Data and Process Visualization Tool for a Machine Learning Framework for ARDS Classification

(Bachelor's Thesis)

ALEXANDER BARON

Motivation

During the SMITH project, the Embedded Software (Informatik i11) chair accumulated many different medical datasets to train machine learning models on the classification of Acute Respiratory Distress Syndrome (ARDS) – a severe inflammatory condition that occurs in Intensive Care Unit patients. ARDS has a high mortality rate of around 40%, which is partially due to the fact that it is often diagnosed too late. Here, AI models have the potential to support doctors in diagnosis, which was researched as part of the ASIC use-case. Since no unified infrastructure for the training and evaluation of these models exists, their comparability is limited. In particular, comparing the data sets used and the resulting performance involves a great deal of manual effort.

State of the art

Many machine learning models for the classification of ARDS have been proposed and described in the literature. Most of them use their own framework for data processing and training the models, which leads to a low comparability. In particular, the lack of data visualization makes the analysis and comparison of AI models a challenge. To increase comparability, a machine learning framework for the development of AI models for ARDS detection was implemented in previous work. This framework, however, does not yet have any visualization capabilities. There are already various approaches for the visualization of patient data, such as the Diagnostic Expert Advisor (DEA), which can be used to view individual patients and parameters. However, a patient group-wide data visualization for the analysis of trained AI models for the ARDS classification does not yet exist.

Objective

The main goal will be to design and develop a graphical user interface that makes it possible to visualize the datasets used in the MLP Framework and allow for an easier comparison of the models results. Thus, improving the understanding of the datasets used and allowing for a better analysis of the trained models. It will also support visualizing the data processing steps, such as patient filtering and parameter imputation. The tool is intended to be useable by ML experienced users as well as medical professionals. To ensure the usability, a medical professional experienced in classifying ARDS will be consulted.

Planned procedure

Currently, further exploration of literature as well as existing libraries for both data visualization and visual configuration is required. Meanwhile, an analysis of requirements will be performed, factoring in both the perspective of a computer scientist and a medical professional. Using these results, the next step will be to design a system that fulfills said requirements. Subsequently, the design and development of the system will begin, keeping in touch with the stakeholders, and requesting feedback on intermediate versions. Finally, the results will be evaluated, and presented in the thesis.

Betreuer: Simon Fonck
15.05.2024

Informatik 11
Embedded Software
RWTH AACHEN UNIVERSITY