Feature selection and data preprocessing methods in the context of AI-based ARDS classification in time-series data

(Bachelor Thesis)

SINA WEDDING

Motivation

Acute Respiratory Distress Syndrome (ARDS) is a severe inflammatory condition with high mortality that is often diagnosed too late or not at all. Artificial Intelligence (AI) tools offer the potential of detecting signs of ARDS and alerting medical personnel to assist physicians and enhance the detection rate. Medical data are often collected over a longer period and with many different parameters, resulting in enormous amounts of data. Certain recorded parameters are not important for ARDS classification or appear redundantly in the data base and thus reduce the performance of ARDS classifiers. Therefore, it is crucial to find the most relevant parameters for predicting ARDS. Feature selection algorithms are well applicable for this problem. Furthermore, the data can also be prepared for the AI using other preprocessing techniques like data sampling to reduce complexity or remove noisy data.

State of the art

Preprocessing and feature selection methods are already used in various medical contexts, including cancer detection and analysis of DNA microarray data. While single feature selection and extraction methods have been applied to ARDS, the focus has been mostly on the evaluation of the classifiers that identify ARDS rather than on the feature selection algorithms. Similarly, different preprocessing methods are often only used while developing classifiers without being analyzed further. The main usage of feature selection methods in the context of ARDS is categorizing ARDS phenotypes and detecting ARDS. Even though some techniques have already been used, there still is a significant gap in research when it comes to comparing the effectiveness of different techniques, especially in analyzing time-series data.

Objective

This bachelor thesis aims to compare different approaches to prepare time-series data for ARDS detection using AI tools, evaluating their impact on effectiveness and performance. There will be a special focus on feature selection methods, as they are already proven to be effective in other medical cases. Additionally, other preprocessing methods including over- and undersampling to address unbalanced data and various data sampling methods for reducing the amount of data used will be assessed. Different already implemented AI tools and several data sets will be used for evaluating the models' results.

Procedure

The first step in the bachelor thesis will be a research on the topics of ARDS and feature selection and preprocessing methods in medical applications. Afterwards, the most promising methods will be chosen and integrated in the already existing framework. In the end, the methods will be run on datasets while comparing and evaluating the performance of the different models.