
MASTER THESIS: **Generation of Synthetic Medical Records**

Limited data access and small sample sizes in medical experiments is a longstanding barrier to data-driven research in the domain of healthcare. One reason behind limited access stem from the fact that most of the medical records of patients composed of personal identifiers, which in combination with potential sensitive medical information, induces privacy concerns.

Therefore, data cannot be shared in most cases among researchers to boost *computational health*. Furthermore, small sample sizes of medical experiments make it difficult to apply most of the machine learning methods. Consequently, generating synthetic data from medical records has the potential to cope with both challenges. Generating synthetic data has been proven to be efficient especially for image data, but for time-series and sequential data it is a more challenging task.

Generative adversarial networks (GAN) have been mainly used for generating synthetic data in the domain of computer vision and some advancements have been made also for other types of data. Thus, this thesis [medGAN](#) and [DoppelGANger](#) (or appropriate other methods) should be examined according to their applicability to generate synthetic data form given medical records from patients suffering from dementia.

General overview of the task:

1. Frame the problem: develop a research question you want to answer
2. Research related work of synthetic data generation of medical records
3. Implement different versions of algorithms to generate synthetic data to a given data set of medical records
4. Analyze the results and apply machine learning methods to predict the medical state of patients based on the synthetic data

Student profile:

- BA in Computer Science, Software Engineering or related
- Strong knowledge of machine learning
- Desirable: experience with GANs

How to apply: please sent a CV and Transcript of Records via email

Contact: Jan-Niklas Voigt-Antons (jan-niklas.voigt-antons@tu-berlin.de) and Vera Schmitt (vera.schmitt@tu-berlin.de)