Montgomery County - Chest X-ray Database

Description: The standard digital image database for Tuberculosis is created by the National Library of Medicine in collaboration with the Department of Health and Human Services, Montgomery County, Maryland, USA. The set contains data from X-rays collected under Montgomery County`s Tuberculosis screening program.

Number of X-rays:

- 58 cases with manifestation of tuberculosis, and
- 80 normal cases.

Image parameters:

- Format: PNG
- Matrix size is 4020 x 4892, or 4892 x 4020.
- The pixel spacing in vertical and horizontal directions is 0.0875 mm.
- Number of gray levels is 12bits.

Image file names are coded as MCUCXR_#####_0/1.png, where '0' represents the normal and '1' represents the abnormal lung.

The clinical readings for X-rays are saved as text file with the same file name as MCUCXR_#####_0/1.txt. Text file contains the patient`s age, sex and abnormality of the lung.

Lung Annotations

We manually generated the "gold standard" segmentations for the chest X-ray under the supervision of a radiologist. We used the following conventions for outlining the lung boundaries: Both posterior and anterior ribs are readily visible in the CXRs; the part of the lung behind the heart is excluded. We follow anatomical landmarks such as the boundary of the hearth, aortic arc/line, and pericardium line; and sharp costophrenic angle that follow the diaphragm boundary. We draw an inferred boundary when the pathology is severe and affects the morphological appearance of the lungs.

The lung boundaries (left and right) are in binary image format and have the same file name as chest X-rays (e.g. …/left/MCUCXR_#####_0/1.png or …/right/MCUCXR_#####_0/1.png).

IRB Details: The dataset were de-identified by the data providers and were exempted from IRB review at their institutions. The data set use and public release were exempted from IRB review (No. 5357) by the NIH Office of Human Research Projections Programs.

Citation and Use:

(a) Please do not share the data set outside of your research group/organization, but forward new data set requests to us.

(b) Please inform us if you find errors or inconsistencies in the data.

(c) It is requested that publications resulting from the use of this data attribute the source (National Library of Medicine, National Institutes of Health, Bethesda, MD, USA) and cite the following publications:

1) Candemir S, Jaeger S, Musco J, Xue Z, Karargyris A, Antani SK, Thoma GR, Palaniappan K. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. IEEE Trans Med Imaging. 2014 Feb;33(2):577-90. doi: 10.1109/TMI.2013.2290491. PMID: 24239990
2) Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan FM, Xue Z, Palaniappan K, Singh RK, Antani SK. Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imaging. 2014 Feb;33(2):233-45. doi: 10.1109/TMI.2013.2284099. PMID: 24108713

Acknowledgement: Staff involved in the collection of this data:

NLM: Data Collection/Curation

- Sameer Antani, PhD
- Sema Candemir, PhD
- Stefan Jaeger, PhD
- Alex Karargyris, PhD
- George R. Thoma, PhD

NLM: Data Validation

- Michael Bonifant
- Ellan Kim

Data provider: Sonia Qasba, M.D., Medical Director, Tuberculosis Control Program, Montgomery County, Maryland, USA.

"Gold Standard" Review: Jonathan Musco, M.D., Department of Radiology, School of Medicine, University of Missouri-Columbia, Columbia, Missouri, USA.

For additional information, please contact: Dr. Sameer Antani, Phone: +1 301-435-3218, Email: sameer *dot* antani AT nih *dot* gov