

Retrospektive Erkennung von ARDS mit Hilfe von Random-Forest Algorithmen

(Bachelorarbeit)



HANNES PIEPER

Motivation

Diese Bachelorarbeit findet im Kontext des SMITH-Projektes im Use-Case ASIC statt. Dieser Use-Case befasst sich mit der Verbesserung der Diagnose und Behandlung von ARDS (Acute Respiratory Distress Syndrome). ARDS hat eine hohe Mortalitätsrate, welche durch die häufig zu späte oder in einigen Fällen ausbleibende Diagnose begründet ist. Durch die fehlende Diagnose werden diese Fälle auch in den, der Forschung zur Verfügung stehenden Datenbanken, nicht richtig kategorisiert. Um die weitere Erforschung von ARDS zu unterstützen, ist eine hohe Qualität des Labelings fundamental. Daher ist es wichtig, eine retrospektive Klassifizierung zu ermöglichen. Dafür bieten sich Random-Forest Algorithmen (RFA) an, da sich diese besonders für Daten mit vielen unterschiedlichen Attributen eignen und ihre Entscheidungen nachvollziehbar bleiben.

Stand der Technik

RFA wurden im Rahmen von mehreren wissenschaftlichen Arbeiten erfolgreich zur Unterstützung der Diagnose von Brustkrebs und Herz-Rhythmus Störungen durch die Analyse von elektronischen Patientendaten eingesetzt. Dies zeigt die generelle Eignung dieser Technik bei der Erkennung von diversen Krankheiten. Darüber hinaus werden bereits verwendete Machine Learning Ansätze wie Gradient Tree Boosting zur erfolgreichen Erkennung von ARDS analysiert. Darüber hinaus gibt es am Lehrstuhl Arbeiten zur Klassifikation mit Hilfe eines Bayes'schen Netzes sowie Deep Learning Modelle zur Erkennung von ARDS Symptomen in Röntgenbildern. Da bisher noch keine Implementierung von RFA für die Klassifizierung von ARDS entwickelt wurden, soll dieser Ansatz auf seine Anwendbarkeit überprüft werden.

Zielsetzung

Ziel dieser Bachelorarbeit ist die Implementierung von RFA zur retrospektiven Klassifizierung von intensivmedizinischen Patientendaten im Hinblick auf das Vorliegen von ARDS. Die RFA werden dabei mithilfe von Teilen von mehreren im SMITH-Projekt vorhandenen Datenbanken trainiert und anschließend auf einem anderen Teil evaluiert und validiert. Außerdem werden sie mit bereits bestehenden Techniken verglichen. Die entwickelten Algorithmen sollen dabei helfen, die ARDS-Diagnose von Patienten in den Datenbanken des SMITH-Projektes zu verifizieren.

Geplante Vorgehensweise

Als erster Arbeitsschritt erfolgt eine Literaturrecherche zu den Themen: ARDS, Verwendung von RFA, Feature-Selection und dem Aufbau der Datenbanken des SMITH-Projektes. Nach einer Vorbearbeitung der Daten erfolgt die Auswahl der relevanten Parameter mithilfe unterschiedlicher Feature Selection Ansätze. Danach werden die RF-Algorithmen trainiert und mit Hilfe von k-fold cross validation validiert und evaluiert.