MA Thesis: Weakly Supervised Federated Learning for Chest X-ray Imaging

Supervisors: Shadi Albarqouni
Albarqouni Lab., Helmholtz AI, Helmholtz Center Munich
AI in Medicine, Technical University of Munich

Abstract.
Detection, localization, and classification of pathologies and findings in Chest X-rays (CXR) is an essential step in the clinical workflow. Automatic data-driven deep learning models have shown promising results with an average radiologist level performance [7] paving the path to mitigate the major challenges in radiology of staff shortage and heavy workloads. In this work, we investigate i) whether leveraging many databases, in a privacy-preserved fashion [10-18], would improve the classification performance for long-tail distributions (highly unbalanced classes), ii) providing more interpretable decisions by localizing and segmenting pathologies without pixel-wise annotations [2].

Databases:
- AI for COVID Database (https://aiforcovid.radiomica.it/) [1] ~850 cases
- CheXpert Data (https://stanfordmlgroup.github.io/competitions/chexpert/) ~240,000 cases
- NIH CXR (https://www.kaggle.com/nih-chest-xrays/data) ~112,000 cases
- OpenI (https://openi.nlm.nih.gov/faq#collection) ~7,470 cases
- RSNA Kaggle (https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/overview) ~30,000 cases
- MIMIC-CXR (https://physionet.org/content/mimic-cxr/2.0.0/) ~250,000 cases
- Tuberculosis (TB) datasets

Roadmap:
- Familiarize yourself with the current literature on:

If you are interested in this project, please write an email to shadi.albarqouni@tum.de
○ Deep Learning with Chest X-rays [7-9, 19-21] → Identify the challenges and research gaps in this area.
○ Federated Learning with non-iid [10,12,13] → understands the non-iid challenges in federated learning.
○ Federated Learning with Medical Imaging [14 - 18] → how FL is applied in medical imaging context.

● Measure the A-distance [4] or Earth Mover's distance [5] between different databases → This would tell if we have any domain shift and non-iid.
● Build baseline models; local models (trained on each database, individually), global model (trained with all databases).
● Develop the proposed method with the help of Weakly-supervised Learning [3]; Multiple Instance Learning [2], and/or Cross-Scale Graph Neural Networks [6].
● Run extensive experiments and analysis
● Write up your thesis

Research Questions:
Q1) Can we train a weakly-supervised federated classifier to jointly classify and localize pathologies in Chest X-ray Imaging?
Q2) Would the model be able to detect any unseen pathologies (out of the distribution)?

Requirements:
- Solid background in Machine/Deep Learning
- Familiar with discriminative deep learning models and SOTA architectures
- Sufficient knowledge of Python programming language and libraries (Scikit-learn)
- Experience with a mainstream deep learning framework such as PyTorch.
- Machine/Deep learning hands-on experience with MONAI framework

What we offer:
- A guest contract at Helmholtz Center Munich
- Access to computational resources

References:

If you are interested in this project, please write an email to shadi.albargouni@tum.de
If you are interested in this project, please write an email to shadi.albarqouni@tum.de