An approach to explainable artificial intelligence in the context of medical care for ARDS patients

(Master's thesis)

ALINA IBACH

Motivation

This Master's thesis takes place within the SMITH project (Smart Medical Technology for Healthcare). A part of the project is the use case ASIC (Algorithmic Surveillance of Intensive Care). One of its goals is to incorporate AI methods into patient care for acute respiratory distress syndrome (ARDS) patients. While AI methods often show great results, the number of AI methods implemented in the clinical reality is low. Without the ability to interpret the results correctly, medical staff does not have a benefit from AI algorithms. One possible approach to close the gap between research and healthcare lies in the field of explainable AI (XAI) - especially, with explanation approaches that keep the user in mind.

State of the art

At the chair Informatik 11, two methods for the detection of ARDS were developed - a Bayesian network and a set of random forest classifiers. Additionally, a convolutional neural network was implemented for the automatic detection of bilateral infiltrates in chest X-Rays, which is an important criterion for ARDS.

In the literature, there is no consensus of what is needed or what characteristics have to be fulfilled for an XAI method to be considered explainable. These questions arise even for AI methods that are considered transparent, e.g. linear regression or Bayesian models. The opposite of transparent methods are black-box as artificial neural networks.

There is a publication that uses, among other methods, random forest to predict the duration of ventilation of patients with ARDS and then uses feature importance methods, LIME, SHAP and DALEX as XAI approaches to display the results. In addition to purely mathematical explanations, approaches that have users' needs in mind, e.g. trustworthiness and fairness, have become more popular.

To our knowledge, there exists no explicit XAI approach for the classification of ARDS patients.

Objective

The goal of this thesis is to create an XAI approach for the classification of ARDS patients from which the user, i.e. medical staff, benefits. For this purpose an assement of existing concepts of XAI is needed. The emphasis lies on XAI approaches that have the potential to be valuable to the user. Some appropriate approaches are chosen that can be applied to the existing AI methods within the scope of the project. A user study is conducted to verify whether the approach meets the needs of the user.

Procedure

Since the idea of focusing explanations on the users' needs is relatively new, the first step is to research the literature from this point of view. The most promising XAI approaches are selected and if needed adapted. At least one of those approaches is implemented and applied to the exiting AI approach of this chair. The results are used to create a small internal user study within the chair. If promising, a small user study is conducted with medical personal.

Betreuer: Simon Fonck
13.12.2022