Informatik 11
Embedded Software

RWTH AACHEN UNIVERSITY

Masterarbeit

Entwicklung eines Frameworks für post-hoc-Erklärbarkeit bei der Fehlererkennung in intensivmedizinischen Daten

Hintergrund

Im Rahmen der Projekte "SMITH - **S**mart **M**edical **I**nformation **T**echnology for **H**ealthcare" und "AIx-Neo-Guard" forscht der Lehrstuhl vermehrt an Algorithmen für die Erkennung und Klassifizierung von Komplikationen bei intensivmedizinischen Patienten.

Im Austausch mit medizinischem Personal und in Umfragen besteht hierbei immer wieder der Bedarf, die Ergebnisse der Algorithmen besser zu verstehen und Ärzten dadurch notwendige Informationen für den Entscheidungsprozess zukommen zu lassen.
Um dieses Ziel zu erreichen dürfen die Algorithmen nicht als „Black-Box“ agieren. Stattdessen sollen ihre Stärken und Schwächen klar sein und die Entscheidungen die zum Ergebnis führen ersichtlich („transparency“) bzw. verständlich („explainability“) sein.
Zur Entwicklung entsprechender Technologien stehen im Projekt verschiedene Datenbanken mit Patienteninformationen, Vitalparametern und Laborwerten zur Verfügung.
Um die Datenbanken für Erkennungs- und Klassifikationsverfahren nutzen zu können, muss zunächst eine ausreichende Datenqualität und Datendichte hergestellt werden. Hierfür wurde am Lehrstuhl ein Analysesystem entwickelt, welches ermöglicht Algorithmen für die Erkennung von fehlerhaft aufgezeichneten Daten und für die Datenimputation auf medizinischen Daten auszuführen.

Um diese beiden Aspekte zu kombinieren sollen solche Transparency und Explainability Methoden für Fehlererkennungsalgorithmen entwickelt werden, damit diese in Zukunft auch für die Erkennung und Klassifizierung adaptiert werden können.

Aufgabenstellung

In dieser Abschlussarbeit soll ein Framework für die Anwendung von post-hoc Erklärbarkeitsmethoden für die Fehlererkennung in intensivmedizinischen Daten entwickelt werden. Das Framework muss hierfür keine GUI besitzen, hierfür genügt die Ausführbarkeit in der Kommandozeile.

Dafür sind folgende Arbeitspunkte notwendig:

- Literaturrecherche zu post-hoc Erklärbarkeitsmethoden im Machine Learning
- Entwickeln eines Frameworks, welches ermöglicht auf bereits vorhandenen Fehlererkennungsverfahren post-hoc Erklärbarkeitsverfahren wie Gegenbeispiele, Saliency-Masken, Regel-Approximierung etc. auszuführen. Hierbei sollen möglichst wenige Annahmen über die Fehlererkennungsverfahren getroffen werden.
- Auswertung der Erklärbarkeitsverfahren mit zwei bereits vorhandenen Fehlererkennungs-Methoden.

Vorkenntnisse

Diese Arbeit richtet sich insbesondere an Studierende aus den Informatik- und Data Science-Studiengängen. Medizin als Anwendungsfach sowie Erfahrung in Machine Learning Verfahren sind wünschenswert, aber nicht notwendig.

Ansprechpartner

Alexander Kruschewky, M. Sc. RWTH kruschewsky@embedded.rwth-aachen.de