

EHR-Driven Label Extraction for AI in Medical Imaging.

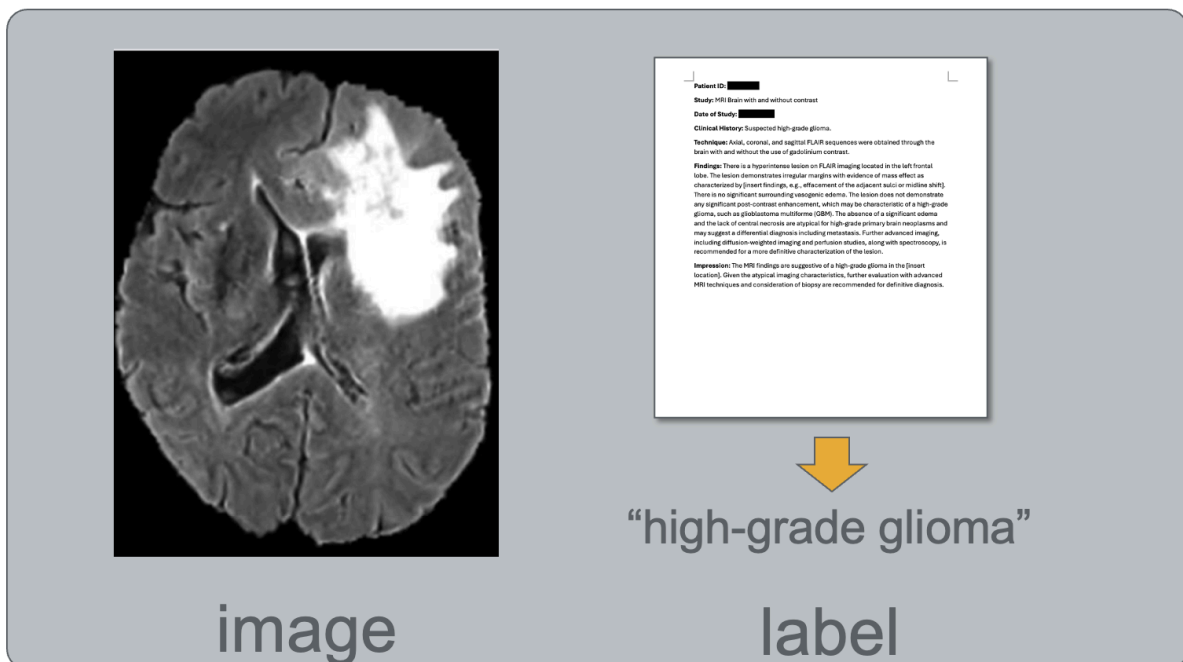
Master Project at the Chair for Digital Health - Machine Learning

Project Motivation:

HPI·Mount Sinai is leading the development of a health data platform named “AI-Ready Mount Sinai” (AIR.MS). This platform links Electronic Health Record (EHR) data from over ten million patients in the Mount Sinai Health System. The data in AIR.MS is linked to the Mount Sinai Imaging Research Warehouse (MS-IRW), developed by the BioMedical Engineering and Imaging Institute (BMEII), which provides millions of de-identified medical images. While we have access to one of the largest clinical data sets worldwide, the bottleneck in advancing medical AI lies in the scarcity of high-quality image annotations. It demands a significant investment of time from seasoned medical professionals and incurs substantial costs. This project aims to address this critical challenge by innovating the label extraction process using AI.

Project Goals:

In this project we aim to leverage text mining, Large Language Models (LLMs) and Vision Language Models (VLMs) to automatically extract image labels from radiologic reports linked to the images, as well as the patient’s Electronic Health Records (EHRs) and Doctors’ notes at health-system scale and use them for the training of supervised learning algorithms. The goal is to create a data set of medical images with labels for a range of supervised learning tasks, including classification, regression, and image captioning and leverage this data set for the training and evaluation of deep neural networks that predict the labels from an image, and generates radiology reports for an image.



Who Can Participate?

Master students from the programs of: Data Engineering, Digital Health, IT-Systems Engineering, Software Systems Engineering

Who Should Participate?

Participants who will find this project especially rewarding include those who:

- *Seek Real-World Impact in the Medical Domain:*

Those who are curious about the intersection of information technology and healthcare will gain hands-on experience in a project that has the potential to impact patient care.

- *Thrive in interdisciplinary Collaborative Settings:*

Collaboration is at the heart of this project. If you work well in teams and are looking to collaborate with international experts in AI and medicine, we welcome you.

- *Come from a Computer Science:*

Students with skills in programming, data analysis, machine learning, deep learning or text mining will find their knowledge directly applicable.

- *Have a Background in the Medical Domain:*

Students with medical knowledge will bring vital domain expertise, bridging the gap between raw data and meaningful clinical interpretation.

Meet the Supervisors:

Benjamin Bergner, PhD Student at HPI (benjamin.bergner@hpi.de) focuses his research on several areas in machine learning for medical imaging and text. His interests include exploring attention mechanisms in deep learning, multiple instance learning, and the interpretability of machine learning models. He is also engaged in research on continual learning and self-supervised learning, with a specific focus on applications in medical imaging and computer vision.

Valentin Fauveau, Senior Programmer Analyst at the BMEII, Mt Sinai (valentin.fauveau@mssm.edu)

I am a senior Programmer Analyst at the Biomedical Engineering and Imaging Institute (BMEII) at Mount Sinai, specializing in Data Science and Biomedical Engineering. With a strong background in Image and Signal Processing, I have extensive experience developing and deploying AI models for the analysis of biomedical data. My role at BMEII involves providing programmatic tools and AI support to various research teams. I am deeply enthusiastic about emerging medical technologies, including medical wearables, VR (virtual reality) treatments, low field MRI acquisitions, and novel MRI sequences designed to enhance patient health assessment.

Christoph Lippert, Professor at HPI, Adjunct Professor at Mt Sinai

(christoph.lippert@hpi.de) has a multifaceted research focus that integrates machine learning, artificial intelligence, medicine, and genomics, focusing on integrating these fields to advance the capabilities of personalized medicine and health risk prediction.

Sumit Shekhar, Postdoc at HPI (sumit.shekhar@hpi.de) has research interests that encompass a diverse range of topics within the field of computer graphics, computer vision,

image and video processing, and interactive systems. His work includes developing interactive tools for media manipulation, enhancing the quality of images and videos, particularly in challenging conditions, and creating systems for collaborative image editing. Additionally, he has an interest in 3D imaging technologies, especially in photo stylization and rendering on mobile devices, as well as the stylization and dynamic representation in live photos.