# Explainable AI (XAI) zur Klassifikation von EKG-Signalen


Diplomand

**Dmitry Grigoriev**


Einleitung: Im Rahmen eines Vorgängerprojekts wurde die Anwendbarkeit von Explainable AI (XAI) zur Erläuterung der Vorhersagen eines Convolutional Neural Networks (CNN) geprüft. Dieses CNN diente zur Klassifizierung von Elektrokardiogrammen (EKGs) zwecks Identifikation verschiedener Signalabweichungen. Ziel war es, einen behandelnden Arzt bei seinen Entscheidungen zu unterstützen, indem relevante Abschnitte im EKG-Signal, die für die Klassifizierung bedeutsam sind, hervorgehoben werden.

In dieser Arbeit wurden weiterführende XAI-Methoden, wie Weiterentwicklungen der Grad-CAM-Methode (Grad-CAM++ und Score-CAM) untersucht. Zudem wurde ein Konzept namens "Right for the Right Reasons" (kurz RRR) untersucht, das darauf abzielt, die Erklärbarkeit der LIME-Methode zu verbessern.

Vorgehen: Im ersten Schritt wurde das Paper "Right for the Right Reasons" analysiert und eine Implementierung auf simulierten Daten aus dem Paper mit der aktuellen Version vom Tensorflow reproduziert. Die Ergebnisse des Papers konnten zu 100% reproduziert werden.

Im zweiten Schritt wurde das Konzept auf ein simuliertes, EKG-ähnliches Signal angewendet, um zu prüfen, ob das Konzept auch für EKG-Signale einsetzbar wäre. Dabei wurde festgestellt, dass die Idee des Papers "Right for the Right Reasons" nicht für EKG-Signale anwendbar ist.

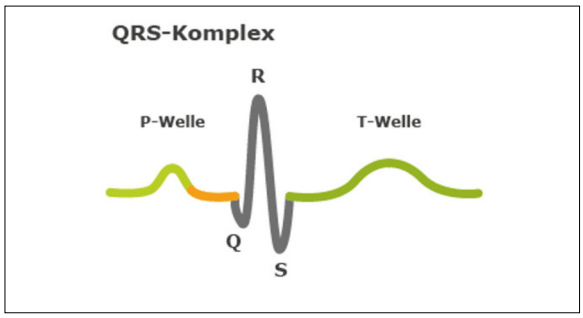
Im letzten Schritt wurde ein neues CNN-Modell entwicklet, welches mehrere Herzrhytmusstörung enerkennen kann. Dazu wurde das Dataset aus "PhysioNet Kardiologie Challenge" verwendet. Die drei XAI Methoden Grad-CAM (Gradient-weighted Class Activation Mapping), Grad-CAM++ und Score-CAM wurden implementiert, auf die EKG-Signale angepasst und deren Resultate verglichen.

Ergebnis: Die Erkenntnisse aus dieser Arbeit verdeutlichen, dass es keine universelle Technik oder Methode gibt, die uneingeschränkte Zuverlässigkeit bietet. Die vielversprechendsten Ergebnisse wurden durch den Einsatz von Score-CAM erzielt, dennoch besteht hier noch Raum für Verbesserungen.

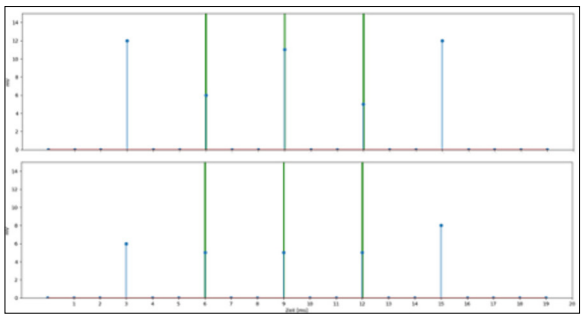
Besondere Aufmerksamkeit wurde der RRR-Methode zuteil, die insbesondere in Kombination mit LIME ihre Effektivität entfaltet hat. Auffällig war, dass die CAM-Methoden von der RRR-Methode nicht beeinflusst wurden.

Hinzu kommt, dass EKG-Signale eindimensionale Daten sind – im Grunde genommen ein Bild mit nur einer Pixelbreite. Diese Besonderheit kann auch für XAI-Methoden problematisch sein, da viele davon auf die Analyse von zweidimensionalen Bildern ausgelegt sind.

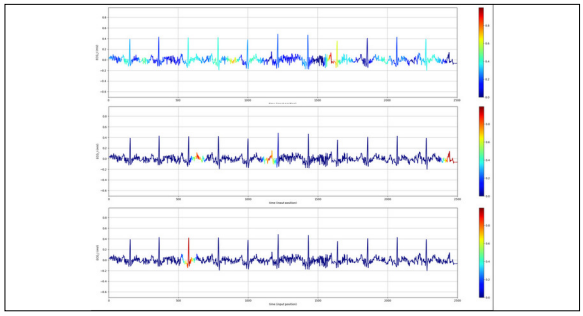
**EKG**
https://tinyurl.com/5xyzjfk4




**LIME mit RRR-Konzept. Erklärung der Predictions des CNNs auf einem eindimensionalen Signal**
Eigene Darstellung



**Vergleich der CAM-Methoden: Grad-CAM, Grad-CAM++ und Score-CAM.**
Eigene Darstellung




Referent
**Hannes Badertscher**

Korreferent
**Gabriel Sidler, Teamup Solutions AG, Zürich, ZH**


Themengebiet
**Data Science**