

**Fachhochschule Aachen**  
**Campus Jülich**

Fachbereich 9  
Medizintechnik und Technomathematik

---

# **Untersuchung der Signalqualität von ballistokardiographischen Signalen mittels Methoden des maschinellen Lernens**

---

**Bachelorarbeit**  
im Studiengang Scientific Programming

von

**Cay Jakob Rahn**  
Matr.-Nr.: 3145495

19. Oktober 2020

1. Prüfer: Prof. Dr. rer. nat. Alexander Voß  
2. Prüfer: Dr.-Ing. Christoph Hoog Antink



# Erklärung

Diese Arbeit ist von mir selbständig angefertigt und verfasst. Es sind keine anderen als die angegebenen Quellen und Hilfsmittel benutzt worden.

---

Ort, Datum

---

Unterschrift



# Abstract

Die Beurteilung der Signalqualität bei ballistokardiographischen Signalen ist für eine Anwendung in der Praxis essentiell, wirft aber besonders bei in Betten aufgenommenen Signalen noch große Fragen auf. Ziel dieser Arbeit ist es, Möglichkeiten der Artefakterkennung bei solchen Signalen mittels maschinellen Lernens zu untersuchen. Grundlage sind bereits entwickelte Verfahren und Fachkenntnis über die Ballistokardiographie

Schwerpunkte der Arbeit:

- Recherche zu verschiedenen Verfahren des Maschinellen Lernens
- Beurteilung der Qualität bereits existierender Verfahren für Langzeitaufnahmen von bettlägerigen Patient\*innen
- Merkmalskonstruktion auf Basis von Domainenexpertise
- Vergleich verschiedener Verfahren und Eingabeparamater
- Untersuchung des Einflusses der Vorverarbeitung des Signals
- Evaluierung und Validierung der Ergebnisse

**Nicht final, nur als Platzhalter**



# Inhaltsverzeichnis

<b>Abkürzungsverzeichnis</b>	<b>ix</b>
<b>Abbildungsverzeichnis</b>	<b>xi</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Ziel der Arbeit . . . . .	2
1.3 Gliederung . . . . .	2
<b>2 Grundlagen</b>	<b>3</b>
2.1 Maschinelles Lernen . . . . .	3
2.1.1 Grundprinzipien . . . . .	3
2.1.2 Mathematischer Hintergrund . . . . .	4
2.1.3 Evaluation und Validierung . . . . .	6
2.1.4 Probleme maschinellen Lernens . . . . .	7
2.1.5 Weitere Lernmodelle des Überwachten Lernens . . . . .	8
2.2 Medizinische Grundlagen . . . . .	10
2.2.1 Kardiorespiratorisches System . . . . .	10
2.2.2 Übersicht Messtechniken . . . . .	11
2.3 Ballistokardiographie . . . . .	11
2.3.1 Medizinischer und technischer Hintergrund . . . . .	11
2.3.2 Einsatzgebiet . . . . .	12
2.3.3 Signaleigenschaften . . . . .	13
<b>3 Signalverarbeitung bei ballistokardiographischen Signalen</b>	<b>19</b>
3.1 Grundsätzliches . . . . .	19
3.2 Detektion von Herzschlägen . . . . .	20
3.3 Artefakterkennung . . . . .	22
3.3.1 Schwellwertbasierte Artefakterkennung . . . . .	25
3.3.2 Maschinelles Lernen mit statistischen Merkmalen . . . . .	26
3.3.3 Ähnlichkeit der Intervallschätzer des CLIE-Algorithmus . . . . .	27
3.4 Messdaten . . . . .	28
3.4.1 Vorverarbeitung . . . . .	28

3.4.2	Annotation der Daten . . . . .	30
<b>4</b>	<b>Analyse</b>	<b>35</b>
4.1	Aufbau und Evaluation der Verfahren . . . . .	35
4.2	Anwendung existierender Verfahren . . . . .	36
4.2.1	Ähnlichkeit der Intervallschätzer des CLIE-Algorithmus . . . . .	36
4.2.2	Schwellwerte für Standardabweichung, Minimum und Maximum	39
4.2.3	Maschinelles Lernen mittels statistischer Merkmale . . . . .	39
4.3	Analyse der Merkmale . . . . .	41
<b>5</b>	<b>Synthese</b>	<b>45</b>
5.1	Eingabeform der Daten . . . . .	45
5.2	Evaluationsmetriken . . . . .	45
5.3	Modelle . . . . .	45
<b>6</b>	<b>Evaluierung der Ergebnisse</b>	<b>47</b>
<b>7</b>	<b>Zusammenfassung und Ausblick</b>	<b>49</b>
7.1	Zusammenfassung . . . . .	49
7.2	Ausblick . . . . .	49



# Abkürzungsverzeichnis

<b>BKG</b>	Ballistokardiographie
<b>EKG</b>	Elektrokardiographie
<b>PPG</b>	Photoplethysmographie
<b>SKG</b>	Seismokardiographie
<b>HR</b>	Herzrate
<b>HRV</b>	Herzratenvariabilität
<b>CLIE</b>	Continuous Local Interval Estimator
<b>SQI</b>	Signal Quality Index
<b>MAE</b>	Mean Absolute Error
<b>MLP</b>	Multilayer-Perzeptron
<b>SVM</b>	Support Vector Machine
<b>LDA</b>	Linear Discriminant Analysis
<b>DT</b>	Decision Tree
<b>RF</b>	Random Forest
<b>PCA</b>	Principal Component Analysis



# Abbildungsverzeichnis

2.1	Darstellung von überwachtem Lernen . . . . .	4
2.2	Übersicht über die Funktionsweise eines allgemeinen im Bett eingebette- ten Ballistokardiographie (BKG)-Systems . . . . .	13
2.3	Beispiel eines typischen BKG-Signals mit Nomenklatur . . . . .	14
2.4	BKG-Aufnahmen in Rücken- und Seitenlage . . . . .	15
2.5	Visualisierung der Variabilität des BKG-Signals . . . . .	17
3.1	Intervallschätzer nach <b>Bruser2013</b> . . . . .	22
3.2	Flussdiagramm eines Algorithmus zur Beurteilung der Signalqualität . . .	23
3.3	Artefakte mit hoher Energie . . . . .	24
3.4	Artefakte mit niedriger Energie . . . . .	24
3.5	Genauigkeit der Herzratenberechnung bei schwellwertbasierter Artefak- terkennung . . . . .	25
3.6	Klassendiagramm der Datenstruktur für die Messdaten . . . . .	29
3.7	Aktivitätsdiagramm der Herzratenschätzung auf dem EKG-Referenzsignal	31
3.8	Elektrokardiographie (EKG)-Herzratenschätzungen auf sich überlappen- den 10-Sekunden-Segmenten für Patient 26 . . . . .	32
3.9	Verteilung der Label je nach maximal zulässiger Abweichung . . . . .	33
3.10	Verteilung der Label pro Patient*in . . . . .	33
3.11	Verteilung von $E_{HR}$ . . . . .	34
4.1	Verteilung von $E_{HR}$ auf dem Validierungsset . . . . .	36
4.2	Fehler und Coverage der Klassifikation nach der Ähnlichkeit der Inter- vallschätzer des CLIE-Algorithmus für verschiedene Schwellwerte im Vergleich zum gesamten Signal und der Annotation . . . . .	37
4.3	Verteilung von $E_{HR}$ bei den als informativ klassifizierten Segmenten . . .	38
4.4	Coverage unter bestimmten Fehlern $E_{HR}$ vor und nach Klassifikation . . .	38
4.5	Fehler und Coverage der Klassifikation für die verschiedenen Modellen des maschinellen Lernens mit statistischen Merkmalen im Vergleich um gesamten Signal und der Annotation . . . . .	40
4.6	Verteilung von $E_{HR}$ bei den als informativ klassifizierten Segmenten . . .	40
4.7	Coverage unter bestimmten Fehlern $E_{HR}$ vor und nach Klassifikation . . .	41

4.8	Korrelationsdiagramm der statistischen Merkmale, $E_{HR}$ und der binären Annotation . . . . .	42
4.9	Dimensionsreduktion der statistischen Merkmale mit einer Principal Component Analysis (PCA) mit linearem Kernel . . . . .	43
4.10	Wichtigkeit der Merkmale für den Random Forest (RF)-Klassifikator . . .	43

# 1 Einleitung

## 1.1 Motivation

Der derzeitige demographische Wandel stellt das Gesundheitssystem vor eine große Herausforderung: Es gibt immer mehr Patient\*innen, die im Alter überwacht und versorgt werden müssen. Eine kontinuierliche autonome Überwachung von Vitalparametern im Krankenhaus oder auch Zuhause erlaubt es, Erkrankungen frühzeitig zu erkennen und zu beobachten, ohne dass große Personalkapazitäten von Nöten sind.

Für diesen Anwendungszweck eignen sich vor allem Messmethoden, die die Patient\*innen im Alltag nicht einschränken und wenig invasiv sind. Im Englischen wird dies mit dem Begriff *unobtrusive* bezeichnet. Da es keine zufriedenstellende deutsche Entsprechung gibt, wird dieser im Folgenden nicht übersetzt verwendet werden. Solche *unobtrusive* Messmethoden beinhalten meist keine Notwendigkeit für direkten Körper- oder Hautkontakt, liefern aber Information über Atmung und Herzschlag. Die Herausforderung bei einem so ermitteltem Signal besteht in der Signalverarbeitung, da Messungenauigkeiten und Alltagsbewegungen zu Störungen im Signal führen. Nicht informatives, also nicht für die Verarbeitung geeignetes Signal muss aber zwingend identifiziert werden, da die Ergebnisse stark verfälscht werden.

Eine solche *unobtrusive* Messmethode ist die Ballistokardiographie (BKG). Sensoren lassen sich beispielsweise in Betten und Stühlen implementieren. Aufgezeichnet werden Aktivitäten des Herzens und der Atmung. Die Signalmorphologie variiert jedoch sowohl zwischen den Patient\*innen als auch innerhalb einer Person sehr stark, wodurch die automatische Beurteilung der Signalqualität erschwert wird. Um eine aussagekräftige Signalverarbeitung zu ermöglichen, ist dies jedoch essentiell. Besonders bei in Betten aufgenommenem Signal ist die Variation des Signals in Kombination mit Artefakten durch Körperbewegungen oder ähnliches problematisch.

### 1.2 Ziel der Arbeit

Das Ziel dieser Arbeit ist es, Möglichkeiten der Beurteilung der Signalqualität von BKG-Signalen mittels maschinellen Lernens zu untersuchen. Im besonderen Fokus liegen dabei Langzeitaufnahmen, sowohl nachts als auch tagsüber, von bettlägerigen Patient\*innen, da diese sich in der Vergangenheit als besonders anfällig für geringe Signalqualität gezeigt haben.

Dafür werden zunächst die vorliegenden Daten aufbereitet und existierende Verfahren der Artefakterkennung für diese getestet und bewertet. Anschließend wird das Signal auf mögliche Merkmale untersucht, die Aussagen über die Signalqualität ermöglichen. Anhand dieses Wissens werden Modelle des maschinellen Lernens ausgewählt und getestet. Auch wird untersucht, welchen Einfluss Annotation und Eingabeform auf das Ergebnis haben.

Langfristig soll ermöglicht werden, Ballistokardiographie (BKG) im medizinischen Alltag auch in unkontrollierten Umgebungen anzuwenden. Ziel dieser Arbeit ist es, für diesen Anwendungszweck eine erste Möglichkeit zur Beurteilung der Signalqualität zu bieten.

### 1.3 Gliederung

Zunächst wird in **Kapitel 2** ein allgemeiner Überblick über den medizinischen und technischen Hintergrund gegeben, der nötig ist, um die vorliegende Arbeit zu verstehen.

In **Kapitel 3** wird das Thema der Verarbeitung ballistokardiographischer Signale näher betrachtet. Dazu gehört die Detektion von Herzschlägen bei BKG-Signalen, existierende Verfahren zur Artefakterkennung und die Vorverarbeitung der vorliegenden Messdaten.

## 2 Grundlagen

Zum Verständnis dieser Arbeit ist grundlegendes Wissen nötig, welches hier in die drei Bereiche Maschinelles Lernen, Medizinische Grundlagen und Ballistokardiographie unterteilt ist.

### 2.1 Maschinelles Lernen

Da in dieser Arbeit Methoden des Maschinellen Lernens verwendet werden, wird im Folgenden eine Übersicht über Techniken, Hintergrund und verschiedene Lernmodelle gegeben.

#### 2.1.1 Grundprinzipien

Maschinelles Lernen ist die „künstliche“ Generierung von Wissen auf Basis von Erfahrung: Aus Beispielen wird gelernt und dieses Wissen nach einer Trainingsphase verallgemeinert. Dafür wird mit Mustererkennung gearbeitet und ein statistisches Modell aufgebaut, das auf den Trainingsdaten beruht. Die Trainingsdaten  $X$  bestehen aus Merkmalsvektoren  $x \in X \subseteq \mathbb{R}^n$ . Gesucht ist eine Funktion  $f$ , die diese Daten beschreibt. Unterschieden wird bei dieser Beschreibung dieser Daten zwischen Klassifikation und Regression. Bei einer Klassifikation werden die Eingabedaten in verschiedene Klassen unterteilt. Bei einer Regression dagegen werden stetige Werte vorhergesagt, es wird die Verteilung der Daten beschrieben. Sie kann z.B. dafür genutzt werden, Verkaufszahlen vorauszusagen. Ein Sonderfall ist die logistische Regression, die die Wahrscheinlichkeit der Angehörigkeit zu einer Klasse benennt.

Es wird zwischen verschiedenen Arten des maschinellen Lernens unterschieden, dem überwachten Lernen, dem unüberwachten Lernen und dem verstärkenden Lernen. Letzteres wird hier nicht betrachtet. Bei überwachtem Lernen bestehen die vorliegenden Daten aus Eingabe-Ausgabe-Paaren  $(x_1, y_1), \dots, (x_n, y_n), x_i \in X$ . Demnach ist nur die Funktion  $f: X \rightarrow Y$  unbekannt. Gesucht ist eine Funktion  $g$ , die  $f$  approximiert. Ein Lernalgorithmus ermittelt diese Funktion aus einem Hypothesenset. Dieses Vorgehen ist in Abbildung

2.1 visualisiert. Bei unüberwachtem Lernen dagegen ist  $Y$  unbekannt; die Eingabedaten haben die Form  $x_1, \dots, x_n, x_i \in X$ . Hier ist ein  $f$  gesucht, das die Daten möglichst gut beschreibt und sie beispielsweise eigenständig in Kategorien einteilt.

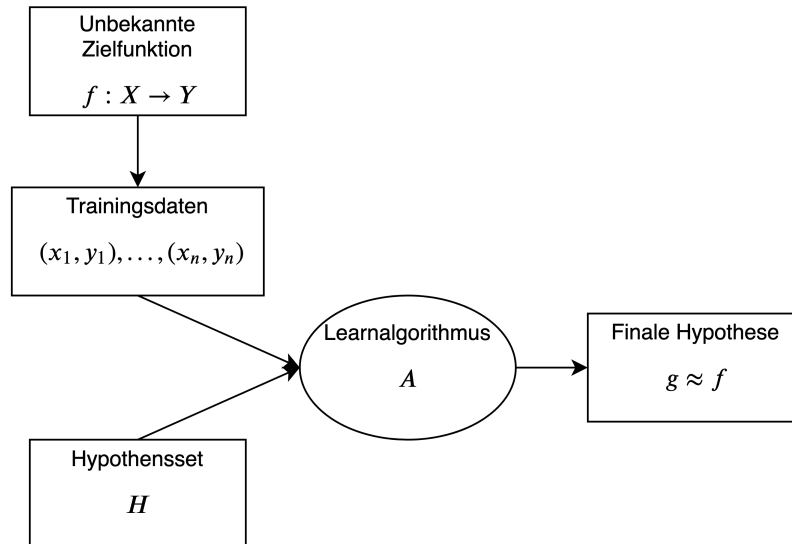


Abbildung 2.1: Darstellung von überwachtem Lernen

In dieser Arbeit wird überwachtes Lernen betrachtet, weshalb die untersuchten Daten annotiert werden müssen.

### 2.1.2 Mathematischer Hintergrund

- auch hier nur supervised betrachtet
- Verteilung der Daten
- linear separierbar oder nicht: 3 Fälle (linear, linear mit Rauschen, nicht linear)
- Abbildung dazu?
- Ziel ist, eine Funktion zu finden, die die verschiedenen Klassen voneinander möglichst genau separiert, wodurch  $y$  ermittelt werden kann bzw. bei einer Regression den Wert möglichst genau vorhersagt, Entscheidungsfunktion genannt
- Lineare Modelle kombinieren Merkmale linear miteinander
- einfachstes Modell für Klassifikation von linear separierbare Daten: Perzeptron
- Komponenten Merkmalsvektor  $x$  unterschiedlich gewichtet
- mit Gewichtungsvektor  $w \in \mathbb{R}^n$



- Ausgabe binärer Klassifikator  $y \in Y = \{-1, 1\}$
- Ermittlung Schwellwert *Bias*  $b$  mit

$$y = +1, \text{ falls } \sum_{i=1}^n w_i \cdot x_i > b$$
$$y = -1, \text{ falls } \sum_{i=1}^n w_i \cdot x_i < b$$

- wenn  $w_0 = b$  und  $x_0 = 1$  durch Umformen Perzeptron Entscheidungsfunktion  $h(x) = \text{sign}(w^T x)$
- $w$  so gewählt, dass Daten korrekt klassifiziert
- wenn Perzeptron künstliches Neuron in Neuronalem Netz wird diese Funktion auch Aktivierungsfunktion genannt und kann variiert werden
- Signum Funktion hat Nachteil des "hartenSSchwellwerts, oft Funktionen wie logistische Funktion mit "weichemSSchwellwert besser
- Lineare Regression: Gerade  $h(x)$  finden, die Datenpunkte beschreibt
- Methode der kleinsten Quadrate, also quadratische Abweichungen zwischen Gerade und Punkten minimieren
- gesucht sind Koeffizienten  $w$ ,

$$\underset{w}{\operatorname{argmin}} \sum_{i=1}^n (w^T x_i - y_i)^2$$

- durch Auflösen  $w = (X^T X)^{-1} X^T$
- bis jetzt nur lineare einfache Modelle betrachten
- bei komplexeren Minimierungsproblemen Methode kleinster Quadrate nicht ausreichend, andere Methoden zur Fehlerminimierung -> Gradientenabstiegsverfahren
- graphische Vorstellung als Hügellandschaft
- In jedem Schritt Ableitung der Kostenfunktion nach jedem Gewicht und Bias berechnet -> Schritt wählen, der Kostenfunktion am stärksten minimiert
- wenn Daten nicht linear separierbar: Kerneltrick möglich
- Transformation (Ersetzen) des Skalarprodukts transformiert implizit den Variablenraum

- Beispiel Gaußscher RBF-Kernel

$$K_{RBF}(x, x') := \exp(-\gamma \|x - x'\|^2) \text{ mit } \gamma > 0$$

- lineare Trennung in höher-dimensionalem Variablenraum
- Berücksichtigung aller Nicht-Linearitäten in der Transformation -> Trennfläche kann wiederum linear sein

### 2.1.3 Evaluation und Validierung

- Modell auf unbekannten Daten validieren
- Hold-Out-Validierung: zufällige Verteilung in Trainings- und Testset
- Kreuzvalidierung
- Daten werden auf v gleich große Mengen (Folds) verteilt
- V Modelle auf allen möglichen Folds trainiert
- validierung auf ausgeschlossenen Fold
- extrem: Leave-One-Out Kreuzvalidierung mit  $v = n$  (Menge der Datenpunkte)
- üblich: v-fache Kreuzvalidierung
- typische Werte  $v=5$  oder  $10$
- auch abhängig von verfügbarer Rechenleistung
- Wofür
- Schätzung von Fehler auf unbekannten Daten
- Wahl von Modellen/Bestimmung von Hyperparametern
- Parameter, die Modellarchitektur bestimmen
- Parameter, die Lernalgorithmus betreffen
- Regularisierungsparameter
- Ergebnis: 3 Datensets
- training, validierung, test
- Validierungsset beeinflusst Lernalgorithmus
- Testset nachdem alle Entscheidungen getroffen wurden

- typisch: Validierung durch zB Kreuzvalidierung auf Testset
- Retraining auf ganzem Trainingsset
- Ablaufdiagramm
- dafür evaluationsmetrik nötig -> Übergang
- verschiedenste Metriken
- begriffe TP, FP, TN, FN
- Accuracy  $ACC = \frac{TP+TN}{TP+TN+FN+FP}$  einfaches Gütemaß, problematisch bei ungleich großen Klassen, da eine hohe Genauigkeit erreicht wird, wenn immer die größere Klasse vorausgesagt wird
- Precision, "bestraft" Falsch-Positive  $PPV = \frac{TP}{TP+FP}$
- Recall, auch Sensitivity genannt, "bestraft" Falsch-Negative  $TPR = \frac{TP}{TP+FN}$
- Oft Recall und Precision in einem Maß zusammengefasst: F1 score harmonisches Mittel aus beiden  $F_1 = 2 * \frac{precision * recall}{precision + recall}$
- nicht balancierte Datensets
- balanced accuracy
- roc kurz für Receiver Operating Characteristic. basiert auf True Positive Rate TPR und False Positive Rate FPR, charakterisiert, wie gut beide Verteilungen durch Schwellwert trennbar sind
- auc Fläche unter der ROC Kurve

### 2.1.4 Probleme maschinellen Lernens

- offensichtlich: Underfitting
- Overfitting -> gute Approximation aber schlechte Generalisierung auf unbekannte Daten
- Einfluss von: Komplexität des Lernmodells, Datenmenge, Kontamination der Daten mit Rauschen
- mehr Datenpunkte mehr gut, Rauschen und Komplexität schlecht
- TODO: Bild
- Regularisierung gegen Overfitting: TODO näher nachlesen

- Stichprobenverzerrung

### 2.1.5 Weitere Lernmodelle des Überwachten Lernens

#### Entscheidungsbäume

- Bild mit Thresholds und Features
- an jedem Knoten Entscheidung ob bestimmtes Feature über Threshold liegt
- generiert z.B. durch rekursives binäres Teilen
- bei Regression mittlere quadratische Abweichung  $MSE$  minimieren, bei Klassifikationsbäumen Messung der "Unreinheit" der Blätter (Fehler = 0 wenn Blatt nur Punkte derselben Klasse, Fehler maximal wenn keine Klasse Mehrheit)
- Optimierung rechnerisch nicht effizient lösbar
- greedy den Schritt, also die Kombination aus Feature und Threshold wählen, der den Fehler am Stärksten reduziert
- Abbruchkriterien: maximal erlaubte Tiefe, Mindestanzahl Datenpunkte in einem Blatt
- keine Featureskalierung nötig

#### Random Forest

- besteht aus mehreren unkorrelierten Entscheidungsbäumen
- Randomisierung bei Erstellung der Bäume
- Mehrheitsentscheidung aller Bäume
- Nach Breiman für jeden Baum im Wald
- $n$  Bootstrap Samples ziehen, also mit Zurücklegen aus dem Trainings-Datensatz gezogen
- von  $M$  Merkmalen  $m \ll M$  Merkmale zufällig gewählt, die als Kriterium für Split infrage kommen
- voller Ausbau des Baums

### Nächste Nachbarn Modelle

- sehr einfache Modelle, öfters zur Schätzung einer Baseline im Einsatz
- Klassifikation anhand der Datenpunkte, die zu klassifizierenden Datenpunkt am nächsten liegen
- dafür Ähnlichkeit quantifizieren
- kein Training, Modell direkt durch Trainingsdaten definiert
- Voronoi-Regionen werden erzeugt: Menge aller Punkte die näher an einem Zentrum liegen als an allen anderen Zentren
- daraus Voronoi-Diagramm (gesammelte Grenzen)
- k-Nächste Nachbarn: das Label der Mehrheit der k nächsten Punkte wird zurückgegeben
- Regression: Mittelwert-Bildung über die Label der k nächsten Nachbarn

### Support Vector Machines

- Rauschen in Daten
- statt einfache Gerade wie bei Perzeptron wird Hyperebene mit maximalem Rand gesucht, die Daten korrekt klassifiziert
- TODO: Bild von Gerade mit Rand
- Ausgangspunkt Perzeptron, Umformulierung in Optimierung mit Nebenbedingung, dass Hyperebene größten Rand besitzt
- bei Support Vector Machine (SVM)s oft oben schon erwähnter Kernel-Trick, also Transformation des Skalarprodukts, damit nicht-lineare Separierung möglich

### Mehrlagiges Perzeptron

- Klasse von künstlichen neuronalen Netzwerken
- 3 Schichten: Eingabe-Schicht, Versteckte Schicht, Ausgabe-Schicht
- außer den Eingabeneuronen wird nicht-lineare Aktivierungsfunktion verwendet

## 2.2 Medizinische Grundlagen

Die vorliegende Arbeit beschäftigt sich mit der Beurteilung der Signalqualität in ballistokardiographischen Signalen. Zum Verständnis der gemessenen Vorgänge und der Problematik in Bezug auf die Signalqualität und dessen Beurteilung ist grundlegendes medizinisches Wissen über die gemessenen Vorgänge und messtechnisches Verständnis nötig. Aufgrund dessen wird hier eine kurze Übersicht über die medizinischen Grundlagen gegeben.

### 2.2.1 Kardiorespiratorisches System

Das kardiorespiratorische System (zusammengesetzt aus *kardia*, deutsch ‘Herz’ und *respiratio*, deutsch ‘Atmung’) setzt sich aus zwei Teilsystemen zusammen, dem kardiovaskulären und dem respiratorischen System, die zusammen die Versorgung der Organe mit Sauerstoff sicherstellen.

Das kardiovaskuläre System umfasst das Herz, die Arterien und die Venen. In einem Zyklus wird das sauerstoffreiche Blut von der linken Herzkammer durch die Arterien zu den Organen gepumpt, wo sich der Sauerstoff zur Versorgung dieser vom Blut löst. Die Venen transportieren das nun kohlenstoffdioxidreiche Blut in die rechte Herzkammer. Von dort wird es zur Lunge geführt, mit Sauerstoff angereichert und in die linke Herzkammer geleitet. Von dort beginnt der Vorgang von Neuem. Die Herzfrequenz ist hierbei ein relevanter messbarer Vitalparameter.

Ein Herzschlag selbst besteht aus zwei Phasen: einer füllenden und einer auswerfenden Phase. Während der Diastole, der Erschlaffungs- und Bluteinströmungsphase, füllen sich die Herzkammern mit Blut. Diese Phase endet mit dem Schließen der Herzklappen und die Systole beginnt. Die Systole ist die Anspannungs- und Blutausröhmungsphase: Die Herzklappen öföfnen sich durch Kontraktion des Herzmuskels und das Blut kann ausströmen.

Das respiratorische System umfasst die Lungen und den Lungenkreislauf. In einem Atemzyklus wird durch gezielte Muskelbewegungen Luft aus der Umgebung eingeatmet. Mit dem eingeatmeten Sauerstoff wird sauerstoffarmes Blut angereichert und anschließend die nun sauerstoffarme Luft ausgeatmet. In diesem Zusammenhang ist der Vitalparameter der Atemfrequenz messbar.

### 2.2.2 Übersicht Messtechniken

Zur Untersuchung der in dieser Arbeit betrachteten Ballistokardiographie (BKG) wird diese oft mit anderen Messmethoden als Referenz aufgenommen. Im Folgenden werden diese zur Einordnung kurz vorgestellt. BKG selbst wird im nächsten Abschnitt separat betrachtet.

Die Elektrokardiographie (EKG) zeichnet die elektrischen Aktivitäten des Herzmuskels auf, indem mit mehreren Elektroden die Spannungsänderung gemessen wird. Hier ist die Herzfrequenz sehr gut ablesbar.

Die Photoplethysmographie (PPG) ist ein optisches Messverfahren, bei dem die Menge des von der Haut reflektierten bzw. transmittierten Lichtes gemessen wird. Dadurch kann die Änderung des Blutvolumens gemessen werden; die Lichtmenge nimmt bei Durchlaufen einer Pulswelle durch die Arterie deutlich ab. Dieses Signal bietet Rückschluss auf Atmung und Herzschlag.

Oft gemeinsam mit dem BKG betrachtet wird die Seismokardiographie (SKG), bei der die Vibration der Wand des Brustkorbs, die durch den Herzschlag entsteht, aufgezeichnet wird. Aufgrund von fehlenden einheitlichen Definitionen wird in der Literatur teils auch der Begriff BKG für SKG genutzt.<sup>1</sup>

## 2.3 Ballistokardiographie

Im Folgenden wird die Ballistokardiographie eingeführt. Das beinhaltet den medizinischen und technischen Hintergrund, das Einsatzgebiet und die Signaleigenschaften.

### 2.3.1 Medizinischer und technischer Hintergrund

Ballistokardiographie (zusammengesetzt aus altgriechisch *ballein*, deutsch ‘werfen’, *kardía*, deutsch ‘Herz’ und *graphein*, deutsch ‘schreiben’) ist die graphische Darstellung der wiederholten, durch den Herzschlag verursachten Bewegungen des menschlichen Körpers. Erstmals schon im 19. Jahrhundert beobachtet<sup>2</sup>, ermöglicht der technische Fortschritt in der Sensortechnik heute aussagekräftige Messungen. Das BKG liefert durch die Aufzeichnung von zirkulierendem Blut und mechanischer Herzaktivität Informationen

---

<sup>1</sup>Inan2015.

<sup>2</sup>Gordon1877.

über die Gesamtleistung des kardiovaskulären Systems.<sup>3</sup> Konkret gemessen wird eine Massenbewegung, die durch die schnelle Beschleunigung des Blutes entsteht, wenn es während des Herzschlages durch die großen Arterien bewegt wird: Bei der Verteilung des Blutes in die peripheren Blutgefäße verschiebt sich das Zentrum der Körpermasse in Richtung der Füße und während der atrialen Systole Richtung Körpermitte. Die BKG-Wellenform entsteht durch diese Schwerpunktverschiebung.

Die Messung dieser Bewegung ist mit verschiedenen Sensortypen, die z.B. hydraulisch oder elektromechanisch auf Druck reagieren, möglich. Sensoren können unter anderem in Waagen, Stühlen und Betten eingebaut werden. Besonders bei im Bett gemessenen Signalen kann oft nicht klar zwischen SKG und BKG unterschieden werden, da sich myokardiale Vibrationen und Massverschiebungen durch den Blutfluss überlagern. Diese gemischten Signale werden in der Literatur teils auch als *cardiac vibration signals* bezeichnet.<sup>4</sup> Da im Bereich der Signalverarbeitung oft nicht zwischen reinem BKG und gemischten Signalen unterschieden wird, wird dies in der vorliegenden Arbeit ebenfalls nicht.

Verschiedene Studien kommen zu unterschiedlichen Ergebnissen bezüglich der Frage, welchen kardiovaskulären Ursprung die einzelnen Signaleile haben. Aufgrund dessen gestaltet sich die detaillierte Interpretation des BKG-Signals als schwierig. Da es neben Informationen zur Herzrate (HR) und Herzratenvariabilität (HRV) ein genauerer Indikator für das Alter des Herzens als Lebensalter ist, hat es trotzdem klinische Relevanz. Außerdem lassen sich durch abnormale Ballistokardiogramme Herzerkrankungen voraussagen, bevor Symptome auftreten. Besonders bei älteren Personen sind diese also eine wichtige Warnung.<sup>5</sup>

### 2.3.2 Einsatzgebiet

Durch diese Beschreibung wird schon deutlich, dass BKG anders als das sehr bekannte EKG ist. Der entscheidende Vorteil des BKGs liegt darin, dass kein einschränkender Körperkontakt wie z.B. durch aufgeklebte Elektroden nötig ist: Es lässt sich in Alltagsgegenständen wie Stühlen aber vor allem auch Betten implementieren, ohne dass es während der Messung zu Einschränkungen im alltäglichen Leben kommt oder medizinisches Fachpersonal anwesend sein muss. Damit gehört es zu den *unobtrusive* Messmethoden und eignet sich gut zur Langzeit- und Trendbeobachtung des Gesundheitszustandes - sowohl im klinischen Kontext als auch Zuhause. Besonders für Patient\*innen mit chronischen

---

<sup>3</sup>Pinheiro2010.

<sup>4</sup>Bruser2013.

<sup>5</sup>Pinheiro2010.



Krankheiten und zur Früherkennung krankhafter Veränderungen bietet eine gesundheitliche Überwachung von Zuhause großes Potential.<sup>6</sup> Je nach Aufbau des Messsystems verändert sich auch die Art der Informationen, die aus dem BKG-Signal gewonnen werden können. Sehr genaue, kontrolliert aufgenommene BKG-Signale ermöglichen eine aussagekräftige Analyse der Morphologie wobei beispielsweise in Betten eingebautes BKG zunächst nur Aussagen zu Herzrate und Herzratenvariabilität bietet. Zusätzlich zu Informationen der Herzaktivitäten ermöglichen Bettsysteme aber auch Informationen über das allgemeine Aktivitätslevel und somit auch über die Schlafqualität.<sup>7</sup> In dieser Arbeit wird es um die Aufzeichnung von BKG-Signalen in Betten gehen. Der Aufbau eines solchen Bettsystems ist in Abbildung 2.2 gezeigt.

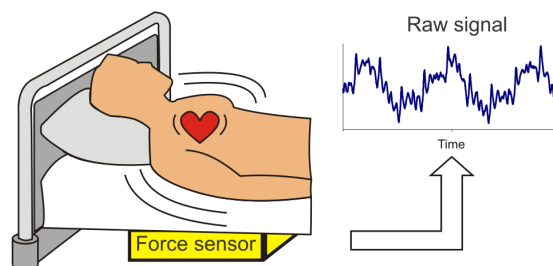


Abbildung 2.2: Übersicht über die Funktionsweise eines allgemeinen im Bett eingebetteten BKG-Systems.<sup>8</sup>

Allerdings ergeben sich neben diesen umfassenden Möglichkeiten auch Nachteile gegenüber konventionellen Messmethoden. Die größte Herausforderung ist eine stark variierende Signalqualität, die sich durch das unkontrollierte Umfeld und die Art der Messung ergibt.

### 2.3.3 Signaleigenschaften

Das gemessene BKG-Signal setzt sich aus Herzaktivitäten, Atmungsaktivitäten und Körperbewegungen zusammen. Gegebenenfalls wird es noch durch Störungen der Messung beeinflusst. Bei einer gesunden Person ohne Störeinflüsse wird die in Abbildung 2.3 abgebildete Wellenform erwartet. Diese Idealform lässt sich in 3 Gruppen unterteilen: Die präsysstolische, wobei diese häufig nicht beachtet wird, die systolische und die diastolische Gruppe unterteilen. Die mit H bis K markierten Extremwerte gehören bei dieser Unterteilung zur systolischen Gruppe, die Wellen L bis N zur diastolischen Gruppe. Die präsysstolische Gruppe, die aus den Wellen F und G besteht, ist in hier nicht abgebildet.

<sup>6</sup>Inan2015.

<sup>7</sup>Bruser2011.

<sup>8</sup>Entnommen aus Bruser2011

I und J werden auch als *ejection waves* bezeichnet. In Bezug auf andere Messmethoden ist zu bemerken, dass die H-Welle nahezu synchron mit dem ersten Herzgeräusch ist. Der Abstand des R-Peaks, des Hochpunkts eines EKGs zur H-Welle variiert im Bereich von 0,2 bis 0,3 Sekunden.<sup>9</sup> Die Amplitude der Wellen ohne Störeinflüsse ist hauptsächlich abhängig von dem Herzzeitvolumen, der Herzkraft und der Geschwindigkeit des Auswurfs.<sup>10</sup>

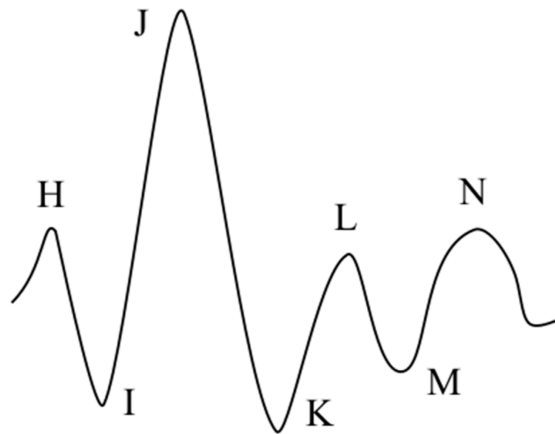


Abbildung 2.3: Beispiel eines typischen BKG-Signals mit Nomenklatur<sup>11</sup>

Im Idealfall wird zwar die oben beschriebene Wellenform erwartet, bei der die Wellen H bis L eine deutliche W-Form bilden, allerdings ist es trotz dieser typischen Form selten, dass alle nicht-systolischen Komponenten sichtbar sind.<sup>12</sup> Es gibt eine starke Variation der Signalmorphologie sowohl zwischen als auch innerhalb von Individuen. Der größte Einfluss ergibt sich durch die verwendeten Sensoren und die Position der Person, also zum Beispiel ob im Stehen, Sitzen oder Liegen gemessen wird.<sup>13</sup> Es gibt Studien die zeigen, dass die intraindividuelle Varianz über serielle Messungen hinweg niedrig ist.<sup>14</sup> Allerdings gilt das nicht, wenn sich die Position der Person verändert. Hierbei reicht es schon, wenn die Person in Rückenlage statt Seitenlage liegt.<sup>15</sup> Aufgrund dieser Variationen in der Signalmorphologie wurden schon in den 1950er Jahren 3 Achsen für die Aufzeichnung des BKGs definiert: Die longitudinale (Kopf-Fuß), die transversale (Seite-Seite) und die dorsoventrale (Rücken-Brust).<sup>16</sup> Zu Beginn maßen die meisten Systeme entlang der longitudinalen Achse, die z.B. der Messung auf einer Waage entspricht. *Unobtrusive* Messsysteme, wie die hier betrachtete Messung in Betten, messen entlang einer

---

<sup>9</sup>DELALLA1950.

<sup>10</sup>Pinheiro2010.

<sup>11</sup>Entnommen aus Albukhari2019 nach Starr1939.

<sup>12</sup>Pinheiro2010.

<sup>13</sup>Sadek2019.

<sup>14</sup>Inan2015.

<sup>15</sup>Bruser2011.

<sup>16</sup>Bruser2011; Inan2015.

Kombination der transversalen und der dorsoventralen Achse - abhängig von der Position der Person. Besonders diese Kombination sorgt für eine große intra- und individuelle Variation des Signals. Abbildung 2.4 verdeutlicht dies durch den direkten Vergleich von BKG-Aufzeichnungen zweier Herzschläge von 2 Personen. Bei jeder dieser beiden Personen wurde in zwei verschiedenen Positionen gemessen.<sup>17</sup> Auch der Ursprung des Signals ist abhängig von der Messachse. Bei longitudinal gemessenem BKG ist der Einfluss des Herzzeitvolumens schon seit 1929 beobachtet.<sup>18</sup> Im Gegensatz dazu ist der Ursprung des in Betten gemessenen BKG-Signals nicht genau bekannt. Das liegt unter anderem daran, dass mechanische Komponenten wie z.B. die Matratze einen schwer zu modellierenden Einfluss haben.

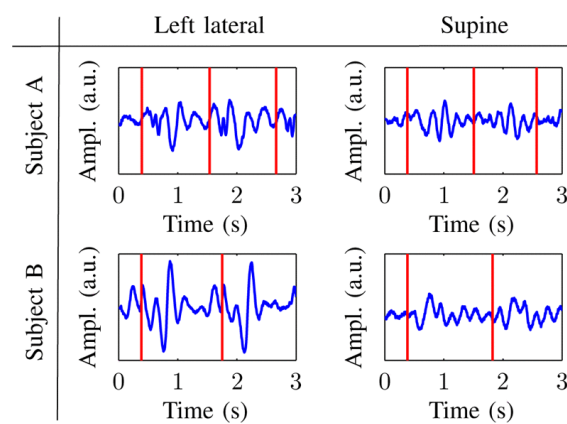


Abbildung 2.4: Hochpass-gefilterte BKG-Aufnahmen von zwei Herzschlägen zweier verschiedener Personen, jeweils in Rücken- und Seitenlage gemessen. Die vertikalen Linien markieren die R-Peaks der EKG-Referenz.<sup>19</sup>

Neben Einflüssen der verwendeten Messachse und der Körperposition beeinflusst auch die Atmung die Signalform. Normale Atmung beeinflusst die Amplitude der *ejection waves* I und J. Bei Atemstillstand dagegen werden die H und J Wellen verzerrt. Auch bei einer gesunden, sich nicht bewegendenden Person, die ihre Atmung kontrolliert, wird kein exakt Schlag für Schlag reproduzierbares Signal erzeugt werden.<sup>20</sup> Von **Zink2017** werden die Einflüsse der Atmung in der vertikalen Achse eines dorsoventralen BKGs als große Schwingungen einer Wellenlänge von fünf bis zehn Sekunden beschrieben. Innerhalb dieser sind kleinere Schwingungen mit höherer Frequenz sichtbar, die jedoch keiner bestimmten Sequenz folgen.<sup>21</sup> Zusätzlich zu dieser schon beschriebenen Variabilität kommt es sehr leicht zum Entstehen von Artefakten. Ursprung ist entweder das Messsystem selbst oder Körperbewegungen. Insgesamt führt Bewegung der Patient\*innen, auch

<sup>17</sup>**Bruser2011.**

<sup>18</sup>**Starr1939.**

<sup>19</sup>Entnommen aus **Bruser2011.**

<sup>20</sup>**Pinheiro2010.**

<sup>21</sup>**Zink2017.**

die der Atmung, zu einem *baseline drift*. Stärkere Bewegungen führen zu einer Massenschiebung, die um ein Vielfaches größer als die gemessenen Vorgänge ist. Aufgrund dessen führt sie immer dazu, dass das Signal stark verzerrt oder sogar vollständig überlagert wird.

Besonders im Vergleich zu anderen kardiorespiratorischen Signalen wie dem EKG und PPG wird deutlich, dass BKG-Signale auch in konsekutiven Messungen deutlich variabler sind. Abbildung 2.5 zeigt dies am Beispiel von BKG-Aufnahmen eines im Bett integrierten Messsystems im Vergleich zum parallel aufgenommenen EKG. Es zeigt sich, dass selbst nach Entfernung von Überlagerungen von Atmung und Bewegung das BKG-Signal eine höhere Variabilität in Bezug auf Amplitudenhöhe, Reihenfolge der Extremwerte und der gesamten Form aufweist.<sup>22</sup> Es wird allerdings angenommen, dass aufeinander folgende Herzschläge sich ähneln<sup>23</sup>. Diese Eigenschaft wird Selbstähnlichkeit genannt. **Bruser2013** nennt als eine mögliche Ausnahme den Fall, dass ein unregelmäßiger Herzschlag mit sehr niedrigem Schlagvolumen einem regulären Herzschlag folgt. In dem Fall ist es möglich, dass die Amplitude im Vergleich so klein ist, dass sie verdeckt wird. Dies ist z.B. bei Vorhofflimmern möglich. Eine Untersuchung von **Rosales2012** zeigt dieses Verhalten der Selbstähnlichkeit nicht bei den kleineren Extremwerten die J umgeben. Dass die Ähnlichkeit um J am größten ist zeigt auch Abbildung 2.5.

---

<sup>22</sup>**Zink2017**.

<sup>23</sup>**Brueser2013**.

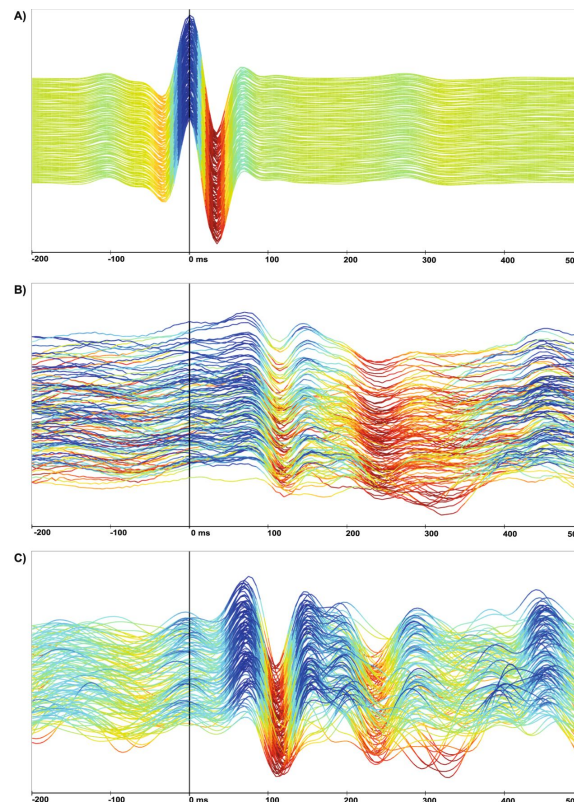


Abbildung 2.5: Diagramm aus 128 konsekutiven Herzschlägen im EKG (A) und BKG (B,C), segmentiert durch das EKG. Die Farben dienen der besseren Visualisierung der Amplituden. (A) EKG-Signal; (B) BKG-Signal mit Überlagerungen durch Atmung und Bewegung; (C) BKG-Signal ohne Bewegungsartefakte und Atmung.<sup>24</sup>

Zusammengefasst lässt sich sagen, dass es sich bei ballistokardiographischen Signalen um nichtstationäre Signale handelt, dessen Ursprung nicht genau bekannt ist. Die Signalform wird von der Messachse, der Position und Körperhaltung der Proband\*innen und dem Messsystem selbst beeinflusst. Besonders bei dem hier im Fokus liegenden Anwendungsfall Bett kommt es sowohl durch die unkontrollierbare Umgebung als auch die Signaleigenschaften selbst zu einer starken Variation der Morphologie und vielen Artefakten im Signal. Trotz dieser Einschränkungen ist die Ballistokardiographie eine Messtechnik, die sich einfach *unobtrusive* in den Alltag einbauen lässt und Aussagen über die Herzrate und die Herzratenvariabilität ermöglicht.

<sup>24</sup>Entnommen aus Zink2017.



## 3 Signalverarbeitung bei ballistokardiographischen Signalen

Im folgenden Kapitel wird eine Einführung in das Thema der Signalverarbeitung und besonders Artefakterkennung bei ballistokardiographischen Signalen gegeben. Hierzu wird zunächst Grundlegendes zu der Verarbeitung kardiorespiratorischer Signale erläutert. Ein in dieser Arbeit verwendeter Algorithmus zur Detektion von Herzschlägen wird vorgestellt. Anschließend wird die Thematik der Artefakterkennung eingeführt und der aktuelle Stand der Wissenschaft bei BKG-Signalen vorgestellt. Hierfür werden drei Verfahren im Detail betrachtet. Darüber hinaus werden die in dieser Arbeit untersuchten Daten vorgestellt und die durchgeführte Vorverarbeitung beschrieben.

### 3.1 Grundsätzliches

Kardiorespiratorische Signale sind durch die quasiperiodische Natur des Herzens und der Lunge selbst quasiperiodisch. Zwei zyklische Vorgänge werden gleichzeitig gemessen, lassen sich aber durch eine Bandpass-Filterung nach ihren unterschiedlichen Frequenzen filtern. Der Normbereich für die Atemfrequenz liegt bei 12 bis 25 Atemzügen pro Minute, alles ober- und unterhalb wird als abnormal betrachtet. Bei der Herzfrequenz wird ein Bereich von 30 bis 200 Schlägen pro Minute erwartet. Dabei entsprechen 30 Schläge pro Minute der Pulsabsenkung in der Nacht, der Ruhepuls selbst ist höher. Grundsätzlich gibt es verschiedene Arten der Signalverarbeitung: Algorithmen, die im Zeitbereich arbeiten, solche, die im Frequenzbereich arbeiten und solche, die beides kombinieren. Dabei ist zu beachten, dass bei frequenzbasierten Algorithmen durch die Analyse von spektralen Eigenschaften zunächst nur durchschnittliche Frequenzen ermittelt werden. Dies ist für einige medizinische Anwendungen ausreichend, für andere wie z.B. die Ermittlung der HRV allerdings nicht. Algorithmen, die auf dem Zeitbereich arbeiten basieren oft auf Wissen über die Morphologie des physiologischen Signals. Durch die Eigenschaften der BKG-Signale, vor allem durch die variable Morphologie ist dies schwieriger als bei anderen kardiorespiratorischen Signalen. **Paalasmaa2015** sagt dazu:

*The properties of the BCG signal vary so much in practice that no simple filtering rule can be devised for an accurate and reliable beat-to-beat interval detection<sup>1</sup>*

Diese Aussage gilt sowohl für die Detektion von Schlag-zu-Schlag Intervallen als auch für die Beurteilung der Signalqualität, bzw. die Artefakterkennung.

## 3.2 Detektion von Herzschlägen

In dieser Arbeit wird der von **Bruser2013** entwickelte Algorithmus, der Continuous Local Interval Estimator (CLIE) verwendet. Aus diesem Grund wird er hier vorgestellt. Der Algorithmus beruht auf der in Kapitel 2.3 erwähnten Annahme, dass sich aufeinander folgenden Herzschläge ähneln und schätzt die Herzrate anhand der Selbstähnlichkeit des Signals.

Der Algorithmus iteriert mit einem *Moving window* über das mit einem Bandpass gefilterte Signal. Es werden zwei Schwellwerte für die Intervalllänge  $T$  genutzt,  $T_{\min}$  und  $T_{\max}$ , basierend auf dem bekanntem Bereich der Herzrate von 30 bis 200 Schlägen pro Minute. Die Länge des Analysefensters  $w_i$  entspricht  $2 \cdot T_{\max}$ , sodass mindestens zwei vollständige Herzschläge enthalten sind.  $f_s$  beschreibt dabei die Abtastrate des untersuchten Signals.

$$w_i[v] = x[n_i + v], v \in \{-T_{\max} * f_s, \dots, T_{\max} * f_s\}$$

In jedem Fenster wird die lokale Intervalllänge  $T_i$  geschätzt und anschließend das Zentrum des Fensters  $n_i$  weiterbewegt:

$$n_{i+1} = n_i + \Delta t * f_s$$

.

Die Schätzung der Intervalllänge beruht auf drei Selbstähnlichkeitsmaßen, die wie folgt definiert sind:

---

<sup>1</sup>Paalasmaa2015.



$$\begin{aligned}
E_{\text{Corr}}[N] &= \frac{1}{N} \sum_{v=0}^N w[v]w[v-N], \\
E_{\text{AMDF}}[N] &= \left( \frac{1}{N} \sum_{v=0}^N |w[v] - w[v-N]| \right)^{-1}, \\
E_{\text{MAP}}[N] &= \max_{v \in \{0, \dots, n\}} (w[v] + w[v-N]).
\end{aligned}$$

Dabei berechnet  $E_{\text{Corr}}$  eine modifizierte Autokorrelationsfunktion, mit  $E_{\text{AMDF}}$  wird die Differenz des Signals bei verschiedenen Abständen miteinbezogen und mit  $E_{\text{MAP}}$  werden die maximale Amplituden von zwei beliebigen Samples über das ganze Fenster berechnet. AMDF steht für *modified average magnitude difference function* und MAP für *maximum amplitude pairs*. Diese Schätzer entsprechen jeweils einer Wahrscheinlichkeitsfunktion, die beschreibt wie wahrscheinlich es ist, dass  $n$  dem tatsächlichen Schlag-zu-Schlag-Intervall entspricht. Durch Skalierung können sie in Wahrscheinlichkeitsdichtefunktionen verwandelt werden. Durch Kombination dieser drei Funktionen wird nun der wahrscheinlichste Wert für  $N$  ermittelt:

$$N_{\text{opt}} = \underset{N}{\operatorname{argmax}} p(N|E_{\text{Corr}}, E_{\text{AMDF}}, E_{\text{MAP}})$$

Nach dem Satz von Bayes kann die Wahrscheinlichkeit, dass  $N$  der tatsächlichen Intervalllänge entspricht, auch wie folgt ausgedrückt werden:

$$p(N|E_{\text{Corr}}, E_{\text{AMDF}}, E_{\text{MAP}}) = \frac{p(E_{\text{Corr}}, E_{\text{AMDF}}, E_{\text{MAP}}|N)p(N)}{p(E_{\text{Corr}}, E_{\text{AMDF}}, E_{\text{MAP}})}$$

Da  $p(E_{\text{Corr}}, E_{\text{AMDF}}, E_{\text{MAP}})$  unabhängig von  $N$  ist, kann es für die Ermittlung der wahrscheinlichsten Intervalllänge  $T_{\text{opt}}$  vernachlässigt werden. Unter den Annahmen, dass die Ergebnisse der drei Schätzer nicht voneinander, sondern nur von  $N$  abhängen und dass  $N$  gleichverteilt ist, kann  $T_{\text{opt}}$  wie folgt ermittelt werden:

$$\begin{aligned}
E_f[N] &= E_{\text{Corr}}[N] \cdot E_{\text{AMDF}}[N] \cdot E_{\text{MAP}}[N], \\
N_{\text{opt}} &= \underset{n}{\operatorname{argmax}} E_f[N]
\end{aligned}$$

Durch lineare Skalierung erhält man auch hier eine Dichtefunktion. Abbildung 3.1 zeigt

die 3 einzelnen Dichtefunktionen und die fusionierte Fusion. In letzterer zeigt ein deutlicher Hochpunkt  $N_{\text{opt}}$ .

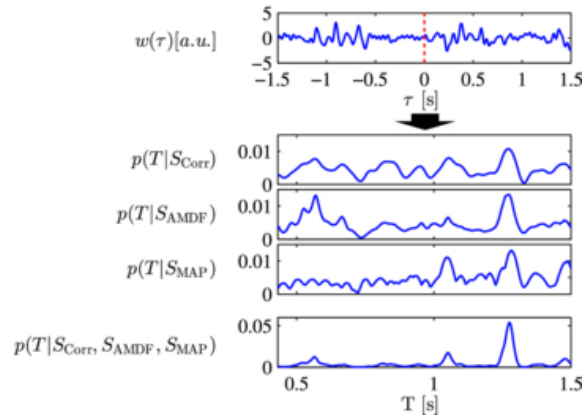


Abbildung 3.1: Die drei Intervallschätzer und ihre Fusion<sup>2</sup>

Nun gibt es für jeden Punkt im Signal eine Schätzung der Intervalllänge. Mit Hilfe der Fenstergröße und dieser Länge können die zu einem Herzschlag gehörenden Hochpunkte ermittelt werden, nämlich die, die die größte kombinierte Amplitude mit dem durch die Intervalllänge gegebenen Abstand besitzen. Für jeden dieser Punkte  $P_k$  existiert nun eine Menge an Schätzungen  $T_k$ , mit der eine robuste Intervallschätzung  $\bar{T}_k = \text{median}(T_k)$  ermittelt werden kann.<sup>3</sup>

### 3.3 Artefakterkennung

Artefakte bzw. Signal von geringer Qualität sind irrelevante Signale mit variierender Amplitude, Frequenz und Dauer, die das physiologische Signal stören<sup>4</sup>. Das Ziel der Artefakterkennung ist es, nur die Teile des Signals zu verarbeiten, die Vitalparameter enthalten. Bewegungsartefakte, Sensorstörungen und ähnliches, die diese Parameter überlagern, sollen die Verarbeitung nicht beeinflussen. Dabei ist die Quelle der Störung selbst irrelevant, allerdings sollten keine medizinisch induzierten Abnormalitäten als gestörtes Signal klassifiziert werden. Insgesamt ist es für aussagekräftige Messungen wünschenswert, eine hohe Coverage zu erreichen, das heißt einen möglichst hohen Anteil an Signal zu verwenden. **Sadek2016** unterscheidet zwischen informativem und nicht-informativem Signal. Informatives Signal enthält mit *noise* gemischtes Signal von guter Qualität aus dem Vitalparameter ohne weiteres extrahiert werden können. Im Gegensatz dazu sind

<sup>2</sup>Entnommen aus **Bruser2013**.

<sup>3</sup>**Bruser2013**.

<sup>4</sup>**Nizami2013**.

die Informationen bei nicht-informativem Signal so mit Artefakten und *noise* vermischt, dass vor Extraktion der Vitalparameter weitere individuelle Verarbeitung nötig ist oder die Extraktion von physiologischen Eigenschaften gar unmöglich ist. Teils wird die Signalqualität auch mit so genannten Signal Quality Indices (SQIs) gemessen, die je nach SQI und Anwendungsfall verschiedene Aussagen haben. Im klinischen Kontext genutzte Artefakterkennung verwendet oft relativ einfaches Preprocessing.<sup>5</sup> Außerdem sind in den meisten Algorithmen bestimmte Informationen direkt oder indirekt *hard coded*. Das kann zum einen etwas wie Typ oder Frequenz der Daten sein, aber auch demographische Informationen über die Patient\*innen wie Alter, Gewicht oder medizinischer Zustand.<sup>6</sup>

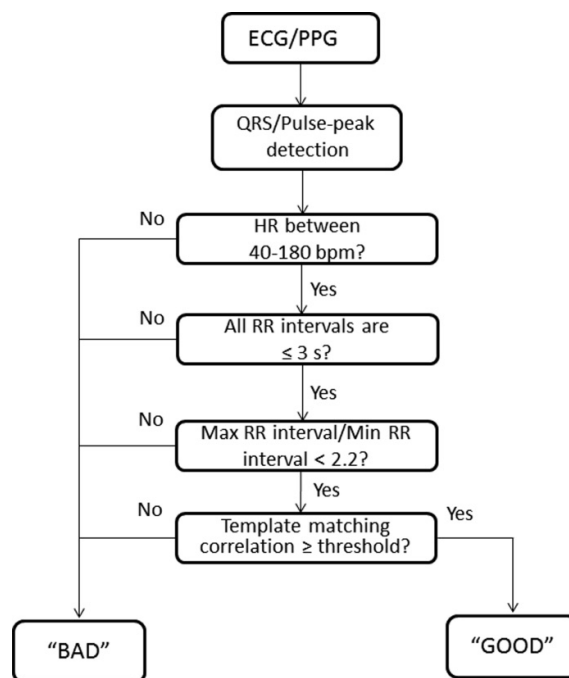


Abbildung 3.2: Flussdiagramm eines Algorithmus zur Beurteilung der Signalqualität<sup>7</sup>

Um einen Eindruck über übliche Beurteilung der Signalqualität bei anderen kardiorespiratorischen Signalen zu bekommen, wird in Abbildung 3.2 ein Beispiel für EKG und PPG gezeigt. In jeweils 10-Sekunden Fenstern wird zunächst eine Segmentierung der Herzschläge durchgeführt und anschließend vier Kriterien überprüft, die jeweils ausreichend sind, um die Signalqualität als schlecht zu klassifizieren. Im ersten Kriterium wird geprüft ob die Herzrate zwischen 40 und 180 Schlägen pro Minute liegt. Im zweiten Schritt wird sichergestellt, dass kein Schlag fehlt, indem geprüft wird, ob alle Intervalle kürzer als drei Sekunden sind. Eine Intervalllänge von 3 Sekunden entspräche über eine Minute einer Herzrate von 20 Schlägen pro Minute. Anschließend wird geprüft, ob die Herzrate nur in

<sup>5</sup>Nizami2013.

<sup>6</sup>Nizami2013.

einem begrenzten Bereich variiert; das Verhältnis der maximalen zur minimalen Intervalllänge muss kleiner als 2.2 sein. Abschließend wird geprüft, ob die Korrelation mit einem erstellten Template einen gewissen Schwellwert nicht unterschreitet. Dieser Algorithmus enthält übliche Techniken der Signalbeurteilung: Eine Begrenzung der akzeptierten Herzrate und Herzratenvariabilität und dem Vergleich mit einem zuvor erstellten Template.

Bei ballistokardiographischen Signalen gestaltet sich auch die Beurteilung der Signalqualität schwieriger als bei anderen kardiorespiratorischen Signalen. Zusätzlich zu der gegebenen Variabilität durch die Atmung ist eine Problematik, dass plötzliche Veränderungen der Signalmorphologie bei Positionsänderungen zuvor erstellte Templates obsolet machen und auch andere Schwellwerte nicht mehr angemessen sind. Bei Artefakten in BKG-Signalen kann zwischen Artefakten mit hoher und mit niedriger Energie unterschieden werden. Artefakte mit hoher Energie, wie in Abbildung 3.3 gezeigt entstehen, weil Bewegungen stärkere Krafteinwirkungen verursachen als Atmung und Herzschlag. Diese Tatsache wird häufig zur Erkennung dieser verwendet. Artefakte mit niedriger Energie, wie in Abbildung 3.4 gezeigt, sind weniger auffällig.



Abbildung 3.3: Artefakte mit hoher Energie

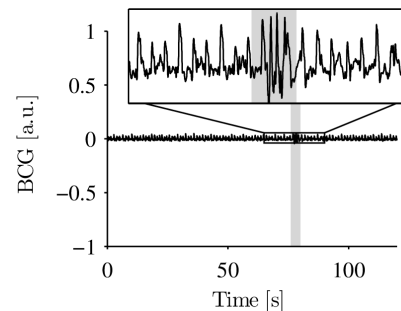


Abbildung 3.4: Artefakte mit niedriger Energie

Bei der Betrachtung verschiedener Ansätze zur Signalverarbeitung und Artefakterkennung bei BKG-Signalen ist auffällig, dass Proband\*innen oft angewiesen werden, sich möglichst wenig zu bewegen. Das ist bei der Messung im Alltag, insbesondere bei Messungen in Betten nicht realistisch. Dazu kommt, dass beim Einsatz im Alltag auch Kontakt mit der Person nicht immer gewährleistet ist, z.B. wenn diese sich aufrichtet. **HoogAntink2020** haben festgestellt, dass bei Messsystemen in Betten besonders Messungen tagsüber große Signaleile von schlechter Qualität aufweisen, deutlich mehr als nachts. Dies ist durch gesteigerte Aktivität tagsüber erklärbar. Ebenfalls im *unobtrusive* Kontext nicht zielführende Methoden zur Artefakterkennung verwenden eine EKG-Referenz.

Im Folgenden werden drei verschiedene Methoden zur Beurteilung der Signalqualität vorgestellt.

### 3.3.1 Schwellwertbasierte Artefakterkennung

In **Pino2015** präsentieren **Pino2015** einen Ansatz für die Erkennung von Körperbewegungen für ein in einen Stuhl eingebettetes BKG-Messsystem. Dafür werden über ein *moving window* Maximum, Minimum, Standardabweichung und Mittelwert ermittelt und daraus 2 Schwellwerte berechnet:

$$T_1 = \frac{\max + \min}{2},$$

$$T_2 = \text{mean} + 1,1 * \text{std}.$$

Ein Segment wird als informativ klassifiziert, wenn  $T_1 \leq T_2$  ist. Die Länge des *moving window* ist mit 200 Samples bei einer Abtastrate von 200 Hz benannt. Untersucht wurden sowohl Freiwillige im Labor, als auch im Krankenhauswartzimmer für eine sehr kurze Messdauer von ein bis zwei Minuten. Mit diesem Ansatz wurde bei mehr als 50 % der Laborgruppe eine Coverage zwischen 87 % und 95 % erreicht. Die Coverage der im Krankenhaus aufgenommenen Gruppe war bedeutend niedriger; hier lagen 50 % der Messungen zwischen 48 % und 95 % Coverage. Zu der Genauigkeit der Herzschlagdetektion auf den akzeptierten Signalteilen wird keine Aussage in Zahlen getroffen, sondern nur der folgende Bland-Altman Graph gezeigt.

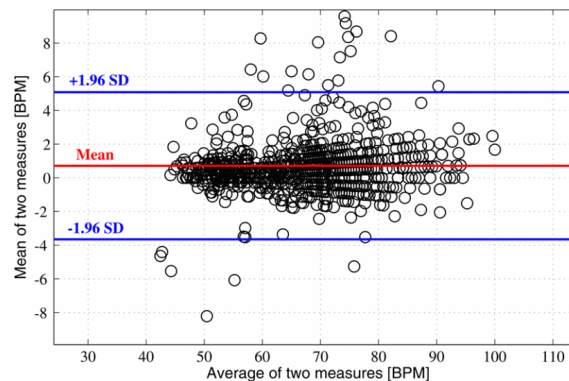


Abbildung 3.5: Bland-Altman Graph zwischen von EKG und BKG berechneter HR.

Hier zeigt sich, dass beim Großteil des hier betrachteten Signals die HR größtenteils mit einer Genauigkeit von  $\pm 3$  Schläge pro Minute bestimmt werden konnte. Allerdings handelt es sich hier um im Sitzen aufgenommenes Signal, bei dem die Variabilität geringer ist als bei in Betten aufgenommenem BKG.

#### 3.3.2 Maschinelles Lernen mit statistischen Merkmalen

Ein Algorithmus zur Beurteilung der Signalqualität mittels maschinellen Lernens wird von **Sadek2016** im Paper **Sadek2016** beschrieben. Betrachtet werden BKG-Signale, die in einem Massagesessel aufgenommen werden, also ebenfalls im Sitzen aufgenommenes Signal, bei dem eine geringere Variabilität als in dem hier untersuchten Anwendungsfall erwartet wird.

Die vorliegenden Daten wurden manuell von Expert\*innen als informativ oder nicht informativ klassifiziert und in 10-Sekunden-Segmente, die sich nicht überlappen, aufgeteilt. Insgesamt waren 58 % der Daten als informativ und 42 % als nicht-informativ gelabelt. Von diesen Segmenten wurden nach einer Bandpass-Filterung auf 1 bis 12 Hz 13 statistische Merkmale berechnet:

- Minimum
- Maximum
- Mittelwert
- Standardabweichung
- Schiefe
- Kurtosis
- Spannweite
- Interquartilspannweite
- mittlere absolute Abweichung
- Anzahl der Nulldurchgänge
- Varianz der lokalen Minima
- Varianz der lokalen Maxima
- Mittelwerte der Signalhüllkurve<sup>8</sup>

Für fünf verschiedene Modelle des maschinellen Lernens wurden jeweils die besten Hyperparameter über Kreuzvalidierung auf den Trainingsdaten ermittelt und die Modelle anschließend mit diesen Hyperparametern trainiert. Anschließend wurden die Modelle

---

<sup>8</sup>Die Signalhüllkurve ist eine glatte Kurve, die die Extrema des Signals umreißt.

auf unbekannten Daten getestet. Das Training und Testen wurde mit getauschten Gruppen wiederholt. Das beste Ergebnis wurde mit einem Random Forest erreicht: Die durchschnittliche Genauigkeit der Kreuzvalidierung betrug 98,13 % bzw. 100 % bei getauschten Gruppen. Auf dem Testset wurde eine Genauigkeit von 92,3 % bzw. 97,99 % erreicht. Weitere Evaluationsmetriken außer eine *Confusion Matrix* für den besten Klassifikator sind nicht gegeben.

Diese Ergebnisse sind sehr gut, allerdings muss bei der Einordnung beachtet werden, dass bei für die Kreuzvalidierung die Segmente zufällig verteilt wurden und nicht beachtet wurde, dass der Algorithmus für aussagekräftige Validierung einzelne Personen nicht kennen sollte. Dadurch ist die Performance auf gänzlich unbekannten Daten weiterhin nicht bekannt und vermutlich schlechter, als die Zahlen es hier vermuten lassen.

### 3.3.3 Ähnlichkeit der Intervallschätzer des CLIE-Algorithmus

Ein weiteres Maß für die Signalqualität basiert auf dem in Kapitel 3.2 vorgestellten Algorithmus zur Intervallschätzung. Dieser Signal Quality Index misst, wie enig sich die drei Intervallschätzer sind. Wenn diese sich uneinig sind, ist der SQI bei 0, je ähnlicher sich die Schätzungen sind, desto höher ist er. Für jedes Fenster  $i$  wird er wie folgt berechnet:

$$q = \frac{E_f[n_{\text{opt}}, i]}{\sum E_f[n, i]}$$

Bei der Ermittlung der konkreten Schlag-zu-Schlag Intervalle wird von dem SQI  $q$  äquivalent zu den geschätzten Intervalllängen der Median berechnet. Schätzungen dessen Qualität unter einem Schwellwert  $q_{th}$  werden verworfen. Die Wahl von  $q_{th}$  hängt von der gewünschten Genauigkeit und Coverage ab, je höher die gewünschte Genauigkeit, desto niedriger die Coverage.

Dieser SQI wird in verschiedenen Papern mit unterschiedlich gewähltem  $q_{th}$  genutzt und erreicht damit bei der BKG-Messung in Betten während der Nacht bei gesunden Patient\*innen gute Ergebnisse. **Bruser2013** erreicht zum Beispiel bei 8 gesunden Proband\*innen durchschnittlich eine Coverage von 85 % bei einem Fehler zur EKG-Referenz von 0,61 % Schlägen pro Minute.<sup>9</sup> Bei Patient\*innen im Krankenhaus zeigte sich, dass sowohl die Ergebnisse bezüglich der Coverage als auch der Genauigkeit bedeutend schlechter sind. So erreichte **HoogAntink2020** bei 14 Patient\*innen durchschnittlich 40% Coverage bei einem Fehler von 2,16 % Schlägen pro Minute. Werden nur die nächtlichen

<sup>9</sup>**Bruser2013**.

Aufnahmen betrachtet, wird eine Coverage von 52 % bei einem Fehler von 1,29 % Schlägen pro Minute erreicht. Bei der Berechnung letztere Ergebnisse wurden zusätzlich Intervallschätzungen verworfen, deren relative Abweichung vom Median der anderen Schätzungen einen bestimmten Schwellwert überschreitet.<sup>10</sup>

Dieser Algorithmus ist der einzige der drei betrachteten, der im Bett aufgenommenes BKG-Signal untersucht. Die Ergebnisse sind bei gesunden, schlafenden Patient\*innen sehr gut, im tatsächlichen Einsatz im Krankenhaus wird allerdings eine deutlich geringere Coverage bei höherem Fehler erreicht.

## 3.4 Messdaten

Die vorliegenden Messdaten wurden in der Gefäßstation des Universitätskrankenhauses in Tampere (Finnland) aufgenommen und wurden im Paper **HoogAntink2020** bereits untersucht. Insgesamt wurden 14 Patient\*innen, zwei weiblich und zwölf männlich, bis zu 24 h überwacht. Alle hatten sich verschiedenen gefäßchirurgischen Eingriffen unterzogen. Das Durchschnittsalter der Patient\*innen betrug 69.57 Jahre und die durchschnittliche Messdauer 17.7 h, wobei sie zwischen 4.46 und 22.96 h variierte.

Gemessen wurde mit einem EMFit QS Bettsensor, der zwischen der Matratze des Krankenhausbettes und dem Bettgestell positioniert wurde. Die Abtastrate des BKG betrug 100 Hz. Das Referenz-EKG wurde mit einem Faros 360 Patientenmonitor mit einer Abtastrate von 1 kHz aufgenommen und hat drei Kanäle. Durch die unterschiedlichen Systemzeiten der beiden Geräte kam es zu einem Drift zwischen den beiden Signalen. Bei den langen Messungen zeigte sich, dass dieser zeitvariant ist.

Es liegen also BKG-Daten und ein 3-kanaliges EKG-Signal vor. Außerdem liegen schon detektierte und nach ihrer Qualität gefilterte Herzschläge mitsamt ihrer Qualität nach dem in Kapitel 3.3.3 beschriebenen SQI, der geschätzten Länge und der Länge des Referenzsignals vor. Zusätzlich gibt es zu jedem Paar von EKG und BKG-Daten einen Vektor, der den zeitvarianten Drift berücksichtigt. Zu jeder Sekunde des BKG-Signals wird die entsprechende Sekunde der EKG-Referenz benannt.

### 3.4.1 Vorverarbeitung

Die verschiedenen Daten liegen in unterschiedlichen Formaten und Dateien vor, die zunächst zusammengeführt und vorverarbeitet werden müssen. Dieser Vorgang sowie die entwickelte Datenstruktur werden im Folgenden beschrieben.

---

<sup>10</sup>**HoogAntink2020.**



Für jede\*n Patient\*in existiert eine .mat-Datei, die das rohe BKG-Signal und die bereits detektierten Intervalle mit Position, Qualität und Referenzlänge enthält. Die Vektoren zum Ausgleich des Drifts liegen ebenfalls in .mat-Dateien vor. Das Referenzsignal ist im .edf-Format gespeichert. Da die Nummerierung von BKG-Daten und EKG-Daten variiert, existiert außerdem eine Zuordnung dieser. Das Ziel der entwickelten Datenstruktur ist es, diese Daten zusammenzuführen. Die Klassenstruktur ist in Abbildung 3.6 visualisiert. Die Klasse *DataSet* enthält alle Daten, die zu einer Person gehören, also das BKG-Signal, das EKG-Signal und Vektor, um den Drift auszugleichen. Alle Patient\*innen sind in *Data* gesammelt. Dort werden auch alle Daten eingelesen und zugeordnet.

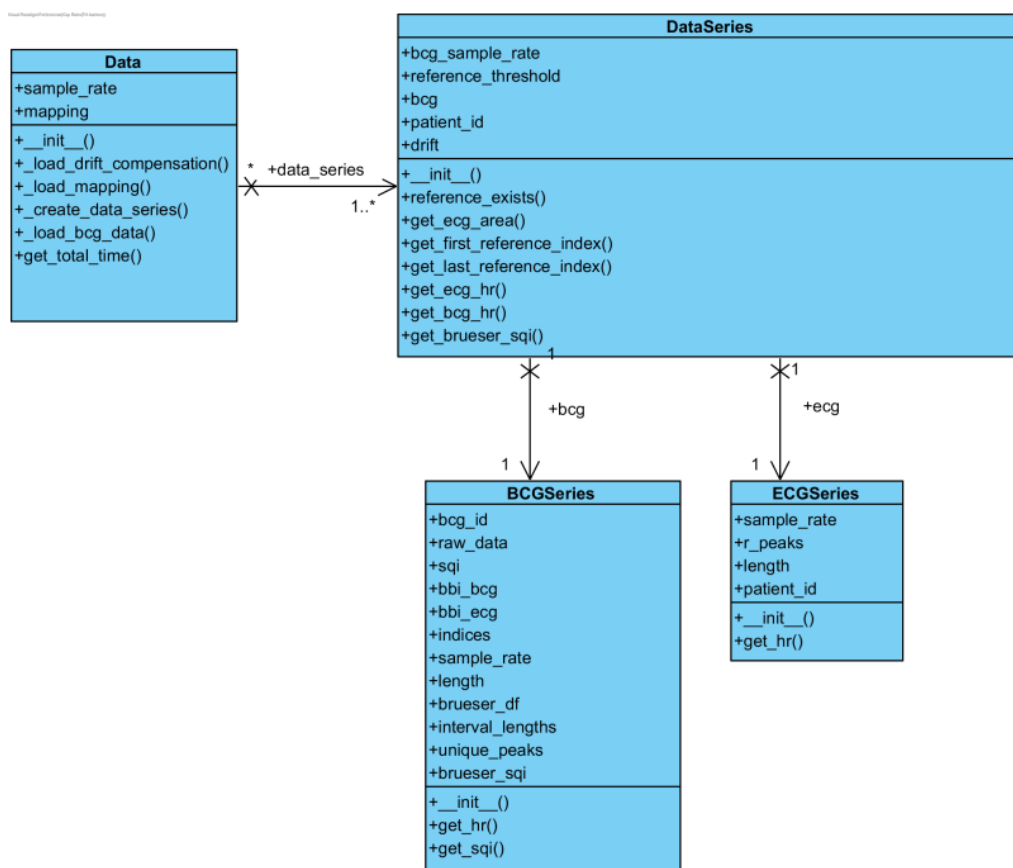


Abbildung 3.6: Klassendiagramm der Datenstruktur für die Messdaten

Um die Herzraten auf Intervallen bestimmen zu können, müssen sowohl das BKG- als auch das EKG-Signal vorverarbeitet werden. Da diese Vorverarbeitung durch die große Datenmenge sehr aufwändig ist, werden die Ergebnisse als .csv-Dateien serialisiert. Beim Laden der Daten kann so geprüft werden, ob die Berechnung wiederholt werden muss oder geladen werden kann.

Bei der Vorverarbeitung des BKG-Signals wird der in Kapitel 3.2 vorgestellte CLIE-Algorithmus angewendet und die ermittelten Indices der Intervalle, die geschätzte Inter-

vallängen und der in Kapitel 3.3.3 beschriebene Qualitätsindex gespeichert. Für das EKG-Signal wird für alle drei Kanäle eine Detektion der QRS-Komplexe durchgeführt, mit denen die Schlag-zu-Schlag Intervalle ermittelt werden können. Dafür wird die Python-Implementierung von **Howell2019** des von **Elgendi2010** beschriebenen Algorithmus<sup>11</sup> genutzt. Dieser zeigt in einer Untersuchung von **Porr2019** gute Ergebnisse.<sup>12</sup>

#### 3.4.2 Annotation der Daten

Die vorliegenden Daten sind nicht annotiert. Es ist im Rahmen dieser Arbeit nicht möglich, die Annotation durch Expert\*innen durchführen zu lassen, weshalb auf das parallel aufgenommene EKG zurückgegriffen wird. Aufgrund des nicht-linearen Drifts der Daten ist eine herzs Schlaggenaue Synchronisierung schwierig. Aus diesem Grund wurde entschieden, die Annotation bereichsweise vorzunehmen. Für jeden Bereich wird zunächst geprüft, ob ein aussagekräftiges Referenzsignal existiert. Falls nicht, wird der entsprechende Bereich ausgeschlossen und nicht weiter verwendet. Ansonsten wird die durchschnittliche Herzrate sowohl für BKG- als auch EKG-Signal berechnet. Für die Berechnung der Herzfrequenz im BKG wird der Median der geschätzten Intervalllängen verwendet. So werden Ausreißer weniger stark gewichtet. Wenn mit den geschätzten Intervalllängen eine Coverage von unter 80 % erreicht wird, wird die Herzrate verworfen, da in diesem Fall nur kleine Teile des Signals aussagekräftig sind.

Auch bei dem Referenzsignal wird die Herzrate über den Median der Intervalllängen ermittelt. Die Intervalllängen entsprechen hier den Abständen der ermittelten R-Peaks. Da drei EKG-Kanäle vorliegen, wird die Herzrate zunächst für jeden Kanal einzeln geschätzt. Falls die geschätzte Herzrate außerhalb des erwarteten Bereichs von 30 bis 200 Schlägen pro Minute liegt, wird der Kanal verworfen. Die Auswahl des Kanals gestaltet sich schwieriger, da auch im Referenzsignal teilweise schlechtes Signal vorliegt. Um dies auszugleichen, wird zur Auswahl des Kanals die Herzrate ebenfalls auf einem doppelt so großem Fenster berechnet. Von diesen drei Herzraten wird die ausgewählt, deren Kanal die geringste Spannweite an Intervalllängen aufweist. Mit Hilfe der geschätzten Herzrate des größeren Fensters wird nun der Kanal für die endgültige Schätzung ausgewählt, der die geringste Abweichung zu der Schätzung des größeren Fensters hat. Der Ablauf ist in Abbildung 3.7 visualisiert.

---

<sup>11</sup>**Elgendi2010**.

<sup>12</sup>**Porr2019**.



Abbildung 3.7: Aktivitätsdiagramm der Herzratenschätzung auf dem EKG-Referenzsignal

Die Auswirkungen dieses Verfahrens sind in Abbildung 3.8 sichtbar. Eine Fusionierung nur auf dem aktuellen Bereich führt zu einigen Fehlern in der Herzratenschätzung, die die Umgebung einbeziehende Fusionierung führt zu deutlich besseren Ergebnissen. Auch ist hier sichtbar, wie groß die Abweichung einzelner Kanäle ist.



Abbildung 3.8: EKG-Herzratenschätzungen auf sich überlappenden 10-Sekunden-Segmenten für Patient 26. Fusioniert lokal zeigt die Fusionierung ohne Miteinbeziehung der Umgebung, Fusioniert Umgebung die oben beschriebene Fusionierung

Auf Basis dieser beiden Herzraten  $HR_{\text{EKG}}$  und  $HR_{\text{BKG}}$  wird nun die Annotation vorgenommen. Dabei wurde sich an der Norm für EKG-Patientenmonitoren orientiert. Diese schreibt vor, dass die Abweichung der Herzrate maximal 10 % oder 5 Schläge pro Minute betragen darf, je nachdem welcher der beiden Werte größer ist<sup>13</sup> und wurde hier so übernommen. Wenn eine höhere oder niedrigere Genauigkeit gefordert ist, kann dies auch auf 15 % und 7,5 Schläge pro Minute oder 5 % und 2,5 Schläge pro Minute angepasst werden. Die entsprechenden Verteilungen der Labels bei einer binären Klassifikation für 10 Sekunden lange Segmente mit 90 % Überlappung für 5, 10 und 15 % maximale Abweichung sind in Abbildung 3.9 zu sehen.

<sup>13</sup>DIN.

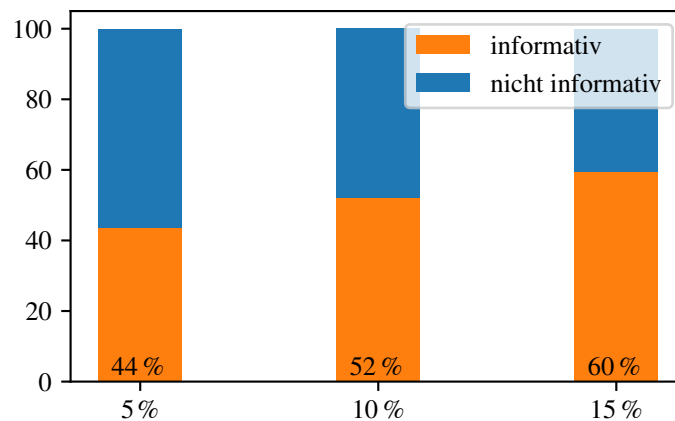


Abbildung 3.9: Verteilung der Label je nach maximal zulässiger Abweichung

Es ist auffällig, dass die Anteile der informativen Segmente stark patient\*innenabhängig sind. In Abbildung 3.10 wird dies für 10 % maximale Abweichung visualisiert.

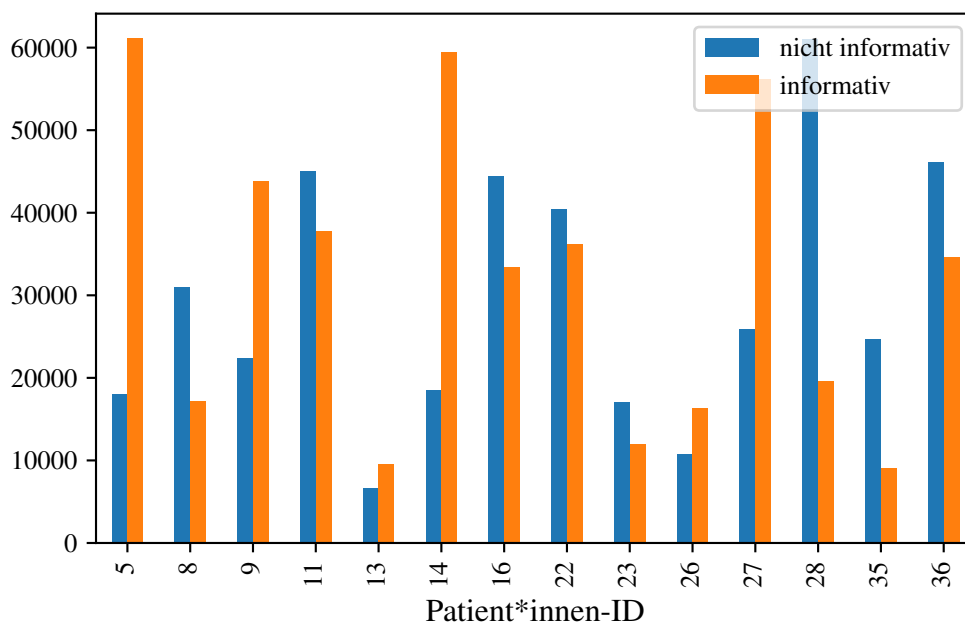


Abbildung 3.10: Verteilung der Label pro Patient\*in für maximal 10 % Abweichung

Neben einer binären Klassifikation ist es auch möglich mit Hilfe dieses Fehlermaßes, im Folgenden  $E_{HR}$  genannt, einen Wertebereich für eine Regression zu erzeugen. Dafür ist es nötig, ein einheitliches Maß für die Unterscheidung von relativer und absoluter Abweichung zu finden, das heißt den absoluten Fehler so umzurechnen, dass er mit der relativen Abweichung vergleichbar ist. Bei Herzraten über 50 Schlägen pro Minute entspricht  $E_{HR}$  dem relativen Fehler  $\Delta HR_{\text{relativ}}$ , bei Herzraten darunter ermöglicht die absolute Abweichung den größeren Spielraum und wird daher für die Annotation verwendet. Eine

Umrechnung dieser erfolgt auf Basis der Überlegung, dass der Schwellwert von 10 % relativer Abweichung einer absoluten Abweichung von 5 Schlägen die Minute entspricht. Daraus lässt sich folgern, dass eine Abweichung von 1 Schlag pro Minute einer relativen Abweichung von 2 % entspräche. Analog lässt sich  $E_{HR}$  bei Herzraten unter 50 Schlägen pro Minute durch die folgende Formel berechnen:

$$E_{HR} \hat{=} \Delta HR_{abs} \cdot 2.$$

Wenn keine Herzrate auf dem BKG-Signal ermittelt werden konnte, wird  $E_{HR}$  auf den maximalen Wert 667 gesetzt, da das der maximal möglichen relativen Abweichung von Herzraten zwischen 30 und 200 Schlägen pro Minute entspricht. Die Verteilung von  $E_{HR}$ , abgebildet in 3.11, zeigt, dass Segmente mit Fehlern von mehr als 20 % und solche mit Fehlern unter 5 % jeweils die beiden größten Gruppen bilden.

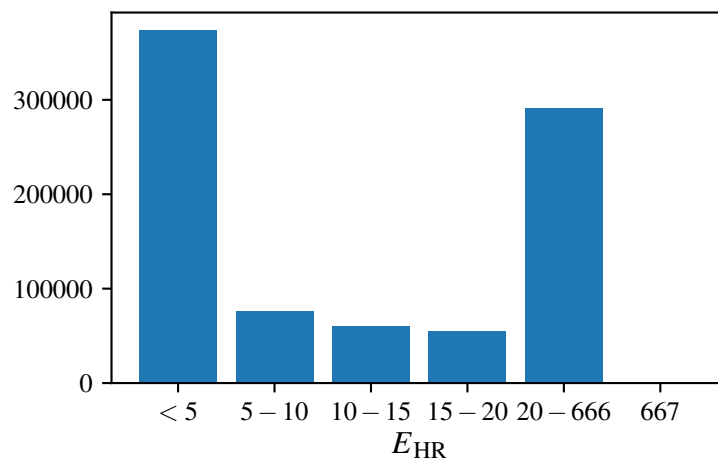


Abbildung 3.11: Verteilung von  $E_{HR}$

## 4 Analyse

In diesem Kapitel werden die vorgestellten Daten näher untersucht. Das beinhaltet sowohl die Anwendung der in Kapitel 3.3 beschriebenen existierenden Verfahren als auch eine Analyse der verwendeten Merkmale.

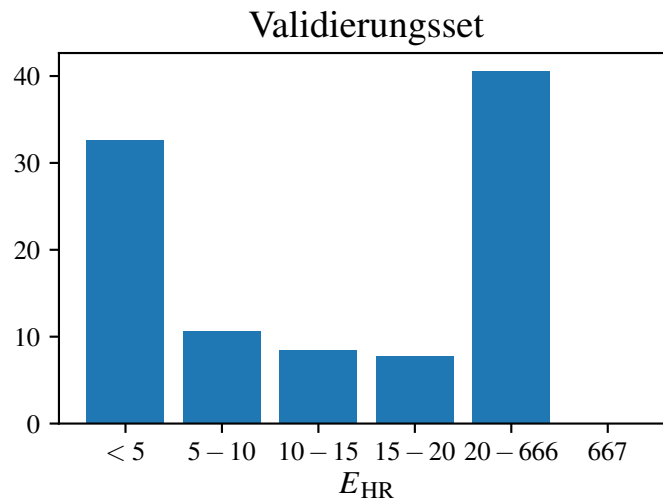
### 4.1 Aufbau und Evaluation der Verfahren

Um Modelle zur Beurteilung der Signalqualität anzuwenden und zu untersuchen, müssen sowohl die verwendeten Merkmale extrahiert als auch die Ergebnisse evaluiert werden. In den hier durchgeführten Untersuchungen wurde die Segmentlänge standardmäßig auf 10 Sekunden festgelegt und der Abstand der Segmente auf 1 Sekunde, wodurch sich die Segmente zu je 90 % überlappen. Diese Segmentlänge wird zum einen häufiger verwendet<sup>1</sup>, zum anderen bietet sie eine ausreichend große Robustheit für die Annotation der Daten. Die für die Verfahren jeweils benötigten Merkmale werden segmentweise extrahiert und serialisiert. Zusätzlich zu den verwendeten Merkmalen werden allgemeine Informationen über die Segmente gespeichert. Dazu gehören Patient\*innen-ID, EKG-Herzrate, BKG-Herzrate, absoluter und relativer Fehler,  $E_{HR}$  und die binäre Annotation. Für letztere wird als Standard  $E_{HR} = 10$  als Schwellwert verwendet.

Diese Daten werden anschließend in Trainings- und Validierungsset aufgeteilt. Um diese beiden Gruppen inhaltlich vollständig zu trennen und auszuschließen, dass auf teilweise bekannten Daten validiert wird, geschieht die Trennung anhand der Patient\*innen-IDs, sodass Segmente einer Person lediglich in einem der beiden Sets verwendet werden. Das Validierungsset in dieser Arbeit entspricht einem Drittel der Patient\*innen, die restlichen zwei Drittel werden zur Datenexploration, zur Merkmalskonstruktion und zum Training verwendet. Die Verteilung von  $E_{HR}$  im Validierungsset ist in Abbildung 4.1 gezeigt. Von den existierenden Verfahren benötigen zwei keine Trainingsphase; um vergleichbare Ergebnisse zu erhalten, wird auch bei diesen bei der Evaluation nur das Validierungsset betrachtet.

---

<sup>1</sup>Yu2020; Sadek2016; Orphanidou2015.

Abbildung 4.1: Verteilung von  $E_{HR}$  auf dem Validierungsset

Wie schon in Kapitel 3.3 beschrieben, sind bei der Verarbeitung von medizinischen Signalen zwei Bereiche wichtig: Die Coverage und die Qualität des Signals, in diesem Fall also die Genauigkeit der geschätzten Herzrate im Vergleich zur Referenz. Diese beiden müssen gegeneinander aufgewogen werden und werden aus diesem Grund beide betrachtet. Die Qualität der Modelle kann zunächst anhand der binären Klassifikation beurteilt werden. Da diese aber keine Auskunft darüber enthält, wie nah klassifiziertes Signal an dem Schwellwert für  $E_{HR}$  liegt, wird letzteres in einer tiefergehenden Evaluation ebenfalls betrachtet. Im Zuge dessen wird sowohl der Fehler auf dem als informativ klassifizierten Signal betrachtet als auch die Coverage in Bezug auf das ganze Signal für verschiedene Fehlergrößen. Insbesondere falsch klassifiziertes Signal, also Falsch-Negative und Falsch-Positive, ist interessant, um zu beurteilen, ob Fehlklassifikationen lediglich im Grenzbereich  $E_{HR} \approx 10$  oder allgemein vorkommen.

## 4.2 Anwendung existierender Verfahren

Zunächst werden die in Kapitel 3.3 beschriebenen existierenden Artefakterkennungsverfahren mit den in dieser Arbeit untersuchten Daten wie oben beschrieben getestet und ihre Leistungsfähigkeit untersucht und bewertet.

### 4.2.1 Ähnlichkeit der Intervallschätzer des CLIE-Algorithmus

Der SQI, der die Ähnlichkeit der Intervallschätzer des CLIE-Algorithmus angibt, wird im Normalfall herzs Schlagweise angewendet. Da die Datenannotation nur bereichsweise vorge-



nommen werden kann, wurde entschieden, ein Segment als informativ zu klassifizieren, wenn mit den Herzschlägen, deren SQI über einem Schwellwert  $q_{th}$  liegt, eine Coverage über einem gegebenen Schwellwert  $c_{th}$  auf dem Segment erreicht wird. Getestete Schwellwerte für die Coverage sind 50 %, 75 % und 100 %. Zum Testen des Algorithmus wurde unter anderem  $q_{th} = 0,4$  gewählt, da dieser Wert auch von **Zink2017** verwendet wird.<sup>2</sup> Zusätzlich wurden  $q_{th} = 0,3$  und  $q_{th} = 0,2$  untersucht, um den Einfluss von  $q_{th}$  einzuordnen. Bei der Berechnung der Merkmale werden für jedes Segment die detektierten Herzschläge extrahiert, deren SQI über  $q_{th}$  liegt. Auf Basis dieser Intervalllängen wird wie in Kapitel 3.4.2 beschrieben die Herzrate, im Folgenden  $HR_{SQI}$  genannt, ermittelt. Extrahierte Merkmale sind damit in diesem Fall  $HR_{SQI}$  und die Coverage  $C_{SQI}$ . Für die Auswertung muss beachtet werden, dass sich die ermittelte Herzrate  $HR_{SQI}$  von der zur Annotation verwendeten Herzrate unterscheidet. Aufgrund der Vergleichbarkeit wird  $HR_{SQI}$  nicht für die Berechnung des Mean Absolute Error (MAE) des Algorithmus verwendet.

Bei einer ersten Betrachtung von MAE, Coverage und Accuracy wird sichtbar, dass die Wahl der Schwellwerte großen Einfluss auf das Ergebnis hat und in jedem Fall Aussagekraft des Signals gegen Coverage eingetauscht wird. Die genauen Ergebnisse sind in Abbildung 4.2. Auch ist deutlich, dass die Accuracy der Klassifikation mit Werten knapp über 0,6 nicht gut ist. Des Weiteren ist die Coverage bei verhältnismäßig kleineren MAE sehr niedrig.

	MAE	Coverage	Accuracy
insgesamt	21,84	-	-
annotiert	3,28	43,22 %	-
$q_{th} = 0,4, c_{th} = 50$	12,30	20,95 %	0,65
$q_{th} = 0,4, c_{th} = 75$	9,57	11,78 %	0,63
$q_{th} = 0,4, c_{th} = 100$	8,76	5,26 %	0,60
$q_{th} = 0,3, c_{th} = 50$	21,35	58,73 %	0,55
$q_{th} = 0,3, c_{th} = 75$	17,97	37,87 %	0,62
$q_{th} = 0,3, c_{th} = 100$	13,44	19,25 %	0,63
$q_{th} = 0,2, c_{th} = 75$	21,38	96,43 %	0,44

Abbildung 4.2: Fehler und Coverage für verschiedene Schwellwerte im Vergleich zum gesamten Signal und der Annotation

Die detailliertere Evaluation wird beispielhaft für die Schwellwerte  $q_{th} = 0.4$  und  $q_{th} = 0.3$  mit  $c_{th} = 75$  vorgestellt. Positiv hervorzuheben ist, dass bei beiden kein Signal als informativ klassifiziert wird, bei denen  $E_{HR}$  maximal ist. Allerdings ist bei über 20 % der mit  $q_{th} = 0.4$  als informativ klassifizierten Segmenten  $E_{HR}$  größer als 20, mit  $q_{th} = 0.3$  bei

<sup>2</sup>Zink2017.

sogar mehr als 35 %, wie auch in Abbildung 4.3 abgebildet. Das bedeutet, dass Falschklassifikationen nicht nur im Randbereich vorkommen. Dies wird noch deutlicher, wenn man sich den MAE auf jeweils auf den Falsch-Positiven und Falsch-Negativen anschaut. So beträgt der MAE für falsch-negative Segmente 3,64, ist also nah an dem MAE aller informativen Segmente von 3,28.

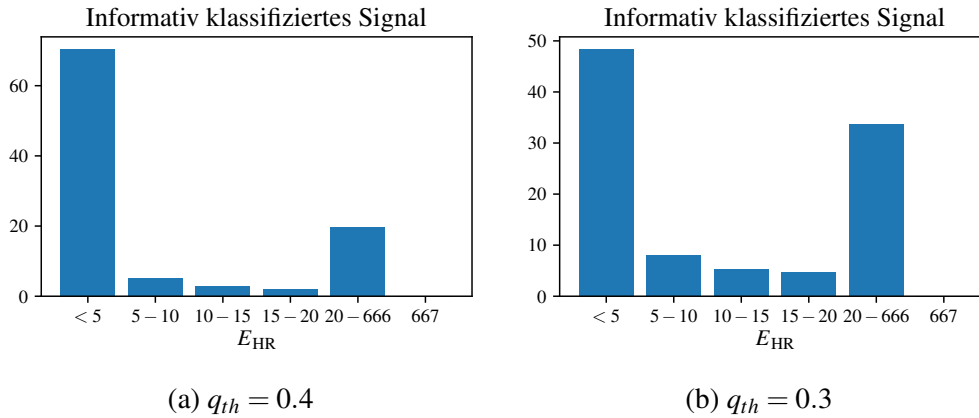


Abbildung 4.3: Verteilung von  $E_{HR}$  bei den als informativ klassifizierten Segmenten

Außerdem wurde untersucht, wie hoch die Coverage unter einem bestimmten Fehler  $E_{HR}$  auf dem gesamten Signal ist, wenn nur die als informativ klassifizierten Segmente verwendet werden. Hier wird besonders sichtbar, wie niedrig die erreichte Coverage ist. So werden mit  $q_{th} = 0,4$  und  $c_{th} = 75$  nur 5,42 % Coverage für  $E_{HR} < 5$  erreicht werden, obwohl das ganze Signal 32,61 % enthält. Mit  $q_{th} = 0,3$  und  $c_{th} = 75$  sind es immerhin 8,82 %, aber auch das liegt deutlich unter dem tatsächlichen Wert. Die Verteilung ist auch in Abbildung 4.4 gezeigt.

	insgesamt	$q_{th} = 0,3, c_{th} = 75$	$q_{th} = 0,4, c_{th} = 75$
$E_{HR} < 5$	32,61 %	8,82 %	5,42 %
$E_{HR} < 10$	43,22 %	15,33 %	7,39 %
$E_{HR} < 15$	51,66 %	20,44 %	8,39 %
$E_{HR} < 20$	59,40 %	24,60 %	9,18 %

Abbildung 4.4: Coverage unter bestimmten Fehlern  $E_{HR}$  vor und nach Klassifikation

Alles in allem zeigt sich, dass trotz einer sehr niedrigen erreichten Coverage der Fehler in der Klassifikation verhältnismäßig hoch ist. Dieses Verfahren ist also bei dem hier vorliegenden Signal nicht ausreichend. Wenn  $q_{th}$  und  $c_{th}$  nicht zu niedrig gewählt werden, kann ein Teil des nicht informativen Signals markiert und somit der Fehler in weiterer Verarbeitung insgesamt reduziert werden.

### 4.2.2 Schwellwerte für Standardabweichung, Minimum und Maximum

Für das zweite getestete Verfahren werden lediglich zwei Schwellwerte  $T_1$  und  $T_2$  benötigt, die auf Standardabweichung, Minimum, Maximum und Durchschnitt des Signals beruhen. Die Klassifikation ist ein einfacher Vergleich. **Pino2015** verwenden sehr kleine Fenster, die hier aufgrund fehlender Robustheit der Annotation nicht verwendet werden können.<sup>3</sup> Stattdessen wird der Algorithmus sowohl auf 10 Sekunden- als auch 4 Sekunden-Segmenten getestet, damit sichtbar wird, ob die Segmentlänge einen Einfluss auf das Ergebnis hat.

Die Tests mit beiden Segmentlängen zeigen, dass dieses Verfahren für die in dieser Arbeit untersuchten Daten nicht nutzbar ist, da bei beiden über 95 % der Daten als informativ klassifiziert werden, darunter auch Segmente, bei denen  $E_{HR}$  maximal ist. Weiterführende Evaluation bietet hier keine weiteren Erkenntnisse. Dieses Verfahren ist damit nicht weiter nutzbar.

### 4.2.3 Maschinelles Lernen mittels statistischer Merkmale

Für das maschinelle Lernen mittels statistischer Merkmale werden die in Kapitel 3.3.2 aufgezählten Merkmale extrahiert. Als Bibliothek für die Modelle des maschinellen Lernens wird *scikit-learn*<sup>4</sup> verwendet. Da die Daten sich grundlegend von den von **Sadek2016** untersuchten unterscheiden, werden die Hyperparameter der Modelle im Zuge dieser Arbeit erneut optimiert. Bei der dafür durchgeführten Kreuzvalidierung wird wie schon bei der Unterteilung in Trainings- und Validierungsset anhand der Patient\*innen-ID geteilt, damit auch diese aussagekräftige Ergebnisse liefert.

Insgesamt zeigen bereits Accuracy und MAE, siehe Abbildung 4.5, eindeutig, dass die Klassifikation mit keinem der Modelle sehr erfolgreich ist. Die Accuracy von im besten Fall 0,57 zeigt, dass die Klassifikation nur minimal besser als reines Raten ist. Teils ist der MAE der als informativ klassifizierten Segmente sogar größer als der auf dem gesamten Signal. In der folgenden tiefergehenden Evaluation werden lediglich RF und Multilayer-Perzeptron (MLP) betrachtet. Zwar sind die Ergebnisse des Decision Tree (DT) auf einem ähnlichen Niveau, aber da ein RF lediglich ein Zusammenschluss mehrerer DT ist, ist eine zusätzliche Analyse nicht lohnenswert.

---

<sup>3</sup>**Pino2015**.

<sup>4</sup>**scikit-learn**.

	MAE	Coverage	Accuracy	Accuracy der Kreuzvalidierung
insgesamt	21,84	-	-	-
annotiert	3,28	43,22 %	-	-
LDA	21,84	84,81 %	0,49	0,51
SVM	22,06	88,14 %	0,47	0,55
DT	19,39	47,20 %	0,53	0,52
RF	17,66	44,20 %	0,57	0,52
MLP	18,41	42,04 %	0,56	0,54

Abbildung 4.5: Fehler und Coverage für die verschiedenen Modelle im Vergleich zum gesamten Signal und der Annotation

Bei einer ähnlich hohen Coverage wie der Annotation ist der deutlich höhere MAE der Klassifikation in falsch-positiven Segmenten begründet. Der Anteil von Segmenten mit einem Fehler  $E_{HR}$  größer als 20 ist bei beiden Modellen mit ca. 35 % ähnlich hoch. Zusätzlich sind 0,006 % der vom DT als informativ klassifizierten Segmente solche, bei den  $E_{HR}$  maximal ist. Das bedeutet, dass die falsch-positiv klassifizierten Segmente bei beiden Klassifikatoren keine Randfälle sind. Der MAE der falsch-negativen Klassifikationen beträgt beim RF 3,48 und beim MLP 3,39, ist also zwar leicht höher als der MAE aller informativen Segmente, aber nicht relevant. Auch bei diesen handelt es sich also nicht um Randfälle.

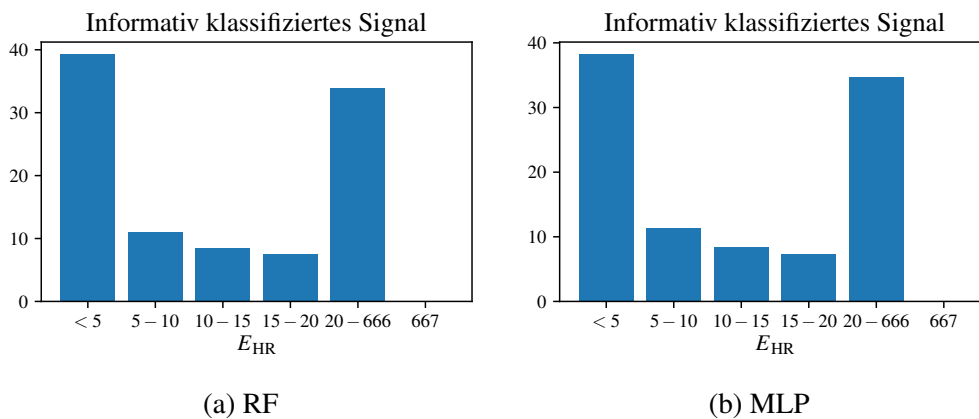


Abbildung 4.6: Verteilung von  $E_{HR}$  bei den als informativ klassifizierten Segmenten

Auch für diese Modelle wird untersucht, wie hoch die Coverage unter einem bestimmten Fehler  $E_{HR}$  bei der Betrachtung des als informativ klassifizierten Signals auf dem ganzen Signal ist. Auffallend ist, dass die Werte knapp halb so hoch wie die tatsächliche Coverage sind. Das zeigt ergänzend zu der gemessenen Accuracy, dass ca. die Hälfte der Segmente, unabhängig von ihren tatsächlichen Qualität, als informativ klassifiziert werden.

	insgesamt	RF	MLP
$E_{HR} < 5$	32,61 %	17,35 %	16,07 %
$E_{HR} < 10$	43,22 %	22,20 %	20,82 %
$E_{HR} < 15$	51,66 %	25,92 %	24,34 %
$E_{HR} < 20$	59,40 %	29,23 %	27,45 %

Abbildung 4.7: Coverage unter bestimmten Fehlern  $E_{HR}$  vor und nach Klassifikation

Insgesamt zeigt sich, dass die Klassifikation mittels statistischer Merkmale bei vorliegenden Daten und Annotation nicht erfolgreich ist. Zwar wurden auf anderen Daten sehr gute Ergebnisse erreicht, aber es bestätigt sich die Vermutung, dass sich dies, vermutlich aufgrund der unterschiedlichen Aufnahmesituation der Daten und der nicht aussagekräftigen Validierung der Ergebnisse für die im Paper untersuchten Daten, nicht übertragen lässt.

## 4.3 Analyse der Merkmale

Obwohl keines der Verfahren für die vorliegenden Daten zufriedenstellende Ergebnisse erzielt, ist besonders bei den Modellen maschinellen Lernens interessant, welchen Einfluss und welchen Informationsgewinn die jeweiligen Merkmale haben bzw. liefern. Aus diesem Grund wird im Folgenden eine explorative Datenanalyse durchgeführt.

Zunächst wird betrachtet, ob die Merkmale untereinander und mit  $E_{HR}$  und der Annotation korreliert sind. Da die Merkmale das Segment gemeinschaftlich statistisch beschreiben ist eine hohe Korrelation untereinander erwartet und zeigt sich auch. Eine Ausnahme bilden Kurtosis, Mittelwert und Schiefe. Außerdem zeigt sich bei der Betrachtung des in Abbildung 4.8 abgebildeten Korrelationsdiagramm, dass die Korrelation zu  $E_{HR}$  und der Annotation nicht signifikant ist. Auch bei paarweiser Visualisierung zeigt sich, dass sich anhand keines untersuchten Merkmalspaars informative von nicht informativen unterscheiden lassen.

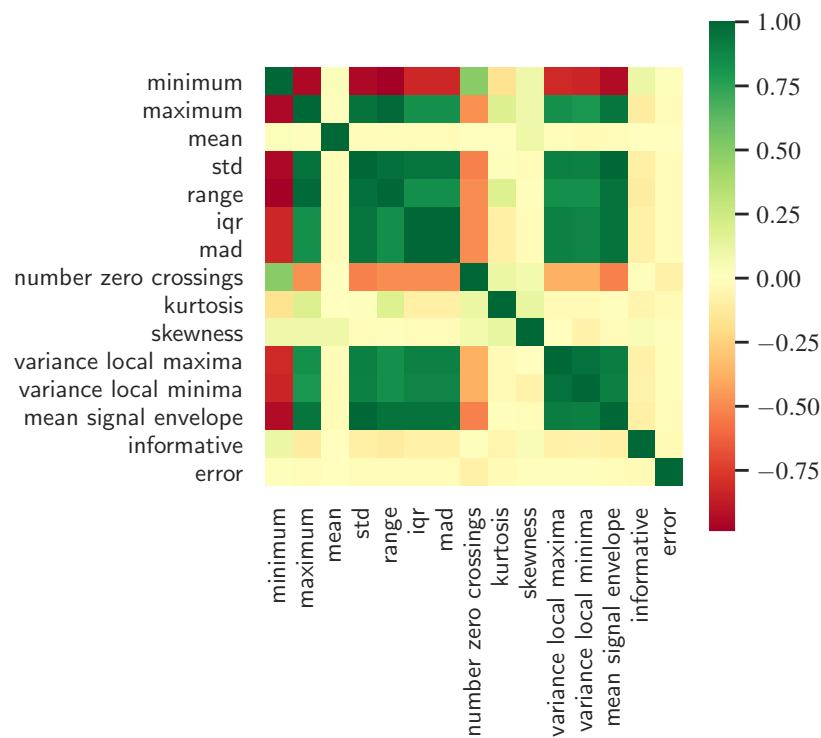


Abbildung 4.8: Korrelationsdiagramm der statistischen Merkmale,  $E_{HR}$  und der binären Annotation

Da die Visualisierung hochdimensionaler Daten schwierig ist und auch Modelle maschinellen Lernens bei einer hohen Merkmalszahl aufwändiger zu trainieren sind, wurde ebenfalls untersucht, wie sich die Daten bei einer Transformation in einen zweidimensionalen Raum verhalten. Untersuchte Transformationen umfassen sowohl unüberwachte Verfahren wie PCA mit verschiedenen Kernel als auch überwachte Verfahren wie eine Dimensionsreduktion mit einer Linear Discriminant Analysis (LDA). Auch die transformierten Daten lassen sich nicht voneinander trennen, wie auch in Abbildung 4.9 beispielhaft für eine PCA mit linearem Kernel gezeigt ist.

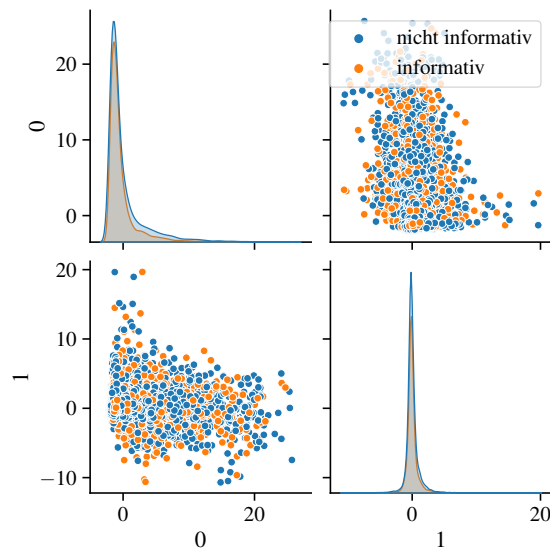


Abbildung 4.9: Dimensionsreduktion der statistischen Merkmale mit einer PCA mit linearem Kernel

Weiteren Einblick ermöglicht die Analyse des RF, da bei diesem ermittelt werden kann, wie wichtig die einzelnen Merkmale jeweils für die Entscheidungsfindung sind. Das Ergebnis zeigt, dass der Durchschnitt des Signals deutlich weniger Einfluss als die anderen betrachteten Merkmale hat. Die Schiefe (skewness) und die Anzahl der Nulldurchläufe sind, wie in Abbildung 4.10 zu sehen, die beiden wichtigsten Merkmale.

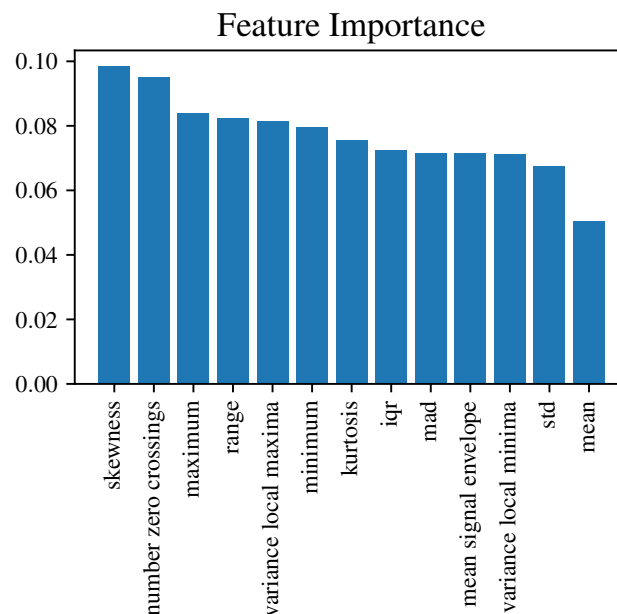


Abbildung 4.10: Wichtigkeit der Merkmale für den RF-Klassifikator

Anhand der statistischen Merkmale kann also keine Klassifikation vorgenommen werden.

Vermutlich sorgt die große Variation der in Betten aufgenommenen BKG-Signale dafür, dass sich die Signale nicht statistisch verallgemeinern lassen. Auch die reine Betrachtung der Ähnlichkeit der Intervallschätzer ist nicht ausreichend. Es müssen also andere Merkmale gefunden werden, die allein oder ergänzend zu den bereits betrachteten eine Aussage zu der Signalqualität ermöglichen.



## **5 Synthese**

### **5.1 Eingabeform der Daten**

### **5.2 Evaluationsmetriken**

### **5.3 Modelle**



## **6 Evaluierung der Ergebnisse**



## **7 Zusammenfassung und Ausblick**

### **7.1 Zusammenfassung**

### **7.2 Ausblick**