This data is distributed under a [Creative Commons Attribution license](https://creativecommons.org/licenses/by/4.0/) CC-BY icon.

**Description of the annotation dataset:**

The annotation collection comprises three parts:

1) Annotation files in JSON (JavaScript Object Notation) format that indicate locations and shapes of 19 lung pattern abnormalities for 336 TB patients.

2) Mask files saved in PNG format for each abnormality per TB patient;

3) A CSV (comma-separated values) file that summarizes lung abnormality types and numbers per TB patient. To the best of our knowledge, this is the first collection of pixel-level annotations of TB-consistent findings in CXRs.

Please refer to our paper Ref. 1 for more details.

Original images associated with the annotations are stored in: <https://data.lhncbc.nlm.nih.gov/public/Tuberculosis-Chest-X-ray-Datasets/Shenzhen-Hospital-CXR-Set/CXR_png/index.html>

**Reference:**

We request 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 publication, which has used the data for parasite detection and classification:

1. Yang, F.; Lu, P.X.; Deng, M.; Wáng, Y.X.J.; Rajaraman, S.; Xue, Z.; Folio, L.R.; Antani, S.K.; Jaeger, S. Annotations of Lung Abnormalities in the Shenzhen Chest X-ray Dataset for Computer-Aided Screening of Pulmonary Diseases. Data 2022, 7, 95. DOI: https://www.mdpi.com/2306-5729/7/7/95.