SDU Robotics

PhD project - Abstract

Natural language processing for information extraction in electronic health records

PhD Jannik Skyttegaard Pedersen jasp@mmmi.sdu.dk
SDU Robotics, The Maersk Mc- Kinney Moller Institute,
University of Southern Denmark

Natural language processing (NLP) is a subfield of artificial intelligence that enables machines to analyze, understand, and manipulate natural language. In the clinical domain, NLP algorithms have the potential to improve patient care, e.g. by extracting information from the unstructured text of medical records. The field of NLP has developed rapidly over the last decade because of the progress in deep learning and more recently by exploiting the Transformer architecture to train big language models. In deep learning, models are trained using a vast amount of labeled data. Collecting these huge datasets are not always feasible because it is time consuming and resource demanding, particularly for specialized domains such as Danish medical records.

This PhD project will explore how medical NLP models can be developed without depending on vast amounts of labeled data. Moreover, an NLP algorithm that can detect bleeding events in the unstructured text of Danish medical records is developed. This has the potential to guide anticoagulatory treatment and increase early detection of severe diseases, and thereby improve patient care.

As a secondary objective, Danish high-quality word embeddings and language models for the medical domain are developed.