset of a CXR repository collected over many years within the tuberculosis control program of the Department of Health and Human Services of Montgomery County. It contains 138 posterior-anterior CXRs, among which 80 CXRs are normal and 58 CXRs are abnormal with manifestations of TB.
Dataset 2: a set collected at Shenzhen No.3 People's Hospital in China during a month period of time. The set contains 340 normal CXRs and 275 abnormal CXRs with TB.

Classification Performance





ROC curve for Dataset 1 and Feature Set A. ROC curve for Dataset 2 and Feature Set A.

Classification Performance (AUC)

	Feature Set A	Feature Set B
Dataset 1	86.9	80.1
Dataset 2	88.0	88.5

Classification Performance for Different Classifiers on Dataset 2 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

Acknowledgment

This research is supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), Lister Hill National Center for Biomedical Communications (LHNCBC).