



Help lifting the treasure chest: make clever use of unstructured reports from nuclear cardiac imaging exams

Project Management and Software Development
for Medical Applications

General Info

Contact Person: Prof. Nekolla, Stephan G.
Solari, Esteban Lucas

Contact Email: stephan.nekolla@tum.de
elucas.solari@tum.de

Project Abstract

We want to develop an NLP (natural language processing) pipeline to retrieve clinical information from SPECT nuclear cardiology reports. It includes a study of possible NLP alternatives (BERT-, NER-, GPT-based); a workflow to curate a database, train and test the models; and a Python implementation on an existing report database.

Background and Motivation

Myocardial perfusion imaging SPECT is one of the most important imaging diagnostic technologies in coronary heart disease, allowing physicians to assess the functionality of the cardiac tissue. Thousands of SPECT studies are performed yearly in our Clinic and there is an increasing interest in retrospective studies with these images. This implies the technical challenge of analyzing thousands of clinical reports and connecting them with imaging data.

One of our previous projects implemented a NER model to **extract clinical parameters** from a **multitude of cardiac SPECT reports**. This approach seems to be sufficient for a small set of parameters, but it requires a thorough rule-based approach for new parameters. New technologies have been developed since, most prominently

those based on pre-trained transformers, to more powerfully perform NLP tasks. Examples include BERT- and GPT-based models.

The goal of this project is to study how to **extract clinical parameters** and **classify cardiac reports** according to their diagnosis. The particular objectives include the selection of a suitable model, the development of a workflow to automate the process of extraction and classification, and a first implementation of such a tool in a previously curated and anonymized dataset for a known clinical problem.

The student will work elbow to elbow with a Masters student working on a related project and will be closely assisted by our team at Nuklearmedizin Clinical Research (Prof. Nekolla's team).

Student's Tasks Description

What's expected:

- a short bibliographic study of the most common NLP methods for parameter extraction and classification;
- a design of a tool to extract clinical parameters and classify cardiac SPECT reports, including a workflow for its implementation, training and testing;
- a Python implementation for a particular set of clinical parameters in a ~2000-report dataset.

What they learn:

Please send the completed proposal to tianyu.song@tum.de, shervin.dehghani@tum.de and felix.tristram@tum.de. Please note that this proposal will be evaluated by the BMC coordinators and will be assigned to a student only in case of acceptance.



- to think on a programming project in a comprehensive way (bibliographic study, design, development, testing)
- to learn to combine a deep understanding of a real-life clinical problem with programming tools and NLP implementations;
- to work in a scientific and collaborative environment, where working in a team should make the tasks more efficient and achievable (learn to discuss strategies, take input, ask for support, communicate –intermediate– findings)

Technical Prerequisites

- experience in Python
- interest and some experience in NLP

References

Eyre H, Chapman AB, Peterson KS, et al. “Launching into clinical space with medspaCy: a new clinical text processing toolkit in Python”. AMIA Annu. Symp. Proc. 2021: <http://arxiv.org/abs/2106.07799>

Rasmy, L., Xiang, Y., Xie, Z. et al. „Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction”. npj Digit. Med. 4, 86 (2021). <https://doi.org/10.1038/s41746-021-00455-y>

Bressem KK, Papaioannou JM, Grundmann P, et al. „MEDBERT.de: A Comprehensive German BERT Model for the Medical Domain”. arXiv Mar 2023. <https://doi.org/10.48550/arXiv.2303.08179>

Please send the completed proposal to tianyu.song@tum.de, shervin.dehghani@tum.de and felix.tristram@tum.de. Please note that this proposal will be evaluated by the BMC coordinators and will be assigned to a student only in case of acceptance.