Digital Health
Personalized Medicine
Hasso-Plattner-Institut · Universität Potsdam

HPI Digital Health

Research Group of Prof. Dr. Erwin Böttinger

Machine Learning on Real-World Health Data with Cloud-based In-Memory Database Computing

Motivation

Electronic health records (EHR) data holds considerable promises for new research insights, improvement of healthcare delivery and more personalised patient treatment.
The Mount Sinai Health System (MSHS) is a modern academic health system serving eight million patients with longitudinal EHR data and the Mount Sinai data warehouse captures real-world healthcare data for more than 10 years. Among the 8 million patients, 50.000 are consented in the BioMe biobank program and have whole genome and/or exome sequence data. Hasso Plattner Institute for Digital Health at Mount Sinai (HPI-MS) was established in March 2019 to connect digital health and engineering competencies at HPI with a fully-digital real-world healthcare and health research enterprise.

Healthcare and research data at MSHS are being migrated to a cloud computing service that is compliant with privacy protection regulation and highly-performant. In-memory database technology (IMDB) is widely recognized for outstanding performance in supporting demanding business applications. However, utility and performance of IMDB technology in cloud implementations of real-world healthcare and health research data supporting deep learning use case application needs to be investigated.

Project Goals

The students will apply machine learning in use cases focusing on real-world applications in heart disease, mental health, or back pain at HPI-MS. Use cases will serve to evaluate performance and utility of IMDB technology in Machine Learning implementations in healthcare and research. Deep learning patient representations may include multi-modal health data, including but not limited to longitudinal EHR clinical data, genomic, imaging, or wearables data.

Envisioned Use Cases:

Hypertension

Our goal is to predict the onset of complications in hypertensive patients—such as stroke and ischemic heart disease—using advanced machine/deep learning methods. Hypertension is one of the most prevalent chronic diseases worldwide. It is an independent predisposing factor for heart failure, coronary artery disease, stroke and renal disease. It is the most important risk factor for cardiovascular morbidity and mortality in industrialized countries.

Mental Health

In current clinical handbooks, mental disorders are described mainly based on symptoms. However, many patients with different disorders share the same biological underpinnings and patients of the same diagnostic category react very differently to the same treatment. The goal of this subproject is to use unsupervised machine learning methods to find categories that are closer to underlying biological or neurological mechanisms as well as subcategories that could inform treatment decisions.

Back Pain

Back pain affects a high percentage of the population and accounts for the majority of sick leave days. This subproject investigates which comorbidities are typically present when patients enter the healthcare system with back pain symptoms, what the short- to medium-term risk of chronification/reappearance of the symptoms is and whether back pain patient populations can be further subclassified with help of multi-modal data.

General Subtasks for each use case would include:

1) Identification of the complications and their onset for the different patient cohorts and saving them as target variables.
2) Extraction of features from the data that will be of interest to predict these complications.
3) Thorough investigation of different models to see which ones perform better to predict these complications.

What you should bring with you

To carry out this project successfully, you will need solid foundations in:

1) Database principles, including normalisation, modelling notation and SQL
2) Basic programming skills (Python scripting)
3) Deep interest in learning more about data in the clinical domain
4) Fundamentals of machine learning (supervised learning, performance evaluation of predictive models)
5) Data visualisation principles

Contact

Get in touch for questions and ideas. We are located at the Digital Health Center on Campus III, Building G2-2, Rudolf-Breitscheid-Str. 187, 14482 Potsdam.

Suparno Datta, M.Sc.
E: suparno.datta@hpi.de
T: 0331/5509-4817

Ariane Morassi Sasso, M.Sc.
E: ariane.morassi-sasso@hpi.de
T: 0331/ 5509-4827

Hanna Drimalla, M.Sc.
E: hanna.drimalla@hpi.de
T: 0331/5509-4829

Dr. Claudia Schurmann
E: claudia.schurmann@hpi.de
T: 0331/5509-4845

Jan Philipp Sachs, MD, M.Sc.
E: jan-philipp.sachs@hpi.de
T: 0331/ 5509-4815

Prof. Dr. Erwin Böttinger
E: erwin.boettinger@hpi.de
T: 0331/5509-163

Ricardo Miotto, PhD
E: ricardo.miotto@mssm.edu

Girish Nadkarni, MD
E: girish.nadkarni@mountsinai.org