Digital Health Center - Machine Learning

Chair: Prof. Dr. Christoph Lippert

Contact:

Jana Fehr (Jana.fehr@hpi.de),

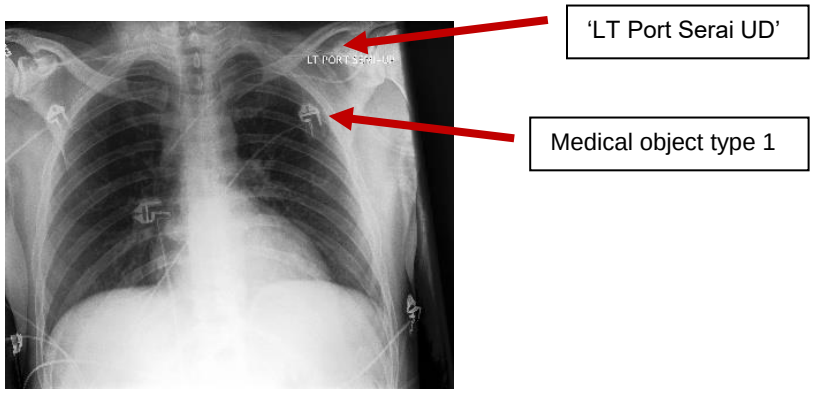
Benjamin Bergner (benjamin.bergner@hpi.de)

Master thesis topic

Automatic annotation of medical objects and text features in chest x-rays

Background:

Deep learning algorithms are able to detect abnormalities in medical images, such as chest x-rays when they are trained on large datasets. Several large open-source databases for chest x-rays have been created, namely ChestXray14, PadChest, CheXpert, to push the development of lung abnormality detection algorithms. One challenge is that chest x-ray images often contain clinical objects (i.e. pace-maker, chest-drains, monitoring cables) or annotated text from the scanner that might bias the algorithmic performance. One study found out that (Oakden-Rayner, 2020) a deep learning algorithm on the ChestXray14 dataset was more likely to identify a pneumothorax when the image contained chest drains (a treatment administered after diagnosis of pneumothorax). On images that did not contain chest drains on the other hand, the algorithm often resulted in false negative predictions. To be valuable for diagnosis, the algorithm should be able to identify pneumothorax without chest drains. Unfortunately, most large medical image datasets do not contain annotations on visible text or medical objects and data scientists therefore are unaware of such biases in their model and not able to test for such. To tackle this problem, we want to develop a tool that can automatically annotate and locate medical objects and text in in chest x-rays. Such an annotation tool can be applied to annotate new datasets and enable data-scientists to test for algorithmic bias on those features.



Task:

The focus of this project will be on Chest X-rays with multiple labels on chest abnormalities. The scope of this project can either focus more on the application or method development, depending on interest and skill-level.

Applied focus:

- Applying a text-detection (i.e. tesseract or EAST) tool to extract text labels from images
- Developing an algorithm to detect medical objects in chest x-rays. For this different approaches are possible: 1) Supervised 2) Semi-supervised
- Applying these tools on different public datasets (ChestXray14, MIMIC_III, PadChest)
- Identifying associations between identified medical objects and pathology labels

Methodological focus:

- Implementing explainability approaches i.e. saliency maps to investigate whether medical objects and text labels are correctly identified.
- Building a classifier to predict lung abnormalities in chest x-rays
  And subsequently implement explainability approaches (i.e. saliency maps) to investigate if the classification algorithm harnesses medical objects or texts to predict abnormality classes.

Requirements:

- Python skills
- Minimum one university lecture on Deep Learning passed
- Experience with Pytorch or Keras
- Github/Gitlab
- Critical thinking
- Bringing own ideas
- Working independently between regular meetings

If you are interested or if you have questions, send us an e-Mail via the address indicated above.

Related Literature

Oakden-Rayner L. Exploring Large-scale Public Medical Image Datasets. Acad Radiol. 2020;**27**:106–12. https://doi.org/10.1016/j.acra.2019.10.006

Badgeley MA, Zech JR, Oakden-Rayner L, Glicksberg BS, Liu M, Gale W, et al. Deep learning predicts hip fracture using confounding patient and healthcare variables. npj Digit Med. 2019;**2**.

Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLoS Med. 2018;**15**:1–17.