Simulated X-ray Radiography for Synthetic TB Data Generation

Preveshin Maduray[1], **Timothy Brooks**[1]

[1] *University of Johannesburg, Auckland Park*
preveshinmaduray@gmail.com

1. Introduction

The project focuses on training Machine Learning (ML) algorithms to detect pulmonary Tuberculosis (TB) infections in clinical imaging. We have identified 5 publicly available datasets of TB infected lung radiographs in medical literature. These datasets have been used repeatedly to train ML algorithms to detect TB.[1] The main reason for improvements in accuracy of TB detection has been the use of increasingly sophisticated algorithms.[1234] To introduce a high volume of new data with known TB indicators, we propose to generate synthetic data from high-resolution digital twin lungs. The study will leverage Hierarchical Phase-Contrast Tomography (HiP-CT) scans which are produced at the European Synchrotron Research Facility (ESRF) beamline BM18. These are high-resolution 3D scans (~20 µm) of whole organs, with regions of sub 2.5 µm resolution. Scans will be taken of ex-vivo healthy and TB-infected lungs. The data will be processed into a large number of 3D models that include a known set of TB indicators. These models will be reprojected into simulated Chest X-rays (CXR) with software using the Geant4 toolkit which simulates the passage of particles through matter. These simulated radiographs with their labelled pathologies will then be used to train classification algorithms for the detection of TB in the lung. This method could be extended to various other organs in future, and as such can assist us in improving the diversity of datasets utilized by the ML community.

2. Results

Figure 1 is a CXR of a patient infected with TB. The dark region on the left side is a pulmonary cavitation. This occurs when normal lung tissue dies from the infection, thus becoming gas filled. Figure 2 is a demonstration of simulating a CXR using a low-resolution 2D model. The X-rays passing from source to detector interact with several tissue classes and scatter realistically. The detector element can record the X-ray dose absorbed to build a 2D radiograph like in figure 1. We intend to use high resolution HiP-CT data which, due to the non-linear progression of TB, can provide a detailed understanding of the evolution of pathologies down to a cellular level. We aim to combine scans into many digital organs which will exhibit differing presentations of the infection and include truth data allowing optimised training of ML algorithms.

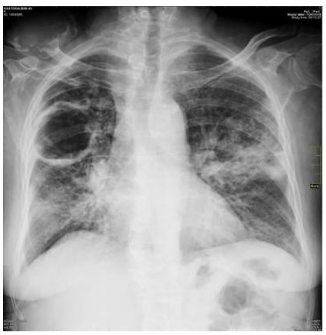

Fig. 1: An X-ray image of a TB infected lung.

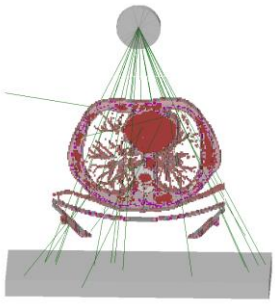

Fig. 2: Geant4 simulation of X-rays passing through a 3D torso sourced from CT.

3. References

[1] T. Rahman *et al*., "Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization," 2020, doi: 10.1109/ACCESS.2020.3031384
[2] Showkatian E, Salehi M, Ghaffari H, Reiazi R, Sadighi N. Deep learning-based automatic detection of tuberculosis disease in chest X-ray images. Pol J Radiol. 2022, doi: 10.5114/pjr.2022.113435. PMID: 35280947; PMCID: PMC8906182
[3] Rajesh, T. Babu, R. R. Nair and P. Pechetti, "Detection of Tuberculosis using a Multi-model Classification Approach on CXR Images," 2022, doi: 10.1109/AIDE57180.2022.10060675
[4] D. Capellán-Martín, J. J. Gómez-Valverde, D. Bermejo-Peláez and M. J. Ledesma-Carbayo, "A Lightweight, Rapid and Efficient Deep Convolutional Network for Chest X-Ray Tuberculosis Detection," 2023, doi: 10.1109/ISBI53787.2023.10230500
[5] Ravimohan, S., Kornfeld, H., Weissman, D., & Bisson, G. P. (2018). Tuberculosis and lung damage: from epidemiology to pathophysiology. https://doi.org/10.1183/16000617.0077-2017