Erklärbarkeitsanalyse von Deep-Learning Modellen zur Erkennung von bilateralen Infiltraten in Röntgenbildern

(Bachelorarbeit)

NICK STETE

Motivation

Das Acute Respiratory Distress Syndrom (ARDS) ist eine schwere Lungenkrankheit, die sich durch eine hohe Mortalität auszeichnet. Ein Grund hierfür liegt in der häufig verspäteten oder ausbleibenden Erkennung der Krankheit. Gemäß der Berlin-Definition stellen bilaterale Infiltrate in Röntgenaufnahmen ein zentrales Diagnosekriterium dar. Deshalb wurden am Lehrstuhl Im Rahmen des Use Cases ASIC der SMITH-Initiative des BMBF neuronale Netze zur Erkennung von ARDS anhand von Röntgenaufnahmen entwickelt. Neuronale Netze werden als Black-Box Modelle angesehen, deren einzelne Entscheidungen nicht nachvollzogen werden können. Im Bereich der Medizin sind Erklärungen aus regulatorischer Sicht jedoch zwingend erforderlich. So schreibt die europäische GDPR Patienten ein Recht auf aussagekräftige Informationen über die genutzten Logiken zu, wenn automatisierte Entscheidungsmodelle genutzt werden.

Stand der Technik

Fehlende Erklärbarkeit ist ein Problem in allen sicherheits- und gesundheitskritischen Domänen. Aus diesem Grund wurden im Rahmen des Forschungsfeldes der erklärbaren künstlichen Intelligenz (XAI) eine Vielzahl an Methoden entwickelt, um die Entscheidungen von Modellen mit maschinellen Lernverfahren erklären zu können. Für Black-Box Modelle wie neuronale Netze werden vor allem Post-Hoc Verfahren genutzt, welche das Modell nach Abschluss des Lernverfahrens untersuchen. Für Erklärungen im Computer Vision Bereich werden in aller Regel Heat/Attention-Maps eingesetzt, welche die Bildregionen mit dem stärksten Einfluss auf die Entscheidung hervorheben. Domänenexperten betrachten diese Erklärungen jedoch oftmals als unzureichend, sodass hier weitere Erklärungen umgesetzt werden müssen.

Zielsetzung

Es sollen geeignete XAI-Methoden zur Erweiterung der bestehenden neuronalen Netze evaluiert und im Rahmen des Use-Cases implementiert werden. Da es sich hierbei sowohl um Convolutional Neuronal Networks (CNN), als auch um Vision Transformer (VIT) handelt, wird zu untersuchen sein, welche Methoden sich für beide Architekturen nutzen lassen und inwiefern man unterschiedliche Verfahren vergleichen kann. Die Ergebnisse sollen in ein am Lehrstuhl parallel entwickelten Framework zur ARDS-Erkennung integriert werden. Eine modulare Entwicklung ist deshalb erforderlich.

Geplante Vorgehensweise

Zu Beginn erfolgt eine Einarbeitung in die Themen ARDS, Deep Learning und XAI. Im Anschluss werden durch eine Literaturrecherche geeignete Methoden für die neuronalen Netze zur ARDS-Erkennung gesucht und analysiert. Diese werden im Anschluss implementiert und evaluiert. Nach Möglichkeit wird die Bewertung durch medizinisches Fachpersonal vorgenommen. Die Vorgehensweise und Ergebnisse werden in einer schriftlichen Ausarbeitung zusammengefasst.

Betreuer: Simon Fonck
15.11.2023

Informatik 11
Embedded Software
RWTH AACHEN UNIVERSITY