

SPARK-BASIERTE ANALYSE VON MONITORING-DATEN INTENSIV ÜBERWACHTER PATIENTEN ZUR IDENTIFIKATION VON ANOMALIEN: EIN USE CASE DER CHARITÉ HEALTH DATA PLATFORM

LINDA REBSTADT – HOCHSCHULE DARMSTADT

MOTIVATION

Herz-Kreislauferkrankungen sind die häufigste Todesursache der Welt, eine frühzeitige Erkennung kann dem entgegenwirken. Zur Erforschung solcher kardiovaskulärer Anomalien spielt die Analyse von Elektrokardiogrammen (EKGs) eine entscheidende Rolle. Durch die Anschaffung des Data Warehouse Connects und der Nutzung der Health Data Plattform ist es der Neurologie der Charité möglich, EKGs in hoher Auflösung über lange Zeiträume für die Forschung vorzuhalten. Jedoch stellen die großen Datenmengen und die technischen Artefakte Herausforderungen bei der Analyse dar. Ziele sind daher das Finden einer Methode zur Elimination der technischen Artefakte *Baseline Wander* und *Power-Line-Interferenz* sowie das Entwickeln eines Modells zur Erkennung abnormaler Herzschläge, um die Datenmenge auf die für medizinische Forscher*innen relevanten Stellen zu begrenzen.

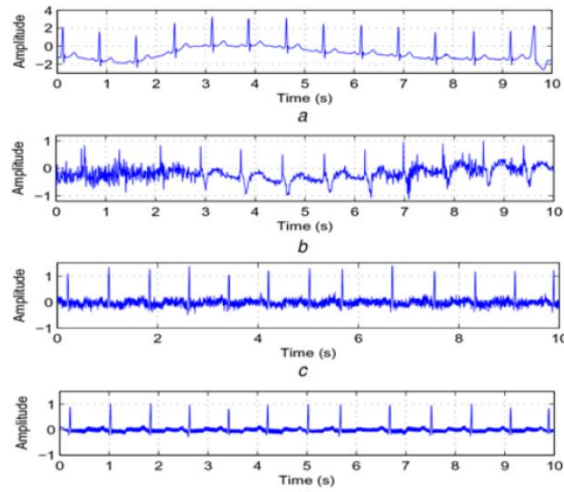


Abbildung 1.1: EKG-Signal mit vier verschiedene technische Artefakten: a) Baseline Wander, b) Muscle Artefact, c) Additive White Gaussian Noise und d) Power-Line-Interferenz; Quelle: Kumar et al., 2020

METHODEN

Zur Elimination der technischen Artefakte wurden Dekompositionen, Filter und Transformationen evaluiert, die auf diese großen Datenmengen mit moderatem Rechenaufwand zuverlässig arbeiten. Diese Methoden setzen sich zusammen aus der *Sparse Signal Decomposition (SSD)*, *Diskrete Wavelet-Transformation (DWT)* und der *Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)*. Aufgrund des hohen Rechenaufwands wurde die CEEMDAN nach der ersten Implementierung nicht weiter in den Vergleich miteinbezogen.

Das Modell zur Herzschlagerkennung setzt zur Reduktion des Aufwands auf den neuen Daten, auf die Prinzipien des *Transfer Learnings*. Es wurden Machine-Learning-Modelle auf der MIT-BIH-Arrhythmia-Datenbank trainiert, anschließend auf die Charité-Daten übertragen und anhand einer Teilmenge der vorhandenen Monitoring-Annotationen evaluiert. Diese Vorgehensweise konnte in einzelne Schritte unterteilt werden:

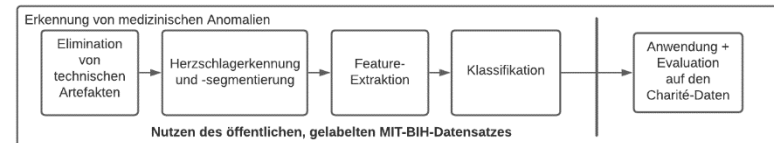
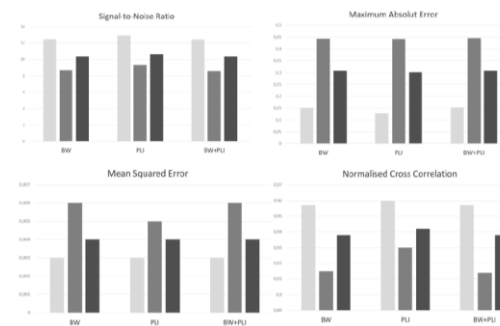


Abbildung 1.2: Ablauf der Schritte nach den Prinzipien des Transfer Learnings.

Die Feature setzen sich aus 19 Hauptkomponenten über aussagekräftige DWT-Koeffizienten und unabhängigen Komponenten des extrahierten Herzschlags zusammen sowie aus vier extrahierten Abständen zwischen den R-Peaks (die in Abb.1.1 zu sehenden Spitzen) eines jeden Herzschlags zu anderen R-Peaks vor und nach dem aktuellen Herzschlag. Als Klassifikationsmodelle wurden die *Lineare Diskriminanzanalyse*, der *K-Nächste-Nachbarn-Klassifizierer* und die *Support Vector Machine (SVM)* eingesetzt.

ERGEBNISSE

Methodenauswahl zur Elimination technischer Artefakte:



Legende: ■ SSD ■ DWT-db4 ■ DWT-dmey

Abbildung 1.3: Die nach Artefakt aggregierten Evaluationsergebnisse, getrennt nach den drei genutzten Metriken.

Die SSD zeigte in allen vier genutzten Metriken die besten Resultate, gefolgt von der DWT mit dem dmey-Wavelet als Basisfunktion. Zusätzlich wurde der Rechenaufwand aufgrund der großen Datenmenge mitbetrachtet. Dort zeigte die DWT die besten Ergebnisse.

Klassifikationsmodell zur Erkennung medizinischer Anomalien:

Elimination der technischen Artefakte

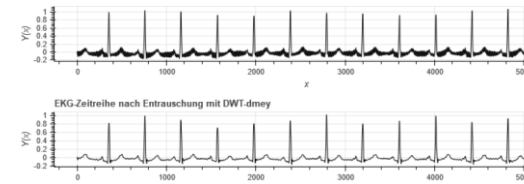


Abbildung 1.4: Elimination von Power-Line-Interferenz mithilfe der DWT

Herzschlagerkennung

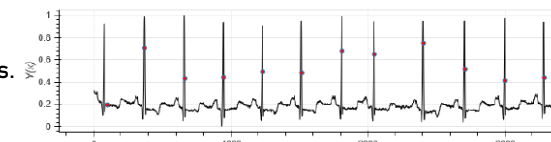


Abbildung 1.5: R-Peak-Erkennung

Die einzelnen Herzschläge wurden über den R-Peak erkannt. (Ergebnisse sind mithilfe der roten Punkte markiert) Die Segmentierung erfolgt über ein definiertes Fenster in einheitlicher Breite. Ein Beispiel eines extrahierten Herzschlags ist durch Abb. 1.6 gegeben.

Herzschlagklassifikation

Zum einen wurde ein Modell zur Klassifikation von sechs Herzschlagtypen trainiert, zum anderen ein Modell zur binären Klassifikation nach „normal“ und „abnormal“. Die höchste Accuracy, Sensitivität und den geringsten Anteil an Falsch-Alarmen zeigte die SVM. Die Ergebnisse auf einem Testdatensatz der MIT-BIH-Datenbank sind der nachfolgenden Tabelle zu entnehmen.

Typen	TP	FP	TN	FN	Sensitivität	Falsch-Alarme	Accuracy
6	4.828	12	1.923	28	99,75 %	0,62 %	99,13 %
2	5.275	26	1.906	29	99,45 %	1,36 %	99,23 %

Die Übertragung auf den Charité-Datensatz zeigte folgende Ergebnisse:

Typen	TP	FP	TN	FN	Sensitivität	Falsch-Alarme	Accuracy
2	42	55	39	4	91,30 %	58,51 %	57,86 %

FAZIT

Insgesamt ist festzuhalten, dass zum praktischen Einsatz auf große Datenmengen sich die DWT besser zur Elimination technischer Artefakte eignet, da der Rechenaufwand deutlich geringer ist, als die der SSD. Wird jedoch nur ein kleiner Datensatz bereinigt, so ist die SSD eher zu empfehlen, da die technischen Artefakte besser eliminiert werden können.

Die Klassifikation von abnormalen Herzschlägen zeigte auf dem MIT-BIH-Datensatz eine sehr hohe Accuracy, eine hohe Sensitivität und einen geringen Anteil an Falsch-Alarmen. Die Übertragung des Modells war jedoch aktuell noch nicht erfolgreich. Grund dafür scheint die ungleiche Auflösung der Aufnahmen zu sein, dessen Resample-Mechanismus zur Angleichung noch nicht gut genug funktionierte.

h_da

HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Kontakt

Linda Rebstadt

Fachbereiche Mathematik und
Naturwissenschaften & Informatik



Referentin

Prof. Dr. Antje Jahn

Korreferent

Prof. Dr. Arnim Malcherek

