

**Konzeption eines Verfahrens zur ähnlichkeitsbasierten Cluster-Analyse von  
Vitaldaten und dessen Validierung am Beispiel von Schmerzmittelgaben nach  
herzchirurgischen Eingriffen**

Dissertation

zur

Erlangung des akademischen Grades

Doktor-Ingenieur (Dr.-Ing.)

der Fakultät für Informatik und Elektrotechnik  
der Universität Rostock

vorgelegt von:

Fabian Schrumpf, geboren am 23. August 1985 in Gera  
aus Leipzig

Gutachter:

Prof. Dr.-Ing. Matthias Sturm (HTWK Leipzig)

Prof. Dr.-Ing. habil. Helmut Beikirch (Universität Rostock)

Prof. Dr. med. Joerg Seeburger (Sana Herzchirurgie Stuttgart)

Prof. Dr. rer. net. habil. Hartmut Ewald (Universität Rostock)

Datum der Einreichung 6. Mai 2019

Datum der Verteidigung: 25. Juni 2020



**Dieses Werk ist lizenziert unter einer  
Creative Commons Namensnennung - Weitergabe unter gleichen Bedingungen 4.0  
International Lizenz.**

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.





## **Vorwort**

Diese Dissertation entstand im Rahmen einer kooperativen Promotion an der Hochschule für Technik, Wirtschaft und Kultur Leipzig und der Universität Rostock. Thematisch war sie im BMBF-Projekt „Nichtinvasive Erfassung und Analyse von Biosignalen zur unterstützenden Bewertung kritischer Vitalzustände“ angesiedelt.

Die wissenschaftliche Betreuung dieser Arbeit erfolgte durch Prof. Dr.-Ing. Helmut Beikirch an der Universität Rostock. Ihm gilt besonderer Dank für die Ermöglichung meiner Promotion unter seiner Betreuung. Besonders bedanken möchte ich mich auch bei Prof. Dr.-Ing. Matthias Sturm für die Unterstützung und Motivation sowie die Möglichkeit, meine Promotion am Laboratory for Biosignal Processing durchzuführen. Nützliche Anregungen bei der Datenauswertung, hilfreiche Diskussionen sowie kleinere und größere Zusätze habe ich auch dem Team des LaBP zu verdanken. Herauszustellen sind hier vor allem Dr.-Ing. Mirco Fuchs, aber natürlich auch allen anderen Mitarbeiter und Studenten der Arbeitsgruppe.

Ein großer Teil dieser Arbeit basiert auf Daten, welche im Rahmen einer Messstudie am Herzzentrum Leipzig erhoben wurden. Ich möchte mich an dieser Stelle bei Prof. Dr. med. Joerg Seeburger und Dr. med. Fabian Emrich dafür bedanken, dies ermöglicht zu haben. Ohne ihre Zusammenarbeit und tatkräftige Unterstützung bei der Studienbeantragung und der anschließenden Datenerhebung wäre die Arbeit in dieser Form nicht möglich gewesen.

Ich möchte mich bei allen Freunden und Familienangehörigen bedanken, die vor allem in der Endphase der Entstehung dieser Arbeit eine große Hilfe waren, sei es durch moralische Unterstützung, die Toleranz meiner gelegentlichen Unzufriedenheit oder das unermüdliche Durchforsten dieses Textes nach Fehlern unterschiedlichster Art. Ohne euch wäre so mancher Tippfehler unerkannt, so manche Formulierung unkorrigiert geblieben. Vielen Dank für eure Geduld und Unterstützung.

Leipzig, im Mai 2019

Fabian Schrumpf



## **Kurzfassung**

Maschinelle Lernverfahren gewinnen in der Gesundheitsversorgung zunehmend an Bedeutung. Im diagnostischen Bereich haben solche Verfahren vor allem bei der medizinischen Bildverarbeitung bereits Einzug gehalten. Im Bereich der Patientenüberwachung nach chirurgischen Eingriffen wird jedoch meist mit einfachen schwellwertbasierten Algorithmen gearbeitet. Dieses Vorgehen ist mit einer großen Anzahl an falschen Alarmen und einer hohen Arbeitsbelastung des Krankenhauspersonals verbunden.

In dieser Arbeit wird deshalb ein neues unüberwachtes maschinelles Lernverfahren entworfen, welches Gesundheitszustände anhand von Vitalparametern erkennen kann. Jeder Gesundheitszustand zeichnet sich durch einen charakteristischen Wertbereich der Vitalparameter aus. Das Verfahren fasst mittels eines iterativen Algorithmus bestehende Zustände zu neuen Gesundheitszuständen anhand ihrer Ähnlichkeit zusammen. Dadurch entsteht eine baumartige Struktur, die eine Unterteilung eines Vitalparameterdatensatzes mit unterschiedlicher zeitlicher Granularität ermöglicht. Die Validierung der Methode wird anhand von simulierten und realen Daten durchgeführt.

Die Anwendbarkeit der Methode im klinischen Umfeld wird anhand der Detektion von Schmerzmittelgaben nach herzchirurgischen Eingriffen demonstriert. Dazu wird eine Messstudie durchgeführt, in der Patienten für 24 Stunden nach der Operation durch zusätzliche Messtechnik überwacht wurden. Ziel ist die Rekonstruktion des Zeitpunktes von protokollierten Schmerzmittelgaben anhand verschiedener Vitalparameter mit Hilfe der neuen Methode. Die gewonnenen Resultate werden mit den Ergebnissen etablierter unüberwachter Lernverfahren verglichen.

Im Ergebnis steht ein hierarchisches Cluster-Verfahren zur Verfügung, das für die Patientenüberwachung im klinischen Umfeld angewendet werden kann. Die Resultate der Messstudie zeigen, dass die Rekonstruktion von Schmerzmittelgaben mit der hier vorgestellten Methode möglich ist. Im Vergleich zu den Referenzverfahren werden vergleichbare und teilweise bessere Ergebnisse erzielt.





## **Abstract**

Although machine learning methods become more important in clinical decision making in general, they are still rarely used in patient supervision. In fact, threshold-based algorithms that alert the clinical staff if a parameter exceeds a certain predefined range are still common practice in patient monitoring. This leads to an increased false alarm rate and, as a consequence, to an increased workload of nurses and doctors.

In this thesis, a new unsupervised machine learning method is introduced that detects health states in patients based on multiple vital parameters. Every health state is characterized by a distinctive range of vital parameters. By using an iterative algorithm, the method combines health states according to a similarity constraint into new arbitrary health states. This results in a tree-like structure that allows for a division of datasets into clusters of varying temporal granularity. Artificial and real-world datasets are used to validate this method.

To demonstrate the applicability of this method in a clinical setting, an observational study has been conducted. Various vital parameters were recorded from patients who underwent cardiac surgery. The new method has been applied in order to reconstruct the administration of analgesics based on these vital parameters. The aim was to detect a change in a patient's health state that was relatable to a recorded administration of analgesic agents. The results obtained with the new algorithm were compared to the results obtained with two established unsupervised learning methods.

The results show that this new hierarchical clustering algorithm is applicable to patient monitoring in a clinical setting. It could be demonstrated that this method is capable of detecting the administration of analgesics based on multiple vital parameters. The obtained results are similar to or better than those acquired with the help of the reference methods.



# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>15</b>
<b>Tabellenverzeichnis</b>	<b>19</b>
<b>Abkürzungsverzeichnis</b>	<b>21</b>
<b>1 Einleitung</b>	<b>23</b>
1.1 Maschinelles Lernen in der Gesundheitsversorgung . . . . .	23
1.2 Aufbau der Arbeit . . . . .	25
<b>2 Forschungsstand</b>	<b>27</b>
2.1 Maschinelles Lernen – Systematischer Überblick . . . . .	27
2.2 Patientenüberwachung – aktueller Stand der Technik . . . . .	29
2.3 Maschinelles Lernen im klinischen Umfeld . . . . .	32
2.4 Referenzverfahren . . . . .	37
2.4.1 Agglomeratives hierarchisches Clustering . . . . .	37
2.4.2 k-means Methode . . . . .	37
2.5 Zusammenfassung und offene Problemstellungen . . . . .	39
<b>3 Ähnlichkeitsbasiertes hierarchisches Clustering von Gesundheitszuständen</b>	<b>41</b>
3.1 Allgemeine Funktionsweise des Algorithmus . . . . .	41
3.2 Ähnlichkeit und Abstandsmaße . . . . .	43
3.2.1 Berechnung der Ähnlichkeit von Gesundheitszuständen . . . . .	43
3.2.2 Höherdimensionale Trennung von Gesundheitszuständen mittels Kernel-Funktionen . . . . .	44
3.3 Cluster-Fusionierungsalgorithmus . . . . .	47
3.4 Cluster-Validitäts Koeffizienten . . . . .	49
3.4.1 Übersicht . . . . .	49
3.4.2 Silhouettenkoeffizient . . . . .	49
3.5 Validierung der Methode anhand von Simulationsdaten . . . . .	51
3.5.1 Simulationsdatensatz und Cluster-Analyse . . . . .	51

3.5.2	Ergebnisse . . . . .	52
3.5.3	Diskussion . . . . .	54
3.6	Validierung der Methode anhand realistischer Daten . . . . .	56
3.6.1	Experimentelles Setup und Datenauswertung . . . . .	56
3.6.2	Ergebnisse . . . . .	57
3.6.3	Diskussion . . . . .	59
3.7	Zusammenfassung . . . . .	61
<b>4</b>	<b>Detektion von Schmerzmittelgaben nach herzchirurgischen Eingriffen</b>	<b>63</b>
4.1	Methode . . . . .	63
4.1.1	Studiendesign . . . . .	63
4.1.2	Verwendete Messtechnik und Messaufbau . . . . .	64
4.1.3	Vorverarbeitung der Biosignale . . . . .	65
4.1.4	Aus den Biosignalen abgeleitete Vitalparameter . . . . .	66
4.1.5	Clustering der Vitalparameter . . . . .	77
4.1.6	Evaluation der Cluster-Lösungen . . . . .	79
4.2	Ergebnisse . . . . .	82
4.2.1	Nutzung von Vitalparametern zur Klassifikation von Gesundheitszuständen . . . . .	82
4.2.2	Ergebnisse der Detektion von Schmerzmittelgaben . . . . .	83
4.2.3	Detektion von Schmerzmittelgaben am Beispiel . . . . .	85
4.3	Diskussion . . . . .	89
4.3.1	Detektion von Schmerzmittelgaben . . . . .	89
4.3.2	Datenerhebung und Vitalparameterextraktion . . . . .	91
4.3.3	Zusammenfassung . . . . .	92
<b>5</b>	<b>Erkennung von obstruktiver Schlafapnoe basierend auf EKG-Parametern</b>	<b>95</b>
5.1	Einleitung . . . . .	95
5.2	Methode . . . . .	96
5.2.1	Verwendete Datenquelle und extrahierte Vitalparameter . . . . .	96
5.2.2	Cluster-Analyse . . . . .	97
5.2.3	Evaluierung und Validierung der Clustering-Ergebnisse . . . . .	98
5.3	Ergebnisse . . . . .	99
5.4	Diskussion . . . . .	101

<b>6</b>	<b>Zusammenfassung und Bewertung der Ergebnisse</b>	<b>103</b>
6.1	Zusammenfassung . . . . .	103
6.2	Bewertung und Einordnung der Ergebnisse bei der Erkennung von Gesundheitszuständen . . . . .	105
<b>7</b>	<b>Ausblick</b>	<b>109</b>
	<b>Literaturverzeichnis</b>	<b>113</b>
<b>A</b>	<b>Evaluierung der EKG-basierten Atemfrequenzbestimmung</b>	<b>131</b>
A.1	Einleitung . . . . .	131
A.2	Methode . . . . .	131
A.3	Ergebnisse . . . . .	133
A.4	Diskussion . . . . .	134
<b>B</b>	<b>Tabellen und Abbildungen</b>	<b>137</b>



## Abbildungsverzeichnis

1.1	Wordcloud bestehend aus den Schlagwörtern zum Thema maschinelles Lernen und dessen klinischen Anwendungsgebieten [1]. . . . .	24
2.1	Anzahl der jährlichen Publikationen, welche die Stichworte „Machine Learning“ oder „Artificial Intelligence“ enthalten [18]. . . . .	28
2.2	Systematik der Algorithmen des maschinellen Lernens und einige konkrete Beispiele. . . . .	29
2.3	Links: Beispieldatensatz $\mathbf{X}$ ( $\mathbf{X} \in \mathbb{R}^{13 \times 2}$ ) bestehend aus 13 Samples, die durch 2 Features beschrieben werden. Die Cluster-Zugehörigkeit ist farblich hervorgehoben. Rechts: Dendrogramm der hierarchischen Zerlegung des Datensatzes links. Die zugrunde liegende Cluster-Struktur ist farblich hervorgehoben. . . . .	38
3.1	Links: Baumstruktur, welche bei der iterativen Einteilung eines Datensatzes in Cluster entsteht Rechts: Iterative Einteilung der Samples in Cluster anhand ihrer Ähnlichkeit. . . . .	41
3.2	Übersicht über den Ablauf des Algorithmus für die ähnlichkeitsbasierte hierarchische Cluster-Methode als UML-Aktivitätsdiagramm . . . . .	42
3.3	Verschiedene Verfahren zur Abstandsberechnung: a) Complete Linkage, b) Single Linkage, c) Average Linkage, d) Median Methode . . . . .	47
3.4	Validierungsdatensätze erzeugt aus fünf bivariaten Gauß-Verteilungen mit variierender Überlappung . . . . .	51
3.5	Ergebnisse des Clusterings der simulierten Datensätze. Links: Optimale Cluster-Einteilung der Datensätze aus Abbildung 3.4, basierend auf SI-Koeffizienten (Tabelle 3.1). Die Marker-Form gibt die tatsächliche Cluster-Zugehörigkeit der Samples an. Rechts: zugehörige Silhouetten-Plots der optimalen Cluster-Lösungen. . . . .	53
3.6	Ablauf des Validierungsexperiments besteht aus zwei Ruhephasen von jeweils $T_{R1} = 5$ min und $T_{R2} = 5$ min Länge sowie einer Belastungsphase der Länge $T_B = 3$ min. . . . .	57

3.7	Cluster-Ergebnisse aller berücksichtigten Probanden. Der Übergang in eine andere experimentelle Phase ist in Schwarz dargestellt. Erkannte Ruhezustände (dunkelblau) und Belastungszustände (rot) sind farblich dargestellt. Darüber hinaus erkannte (Zwischen-) Zustände sind andersfarbig markiert (siehe auch Abb. 3.8). . . . .	58
3.8	Verteilung der Parameter HR und Q-TE in den erkannten Gesundheitszuständen für die Probanden <i>ptt002</i> und <i>ptt009</i> : Die Farben korrespondieren dabei mit den Farben der Gesundheitszustände in Abbildung 3.7. . . . .	59
4.1	Messaufbau zur Aufzeichnung von Biosignalen bei der Durchführung der Messstudie am Herzzentrum Leipzig. . . . .	65
4.2	Überblick über alle Verarbeitungsschritte der während der Messstudie am Herzzentrum Leipzig aufgezeichneten Biosignale. . . . .	67
4.3	Extraktion des ANI aus dem Tachogramm. . . . .	70
4.4	Oben: Baseline-korrigiertes EKG-Signal mit markierten R- und S-Peaks zur Bestimmung der R-Peak Amplitude sowie der RR-Intervalle. Unten: Atemsignal basierend auf der Modulation der R-Peak-Amplitude (gelb) sowie auf Basis der Variation der RR-Intervalle (blau). . . . .	72
4.5	Links: Mit höherfrequenten Störungen überlagertes Atemsignal (Beispiel). Rechts: Autokorrelationsfunktion dieses Atemsignals. Durch hochfrequente Störungen entstehen zusätzliche lokale Maxima (rot markiert), die die Schätzung der Atemfrequenz verfälschen. . . . .	73
4.6	Beispielhafte Darstellung der PTT: Zeit zwischen der elektrischen Erregung des Herzens (rote Markierung in der oberen Abbildung) und der Ankunft der Pulswelle an der Körperperipherie. Der maximale Anstieg des Finger-PPG (rote Markierung untere Abbildung) wird als Ankunft der Pulswelle definiert.	74
4.7	Zerlegung des EDA-Signals (schwarz) in einen niederfrequenten tonischen Anteil (orange) und einen höherfrequenten phasischen Anteil (blau). Gelb: Durch Dekonvolution geschätztes Treibersignal, welches mit der sudomotorischen Aktivierung der Schweißdrüsen assoziiert ist. . . . .	76
4.8	Werte Verteilung der zur Detektion von Gesundheitszuständen verwendeten Vitalparameter am Beispiel von Proband VD021. Die dargestellten Werte decken fünf Stunden des Aufnahmezeitraums ab. . . . .	78
4.9	Illustration des Parameters $t_d$ als zeitlichem Abstand zwischen protokollierter Schmerzmittelgabe (2:00 Uhr) und erkanntem Zustandsübergang (2:30 Uhr). . . . .	80



4.10	Verteilung der Rekonstruktionsqualität $Q$ für die Verfahren SHC, k-means und AHC. Die statistische Signifikanz der Unterschiede wurde mittels multiplen Wilcoxon-Vorzeichen-Rang-Tests und einem Signifikanzniveau von $p^* = 0,0167$ überprüft. . . . .	83
4.11	Verteilung der Teil-Scores des Güteparameters $Q$ . . . . .	84
4.12	Links: Verlauf der Zustandsübergänge basierend auf der als optimal eingeschätzten Cluster-Lösung der Vitaldaten von Proband <i>VD011</i> . Markierung des protokollierten Zeitpunkts der Schmerzmittelgabe in blau. Die Färbung der Cluster repräsentiert den Median des Vitalparameters $P_{VLF}/P_{HF}$ im jeweiligen Cluster. Die Verteilung aller Vitalparameter in den einzelnen Zuständen ist in Abbildung 4.13 dargestellt. Rechts: Spider-Plot der Teilparameter, welche zusammen den Güteparameter $Q$ bilden. . . . .	86
4.13	Verteilung der in Abbildung 4.12 zur Gesundheitszustandserkennung verwendeten Vitalparameter für die Cluster-Verfahren SHC, k-means und AHC. Die Färbung der Boxen korrespondiert mit den Farben der jeweiligen Gesundheitszustände in Abbildung 4.12. . . . .	88
5.1	Oben: Spektrogramm des NN-Zeitsignals ( $f_s = 2.4$ Hz) basierend auf einer Fensterlänge von 30 Minuten und einer Schrittweite von 30 Sekunden. Unten: Minutengenaue Klassifizierung des Signals in Normal (N) und OSA (A). . . . .	97
5.2	a) Vergleich der maximalen SI-Indizes für alle betrachteten Methoden b) Vergleich der ARI-Werte aller betrachteten Methoden. Die statistische Signifikanz der Unterschiede im Median wurde mit Hilfe des Wilcoxon-Vorzeichen-Rang-Tests mit Bonferroni-Korrektur überprüft. . . . .	99
A.1	Variation der durch das Metronom vorgegebenen Atemfrequenz während des Versuchs. Die Atemfrequenz wurde ausgehend von 10 bpm in fünf Schritten auf 30 bpm angehoben. . . . .	132
A.2	Probandenspezifische Fehlerverteilungen bei der Extraktion der Atemfrequenz aus dem EKG. . . . .	133
B.1	Verlauf der Zustandsübergänge aus Kapitel 4, welche von den Verfahren SHC sowie den Referenzverfahren detektiert wurden. Die Färbung repräsentiert die Mediane der Vitalparameter pNN20, Atemfrequenz und ANI in den erkannten Gesundheitszuständen. . . . .	144

B.2 Vergleich zwischen manschettenbasierter Blutdruckmessung und PTT-basierter Blutdruckschätzung (systolischer Wert) für Probanden mit vollständigen EKG- und PPG-Signalen. . . . . 145

## Tabellenverzeichnis

2.1	Zum Thema maschinelles Lernen im klinischen Umfeld seit dem Jahr 2000 veröffentlichte Arbeiten. Ausgewählt wurden Veröffentlichungen, die mit Hilfe unüberwachter Lernverfahren anhand von physiologischen Daten Aussagen bzw. Vorhersagen über den Gesundheitszustand eines Patienten treffen. . . . .	33
3.1	SI-Koeffizienten der aus den Datensätzen berechneten Cluster-Lösungen für $K = 2 \dots 10$ . Lösungen, für die der berechnete SI-Koeffizient maximal wird, sind fett dargestellt. . . . .	52
4.1	Tabellarische Übersicht über die verwendeten Parameter aus dem Zeit- und Frequenzbereich. Der EKG-Datensatz wird hierfür in Zeitfenster konstanter Länge und Überlappung unterteilt. Die Parameter werden für jedes Zeitfenster separat berechnet. . . . .	69
4.2	p-Werte nach dem Vergleich der Methoden SHC, k-means und AHC hinsichtlich des Gütekriteriums Q und der darin enthaltenen Parameter. Es wurde der Wilcoxon-Vorzeichen-Rang-Test mit Bonferroni-Korrektur ( $p^* \leq 0,0167$ ) durchgeführt. Statistisch signifikante Ergebnisse sind kursiv dargestellt. . .	84
4.3	Gegenüberstellung der von den Cluster-Methoden identifizierten Schmerz- und Normalzuständen. Die Zustände werden anhand der Mediane ihrer Vitalparameter verglichen. . . . .	87
5.1	Cluster-Ergebnisse der drei betrachteten Cluster-Verfahren. Dargestellt sind jeweils die Lösungen mit der Cluster-Anzahl K, für die der SI-Koeffizient maximal wird. . . . .	100
A.1	Median und Varianz der Fehler der betrachteten Methoden für alle Probanden.	134
B.1	SI-Koeffizienten der Cluster-Lösungen ( $K = 2 \dots 10$ ) aller ausgewerteten Probanden. Der höchste SI-Koeffizient jedes Probanden ist fett dargestellt und repräsentiert die optimale Cluster-Lösung. . . . .	138

B.2	Auflistung der zur Klassifikation von Gesundheitszuständen verwendeten Vitalparameter und Biosignale für jeden Probanden mit protokollierter Schmerzmittelgabe. . . . .	139
B.3	Ergebnisse der Cluster-Analyse zu Zeitpunkten von Medikamentengaben. Analysiert wurde jeweils ein Zeitraum zwischen zwei Stunden vor und nach einer Schmerzmittelgabe mit den Verfahren SHC, k-means und AHC. Abgebildet sind die Cluster-Lösungen mit der Cluster-Anzahl K, für die der SI-Koeffizient maximal wird. Weiterhin wird für jede gefundene Lösung der ermittelte Q-Parameter zur Einschätzung der Rekonstruktionsqualität der Schmerzmittelgabe aufgelistet. . . . .	143

## Abkürzungverzeichnis

ACB-OP	.Aortocoronary Bypass Operation
AHC	.agglomeratives hierarchisches Clustering
ANI	.Analgesia Nociception Index
APACHE	.Acute Physiology And Chronic Health Evaluation
ARI	.adjusted Rand index
BDC	.Body Correlation Factor
bpm	.Beats per Minute
EDA	.Elektrodermale Aktivität
EEG	.Elektroenzephalogramm
EKG	.Elektrokardiogramm
FFT	.Fast Fourier Transformation
FIR	.Finite Impulse Response
HR	.Herzrate
HRV	.Herzratenvariabilität
IC	.Intermediate Care
ICV	.Inter-Cluster-Variabilität
MIDCAB	.minimally invasive direct coronary artery bypass
OPCAB	.off-pump coronary artery bypass
OSA	.obstruktive Schlafapnoe
PCA	.Principal Component Analysis
PPG	.Photoplethysmogramm

PTT	. . . . .	.Pulse Transit Time
PWV	. . . . .	.Pulse Wave Velocity
RSA	. . . . .	.respiratorische Sinusarrhythmie
SHC	. . . . .	.Similarity-based Hierarchical Cluster Analysis
SI	. . . . .	.Silhouettenkoeffizient
SpO2	. . . . .	.Sauerstoffsättigung des Blutes
USB	. . . . .	.Universal Serial Bus
VES	. . . . .	.ventrikuläre Extrasystole

# Kapitel 1

## Einleitung

### 1.1 Maschinelles Lernen in der Gesundheitsversorgung

„Big Data“, „Machine Learning“ und „künstliche Intelligenz“ gehören zu den am häufigsten gebrauchten Modeworten des noch jungen 21. Jahrhunderts [1]. Glaubt man den Zahlen von Google Trends, verzehnfachte sich das Nutzerinteresse an diesen Begriffen im Zeitraum von 2008 bis heute. Einen Schub erhielt der Forschungsbereich „künstliche Intelligenz“ spätestens als es möglich wurde, künstliche neuronale Netze praktisch umzusetzen. Dieses Konzept der Übertragung der Struktur des menschlichen Gehirns in Hard- und Software wurde bereits in den 50er Jahren vorgeschlagen [2]. Allerdings fehlte es damals noch an den technischen Möglichkeiten, diese Technologie einer praktischen Anwendung zukommen zu lassen wodurch dieses Forschungsfeld zunächst nicht Ziel größerer öffentlicher Aufmerksamkeit wurde. Erst Ende der 90er Jahre stand die nötige Rechenleistung zu Verfügung, um mit Hilfe komplexer künstlicher neuronale Netze – jetzt auch unter dem Begriff „Convolutional Neural Networks“ oder „Deep Learning“ bekannt – maschinelles Lernen auf praktische Probleme anzuwenden und einem größeren Fachpublikum zugänglich zu machen. Hier waren vor allem die Arbeiten der Forschergruppe um den britischen Informatiker und Kognitionspsychologen Geoffrey Hinton wegweisend, welche mehrlagige künstliche neuronale Netze um den Backpropagation Algorithmus erweiterten und damit effektiv trainierbar machten.

Aus den Anfängen und ersten vorsichtigen Schritten auf dem Weg zur künstlichen Intelligenz hat sich mittlerweile eine ganze Industrie entwickelt. Die Autoren des Magazins *The Economist* prägten in einem Artikel aus dem Jahre 2017 das treffende Zitat: „Data are to this century what oil was to the last one: a driver of growth and change“. Tatsächlich haben sich mittlerweile breite Anwendungsfelder für maschinelles Lernen und Big Data eröffnet. Neben Anwendungsbereichen in der Vorhersage klimatischer Ereignisse, der Detektion von Bilanzfälschungen oder der Erkennung von Umweltkatastrophen profitiert vor allem der medizinische Bereich stark von Fortschritten im Bereich der künstlichen Intelligenz [3, 4, 5, 6, 7]. Einige dieser Anwendungsgebiete sind als Schlagwörter in Abbildung 1.1 zu sehen. So lassen sich Algorithmen zur Vorhersage von Epidemien verwenden. Einigen





Möglichkeit geschaffen werden, den gesundheitlichen Zustand eines Patienten unter der Berücksichtigung multipler Vitalparameter zu verfolgen. Es soll damit ein Beitrag zur Verbesserung der Patientenüberwachung im klinischen Umfeld geleistet werden.

## 1.2 Aufbau der Arbeit

Die Arbeit ist in fünf weitere Kapitel gegliedert. In Kapitel 2 soll zunächst auf den aktuellen Stand der Forschung im Bereich des maschinellen Lernens eingegangen werden. Spezielles Augenmerk liegt dabei auf der Anwendung dieser Technik im klinischen Umfeld bei der Überwachung von Patienten. Es werden beispielhaft Studien präsentiert, in denen maschinelle Lernverfahren bereits zur Überwachung von Patienten im Krankenhausumfeld eingesetzt werden. Auf Basis dieser Studien sollen offene Problemstellungen bei der Patientenüberwachung aufgedeckt und so eine Aufgabenstellung für diese Arbeit abgeleitet werden. Es werden weiterhin zwei etablierte Algorithmen vorgestellt, die bereits häufig in diesem Kontext Anwendung finden. Diese Verfahren dienen im weiteren Verlauf der Arbeit als Referenzverfahren.

In Kapitel 3 wird schließlich die ähnlichkeitsbasierte hierarchische Cluster-Analyse (Similarity-based Hierarchical Cluster Analysis (SHC)) als neuartiges Verfahren zur Klassifikation von Gesundheitszuständen eingeführt. Eingangs wird zunächst ein Überblick über die Funktionsweise des Verfahrens gegeben. Anschließend wird auf einzelne wichtige Aspekte des Verfahrens genauer eingegangen. Im weiteren Verlauf des Kapitels werden zwei Studien zur Validierung des Verfahrens präsentiert. Die erste Studie erbringt den Nachweis der Anwendbarkeit des Verfahrens anhand des Clusterings verschiedener künstlich erzeugter Datensätze. Um die Effektivität der Methode auch an realen Daten zu demonstrieren, wird im Anschluss ein einfaches Experiment vorgestellt, an dessen Beispiel die Klassifikation von Gesundheitszuständen demonstriert werden soll.

Den Inhalt von Kapitel 4 bildet eine umfangreichere Messstudie, welche in Zusammenarbeit mit dem Herzzentrum Leipzig konzipiert und durchgeführt wurde. Ziel der Studie ist die Detektion von Gesundheitszuständen in einem Datensatz, welcher aus Biosignalen und Vitalparameter unterschiedlichster Art besteht. Der konkrete Anwendungsfall liegt in der Erkennung von Schmerzmittelgaben nach einem herzchirurgischen Eingriff. Weiterhin dient die Studie als Testfall für die in Kapitel 3 vorgestellte Methode zur Klassifikation von Gesundheitszuständen. Es wird zunächst die Verarbeitungs-Pipeline zur Extraktion der Vitalparameter aus den aufgenommenen Biosignalen erläutert. Hier wird besonderes Augenmerk auf die Berechnung von Parametern der Herzratenvariabilität (HRV) aus dem Zeit- und Frequenzbereich gelegt. Außerdem soll ein Güteparameter entworfen werden,

mit dessen Hilfe die Effektivität von maschinellen Lernverfahren bei der Detektion von Schmerzmittelgaben eingeschätzt werden kann. Dieser stellt eine Erweiterung eines im Bereich der Cluster-Analyse etablierten Bewertungsmaßes dar, berücksichtigt aber noch weitere Faktoren, die für die Auswertung der Studie wichtig sind. Im Anschluss werden die Ergebnisse der Detektion, welche mit Hilfe der SHC-Methode sowie die den beiden Referenzverfahren gewonnen wurden, vergleichend gegenüber gestellt und diskutiert.

Über das Thema der Arbeit hinaus wird in Kapitel 5 schließlich ein weiterer Anwendungsfall für die in Kapitel 3 konzipierte Klassifizierungsmethode vorgestellt. Mit Hilfe von Biosignalen sollen Phasen obstruktiver Schlafapnoe zeitlich rekonstruiert werden. Hierfür wird auf existierende Biosignale aus einer öffentlich zugänglichen Datenbank zurückgegriffen. Anhand von Vitalparametern, die aus den Biosignalen abgeleitet wurden, sollen Gesundheitszustände identifiziert werden. Es wird untersucht, inwiefern Zustandsübergänge mit On- bzw. Offset von Phasen obstruktiver Schlafapnoe übereinstimmen.

Kapitel 6 widmet sich der Diskussion der Ergebnisse dieser Arbeit. Die Ergebnisse der einzelnen Studien werden zusammengefasst und miteinander in Zusammenhang gebracht. Weiterhin soll eine vergleichende Einschätzung der Anwendbarkeit der SHC-Methode und der Referenzverfahren gegeben werden. Dabei sollen nicht nur die Ergebnisse der Validierungs- und Messstudien Berücksichtigung finden. Vielmehr sollen auch die algorithmischen Eigenschaften der Verfahren verglichen werden und auf ihre Eignung für die Detektion von Gesundheitszuständen hin untersucht werden.

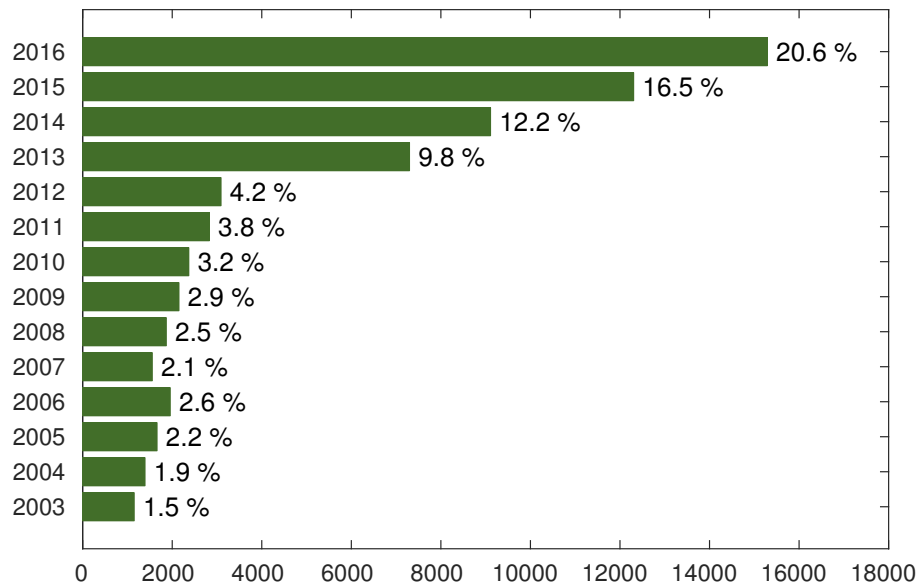
Kapitel 7 bildet den Abschluss der Arbeit. In diesem Abschnitt soll auf die noch offenen Probleme bei der weiteren Verbesserung des Verfahrens eingegangen werden. Es werden Vorschläge zur Lösung verbleibender Probleme unterbreitet. Außerdem werden mögliche zukünftige Anwendungsgebiete der SHC-Methode skizziert.

## **Kapitel 2**

### **Forschungsstand**

#### **2.1 Maschinelles Lernen – Systematischer Überblick**

Im Laufe der Evolution hat das menschliche Gehirn die Fähigkeit des komplexen Lernens durch Abstraktion entwickelt. Als Abstraktion wird die Schaffung eines mentalen Abbildes eines Sachverhaltes oder Zusammenhanges in der realen Welt verstanden, bei der unwesentliche Elemente oder Merkmale ausgeklammert werden, um so zu einem allgemeinen Verständnis zu kommen [15]. Unter maschinellem Lernen wird in diesem Zusammenhang die Suche nach einer allgemeinen Beschreibung und die Identifikation von Mustern in einem gegebenen Datensatz durch eine Maschine bzw. einen Computer verstanden [16]. Ziel ist beispielsweise die Unterteilung eines Datensatzes in verschiedene Klassen oder die Identifikation von Clustern. Im Gesundheitswesen könnte das zum Beispiel die Einteilung von Patienten in Gruppen anhand mehrerer klinischer Parameter sein. Es können zwei Arten des komplexen Lernens unterschieden werden: deduktives und induktives Lernen. Beim deduktiven Lernen werden zunächst allgemeine Regeln, Gesetze und Definitionen zur Beschreibung eines Sachverhaltes aufgestellt. Diese können dann auf beobachtete Daten angewandt werden. Dies entspricht der expliziten Definition der Eigenschaften eines Datensatzes durch den Menschen beispielsweise in einem Computerprogramm. Der Mensch kann also der Maschine bestimmte Merkmale definieren, anhand derer sie Patienten in Gruppen einteilen kann. In der Mehrzahl der Anwendungsfälle sind diese Merkmale jedoch nicht *a-priori* bekannt. Weiterhin sind typische Datensätze hochdimensional und sehr umfangreich was es dem Menschen erschwert aussagekräftige und allgemeingültige Eigenschaften zu identifizieren. Beim maschinellen Lernen wird daher ein induktiver Ansatz verfolgt. Hier lernt die Maschine ohne explizite Programmierung komplexe Zusammenhänge anhand von Beispielen [17]. Mit Hilfe geeigneter Lernverfahren können so die den Daten zugrunde liegenden Muster abgeleitet werden. Die theoretischen Grundlagen zur Übertragung des induktiven Lernens auf eine Maschine wurden bereits im frühen 19. Jahrhundert durch die Arbeiten von Pierre-Simon Laplace (Formulierung des Satz von Bayes) und Adrien-Marie Legendre (Methode der kleinsten Quadrate) gelegt. Allerdings steht die für eine praktische Realisierung notwendige Rechenleistung erst seit Mitte des 20. Jahrhunderts zur Verfügung.

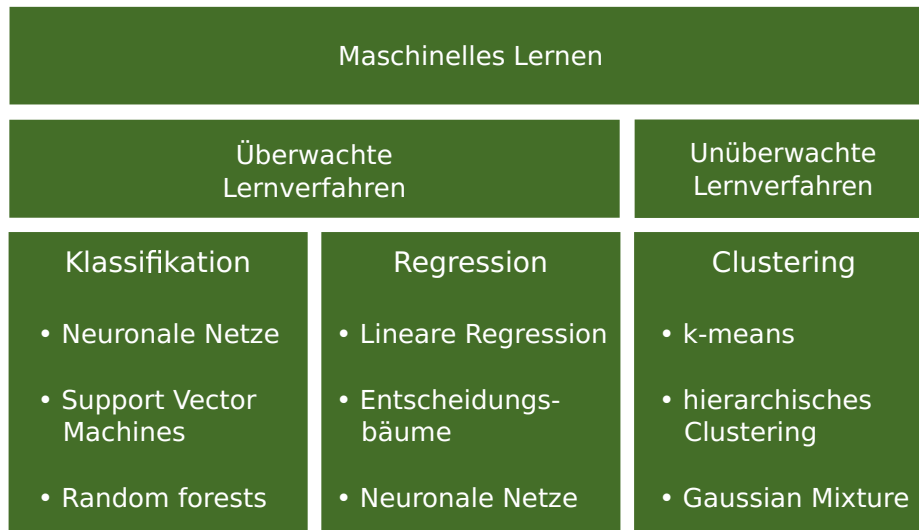


**Abbildung 2.1:** Anzahl der jährlichen Publikationen, welche die Stichworte „Machine Learning“ oder „Artificial Intelligence“ enthalten [18].

Aufgrund dessen und aufgrund des in den letzten Jahren stark gestiegenen Datenvolumens durch die immer schneller voranschreitende Digitalisierung der Gesellschaft rückt maschinelles Lernen und künstliche Intelligenz immer stärker in den Fokus von Forschung und Entwicklung. Wie aus Abbildung 2.1 ersichtlich ist, wurden allein in den Jahren 2013 bis 2016 rund 60 % aller bisher erschienenen Publikationen zu diesem Thema veröffentlicht.

Aus dieser Menge an wissenschaftlichen Arbeiten ist eine Vielzahl von maschinellen Lernverfahren entstanden. Wie in Abbildung 2.2 zu sehen ist, können diese grob in überwachte und unüberwachte Lernverfahren eingeteilt werden. Bei überwachten Lernverfahren werden zu den Eingangsdaten die korrekten Ausgangsdaten übergeben. Bei Klassifizierungsverfahren (z.B. neuronale Netze) bedeutet dies beispielsweise, dass die korrekte Klassenzugehörigkeit jedes Samples bekannt ist. Der Algorithmus versucht während des Trainings anhand der Daten Gesetzmäßigkeiten und Muster zu identifizieren, die es ihm ermöglichen die gegebenen Ausgangsdaten korrekt nachzubilden. Ziel ist es, anschließend neue Daten ohne bekannte Klassenzugehörigkeit korrekt klassifizieren zu können [19].

Im Gegensatz hierzu arbeiten unüberwachte Lernverfahren mit Daten ohne Feedback in Form von korrekten Klassenzugehörigkeiten. Solche Verfahren finden in der medizinischen Überwachung und Nachsorge Anwendung. Hier stellt sich oft die Frage, ob zwischen den Untersuchungsobjekten (z.B. Patienten oder Vitaldaten zu unterschiedlichen Zeitpunkten) Ähnlichkeiten bestehen. Ziel ist es, diese Untersuchungsobjekte in Gruppen (Cluster) einzuteilen, die hinsichtlich der betrachteten Merkmale (Features) möglichst homogen sind.



**Abbildung 2.2:** Systematik der Algorithmen des maschinellen Lernens und einige konkrete Beispiele.

Gleichzeitig soll erreicht werden, dass die Verteilung der Features zwischen den Clustern möglichst heterogen ist, um eine maximale Unterscheidbarkeit zu erreichen [20]. In der Gruppe der unüberwachten Lernverfahren existiert eine Unterteilung in hierarchische und partitionierende Methoden [21]. Hierarchische Verfahren erreichen eine Strukturierung des Datensatzes durch iteratives Teilen bzw. Kombinieren von existierenden Clustern. Im Gegensatz dazu versuchen partitionierende Methoden eine direkte Unterteilung eines Datensatzes in nicht überlappende Cluster. Ziel ist dabei die Minimierung einer bestimmten Zielfunktion, die zwischen den verschiedenen partitionierenden Verfahren variieren kann.

Die im Rahmen dieser Arbeit vorgeschlagene hierarchische Cluster-Methode ist bei den unüberwachten Lernverfahren einzuordnen. Um einen validen Vergleich zwischen der hier vorgestellten Methode und anderen Referenzverfahren anstellen zu können, werden auch die Referenzverfahren aus der Klasse der unüberwachten Lernverfahren ausgewählt. Andere in den Bereich des überwachten Lernens gehörige Verfahren (z.B. künstliche neuronale Netze, Support Vector Machines oder Decision Trees) werden deshalb im Rahmen dieser Arbeit nicht weiter betrachtet.

## 2.2 Patientenüberwachung – aktueller Stand der Technik

Derzeitiger Stand der Technik bei der Überwachung von Patienten im klinischen Umfeld sind Patientenmonitore, welche eine Vielzahl von Vitalparametern mit Hilfe von Sensoren messen. Für den klinischen Alltag sind vor allem Blutdruck (invasive Messung mittels Arte-

rienkatheter oder Manschette), Herzrate (HR), Atemfrequenz und Sauerstoffsättigung des Blutes (SpO<sub>2</sub>) [22]. Diese Parameter werden in regelmäßigen Abständen durch das Pflegepersonal vom Patientenmonitor protokolliert. Das Pflegepersonal entscheidet bei signifikanten Abweichungen von vorher definierten Normwerten über weitere Behandlungsschritte oder gegebenenfalls eine Verlegung des Patienten auf eine Station mit intensiverer medizinischer Betreuung [23, 24, 25]. Bei engmaschiger Überwachung, d.h. häufiger Überprüfung der Vitalparameter eines Patienten, kann so schnell auf kritische Gesundheitszustände reagiert werden [26]. Dies ist jedoch nur dann gegeben, wenn sich das Pflegekräfte-Patienten Verhältnis in einem ausgewogenen Bereich befindet. Dieser liegt, je nach Krankenhaus und Patientenstruktur, bei mindestens 1:4 [27]. Aufgrund der immer älter werdenden Gesellschaft sind die Gesundheitssysteme mit immer mehr Patienten konfrontiert. Zusätzlich weisen viele dieser Patienten immer schwerere und langwierigere Erkrankungen mit komplizierten Therapien auf [28]. Dieser Trend hat dazu geführt, dass beispielsweise das deutsche Gesundheitssystem seit einigen Jahren mit einem Mangel an Pflegekräften zu kämpfen hat [29]. Die vorhandenen Pflegekräfte stehen daher einer vielfach stark erhöhten Arbeitsbelastung gegenüber. Dies führt dazu, dass beispielsweise bei der Patientenüberwachung die Kontrolle und Aufzeichnung der Vitalparameter oftmals in zu großen zeitlichen Abständen sowie unvollständig erfolgt [30]. Daraus kann ein gravierender Rückgang der Pflegequalität resultieren. In Studien konnte ein Zusammenhang zwischen der Arbeitsbelastung des Pflegepersonals in Kliniken und der Mortalität von intensivmedizinisch betreuten Patienten gezeigt werden [31, 32, 33, 34, 35]. Eine Überprüfung der Vitalparameter eines Patienten in zu großen zeitlichen Abständen kann also die Auftrittswahrscheinlichkeit kritischer Komplikationen erhöhen.

Um diesem Effekt entgegen zu wirken, bieten aktuelle Patientenmonitore die Möglichkeit der schwellwertbasierten Überwachung. Das bedeutet, dass ein Alarm ausgelöst wird, sobald einer oder mehrere der überwachten Vitalparameter einen vorher definierten Schwellwert über- oder unterschreiten. Dabei existieren verschiedene Arten von Alarmen, die das Pflegepersonal bei der Priorisierung dieser unterstützen sollen [36]. Es wurde gezeigt, dass dieses Vorgehen mit einer hohen Zahl an falschen Alarmen verbunden ist [37]. Dies führt zu einer weiteren Erhöhung der Arbeitsbelastung des Krankenhauspersonals [38]. Weiterhin konnte gezeigt werden, dass es bei zu häufig auftretenden Fehlalarmen zu einer Abstumpfung des Personals kommt, was wiederum die Gefahr für verpasste kritische Gesundheitszustände erhöht [36, 39].

In der Vergangenheit wurden bereits Versuche unternommen, die Patientenüberwachung zu verbessern. So wurde zunächst durch die Verwendung Bayes'scher Netze eine Fusionie-

rung verschiedener physiologischer Parameter durchgeführt und so eine Abkehr von rein schwellwertbasierten Verfahren versucht [40]. Dieser Ansatz sollte eine datengetriebene Erkennung verschiedener klinisch relevanter Ereignisse wie z.B. Hypervolämie oder ungenügende Narkotisierung des Patienten realisieren. Weitere Innovationen auf diesem Gebiet bestanden in der Konzipierung einer Software-Architektur für einen intelligenten Monitor für kardiovaskuläre Parameter. Auch dieses System nutzt physiologische Parameter aus unterschiedlichen Quellen um anhand von deren Änderungen ereignisgesteuert Handlungs- und Therapieempfehlungen für das Klinikpersonal bereitzustellen [41]. Außerdem stand die Vorhersage der gesundheitlichen Entwicklung von Patienten direkt nach chirurgischen Eingriffen im Fokus. Hier versuchten Koski et al. den Trendverlauf von Biosignalen mit Hilfe von Medianfiltern unterschiedlicher Länge und einem linearen Regressionsansatz vorherzusagen. Der Abstand der Vorhersagewerte von den später tatsächlich gemessenen Werten diente dabei als Maß für die Zuverlässigkeit des geschätzten Trends. Eine Interpretation des Trends hinsichtlich seiner Bedeutung für den Patientenzustand wurde jedoch nicht durchgeführt [42]. Weitere Arbeiten, wie das SIMON-Projekt (Signal Interpretation and Monitoring), versuchen einen ontologischen Ansatz um die Alarmierung und Vorhersage von kritischen Ereignissen zu verbessern [43]. Dabei handelt es sich um ein System, welches eine Kombination von schwellwertbasierten Analysen und softwaretechnisch realisiertem Expertenwissen verwendet, um Handlungsempfehlungen zu geben bevor kritische Gesundheitszustände auftreten [44, 45]. Andere Autoren versuchten die schwellwertbasierte Alarmierung zu verbessern. Becker et al. schlugen ein System vor, das die Entscheidungsfindung von Anästhesisten während einer Operation unterstützen sollte. Mittels Fuzzy Logic wurden Vitalparameter in Variablen übersetzt, die eine Auskunft über den aktuellen Patientenzustand lieferten. Durch eine geeignete graphische Darstellung wurde der Benutzer über kritische Parameterwerte und daraus resultierende negative Folgen für den Patienten informiert [46, 47].

Um dem Krankenhauspersonal die Einschätzung der Schwere einer Erkrankung zu erleichtern, wurde vor allem auf Intensivstationen der Acute Physiology And Chronic Health Evaluation (APACHE)-Score sowie später dessen Erweiterungen APACHE-II, APACHE-III und APACHE-IV eingeführt [48, 49, 50, 51]. Dieses Punktesystem berücksichtigt neben kardiorespiratorischen auch renale, gastrointestinale, hämatologische, neurologische und demographische Kenngrößen. Jeder Parameter wird mit 0 bis 4 Punkten, je nach Stärke der Abweichung vom Normalbereich, gewichtet. Weiterhin wird die medizinische Vorgeschichte eines jeden Patienten berücksichtigt. In Validierungsstudien wurde eine positive Korrelation zwischen Gesamtpunktzahl und Mortalitätsrate in verschiedenen Patientenpopulationen

nachgewiesen [52, 53, 54]. Auf diese Weise kann das Risiko für schwerwiegende Komplikationen für jeden Patienten abgeschätzt und die Behandlung entsprechend angepasst werden. Die erste Berechnung des APACHE-Scores erfolgt 24 Stunden nach der Aufnahme auf der Intensivstation [48]. Aufgrund dieser Tatsache ermöglicht der APACHE-Score zwar eine grundlegende Einschätzung des Patientenzustandes bzw. die Zuordnung eines Patienten zu einer bestimmten Risikogruppe, erlaubt jedoch keine kontinuierliche Überwachung des Patientenzustandes. Weiterhin sind regelmäßige Aktualisierungen der Scores sowie Anpassungen an regionale Gegebenheiten notwendig, um eine zuverlässige Risikovorhersage treffen zu können [55].

### 2.3 Maschinelles Lernen im klinischen Umfeld

Wie in Abschnitt 2.1 erläutert wurde, hat das akademische Interesse am maschinellen Lernen in den letzten Jahren stark zugenommen. Auch im Bereich des Gesundheitswesens spielen maschinelle Lernverfahren sowohl in der klinischen Vor- und Nachsorge sowie beim Patienten-Management eine immer größere Rolle.

Aktuelle Arbeiten und Verfahren zur Patientenüberwachung kommen grundsätzlich in zwei verschiedenen Bereichen zur Anwendung: Überwachung von Patienten im klinischen Umfeld sowie beim Telemonitoring, also der Überwachung im häuslichen Umfeld der Patienten. Für beide Bereiche sollen im Folgenden beispielhafte Entwicklungen und Veröffentlichungen vorgestellt werden. Dabei soll der thematische Schwerpunkt auf klinischen Überwachungssystemen auf Basis unüberwachter maschineller Lernverfahren liegen. In Tabelle 2.1 sind die in diesem Abschnitt betrachteten Literaturquellen zusammengefasst.

Bei der Überwachung des Patienten während des stationären Aufenthalts im Krankenhaus existieren zwei Anwendungsfälle. Zum einen wird versucht mit Hilfe unüberwachter Lernverfahren eine intelligente Überwachung und Alarmierung zu realisieren. Zum anderen werden solche Verfahren angewendet, um patientenübergreifend Muster in Vitaldaten zu finden. Ziel ist es, Hochrisikopatienten mit speziellen pflegerischen Anforderungen zu identifizieren, um so den Einsatz der Krankenhausressourcen zu optimieren und das Auftreten von Komplikationen zu reduzieren. Allen Veröffentlichungen ist gemein, dass eine Klassifizierung des Gesundheitszustandes des Probanden angestrebt wird, welche auf Mustern in deren Vitalparametern basieren und sonst ohne zusätzliches Wissen auskommt.

Zur Reduzierung der Falsch-Positiv-Rate bei der Alarmierung und damit Entlastung des Krankenhauspersonals kommen vermehrt maschinelle Lernverfahren zum Einsatz. Tarassenko et al. berechnen mit Hilfe eines probabilistischen Modells einen Score, der die Abweichung der Vitalparameter von einem Normbereich angibt. Auf diese Weise konnte



Paper	ML-Verfahren	Parameter	Anwendungsgebiet
Tarassenko et al. 2006 [56]	k-means	HR, Atemfrequenz, Blutdruck, SpO2, Hauttemp.	Alarmierung
Apiletti et al. 2009 [57]	u.a. k-means	HR, Blutdruck, SpO2	Alarmierung
Baig et al. 2016 [58]	k-means, fuzzy c-means	HR, Blutdruck, PPG	Alarmierung
Anliker et al. 2004 [59]	schwellwertbas. Verfahren	EKG, SpO2, Blutdruck, Beschleunigung, Hauttemp.	Alarmierung
Bose et al. 2018 [60]	k-means, hier. Clustering	HR, Atemfrequenz, SpO2	Zustandsklassifikation
Vranas et al. 2017 [61]	hier. Clustering	Patienten-Scores	Zustandsklassifikation
Pimentel et al. 2013 [62]	hier. Clustering	HR, Atemfrequenz	Zustandsklassifikation
Sorani et al. 2007 [63]	hier. Clustering	u.a. HR, SpO2, Blutdruck	Zustandsklassifikation
Cohen et al. 2010 [64]	hier. Clustering	u.a. HR, SpO2, Blutdruck	Zustandsklassifikation
Xu et al. 2015 [65]	k-means	EEG, EKG, EMG, EDA	Zustandsklassifikation
Paoletti et al. 2006 [66]	mod. k-means	EKG	Mustererkennung

**Tabelle 2.1:** Zum Thema maschinelles Lernen im klinischen Umfeld seit dem Jahr 2000 veröffentlichte Arbeiten. Ausgewählt wurden Veröffentlichungen, die mit Hilfe unüberwachter Lernverfahren anhand von physiologischen Daten Aussagen bzw. Vorhersagen über den Gesundheitszustand eines Patienten treffen.

die Falsch-Positiv-Rate bei der Alarmierung aufgrund kritischer Gesundheitszustände auf 0,05 reduziert werden. Das System mit dem Namen BioSign soll auf der Normalstation von Krankenhäusern eingesetzt werden [56]. Apiletti et al. verwenden die mittel- und langfristige Änderungsgeschwindigkeit der Vitalparameter eines Probanden zusätzlich zur Abweichung von der Norm als Indikator für kritische Gesundheitszustände. Diese Informationen werden mit Hilfe unüberwachter Lernverfahren einem diskreten Risiko-Level zugeordnet. Dabei erzielte der k-means Algorithmus vor allem bei der Identifikation von Zuständen mit großem Risiko eine hohe Genauigkeit [57]. Ein weiteres System, welches vor allem zur Erkennung von Komplikationen im kardiovaskulären System angewendet werden soll, wurde von Baig et al. vorgeschlagen. Die Autoren vergleichen verschiedene unüberwachte Lernverfahren (k-means und fuzzy c-means) bei der Detektion von verschiedenen Arrhythmie-Typen: Bradykardie, Tachykardie, Hypertonie, Hypotonie sowie Hypervolämie. Die beiden betrachteten Algorithmen schnitten dabei ähnlich gut ab und erreichten eine Übereinstimmung von 90 % mit der Diagnose eines menschlichen Experten [58].

Weitere Veröffentlichungen haben die Identifikation von Risikopatienten oder die allgemeine Klassifikation von Gesundheitszuständen zum Ziel. Bose et al. versuchen bei Patienten nach herzchirurgischen Eingriffen kardiorespiratorische Instabilität vorherzusagen. Das verwendete ähnlichkeitsbasierte k-means Clustering identifizierte drei Patienten-Cluster anhand des Verlaufs von HR, Atemfrequenz und SpO<sub>2</sub>. Allerdings konnten keine klinisch signifikanten Unterschiede in den Vitalparametern der einzelnen Cluster gefunden werden. Die Autoren führen dies auf eine zu große Heterogenität der Patienten in den einzelnen Clustern zurück und argumentieren, dass die Einteilung in drei Cluster noch zu grob und eventuell auch die Stichprobenanzahl zu klein ist [60]. Vranas et al. nutzen agglomeratives hierarchisches Clustering (AHC) als Methode für die Identifikation von Hochrisikopatienten auf einer Intensivstation. Als Datengrundlage dienen verschiedene Patienten-Scores, von denen einige auf Basis von Vitalparametern erstellt wurden. Die Einteilung der Patientenkohorte in sechs Cluster wurde dabei von den Autoren als optimal angesehen. Diese Cluster unterschieden sich signifikant z.B. hinsichtlich schwere der Erkrankung sowie Pflegeanforderung und Mortalität. Allerdings spielen hier Vitalparameter nur indirekt eine Rolle, da die meisten Scores auf Patientencharakteristika (z.B. demographische Daten, Art der Erkrankung, Laborbefunde) basieren [61]. Einen ähnlichen Ansatz verfolgen Pimentel et al. bei der Identifikation von Risikopatienten. Hier wird aus den Verläufen der aufgezeichneten Biosignale für jeden Patienten eine multimodale Trajektorie berechnet, welche den Verlauf des Patientenzustandes im Verlauf der Zeit beschreibt. Es wird dann versucht, Patienten mit ähnlichen Trajektorien mit Hilfe des agglomerativen hierarchischen Clusterings zu-

sammenzufassen. Den Autoren ist es gelungen, vier Patienten-Cluster zu identifizieren, von denen einer für Patienten stand, die später Komplikationen entwickelten [62]. Die Forschergruppe um Sorani et al. untersuchten den prognostischen Wert von Vitalparametern für die mittel- und langfristige gesundheitliche Entwicklung von Patienten, welche ein Schädel-Hirn Trauma erlitten [63]. Sie beobachteten dazu 23 Probanden über mehrere Tage und erhoben eine Reihe von Vitalparametern. Der dabei entstehende Datensatz wurde mit Hilfe eines hierarchischen Cluster-Verfahrens in drei Cluster unterteilt. Diese Cluster-Anzahl wurde von den Autoren fest definiert. Es stellte sich heraus, dass diese Klassifizierung als Prädiktor für den weiteren Verlauf sowie für die Schwere der erlittenen Verletzung geeignet war. Die Autoren schlussfolgerten, dass die Cluster-Analyse von Vitalparametern einen Beitrag zur Patientenklassifikation und einer effektiven Therapieplanung leisten kann. Im Artikel von Cohen et al. wird zum ersten Mal der Begriff „Gesundheitszustand“ im Sinne dieser Arbeit verwendet [64]. Die Autoren unterzogen Vitaldaten, die von 17 Patienten einer Intensivstation erhoben wurden, einer Cluster-Analyse. Jedes erkannte Cluster wurde als Gesundheitszustand angesehen, welcher sich durch charakteristische Wertebereiche der betrachteten Vitalparameter auszeichnete. Weiterhin wurden die erkannten Zustände als zeitlich kompakt angesehen. Das bedeutet, dass ein Proband im Laufe der Zeit verschiedene Gesundheitszustände durchlief. Die an der Studie teilnehmenden Personen wurden über einen längeren Zeitraum beobachtet. Eine der wichtigsten Erkenntnisse dieser Studie war, dass alle Probanden, die während ihres Aufenthaltes im Krankenhaus verstarben am Ende des Messzeitraumes den gleichen Gesundheitszustand annahmen. Die Autoren schlussfolgerten daher, dass dieser Zustand als Prädiktor für Mortalität nach schweren Verletzungen geeignet war.

Xu et al. entwickelten ein System zur Bestimmung des Stressindex unter mentaler und physischer Belastung. Die Probanden durchliefen abwechselnd verschiedene Phasen von Ruhe und Belastung. Der k-means Algorithmus wurde genutzt, um die Belastungs- und Ruhephasen anhand der parallel aufgezeichneten Vitaldaten zu identifizieren. Es wurden zwei signifikant unterschiedliche Cluster ermittelt, von denen einer mit Stress assoziiert werden konnte. Die automatischen Klassifizierungsergebnisse wurden mit der Selbsteinschätzung der Probanden (State-Trait Anxiety Inventory (STAI) Index [67]) verglichen. Stresszustände konnten mit Hilfe des vorgestellten Verfahrens mit einer Sensitivität von 0,85 rekonstruiert werden [65].

Ein weiteres Anwendungsgebiet für maschinelles Lernen liegt in der Arrhythmiedetektion bei Herzpatienten. Paoletti et al. analysieren hierfür die QRS-Komplexe von Patienten. Diese werden aus dem Elektrokardiogramm (EKG) extrahiert und zu einer Datenmatrix

zusammengesetzt. Anschließend wird die Datenmatrix einer Hauptkomponentenanalyse (Principal Component Analysis (PCA)) unterzogen. Es handelt sich hierbei um ein halbautomatisches Verfahren, da der Benutzer die Komponenten, welche als Datengrundlage für das anschließende k-means Clustering dienen sollen, selbst auswählen muss. Den Autoren ist es mit diesem Verfahren gelungen, Arrhythmien mit einer Genauigkeit von 95 % zu detektieren [66].

Im zweiten medizinischen Anwendungsgebiet, dem Telemonitoring, werden mobile Messsysteme im häuslichen Umfeld des Patienten eingesetzt. Diese kommunizieren meist drahtlos mit einer Verarbeitungseinheit, welche die gesammelten Daten zur Auswertung an einen zentralen Server sendet. Ziel des Einsatzes solcher Systeme ist es, den Patienten eine Genesung im gewohnten Umfeld bei gleichzeitig adäquater medizinischer Betreuung zu ermöglichen. Weiterhin soll der effektive Einsatz von Krankenhausressourcen z.B. bei der Planung der Bettenbelegung ermöglicht werden.

Zu diesem Zweck haben Bidagardie et al. ein System für die Detektion kritischer Gesundheitszustände nach einem herzchirurgischen Eingriff entwickelt. Ein tragbares Messsystem zeichnet HR und HRV des Patienten kontinuierlich auf. Mit Hilfe des k-means Algorithmus wurde versucht, Muster zu finden, die für den Zustand des Patienten zu einem bestimmten Zeitpunkt charakteristisch sind [68]. Einen ähnlichen Ansatz verfolgen Clifton et al. mit einem System zur Überwachung von ambulant behandelten Patienten. Dabei werden verschiedene drahtlose Sensoren zur Aufzeichnung von Biosignalen verwendet. Diese Signale werden per Bluetooth an das Smartphone des Patienten weitergeleitet und per IP-Verbindung an einen zentralen Server zur Auswertung gesendet [69]. Mit Hilfe überwachter Lernverfahren werden von der Norm des Patienten abweichende Vitalparameter erkannt und eine Alarmierung des Krankenhauspersonals durchgeführt. Auch das von Ho et al. entwickelte System zur Erkennung von Arrhythmien im EKG kann als mobiles System im klinischen und heimischen Umfeld eingesetzt werden [70]. Auch hier werden mit Hilfe eines mobilen Sensorsystems das EKG des Patienten aufgezeichnet und per IP-Verbindung an einen Auswerterechner im Krankenhaus übertragen. Dort findet mit Hilfe überwachter Lernverfahren eine Klassifikation des EKG statt. Im Falle einer erkannten Arrhythmie soll das Krankenhauspersonal alarmiert werden.

Die hier aufgeführten Literaturbeispiele stellen selbstverständlich keinen erschöpfenden Überblick über die Anwendung von maschinellem Lernen im Bereich der Gesundheitsversorgung dar. Sie zeigen jedoch das große Interesse der Forschungsgemeinschaft an diesem Thema. Sie zeigen weiterhin, dass ähnlichkeitsbasierte Clustering-Verfahren bei der Mustererkennung in Biosignalen anwendbar sind.

## 2.4 Referenzverfahren

### 2.4.1 Agglomeratives hierarchisches Clustering

Die AHC-Methode hat das Ziel, einen Datensatz  $\mathbf{X}$  ( $\mathbf{X} \in \mathbb{R}^{N \times M}$ ) bestehend aus  $N$  Samples und  $M$  Merkmalen (Features) in Cluster zu gruppieren. Die Bildung von Clustern erfolgt in einem iterativen Prozess. Zunächst werden alle  $N$  Samples des Datensatzes jeweils als Cluster angesehen. In jedem Schritt des Algorithmus werden nun die beiden Cluster-Paare zu einem neuen Cluster zusammengefasst, die die größte Ähnlichkeit bzw. geringste Distanz zueinander aufweisen. Dies wird solange wiederholt, bis nur noch ein Cluster übrig bleibt, der alle Samples des Datensatzes umfasst. In Abbildung 2.3 wird die AHC-Methode auf einen simulierten Datensatz angewandt. Dieses Verfahren zählt zur hierarchischen Gruppe unter den unüberwachten Lernverfahren. Da bei diesem iterativen Algorithmus in jedem Schritt nur einzelne Cluster auf ihre jeweilige Kombinierbarkeit untersucht werden und andere Teile des Datensatzes außer Acht gelassen werden, kann von einer iterativen lokalen Optimierung der Cluster-Struktur gesprochen werden.

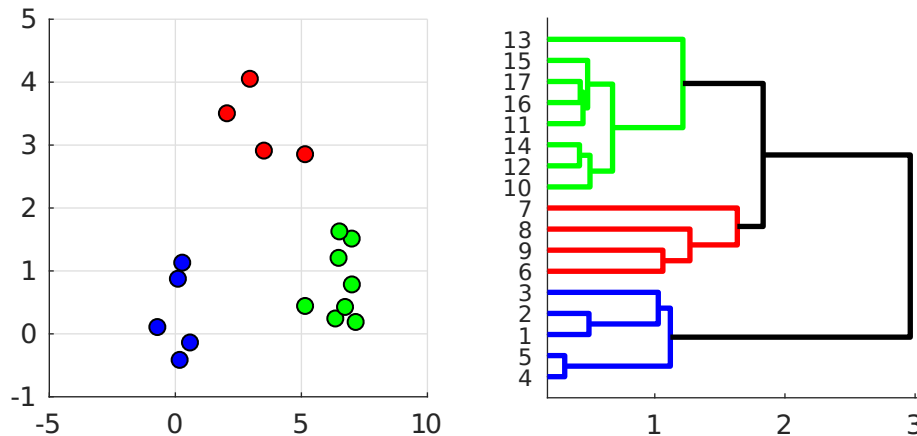
Die bei diesem Verfahren entstehende baumartige Struktur kann als Dendrogramm dargestellt werden. Ein solcher ist in Abbildung 2.3 rechts dargestellt. Die Wurzel des Baums stellt dabei ein einziges Cluster dar, das den gesamten Datensatz  $\mathbf{X}$  umfasst. Die weiter oben liegenden Knoten repräsentieren die Unterteilung dieses Clusters in immer kleiner werdende Teil-Cluster. Schlussendlich stellen die Blätter des Baums Cluster dar, in denen sich jeweils nur noch ein Sample des Datensatzes befindet [71].

### 2.4.2 k-means Methode

Bei dem k-means Algorithmus handelt es sich um ein Verfahren zur Partitionierung eines Datensatzes in eine Anzahl Cluster, die a-priori festgelegt werden muss. In diesem Punkt unterscheidet er sich wesentlich vom agglomerativen hierarchischen Cluster-Verfahren. Ziel ist es, die Cluster-Zugehörigkeit der Samples so zu wählen, dass die Intra-Cluster-Varianz minimiert wird. Ziel des k-means Verfahrens ist also eine globale Optimierung der Cluster-Zugehörigkeit der Samples. Die Einteilung eines Datensatzes  $\mathbf{X}$  in  $K$  Cluster läuft in drei Schritten ab:

1. Seeding: Initialisierung der  $K$  Cluster-Zentren  $\mu_k$ .
2. Zuweisung aller Samples  $\mathbf{x}_i$  zu dem Cluster  $C_k$ , dessen Zentrum am nächsten liegt.

$$k_i = \underset{k}{\operatorname{argmin}}(\mu_k - \mathbf{x}_i) \quad (2.1)$$



**Abbildung 2.3:** Links: Beispieldatensatz  $\mathbf{X}$  ( $\mathbf{X} \in \mathbb{R}^{13 \times 2}$ ) bestehend aus 13 Samples, die durch 2 Features beschrieben werden. Die Cluster-Zugehörigkeit ist farbig hervorgehoben. Rechts: Dendrogramm der hierarchischen Zerlegung des Datensatzes links. Die zugrunde liegende Cluster-Struktur ist farbig hervorgehoben.

Hier wird die euklidische Distanz als Abstandsmaß verwendet. Es sind auch andere Abstandsmaße verwendbar. Diese sind in Abschnitt 3.2 näher erläutert.

### 3. Neuberechnung der Cluster-Zentren und Rückkehr zu Punkt 2

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} (\mathbf{x}_i) \quad \forall \mathbf{x}_i \in C_k \quad (2.2)$$

Hier repräsentiert  $N_k$  die in Cluster  $k$  enthaltene Anzahl Samples.

Der Algorithmus wird solange ausgeführt, bis sich die Zuordnung der Samples zu den Clustern nicht mehr ändert.

Die Qualität des Clustering-Ergebnisses sowie die Konvergenzzeit des Algorithmus hängen stark von der Wahl der initialen Cluster-Zentren ab. Die ursprüngliche k-means Methode initialisierte die Cluster-Zentren zufällig. Es existieren verschiedenste Ansätze, welche die Initialisierung des Algorithmus verbessern. Exemplarisch soll hier der auch in MATLAB verwendete k-means++ Algorithmus vorgestellt werden [72]. Die Initialisierung mittels des k-means++ Algorithmus erfolgt durch folgende Arbeitsschritte:

1. Wähle ein zufälliges Sample als Cluster-Zentrum  $\mu_1$  aus dem Datensatz  $\mathbf{X}$ .
2. Berechne für jedes Sample  $\mathbf{x}_i$  den Abstand  $D(\mathbf{x})$  zum nächst gelegenen Cluster-Zentrum.

3. Wahl eines neuen Cluster-Zentrums aus  $\mathbf{X}$  mit der Wahrscheinlichkeit

$$\frac{D(\mathbf{x}_i)}{\sum_{\mathbf{x}} D(\mathbf{x}_i)} \quad (2.3)$$

4. Wiederhole Schritte 2 und 3 bis die gewünschte Anzahl an initialen Cluster-Zentren gefunden wurde.
5. Führe normalen k-means Algorithmus aus.

Diese Form der Initialisierung hat den Vorteil, dass die Wahrscheinlichkeit, mit der ein Sample als Cluster-Zentrum gewählt wird, proportional mit seiner Entfernung zum nächsten bereits existierenden Cluster-Zentrum steigt. Das sorgt für eine gleichmäßige Verteilung der initialen Cluster-Zentren über den vom Datensatz aufgespannten Raum. Auf diese Weise können in kürzerer Zeit bessere Clustering-Ergebnisse erzielt werden [72].

Mit der AHC- und der k-means Methode wurden zwei Vertreter der Klasse der unüberwachten Lernverfahren vorgestellt. Es handelt sich hierbei um zwei Verfahren, die jeweils der hierarchischen bzw. partitionierenden Klasse zuzuordnen sind. Sie finden breite Anwendung beispielsweise in den Bereichen Bildverarbeitung, Biosignalverarbeitung und Data Mining [73]. Im weiteren Verlauf dieser Arbeit dienen sie als Referenzverfahren.

## 2.5 Zusammenfassung und offene Problemstellungen

Dieses Kapitel hatte die Darlegung des aktuellen Standes der Technik bei der Anwendung von maschinellen Lernverfahren im klinischen Umfeld zum Ziel. Nach einer allgemeinen Einführung in das Thema und einer systematischen Einordnung einzelner Verfahren wurden eine Reihe von Arbeiten vorgestellt, welche die Anwendung des maschinellen Lernens bei der Überwachung stationär behandelter Patienten zum Thema haben. Auswahlkriterien waren sowohl die Verwendung von Vitaldaten (EKG, PPG, EDA etc.) als auch die Anwendung unüberwachter Lernverfahren zur Mustererkennung in diesen Vitaldaten. Es wurden insgesamt drei verschiedene Anwendungsfälle identifiziert: (i) Alarmierung des Krankenhauspersonals beim Auftreten kritischer Gesundheitszustände, (ii) Erkennung und Klassifikation von Hochrisikopatienten im intensivmedizinischen Umfeld und (iii) das Telemonitoring von Patienten im Krankenhaus oder Zuhause. Insgesamt betrachtet leisten alle in diesem Kapitel aufgeführten Arbeiten einen wichtigen Beitrag zur Erhöhung der Patientensicherheit sowie damit einhergehend auch in vielen Fällen zur Entlastung des Krankenhauspersonals. Jedoch zielen die meisten Systeme auf sehr spezifische Krankheitsbilder oder Anwendungsfelder ab. Die Repräsentation des Gesundheitszustandes des Patienten

besteht bei intelligenten Systemen, welche der Alarmierung dienen, in einem Score. Dieser bietet dem Krankenhauspersonal zwar eine instantane Einschätzung des Zustandes, lässt aber Entwicklungen der Vitalparameter, die zu diesem Zustand geführt haben, weitgehend außer Acht. Weiterhin nutzen die meisten Verfahren nur einen Bruchteil der bei der Patientenüberwachung zur Verfügung stehenden Parameter. So sind HR, Atemfrequenz, SpO<sub>2</sub> sowie der Blutdruck die am häufigsten betrachteten Parameter. Andere aussagekräftige Messgrößen wie der Hautleitwert oder weitere aus dem EKG abgeleitete Parameter aus dem Zeit- und Frequenzbereich werden hingegen selten genutzt.

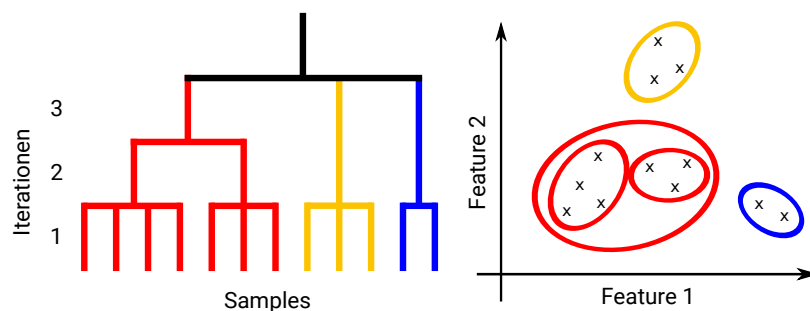
Im Rahmen dieser Arbeit soll deshalb eine neue Methode vorgestellt werden, welche anhand von Vitalparametern verschiedener Art Gesundheitszustände klassifizieren kann. Hierfür soll ein hierarchisches Verfahren konzipiert werden, das Gesundheitszustände anhand ihrer Ähnlichkeit zu neuen Zuständen zusammenfassen kann. Auf diese Weise soll eine Beschreibung des Patientenzustandes auf beliebigen Zeitskalen ermöglicht werden. Perspektivisch soll das Krankenhauspersonal in die Lage versetzt werden, den aktuellen Zustand des Patienten und dessen Entwicklung in der Vergangenheit mit unterschiedlicher zeitlicher Granularität zu erfassen und bewerten zu können. Dabei sollen durch das Verfahren keine aktiven Diagnosen gestellt werden. Den Ärzten soll vielmehr ein Werkzeug zur Verfügung gestellt werden, mit dessen Hilfe Gesundheitszustände aus großen Datensätzen extrahiert werden können. Anders als bei der schwellwertbasierten Überwachung einzelner Parameter soll ein ganzheitlicher Überblick über den Patientenzustand ermöglicht werden, welcher als Basis für individualisierte Entscheidungen des Krankenhauspersonals dienen kann. Die damit realisierbare Berücksichtigung von Parametern unterschiedlichster Skalen soll prinzipiell auch die Bewertung komplexer Zustände ermöglichen.



## Kapitel 3

# Ähnlichkeitsbasiertes hierarchisches Clustering von Gesundheitszuständen

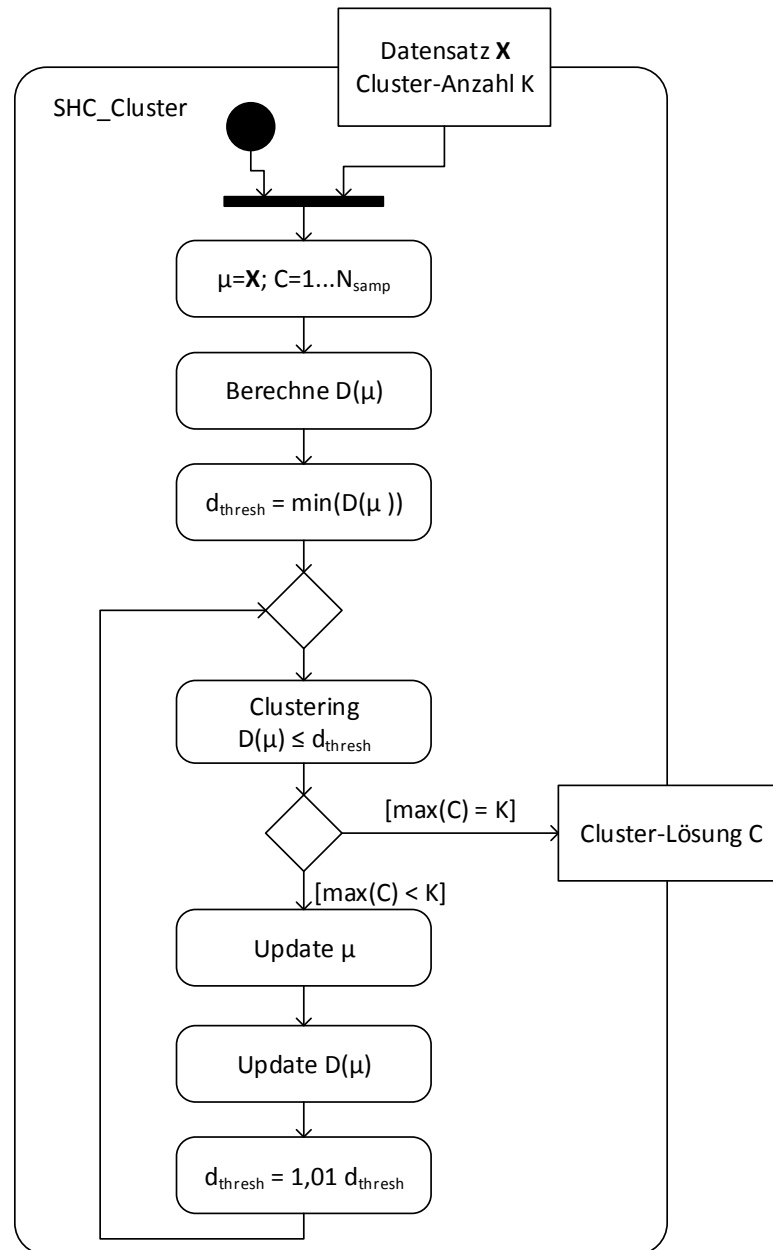
### 3.1 Allgemeine Funktionsweise des Algorithmus



**Abbildung 3.1:** Links: Baumstruktur, welche bei der iterativen Einteilung eines Datensatzes in Cluster entsteht. Rechts: Iterative Einteilung der Samples in Cluster anhand ihrer Ähnlichkeit.

Bei der SHC Methode handelt es sich um ein neuartiges Verfahren, welches anhand von Vitaldaten Gesundheitszustände identifizieren kann. Die Methode zählt zu den unüberwachten maschinellen Lernverfahren. Die für die Klassifizierung verwendeten Vitaldaten werden dabei aus verschiedensten Biosignalen abgeleitet. Ein möglicher Anwendungsfall liegt in der postoperativen Überwachung von stationär aufgenommenen Patienten. Indem dem Krankenhauspersonal ein ganzheitlicher Überblick über den Patientenzustand, basierend auf multiplen Vitalparametern, gegeben wird, kann eine Verbesserung der Patientenüberwachung erreicht werden.

Wie in Abbildung 3.2 dargestellt, identifiziert die Methode Gesundheitszustände durch iterative Teilung eines Datensatzes  $\mathbf{X}$  ( $\mathbf{X} \in \mathbb{R}^{N \times M}$ ) in eine definierte Anzahl von  $K$  Clustern. Jedes Cluster soll dabei als separater Gesundheitszustand aufgefasst werden. Durch die iterative Teilung des Datensatzes wird eine Einteilung der zeitlichen Entwicklung des Patientenzustandes mit unterschiedlicher Granularität erreicht. Abbildung 3.1 demonstriert dies an einem einfachen Beispiel. Zunächst wird jedes Sample des Datensatzes  $\mathbf{X}$  als separates Cluster angesehen. Als nächstes wird die Abstandsmatrix  $\mathbf{D}$  der Cluster berechnet. Die



**Abbildung 3.2:** Übersicht über den Ablauf des Algorithmus für die ähnlichkeitsbasierte hierarchische Cluster-Methode als UML-Aktivitätsdiagramm

Abstandsmatrix  $\mathbf{D}$  ist symmetrisch, wobei die Elemente  $D(i, k)$  den Abstand zwischen den Clustern  $i$  und  $k$  darstellen. Es ist die Verwendung verschiedenster Abstandsmaße möglich (siehe auch Kapitel 3.2). Im Folgenden soll die quadratische euklidische Distanz verwendet werden. Im ersten Schritt des Algorithmus werden Cluster, deren Abstand einen initial gewählten Ähnlichkeitsschwellwert unterschreiten, zu einem neuen Cluster zusammengefasst. Als initialer Schwellwert wird der kleinste Abstand zwischen den Samples des Datensatzes  $\mathbf{X}$  verwendet:

$$d_{thresh} = \min(\mathbf{D}) \quad (3.1)$$

Nach der ersten Iteration des Verfahrens ergeben sich bereits einige Cluster, die aus mehreren Samples bestehen. Im nächsten Schritt wird die Abstandsmatrix  $\mathbf{D}$  aktualisiert. Hierfür werden die Mediane  $\mu$  ( $\mu \in \mathbb{R}^{N_{clusters} \times M}$ ) aller Cluster berechnet. Ausgehend von den Cluster-Medianen wird zusätzlich die Distanzmatrix  $\mathbf{D}(\mu)$  aktualisiert. Der Ähnlichkeitsschwellwert wird für die nächste Iteration des Algorithmus um 1 % erhöht.

$$d_{thresh} := 1,01d_{thresh} \quad (3.2)$$

Es werden nun wiederum Cluster zusammengefasst, deren Ähnlichkeit  $d_{thresh}$  unterschreiten. Der Algorithmus läuft theoretisch so lange, bis nur noch ein großes Cluster existiert, das alle Samples des Datensatzes enthält. Alternativ kann abgebrochen werden, wenn eine gewünschte Cluster-Anzahl  $K$  erreicht ist oder ein speziell gewähltes Gütekriterium sein Optimum erreicht hat. Auf diese Weise entsteht – ähnlich wie beim agglomerativen hierarchischen Clustering-Verfahren – eine baumähnliche Struktur. Die Innovation dieses Verfahrens gegenüber der traditionellen hierarchischen Cluster-Methode besteht darin, dass potentiell mehr als zwei Cluster in einem Schritt zusammengefasst werden können. Damit können Zustände, die aufgrund ihrer Ähnlichkeit nah beieinander liegen, in einem Schritt zusammengefasst werden. Diesem Vorgehen liegt die Annahme zugrunde, dass sich einzelne Gesundheitszustände durch einen charakteristischen Satz an Vitalparameter auszeichnen.

## 3.2 Ähnlichkeit und Abstandsmaße

### 3.2.1 Berechnung der Ähnlichkeit von Gesundheitszuständen

Um die Objekte eines Datensatzes in Cluster einteilen zu können, muss zunächst ein Maß definiert werden, das deren Abstand zueinander beschreibt. Dabei kann jedes Sample eines Datensatzes  $\mathbf{X}$ , bestehend aus  $M$  Features, als Punkt in einem  $M$ -dimensionalen Raum angesehen werden. Der Abstand zwischen zwei Samples A und B in diesem Raum wird als

deren Ähnlichkeit definiert. Dabei werden Samples, welche sehr nahe beieinander liegen, als ähnlich, Samples, die weit auseinander liegen, hingegen als unähnlich betrachtet. Es ist zu unterscheiden, ob die Daten metrischer, nominaler oder binärer Art sind. Bei metrischen Daten, also Daten, deren Features einen kontinuierlichen Wertebereich aufweisen, werden dafür Minkowski-Abstände berechnet [74]. Für zwei Samples  $\mathbf{x}_a$  und  $\mathbf{x}_b$  aus einem Datensatz  $\mathbf{X}$  ( $\mathbf{X} \in \mathbb{R}^{N \times M}$ ), bestehend aus  $N$  Samples und  $M$  Merkmalen (Features), ist dieser wie folgt definiert

$$d_p(\mathbf{x}_a, \mathbf{x}_b) = \left( \sum_{i=1}^M |x_{i,a} - x_{i,b}|^p \right)^{1/p} ; \forall p \geq 1, p \in \mathbb{N} \quad (3.3)$$

Für den Fall  $p = 2$  ergibt sich aus Gleichung 3.3 für unkorrelierte Features die euklidische Distanz.

$$d_e(\mathbf{x}_a, \mathbf{x}_b) = \sqrt{\sum_{i=1}^M (x_{i,a} - x_{i,b})^2} = \sqrt{(\mathbf{x}_a - \mathbf{x}_b)^T (\mathbf{x}_a - \mathbf{x}_b)} \quad (3.4)$$

Bei zusätzlicher Berücksichtigung der Korrelation der Features des Datensatzes ergibt sich die Mahalanobis-Distanz [75]:

$$d_m(\mathbf{x}_a, \mathbf{x}_b) = \sqrt{(\mathbf{x}_a - \mathbf{x}_b)^T \mathbf{C}^{-1} (\mathbf{x}_a - \mathbf{x}_b)} \quad (3.5)$$

Hier stellt  $\mathbf{C} \in \mathbb{R}^{M \times M}$  die Kovarianzmatrix des Datensatzes  $\mathbf{X}$  dar. Bei der euklidischen Distanz werden durch die Definition  $\mathbf{C} = \mathbb{1}$  die Features als unkorreliert angenommen. Es können, je nach Wahl des Parameters  $p$ , in Gleichung 3.3 noch eine ganze Reihe weiterer Abstandsmaße definiert werden.

Mit Hilfe der Abstandsmaße in den Gleichungen 3.4 und 3.5 kann nun die Abstandsmatrix  $\mathbf{D} \in \mathbb{R}^{N \times N}$  definiert werden, bei der jedes Element  $D_{ij}$  einen Abstand  $d(\mathbf{x}_i, \mathbf{x}_j)$  der beiden Samples  $\mathbf{x}_i$  und  $\mathbf{x}_j$  beschreibt. Die Matrix besitzt zusätzlich folgende Eigenschaften:

- Symmetrie:  $D_{ij} = D_{ji}$
- Positive Definitheit:  $D_{ij} > 0$  und  $D_{ij} = 0 \iff i = j$

### 3.2.2 Höherdimensionale Trennung von Gesundheitszuständen mittels Kernel-Funktionen

Bei realen Datensätzen kommt es häufig zu Cluster-Konstellationen, die im Ursprungsraum  $\mathbb{R}^{N \times M}$  nicht linear separierbar sind. Um diesem Umstand gerecht zu werden, soll hier eine Methode vorgestellt werden, die aus anderen maschinellen Lernverfahren (z.B Support Vec-

tor Machines) als *Kernel-Trick* bekannt ist [76]. Dieses Verfahren wird für die Verwendung mit der SHC-Methode angepasst.

Ziel des *Kernel-Tricks* ist die Überführung des Datensatzes  $\mathbf{X}$  aus dem  $\mathbb{R}^{N \times M}$  in einen höherdimensionalen Raum  $\mathbb{R}^{N \times P}$  ( $P \gg K$ ) mittels einer nichtlinearen Transformationsfunktion  $\phi(\mathbf{X})$ . Diese soll für die Verwendung mit der SHC-Methode angepasst werden.

$$\tilde{\mathbf{X}} = \phi(\mathbf{X}) \quad (3.6)$$

Der Methode liegt die Annahme zugrunde, dass der Datensatz  $\tilde{\mathbf{X}}$  im P-dimensionalen Bildraum besser trennbar ist als im M-dimensionalen Ursprungsraum.

Die Transformation des Datensatzes in Gleichung 3.6 kann sehr rechenaufwendig werden. Dies ist vor allem dann der Fall, wenn die Dimensionalität des Bildraums sehr groß oder unendlich ist. Die Tatsache, dass die Ähnlichkeit von Gesundheitszuständen auf der Ähnlichkeit bzw. auf Abständen sowohl im Ursprungs- als auch im Bildraum basiert, bietet dabei einen Weg, dieses Problem zu umgehen. Viele Abstandsmaße können in Form von Skalarprodukten formuliert werden. Auf diese Weise ist es möglich, den Abstand zweier Cluster im hochdimensionalen Bildraum implizit zu berechnen, ohne den Ursprungsraum zu verlassen. Im Folgenden soll dieses Prinzip am Beispiel der euklidischen Distanz demonstriert werden. Aus Gleichung 3.4 ergibt sich durch Quadrieren die quadratische euklidische Distanz.

$$d_e^2(\mathbf{x}_a, \mathbf{x}_b) = (\mathbf{x}_a - \mathbf{x}_b)^T (\mathbf{x}_a - \mathbf{x}_b) \quad (3.7)$$

Durch Ausmultiplizieren kann dieser Abstand in eine Form überführt werden, die nur noch aus Skalarprodukten besteht.

$$d_e^2 = \mathbf{x}_a \mathbf{x}_a^T + \mathbf{x}_b \mathbf{x}_b^T - 2\mathbf{x}_a \mathbf{x}_b^T \quad (3.8)$$

Die Anwendung der nichtlinearen Transformationsfunktion auf Gleichung 3.8 ergibt:

$$d_e^2 = \phi(\mathbf{x}_a)\phi(\mathbf{x}_a)^T + \phi(\mathbf{x}_b)\phi(\mathbf{x}_b)^T - 2\phi(\mathbf{x}_a)\phi(\mathbf{x}_b)^T \quad (3.9)$$

Aus Gleichung 3.9 wird ersichtlich, dass die Transformation des Datensatzes von der Dimensionalität des Bildraums unabhängig ist, da der folgende Zusammenhang gilt:

$$\phi(\mathbf{X}) \in \mathbb{R}^{N \times P} \rightarrow \phi(\mathbf{X})^T \in \mathbb{R}^{P \times N} \rightarrow \phi(\mathbf{X})\phi(\mathbf{X})^T \in \mathbb{R}^{N \times N} \quad (3.10)$$

Wenn also eine sogenannte Kernel-Funktion in der Form  $K(\mathbf{x}_a, \mathbf{x}_b) = \langle \phi(\mathbf{x}_a)\phi(\mathbf{x}_b) \rangle$  gefunden werden kann, dann können die Abstände der Elemente eines Datensatzes im hochdimensionalen Bildraum auf implizite Weise berechnet werden, ohne den Ursprungsraum zu verlassen.

Die einfachste Form eines Kernel ist der lineare Kernel. Dieser besteht nur aus dem Skalarprodukt der beiden Datenvektoren.

$$K(\mathbf{x}_a, \mathbf{x}_b) = \langle \mathbf{x}_a, \mathbf{x}_b \rangle \quad (3.11)$$

Ein weiteres Beispiel für eine Kernel-Funktion ist unter der polynomische Kernel.

$$K(\mathbf{x}_a, \mathbf{x}_b) = (\mathbf{x}_a \mathbf{x}_b + c)^d \quad (3.12)$$

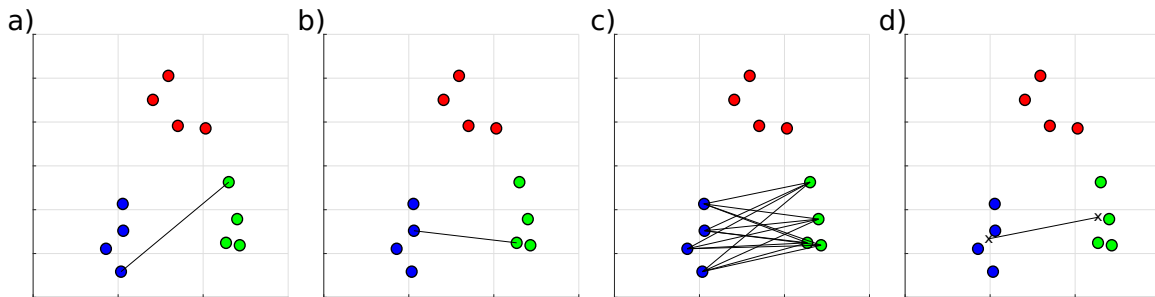
Dabei steuert der Parameter  $d$  den Grad des Polynoms und damit die Dimension des Bildraums. Der Parameter  $c$  kontrolliert die Gewichtung der Terme höherer Ordnung gegenüber denen niedrigerer Ordnung. Für  $d = 1$  ergibt sich beispielsweise der lineare Kernel aus Gleichung 3.12.

Weiterhin existieren Radial-Basis-Function Kernels (RBF-Kernel), die eine Transformation in einen unendlich dimensionalen Bildraum realisieren. Dieser soll auch für alle folgenden Auswertungen mittels der agglomerativen hierarchischen Cluster-Methode verwendet werden.

$$K(\mathbf{x}_a, \mathbf{x}_b) = \exp\left(-\frac{\|\mathbf{x}_a - \mathbf{x}_b\|^2}{2\sigma^2}\right) \quad (3.13)$$

Anschaulich beschrieben handelt es sich hierbei um eine multivariate Gauß-Funktion, welche die Ähnlichkeit der beiden Samples bzw. Cluster  $\mathbf{x}_a$  und  $\mathbf{x}_b$  angibt. Für sehr nahe beieinander gelegene Cluster geht  $K(\mathbf{x}_a, \mathbf{x}_b)$  gegen 1, für weit voneinander entfernt liegende Cluster gegen 0. Der Parameter  $\sigma$  steuert dabei die Entfernung, für die der RBF-Kernel gegen 0 strebt. Insofern ist der RBF-Kernel an sich bereits eine Art Ähnlichkeitsmaß zweier Cluster. Eingesetzt in Gleichung 3.9 und nach einigen Vereinfachungen ergibt sich schließlich die Distanz zweier Cluster im Bildraum bei der Verwendung des RBF-Kernels.

$$d_e(\mathbf{x}_a, \mathbf{x}_b) = 2 \left(1 - \exp\left(-\frac{\|\mathbf{x}_a - \mathbf{x}_b\|^2}{2\sigma^2}\right)\right) \quad (3.14)$$



**Abbildung 3.3:** Verschiedene Verfahren zur Abstandsberechnung: a) Complete Linkage, b) Single Linkage, c) Average Linkage, d) Median Methode

### 3.3 Cluster-Fusionierungsalgorithmus

Ausgehend von der Definition des Abstandsmaßes zwischen zwei Punkten im hochdimensionalen Raum des Datensatzes  $\mathbf{X}$  muss nun der Abstand zweier Cluster, welche jeweils aus mehreren Samples bestehen, definiert werden. Hierfür existieren verschiedene Strategien. Diese kommen – teilweise mit Abwandlungen – in allen hierarchischen Cluster-Methoden zur Anwendung [77]. Sie sollen auch im Rahmen der SHC-Methode zum Einsatz kommen. Die Strategien dienen der möglichst optimalen Kombination von Clustern zu neuen, größeren Clustern. Die für diese Arbeit wichtigen Strategien sind in Abbildung 3.3 mit jeweils einem Beispiel illustriert.

Bei der *Complete-Linkage Methode* werden zunächst alle paarweisen Abstände der Samples der beiden Cluster berechnet. Anschließend wird die größte Distanz ermittelt und als Abstand der beiden Cluster verwendet (Abbildung 3.3 a). Diese Methode neigt verstärkt zur Dilatation, d.h. zur Bildung gleich großer Cluster [20].

Die *Single-Linkage Methode* verwendet die kleinste paarweise Distanz zweier Cluster als Cluster-Abstand (Abbildung 3.3 b). Es handelt sich hierbei um eine kontrahierende Methode, da sie zur Bildung weniger großer Cluster neben einer Vielzahl kleinerer Cluster neigt [20].

Die *Average-Linkage Methode* stellt einen Kompromiss aus Complete-Linkage und Single-Linkage Methode dar. Hier wird die Distanz zweier Cluster aus dem Durchschnitt aller paarweiser Distanzen der Objekte der Cluster berechnet (Abbildung 3.3 c) [78]. Diese Methode wird als konservativ bezeichnet, da sie weder zur Kontraktion noch zur Dilatation neigt.

Eine weitere Methode der Fusionierung von Clustern ist die *Median Methode*. Hier wird zunächst der Median jedes Clusters berechnet. Der Abstand zweier Cluster berechnet sich dann aus dem Abstand ihrer Mediane (Abbildung 3.3 d). Auch diese Methode kann wie die Average-Linkage Methode als konservativ bezeichnet werden [20].

Wie in den Abbildungen 3.1 und 3.2 dargestellt, fasst der SHC-Algorithmus in jeder Iteration zwei oder mehrere Cluster zu neuen Clustern zusammen. Dies kann vor allem bei der Kombination von großen Clustern zu einer stark ansteigenden Heterogenität innerhalb der Cluster führen. Um diesen Effekt abzumildern, wird in Anlehnung an Ward et al. eine Strategie angewandt, die den Anstieg der Inter-Cluster-Variabilität (ICV) bei der Kombination berücksichtigt [79, 80]. Das ursprüngliche Verfahren ist für die Kombination von maximal zwei Clustern konzipiert. Da die SHC-Methode aber in einem einzigen Iterationsschritt mehrere Cluster miteinander fusionieren kann, sind hier Anpassungen vorzunehmen.

Unter der ICV eines Clusters  $A$ , bestehend aus  $N_A$  Samples, wird die Summe der quadratischen Abweichung  $E_A$  aller Samples von dessen Mittelwert verstanden.

$$E_A = \sum_{i=1}^{N_A} (\mathbf{x}_{Ai} - \mu_A)^2 \quad (3.15)$$

Die Kombination des Clusters  $A$  mit einem anderen Cluster  $B$  führt zu einer Erhöhung der ICV  $I$ .

$$I = E_R - E_A - E_B \quad (3.16)$$

Hier bezeichnet  $E_R$  die ICV des Clusters, welcher nach der Vereinigung  $A \cup B$  entsteht. Da der Algorithmus potentiell mehrere Cluster auf einmal kombinieren kann, kann Gleichung 3.16 für die Kombination von  $J$  Clustern allgemeiner formuliert werden.

$$I = E_R - \sum_{j=1}^J E_j \quad (3.17)$$

Durch Einsetzen von Gleichung 3.15 in Gleichung 3.17 ergibt sich:

$$I = \sum_{i=1}^{N_R} (\mathbf{x}_{Ri} - \mu_R)^2 - \sum_{j=1}^J \left( \sum_{i=1}^{N_j} (\mathbf{x}_{ji} - \mu_j)^2 \right) \quad (3.18)$$

Im Idealfall gilt  $E_A + E_B = E_R$  und somit  $I = 0$ . Da dies häufig nicht der Fall ist, führt der Algorithmus in jeder Iteration nur jene Cluster-Zusammenfassung aus, welche das kleinste  $I$  aufweist und somit die ICV am wenigsten erhöht.



## 3.4 Cluster-Validitäts Koeffizienten

### 3.4.1 Übersicht

Die Cluster-Validierung ist ein wichtiger Schritt bei der Beantwortung der Frage, wie gut einzelne Cluster-Lösungen die Struktur eines gegebenen Datensatzes nachbilden. Da kein optimales Cluster-Verfahren existiert, müssen Möglichkeiten gefunden werden, die Ergebnisse unterschiedlicher Verfahren – oder eines Verfahrens mit unterschiedlicher Konfiguration – zu vergleichen. Beim Training von überwachten Lernverfahren wird hierfür die von den maschinellen Lernverfahren ermittelte Cluster-Lösung mit der tatsächlichen Cluster-Einteilung verglichen. Zu diesem Zweck existieren eine Vielzahl von Cluster-Validitäts-Koeffizienten (CV-Koeffizienten) wie z.B. der Rand-Index, der Adjusted Rand-Index sowie der Jaccard-Koeffizient, welche alle die Ähnlichkeit der berechneten Lösung mit der tatsächlichen Cluster-Lösung angeben [81, 82].

Bei unüberwachten Lernverfahren ist die tatsächliche Anzahl der Cluster in einem Datensatz in den meisten Fällen jedoch *a-priori* unbekannt. Daher werden üblicherweise eine Reihe verschiedener Cluster-Lösungen mit unterschiedlicher Cluster-Anzahl berechnet. Unter diesen wird anschließend mit Hilfe der Cluster-Validierung die optimale Lösung ermittelt. Hierfür existieren Koeffizienten, welche die Kompaktheit und Separation einer Cluster-Lösung bewerten. Wichtige Indizes sind vor allem der Silhouettenkoeffizient (SI)-Koeffizient [83], Dunn's Index [84] oder der Calinski-Harabasz-Index [85]. Eine weitere Klasse von CV-Koeffizienten bewertet die Stabilität einer Cluster-Lösung über eine Vielzahl von Durchläufen auf nur einem Datensatz. Diese als *Consensus Clustering* bekannte Methode berechnet eine Reihe von Cluster-Einteilungen mit einer bestimmten Cluster-Anzahl auf verschiedenen Versionen des Originaldatensatzes. Diese Versionen werden durch zufällige Auswahl von Samples aus dem Originaldatensatz erzeugt (Subsampling-Datensätze). Eine gute Cluster-Einteilung zeichnet sich durch eine hohe Stabilität der Einteilung aus. Das bedeutet, dass der überwiegende Teil aller Samples in jeder Lösung dem gleichen Cluster zugeordnet wird [86, 87].

In dieser Arbeit soll der SI-Koeffizient zur Bewertung von Cluster-Lösungen verwendet werden, da er sehr leicht interpretierbar ist und im Umfeld des unüberwachten maschinellen Lernens eine gute Verlässlichkeit zeigt und breite Anwendung findet [88].

### 3.4.2 Silhouettenkoeffizient

Für die Berechnung des SI-Koeffizienten wird für jedes Sample  $x_i$  zunächst folgende euklidische Abstände ermittelt:

- $d(\mathbf{x}_i, A) = \|\mathbf{x}_i - \mu_A\|_2$ : Der Abstand jedes Samples zum Median des Clusters  $A$ , dem es während des Clustering zugeordnet wurde. Diese Werte bestimmen auch die Kompaktheit der Cluster-Lösungen. Kleine Werte für  $d(\mathbf{x}_i, A)$  repräsentieren einen sehr kompakten Cluster.
- $d(\mathbf{x}_i, B) = \|\mathbf{x}_i - \mu_B\|_2$ : Der Abstand des Samples zum Median des nächstgelegenen Clusters  $B$ . Hierdurch wird die Separation der Cluster beschrieben. Große Werte für  $d(\mathbf{x}_i, B)$  bezeichnen Samples, die sich in großer Entfernung zu anderen Clustern befinden und daher als gut separierbar angesehen werden können.

Aus diesen Abständen lässt sich für jedes Sample im Datensatz ein SI-Koeffizient berechnen.

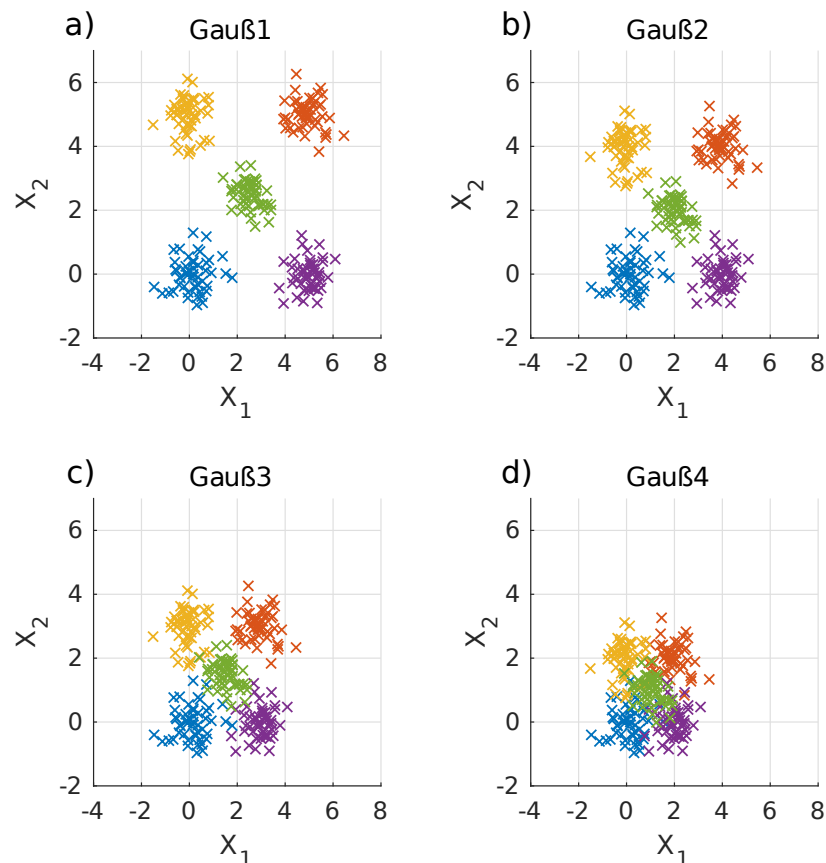
$$s_i = \begin{cases} 0 & \text{wenn } d(\mathbf{x}_i, A) = d(\mathbf{x}_i, B) = 0 \\ \frac{d(\mathbf{x}_i, B) - d(\mathbf{x}_i, A)}{\max(d(\mathbf{x}_i, A), d(\mathbf{x}_i, B))} & \text{sonst} \end{cases} \quad (3.19)$$

Aus Gleichung 3.19 ist ersichtlich, dass der SI einen Wertebereich von  $-1 \leq s_i \leq 1$  aufweist. Ein Wert  $s_i < 0$  weist auf Samples hin, die näher am Nachbar-Cluster liegen als an dem Cluster, dem sie zugewiesen wurden und stehen für eine suboptimale Cluster-Lösung. Aufgrund dieses beschränkten Wertebereichs ist der SI-Koeffizient als Bewertungsmaß für Cluster-Lösungen sehr einfach interpretierbar. Der SI-Koeffizient für den gesamten Datensatz kann als Mittelwert aller  $s_i$  berechnet werden.

$$SI = \frac{1}{N} \sum_{i=1}^N s_i \quad (3.20)$$

Dieser Wert kann für verschiedene Cluster-Lösungen mit jeweils unterschiedlicher Cluster-Anzahl  $K$  berechnet werden. Das optimale Ergebnis ist dann jenes, welches den maximalen SI-Koeffizienten aufweist. Im Bereich der Cluster-Analyse wird ein Cluster-Ergebnis dann als hinreichend valide angesehen, wenn der SI-Koeffizient größer oder gleich 0,6 ist. [89, 90, 91].

In der folgenden Validierung des Algorithmus wird das Funktionsprinzip des SI-Koeffizienten an einem einfachen Beispiel illustriert.



**Abbildung 3.4:** Validierungsdatensätze erzeugt aus fünf bivariaten Gauß-Verteilungen mit variierender Überlappung

### 3.5 Validierung der Methode anhand von Simulationsdaten

#### 3.5.1 Simulationsdatensatz und Cluster-Analyse

Die Validierung der Methode wurde anhand von simulierten Datensätzen durchgeführt. Diese sind in Abbildung 3.4 dargestellt. Wie bereits eingangs erwähnt, ist bei Daten, welche unter realen Bedingungen aufgenommen wurden, mit keiner ausgeprägten Struktur zu rechnen. Dies ist darauf zurückzuführen, dass Vitalparameter sich über einen längeren Zeitraum in der Regel nur langsam ändern. Daher kann es zu fließenden Übergängen zwischen Gesundheitszuständen, also einer Überlappung der einzelnen Cluster, kommen. Die Methode soll diese Cluster trotz ihrer Überlappung erkennen können. Um dies zu testen, wurden vier Datensätze, bestehend aus jeweils fünf Clustern, erzeugt. Jedes Cluster bestand aus 50 Samples, die aus einer bivariaten Gauß-Verteilung mit einer Kovarianzmatrix  $\Sigma = 0,25\mathbf{I}$  gezogen wurden. Die Cluster 1 bis 4 wurden an den Ecken eines Quadrates mit jeweils unterschiedlichen Kantenlängen angeordnet. Cluster 5 lag jeweils in der Mitte des Quadrats.

Datensatz	Cluster-Anzahl K								
	2	3	4	5	6	7	8	9	10
<i>Gauß1</i>	0,42	0,62	0,76	<b>0,91</b>	0,85	0,78	0,7	0,63	0,62
<i>Gauß2</i>	0,41	0,59	0,72	<b>0,85</b>	0,8	0,74	0,67	0,6	0,58
<i>Gauß3</i>	0,39	0,54	0,61	<b>0,73</b>	0,68	0,64	0,58	0,55	0,53
<i>Gauß4</i>	0,33	0,41	0,46	0,42	<b>0,54</b>	0,51	0,46	0,45	0,41

**Tabelle 3.1:** SI-Koeffizienten der aus den Datensätzen berechneten Cluster-Lösungen für  $K = 2 \dots 10$ . Lösungen, für die der berechnete SI-Koeffizient maximal wird, sind fett dargestellt.

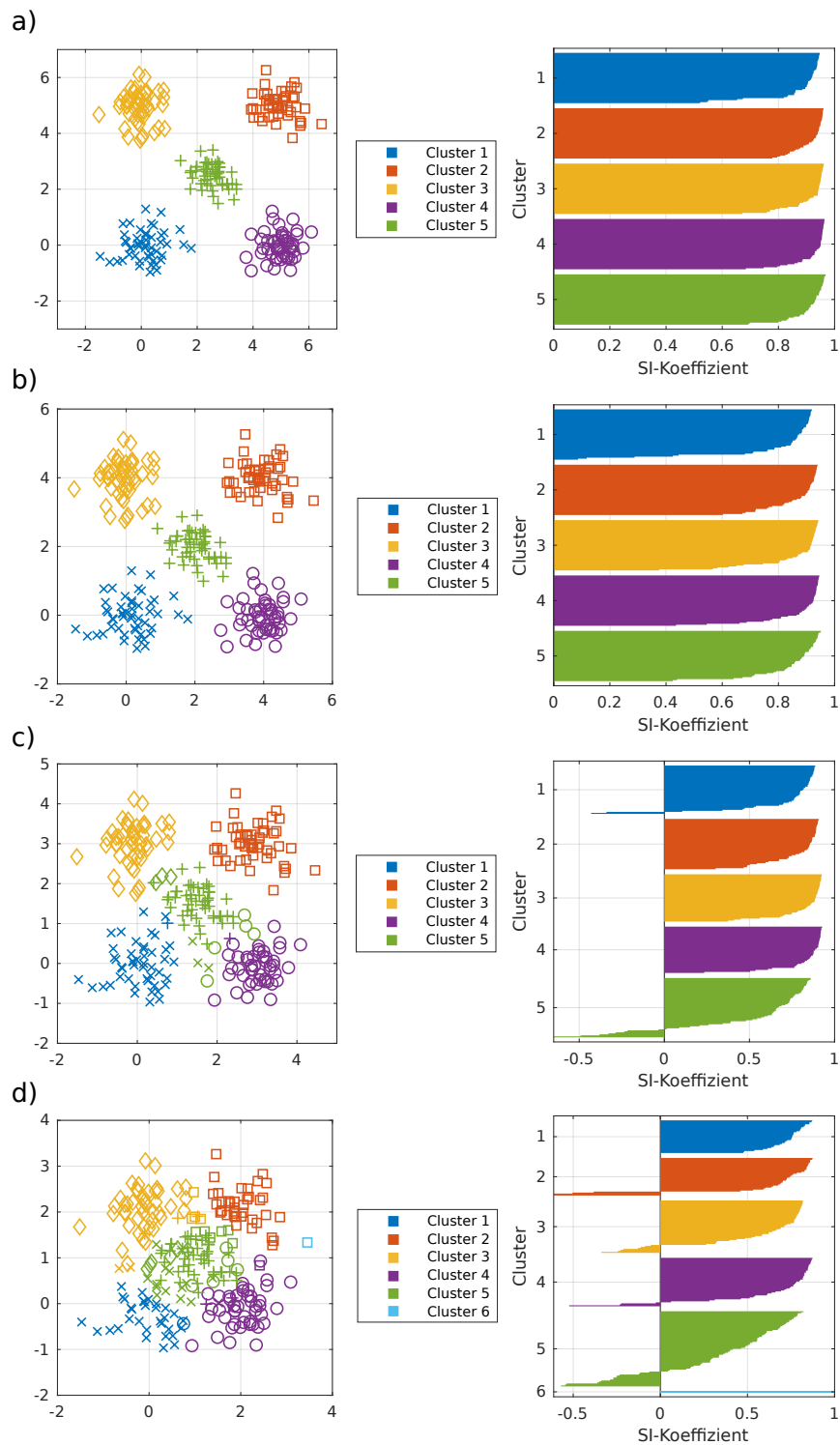
Durch Variierung der Abstände der Cluster zueinander, respektive der Kantenlänge des Quadrates, ergaben sich Datensätze mit unterschiedlich stark ausgeprägter Struktur. Basierend darauf, wurden mit Hilfe der SHC-Methode Cluster-Lösungen für  $K = 2 \dots 10$  Cluster berechnet. Zur Kombination von Clustern wurde die Median-Methode unter Verwendung des linearen Kernels angewandt. Für jede dieser Lösungen wurde ein SI-Koeffizient berechnet. Die Lösung, für die der SI-Koeffizient maximal wurde, wurden als optimale Cluster-Lösung verwendet.

### 3.5.2 Ergebnisse

Die Ergebnisse des Clusterings der Datensätze *Gauß1*, *Gauß2*, *Gauß3* und *Gauß4* sind in Abbildung 3.5 zu sehen. Es handelt sich dabei um simulierte Daten, welche fünf Cluster mit unterschiedlich starkem Überlappungsgrad enthalten. Es wurden Cluster-Lösungen mit  $K = 2 \dots 10$  Clustern auf ihre Validität untersucht. Dazu wurde für jede Lösung der SI-Koeffizient berechnet (siehe Tabelle 3.1).

Der Datensatz *Gauß1* weist den größten Inter-Cluster-Abstand auf. Die Cluster 1 bis 4 liegen an den Ecken eines Quadrates mit der Kantenlänge 4. Cluster 5 liegt in der Mitte des Quadrats. Anhand der SI-Koeffizienten in Tabelle 3.1 ist zu sehen, dass das Clustering für  $K = 5$  eine optimale Cluster-Lösungen unter allen untersuchten Lösungen darstellt. In Abbildung 3.5 a) ist das für  $K = 5$  gefundene Ergebnis des Datensatzes *Gauß1* zu sehen. Es ist erkennbar, dass der Algorithmus die Cluster-Einteilung aus Abbildung 3.4 a) korrekt rekonstruiert. Der dazugehörige SI-Plot (Abbildung 3.5 a) rechts) zeigt die zur optimalen Cluster-Lösung gehörigen SI-Koeffizienten der Samples. Aufgrund der klaren Trennung bewegen sich die SI-Werte zwischen 0,5 und 1.

Ein ähnliches Ergebnis liefert das Cluster-Verfahren mit dem Datensatz *Gauß2* (Abbildung 3.5 b). Der SI-Koeffizient erreicht mit 0,85 für  $K = 5$  ein Maximum. Damit entspricht auch



**Abbildung 3.5:** Ergebnisse des Clusterings der simulierten Datensätze. Links: Optimale Cluster-Einteilung der Datensätze aus Abbildung 3.4, basierend auf SI-Koeffizienten (Tabelle 3.1). Die Marker-Form gibt die tatsächliche Cluster-Zugehörigkeit der Samples an. Rechts: zugehörige Silhouetten-Plots der optimalen Cluster-Lösungen.

hier die erkannte Cluster-Anzahl genau der tatsächlich vorhandenen Cluster-Anzahl. Die Cluster-Struktur wird korrekt rekonstruiert. Es treten keine falsch zugeordneten Samples auf. Wie im zugehörigen SI-Plot zu sehen ist, werden die meisten Samples sehr klar ihrem jeweiligen Cluster zugeordnet. Aufgrund der kleiner werdenden Inter-Cluster-Abstände des Datensatzes weisen einige Samples kleinere SI-Koeffizienten als noch beim Datensatz *Gauß1* auf. Es ist zu vermuten, dass dies vor allem Samples am Rand der Cluster betrifft.

Im Datensatz *Gauß3* haben die fünf Cluster einen noch geringeren Abstand zueinander und überlappen einander teilweise. Zwar wird der SI-Koeffizient für  $K = 5$  maximal ( $SI_{K=5} = 0,73$ ), allerdings kommt es zu einer ersten fehlerhaften Zuordnungen der Samples (Abbildung 3.5 c). Es ist zu sehen, dass Cluster 5 in Cluster 4 hinein gewachsen ist. Auch aus Cluster 3 wurden einige Samples fehlerhaft dem Cluster 5 zugeordnet. Anhand des dazugehörigen SI-Plots ist deutlich zu sehen, dass einige Samples, welche Cluster 5 zugeprochen wurden, negative SI-Koeffizienten aufweisen. Dies deutet auf eine suboptimale Zuordnung dieser Samples hin. Auch bei Cluster 1 ist dies, wenn auch in geringerem Maße, zu sehen.

Der Datensatz *Gauß4* hat von allen betrachteten Datensätzen den geringsten Inter-Cluster-Abstand und die größte Cluster-Überlappung. In Tabelle 3.1 ist zu sehen, dass das SI-Maß bei einer Cluster-Lösung mit  $K = 6$  am geringsten ist, was nicht der tatsächlichen Anzahl an Clustern im Datensatz *Gauß4* entspricht. Abbildung 3.5 d) zeigt aber, dass für diese Cluster-Anzahl die Cluster-Einteilung trotzdem weitestgehend dem Originaldatensatz entspricht. Das sechste Cluster besteht aus nur einem Sample. Weiterhin kann der Abbildung entnommen werden, dass die Zahl der falsch zugeordneten Samples vor allem am Rande von Cluster 5 weiter zugenommen hat. Aufgrund der stärkeren Überlappungen kommt es auch am Rand aller anderen Cluster zu falsch zugeordneten Samples. Dies hat zur Folge, dass diese Samples einen SI-Koeffizienten kleiner Null erhalten. Weiterhin ist aufgrund dessen auch der mittlere SI-Koeffizient der Cluster Lösung mit  $SI_{K=6} = 0,54$  der kleinste optimale Wert aller betrachteten Datensätze.

### 3.5.3 Diskussion

Zur Validierung der ähnlichkeitsbasierten hierarchischen Clustering-Methode wurden vier Datensätze mit unterschiedlich stark ausgeprägter Struktur erzeugt. Diese Datensätze bestanden jeweils aus fünf Clustern mit variierendem Grad der Überlappung. Die Verteilung der Samples in jedem Cluster entsprach jeweils einer bivariaten Gaußverteilung. Auf Basis dieser Datensätze wurden Cluster-Lösungen mit einer variierenden Cluster-Anzahl ( $K = 2 \dots 10$ ) erzeugt. Mit Hilfe des SI-Koeffizienten wurden diese bewertet. Diejenige Lösung, welche

den höchsten SI-Wert aufwies, wurde als optimale Lösung verwendet. Es sollte gezeigt werden, dass die Cluster-Methode in der Lage ist, die korrekte Cluster-Struktur in diesen Datensätzen zu rekonstruieren.

Es hat sich gezeigt, dass vor allem bei gut separierten Datensätzen (*Gauß1* und *Gauß2*) die rekonstruierte Cluster-Struktur der tatsächlichen Struktur perfekt entsprach. Der SI-Koeffizient wurde in beiden Fällen für  $K = 5$  maximal.

Erste Fehler bei der Zuordnung der Samples traten bei höherer Überlappung der Cluster auf. Wie anhand des Datensatzes *Gauß3* zu sehen ist, kommt es vor allem im Bereich des mittleren Clusters 5 zu Fehlern. Hier wurden einige Samples fälschlich den Eck-Clustern 1 bis 4 zugesprochen. Trotzdem ist der Algorithmus in der Lage, die Struktur weitgehend zu rekonstruieren. Der SI-Koeffizient wurde für  $K = 5$  maximal.

Der Datensatz *Gauß4* weist die höchste Cluster-Überlappung auf. Bei der Rekonstruktion der Struktur kam es zu den meisten Fehlern aller verwendeter Datensätze. Diese traten auch hier wieder an den Rändern des mittleren Clusters auf. Der SI-Koeffizient wurde für  $K = 6$  maximal. Dies entspricht nicht der eigentlichen Cluster-Anzahl. Die Lösung ist jedoch trotzdem als gut zu bewerten, da die meisten Samples richtig in die Cluster 1 bis 5 eingeteilt wurden. Der zusätzliche sechste Cluster besteht nur aus einem Sample (Abbildung 3.5 d) hellblau markiertes Sample).

Zusammenfassend kann festgestellt werden, dass die Methode sehr gute Ergebnisse für gut separierte Cluster erzielt. Bei Datensätzen mit geringer ausgeprägter Struktur kommt es zu fehlerhaften Zuordnungen im Bereich der Cluster-Überlappungen. Zusätzlich wird bei starker Überlappung eine höhere Cluster-Anzahl benötigt, um die Struktur des Datensatzes abzubilden. Diese Tatsache kann aus funktioneller Sicht als Fehler angesehen werden, erscheint aus einer physiologischen Perspektive jedoch als sinnvoll. Wie eingangs erwähnt, ändern sich Vitalparameter selten sprunghaft. Vielmehr ist mit einem graduellen Übergang zwischen Gesundheitszuständen (vor allem über eine lange Zeit gesehen) zu rechnen. Daher bildet wohl vor allem der Datensatz *Gauß4* die Realität sehr gut ab. Vor diesem Hintergrund ist es physiologisch plausibel, innerhalb eines Datensatzes zusätzliche Übergangs-Cluster zwischen großen Cluster-Strukturen zu finden. Abhängig von der Geschwindigkeit der Vitalparameteränderung sollten diese Übergangs-Cluster jedoch nur eine relative kleine Anzahl an Samples enthalten.

Im Folgenden wird die Validität des Algorithmus anhand eines realistischen Datensatzes, bestehend aus echten Vitaldaten, weiter untersucht.

## 3.6 Validierung der Methode anhand realistischer Daten

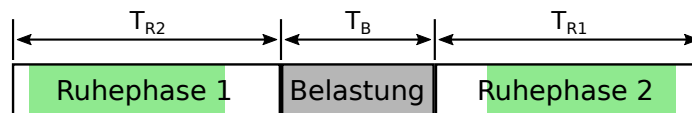
### 3.6.1 Experimentelles Setup und Datenauswertung

Um das Clustering-Verfahren mit realistischen Daten zu evaluieren, wurde ein Experiment durchgeführt, bei dem menschliche Probanden Phasen unterschiedlicher physischer Belastung durchliefen. Der experimentelle Ablauf ist in Abbildung 3.6 dargestellt. Ziel ist es, anhand der simultan aufgenommenen Vitaldaten diese Phasen mit Hilfe der ähnlichkeitsbasierten hierarchischen Cluster-Methode zu rekonstruieren.

Es wurden insgesamt 33 gesunde Probanden aus den Studenten und Mitarbeitern der Arbeitsgruppe rekrutiert. Die Kohorte bestand aus 23 Männern und 10 Frauen mit einem mittleren Alter von 26,8 Jahren ( $\sigma = 5,5$  Jahre). Alle Probanden nahmen freiwillig am Experiment teil. Die beiden Ruhephasen verbrachten die Probanden liegend. Während der Belastungsphase führten die Probanden Kniebeugen durch. Im Verlauf des Experiments wurde mittels eines PLUX biosignals researchers<sup>1</sup> ein Einkanal-EKG ( $f_s = 900$  Hz) aufgezeichnet. Weiterhin wurden die Übergänge zwischen den Belastungsphasen mittels eines Knopfdrucks in der EKG-Aufzeichnung markiert. Die Vorverarbeitung des EKG-Signals bestand aus einer Baseline-Korrektur sowie der Unterdrückung der 50 Hz Interferenz mittels eines FIR-Filters [92]. Anschließend wurde das EKG-Signal in Abschnitte von 60 Sekunden Länge und einer Überlappung von 55 Sekunden unterteilt. In jedem Messfenster wurden die QRS-Komplexe mittels eines Hamilton-Tompkins QRS-Detektors bzw. eines wavelet-basierten QRS-Detektors extrahiert [93, 94]. Außerdem wurde eine T-Wellen-Detektion durchgeführt [95]. Basierend auf den detektierten QRS-Komplexen sowie den T-Wellen wurden folgende Parameter ermittelt: (i) die mittlere Herzrate (HR) und (ii) die Zeitdauer zwischen dem Onset des QRS-Komplexes und dem Offset der darauf folgenden T-Welle (Q-TE). Es ist zu erwarten, dass physische Belastung zu einer Erhöhung der Herzrate sowie einer Verkürzung der Q-TE-Zeit führt [96].

Die abgeleiteten Parameter bildeten den Datensatz für das anschließende Clustering. Dabei wurde die Median-Methode unter Verwendung des linearen Kernels angewandt. Für jeden Probanden wurden Cluster-Lösungen mit  $K = 2 \dots 10$  Clustern erstellt. Für jede dieser Lösungen wurde ein SI-Koeffizient berechnet. Die Lösung mit dem größten SI-Koeffizienten wurde als endgültige Lösung für den jeweiligen Probanden angenommen.





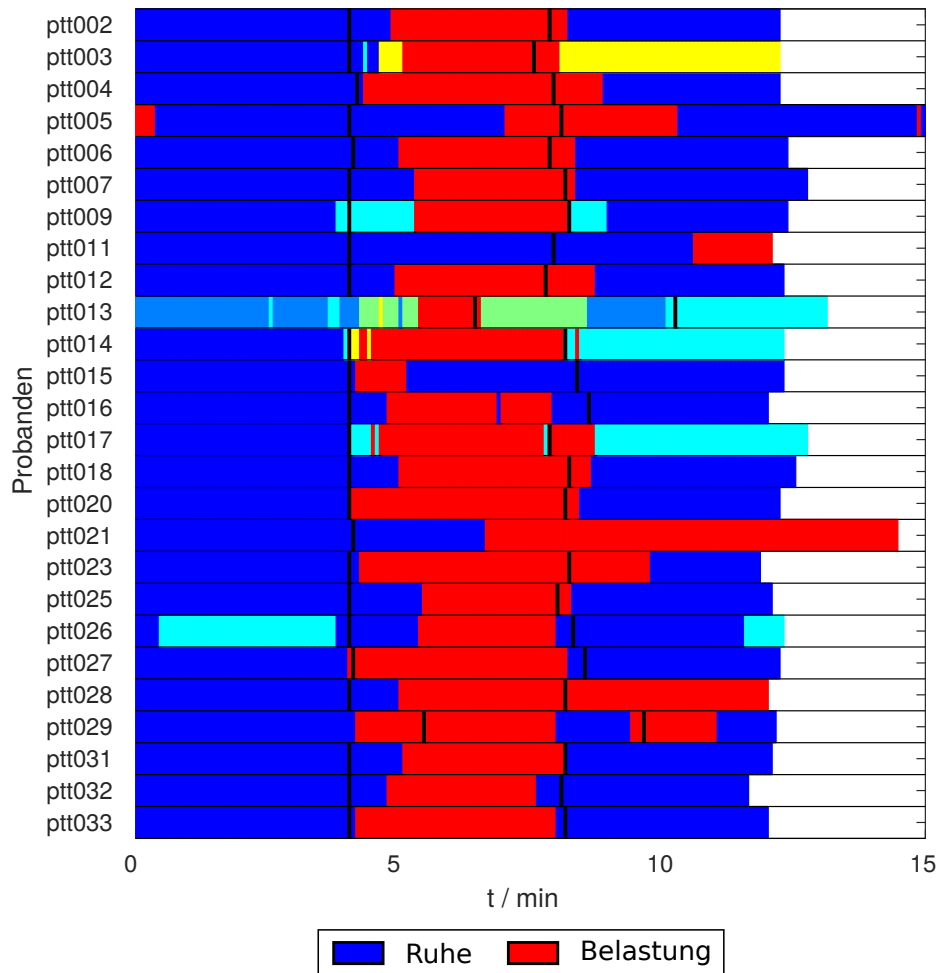
**Abbildung 3.6:** Ablauf des Validierungsexperiments besteht aus zwei Ruhephasen von jeweils  $T_{R1} = 5$  min und  $T_{R2} = 5$  min Länge sowie einer Belastungsphase der Länge  $T_B = 3$  min.

### 3.6.2 Ergebnisse

Bei 7 der 33 Probanden traten während der Belastungsphase starke Bewegungsartefakte im EKG auf. Dies führte zu einer sehr fehlerbehafteten QRS-Detektion und zum Ausschluss dieser Probanden von der weiteren Analyse. Abbildung 3.7 zeigt für die verbleibenden 26 Probanden farbkodiert die Cluster-Lösungen, welche anhand des SI-Koeffizienten als optimal identifiziert wurden. In Tabelle B.1 sind die zu allen berechneten Cluster-Lösungen gehörigen SI-Koeffizienten zu finden. Außer bei den Probanden *ptt011* und *ptt021* konnten bei der Mehrzahl der Probanden plausible Ergebnisse mit zwei unterschiedlichen Gesundheitszuständen gefunden werden. Es ist sofort ersichtlich, dass die Methode am Anfang jeder Aufzeichnung konsistent über alle Probanden einen Ruhezustand identifizieren konnte (dargestellt in dunkelblau). Während dieser Phase befanden sich die Probanden in Ruhelage, weshalb sich die aufgezeichneten Vitalparameter nur wenig änderten.

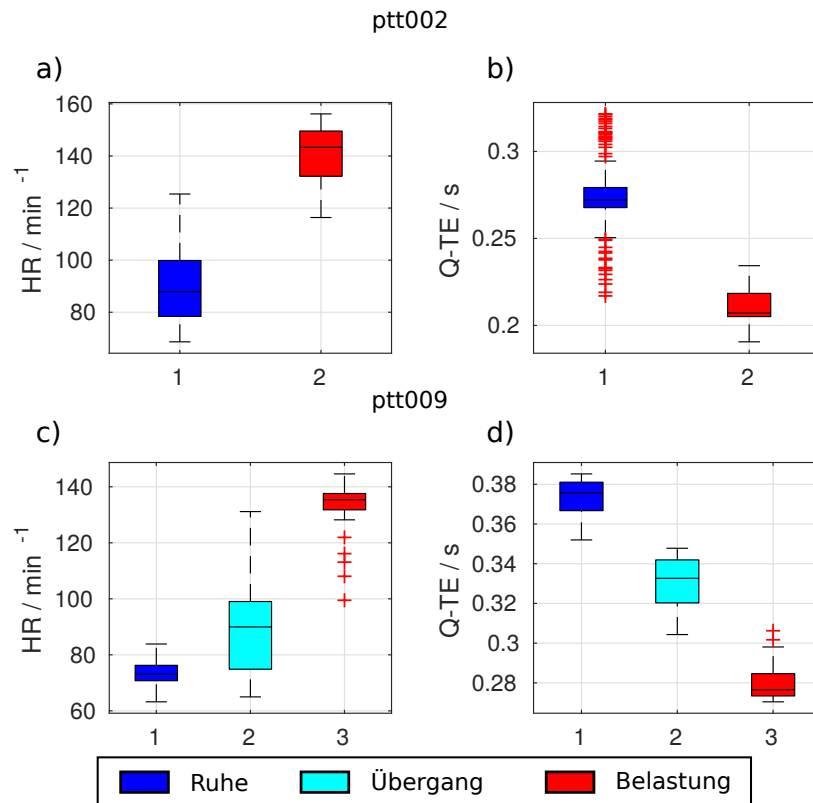
In Abbildung 3.8 sind die Verteilungen der Vitalparameter in den einzelnen Gesundheitszuständen für zwei Probanden beispielhaft dargestellt. Proband *ptt002* (Abbildung 3.8 a) und b)) zeigt in der ersten erkannten Ruhephase eine Herzrate im Bereich von 80 bis 100 Beats per Minute (bpm) sowie eine Q-TE Zeit im Bereich von 200 bis 300 Millisekunden. An die erste Ruhephase schließt sich eine Phase physischer Belastung an, deren Beginn durch einen Tastendruck markiert wurde. Es ist zu sehen, dass die Methode bei den meisten Probanden einen Wechsel des Gesundheitszustandes in zeitlicher Nähe zu diesem Tastendruck erkannt hat. Diese Belastungsphase (rot) ist durch eine erhöhte Herzrate sowie eine verkürzte Q-TE-Zeit charakterisiert. In den Abbildungen 3.8 a) und b) ist zu erkennen, dass sich die Herzrate während der Belastungsphase signifikant erhöhte. Außerdem war ein starker Rückgang der Q-TE Zeit zu verzeichnen. Das Ende des erkannten Belastungszustandes ist bei den meisten Probanden wiederum in zeitlicher Nähe zum zweiten Tastendruck zu finden, welcher das Ende der Belastungsphase markiert. In der abschließenden zweiten Ruhephase kehrten die Vitalparameter langsam wieder zu ihren Ruhewerten zurück. Daher wird auch wieder der Gesundheitszustand vor der Belastungsphase (dunkelblau) eingenommen. Es ist ersichtlich, dass der Übergang von Ruhezustand zu Belastungszustand und wieder zum Ruhezustand

<sup>1</sup>PLUX wireless biosignals, Lisabon, Portugal (<http://www.biosignalsplux.com/en/>)



**Abbildung 3.7:** Cluster-Ergebnisse aller berücksichtigten Probanden. Der Übergang in eine andere experimentelle Phase ist in Schwarz dargestellt. Erkannte Ruhezustände (dunkelblau) und Belastungszustände (rot) sind farblich dargestellt. Darüber hinaus erkannte (Zwischen-) Zustände sind andersfarbig markiert (siehe auch Abb. 3.8).

zurück oftmals zeitverzögert zum dazugehörigen Tastendruck erfolgt. Dies ist zum Beispiel bei Proband *ptt006* der Fall. In Abbildung 3.7 ist zu sehen, dass bei den meisten Probanden eine Einteilung des Datensatzes in zwei Cluster (Ruhe- und Belastungszustand) als optimal angesehen wird. Dennoch gibt es Probanden (z.B. *ptt009* und *ptt017*), bei denen eine größere Anzahl von Clustern als optimale Einteilung des Datensatzes angesehen wurde. Diese zusätzlichen Cluster sind oft im Bereich vor und nach dem erkannten Belastungszustand zu finden. Bei Proband *ptt009* sind diese Cluster in Abbildung 3.7 in hellblau markiert. In den Abbildungen 3.8 c) und d) ist zu erkennen, dass diese zusätzlichen Cluster Übergangszustände darstellen. So zeigt die HR des Übergangszustandes bei Proband *ptt009* Werte zwischen 75 und 100 bpm, welche zwischen Ruhezustand (ca. 70 bpm) und Belas-



**Abbildung 3.8:** Verteilung der Parameter HR und Q-TE in den erkannten Gesundheitszuständen für die Probanden *ptt002* und *ptt009*: Die Farben korrespondieren dabei mit den Farben der Gesundheitszustände in Abbildung 3.7.

tungszustand (ca. 135 bpm) liegen. Auch die Q-TE Zeit des Übergangszustandes befindet sich zwischen der des Ruhezustandes und der des Belastungszustandes.

Ein weiterer Zustand konnte bei Proband *ptt003* während der zweiten Ruhephase identifiziert werden. Dieser Zustand ist in Abbildung 3.7 gelb markiert. Auch hierbei handelt es sich um einen Ruhezustand, wobei sich die Parameter vom ersten Ruhezustand (blau) wertemäßig unterscheiden. Der Proband kehrte nach der Belastungsphase nur sehr langsam in den Ruhezustand zurück, weshalb die SHC-Methode hier einen separaten Gesundheitszustand identifizierte.

### 3.6.3 Diskussion

Die Ergebnisse des Clusterings zeigen, dass die hier vorgestellte ähnlichkeitsbasierte hierarchische Cluster-Methode in der Lage ist, unterschiedliche Belastungsphasen anhand von Vitaldaten zu erkennen. Die Mehrzahl der analysierten Datensätze wurde in zwei Cluster unterteilt, von denen ein Cluster die beiden Ruhephasen und der zweite Cluster die Belastungsphase repräsentierte. Diese sind in Abbildung 3.7 jeweils in dunkelblau und rot

dargestellt. Zustandsübergänge fanden in zeitlicher Nähe zu den tatsächlichen Änderungen der physischen Belastung statt.

Trotzdem stimmen die Zustandsübergänge mit den Übergängen der experimentellen Phasen nie vollkommen überein. Der Grund hierfür liegt zum Teil im Design des Experiments. Der Tastendruck markiert den Beginn der Belastungsphase nicht mit hundertprozentiger Genauigkeit. Bis die Probanden tatsächlich mit den Kniebeugen begannen, vergingen in der Regel einige Sekunden. Weiterhin benötigt der menschliche Körper einige Zeit, um auf die veränderte physische Belastung zu reagieren. Abhängig von der körperlichen Fitness des jeweiligen Probanden erhöht sich beispielsweise die Herzrate mehr oder weniger schnell bzw. kehrt mehr oder weniger schnell zu den Ruhewerten zurück. Dies ist der Grund, weshalb der Übergang in den Belastungszustand immer nach dem ersten Knopfdruck detektiert wurde. Dies ist ebenfalls der Grund dafür, dass in den meisten Fällen die Rückkehr in den Ruhezustand nach dem zweiten Knopfdruck detektiert wurde. Auch hier vergeht Zeit, bis der Körper infolge der nachlassenden physischen Belastung wieder in den Ruhezustand zurückkehrt.

Bei einige Probanden wurden außer den beiden Ruhe- und Belastungszuständen weitere Gesundheitszustände identifiziert, die jedoch häufig nur kurz andauern. Diese sind in den meisten Fällen zeitlich vor und nach dem Belastungszustand zu finden. Es handelt sich hierbei um eine Form von Übergangszuständen. Der Kreislauf der jeweiligen Probanden reagierte besonders langsam auf die sich ändernde körperliche Anstrengung, was den Algorithmus dazu veranlasste, diese langsame Veränderung als eigenen Cluster zu identifizieren. Dies weist auch auf einen weniger stark strukturierten Datensatz hin. Dieser Effekt ist in Abbildung 3.8 c) und d) zu sehen. Hier existiert zwischen den Ruhe- und Belastungszuständen (jeweils dunkelblau und rot) ein weiterer Zustand (hellblau). Die Vitalparameter dieses Zustandes zeigen einen Wertebereich zwischen Belastung und Ruhe. Diese Beobachtung ist konsistent mit den Ergebnissen aus Abschnitt 3.5, wo bei einem Datensatz mit starker Cluster-Überlappung (*Gauß4*) eine größere Anzahl Cluster zur Rekonstruktion der Struktur des Datensatzes nötig war.

Mit Hilfe dieses einfachen Experiments konnte die ähnlichkeitsbasierte hierarchische Cluster-Methode validiert werden. Es ist gelungen, für jeden der ausgewerteten Probanden zeitlich kompakte Gesundheitszustände zu identifizieren. Es wurden realistische Datensätze verwendet, welche auf aus dem EKG abgeleiteten Parametern basieren. Dass die während des Clusterings gefundenen Gesundheitszustände tatsächlich mit den Belastungsphasen im experimentellen Ablauf korrespondieren, konnte bewiesen werden. Die zeitliche Abfolge der Gesundheitszustände sowie deren Übergänge zeigte eine hohe zeitliche Korrelation mit

den Belastungsphasen (Abbildung 3.7). Außerdem zeigten auch die Vitalparameter plausible Werte in den einzelnen erkannten Zuständen (Abbildung 3.8). Damit konnte nachgewiesen werden, dass die hier vorgestellte Methode mit Daten funktioniert, welche in der Form auch regelmäßig im klinischen Umfeld erhoben werden. Weiterführende Untersuchungen – welche hier jedoch nicht im Detail dargestellt werden sollen – haben die Methode mit etablierten unüberwachten Lernverfahren verglichen. Dabei wurden die k-means Methode und die agglomerative hierarchische Cluster-Methode jeweils auf den gleichen Datensatz angewandt (für Details zu den Verfahren siehe die Abschnitte 2.4.1 und 2.4.2). Diese Referenzverfahren lieferten vergleichbare oder schlechtere Ergebnisse hinsichtlich der Genauigkeit der Rekonstruktion der Belastungszustände [97].

### **3.7 Zusammenfassung**

In diesem Kapitel wurde ein unüberwachtes maschinelles Lernverfahren vorgestellt, mit welcher sich Vitalparameter in zeitlich kompakte Gesundheitszustände anhand ihrer Ähnlichkeit einteilen lassen. Die Methode kombiniert diese Gesundheitszustände zu neuen Zuständen anhand ihrer Abstände im von den Vitaldaten aufgespannten Raum. Dafür wurde auf bekannte und etablierte Abstandsmaße im euklidischen Vektorraum und Cluster-Kombinationsstrategien zurückgegriffen. Diese wurden für die Verwendung mit dem neuen Verfahren angepasst. Auf diese Weise gelingt eine multimodale Repräsentation der Entwicklung des Patientenzustandes über der Zeit. Aufgrund der Kombination von Gesundheitszuständen zu neuen Gesundheitszuständen entsteht eine hierarchische Struktur. Dies erlaubt eine Einteilung der Entwicklung des Patientenzustandes mit variabler zeitlicher Granularität.

Zur Entwicklung des Verfahren wurde zunächst ein Überblick über den Ablauf des Verfahrens gegeben. Anschließend wurde auf einige wichtige Aspekte dieses Ablaufs genauer eingegangen. Um einzelne Cluster-Lösungen miteinander zu vergleichen und die optimale Cluster-Einteilung eines Datensatzes zu finden, wurde der SI-Koeffizient beschrieben. Dabei handelt es sich um ein etabliertes Gütekriterium, das im Bereich des unüberwachten maschinellen Lernens häufig eingesetzt wird und gute Resultate liefert.

Zur Validierung der Methode wurden zwei Studien konzipiert. Zunächst wurden simulierte Datensätze mit bekannter Cluster-Anzahl und unterschiedlich stark ausgeprägter Struktur erstellt. Die Variation der Struktur wurde durch eine variierende Cluster-Überlappung im Datensatz realisiert. Bei Datensätzen mit geringer Cluster-Überlappung war das Verfahren in der Lage die Cluster-Struktur perfekt zu rekonstruieren. Bei stärker werdender Überlappung zeigte sich jedoch, dass einzelne Samples dem falschen Cluster zugeordnet wurden. Dies war vor allem im Bereich der Überlappung vermehrt der Fall. Zusätzlich wurden bei un-

strukturierten Datensätzen mehr Cluster erkannt als tatsächlich vorhanden waren. Trotzdem wurde die Struktur des Datensatzes gut rekonstruiert, da die zusätzlichen Cluster nur wenige Samples beinhalteten.

Der zweite Teil der Validierung bestand in der Erzeugung eines strukturierten Datensatzes aus realen Vitaldaten. Hierzu wurde ein Experiment mit menschlichen Probanden durchgeführt. Die Struktur des Datensatzes wurde erzeugt, indem die Probanden Phasen von körperlicher Belastung und Ruhe durchliefen. Ziel war es, mit Hilfe des hier vorgestellten Cluster-Verfahrens diese Phasen zu rekonstruieren. Für die Mehrzahl der untersuchten Probanden wurden zwei Gesundheitszustände gefunden, welche eine hohe zeitliche Korrelation mit den tatsächlichen Belastungsphasen des Experiments aufwiesen. Bei einigen Probanden wurden zusätzliche Gesundheitszustände gefunden, welche als Übergangszustände zwischen Ruhe- und Belastungsphasen angesehen werden können.

Insgesamt haben die durchgeführten Untersuchungen gezeigt, dass die hier vorgestellte ähnlichkeitsbasierte hierarchische Cluster-Methode zur Identifikation von Gesundheitszuständen geeignet ist. Dabei konnten signifikante Unterschiede in den betrachteten Vitalparametern zwischen Ruhe- und Belastungszustand festgestellt werden.

## **Kapitel 4**

### **Detektion von Schmerzmittelgaben nach herzchirurgischen Eingriffen**

#### **4.1 Methode**

##### **4.1.1 Studiendesign**

Die Messstudie wurde von Februar 2017 bis Juli 2017 in Zusammenarbeit mit dem Herzzentrum Leipzig durchgeführt. Es wurden 25 männliche Probanden (mittleres Alter: 64,8 Jahre) mit geplanter koronarer Bypass-Operation (Aortocoronary Bypass Operation (ACB-OP), minimally invasive direct coronary artery bypass (MIDCAB), off-pump coronary artery bypass (OPCAB)) rekrutiert. Die Probanden wurden durch den behandelnden Arzt am Tag vor der OP über Zweck und Inhalt der Studie aufgeklärt und gaben ihre schriftliche Einwilligung zur Teilnahme. Die Studie wurde am 20. Juni 2016 durch die Ethik-Kommission der Universität Leipzig (Aktenzeichen 203-16/ek) zur Durchführung zugelassen. Alle personenbezogenen Daten wurden anonymisiert erhoben. Die Aufzeichnung wichtiger Vitalparameter begann zeitnah nach Abschluss des Eingriffs und der Verlegung des Patienten auf die Intermediate Care (IC)-Station. Bei der IC-Station handelt es sich nicht um eine klassische Intensivstation mit entsprechend engmaschiger Betreuung. Dementsprechend ist die Personaldecke geringer und die einzelnen Pflegekräfte meist jeweils für mehrere Patienten zuständig. Trotzdem kommt der Überwachung der Patienten in den Stunden nach der Operation eine große Bedeutung zu. Durch eine kontinuierliche Aufzeichnung vor allem der Herzaktivität kann schnell und effektiv auf Komplikationen reagiert werden. Auch andere Parameter (Blutdruck, Blutsauerstoffsättigung etc.) werden erhoben, was zu einer sehr großen Menge an Daten und Informationen führt, die vom Pflegepersonal interpretiert werden müssen. Hier bietet sich ein Anwendungsfall für maschinelle Lernverfahren. Im folgenden Kapitel soll daher die Anwendbarkeit der ähnlichkeitsbasierten hierarchischen Cluster-Methode demonstriert werden. Konkret soll die Wirkung von Schmerzmitteln, welche dem Patienten zur Behandlung postoperativer Schmerzen verabreicht werden, anhand verschiedener Vitalparameter detektiert werden. Es wird eine Studie in Zusammenarbeit mit dem Herzzentrum Leipzig durchgeführt, bei der Patienten während der ersten 24 Stunden nach der Bypass-Operation

durch ein mobiles Messsystem zusätzlich überwacht werden. Aus den aufgezeichneten Daten werden Vitalparameter abgeleitet, anhand derer die Zeitpunkte der Schmerzmittelgaben rekonstruiert werden sollen. Hierfür sollen für die abgeleiteten Vitalparameter Trends definiert werden, die für eine Gabe von Schmerzmitteln typisch sind. Die dabei gewonnenen Ergebnisse werden mit denen der in Kapitel 2 vorgestellten Referenz-Methoden verglichen.

Da die Überwachungsgeräte der IC-Station kein kontinuierliches Speichern der Biosignale (bspw. des EKG) zuließen, wurden die Patienten mit einem zusätzlichen Messsystem (PLUX Biosignals researcher) ausgestattet. Dieses System erlaubt die mehrkanalige Aufzeichnung von Biosignalen mit hoher zeitlicher Auflösung. Es wurden folgende Signale mit einer Abtastfrequenz von  $f_s = 1000$  Hz erfasst:

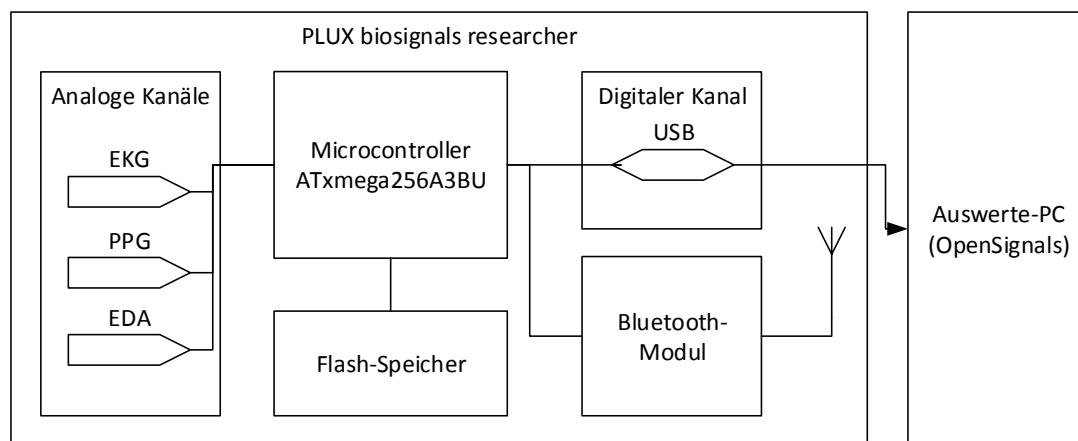
1. EKG
2. Photoplethysmogramm (PPG)
3. Elektrodermale Aktivität (EDA)

Weiterhin wurde während der Messung durch das Pflegepersonal ein Protokoll ausgefüllt. Einzutragen waren Daten zu Art und Zeitpunkt von Schmerzmittelgabe, Mobilisierung und Nahrungsaufnahme des Probanden. Außerdem sollte in regelmäßigen Abständen, mindestens jedoch einmal pro Schicht, der Patient nach seiner Schmerzintensität auf einer numerischen Rating-Skala von 1 bis 10 befragt werden. Hier steht „0“ für „keine Schmerzen“ und „10“ für „stärkste vorstellbare Schmerzen“. Zusätzlich wurden nach Abschluss der Messung weitere Schmerzmittelgaben sowie die stündlich gemessenen Blutdruckwerte aus der Patientenakte protokolliert.

#### 4.1.2 Verwendete Messtechnik und Messaufbau

In Abbildung 4.1 ist die zur Aufzeichnung von Biosignalen verwendete Hardware abgebildet. Es wurde das Messsystem *biosignals researcher* der Firma PLUX Wireless Biosignals verwendet. Dabei handelt es sich um ein mobiles System mit bis zu acht analogen Messkanälen und einem digitalen Messkanal, welche Biosignale mit einer Abtastfrequenz von bis zu 1 kHz messen kann. Zur Speicherung von Messdaten ist ein 8 GB großer interner Flash-Speicher vorhanden. Die Kommunikation mit der Messsoftware OpenSignals, welche auf einem PC (Ubuntu Linux 16.04 LTS) läuft, wird über Bluetooth (Bluetooth 2.0 + EDR) realisiert. Über den digitalen Kanal steht mit Hilfe eines Adapters auch eine Kommunikationsmöglichkeit über Universal Serial Bus (USB) zur Verfügung.





**Abbildung 4.1:** Messaufbau zur Aufzeichnung von Biosignalen bei der Durchführung der Messstudie am Herzzentrum Leipzig.

Vor Beginn der Messung wurde das PLUX Messsystem über Bluetooth auf eine zeitgesteuerte Aufzeichnung der in Abschnitt 4.1.1 aufgelisteten Parameter programmiert. Hierfür diente die vom Hersteller bereit gestellte Software OpenSignals. Nachdem alle Sensoren am Probanden angebracht waren, wurde die Messung mittels Knopfdruck gestartet. Das System beendete die Messung selbstständig nach Ablauf von 24 Stunden. Anschließend wurden die Daten über USB-Verbindung ausgelesen. In Übereinstimmung mit dem Ethikantrag wurden die Daten pseudonymisiert abgespeichert. Die weitere Verarbeitung erfolgte mit Hilfe von MATLAB in der Version R2017b.

### 4.1.3 Vorverarbeitung der Biosignale

Nach Herunterladen der Daten vom Messsystem wurde zunächst mittels eines IIR-Filters in allen Biosignalen das 50-Hz Störsignal unterdrückt. Im EKG wurde mittels eines Finite Impulse Response (FIR)-Tiefpassfilters ( $f_{cut} = 70$  Hz, Hann-gewichtet) und eines wavelet-basierten Filters hochfrequente Störanteile unterdrückt [98, 99, 100]. Die anschließende Baseline-Korrektur wurde mit Hilfe eines zeitabhängigen FIR-Filters durchgeführt [92]. Den nächsten Schritt bildete die Extraktion der QRS-Komplexe aus dem EKG-Signal. Zu diesem Zweck kamen ein Hamilton-Tompkins QRS-Detektor und ein wavelet-basierter QRS-Detektor zum Einsatz [93, 94]. Bei der optischen Kontrolle der EKG-Annotation zeigte sich eine hohe Zahl an Falsch-positiv-Ergebnissen bei der Detektion der QRS-Komplexe. Vor allem stark ausgeprägte T-Wellen wurden durch beide Detektoren häufig als R-Zacke erkannt. Dies hängt mit einer teilweise ungünstigen Elektrodenposition bei der Aufzeich-

nung des EKG zusammen. Aufgrund der bereits durch das Pflegepersonal angebrachten Überwachungstechnik und der Operationswunde war der Platz für die Elektroden des zusätzlichen Überwachungssystems sehr beschränkt und deswegen die Elektrodenposition oft suboptimal. Daher wurde ein halbautomatisches Verfahren basierend auf neuronalen Netzen entwickelt, welches in der Lage war diese Falsch-Positiv-Ergebnisse anhand ihres Zeitverlaufes zu erkennen. Mittels optischer Kontrolle wurde diese Vorauswahl überprüft und gegebenenfalls korrigiert.

Zur Störanteilunterdrückung im PPG wurde ein FIR-Filter ( $f_{cut} = 40$  Hz, Hann-gewichtet) verwendet. Die Erkennung der systolischen und diastolischen Peaks sowie des Beginns der Pulswelle und des Punktes des maximalen Anstieges erfolgte mit Hilfe eines schwellwertbasierten Verfahrens [101].

Die Vorverarbeitung des EDA-Signal bestand aus einer Tiefpassfilterung (FIR-Filter,  $f_{cut} = 3$  Hz, Hann-gewichtet).

In Abbildung 4.2 sind diese Arbeitsschritte im Bereich „Vorverarbeitung“ zu sehen. Sie dienten der Vorbereitung der Biosignale für die anschließende Extraktion der Vitalparameter. Diese bildeten den Eingangsdatensatz für die anschließende Cluster-Analyse zur Detektion von Gesundheitszuständen.

Die gefilterten Biosignale wurden für die Ableitung von Vitalparametern in Messfenster von 300 Sekunden Länge unterteilt. Aufeinander folgende Messfenster wiesen eine Überlappung von 96 % auf, sodass ca. alle 10 Sekunden ein neuer Parameterwert zur Verfügung steht. Im Folgenden werden die Parameter beschrieben, die für die Gesundheitszustandserkennung aus jedem der Messfenster abgeleitet wurden.

#### **4.1.4 Aus den Biosignalen abgeleitete Vitalparameter**

##### **HRV-Parameter**

Im nächsten Schritt wurden aus den gefilterten Biosignalen Vitalparameter abgeleitet. In Abbildung 4.2 findet sich unter „Parameterextraktion“ eine Übersicht über alle Werte, die dabei eine Rolle spielten. Bei der Klassifizierung von Gesundheitszuständen spielt die Analyse der HRV eine wichtige Rolle. Die HRV beinhaltet alle Vorgänge, die die Herzfrequenz eines Organismus in Ruhe beeinflussen, also nicht mit körperlicher Belastung in Zusammenhang stehen. Taktgeber des Herzschlags ist der Sinusknoten. Dies ist der anatomische Bereich des Herzens, an dem die elektrische Erregung zuerst entsteht [102]. Spontane Änderungen der Aktivität des Sinusknotens werden hauptsächlich durch das vegetative Nervensystem gesteuert. Hier spielen vor allem die Aktivitäten des sympathischen und parasympathischen Nervensystems eine Rolle [103]. Diese beiden Teile wirken antagonistisch. Der Sympathi-

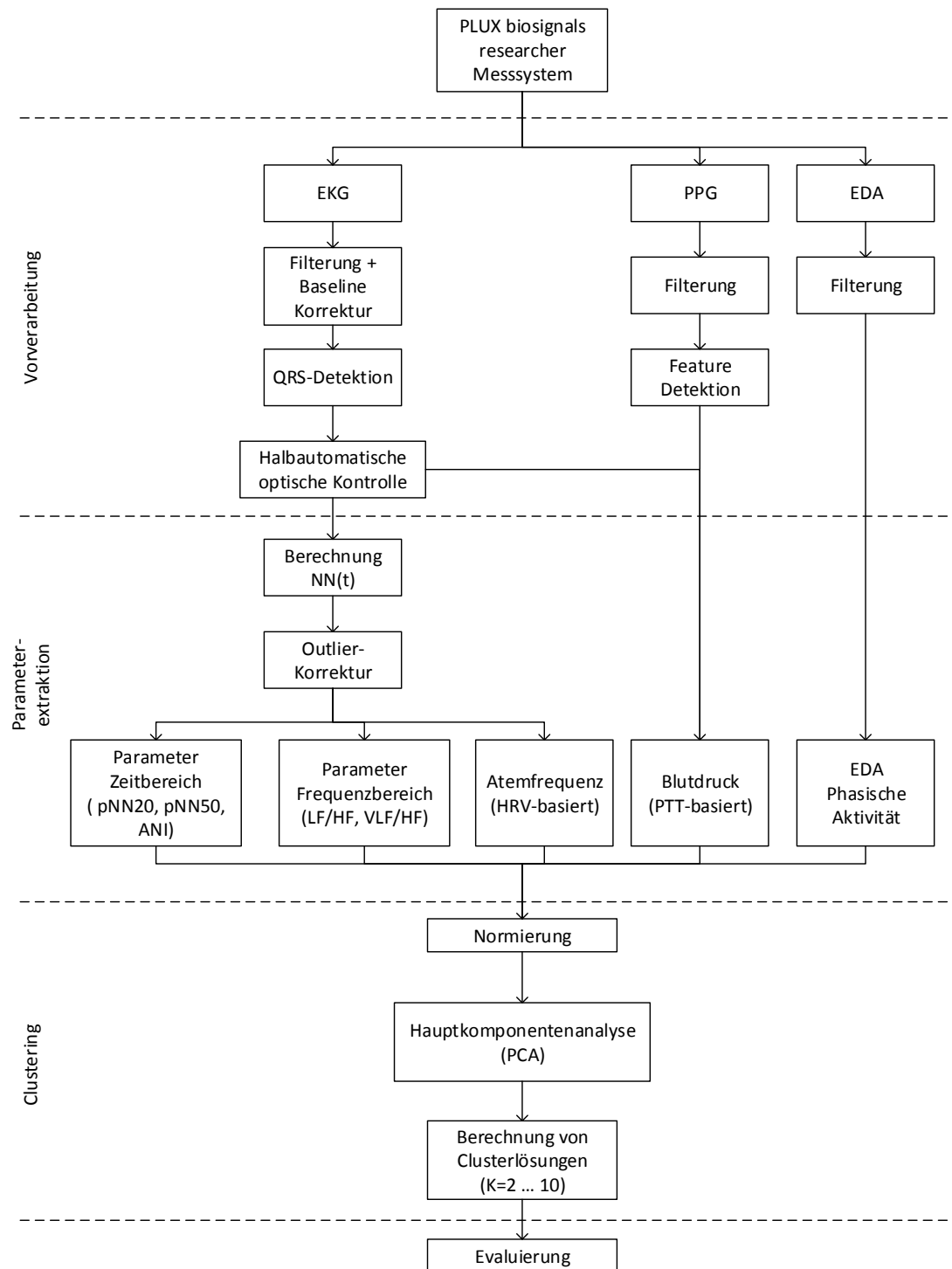


Abbildung 4.2: Überblick über alle Verarbeitungsschritte der während der Messstudie am Herzzentrum Leipzig aufgezeichneten Biosignale.

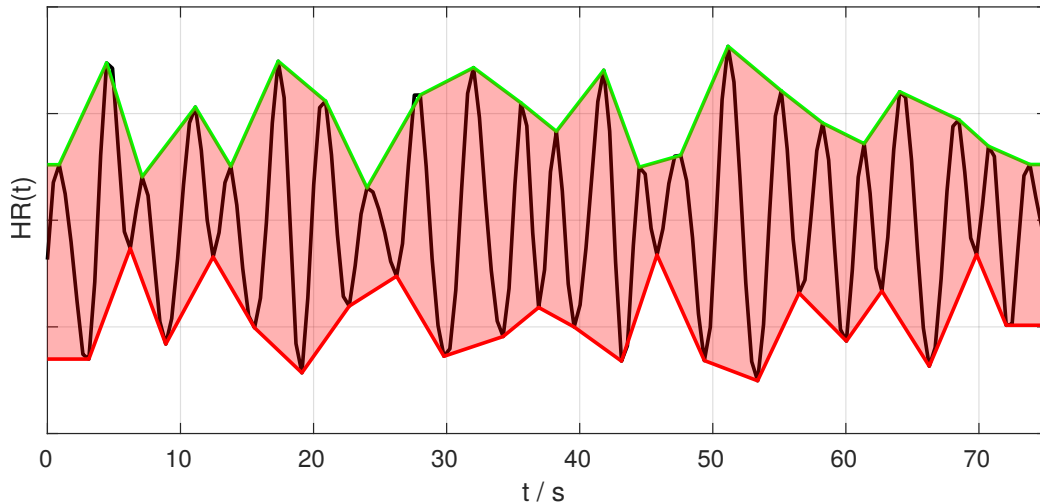
kus bewirkt durch eine verstärkte Aktivität eine Erhöhung der Herzrate. Weiterhin hat der Sympathikus steigernde Wirkung auf den Blutdruck sowie auf den Stoffwechsel und den Tonus der Skelettmuskulatur. Er bewirkt also eine Steigerung der Leistungsbereitschaft des Organismus unter Stress. Der auch als „Ruhenerve“ bezeichnete Parasympathikus bewirkt eine Verlangsamung des Herzschlages und steht für die Erholung des Körpers [104]. Es wird erwartet, dass körperlicher Stress durch postoperative Schmerzen sowie die Erholung nach Gabe von Schmerzmitteln sich jeweils auf die Balance aus sympathischer und parasympathischer Aktivität auswirkt [105]. In verschiedenen Studien konnte gezeigt werden, dass körperliche Schmerzen und negative Emotionen zu einer erhöhten sympathischen und einer verminderten parasympathischen Aktivität führen [106, 107]. Die Analyse der Herzratenvariabilität kann hierbei als Werkzeug dienen, diese Aktivität des vegetativen Nervensystems zu analysieren [107]. Daher sollen aus dem EKG mit Hilfe der HRV-Analyse Parameter abgeleitet werden, welche die Balance aus sympathischer und parasympathischer Aktivität widerspiegeln. Die „European Society of Cardiology“ gibt in ihren Leitlinien eine Reihe von Kennwerten an, die sich zur Analyse der HRV eignen [108]. Eine Übersicht über die Parameter, die in dieser Studie verwendet werden sollen, ist in Tabelle 4.1 zu sehen.

Grundlage aller HRV-Parameter ist das Tachogramm  $NN(t)$ , welches die NN-Intervalle, also den zeitlichen Abstand aufeinander folgender Herzschläge, beschreibt. Bei Bypass-Operationen kann es infolge des chirurgischen Eingriffs zu Herzarrhythmien kommen [109]. Vor allem ventrikuläre Extrasystole (VES) beeinflussen das Tachogramm und die anschließende HRV-Analyse. Das Auftreten von VES äußert sich in einem sprunghaften Anstieg der Herzrate und einem anschließenden ebenso sprunghaften Abfall. Um den Einfluss derartiger Artefakte zu minimieren, wurde direkt im Anschluss an die Berechnung des Tachogramms eine Outlier-Detektion durchgeführt. Hierfür wurden zwei Methoden angewandt. Die erste Methode beinhaltete die Klassifikation von Samples als Ausreißer, die ein definiertes Vielfaches des Interquartilabstandes überschritten. Die zweite Methode war der Ausreißertest nach Grubbs [110]. Die Wahl eines der beiden Verfahren und die Einstellung der notwendigen Parameter wurden nach visueller Inspektion des Tachogramms vorgenommen. Bei der Entfernung der Outlier entstanden Lücken im Signal. Das Auffüllen dieser Lücken geschah mit Hilfe linearer Interpolation. Aus dem so bereinigten Tachogramm wurden Parameter aus dem Zeit- und Frequenzbereich berechnet, welche im Folgenden erläutert werden.

Der Parameter  $SDNN$  beschreibt die Standardabweichung der NN-Intervalle in einem Messfenster. Die Kennwerte  $NN20$  und  $NN50$  geben die Anzahl an NN-Intervallen an, die sich um mehr als 20 ms bzw. 50 ms von ihrem jeweiligen vorhergehenden NN-Intervall

	Parameter	Einheit	Definition
Statistische Maße	SDNN	ms	Standardabweichung aller RR-Intervalle einer Messung bzw. eines Messfensters.
	NN50		Anzahl der aufeinander folgenden RR-Intervalle innerhalb eines Messfensters, die sich in ihrer Länge um mehr als 50 ms unterscheiden.
	pNN50		Anteil aller aufeinander folgender RR-Intervalle, die sich in ihrer Länge um mehr als 50 ms unterscheiden an der Gesamtzahl der RR Intervalle eines Messfensters.
	NN20		Anzahl der aufeinander folgenden RR-Intervalle innerhalb eines Messfensters, die sich in ihrer Länge um mehr als 20 ms unterscheiden.
	pNN20		Anteil aller aufeinander folgender RR-Intervalle, die sich in ihrer Länge um mehr als 20 ms unterscheiden an der Gesamtzahl der RR Intervalle eines Messfensters.
	ANI		Analgesia Nociception Index; beschreibt die Determiniertheit des Zeitsignals NN(t) im Bereich 0,15 Hz bis 0,5 Hz. Dieser Wert sinkt beim Auftreten von Schmerzen.
Frequenzmaße	$P_{VLF}$	$ms^2$	Leistung im VLF-Frequenzband des NN-Zeitsignals im Bereich $f < 0.04$ Hz; von sympathischer Aktivität dominiert.
	$P_{LF}$	$ms^2$	Leistung im LF-Frequenzband des NN-Zeitsignals im Bereich $0.04 \text{ Hz} \leq f < 0.15 \text{ Hz}$ ; sowohl sympathische als auch parasympathische Anteile.
	$P_{HF}$	$ms^2$	Leistung im HF-Frequenzband des NN-Zeitsignals im Bereich $0.15 \text{ Hz} \leq f < 0.4 \text{ Hz}$ ; von parasympathischer Aktivität dominiert.
	$P_{VLF}/P_{HF}$		Verhältnis aus VLF- und HF-Leistung.
	$P_{LF}/P_{HF}$		Verhältnis aus LF- und HF-Leistung.

**Tabelle 4.1:** Tabellarische Übersicht über die verwendeten Parameter aus dem Zeit- und Frequenzbereich. Der EKG-Datensatz wird hierfür in Zeitfenster konstanter Länge und Überlappung unterteilt. Die Parameter werden für jedes Zeitfenster separat berechnet.



**Abbildung 4.3:** Extraktion des ANI aus dem Tachogramm.

unterscheiden. Ausgehend hiervon geben die Parameter  $pNN20$  und  $pNN50$  den jeweiligen prozentualen Anteil der abweichenden NN-Intervalle an der Gesamtzahl der NN-Intervalle in einem Zeitfenster an. Diese Kennwerte stellen ein Maß für die parasympathische Aktivität dar. Hohe Werte stehen für verstärkte parasympathische Aktivität [111]. Die Parameter  $pNN20$  und  $pNN50$  wurden aus dem Tachogramm für die Extraktion von Gesundheitszuständen ermittelt.

Jeanne et al. schlugen im Jahr 2009 den Analgesia Nociception Index (ANI) vor. Dieser Parameter basiert auf einer Analyse des hochfrequenten ( $0.15 \text{ Hz} \leq f \leq 0.5 \text{ Hz}$ ) Signalanteils des Tachogramms im Zeitbereich und beschreibt den chaotischen Charakter des Zeitsignals  $NN(t)$ . Die Autoren beobachteten, dass Fluktuationen im Tachogramm während gut eingestellter Schmerzmitteldosen hauptsächlich von respiratorischen Einflüssen dominiert werden. Während des Auftretens von Schmerzen nehmen diese Fluktuationen einen unregelmäßigen und chaotischen Charakter an [112]. In Studien konnte ein negativer linearer Zusammenhang zwischen dem ANI und den von Patienten angegebenen Schmerz-Scores nachgewiesen werden [113, 114]. In Abbildung 4.3 ist die Extraktion des ANI beispielhaft an einem Tachogramm dargestellt. Auch dieser Parameter wurde für die weitere Verarbeitung extrahiert.

Weitere Vitalparameter können durch Analyse des Tachogramms im Frequenzbereich extrahiert werden. Das Spektrum von  $NN(t)$  kann in einzelne Frequenzbänder unterteilt werden. Die Leistung in diesen Frequenzbändern steht ebenfalls als Maß für die Aktivität des vegetativen Nervensystems. Niederfrequente Spektralanteile bis 0.15 Hz werden mit überwiegend parasympathischer Aktivität assoziiert, höherfrequente Spektralanteile im Be-

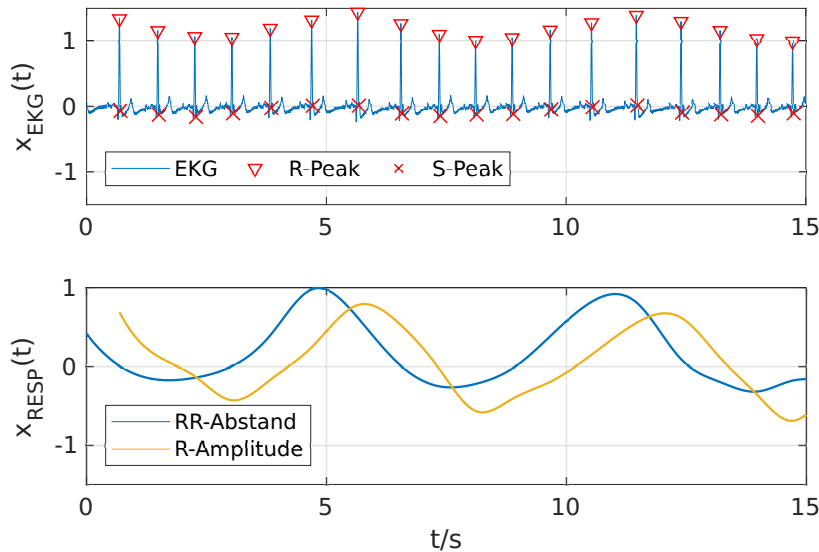
reich 0.15 Hz bis 0.4 Hz hingegen mit einer dominierenden sympathischen Aktivität [115]. Da es sich bei  $NN(t)$  um ein Zeitsignal mit variabler Abtastfrequenz handelt, wurde das Lomb-Scargle Periodogramm zur Berechnung des Leistungsspektrums verwendet [116]. Dies hat den Vorteil, dass das für die Fast Fourier Transformation (FFT) notwendige Resampling entfällt und so genauere Werte für die frequenzbasierten Vitalparameter entstehen [117]. Die Verhältnisse  $P_{VLF}/P_{HF}$  und  $P_{LF}/P_{HF}$  lassen daher Aussagen über die Balance aus sympathischer und parasympathischer Aktivität zu. Hohe Werte der Verhältnisse stehen vor allem für sympathische, kleine Werte für dominierende parasympathische Aktivität [108].

### Atemfrequenz

Es existieren verschiedene Methoden, die Atemfrequenz aus dem EKG zu extrahieren. So kann das Atemsignal aus der Modulation der R-Peak-Amplitude geschätzt werden [118]. Ein- und Ausatembewegungen heben eine Veränderung der elektrischen Eigenschaften des Brustkorbes zur Folge. Dies führt zu einer sich ändernden Amplitude der R-Zacke, welche sich im Gleichtakt mit der Atembewegung befindet. Ein weiteres Verfahren macht sich die respiratorische Sinusarrhythmie (RSA) zunutze [119]. Dieser Effekt beschreibt das Beschleunigen und Verlangsamen der Herzrate während der Ein- und Ausatembewegungen. Im Folgendem soll die RSA zur Extraktion des Atemsignals verwendet werden, da die RR-Intervalle aufgrund der Prominenz der R-Zacken im EKG mit hoher Genauigkeit detektiert werden können.

In einem ersten Schritt wurde durch Demodulation das Respirationssignal  $x_{resp}$  aus dem mit der Frequenz  $f_s$  abgetasteten EKG-Signal  $x_{EKG}$  erzeugt. Dazu wurden die mit Hilfe des Hamilton-Tompkins Algorithmus [120] extrahierten QRS-Komplexe verwendet (Abbildung 4.4 oben). Anschließend wurde der Abstand aller aufeinander folgender R-Peaks berechnet und mit kubischen C-Splines interpoliert. Das Resultat war ein Respirationssignal mit der gleichen Abtastfrequenz wie das zugrunde liegende EKG-Signal. In Abbildung 4.4 unten ist dieses Respirationssignal beispielhaft abgebildet. Zum Vergleich ist ein aus der Amplitude der R-Zacke extrahiertes Atemsignal dargestellt [121]. Beide Signale weisen, bis auf eine zeitliche Verschiebung, eine hohe Korrelation auf.

Zur Schätzung der Atemfrequenz muss die Fundamentalfrequenz dieses Atemsignals ermittelt werden. Hierfür existieren verschiedene etablierte Verfahren. Mittels einer FFT kann im Frequenzbereich die spektrale Komponente mit der größten Amplitude als Atemfrequenz interpretiert werden [118]. Die Genauigkeit dieser Methode hängt jedoch von der Frequenzauflösung der FFT ab. Besonders bei kurzen Signalabschnitten ist somit eine Hohe



**Abbildung 4.4:** Oben: Baseline-korrigiertes EKG-Signal mit markierten R- und S-Peaks zur Bestimmung der R-Peak Amplitude sowie der RR-Intervalle. Unten: Atemsignal basierend auf der Modulation der R-Peak-Amplitude (gelb) sowie auf Basis der Variation der RR-Intervalle (blau).

Exaktheit bei der Berechnung der Atemfrequenz nicht gegeben. Diesem Effekt kann mit zero-padding des Zeitsignals entgegen gewirkt werden.

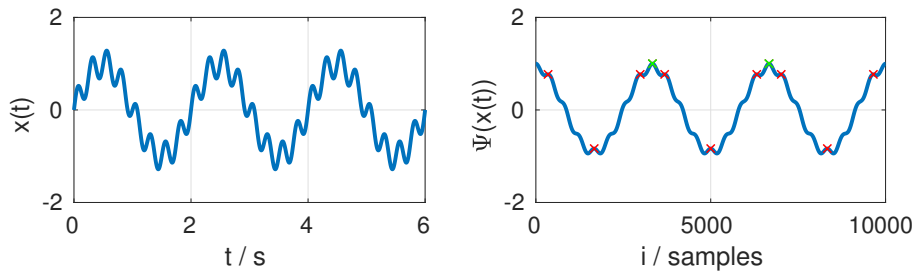
Weiterhin kann eine Nulldurchgangsanalyse des Atemsignals im Zeitbereich angefertigt werden. Aus dem Abstand dieser Nulldurchgänge kann dann die mittlere Atemfrequenz im betrachteten Zeitraum geschätzt werden [122].

Der hier vorgestellte neue Ansatz schätzt die Atemfrequenz anhand der Autokorrelationsfunktion des Atmungssignals. Dieses Vorgehen ist ähnlich der als „Pitch-Detection“ bekannten Methode zur Berechnung der Grundfrequenz eines Audio-Signals [123]. Die Autokorrelationsfunktion  $\Psi(i)$  eines Signals  $x_{resp}(n)$  beschreibt dessen Ähnlichkeit zu einer zeitlich verschobenen Version seiner selbst und wird nach folgender Formel berechnet:

$$\Psi(i) = \sum_{n=0}^N x_{resp}(n)x_{resp}(n-i) \quad (4.1)$$

Hierbei entspricht  $N$  der Anzahl der Abtastwerte des Signals und  $i$  der zeitlichen Verschiebung. Bei periodischen Signalen – wie z.B. einem Atemsignal – weist die Korrelationsfunktion periodische Maxima auf. Die Abstände, mit denen diese Maxima auftreten, entsprechen ganzzahligen Vielfachen der Grundfrequenz des Atemsignals. Diese zeitlichen Abstände  $t_i$  wurden mittels eines einfachen Peak-Detection Algorithmus aus  $\Psi(i)$  extrahiert. Durch Artefakte im EKG können jedoch weitere lokale Maxima in der Autokorrelationsfunktion entstehen. Diese verfälschen die berechnete Atemfrequenz. In Abbildung 4.5 ist





**Abbildung 4.5:** Links: Mit höherfrequenten Störungen überlagertes Atemsignal (Beispiel). Rechts: Autokorrelationsfunktion dieses Atemsignals. Durch hochfrequente Störungen entstehen zusätzliche lokale Maxima (rot markiert), die die Schätzung der Atemfrequenz verfälschen.

dieser Effekt beispielhaft dargestellt. Die Autokorrelationsfunktion eines mit hochfrequenten Störungen überlagerten Atemsignals weist eine Reihe zusätzlicher lokaler Maxima auf (rot markiert). Um diese von der Schätzung der Atemfrequenz auszuschließen, wurde ein empirisch ermittelter Schwellwert eingeführt, welcher eine Mindestamplitude für lokale Maxima gegenüber benachbarten lokalen Minima definiert. Nur bei Überschreitung dieses Schwellwerts wurde ein Maximum für die Schätzung der Atemfrequenz verwendet. Um zusätzlich den Einfluss einzelner stark abweichender Zeitabstände zu minimieren, wurde die Dauer eines Atemzuges anhand des Medians der ermittelten Zeitabstände  $t_i$  ermittelt. Die Atemfrequenz wurde dann mittels Gleichung 4.2 in Atemzügen pro Minute (bpm) berechnet.

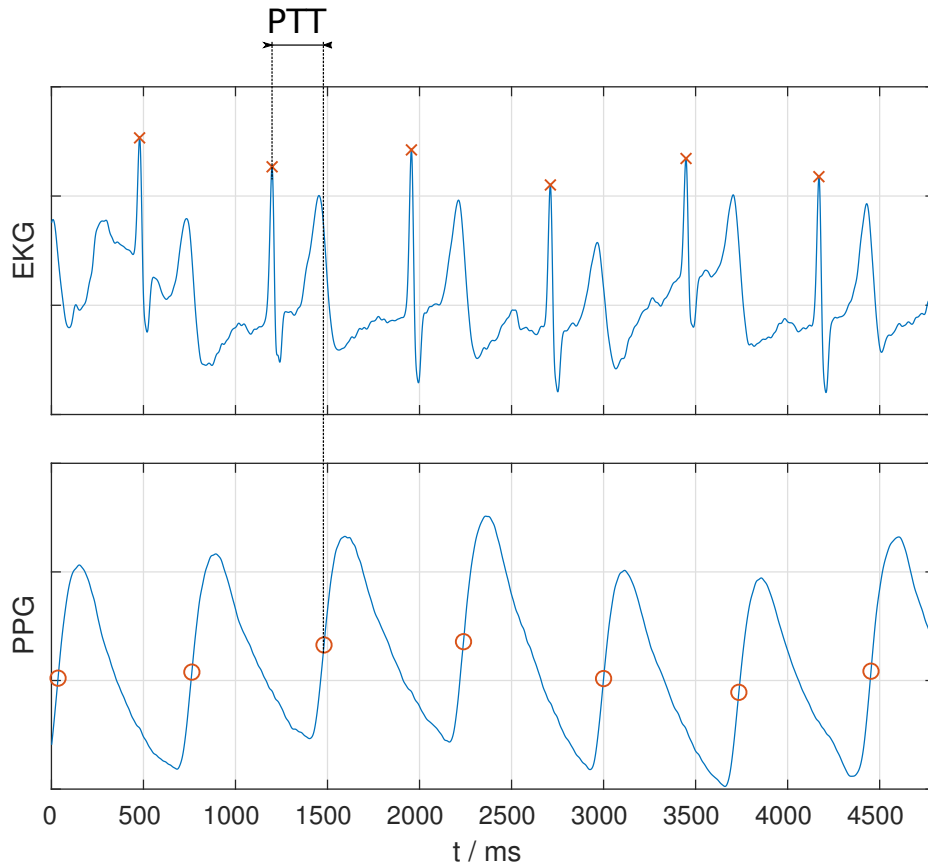
$$f_{resp} = \frac{60}{f_s} T_{resp} \quad (4.2)$$

In Anhang A wird eine einfache Studie vorgestellt, die die Evaluierung der Autokorrelationsmethode zum Ziel hat. Außerdem werden weitere Methoden zur Schätzung der Atemfrequenz aus Atemsignalen hinzugezogen, um die Ergebnisse der Autokorrelationsmethode einordnen zu können.

Die Behandlung von Schmerzen mit Opioiden kann zu einer gefährlichen Verringerung der Atemfrequenz bis hin zum Atemstillstand führen. Diesem Effekt wird mit der Gabe zusätzlicher Medikamente entgegengewirkt [124]. Es wird daher mit einer moderaten Reduktion der Atemfrequenz infolge der Schmerzmittelgabe gerechnet.

## Blutdruck

Da eine manschettensbasierte Messung für eine kontinuierliche Erfassung des Blutdrucks ungeeignet ist, sollte dieser aus der Pulse Transit Time (PTT) geschätzt werden. Die PTT ist die Zeit, die zwischen der elektrischen Erregung des Herzens (markiert durch die R-Zacke) und der Ankunft der dadurch erzeugten Pulswelle in den Extremitäten vergeht. In Abbildung



**Abbildung 4.6:** Beispielhafte Darstellung der PTT: Zeit zwischen der elektrischen Erregung des Herzens (rote Markierung in der oberen Abbildung) und der Ankunft der Pulswelle an der Körperperipherie. Der maximale Anstieg des Finger-PPG (rote Markierung untere Abbildung) wird als Ankunft der Pulswelle definiert.

4.6 ist dies exemplarisch dargestellt. Es existieren eine Reihe von Publikationen, welche EKG und/oder PPG zur Schätzung des Blutdrucks verwenden [125, 126, 127]. Im Rahmen dieser Arbeit wurde das von Gesche et al. publizierte Verfahren verwendet [125]. Aus der PTT in Millisekunden, der Körpergröße des Probanden  $h$  in Zentimetern und des Body Correlation Factor (BDC) als konstanten Faktor konnte die Pulse Wave Velocity (PWV) berechnet werden.

$$PWV = \frac{BDC \times h}{PTT} \quad (4.3)$$

Der BDC wurde mit 0,5 als konstant angenommen. Der geschätzte Blutdruckwert ergab sich dann mittels einer empirisch bestimmten Abbildungsvorschrift.

$$BP_{ptt} = 700 \times PWV \times \exp(-PWV) + 766000 \times PWV^9 - (BP_{ptt,cal} - BP_{cal}) \quad (4.4)$$

Der letzte Term in Gleichung 4.4 verschiebt die Kurve auf der  $y$ -Achse und ist probanden-spezifisch. Er dient der Kalibrierung des Verfahrens. Am Beginn jeder Messung musste ein Blutdruckwert  $BP_{ptt,cal}$  mit herkömmlichen Mitteln gemessen werden. Zusammen mit dem gleichzeitig bestimmten PTT-basierten Blutdruckwert  $BP_{ptt}$  ergab sich so eine individuelle Kalibrierung für jeden Probanden.

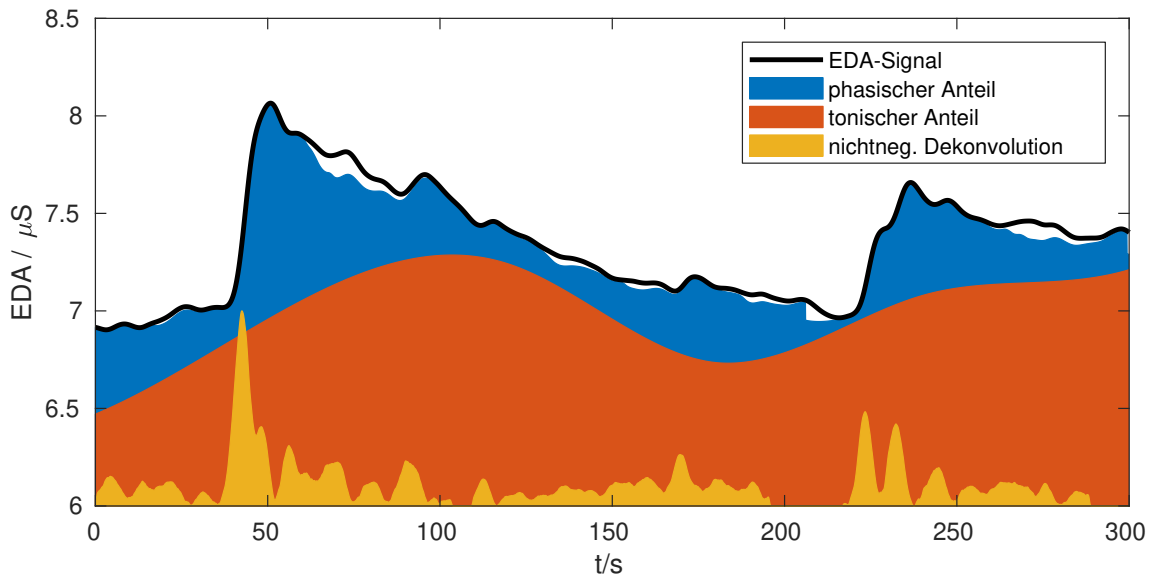
Aus der Krankenakte jedes Probanden wurden die stündlich mittels Arterienkatheter gemessenen systolischen und diastolischen Blutdruckwerte notiert. Somit konnte eine stündliche Rekalibrierung der Blutdruckschätzung durchgeführt werden. Im Falle eines Ausfalls von EKG und/oder PPG über einen längeren Zeitraum wurde der Blutdruckparameter nicht in die weitere Verarbeitung und Analyse einbezogen. In Abbildung B.2 ist ein Vergleich zwischen invasiver Blutdruckmessung und PTT-basierter Blutdruckschätzung für alle Probanden abgebildet, bei denen die beiden benötigten Messkanäle in ausreichender Qualität vorhanden waren.

Die Behandlung von postoperativen Schmerzen mittels Analgetika geht oftmals mit einer Reduzierung des Blutdrucks einher [128, 129].

### **Elektrodermale Aktivität – Phasische Aktivität**

Teile des Sympathikus – im Gegensatz zum Parasympathikus – innervieren auch die Schweißdrüsen in den oberen Hautschichten vor allem im Bereich der Extremitäten [130]. Stress durch postoperative Schmerzen und die damit einhergehende Steigerung der sympathischen Aktivität führen deshalb zu einer erhöhten Schweißproduktion. Diese senkt den Hautleitwert an den betroffenen Stellen. Deswegen eignet sich die Messung der EDA zur Messung der Stressbelastung [131]. Es existiert eine Reihe von Studien, die einen Zusammenhang zwischen Änderungen des Hautleitwertes und Stress infolge körperlicher Schmerzen nachweisen konnten [132, 133, 134]. Die EDA soll deshalb als weiteres Biosignal für die Detektion der Wirkung von Schmerzmitteln verwendet werden.

Ein typisches EDA-Zeitsignal ist aus einem sich langsam verändernden tonischen und einem schnelleren phasischen Teil zusammengesetzt. In Abbildung 4.7 sind diese zusammen mit einem beispielhaften EDA-Signal jeweils in Orange und Blau dargestellt. Für die Messung des Stresslevels ist vor allem der phasische Anteil interessant. Dieser ist durch einen schnellen Anstieg bis zu einem Spitzenwert und einer anschließend langsamen Rückkehr zur Grundlinie des Signals charakterisiert. Es konnte belegt werden, dass die Amplitude dieser Signalspitzen sowie deren Häufigkeit mit emotionalem und körperlichem Stress korreliert [135]. Die Amplitude einer EDA-Spitze wird dabei als die Differenz zwischen dem Maximum des Spitzenwertes und dem vorhergehenden Minimum definiert [136]. Im Falle



**Abbildung 4.7:** Zerlegung des EDA-Signals (schwarz) in einen niederfrequenten tonischen Anteil (orange) und einen höherfrequenten phasischen Anteil (blau). Gelb: Durch Dekonvolution geschätztes Treibersignal, welches mit der sudomotorischen Aktivierung der Schweißdrüsen assoziiert ist.

kurz aufeinander folgender EDA-Signalspitzen kann die Messung der Amplitude erschwert sein. Das ist vor allem dann der Fall, wenn eine Signalspitze in den langsam anfallenden Endteil der vorherigen Signalspitze fällt [137]. Um diese voneinander trennen zu können, wird eine von Benedek et al. entwickelte Methode verwendet, die auf einer mathematischen Entfaltung des EDA-Signals basiert. Ziel ist eine Trennung von sich überlagernden phasischen Signalspitzen (gelb in Abbildung 4.7) [138]. Durch steigende parasympathische Aktivität infolge der Entspannung nach Schmerzmittelgabe wird daher eine Reduktion der elektrodermalen Aktivität erwartet [131].

Die Extraktion des phasischen Anteils des EDA-Signals wurde mit Hilfe der Software „Ledalab“<sup>1</sup> durchgeführt [139]. Die weitere Verarbeitung im Rahmen der hier vorgestellten Studie besteht in der Mittelung aller in einem Messfenster gefundenen EDA-Amplituden.

Ein optimaler Messort sind die Handinnenflächen, da hier eine hohe Schweißdrüsendichte zu finden ist. Dieser Messort war jedoch mit einer hohen Ausfallrate der Elektroden verbunden und schränkte den Komfort der Probanden stark ein. Daher wurde das linke Handgelenk als Messort gewählt. Trotzdem kam es bei einigen Probanden zur Ablösung der Elektroden und damit zu einem Ausfall des EDA-Signals während der Messung. In einem solchen Fall wurde das EDA-Signal und die daraus abgeleiteten Parameter nicht zur weiteren Analyse verwendet.

<sup>1</sup>[www.ledalab.de](http://www.ledalab.de)

## Zusammenfassung

Mit Hilfe der HRV-Analyse sowie einer Blutdruckschätzung und der Auswertung der EDA wurden Parameter gewonnen, mit deren Hilfe der Schmerzzustand sowie die Wirkung der Schmerzmittel erfasst werden sollten. Es wurden folgende Beeinflussungen der Vitalparameter als direkte Folge der Schmerzmittelgabe erwartet:

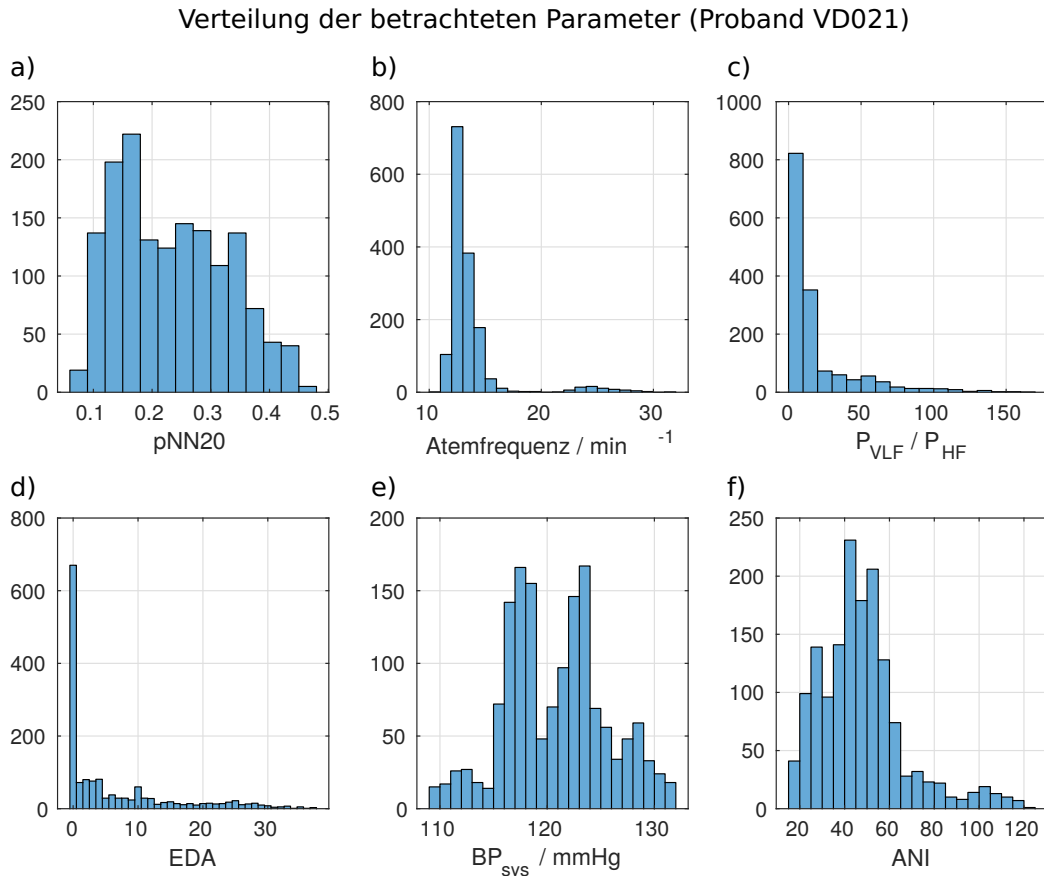
- **HRV-Parameter aus dem Zeitbereich:** Aus dem Zeitbereich wurden die Parameter pNN50 bzw. pNN20 berechnet. Infolge steigender parasymphathischer Aktivität nach Schmerzmittelgabe wurde mit einer Erhöhung der HRV und damit mit einer Erhöhung dieser Parameter gerechnet [111].
- **HRV-Parameter aus dem Frequenzbereich:** Die Frequenzauswertung der HRV bestand in der Berechnung der Parameter  $P_{LF}/P_{HF}$  bzw.  $P_{VLF}/P_{HF}$ . Es wurde erwartet, dass sich diese Werte infolge der Schmerzmittelgabe verringern [108]. Ein weiterer Parameter aus dem Frequenzbereich ist der ANI-Wert. In Folge der Schmerzbehandlung wurden hier steigende Werte erwartet [112].
- **Atemfrequenz:** Infolge von Schmerzmittelgabe wurde eine Verringerung der Atemfrequenz erwartet [124].
- **Blutdruck:** Die Verabreichung von Schmerzmitteln hat unter anderem eine Erweiterung der Blutgefäße zur Folge. Es wurde daher erwartet, dass der Blutdruck infolge der Schmerzmittelgabe sinkt [128, 129].
- **EDA-Parameter:** Aufgrund steigender parasymphathischer Aktivität wurde eine Reduktion der elektrodermalen Aktivität erwartet. [131].

### 4.1.5 Clustering der Vitalparameter

Im letzten Schritt bildeten die berechneten Vitalparameter die Datengrundlage für die Erkennung von Gesundheitszuständen mit Hilfe der Cluster-Verfahren. In Abbildung 4.2 sind die dafür notwendigen Arbeitsschritte unter „Clustering“ abgebildet.

Vor dem Clustering wurden die Zeitverläufe der abgeleiteten Vitalparameter zunächst geglättet, um kurzzeitige Schwankungen und Outlier auszugleichen. Dies geschah mittels eines Medianfilters der Länge 1500 s. Um alle Parameter auf den gleichen Wertebereich zu transformieren, wurde eine z-Transformation zur Normierung durchgeführt.

Für die Wahl des Ähnlichkeitsmaßes beim Clustering des Datensatzes muss zunächst die Verteilung der erhobenen Vitaldaten betrachtet werden. Einige Parameter wiesen einen



**Abbildung 4.8:** Werteverteilung der zur Detektion von Gesundheitszuständen verwendeten Vitalparameter am Beispiel von Proband VD021. Die dargestellten Werte decken fünf Stunden des Aufnahmezeitraums ab.

begrenzten Wertebereich auf. So nimmt der Parameter pNN20 nur Werte zwischen 0 und 1 und der Parameter  $P_{VLF}/P_{HF}$  nur Werte größer 0 an. In Abbildung 4.8 ist die Verteilung der Parameter für den Probanden VD021 beispielhaft dargestellt. Daraus ist weiterhin ersichtlich, dass z.B. die Atemfrequenz, aber auch der Blutdruck keiner symmetrischen Verteilung folgten. Daher kamen keine Abstandsmaße in Frage, die eine Normalverteilung der ihr zugrunde liegenden Daten annehmen (z.B. Mahalanobis-Distanz). Es wurde daher die euklidische Distanz als Maß für die Ähnlichkeit zwischen Samples bzw. Clustern verwendet.

Zur Dekorrelation der Daten und damit zur Reduktion der Anzahl der für das Clustering notwendigen Features wurde eine PCA durchgeführt. Für das Clustering wurden so viele Hauptkomponenten berücksichtigt, wie zur Erklärung von mindestens 90 % der Varianz des Datensatzes nötig waren. Zur Kombination der Samples und Cluster wurde die Median-Methode verwendet. Zur Berechnung des Abstandsmaßes wurde ein linearer Kernel verwendet.

Das Clustering der Vitalparameter fand nicht mit Hilfe des kompletten Datensatzes statt. Vielmehr wurden jeweils zwei Stunden vor und nach jeder protokollierten Schmerzmittelgabe zur Analyse verwendet. Insgesamt wurden für alle Probanden 73 Schmerzmittelgaben erfasst. Dies führte zu 73 Datensätzen, welche mit Hilfe der Cluster-Methoden analysiert werden sollten. Es wurden Cluster-Lösungen mit  $K = 2 \dots 10$  Clustern mit Hilfe der ähnlichkeitsbasierten Cluster-Methode und weiterhin mit Hilfe der k-means und AHC-Methode berechnet.

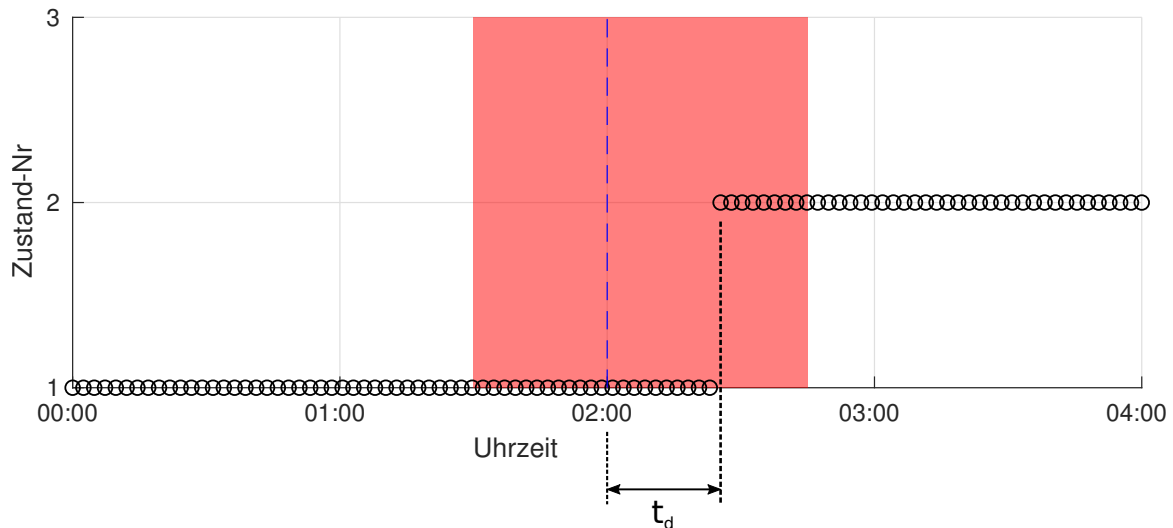
#### 4.1.6 Evaluation der Cluster-Lösungen

Zur Bewertung und zum Vergleich unterschiedlicher Cluster-Lösungen wurde im Abschnitt 3.4.2 bereits der Silhouettenkoeffizient (SI) eingeführt. Dieser Cluster-Validitätsindex gibt die Kompaktheit und Separation einer Cluster-Lösung an. Um die Exaktheit der Rekonstruktion von Schmerzmittelgaben einschätzen zu können, mussten jedoch noch weitere Faktoren berücksichtigt werden. Für die Verwendung im Rahmen dieser Studie wurde ein Score entworfen, der eine Cluster-Lösung anhand dreier zusätzlicher Merkmale bewertet. Jedes dieser Merkmale geht als Teil-Score zu gleichen Teilen in die Bewertung ein und wurde auf einen Wertebereich von 0 (schlechteste Bewertung) bis 1 (beste Bewertung) festgelegt. Folgende Eigenschaften der Cluster-Lösungen wurden mit Hilfe von Teil-Scores zusätzlich untersucht:

1. **Zeitliche Exaktheit der Rekonstruktion der Schmerzmittelgabe  $T_e$ :** Von dem Ergebnis des Clusterings wurde gefordert, dass in zeitlicher Nähe zur Schmerzmittelgabe ein Wechsel des Gesundheitszustandes stattfindet. Die Zeit  $t_d$  gibt den zeitlichen Abstand der Schmerzmittelgabe zum Zeitpunkt  $T_M$  zur Zustandsänderung zum Zeitpunkt  $T_Z$  an.

$$t_d = T_Z - T_M \quad (4.5)$$

In Abbildung 4.9 ist die Berechnung dieses Wertes an einem Beispiel illustriert. Der Zustandsübergang erfolgt hier 30 min nach der protokollierten Medikamentengabe. Um Verzögerungen der Wirkung des Schmerzmittels und Ungenauigkeiten bei der Protokollierung der Verabreichung durch das Krankenhauspersonal zu berücksichtigen, wurde ein Zustandswechsel im Bereich 30 Minuten vor und 45 Minuten nach dem protokollierten Zeitpunkt noch als mit der Schmerzmittelgabe assoziiert gewertet. Dieser Zeitbereich ist in Abbildung 4.9 rot hinterlegt. Aus dem Parameter  $t_d$  wird



**Abbildung 4.9:** Illustration des Parameters  $t_d$  als zeitlichem Abstand zwischen protokollierter Schmerzmittelgabe (2:00 Uhr) und erkanntem Zustandsübergang (2:30 Uhr).

durch Normierung auf den Toleranzbereich um die protokollierte Medikamentengabe der Teilparameter  $T_e$  berechnet.

$$T_e = \begin{cases} 1 - \frac{|t_d|}{30 \text{ min}} & t_d \leq 0 \\ 1 - \frac{|t_d|}{45 \text{ min}} & t_d > 0 \end{cases} \quad (4.6)$$

Wurde innerhalb dieser Zeitspanne kein Zustandswechsel detektiert, bekamen alle folgenden Teil-Scores den Wert 0. Der Gesamt-Score wurde somit ebenfalls auf Null gesetzt.

- 2. Veränderung der Vitalparameter P:** In Abschnitt 4.1.4 wurde dargelegt, wie sich die betrachteten Vitalparameter unter Schmerz bzw. nach Gabe von Schmerzmitteln ändern. Es wurde daher die Anzahl  $N_p$  der Parameter bezogen auf die Gesamtzahl  $N_{param}$  der betrachteten Parameter ermittelt, deren Median sich nach Schmerzmittelgabe statistisch signifikant in die erwartete Richtung verändert.

$$P = \frac{N_p}{N_{param}} \quad (4.7)$$

Die statistische Signifikanz dieser Änderung wurde mit Hilfe des Wilcoxon-Rangsummentests untersucht [140, 141].



- 3. Komplexität der Cluster-Lösung  $K_e$ :** Die Cluster-Methode soll die Verabreichung von Schmerzmedikamenten erkennen und dabei mit möglichst wenigen unterschiedlichen Zuständen bzw. Clustern auskommen. Das Optimum liegt hier bei nur  $N_K = 2$  Zuständen, von denen jeweils einer für Schmerz bzw. Erholung nach Schmerzmittelgabe steht. Eine Abweichung von dieser optimalen Cluster-Anzahl geht mit einer Verringerung dieses Teil-Scores einher.

$$K_e = 2^{-(N_k-2)} \quad (4.8)$$

- 4. Kompaktheit und Separation der Cluster-Lösung  $SI$ :** Der  $SI$ -Koeffizient beurteilt die Cluster-Lösung hinsichtlich ihrer Kompaktheit und der Separation der Cluster untereinander. Der Wertebereich wird auf  $0 \leq SI \leq 1$  begrenzt, da  $SI < 0$  als suboptimale Cluster-Einteilungen angesehen werden.

Die Teil-Scores aus den Gleichungen 4.6 bis 4.8 wurden zusammen mit dem  $SI$ -Koeffizienten zu einem Score-Vektor  $\mathbf{q}$  kombiniert. Die 1-Norm dieses Vektors ergibt dann den Gesamt-Score  $Q$  der Cluster-Lösung im Wertebereich 0 bis 1.

$$Q = \frac{1}{4} \|\mathbf{q}\|_1 = \frac{1}{4} \left\| \begin{pmatrix} K_e \\ T_e \\ P \\ SI \end{pmatrix} \right\|_1 \quad (4.9)$$

Jede Abweichung eines der Teil-Scores vom optimalen Wert führt zu einem Score  $Q < 1$ . Für den Fall, dass bei einem detektierten Zustandsübergang in zeitlicher Nähe einer Schmerzmittelgabe keiner der verwendeten Vitalparameter eine statistisch signifikante Änderung aufweist ( $P = 0$ ), wird der Güteparameter  $Q$  auf Null gesetzt. In diesem Fall kann davon ausgegangen werden, dass die Zustandsänderung nicht mit der Schmerzmittelgabe in Zusammenhang steht. Weiterhin führt eine nur schwach strukturierte Cluster-Lösung ( $SI < 0,6$ ) zu einem  $Q = 0$  [89, 90, 91].

Auf diese Weise lassen sich Cluster-Lösungen mit unterschiedlichen Cluster-Anzahlen  $K$  vergleichen und gleichzeitig weitere Aspekte hinsichtlich der Exaktheit der Rekonstruktion des Zeitpunkts der Schmerzmittelgabe berücksichtigen. Weiterhin werden Zustandsübergänge, welche in zeitlicher Nähe zu einer protokollierten Schmerzmittelgabe auftreten, hinsichtlich der Änderung der betrachteten Vitalparameter beurteilt. Der Vorteil dieser Vorgehensweise besteht in der Unabhängigkeit von der verwendeten Cluster-Methode.

Auf diese Weise ist der Vergleich mit den Ergebnissen der Referenzmethoden auf diese Weise möglich.

## **4.2 Ergebnisse**

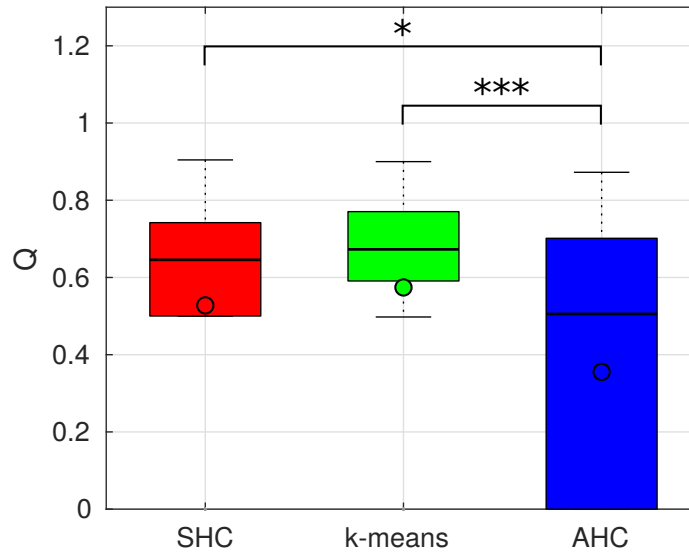
### **4.2.1 Nutzung von Vitalparametern zur Klassifikation von Gesundheitszuständen**

Bei der Ableitung der für die Extraktion der Vitalparameter notwendigen Biosignale traten verschiedene Schwierigkeiten auf. So wurde das Platzieren der EKG-Elektroden an der optimalen Stelle durch das Vorhandensein der Operationswunde im Bereich des Sternums und die bereits existierende Überwachungstechnik des Patientenmonitors erschwert. In der Folge konnte ein qualitativ hochwertiges EKG-Signal nicht in jedem Fall abgeleitet werden.

Ein weiteres Problem waren Sensorausfälle während der Messung. Während der Mobilisierung und Pflege der Patienten kam es beispielsweise in einigen Fällen zum Ablösen der Klebeelektroden und damit zum langfristigen Ausfall des entsprechenden Messkanals, wenn dies nicht zeitnah vom Pflegepersonal bemerkt und behoben wurde. Dies betraf hauptsächlich die Messung der peripheren Pulswelle, welche mit Hilfe eines Finger-Clips durchgeführt wurde. Ein Ausfall des PPG-Signals hatte auch zur Folge, dass die PTT-basierte Blutdruckschätzung nicht durchgeführt werden konnte. Zur Berechnung der Pulswellenlaufzeit zwischen Herz und Extremitäten müssen sowohl EKG als auch PPG vorhanden sein. Bei Patienten, bei denen die PTT zur Blutdruckschätzung auswertbar war, wurde zusätzlich die Übereinstimmung zwischen PTT-basierter Blutdruckschätzung und den invasiv gemessenen Blutdruckwerten an den jeweiligen Stützstellen untersucht. Bei zu großen Abweichungen wurde der Blutdruckparameter nicht zur weiteren Klassifikation von Gesundheitszuständen verwendet.

Auch das Ablösen der EDA-Elektroden war ein häufiger auftretendes Problem. Weiterhin konnten die Elektroden aus Gründen des Patientenkomforts nur am Unterarm befestigt werden. Der ideale Messort wären die Handinnenflächen gewesen. Auch dieses Signal wurde nur ausgewertet, wenn es über den gesamten Zeitraum einer Messung vollständig und ohne Unterbrechung vorlag.

In Tabelle B.2 sind für jeden Probanden die Parameter aufgelistet, die als Datengrundlage zur Klassifikation von Gesundheitszuständen verwendet wurden. Es sind nur Probanden aufgeführt, bei denen Schmerzmittelgaben protokolliert wurden und zu diesen Zeitpunkten auswertbare Biosignale vorlagen. Weiterhin wurden Medikamentengaben von der Auswertung ausgeschlossen, bei denen eine patientengesteuerte Schmerzbehandlung durchgeführt



**Abbildung 4.10:** Verteilung der Rekonstruktionsqualität  $Q$  für die Verfahren SHC, k-means und AHC. Die statistische Signifikanz der Unterschiede wurde mittels multiplen Wilcoxon-Vorzeichen-Rang-Tests und einem Signifikanzniveau von  $p^* = 0,0167$  überprüft.

wurde, da in diesem Fall eine korrekte zeitliche Zuordnung der Schmerzmittelgabe nicht möglich ist.

Insgesamt wurden für die in Tabelle B.2 aufgelisteten Probanden 73 Medikamentengaben protokolliert.

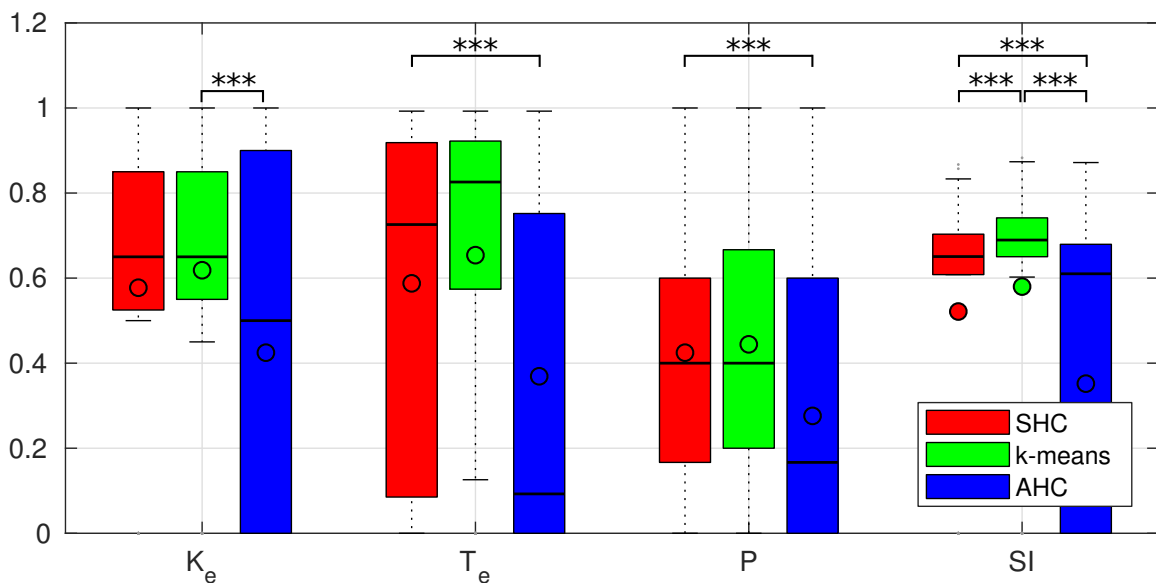
#### 4.2.2 Ergebnisse der Detektion von Schmerzmittelgaben

In Abbildung 4.10 ist die Verteilung des Güteparameters  $Q$  für alle untersuchten Verfahren abgebildet. Aus den Boxplots ist erkennbar, dass die Verfahren SHC und k-means einen ähnlichen Median aufweisen, wobei das k-means Verfahren einen etwas kleineren Interquartilabstand zeigt. Das AHC-Verfahren zeigt bei größerem Interquartilabstand einen geringeren Median als die beiden anderen Verfahren. Der große Interquartilabstand kommt durch eine hohe Anzahl an nicht detektierten Schmerzmittelgaben und daraus resultierendem  $Q = 0$  zustande. Daraus kann geschlossen werden, dass außer dem AHC-Verfahren alle Verfahren eine sehr ähnliche Leistung bei der Rekonstruktion von Schmerzmittelgaben anhand von Vitalparametern erzielen. Zur statistischen Überprüfung dieser Aussagen wurden zwischen den  $Q$ -Werten der Verfahren wurden drei Wilcoxon-Vorzeichen-Rang-Tests durchgeführt. Mittels Bonferroni-Korrektur wurde dabei der Alphafehler-Kumulierung entgegen gewirkt [142]. Dies bedeutete eine Absenkung des Signifikanzniveaus von  $p \leq 0.05$  auf:

$$p^* \leq \frac{p}{3} = 0,0167 \quad (4.10)$$

	SHC vs. k-means	SHC vs. AHC	k-means vs. AHC
$Q$	0,97	0	0
$K_e$	0,86	0,02	0
$T_e$	0,85	0	0
$P$	0,64	0	0
$SI$	0	0	0

**Tabelle 4.2:** p-Werte nach dem Vergleich der Methoden SHC, k-means und AHC hinsichtlich des Gütekriteriums  $Q$  und der darin enthaltenen Parameter. Es wurde der Wilcoxon-Vorzeichen-Rang-Test mit Bonferroni-Korrektur ( $p^* \leq 0,0167$ ) durchgeführt. Statistisch signifikante Ergebnisse sind kursiv dargestellt.



**Abbildung 4.11:** Verteilung der Teil-Scores des Güteparameters  $Q$

In Abbildung 4.10 ist zu sehen, dass nur zwischen AHC-Methode und k-means bzw. SHC Methode ein statistisch signifikanter Unterschied besteht. In Tabelle 4.2 sind die Ergebnisse der statistischen Signifikanztest quantitativ aufgeführt. Eine vollständige Auflistung der  $Q$ -Werte sowie der zu den Cluster-Lösungen gehörigen SI-Koeffizienten für alle Probanden und alle Schmerzmittelgaben findet sich in Tabelle B.3 im Anhang. Insgesamt betrachtet war bei 15 % aller Datensätze keines der Verfahren in der Lage, die Schmerzmittelgabe zu erkennen. Die Erkennung schlug weiterhin bei der SHC-Methode in 24 %, bei der k-means-Methode in 19 % und bei der AHC-Methode in 49 % aller Fälle fehl.

Die Verteilungen der einzelnen Teil-Scores bei den verwendeten Cluster-Verfahren sind in Abbildung 4.11 und Tabelle 4.2 aufgeführt. Auch hier ist ersichtlich, dass die Verfahren

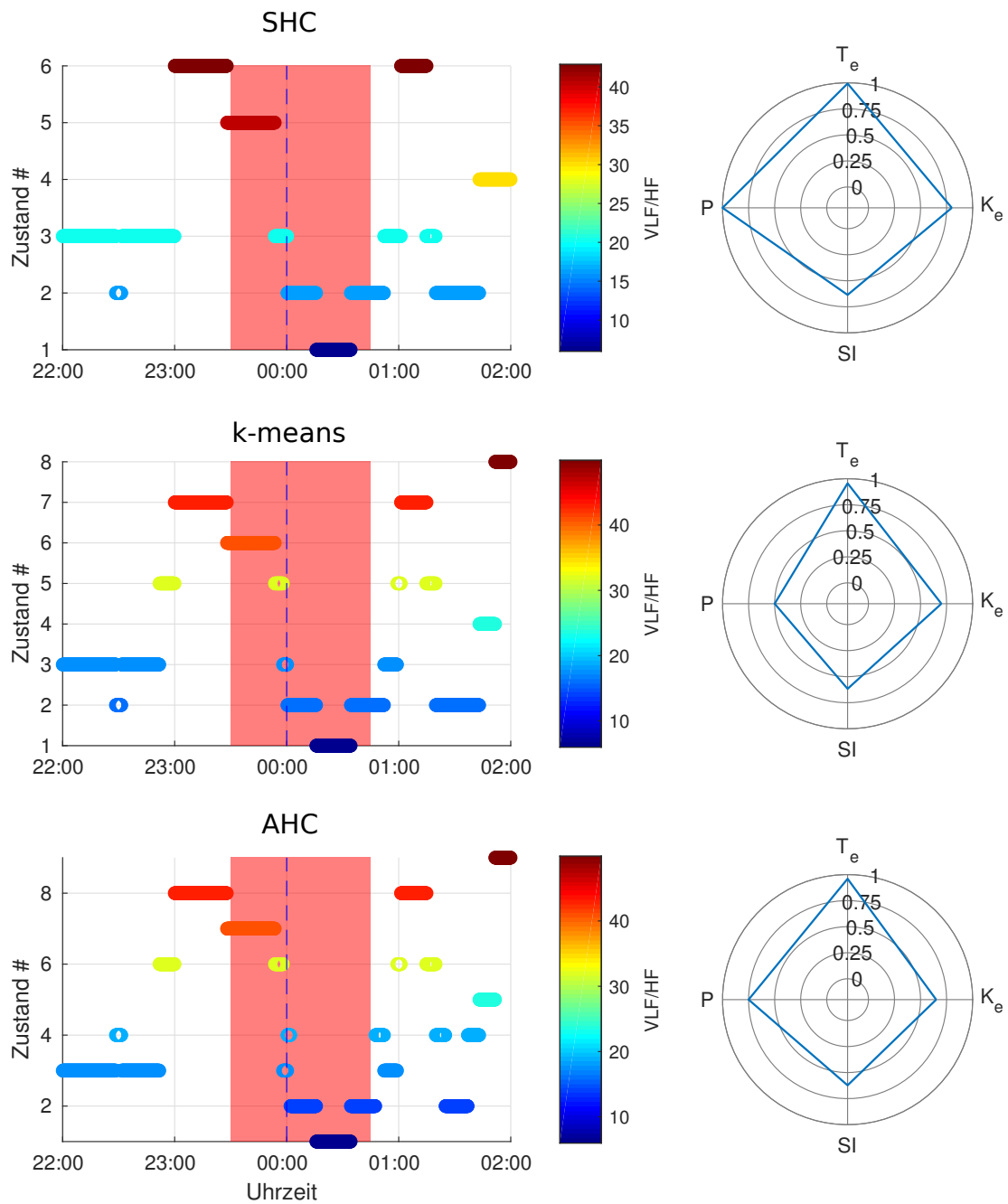
SHC und k-means sehr dicht beieinander liegen. Vor allem die Verteilungen der Parameter  $K_e$ ,  $T_e$  und  $P$  ähneln sich bei beiden Verfahren stark. Die AHC-Methode zeigt von allen Verfahren den größten Interquartilabstand, da bei dieser Methode die Anzahl der nicht erkannten Medikamentengaben ( $Q = 0$ ) von allen Verfahren am größten ist. Auch die Mediane der Teilparameter sind bei der AHC-Methode geringer als beim SHC- und k-means-Verfahren. Diese Unterschiede sind statistisch signifikant. Der größte Unterschied zwischen SHC-Methode und k-means-Methode zeigt sich im Vergleich der SI-Koeffizienten.

### 4.2.3 Detektion von Schmerzmittelgaben am Beispiel

Im Folgenden soll nun die Detektion von Schmerzmittelgaben an einem Beispiel näher betrachtet werden. In Abbildung 4.12 ist das Cluster-Ergebnis für einen Ausschnitt des Datensatzes von Proband *VD011* dargestellt. Es wurden mit Hilfe der drei Cluster-Methoden versucht, die Medikamentengabe um 0 Uhr zu rekonstruieren. Der hierfür verwendete Ausschnitt aus dem Datensatz deckte den Zeitraum von 22 Uhr bis 2 Uhr ab. Die zum Clustering verwendete Datengrundlage bestand aus den Vitalparametern  $p_{NN50}$ ,  $v_{LF/HF}$ , Atemfrequenz sowie ANI.

Aus Abbildung 4.12 kann entnommen werden, dass alle Verfahren einen Zustandsübergang um 0 Uhr detektierten. Zu diesem Zeitpunkt fand offenbar ein Übergang von einem Schmerzzustand zu einem Normalzustand statt. Der Zustand vor der Schmerzmittelgabe kann daher zunächst als Schmerzzustand angesehen werden. Der Zustand nach der Schmerzmittelgabe wird als Normalzustand angesehen. Daneben existieren weitere Zustände. Der SHC-Algorithmus erkannte vier weitere Zustände, k-means und AHC Methode jeweils sechs und fünf. In Tabelle 4.3 sind die als Schmerz- und Normalzustand identifizierten Cluster der einzelnen Verfahren zusammen mit den jeweiligen Medianen der Vitalparameter aufgelistet.

Die Färbung der Zustände in Abbildung 4.12 repräsentiert den Median des Parameters  $P_{VLF/P_{HF}}$  in den jeweiligen Zuständen. Dieser Parameter gibt das Verhältnis aus nieder- zu hochfrequenter Leistung der HRV des Probanden an. Hohe Werte stehen für verstärkte sympathische Aktivität und damit für ein hohes Maß an Stress und körperlicher Belastung. Damit soll beispielhaft der zeitliche Verlauf der Vitalparameter, welcher Grundlage für die Detektion von Gesundheitszuständen ist, illustriert werden. Es ist zu erkennen, dass der Wert des  $P_{VLF/P_{HF}}$ -Parameters ab ca. 23:00 Uhr deutlich anstieg. Der ab 23:30 Uhr zu beobachtende Rückgang dieses Wertes steht dabei mit hoher Wahrscheinlichkeit in Zusammenhang mit der für 0:00 Uhr protokollierten Schmerzmittelgabe. In Abbildung B.1 in Anhang C ist die optimale Cluster-Lösung jedes Algorithmus mit einer Einfärbung der



**Abbildung 4.12:** Links: Verlauf der Zustandsübergänge basierend auf der als optimal eingeschätzten Cluster-Lösung der Vitaldaten von Proband *VD011*. Markierung des protokollierten Zeitpunkts der Schmerzmittelgabe in blau. Die Färbung der Cluster repräsentiert den Median des Vitalparameters  $P_{VLF/P_{HF}}$  im jeweiligen Cluster. Die Verteilung aller Vitalparameter in den einzelnen Zuständen ist in Abbildung 4.13 dargestellt. Rechts: Spider-Plot der Teilparameter, welche zusammen den Güteparameter  $Q$  bilden.

	SHC		k-means		AHC	
	Schmerz	Normal	Schmerz	Normal	Schmerz	Normal
Cluster	3	2	3	2	3	4
pNN20	0,007	0,013	0,006	0,013	0,006	0,009
$P_{VLF}/P_{HF}$	20	16	17	16	17	18
resp / min <sup>-1</sup>	15,2	15	15	15	15	14.4
ANI	31	36	32	36	32	33

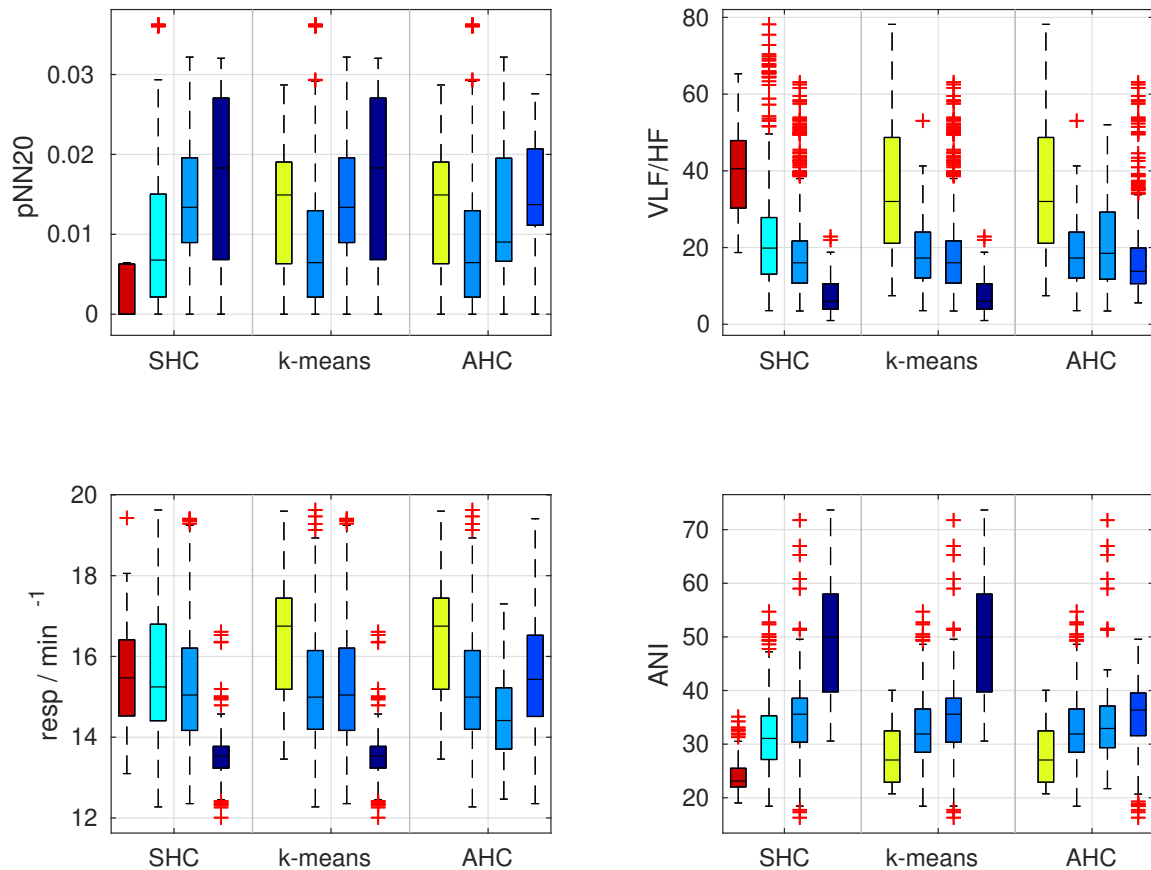
**Tabelle 4.3:** Gegenüberstellung der von den Cluster-Methoden identifizierten Schmerz- und Normalzuständen. Die Zustände werden anhand der Mediane ihrer Vitalparameter verglichen.

Zustände in Abhängigkeit der Mediane der verbleibenden drei Vitalparameter pNN50, ANI und Atemfrequenz abgebildet.

In Abbildung 4.12 rechts sind die Werte der Güte-(Teil-)Parameter in einem Spider-Plot dargestellt. Es ergaben sich aufgrund der erfolgreichen, ungefähr zeitgleichen Detektion der Schmerzmittelgabe ähnliche Werte für alle Parameter. Der Teilparameter  $K_e$  war beim SHC-Algorithmus am höchsten, da dieser mit der geringsten Anzahl Zustände bei der Erklärung der Struktur des Datensatzes auskam. Die Verfahren SHC und k-means erzielten einen höheren Wert beim SI-Koeffizienten als das AHC-Verfahren. Anders als bei den Referenzverfahren unterschieden sich beim SHC-Verfahren alle vier betrachteten Vitalparameter signifikant zwischen Schmerz- und Normalzustand. K-means und AHC-Verfahren erkannten jeweils nur einen signifikanten Unterschied in zwei bzw. drei der Vitalparameter.

Die Werteverteilung der Vitalparameter ist in Abbildung 4.13 dargestellt. Hier wird nochmals deutlich, dass sich bei der SHC-Methode alle Parameter in eine Richtung entwickelten, die für eine ansteigende parasymphatische Aktivität in Folge der Schmerzmittelgabe steht. Beispielsweise erhöhte sich der Parameter ANI beim Übergang vom Schmerz- in den Normalzustand von 31 auf 35. Auch der Parameter pNN50 verzeichnete einen leichten Anstieg beim Übergang vom Schmerz- in den Normalzustand. Der Abfall der Atemfrequenz war weniger stark ausgeprägt, aber vorhanden. Beim k-means Verfahren konnten bezüglich der Parameter  $P_{VLF}/P_{HF}$  und Atemfrequenz keine statistisch signifikanten Unterschiede in den Zuständen nachgewiesen werden. Beim AHC-Verfahren zeigte sich hinsichtlich der Atemfrequenz kein statistisch signifikanter Unterschied.

Bei der genaueren Analyse des zeitlichen Verlaufs der Gesundheitszustände in Abbildung 4.12 links fällt auf, dass die Schmerzzustände, welche unmittelbar vor der Medikamentengabe erkannt wurden, von kurzer Dauer waren. So war der Zustand 3, welcher von k-means



**Abbildung 4.13:** Verteilung der in Abbildung 4.12 zur Gesundheitszustandserkennung verwendeten Vitalparameter für die Cluster-Verfahren SHC, k-means und AHC. Die Färbung der Boxen korrespondiert mit den Farben der jeweiligen Gesundheitszustände in Abbildung 4.12.

und AHC-Methode als Schmerzzustand erkannt wurde nur 140 Sekunden lang. Auch bei der SHC-Methode dauerte der Schmerzzustand lediglich 7 Minuten. Ein umfassenderer Überblick über die Entwicklung der Vitalparameter kann durch die Betrachtung der Gesundheitszustände unmittelbar vor dem Schmerzzustand bzw. nach dem Normalzustand gewonnen werden. Diese sind in Abbildung 4.13 in ihren jeweiligen Farben dargestellt. Es ist zu sehen, dass die Vitalparameter bereits vor dem Übergang in den Schmerzzustand einen Verlauf einnahmen, der für eine ansteigende parasympathische Aktivität charakteristisch ist. So zeigte der ANI-Parameter in Zustand 5 des SHC-Algorithmus einen geringeren Median als im Zustand 3, welcher sich als Schmerzzustand unmittelbar vor dem Zeitpunkt der Medikamentengabe einstellte. Auch der  $P_{VLF}/P_{HF}$ -Parameter sank bereits ab 23:30 Uhr von Zustand 5 über die Zustände 3 und 2 bis Zustand 1 von 40 auf 5 ab. Die Atemfrequenz zeigte erst ab Zustand 1 bei allen Verfahren einen substantiellen Rückgang. Als einziges Verfahren zeigte der AHC-Algorithmus beim Übergang von Zustand 3 in Zustand 4 sogar



einen leichten Anstieg der Atemfrequenz. Aus der Verteilung der Vitalparameter ist insgesamt ersichtlich, dass ab 23:00 Uhr eine Veränderung der Vitalparameter stattfand, die für erhöhte sympathische Aktivität charakteristisch war. Ursache hierfür könnten stärker werdende Schmerzen sein. Bereits ab 23:30 Uhr bis ca. 0:15 Uhr setzte dann eine Entwicklung ein, die durch zunehmende parasympathische Aktivität gekennzeichnet war. Daraus kann geschlossen werden, dass die Gabe des Schmerzmedikaments wahrscheinlich bereits vor dem protokollierten Zeitpunkt stattgefunden hat. Im weiteren Zeitverlauf war um 1:00 Uhr ein erneuter und kurzfristiger Anstieg der sympathischen Aktivität zu verzeichnen. Diese ging kurze Zeit später zurück und es wurde der Normalzustand eingenommen. Zum Zeitpunkt des Rückgangs waren jedoch keine weiteren Schmerzmittelgaben in der Patientenakte protokolliert.

### **4.3 Diskussion**

#### **4.3.1 Detektion von Schmerzmittelgaben**

Die vorliegende Studie befasste sich mit der Rekonstruktion von Schmerzmittelgaben nach herzchirurgischen Eingriffen. Dazu wurde eine Messstudie durchgeführt, die die Messung einer Reihe von Biosignalen für eine Dauer von 24 Stunden unmittelbar nach dem operativen Eingriff beinhaltete. Aus diesen Biosignalen wurden Vitalparameter extrahiert. Mit Hilfe der Vitalparameter sollte der Zeitpunkt der Schmerzmittelgabe erkannt werden. Die im Rahmen dieser Arbeit vorgestellte ähnlichkeitsbasierte hierarchische Cluster-Methode (SHC) wurde für Erkennung von Gesundheitszuständen anhand des Vitalparameterdatensatzes verwendet. Die etablierten Verfahren AHC und k-means dienten als Referenzmethoden. Es wurde erwartet, dass in zeitlicher Nähe zur Medikamentengabe ein Gesundheitszustandsübergang von einem Schmerzzustand zu einem Normalzustand stattfindet. Die Überprüfung dieser Hypothese bestand aus einer Bewertung der Cluster-Ergebnisse sowohl hinsichtlich zeitlicher Exaktheit der Rekonstruktion als auch hinsichtlich der Plausibilität der Vitalparameteränderung zum Zeitpunkt der erkannten Schmerzmittelgabe. Dazu wurde zunächst ein vergleichender Überblick über die Rekonstruktionsergebnisse gegeben. Anschließend wurden die Ergebnisse eines Datensatzes als Beispiel genauer betrachtet.

Die Ergebnisse haben gezeigt, dass die SHC Methode grundsätzlich in der Lage ist, anhand der Änderung der Vitalparameter Zeitpunkte zu erkennen, zu denen dem Patienten mit einiger Wahrscheinlichkeit Schmerzmittel verabreicht wurden. Es hat sich weiterhin gezeigt, dass die SHC Methode hinsichtlich des Qualitätsparameters Q signifikant besser als die AHC-Methode arbeitet. Ein Vergleich mit der k-means Methode ergab keinen statistisch

signifikanten Unterschied. Eine detaillierte Untersuchung der einzelnen Teilparameter, welche das Gütemaß  $Q$  ergeben, bestätigte dieses Ergebnis. Die SHC-Methode erzielte auch bei den Teil-Scores signifikant höhere Werte als die AHC-Methode. Signifikante Unterschiede zwischen SHC- und k-means Methode waren nur hinsichtlich des SI-Koeffizienten zu verzeichnen.

Bei der Betrachtung der Komplexität der ermittelten Cluster-Lösungen fällt auf, dass – anders als die Anschauung zunächst suggeriert – in den meisten Fällen nicht nur zwischen zwei Clustern unterschieden werden kann, von denen jeweils einer einen Schmerzzustand und der andere einen Normalzustand repräsentiert. SHC- und k-means Methode ermittelten beide bei allen Datensätzen im Durchschnitt  $K = 8$  als optimale Cluster-Anzahl. Im Krankenhausumfeld lassen sich die äußeren Bedingungen, unter denen eine Messung stattfindet, nur in begrenztem Maße kontrollieren. Während der ersten Stunden nach der Operation durchläuft der Patient eine Vielzahl von Behandlungen. Beispielsweise werden Medikamente verabreicht, die nicht der Schmerzbehandlung dienen, aber trotzdem einen Einfluss auf die Vitalparameter haben können. Weiterhin werden Patienten von der Intermediate-Care Station auf die Normalstation verlegt, sobald sich ihr Zustand weitgehend stabilisiert hat. Dort wird zeitnah mit der Mobilisierung begonnen, um den Rehabilitationsprozess zu unterstützen. All diese Vorgänge wirken sich auf den zeitlichen Verlauf der Vitalparameter des Patienten aus. Daher kann nicht davon ausgegangen werden, dass zwei Zustände zur Erklärung der datensatzimmanenten Struktur ausreichen.

Bei der Bewertung der zeitlichen Exaktheit wurde um den protokollierten Zeitpunkt der Medikamentengabe ein Toleranzfenster gelegt, innerhalb dessen eine detektierte gültige Zustandsänderung noch als mit der Medikamentengabe assoziiert betrachtet werden kann. Die Tatsache, dass die Patientenakte protokollierte Medikamentengaben immer zur vollen Stunde enthält, legte die Vermutung nahe, dass hier nicht der exakte Zeitpunkt der Verabreichung des Medikaments notiert wurde. Vor diesem Hintergrund scheint die Wahl eines Toleranzzeitfensters 30 Minuten vor und 45 Minuten nach der Medikamentengabe sinnvoll. Weiterhin bedeutet dies, dass Abweichungen des  $T_e$ -Parameters von Eins entweder auf eine ungenaue Erkennung der Schmerzmittelbehandlung oder eine zeitlich genaue Erkennung einer ungenau protokollierten Behandlung zurückzuführen ist. Der  $T_e$ -Parameter liegt bei den Verfahren SHC und k-means bei Datensätzen, in denen eine Schmerzmittelgabe erfolgreich erkannt werden konnte, im Bereich von 0,6 bis 0,9. Dies entspricht einer zeitlichen Abweichung von 5 bis 18 Minuten von der protokollierten Medikamentengabe. Da die Werteverteilungen beider Verfahren keinen statistisch signifikanten Unterschied aufweisen,

kann davon ausgegangen werden, dass eine tatsächliche Medikamentengabe in diesem Zeitbereich zuverlässig erkannt werden konnte.

Der Parameter  $P_e$  beschreibt den Anteil der Vitalparameter, die sich bei einem mit Analgesie assoziierten Gesundheitszustandsübergang in die erwartete Richtung verändern. Hierbei erwiesen sich vor allem die HRV-Parameter aus dem Frequenzbereich als zuverlässig. Die EDA-basierten Parameter waren – sofern das EDA Signal auswertbar war – ebenfalls in den meisten Fällen aussagekräftig. Dabei spielt der Messort der elektrodermalen Aktivität am linken Unterarm des Probanden eine Rolle. Hier ist die Schweißdrüsendichte geringer als an der Handinnenfläche. Die Veränderung der EDA aufgrund schmerzinduzierter emotionaler und körperlicher Belastung war daher nicht bei jedem Patienten ausreichend stark vorhanden.

Anhand des Beispieldatensatzes von Proband *VD011* konnte die Klassifizierung von Gesundheitszuständen in Abhängigkeit der verwendeten Vitalparameter genauer untersucht werden. Es wurden neben der Atemfrequenz vor allem HRV-Parameter aus dem Zeit- und Frequenzbereich verwendet. Hier ist noch einmal deutlich geworden, dass die Änderung des Gesundheitszustandes nicht plötzlich, sondern vielmehr graduell verläuft. Es konnte ein Zustandsübergang in zeitlicher Nähe zur protokollierten Medikamentengabe detektiert werden. Auch die Werte der Vitalparameter änderten sich wie erwartet. Eine Betrachtung eines erweiterten Zeitraumes um den erkannten Zeitpunkt der Medikamentengabe offenbarte jedoch, dass diese Entwicklung schon viel früher einsetzte und sich auch noch einige Zeit nach der Medikamentengabe fortsetzte. Dies resultierte in der Klassifikation weiterer Gesundheitszustände mit entsprechend charakteristischen Wertebereichen der betrachteten Parameter.

### 4.3.2 Datenerhebung und Vitalparameterextraktion

Mit Hilfe des PLUX Messsystems konnten Biosignale mit hoher Auflösung ( $f_s = 1 \text{ kHz}$ ) aufgenommen werden. Das ermöglicht eine hohe Genauigkeit sowohl bei der Erfassung von QRS-Komplexen aus dem EKG-Signal als auch bei der Bestimmung von Parametern im Rahmen der HRV-Analyse. Aufgrund der teilweise suboptimalen Elektrodenpositionen und auch aufgrund der Tatsache, dass nur ein Einkanal-EKG aufgenommen wurde, musste sich die Analyse des Signals auf die HRV-Parameter, welche aus dem Tachogramm gewonnen wurden, beschränken. Da die Elektrodenkonfiguration nicht mit den gängigen Standardableitungen für EKG-Signale (z.B. Ableitungen nach Eindhoven oder Goldberger) übereinstimmte, konnten keine morphologischen Eigenschaften ausgewertet werden. Es konnte jedoch in Studien gezeigt werden, dass Veränderungen in der Form, z.B. der P- und T-Wellen, des EKG-Signals als Marker für postoperative Komplikationen dienen können

[143, 144]. Hierfür werden jedoch EKG-Ableitungen mit bis zu 12 Kanälen benötigt. Es wäre daher vorteilhaft, in Zukunft die vorhandene Messtechnik zu nutzen, welche ohnehin zur Überwachung des Patienten eingesetzt wird. Dies war jedoch aufgrund der fehlenden Möglichkeit der Aufzeichnung des Roh-EKG-Signals nicht möglich.

Durch die Verwendung zusätzlicher physiologischer Parameter könnte die Detektion von Schmerzmittelgaben weiter verbessert werden. Wie bereits erwähnt, konnten aus der auf der IC-Station vorhandenen Messtechnik keine Rohdaten zur weiteren Verarbeitung entnommen werden. Durch die Verwendung des PLUX-Messsystems war nur die Messung einer begrenzten Anzahl von Parametern möglich. Durch die Ableitung von HRV-Parametern konnten jedoch aussagekräftige Indikatoren gewonnen werden. Diese waren jedoch oftmals stark von der Zuverlässigkeit der EKG-Verarbeitung, insbesondere der QRS-Detektion abhängig. Alternative Parameter könnten arterielle Blutdruckwerte sein, welche bei der intensivmedizinischen Überwachung häufig kontinuierlich gemessen werden [145]. Weiterhin könnten aus dem Elektroenzephalogramm (EEG) abgeleitete evozierte Potentiale verwendet werden, um mit Hilfe der SHC-Methode die Anästhesietiefe und das Schmerzempfinden eines Patienten abzuschätzen. Studien haben gezeigt, dass durch Schmerz induzierte EEG-Signalamplituden mit der Schmerzstärke korrelieren und durch die Gabe von Schmerzmitteln reduziert werden können [146, 147].

Um den Einfluss weiterer, nicht mit schmerzinduziertem Stress assoziierter Faktoren zu minimieren, ist die Wahl eines weniger komplexen Setups ratsam. Für zukünftige Untersuchungen sollten kürzere Datensätze in Umgebungen mit weniger Störeinflüssen als auf einer IC-Station gewählt werden. Mögliche Szenarien sind Operationen, bei denen unter lokaler Betäubung gearbeitet wird. Hier würden sich beispielsweise Katheterablationen anbieten, die unter Lokalanästhesie durchgeführt werden und bei denen Patienten bei Bedarf Schmerzmittel verabreicht werden.

### **4.3.3 Zusammenfassung**

Bei 11 der 73 analysierten Datensätze konnte keines der Verfahren die Medikamentengabe erfolgreich detektieren. Dies waren Datensätze, in denen entweder keine Beeinflussung der Vitalparameter zu erkennen war oder eine Reaktion außerhalb des Toleranzzeitraumes stattfand. Trotzdem legen die Ergebnisse nahe, dass die Detektion von Schmerzmittelgaben anhand von Vitalparametern, welche vor allem auf der HRV-Analyse des EKG basieren, möglich ist. Es ist zu betonen, dass die Erkennung von Schmerzmittelgaben nicht als konkreter Anwendungsfall verstanden werden soll. Ziel war es vielmehr zu zeigen, dass die Methode für die Erkennung von Gesundheitszuständen und deren Änderungen aufgrund

äußerer Einflüsse geeignet ist. Im Hinblick auf das Schmerz-Management liegt der praktische Nutzen dieser Methode vor allem in der Überwachung von Patienten, welche sich selbst nicht mitteilen können. Hier sind beispielsweise Patienten mit eingeschränkten geistigen Fähigkeiten, Säuglinge, aber auch Patienten, die nach einer Operation noch nicht wieder bei vollem Bewusstsein sind, zu nennen.

Mit dieser Studie konnte demonstriert werden, dass das SHC-Verfahren im klinischen Umfeld anwendbar ist. Die Eingangsdaten der Methode sind nicht auf EKG-Parameter beschränkt. Daher sind auch andere Anwendungsfälle denkbar, wie z.B die telemedizinische Überwachung von Patienten zu Hause



## **Kapitel 5**

### **Erkennung von obstruktiver Schlafapnoe basierend auf EKG-Parametern**

#### **5.1 Einleitung**

Die Anwendung von maschinellen Lernverfahren im Allgemeinen und der Similarity-based Hierarchical Cluster Analysis (SHC)-Methode im Besonderen ist nicht auf die Erkennung von Schmerzmittelgaben beschränkt. Im Zuge der Recherchen und Untersuchungen für diese Arbeit entstand ein weiteres Anwendungsgebiet für die neu konzipierte Methode. Über das Thema der Arbeit hinaus soll die Methode zur Erkennung einer in der Bevölkerung weit verbreiteten schlafbezogenen Atemstörung verwendet werden.

Die obstruktive Schlafapnoe (OSA) ist ein medizinischer Zustand, der durch wiederkehrende Atembeschwerden oder Atemstillstand während des Schlafes aufgrund partieller oder kompletter Blockade der Atemwege gekennzeichnet ist. Hiervon sind in Deutschland 3-7 % der Männer und 2-5 % der Frauen betroffen [148]. Risikofaktoren sind unter anderem, wie auch bei den koronaren Herzerkrankungen, fortgeschrittenes Alter und Übergewicht bzw. Adipositas [149]. Die folgenreichsten Symptome während OSA-Phasen sind Sauerstoffmangel, Schlafstörungen sowie Hypertonie und eine erhöhte Herzrate aufgrund erhöhter sympathischer Aktivität [150]. Aufgrund der beeinträchtigten Schlafqualität kommt es als direkte Folge zu einer Verringerung der geistigen Leistungsfähigkeit. Außerdem steigt die Gefahr von mitunter tödlich verlaufenden Unfällen an [151, 152]. Langfristig ist bei betroffenen Personen mit einer Erhöhung des Herzanfallrisikos (140 %), Schlaganfallrisikos (60 %) und der Erhöhung des Risikos einer koronaren Herzerkrankung (30 %) zu rechnen [153]. Die Diagnose von OSA findet meist mittels Polysomnographie in einem Schlaflabor statt [154]. Die am häufigsten untersuchten Parameter sind:

- Elektrokardiogramm (HRV-Parameter)
- Tonaufzeichnung zur Detektion von Schnarchen
- Elektroenzephalogramm und Elektrookulogramm zur Bestimmung von Schlafphasen
- Bewegungen während des Schlafs

- Brustkorb- und Bauchbewegungen zur Messung der Atemanstrengungen
- Messung der Atemfrequenz mittels Gesichtsmaske

Im Folgenden soll die Detektion von OSA mit Hilfe maschineller Lernverfahren durchgeführt werden. Datengrundlage bilden Elektrokardiogramm (EKG)-basierte Vitalparameter, welche der PhysioNet Datenbank entnommen wurden [155, 156]. Das Auftreten von OSA sollen mit Hilfe der SHC-Methode zeitlich so genau wie möglich rekonstruiert werden. Die mit dieser Methode gewonnenen Ergebnisse werden mit denen der Referenzverfahren aus Kapitel 2 verglichen.

## 5.2 Methode

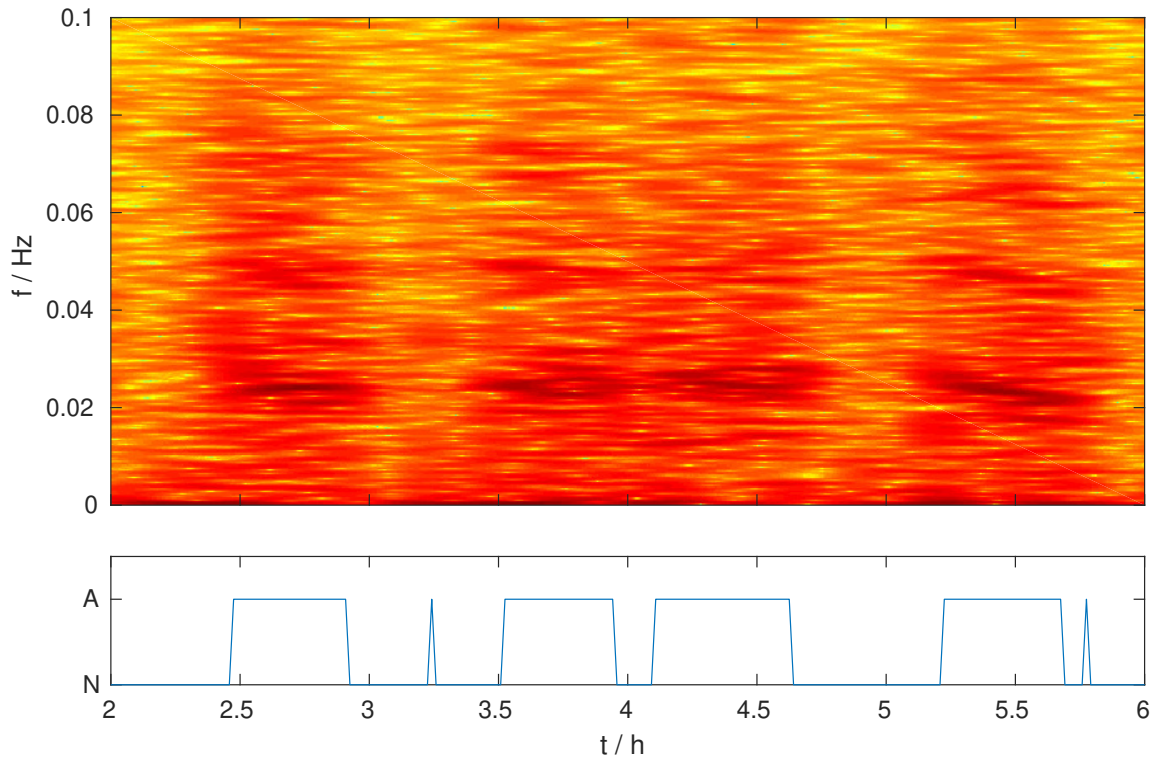
### 5.2.1 Verwendete Datenquelle und extrahierte Vitalparameter

In Ermangelung eines eigenen Schlaflabors wird auf die PhysioNet-Datenbank als Datenquelle zurück gegriffen [155]. Die Datenbank besteht aus 70 Datensätzen der Länge sieben bis zehn Stunden. Jeder Datensatz besteht aus einem EKG-Signal. Weiterhin enthalten einige Datensätze zusätzlich noch ein Respirationssignal sowie eine kontinuierliche Messung der Sauerstoffsättigung des Blutes. Bei 35 Probanden wurden Phasen mit OSA durch einen Experten minutengenau markiert. Diese Markierungen sollen als Goldstandard für die Detektion von Phasen mit OSA dienen.

Wie bereits in Kapitel 4 dargelegt wurde, können mittels einer HRV-Analyse im Frequenzbereich Aussagen über die Balance aus sympathischer und parasympathischer Nervenaktivität getroffen werden [107]. Weiterhin ist bekannt, dass auftretende OSA mit erhöhter sympathischer Aktivität einhergeht [150]. Zywiets et al. haben daher einen Ansatz gewählt, der völlig auf der Auswertung der VLF-Komponente (0.013 Hz - 0.0375 Hz) der Herzrate beruht. In diesem Frequenzbereich haben die Autoren den größten Unterschied zwischen normalen und pathologischen Schlafzuständen feststellen können [157]. Die VLF-Komponente der Herzrate lässt sich mit Hilfe eines Spektrogramms berechnen. Ein solches ist in Abbildung 5.1 abgebildet. Ein Vergleich der Frequenzanteile um 0.025 Hz während normaler und OSA-Phasen zeigt einen deutlichen Anstieg der Amplitude in diesem Frequenzbereich mit Beginn der OSA-Phase.

Die Verarbeitung des Roh-EKG-Signals begann mit der Detektion der QRS-Komplexe mittels des Pan-Tompkins Algorithmus [93]. Daran schloss sich die Berechnung des Tachogramms  $NN(t)$  aus dem zeitlichen Abstand von aufeinander folgenden R-Zacken an. Outlier, welche sich durch Extrasystolen oder nicht detektierte QRS-Komplexe ergaben,





**Abbildung 5.1:** Oben: Spektrogramm des NN-Zeitsignals ( $f_s = 2.4$  Hz) basierend auf einer Fensterlänge von 30 Minuten und einer Schrittweite von 30 Sekunden. Unten: Minutengenaue Klassifizierung des Signals in Normal (N) und OSA (A).

wurden aus dem Tachogramm entfernt. Dies betraf alle Werte, die sich außerhalb eines individuell festgelegten Vielfachen des Interquartilabstandes befanden. Um eine Frequenzanalyse durchführen zu können, musste das Tachogramm durch Spline-Interpolation in ein äquidistant abgetastetes Signal überführt werden. Die für die Interpolation gewählte neue Abtastfrequenz betrug  $f_s = 2.4$  Hz. Zur Berechnung der für die Cluster-Analyse benötigten Vitalparameter wurde das interpolierte Signal in Messfenster von 30 Minuten Länge und einer Überlappung von 29,5 Minuten unterteilt. In jedem Messfenster wurde die Herzratenvariabilität (HRV) berechnet. Außerdem wurde das Signal durch eine FFT in den Frequenzbereich überführt und es wurde die Leistung  $P_{VLF}$  im VLF-Band zwischen 0.013 Hz und 0.0375 Hz berechnet. Diese beiden Parameter bildeten die Eingangswerte für die anschließende Cluster-Analyse.

### 5.2.2 Cluster-Analyse

Zur Detektion von Clustern wurde die SHC-Methode mit folgenden Einstellungen verwendet:

- Fusionierungsmethode: Median
- Kernel-Funktion: linearer Kernel
- Berechnung von Cluster-Lösungen mit  $K = 2 \dots 10$

Als Referenzverfahren dienten die bereits in Abschnitt 2.4.2 und 2.4.1 beschriebenen unüberwachten Cluster-Methoden k-means und agglomeratives hierarchisches Clustering (agglomeratives hierarchisches Clustering (AHC)). Für alle Probanden wurden mit jeder Cluster-Methode drei Cluster-Lösungen mit jeweils unterschiedlichem Eingangsdatensatz berechnet:

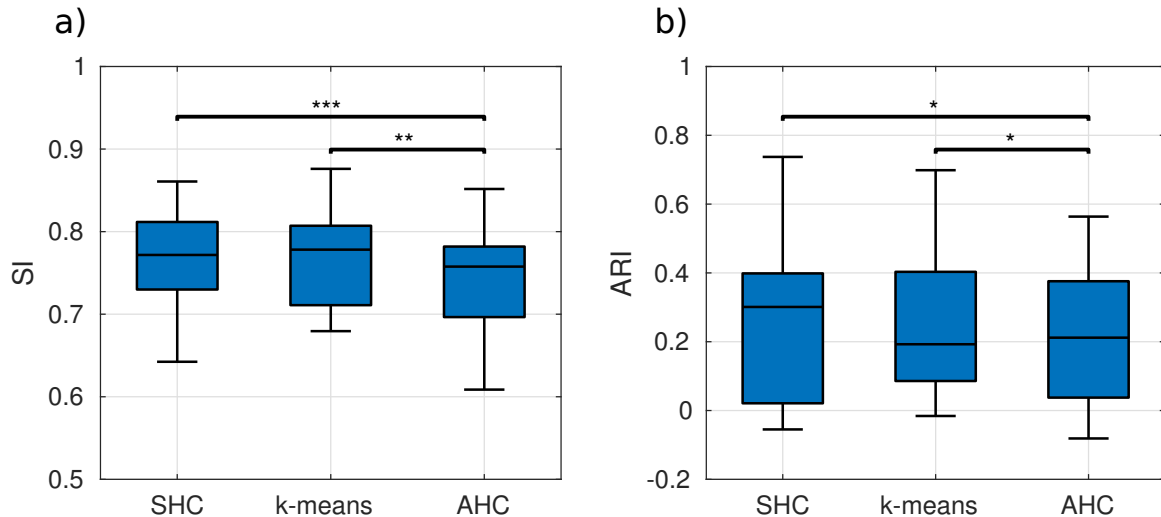
1. HRV
2.  $P_{VLF}$
3.  $P_{VLF}$  und HRV

Das Ziel bestand darin, die Kombination an Vitalparametern zu finden, die zur effektivsten Cluster-Einteilung führt.

### 5.2.3 Evaluierung und Validierung der Clustering-Ergebnisse

Die Evaluierung der Cluster-Ergebnisse erfolgte mit Hilfe des Silhouettenkoeffizient (SI)-Koeffizienten. Für jede der drei Vitalparameterkombinationen wurde zunächst die Cluster-Anzahl  $K$  ermittelt, für die der SI-Koeffizient maximal wird. Dies ergab einen maximalen SI-Koeffizienten pro Vitalparameterkombination. Anschließend wurde jene Vitalparameterkombination als endgültige Lösung gewählt, welche den größten maximalen SI-Koeffizienten aufwies.

Des Weiteren musste die Übereinstimmung der gefundenen optimalen Cluster-Lösung mit dem Goldstandard evaluiert werden. Die perfekte Lösung stellt dabei die Einteilung des Datensatzes in zwei Cluster dar, von denen einer das Auftreten von OSA zeitlich perfekt rekonstruiert. Es konnte allerdings nicht davon ausgegangen werden, dass die laut SI-Koeffizient optimale Cluster-Lösung mit nur zwei Clustern/Zuständen auskommt. Daher sind klassische Evaluationsmetriken für binäre Klassifikationen (z.B Area under Curve der Receiver Operating Characteristic) nicht anwendbar. Aus diesem Grund wurde für die Validierung der gefundenen Cluster auf den adjusted Rand index (ARI) zurückgegriffen [81]. Dieser bildet die Ähnlichkeit zweier Cluster-Einteilungen  $A$  und  $B$  auf den Wertebereich 0 (keine Ähnlichkeit) bis 1 (identisch) ab. Die Einschätzung der Ähnlichkeit basiert



**Abbildung 5.2:** a) Vergleich der maximalen SI-Indizes für alle betrachteten Methoden b) Vergleich der ARI-Werte aller betrachteten Methoden. Die statistische Signifikanz der Unterschiede im Median wurde mit Hilfe des Wilcoxon-Vorzeichen-Rang-Tests mit Bonferroni-Korrektur überprüft.

hauptsächlich auf der Anzahl von Datenpunktpaaren, die in den beiden Cluster-Lösungen  $A$  und  $B$  jeweils gemeinsam in ein Cluster oder in unterschiedliche Cluster eingeordnet werden. Existieren jedoch beispielsweise Datenpunktpaare, die in  $A$  gemeinsam in ein Cluster eingruppiert werden, in  $B$  jedoch nicht, senkt dies den ARI entsprechend, da beide Partitionen nicht mehr identisch sind. Dieses Vorgehen macht einen Vergleich von Cluster-Lösungen mit unterschiedlichen Cluster-Anzahlen möglich. Die statistische Signifikanz von beobachteten Unterschieden in den Verteilungen von SI und ARI über alle betrachteten Methoden wurde mittels des Wilcoxon-Vorzeichen-Rang-Test für abhängige und nichtnormalverteilte Stichproben überprüft [140]. Zur Vermeidung der Alphafehler-Kumulierung wurde das Signifikanzniveau mittels Bonferroni-Korrektur auf  $p \leq 0,0167$  festgelegt [142].

### 5.3 Ergebnisse

In Tabelle 5.1 sind die Werte für den maximalen SI-Koeffizienten sowie die jeweilige Cluster-Anzahl pro Proband und Cluster-Verfahren aufgelistet. Weiterhin sind die zu jeder optimalen Cluster-Lösung gehörigen ARI-Werte dargestellt. In Abbildung 5.2 sind die Verteilungen von SI-Koeffizient und ARI für jede Methode vergleichend dargestellt. Es ist zu sehen, dass die Verteilungen der SI-Koeffizienten für alle Methoden ähnlich sind. Unterschiede zeigen sich nur beim Vergleich der AHC-Methode mit den beiden anderen Verfahren. Der SI-Koeffizient der AHC Methode zeigt einen signifikant geringeren Median als AHC und k-means Methode. Da jedoch die meisten SI-Werte im Bereich zwischen 0,7

Proband	SHC			k-means			AHC		
	K	SI	ARI	K	SI	ARI	K	SI	ARI
a01	2	0,689	-0,036	2	0,686	-0,016	3	0,640	0,050
a02	2	0,857	0,397	2	0,863	0,426	5	0,700	0,182
a03	4	0,771	0,299	3	0,807	0,325	3	0,781	0,271
a04	2	0,829	0,737	2	0,813	0,641	2	0,783	0,564
a05	2	0,777	0,268	2	0,790	0,301	2	0,776	0,320
a06	2	0,795	0,028	2	0,750	0,132	2	0,792	0,030
a07	2	0,779	-0,014	3	0,712	0,054	2	0,779	-0,014
a08	3	0,668	0,265	2	0,718	0,229	2	0,663	0,212
a09	2	0,642	-0,055	2	0,679	0,123	2	0,623	-0,081
a10	6	0,657	0,346	4	0,692	0,271	4	0,609	0,288
a11	3	0,677	0,301	3	0,681	0,313	2	0,627	0,353
a12	2	0,792	0,395	3	0,783	0,135	3	0,735	0,106
a13	2	0,764	0,388	2	0,802	0,467	2	0,764	0,388
a14	2	0,768	0,237	6	0,772	0,098	8	0,733	0,105
a15	4	0,757	0,124	3	0,768	0,169	2	0,745	-0,076
a16	2	0,781	-0,044	9	0,793	0,058	2	0,760	-0,048
a17	2	0,810	0,453	2	0,790	0,392	2	0,790	0,392
a18	2	0,815	-0,053	10	0,696	0,017	2	0,827	-0,057
a19	2	0,838	0,699	2	0,838	0,699	3	0,812	0,561
a20	3	0,750	0,376	2	0,767	0,193	3	0,757	0,370
b01	2	0,756	0,531	4	0,745	0,061	2	0,758	0,154
b02	2	0,848	0,402	2	0,876	0,621	2	0,852	0,428
b03	2	0,791	0,636	2	0,807	0,584	5	0,769	0,455
b04	3	0,772	0,007	3	0,778	0,009	2	0,713	0,012
b05	2	0,674	0,331	2	0,708	0,171	2	0,689	0,359

**Tabelle 5.1:** Cluster-Ergebnisse der drei betrachteten Cluster-Verfahren. Dargestellt sind jeweils die Lösungen mit der Cluster-Anzahl K, für die der SI-Koeffizient maximal wird.

und 0,8 liegen, kann insgesamt von einer guten Partitionierung der Datensätze in Cluster ausgegangen werden.

Die Verteilungen der ARI-Werte bei den jeweiligen Methoden (Abbildung 5.2) zeigt nur zwischen SHC-Methode und AHC-Methode sowie zwischen k-means Methode und AHC-Methode signifikante Verbesserungen. Beim Vergleich zwischen SHC-Methode und k-means Methode ergeben sich keine statistisch signifikanten Unterschiede. Insgesamt ist festzustellen, dass die ARI-Werte bei allen Verfahren im niedrigen Bereich liegen. Zusätzlich existieren einige Datensätze (z.B. a01, a07, a09), bei denen ARI-Werte nahe Null auftreten, was auf eine große Diskrepanz zwischen optimaler Cluster-Lösung und Goldstandard hinweist. Auf der anderen Seite existieren auch Probanden, bei denen die Cluster-Verfahren einen hohen ARI-Wert erzielen (z.B. Probanden a04, a19, b03).

Bei der Bewertung der Cluster-Lösungen mit Hilfe des SI-Koeffizienten wird die Einteilung der Datensätze in zwei unterschiedliche Cluster am häufigsten als optimale Lösung ausgewählt. Dies entspricht der tatsächlichen Struktur des zugrunde liegenden Goldstandards. In etwa 28 % (SHC) bis 44 % (AHC) aller Datensätze werden anhand des SI-Koeffizienten Cluster-Lösungen mit  $K > 2$  als optimale Unterteilung des jeweiligen Parametersatzes angesehen. Dies schlägt sich in einem tendenziell niedrigeren ARI-Wert nieder. Jedoch gehen Cluster-Einteilungen mit  $K = 2$  nicht in jedem Fall mit hohen ARI-Werten einher. Dies zeigt, dass in vielen Fällen immer noch OSA-Phasen der Normal-Phase zugeordnet werden und anders herum.

## **5.4 Diskussion**

In dieser Studie wurde der Versuch unternommen, anhand von Vitalparametern Phasen von OSA mittels unüberwachter Lernverfahren zu detektieren. Die dafür verwendeten Daten wurden der Physionet-Datenbank entnommen. Ziel war das Aufzeigen eines weiteren Anwendungsfalles für die ähnlichkeitsbasierte hierarchische Cluster-Analyse. Als Referenzverfahren wurden, wie bereits in voran gegangenen Kapiteln, die k-means sowie die AHC-Methode herangezogen.

Die verwendeten Vitalparameter wurden mit Hilfe einer Zeit- und Frequenzanalyse des EKG-Signals gewonnen. Im Frequenzbereich wurde die Leistung im VLF-Band ausgewertet. In diesem Frequenzbereich wird der Einfluss des parasympathischen Nervensystems auf die Herzrate am deutlichsten. Es wurde erwartet, dass mit dem Eintreten von OSA auch die Leistung in diesem Band zunimmt. Für einen Großteil der Probanden war dies der Fall. Es wäre also zu erwarten gewesen, dass die Cluster-Verfahren zu einer exakten Trennung von normalen und pathologischen Schlafzuständen zu unterscheiden in der Lage sind. Trotzdem

ist an den ARI-Werten zu erkennen, dass dies bei einigen Datensätzen nur unzureichend funktioniert. Dies hat im Wesentlichen zwei Ursachen. Der Übergang von der Normalphase zu einer OSA-Phase war in vielen Fällen fließend. Dies führte zu einer schlechten Trennbarkeit der Datensätze in die entsprechenden Cluster. Der erkannte Übergang von einer Phase zu nächsten lag also nicht notwendigerweise am tatsächlichen Übergang, was zu einer Verminderung des ARI-Wertes führte. Zusätzlich werden, wie bereits in den Abschnitten 3.5 und 3.6 demonstriert, solche fließenden Übergänge als Übergangs-Cluster erkannt. Diese Erhöhung der erkannten Cluster-Anzahl geht auch mit einer Verminderung des ARI-Wertes einher.

Insgesamt betrachtet reichen die Leistungen der hier verwendeten unüberwachten Lernverfahren hinsichtlich der Genauigkeit bei der Detektion von OSA nicht an die Leistungen von überwachten Lernverfahren heran. Zywiets et al., aus deren Publikation auch der Ansatz über die Verwendung von HRV-Parametern stammt, geben eine Genauigkeit von 93,5 % an, welche mit Hilfe einer Diskriminanzanalyse erreicht wurde. Auch andere Autoren erzielen gute bis sehr gute Ergebnisse mit Hilfe von Parametern, welche aus dem EKG abgeleitet wurden [158, 159, 160].

## **Kapitel 6**

### **Zusammenfassung und Bewertung der Ergebnisse**

#### **6.1 Zusammenfassung**

Ziel dieser Dissertation war die Einführung einer neuen hierarchischen Cluster-Methode für die Anwendung bei der postoperativen Patientenüberwachung im klinischen Umfeld. Dem Verfahren liegt die Annahme zugrunde, dass sich die gesundheitliche Entwicklung eines Patienten durch eine Reihe von Zuständen beschreiben lässt. Ein Zustand ist dabei durch einen einen Satz an Vitalparametern charakterisiert. Durch die zeitliche Veränderung dieser Vitalparameter kommt es zu Übergängen zwischen Zuständen. Es wird weiterhin angenommen, dass gefundene Zustände anhand ihrer Ähnlichkeit zu neuen Gesundheitszuständen zusammengefasst werden können. Der Begriff der Ähnlichkeit wird hier im mathematischen Sinne als Abstand zweier Punkte, welche durch die jeweiligen charakteristischen Vitalparameter der beteiligten Zustände definiert werden, im N-dimensionalen Raum verstanden. Auf diese Weise entsteht ein hierarchisches Modell, mit dessen Hilfe eine Betrachtung der gesundheitlichen Entwicklung eines Patienten mit variabler zeitlicher Granularität möglich ist.

Nach einer Einleitung in Kapitel 1 wurde in Kapitel 2 zunächst eine kurze Einführung in das Gebiet des maschinellen Lernens gegeben. Der Fokus im ersten Teil dieses Abschnitts lag vor allem auf einer systematischen Einteilung der wichtigsten maschinellen Lernverfahren und der Herausarbeitung des unüberwachten maschinellen Lernens als Mittel zur Klassifikation von Vitaldaten. Als Beispiel für unüberwachte Lernverfahren wurden zwei etablierte Verfahren beschrieben, welche im weiteren Verlauf der Arbeit auch als Referenzmethoden dienten. Thema des zweiten Abschnitts war der aktuelle Forschungsstand bei der Überwachung von Patienten im Krankenhausumfeld. Am Anfang stand zunächst eine Beschreibung des Ist-Zustandes, welcher aus der schwellwertbasierten Überwachung des Patientenzustands mittels verschiedenster Patientenmonitore besteht. Im Anschluss wurden ausgewählte Forschungsarbeiten vorgestellt, welche mit Hilfe maschineller Lernverfahren eine Verbesserung der Patientenüberwachung oder Rehabilitation der Patienten zum Ziel hatten. Das Hauptaugenmerk lag hier vor allem auf Arbeiten, welche kardiorespiratorische Parameter und unüberwachte Lernverfahren benutzten. Aus den Ergebnissen dieses Kapitels

wurde die multimodale Auswertung von Vitaldaten zur Gesundheitszustandserfassung bei Patienten als offene Problemstellung identifiziert.

Kapitel 3 der Arbeit befasste sich mit der Einführung der ähnlichkeitsbasierten hierarchischen Cluster-Methode (SHC) als neuartiges Verfahren zur Überwachung des Gesundheitszustandes von stationär aufgenommenen Patienten. Es wurde zunächst ein Überblick über den Ablauf des Klassifizierungsalgorithmus gegeben. Einzelne wichtige Schritte wie z.B. die Definition des Abstandsmaße, der Cluster-Fusionierungsalgorithmus oder die Bewertung der gefunden Cluster-Lösung waren im Anschluss Gegenstand der Betrachtung. Da diese Schritte teilweise für die Verwendung mit der neuen Methode angepasst werden mussten, waren sie Gegenstand einer detaillierteren Betrachtung. Zur Validierung der neuen Methode wurden zwei Experimente konzipiert und durchgeführt. Im ersten Experiment wurden vier künstliche Datensätze verwendet. Diese bestanden aus Samples, welche in mehrere Cluster mit unterschiedlich ausgeprägter Überlappung unterteilt waren. Ziel war die korrekte Rekonstruktion der Cluster-Zugehörigkeit der einzelnen Samples unabhängig von der Überlappung. Hier konnte demonstriert werden, dass die SHC-Methode in der Lage ist, auch stark überlagerte Cluster voneinander zu trennen. Lediglich im Grenzbereich der verschiedenen Cluster wurden einige Samples den falsch zugeordnet. Im zweiten Experiment wurden reale Daten zur Klassifizierung verwendet. Das experimentelle Setup bestand aus verschiedenen Phasen unterschiedlicher physischer Belastung, welche von Probanden durchlaufen wurden. Die SHC-Methode wurde zur Rekonstruktion dieser Belastungsphasen anhand der während des Experiments abgeleiteten EKG-Signale verwendet. Aus einer Anzahl von Lösungen mit unterschiedlichen Cluster-Anzahlen wurde diejenige ausgewählt, welche sich hinsichtlich des SI-Koeffizienten als optimal herausstellte. Im Ergebnis konnte zunächst innerhalb eines einfach experimentellen Setups gezeigt werden, dass die SHC-Methode Gesundheitszustände anhand von Vitaldaten identifizieren kann.

In Kapitel 4 wurde die Methode in einem komplexen Szenario getestet, welche dem eigentlichen Anwendungsfall – der Patientenüberwachung im klinischen Umfeld – so nahe wie möglich kommen sollte. Es wurde eine Messstudie in Kooperation mit dem Herzzentrum Leipzig geplant und durchgeführt. Die Probandenkohorte setzte sich aus Patienten zusammen, welche sich einer koronaren Bypass-Operation unterzogen. Die Messung wurde während der ersten postoperativen 24 Stunden mit Hilfe eines zusätzlichen Messsystems (biosignals researcher, PLUX Wireless Biosignals S.A.) durchgeführt. Die erhobenen Daten umfassten die Biosignale EKG, Photoplethysmogramm (PPG) und Elektrodermale Aktivität (EDA) sowie die Zeitpunkte von Schmerzmittelgaben während der Dauer der Messung. Ziel der Studie war die Rekonstruktion dieser Schmerzmittelgaben mit Hilfe der



SHC-Methode. Zu diesem Zweck wurden aus den Biosignalen eine Vielzahl von Vitalparametern aus dem Zeit- und Frequenzbereich errechnet. Diese bildeten die Datengrundlage für Cluster-Einteilungen mit unterschiedlichen Cluster-Anzahlen. Der SI-Koeffizient bildete die Entscheidungsgrundlage für eine der errechneten Cluster-Lösungen. Die Erkennung der Medikamentengabe in der ausgewählten optimalen Cluster-Lösung wurde mit einem neu entwickelten Gütemaß bewertet. Um die Ergebnisse der SHC-Methode einordnen und vergleichen zu können, wurden die gleichen Untersuchungen nochmals mit den beiden Referenzmethoden AHC und k-means durchgeführt. Die mit Hilfe dieser Methoden erzeugten Cluster-Lösungen wurden mit dem gleichen Gütemaß bewertet und mit den Ergebnissen der SHC-Methode verglichen. Im Vergleich zwischen den Methoden zeigte sich, dass das AHC-Verfahren signifikant schlechtere Ergebnisse bei der Rekonstruktion von Gesundheitszuständen lieferte als die beiden anderen Verfahren. Zwischen SHC- und k-means Methode waren keine signifikanten Unterschiede zu beobachten. Unterschiede ergaben sich lediglich hinsichtlich einiger Teilparameter.

In Kapitel 5 wurde ein zusätzliches Anwendungsgebiet für die SHC-Methode vorgestellt. Anhand von verschiedenen Vitalparametern wurde versucht, OSA-Phasen. Die dafür verwendeten Vitaldaten wurden der PhysioNet-Datenbank entnommen. Mit dem ARI wurde ein Gütekriterium ausgewählt, welches die Ähnlichkeit des durch die SHC-Methode ermittelten Gesundheitszustandsverlaufs mit dem tatsächlich vorliegenden OSA-Phasen angibt. Die gewonnenen Ergebnisse wurden mit denen der Referenzverfahren verglichen.

## **6.2 Bewertung und Einordnung der Ergebnisse bei der Erkennung von Gesundheitszuständen**

In dieser Arbeit wurde mit der SHC-Methode ein neues hierarchisches Verfahren zur Klassifizierung von Gesundheitszuständen anhand von Vitalparametern vorgestellt. Die Anwendbarkeit und Effektivität wurde im Rahmen von mehreren Studien untersucht. Zum einen bestand die Validierung der SHC-Methode darin, physische Belastungszustände mit Hilfe eines Datensatzes von EKG-Parametern zu erkennen. Zum anderen wurde ein Messstudie am Herzzentrum Leipzig mit einem umfangreicheren Setup durchgeführt. Hier wurde versucht, anhand von Änderungen der erkannten Gesundheitszustände auf die Gabe von Analgetika infolge postoperativer Schmerzen zu schließen. Es wurden zwei weitere Verfahren des unüberwachten maschinellen Lernens als Referenzverfahren verwendet.

Die Ergebnisse zeigen zunächst, dass es grundsätzlich möglich ist, einen Vitaldatensatz mit Hilfe unüberwachter Lernverfahren in sinnvolle Gesundheitszustände zu unterteilen. Aus den Resultaten der Validierungsstudie geht hervor, dass die SHC-Methode Phasen unter-

schiedlicher physischer Belastung als separate Gesundheitszustände gut voneinander trennen kann. So konnte beispielsweise in der Validierungsstudie bei dem als Belastungszustand erkannten Cluster eine deutlich erhöhte Herzrate (HR) im Vergleich zum Grundzustand festgestellt werden. Da es sich hierbei um eine Studie zur Überprüfung der Eignung der SHC-Methode für die Detektion von Gesundheitszuständen handelt, wurden noch keine Vergleiche mit Ergebnissen von Referenzmethoden angestellt.

Im Rahmen der Messstudie konnte mit Hilfe der SHC-Methode Zustände identifiziert werden, die sich mit postoperativen Schmerzen in Verbindung bringen ließen. Weiterhin konnten Zustandsänderungen in zeitlicher Nähe zu protokollierten Schmerzmittelgaben beobachtet werden. Die Ergebnisse der Verfahren wurden mit Hilfe eines für diesen Zweck entwickelten Güteparameters  $Q$  beurteilt und verglichen. Die AHC-Methode schnitt dabei am schlechtesten ab. Zwischen SHC und k-means Methode gab es hinsichtlich des Güteparameters  $Q$  marginale Unterschiede, welche statistisch nicht signifikant waren. Der Hauptunterschied zwischen beiden Methoden lag im höheren SI-Koeffizienten bei der k-means Methode bei ansonsten gleich guter zeitlicher Genauigkeit bei der Erkennung von Schmerzmittelgaben (Parameter  $T_e$ ). Diese Abweichung in den SI-Koeffizienten liegt in den unterschiedlichen methodischen Ansätzen der Verfahren begründet. Während das k-means Verfahren auf eine globale Optimierung der Cluster-Struktur abzielt, ist die Optimierung des SHC-Verfahrens durch die iterative Kombination von Clustern eher lokaler Natur. Diese lokale Cluster-Optimierung führt zu geringeren SI-Koeffizienten, da bei der Kombination einzelner Cluster immer nur ein Teil des Datensatzes berücksichtigt wird. Da sich beide Methoden hinsichtlich des Teilparameters  $T_e$  nicht signifikant unterscheiden, scheint dies jedoch kein Nachteil der SHC-Methode gegenüber dem k-means Verfahren zu sein. Bei der Erkennung von obstruktiver Schlafapnoe (OSA) in Anhang 5 zeigt sich ein ähnliches Bild wie bei der Mess- und Validierungsstudie. Hier schnitt SHC-Verfahren hinsichtlich des ARI-Güteparameters signifikant besser als das AHC-Verfahren ab. Zwischen SHC- und k-means Methode konnte jedoch kein Unterschied festgestellt werden.

Aus den Ergebnissen der Messstudie kann also geschlossen werden, dass maschinelle Lernverfahren die Wirkung von Analgetika, welche sich durch einen Gesundheitszustandsübergang auszeichnen, erkennen können. Ferner kann geschlussfolgert werden, dass es keinen statistisch signifikanten Unterschiede zwischen hierarchischen – konkret der SHC-Methode – und partitionierenden (k-means Methode) Verfahren hinsichtlich der Rekonstruktionsleistung gibt. Die hier vorgestellte SHC-Methode scheint also mindestens so gut für die Detektion von Gesundheitszuständen geeignet zu sein wie die Referenzmethode.

Es existieren jedoch Eigenschaften von hierarchischen Cluster-Methoden, speziell der SHC-Methode, die diese gegenüber partitionierenden Verfahren im Vorteil erscheinen lassen. Hierarchische Verfahren treffen nur sehr wenige Annahmen über die Organisation der Cluster-Struktur eines Datensatzes oder die Verteilung von Samples innerhalb dieser Cluster. Daher lassen sich Verfahren wie die SHC-Methode auf alle Datensätze anwenden, bei denen sich in irgendeiner Form paarweise Distanzen bzw. Ähnlichkeiten zwischen den darin enthaltenen Samples und Clustern berechnen lassen. Auf diese Weise ist auch das Clustering von nominalen und ordinalen Daten möglich [161, 162, 163]. Solche Daten treten beispielsweise bei der Verarbeitung von demographischen Informationen von Patienten auf. Dabei muss erfüllt sein, dass sich im Wertebereich der Parameter ein Abstandsmaß definieren lässt. Dies geschieht durch Unterteilung des Wertebereichs der Variable in passende Intervalle. Spezielles Augenmerk muss hier auf die Unterteilung des auftretenden Wertebereichs und eine sinnvolle Übersetzung in numerische Äquivalente gelegt werden [162]. Der k-means Algorithmus kann diese Datentypen nicht direkt verwenden sodass Abwandlungen dieses Algorithmus verwendet werden müssen [164, 165].

Ein weiterer wichtiger Unterschied liegt in den Anfangsbedingungen der Algorithmen. Die Initialisierung des k-means Verfahrens besteht in der *a-priori* Definition der Anzahl der im Datensatz zu erwartenden Cluster. Diese Anzahl ist bei einer explorativen Analyse umfangreicher Datensätze mit einer Vielzahl von unabhängigen Variablen im Vorfeld meist nicht bekannt. Im Gegensatz dazu ist bei der SHC-Methode die Angabe der Cluster-Anzahl im Vorfeld nicht notwendig. Da sich eine Cluster-Lösung mit einer bestimmten Cluster-Anzahl  $K$  iterativ durch Zusammenfassen aus der vorhergehenden Cluster-Lösung ergibt, muss lediglich ein Abbruchkriterium definiert werden. Bei der Validierungsstudie in Abschnitt 3.6 konnte anschaulich demonstriert werden, dass selbst bei einem scheinbar binären Klassifikationsproblem (Belastung/Ruhe) mehr als zwei Cluster zur Strukturierung des zugrunde liegenden Datensatzes notwendig sein können. Vor allem Vitalparameter werden immer von mehr als einer Einflussvariable moduliert, was Aussagen über die Struktur eines Datensatzes im Vorfeld der Cluster-Analyse erschwert. Das gleiche Bild bot sich auch bei der Auswertung der Ergebnisse der Messstudie. Bei den Datensätzen, bei denen eine Rekonstruktion der Schmerzmittelgabe erfolgreich war, lag die laut SI-Koeffizient optimale Cluster-Anzahl bei  $K \geq 2$ . Eine pauschale Strukturierung der Datensätze in Schmerz- und Normalzustand ist also auch hier nicht möglich. Weiterhin werden die Cluster-Zentren bei der k-means Methode zufällig initialisiert. Dies führt bei mehreren Durchläufen des Algorithmus zu unterschiedlichen Ergebnissen bei ansonsten gleich bleibenden Rahmenbedingungen. Die iterative Unterteilung eines großen Datensatzes in eine Menge von Zuständen unter-

schiedlicher Granularität ist damit nahe an der klinischen Anwendung. Diesem Umstand trägt die SHC-Methode besser Rechnung als die k-means Methode.

Wie bereits am Beginn der Arbeit erwähnt, findet das maschinelle Lernen in der Medizintechnik bereits breite Anwendung. In der klinischen Praxis bei der Überwachung von Patienten sind maschinelle Lernverfahren aber, abgesehen von wenigen Ausnahmen, noch nicht angekommen. Der aktuelle Stand der Technik ist nach wie vor von schwellwertbasierten Methoden bestimmt, die oft nur einzelne Vitalparameter im Blick haben und damit wenig spezifisch auf akute Verschlechterungen des Patientenzustandes reagieren können. Trotzdem werden, wie in Kapitel 2 gezeigt, die Anwendung multimodaler Ansätze und maschinellen Lernens in der Fachwelt als sinnvoller Beitrag zur Erhöhung der Patientensicherheit sowie zur Reduktion der Arbeitsbelastung des Krankenhauspersonals angesehen. Mit der SHC-Methode wurde ein Werkzeug vorgestellt, das eine neue Form der Patientenüberwachung ermöglicht. Der aktuelle Zustand des Patienten wird nicht nur durch den derzeitigen Wert der Vitalparameter charakterisiert. Durch die Beschreibung der Entwicklung des Patienten in Form von Gesundheitszuständen ist damit eine ganzheitliche Bewertung des Patienten auf der Basis multipler Vitalparameter möglich. Das Prinzip der Zerlegung eines Vitalparameterdatensatzes in eine hierarchische Struktur aus Gesundheitszuständen erwies sich dabei als geeignetes Mittel, um eine Beschreibung der Vitalparameterentwicklung mit variabler zeitlicher Auflösung zu erreichen. Es können damit sowohl langfristige Veränderungen der Vitalparameter (z.B. die Erholung nach einem chirurgischen Eingriff) als auch kurzfristige Reaktionen (z.B. auf die Verabreichung bestimmter Medikamente) einfach beobachtet werden. Somit steht dem Krankenhauspersonal ein neuartiges Werkzeug zur Verfügung, das eine frühzeitige Erkennung potentiell kritischer Gesundheitszustände ermöglicht. Außerdem kann das Verfahren einen Beitrag zur Verbesserung der Patientensicherheit und dem effektiven Einsatz von Krankenhausressourcen leisten.

## **Kapitel 7**

### **Ausblick**

Mit der SHC-Methode wurde ein Verfahren vorgestellt, welches zur Detektion von Gesundheitszuständen anhand von Vitalparametern einsetzbar ist. Im Rahmen dieser Arbeit wurde das Verfahren so weit entwickelt, dass es medizinisch relevante Ereignisse aus einem komplexen Datensatz bestehend aus verschiedenen Biosignalen erkennen kann. Der Zweck des Verfahrens ist jedoch allgemein die Erkennung von Gesundheitszuständen. Dafür sind noch weitere Forschungsarbeiten nötig. Beispielsweise wäre neben der Erkennung von Gesundheitszuständen auch eine Vorhersage dieser interessant. Hierzu wäre die Integration der gefundenen Zustände in einen Markow-Prozess denkbar [166]. Das Prinzip der Zustandsbeschreibung und -vorhersage mit Hilfe von Markow-Ketten als Hilfsmittel zur Entscheidungsfindung hat bereits in vielen Bereichen Anwendung. In der Gesundheitsversorgung wird dieses Verfahren bisher hauptsächlich bei der Analyse und Optimierung von administrativen Prozessen verwendet [167, 168, 169]. Die Vorhersage von Gesundheitszuständen anhand vergangener Zustände ist jedoch auch denkbar [170, 171].

Im Rahmen der Messstudie am Herzzentrum Leipzig (Kapitel 4) wurden mit Hilfe der SHC-Methode Aussagen über Gesundheitszustände auf Probandenebene getroffen. Indem Reaktionen der Vitalparameter auf die Gabe von Schmerzmedikamenten untersucht wurden, wurde das Augenmerk vor allem auf kurzfristige Entwicklung des Patientenzustandes gelegt. Untersuchungen längerfristiger Entwicklungen mit Hilfe der SHC-Methode können Aussagen über den generellen Rehabilitationsverlauf des Patienten ermöglichen. Dadurch können Vorhersagen über später auftretende Komplikationen getroffen werden.

Interessant wäre auch ein probandenübergreifender Vergleich. Eine solche Analyse könnte Ähnlichkeiten in den Entwicklungen der Patienten in den ersten 24 Stunden nach einem chirurgischen Eingriff aufdecken. Damit könnte beispielsweise, vergleichbar zu den Untersuchungen in [64], die Frage beantwortet werden, ob es Zustände gibt, die alle Patienten nach der Operation durchlaufen. Auf diese Weise könnte es mittelfristig möglich sein, Anomalien zu identifizieren, die einen prognostischen Wert für die weitere Entwicklung haben. Auch eine längere Begleitung der Patienten über die Entlassung aus dem Krankenhaus hinaus könnte helfen, Gesundheitszustände zu identifizieren, die typisch für eventuelle langfristige

Komplikationen sind. In der Onkologie werden maschinelle Lernverfahren beispielsweise zur Langzeitprognose eingesetzt [172, 173, 174]. Nach einer umfassenden Literaturrecherche fanden sich nur wenige Studien, die dies auch in der Herzchirurgie angewandt haben, sodass hier möglicherweise ein bisher noch unbeachtetes Anwendungsgebiet liegt. Für diesen Zweck ist die derzeitige Anzahl an Datensätzen in der Messstudie jedoch noch zu gering. Zukünftige Untersuchungen müssten daher mit einer größeren Patientenkohorte durchgeführt werden, um aussagekräftig zu sein.

Weiterhin sind vergleichende Untersuchungen hinsichtlich der verwendeten Distanzmaße und Cluster-Validitätsmaße anzustellen. In allen Studien wurde die euklidische Distanz zur Berechnung der Ähnlichkeit von Gesundheitszuständen verwendet. Da Vitalparameter aber möglicherweise auch miteinander korreliert sein können, stellt die Mahalanobis-Distanz eine Alternative dar. Auch andere Distanzmaße wie z.B. die Manhattan-Distanz können Teil von weiteren Untersuchungen sein. Auf diese Weise wäre auch die Verwendung von ordinalskalierten Parametern denkbar. Der SI-Koeffizient wurde zur Bewertung von berechneten Cluster-Lösungen eingesetzt. Er stellt aber bei weitem nicht das einzige Maß zur Cluster-Validierung dar. In der Molekularbiologie wird beispielsweise das Consensus-Clustering zur Klassifizierung von Gensequenzen oder Krebszellen eingesetzt [87, 175, 176].

Wie im Abschnitt 4.3 dargelegt, sind neben HRV-Parametern auch weitere physiologische und nichtphysiologische Kenngrößen für die Zustandsbeschreibung eines Patienten notwendig. Im Rahmen eines drittmittelgeförderten Projektes wird am LaBP ein kontaktloses System für das Patientenmonitoring entwickelt. Mit Hilfe von im Raum verteilten Kameras sollen Parameter von Patienten auf der Normalstation eines Krankenhauses erhoben werden. Diese umfassen physiologische Daten wie Herzrate, Atemfrequenz oder Blutdruck [121, 177, 178]. Weiterhin sollen auf optischem Wege Verhaltensdaten erhoben werden um Aktivitätsintensität und -muster verfolgen zu können [179, 180]. Mit Hilfe dieser Daten und der SHC-Methode soll ein kontaktloses System zur Patientenüberwachung realisiert werden.

Es sind also noch einige Schritte zu unternehmen, bis eine praktische Anwendung des Verfahrens im klinischen Alltag möglich ist. Vor allem die geeignete Visualisierung der Ergebnisse stellt eine wichtige Hürde auf dem Weg dorthin dar. Die Cluster-Ergebnisse und relevanten Vitalparameter müssen optisch so aufbereitet werden, dass sie für medizinisches Personal leicht interpretierbar sind und miteinander in Beziehung gesetzt werden können. Auch sollten aus dem Satz an erhobenen Parametern nur die für die jeweilige Zustandserkennung Relevantesten visualisiert werden. Auf diesem Gebiet sind noch einige Forschungsarbeiten nötig. Auch die Laufzeit des Verfahrens muss verbessert werden. Dies ist vor allem vor dem Hintergrund der Verwendung der Methode zur kontaktlosen

Überwachung von Patienten von großer Bedeutung. Momentan werden die Daten offline analysiert. In der klinischen Anwendung sind jedoch auch Echtzeitanalysen von Bedeutung. Da die Methode zu den hierarchischen Verfahren zählt, benötigt vor allem die Berechnung der Ähnlichkeitsmatrix  $\mathbf{D}$  in jedem Iterationsschritt viel Rechenleistung. Auch hier müssen Schritte unternommen werden, um die nötigen Berechnungen effektiver zu gestalten.





## Literatur

- [1] D. Fagella. (2018). „7 Applications of Machine Learning in Pharma and Medicine,“ Adresse: [www.techemergence.com/machine-learning-in-pharma-medicine/](http://www.techemergence.com/machine-learning-in-pharma-medicine/) (besucht am 19.11.2018).
- [2] F. Rosenblatt, „The perceptron: a probabilistic model for information storage and organization in the brain.,“ *Psychological review*, Jg. 65, Nr. 6, S. 386, 1958.
- [3] R. C. Deo und M. Şahin, „Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia,“ *Atmospheric Research*, Jg. 153, S. 512–525, 2015.
- [4] K. Rasouli, W. W. Hsieh und A. J. Cannon, „Daily streamflow forecasting by machine learning methods with weather and climate inputs,“ *Journal of Hydrology*, Jg. 414, S. 284–293, 2012.
- [5] A. L. Pyayt, I. I. Mokhov, B. Lang, V. V. Krzhizhanovskaya, R. J. Meijer u. a., „Machine learning methods for environmental monitoring and flood protection,“ *World Academy of Science, Engineering and Technology*, Jg. 5, Nr. 4, S. 118–123, 2011.
- [6] H. Shao, H. Zhao und G.-R. Chang, „Applying data mining to detect fraud behavior in customs declaration,“ in *Proceedings of the International Conference on Machine Learning and Cybernetics*, IEEE, Bd. 3, 2002, S. 1241–1244.
- [7] J. Perols, „Financial statement fraud detection: An analysis of statistical and machine learning algorithms,“ *Auditing: A Journal of Practice & Theory*, Jg. 30, Nr. 2, S. 19–50, 2011.
- [8] E. Aramaki, S. Maskawa und M. Morita, „Twitter catches the flu: detecting influenza epidemics using Twitter,“ in *Proceedings of the conference on empirical methods in natural language processing*, Association for Computational Linguistics, 2011, S. 1568–1576.
- [9] A. Culotta, „Towards detecting influenza epidemics by analyzing Twitter messages,“ in *Proceedings of the First Workshop on Social Media Analytics*, Association for Computing Machinery, 2010, S. 115–122.

- [10] P. Guo, T. Liu, Q. Zhang, L. Wang, J. Xiao, Q. Zhang, G. Luo, Z. Li, J. He, Y. Zhang u. a., „Developing a dengue forecast model using machine learning: A case study in China,“ *PLoS Neglected Tropical Diseases*, Jg. 11, Nr. 10, e0005973, 2017.
- [11] E. I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E. R. Melhem und C. Davatzikos, „Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme,“ *Magnetic Resonance in Medicine*, Jg. 62, Nr. 6, S. 1609–1618, 2009.
- [12] H. G. Schnack, M. Nieuwenhuis, N. E. van Haren, L. Abramovic, T. W. Scheewe, R. M. Brouwer, H. E. H. Pol und R. S. Kahn, „Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects,“ *Neuroimage*, Jg. 84, S. 299–306, 2014.
- [13] E. I. Zacharaki, V. G. Kanas und C. Davatzikos, „Investigating machine learning techniques for MRI-based classification of brain neoplasms,“ *International Journal of Computer Assisted Radiology and Surgery*, Jg. 6, Nr. 6, S. 821–828, 2011.
- [14] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, A. D. N. Initiative u. a., „Machine learning framework for early MRI-based Alzheimer’s conversion prediction in MCI subjects,“ *Neuroimage*, Jg. 104, S. 398–412, 2015.
- [15] P. Prechtl und F.-P. Burkard, *Metzler Lexikon Philosophie: Begriffe und Definitionen*. Springer-Verlag, 2015.
- [16] T. G. Dietterich und R. S. Michalski, „A comparative review of selected methods for learning from examples,“ in *Machine Learning, Volume I*, Elsevier, 1983, S. 41–81.
- [17] T. M. Mitchell, *The discipline of machine learning*. Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2006, Bd. 9.
- [18] Clarivate Analytics, Hrsg. (2018). „Web of Science,“ Adresse: [www.webofknowledge.com](http://www.webofknowledge.com) (besucht am 28.02.2018).
- [19] P. Louridas und C. Ebert, „Machine Learning,“ *IEEE Software*, Jg. 33, Nr. 5, Sep. 2016.
- [20] K. Backhaus, B. Erichson, W. Plinke und R. Weiber, *Multivariate Analysemethoden: eine anwendungsorientierte Einführung*. Springer-Verlag, 2015.
- [21] A. K. Jain und R. C. Dubes, *Algorithms for Clustering Data*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988.

- [22] M. Elliott und A. Coventry, „Critical care: the eight vital signs of patient monitoring,“ *British Journal of Nursing*, Jg. 21, Nr. 10, S. 621–625, 2012.
- [23] P. S. Chan, R. Jain, B. K. Nallmothu, R. A. Berg und C. Sasson, „Rapid response teams: a systematic review and meta-analysis,“ *Archives of Internal Medicine*, Jg. 170, Nr. 1, S. 18–26, 2010.
- [24] J. Chen, R. Bellomo, A. Flabouris, K. Hillman und S. Finfer, „The relationship between early emergency team calls and serious adverse events,“ *Critical Care Medicine*, Jg. 37, Nr. 1, S. 148–153, 2009.
- [25] K. Hillman, J. Chen, M. Cretikos, R. Bellomo, D. Brown, G. Doig, S. Finfer, A. Flabouris, M. S. Investigators u. a., „Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial,“ *Lancet*, Jg. 365, Nr. 9477, S. 2091–2097, 2005.
- [26] I. Mitchell, C. Van Leuvan u. a., „Missed opportunities? An observational study of vital sign measurements,“ *Critical Care and Resuscitation*, Jg. 10, Nr. 2, S. 111, 2008.
- [27] C. Herting, E. Herting, S. Kluge, A. Markewitz, G. Marx, E. Muhl, T. Nicolai, K. Notz, V. Parvu, M. Quintel, E. Rickels, D. Schneider, K. Steinmeyer Bauer, G. Sybrecht und T. Welte, „Intermediate Care Station – Empfehlungen zur Ausstattung und Struktur (Langversion),“ Deutsche Interdisziplinäre Vereinigung für Intensiv- und Notfallmedizin, 2017.
- [28] D. Edvardsson, R. Nay u. a., „Acute care and older people: challenges and ways forward,“ *The Australian Journal of Advanced Nursing*, Jg. 27, Nr. 2, S. 63–69, 2009.
- [29] J. Böcken und T. Kostera, „Faktencheck Pflegepersonal im Krankenhaus – Internationale Empirie und Status quo in Deutschland,“ Bertelsmann Stiftung, 2017.
- [30] M. Cardona-Morrell, M. Prgomet, R. Lake, M. Nicholson, R. Harrison, J. Long, J. Westbrook, J. Braithwaite und K. Hillman, „Vital signs monitoring and nurse–patient interaction: A qualitative observational study of hospital practice,“ *International Journal of Nursing Studies*, Jg. 56, S. 9–16, 2016.
- [31] L. H. Aiken, D. M. Sloane, L. Bruyneel, K. Van den Heede, P. Griffiths, R. Busse, M. Diomidous, J. Kinnunen, M. Kózka, E. Lesaffre u. a., „Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study,“ *Lancet*, Jg. 383, Nr. 9931, S. 1824–1830, 2014.

- [32] L. H. Aiken, S. P. Clarke, D. M. Sloane, E. T. Lake und T. Cheney, „Effects of hospital care environment on patient mortality and nurse outcomes,“ *The Journal of Nursing Administration*, Jg. 38, Nr. 5, S. 223, 2008.
- [33] A. Lee, Y. S. L. Cheung, G. M. Joynt, C. C. H. Leung, W.-T. Wong und C. D. Gomersall, „Are high nurse workload/staffing ratios associated with decreased survival in critically ill patients? A cohort study,“ *Annals of Intensive Care*, Jg. 7, Nr. 1, S. 46, 2017.
- [34] J. B. Cooper, R. S. Newbower und R. J. Kitz, „An analysis of major errors and equipment failures in anesthesia management: considerations for prevention and detection.,“ *Anesthesiology*, Jg. 60, Nr. 1, S. 34–42, 1984.
- [35] T. Short, A. O’regan, J. Lew und T. Oh, „Critical incident reporting in an anaesthetic department quality assurance programme,“ *Anaesthesia*, Jg. 48, Nr. 1, S. 3–7, 1993.
- [36] J. Edworthy und E. Hellier, „Fewer but better auditory alarms will improve patient safety,“ *BMJ Quality & Safety*, Jg. 14, Nr. 3, S. 212–215, 2005.
- [37] F. Schmid, M. S. Goepfert und D. A. Reuter, „Patient monitoring alarms in the ICU and in the operating room,“ *Critical Care*, Jg. 17, Nr. 2, 2013.
- [38] L. Varpio, C. Kuziemsky, C. MacDonald und W. J. King, „The Helpful or Hindering Effects of In-Hospital Patient Monitor Alarms on Nurses,“ *CIN: Computers, Informatics, Nursing*, Jg. 30, Nr. 4, S. 210–217, 2012.
- [39] S. Sendelbach und M. Funk, „Alarm fatigue: a patient safety concern,“ *AACN Advanced Critical Care*, Jg. 24, Nr. 4, S. 378–386, 2013.
- [40] I. A. Beinlich, H. J. Suermondt, R. M. Chavez und G. F. Cooper, „The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks,“ in *AIME 89*, Springer-Verlag Berlin Heidelberg, 1989, S. 247–256.
- [41] M. Factor, D. H. Gelernter, C. E. Kolb, P. L. Miller und D. F. Sittig, „Real-time data fusion in the intensive care unit,“ *Computer*, Jg. 24, Nr. 11, S. 45–54, 1991.
- [42] E. M. Koski, A. Mäkivirta, T. Sukuvaara und A. Kari, „Development of an expert system for haemodynamic monitoring: computerized symbolization of on-line monitoring data,“ *International journal of clinical monitoring and computing*, Jg. 8, Nr. 4, S. 289–293, 1991.
- [43] S. Uckun, B. M. Dawant und D. P. Lindstrom, „Model-based diagnosis in intensive care monitoring: the YAQ approach,“ *Artificial Intelligence in Medicine*, Jg. 5, Nr. 1, S. 31–48, 1993.

- [44] S. Uckun und B. M. Dawant, „Qualitative modeling as a paradigm for diagnosis and prediction in critical care environments,“ *Artificial Intelligence in Medicine*, Jg. 4, Nr. 2, S. 127–144, 1992.
- [45] B. Dawant, S. Uckun, E. Manders und D. Lindstrom, „The SIMON project: model-based signal acquisition, analysis, and interpretation in intelligent patient monitoring,“ *IEEE Engineering in Medicine and Biology Magazine*, Jg. 12, Nr. 4, S. 82–91, 1993.
- [46] K. Becker, G. Rau, H. Kaesmacher, M. Petermeyer, G. Kalff und H.-J. Zimmermann, „Fuzzy logic approaches to intelligent alarms,“ *IEEE Engineering in Medicine and Biology Magazine*, Jg. 13, Nr. 5, S. 710–716, 1994.
- [47] K. Becker, B. Thull, H. Käsmacher-Leidinger, J. Stemmer, G. Rau, G. Kalff und H.-J. Zimmermann, „Design and validation of an intelligent patient monitoring and alarm system based on a fuzzy logic process model,“ *Artificial intelligence in medicine*, Jg. 11, Nr. 1, S. 33–53, 1997.
- [48] W. A. Knaus, J. E. Zimmerman, D. P. Wagner, E. A. Draper und D. E. Lawrence, „APACHE-acute physiology and chronic health evaluation: a physiologically based classification system.,“ *Critical care medicine*, Jg. 9, Nr. 8, S. 591–597, 1981.
- [49] W. A. Knaus, E. A. Draper, D. P. Wagner und J. E. Zimmerman, „APACHE II: a severity of disease classification system.,“ *Critical care medicine*, Jg. 13, Nr. 10, S. 818–829, 1985.
- [50] W. A. Knaus, D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P. G. Bastos, C. A. Sirio, D. J. Murphy, T. Lotring, A. Damiano u. a., „The APACHE III prognostic system: risk prediction of hospital mortality for critically III hospitalized adults,“ *Chest*, Jg. 100, Nr. 6, S. 1619–1636, 1991.
- [51] J. E. Zimmerman, A. A. Kramer, D. S. McNair und F. M. Malila, „Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today’s critically ill patients,“ *Critical care medicine*, Jg. 34, Nr. 5, S. 1297–1310, 2006.
- [52] A. P. Nassar Junior, A. O. Mocelin, F. M. Andrade, L. Brauer, F. P. Giannini, A. L. B. Nunes und C. A. Dias, „SAPS 3, APACHE IV or GRACE: which score to choose for acute coronary syndrome patients in intensive care units?“ *Sao Paulo Medical Journal*, Jg. 131, Nr. 3, S. 173–178, 2013.

- [53] V.-C. Wu, H.-B. Tsai, Y.-C. Yeh, T.-M. Huang, Y.-F. Lin, N.-K. Chou, Y.-S. Chen, Y.-Y. Han, A. Chou, Y.-H. Lin u. a., „Patients supported by extracorporeal membrane oxygenation and acute dialysis: acute physiology and chronic health evaluation score in predicting hospital mortality,“ *Artificial organs*, Jg. 34, Nr. 10, S. 828–835, 2010.
- [54] B. Khwannimit und R. Bhurayanontachai, „A comparison of the performance of Simplified Acute Physiology Score 3 with old standard severity scores and customized scores in a mixed medical-coronary care unit.,“ *Minerva anesthesiologica*, Jg. 77, Nr. 3, S. 305–312, 2011.
- [55] J. I. Salluh und M. Soares, „ICU severity of illness scores: APACHE, SAPS and MPM,“ *Current opinion in critical care*, Jg. 20, Nr. 5, S. 557–565, 2014.
- [56] L. Tarassenko, A. Hann und D. Young, „Integrated monitoring and analysis for early warning of patient deterioration,“ *British Journal of Anaesthesia*, Jg. 97, Nr. 1, S. 64–68, 2006.
- [57] D. Apiletti, E. Baralis, G. Bruno und T. Cerquitelli, „Real-Time Analysis of Physiological Data to Support Medical Applications,“ *IEEE Transactions on Information Technology in Biomedicine*, Jg. 13, Nr. 3, S. 313–321, 2009.
- [58] M. M. Baig, H. G. Hosseini und M. Linden, „Machine learning-based clinical decision support system for early diagnosis from real-time physiological data,“ in *Proceedings of the 2016 IEEE Region 10 Conference (TENCON)*, IEEE, 2016, S. 2943–2946.
- [59] U. Anliker, J. Ward, P. Lukowicz, G. Troster, F. Dolveck, M. Baer, F. Keita, E. Schenker, F. Catarsi, L. Coluccini, A. Belardinelli, D. Shklarski, M. Alon, E. Hirt, R. Schmid und M. Vuskovic, „AMON: A Wearable Multiparameter Medical Monitoring and Alert System,“ *IEEE Transactions on Information Technology in Biomedicine*, Jg. 8, Nr. 4, S. 415–427, 2004.
- [60] E. L. Bose, G. Clermont, L. Chen, A. W. Dubrawski, D. Ren, L. A. Hoffman, M. R. Pinsky und M. Hravnak, „Cardiorespiratory instability in monitored step-down unit patients: using cluster analysis to identify patterns of change,“ *Journal of Clinical Monitoring and Computing*, Jg. 32, Nr. 1, S. 117–126, 2018.
- [61] K. C. Vranas, J. K. Jopling, T. E. Sweeney, M. C. Ramsey, A. S. Milstein, C. G. Slatore, G. J. Escobar und V. X. Liu, „Identifying Distinct Subgroups of ICU Patients: A Machine Learning Approach,“ *Critical Care Medicine*, Jg. 45, Nr. 10, S. 1607–1615, 2017.

- [62] M. A. Pimentel, D. A. Clifton und L. Tarassenko, „Gaussian process clustering for the functional characterisation of vital-sign trajectories,“ in *IEEE International Workshop on Machine Learning for Signal Processing, MLSP*, IEEE, 2013, S. 1–6. eprint: 13.
- [63] M. D. Sorani, J. C. Hemphill, D. Morabito, G. Rosenthal und G. T. Manley, „New approaches to physiological informatics in neurocritical care,“ *Neurocritical Care*, Jg. 7, Nr. 1, S. 45–52, 2007.
- [64] M. J. Cohen, A. D. Grossman, D. Morabito, M. M. Knudson, A. J. Butte und G. T. Manley, „Identification of complex metabolic states in critically injured patients using bioinformatic cluster analysis,“ *Critical Care*, Jg. 14, Nr. 1, R10, 2010.
- [65] Q. Xu, T. L. Nwe und C. Guan, „Cluster-Based Analysis for Personalized Stress Evaluation Using Physiological Signals,“ *IEEE Journal of Biomedical and Health Informatics*, Jg. 19, Nr. 1, S. 275–281, 2015.
- [66] M. Paoletti und C. Marchesi, „Discovering dangerous patterns in long-term ambulatory ECG recordings using a fast QRS detection algorithm and explorative data analysis,“ *Computer Methods and Programs in Biomedicine*, Jg. 82, Nr. 1, S. 20–30, 2006.
- [67] C. D. Spielberger und S. J. Sydeman, „State-Trait Anxiety Inventory and State-Trait Anger Expression Inventory.,“ in *The use of psychological testing for treatment planning and outcome assessment*, M. Maruish, Hrsg., Lawrence Erlbaum Associates, Inc, 1994, S. 292–321.
- [68] N. Bidargaddi, A. Sarela und I. Korhonen, „Physiological state characterization by clustering heart rate, heart rate variability and movement activity information,“ in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Bd. 2008, IEEE, Aug. 2008, S. 1749–1752.
- [69] L. Clifton, D. A. Clifton, M. A. Pimentel, P. J. Watkinson und L. Tarassenko, „Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors,“ *IEEE Journal of Biomedical and Health Informatics*, Jg. 18, Nr. 3, S. 722–730, 2014.
- [70] T.-W. Ho, C.-W. Huang, C.-M. Lin, F. Lai, J.-J. Ding, Y.-L. Ho und C.-S. Hung, „A Telesurveillance System With Automatic Electrocardiogram Interpretation Based on Support Vector Machine and Rule-Based Processing,“ *JMIR Medical Informatics*, Jg. 3, Nr. 2, Mai 2015.

- [71] W. H. Day und H. Edelsbrunner, „Efficient algorithms for agglomerative hierarchical clustering methods,“ *Journal of Classification*, Jg. 1, Nr. 1, S. 7–24, 1984.
- [72] D. Arthur und S. Vassilvitskii, „k-means++: The advantages of careful seeding,“ in *Proceedings of the 18. Annual ACM-SIAM Symposium on Discrete Algorithms*, Society for Industrial und Applied Mathematics, 2007, S. 1027–1035.
- [73] A. K. Jain, M. N. Murty und P. J. Flynn, „Data clustering: a review,“ *ACM computing surveys (CSUR)*, Jg. 31, Nr. 3, S. 264–323, 1999.
- [74] F. Murtagh und P. Contreras, „Algorithms for hierarchical clustering: an overview,“ *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Jg. 2, Nr. 1, S. 86–97, 2012.
- [75] P. C. Mahalanobis, „On the generalised distance in statistics,“ in *Proceedings National Institute of Science, India*, Bd. 2, Apr. 1936, S. 49–55.
- [76] T. Hofmann, B. Schölkopf und A. J. Smola, „Kernel methods in machine learning,“ *The Annals of Statistics*, S. 1171–1220, 2008.
- [77] B. Cabrer, D. Contreras und E. Miravete, „Aggregation in input–output tables: how to select the best cluster linkage,“ *Economic Systems Research*, Jg. 3, Nr. 2, S. 99–110, 1991.
- [78] M. Wiedenbeck, C. Züll u. a., „Klassifikation mit Clusteranalyse: Grundlegende Techniken hierarchischer und k-means-Verfahren,“ *ZUMA how-to-Reihe*, Jg. 10, Nr. 2, S. 18, 2001.
- [79] J. H. Ward Jr, „Hierarchical grouping to optimize an objective function,“ *Journal of the American Statistical Association*, Jg. 58, Nr. 301, S. 236–244, 1963.
- [80] D. Wishart, „256. Note: An algorithm for hierarchical classifications,“ *Biometrics*, S. 165–170, 1969.
- [81] W. M. Rand, „Objective criteria for the evaluation of clustering methods,“ *Journal of the American Statistical Association*, Jg. 66, Nr. 336, S. 846–850, 1971.
- [82] M. Levandowsky und D. Winter, „Distance between sets,“ *Nature*, Jg. 234, Nr. 5323, S. 34, 1971.
- [83] P. J. Rousseeuw, „Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,“ *Journal of Computational and Applied Mathematics*, Jg. 20, S. 53–65, 1987.



- [84] J. C. Dunn, „A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters,“ *Journal of Cybernetics*, Jg. 3, 3 1973.
- [85] T. Caliński und J. Harabasz, „A dendrite method for cluster analysis,“ *Communications in Statistics – Theory and Methods*, Jg. 3, Nr. 1, S. 1–27, 1974.
- [86] Y. Şenbabaoğlu, G. Michailidis und J. Z. Li, „Critical limitations of consensus clustering in class discovery,“ *Scientific Reports*, Jg. 4, S. 6207, 2014.
- [87] S. Monti, P. Tamayo, J. Mesirov und T. Golub, „Consensus clustering: A resampling based method for class discovery and visualization of gene expression microarray data,“ *Machine Learning*, Jg. 52, Nr. 1, S. 91–118, 2003.
- [88] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez und I. Perona, „An extensive comparative study of cluster validity indices,“ *Pattern Recognition*, Jg. 46, Nr. 1, S. 243–256, 2013.
- [89] R. Gutkin, C. Green, S. Vangrattanachai, S. Pinho, P. Robinson und P. Curtis, „On acoustic emission for failure investigation in CFRP: Pattern recognition and peak frequency analyses,“ *Mechanical Systems and Signal Processing*, Jg. 25, Nr. 4, S. 1393–1407, 2011.
- [90] L. Lovmar, A. Ahlford, M. Jonsson und A.-C. Syvänen, „Silhouette scores for assessment of SNP genotype clusters,“ *BMC Genomics*, Jg. 6, Nr. 1, S. 35, 2005.
- [91] L. Li, S. V. Lomov, X. Yan und V. Carvelli, „Cluster analysis of acoustic emission signals for 2D and 3D woven glass/epoxy composites,“ *Composite Structures*, Jg. 116, S. 286–299, 2014.
- [92] F. Zhang, S. Chen, H. Zhang, X. Zhang und G. Li, „Bioelectric signal detrending using smoothness prior approach,“ *Medical Engineering & Physics*, Jg. 36, Nr. 8, S. 1007–1013, 2014.
- [93] J. Pan und W. J. Tompkins, „A real-time QRS detection algorithm,“ *IEEE Transactions Biomedical Engineering*, Jg. 32, Nr. 3, S. 230–236, 1985.
- [94] Cuiwei Li, Chongxun Zheng und Changfeng Tai, „Detection of ECG characteristic points using wavelet transforms,“ *IEEE Transactions on Biomedical Engineering*, Jg. 42, Nr. 1, S. 21–28, Okt. 1995.
- [95] H. Chatterjee, R. Gupta und M. Mitra, „Real Time P and T Wave Detection from Ecg using Fpga,“ *Proceedings of the 2nd International Conference on Computer, Communication, Control and Information Technology*, Jg. 4, S. 840–844, 2012.

- [96] T. Hijzen und J. L. Slangen, „The electrocardiogram during emotional and physical stress.,“ *International Journal of Psychophysiology*, Jg. 2, Nr. 4, S. 273–279, 1985.
- [97] F. Schrumpf, G. Bausch, M. Sturm und M. Fuchs, „Similarity based hierarchical clustering of physiological parameters for the identification of health states-a feasibility study,“ in *Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2017, S. 458–462.
- [98] D. L. Donoho, „De-noising by soft-thresholding,“ *IEEE Transactions on Information Theory*, Jg. 41, Nr. 3, S. 613–627, 1995.
- [99] P. Agante und J. M. De Sá, „ECG noise filtering using wavelets with soft-thresholding methods,“ in *Computers in Cardiology, 1999*, IEEE, 1999, S. 535–538.
- [100] G. U. Reddy, M. Muralidhar und S. Varadarajan, „ECG De-Noising using improved thresholding based on Wavelet transforms,“ *International Journal of Computer Science and Network Security*, Jg. 9, Nr. 9, S. 221–225, 2009.
- [101] M. Elgendi, „On the analysis of fingertip photoplethysmogram signals,“ *Current Cardiology Reviews*, Jg. 8, Nr. 1, S. 14–25, 2012.
- [102] J. Steffel und T. Lüscher, „Anatomie und Physiologie des Herz-Kreislauf-Systems,“ in *Herz-Kreislauf*, 2. Aufl. Springer-Verlag Berlin Heidelberg, 2014, S. 1–7.
- [103] B. M. Saykrs, „Analysis of heart rate variability,“ *Ergonomics*, Jg. 16, Nr. 1, S. 17–32, 1973.
- [104] W. Jänig, „Vegetatives Nervensystem,“ in *Neuro- und Sinnesphysiologie*, 5. Aufl. Springer Medizin Verlag Heidelberg, 2006, S. 132–172.
- [105] K. Hamunen, V. Kontinen, E. Hakala, P. Talke, M. Paloheimo und E. Kalso, „Effect of pain on autonomic nervous system indices derived from photoplethysmography in healthy volunteers,“ *British Journal of Anaesthesia*, Jg. 108, Nr. 5, S. 838–844, 2012.
- [106] H. A. Demaree, J. L. Robinson, D. E. Everhart und B. J. Schmeichel, „Resting RSA is associated with natural and self-regulated responses to negative emotional stimuli,“ *Brain and Cognition*, Jg. 56, Nr. 1, S. 14–23, 2004.
- [107] B. M. Appelhans und L. J. Luecken, „Heart rate variability and pain: associations of two interrelated homeostatic processes,“ *Biological Psychology*, Jg. 77, Nr. 2, S. 174–182, 2008.

- [108] M. Malik, A. J. Camm, J. T. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. E. Kleiger, F. Lombardi, A. Malliani, A. J. Moss, J. N. Rottman, G. Schmidt, P. J. Schwartz und D. H. Singer, „Heart rate variability. Standards of measurement, physiological interpretation, and clinical use,“ *European Heart Journal*, Jg. 17, Nr. 3, S. 354–381, 1996.
- [109] O. Ormerod, C. McGregor, D. Stone, C. Wisbey und M. Petch, „Arrhythmias after coronary bypass surgery,“ *Heart*, Jg. 51, Nr. 6, S. 618–621, 1984.
- [110] F. E. Grubbs u. a., „Sample criteria for testing outlying observations,“ *The Annals of Mathematical Statistics*, Jg. 21, Nr. 1, S. 27–58, 1950.
- [111] F. Shaffer und J. Ginsberg, „An overview of heart rate variability metrics and norms,“ *Frontiers in Public Health*, Jg. 5, S. 258, 2017.
- [112] M. Jeanne, R. Logier, J. De Jonckheere und B. Tavernier, „Validation of a graphic measurement of heart rate variability to assess analgesia/nociception balance during general anesthesia,“ in *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2009, S. 1840–1843.
- [113] M. Le Guen, M. Jeanne, K. Sievert, M. Al Moubarik, T. Chazot, P. Laloe, J. Dreyfus und M. Fischler, „The Analgesia Nociception Index: a pilot study to evaluation of a new pain parameter during labor,“ *International Journal of Obstetric Anesthesia*, Jg. 21, Nr. 2, S. 146–151, 2012.
- [114] T. Ledowski, W. Tiong, C. Lee, B. Wong, T. Fiori und N. Parker, „Analgesia nociception index: evaluation as a new parameter for acute postoperative pain,“ *British Journal of Anaesthesia*, Jg. 111, Nr. 4, S. 627–629, 2013.
- [115] G. Parati, G. Mancina, M. D. Rienzo, P. Castiglioni, J. A. Taylor und P. Studinger, „Cardiovascular variability is/is not an index of autonomic control of circulation,“ *Journal of Applied Physiology*, Jg. 101, Nr. 2, S. 690–691, 2006.
- [116] J. T. VanderPlas, „Understanding the Lomb–Scargle Periodogram,“ *The Astrophysical Journal Supplement Series*, Jg. 236, Nr. 1, S. 16, 2018.
- [117] G. D. Clifford und L. Tarassenko, „Quantifying errors in spectral estimates of HRV due to beat replacement and resampling,“ *IEEE Transactions on Biomedical Engineering*, Jg. 52, Nr. 4, S. 630–638, 2005.
- [118] G. B. Moody und R. G. Mark, „Derivation of respiratory signals from multi-lead ECGs,“ *Computers in Cardiology*, Jg. 12, S. 113–116, 1985.

- [119] J. A. Hirsch und B. Bishop, „Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate,“ *American Journal of Physiology-Heart and Circulatory Physiology*, Jg. 241, Nr. 4, H620–H629, Okt. 1981.
- [120] P. Hamilton und W. Tompkins, „Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database,“ *IEEE Transactions on Biomedical Engineering*, Jg. 33, Nr. 12, S. 1157–1165, 1986.
- [121] F. Schrupf, M. Sturm, G. Bausch und M. Fuchs, „Derivation of the respiratory rate from directly and indirectly measured respiratory signals using autocorrelation,“ *Current Directions in Biomedical Engineering*, Jg. 2, Nr. 1, S. 241–245, 2016.
- [122] W. Karlen, C. Mattiussi und D. Floreano, „Sleep and wake classification with ECG and respiratory effort signals,“ *IEEE Transactions on Biomedical Circuits and Systems*, Jg. 3, Nr. 2, S. 71–78, 2009.
- [123] P. McLeod und G. Wyvill, „A smarter way to find pitch,“ in *Proceedings of International Computer Music Conference, ICMC*, 2005.
- [124] K. Pattinson, „Opioids and the control of respiration,“ *British Journal of Anaesthesia*, Jg. 100, Nr. 6, S. 747–758, 2008.
- [125] H. Gesche, D. Grosskurth, G. Kuchler und A. Patzak, „Continuous blood pressure measurement by using the pulse transit time: comparison to a cuff-based method,“ *European Journal of Applied Physiology*, Jg. 112, Nr. 1, S. 309–315, 2012.
- [126] A. Hennig, H. Gesche, I. Fietze, T. Penzel, M. Glos und A. Patzak, „Messung von apnoebezogenen Blutdruckänderungen mittels Pulstransitzeit und Penaz-Prinzip,“ *Atemwegs und Lungenkrankheiten*, Jg. 38, Nr. 11, S. 447, 2012.
- [127] D. B. McCombie, A. T. Reisner und H. H. Asada, „Adaptive blood pressure estimation from wearable PPG sensors using peripheral artery pulse wave velocity measurements and multi-channel blind identification of local arterial dynamics,“ in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2006, S. 3521–3524.
- [128] D. M. Biondi, J. Xiang, M. Etropolski und B. Moskovitz, „Evaluation of blood pressure and heart rate in patients with hypertension who received tapentadol extended release for chronic pain: a post hoc, pooled data analysis,“ *Clinical Drug Investigation*, Jg. 34, Nr. 8, S. 565–576, 2014.
- [129] A. Chen und M. A. Ashburn, „Cardiac effects of opioid therapy,“ *Pain Medicine*, Jg. 16, S. 27–31, 2015.

- [130] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster und U. Ehlert, „Discriminating stress from cognitive load using a wearable EDA device,“ *IEEE Transactions on information technology in biomedicine*, Jg. 14, Nr. 2, S. 410–417, 2010.
- [131] B. G. Wallin, „Sympathetic nerve activity underlying electrodermal and cardiovascular reactions in man,“ *Psychophysiology*, Jg. 18, Nr. 4, S. 470–476, 1981.
- [132] T. Ledowski, J. Bromilow, J. Wu, M. Paech, H. Storm und S. Schug, „The assessment of postoperative pain by monitoring skin conductance: results of a prospective study,“ *Anaesthesia*, Jg. 62, Nr. 10, S. 989–993, 2007.
- [133] T. Ledowski, J. Bromilow, M. Paech, H. Storm, R. Hacking und S. Schug, „Monitoring of skin conductance to assess postoperative pain intensity,“ *BJA: British Journal of Anaesthesia*, Jg. 97, Nr. 6, S. 862–865, 2006.
- [134] M. L. Chiswick, „Assessment of pain in neonates,“ *The Lancet*, Jg. 355, Nr. 9197, S. 6–8, 2000.
- [135] L. Lidberg und B. G. Wallin, „Sympathetic skin nerve discharges in relation to amplitude of skin resistance responses,“ *Psychophysiology*, Jg. 18, Nr. 3, S. 268–270, 1981.
- [136] W. Boucsein, *Elektrodermale Aktivität – Grundlagen, Methoden und Anwendungen*. Springer Verlag, 2013, S. 155–173.
- [137] W. W. Grings und A. M. Schell, „Magnitude of electrodermal response to a standard stimulus as a function of intensity and proximity of a prior stimulus,“ *Journal of Comparative and Physiological Psychology*, Jg. 67, Nr. 1, S. 77, 1969.
- [138] M. Benedek und C. Kaernbach, „A continuous measure of phasic electrodermal activity,“ *Journal of Neuroscience Methods*, Jg. 190, Nr. 1, S. 80–91, 2010.
- [139] M. Benedek und C. Kaernbach, „Decomposition of skin conductance data by means of nonnegative deconvolution,“ *Psychophysiology*, Jg. 47, Nr. 4, S. 647–658, 2010.
- [140] F. Wilcoxon, „Individual comparisons by ranking methods,“ *Biometrics Bulletin*, Jg. 1, Nr. 6, S. 80–83, 1945.
- [141] H. B. Mann und D. R. Whitney, „On a test of whether one of two random variables is stochastically larger than the other,“ *The Annals of Mathematical Statistics*, S. 50–60, 1947.
- [142] H. Abdi, „Bonferroni and Šidák corrections for multiple comparisons,“ *Encyclopedia of Measurement and Statistics*, Jg. 3, S. 103–107, 2007.

- [143] B. J. Drew, M. M. Pelter, D. E. Brodnick, A. V. Yadav, D. Dempel und M. G. Adams, „Comparison of a new reduced lead set ECG with the standard ECG for diagnosing cardiac arrhythmias and myocardial ischemia,“ *Journal of Electrocardiology*, Jg. 35, Nr. 4, S. 13–21, 2002.
- [144] M. K. Das, C. Saha, H. El Masry, J. Peng, G. Dandamudi, J. Mahenthiran, P. McHenry und D. P. Zipes, „Fragmented QRS on a 12-lead ECG: a predictor of mortality and cardiac events in patients with coronary artery disease,“ *Heart Rhythm*, Jg. 4, Nr. 11, S. 1385–1392, 2007.
- [145] P. S. Chawla und M. S. Kochar, „Effect of pain and nonsteroidal analgesics on blood pressure.,“ *WMJ: official publication of the State Medical Society of Wisconsin*, Jg. 98, Nr. 6, S. 22–5, 1999.
- [146] G. Untergehrer, D. Jordan, S. Eyl und G. Schneider, „Effects of propofol, sevoflurane, remifentanyl, and (s)-ketamine in subanesthetic concentrations on visceral and somatosensory pain-evoked potentials,“ *Anesthesiology: The Journal of the American Society of Anesthesiologists*, Jg. 118, Nr. 2, S. 308–317, 2013.
- [147] M. C. Lee, A. Mouraux und G. D. Iannetti, „Characterizing the cortical activity through which pain emerges from nociception,“ *Journal of Neuroscience*, Jg. 29, Nr. 24, S. 7909–7916, 2009.
- [148] N. M. Punjabi, „The epidemiology of adult obstructive sleep apnea,“ *Proceedings of the American Thoracic Society*, Jg. 5, Nr. 2, S. 136–143, 2008.
- [149] T. Young, J. Skatrud und P. E. Peppard, „Risk factors for obstructive sleep apnea in adults,“ *Jama*, Jg. 291, Nr. 16, S. 2013–2016, 2004.
- [150] V. K. Somers, M. E. Dyken, M. P. Clary und F. M. Abboud, „Sympathetic neural mechanisms in obstructive sleep apnea.,“ *The Journal of Clinical Investigation*, Jg. 96, Nr. 4, S. 1897–1904, 1995.
- [151] L. J. Findley, M. E. Unverzagt und P. M. Suratt, „Automobile Accidents Involving Patients with Obstructive Sleep Apnea,“ *American Review Respiratory Disease*, Jg. 138, S. 337–340, 1988.
- [152] D. W. Beebe und D. Gozal, „Obstructive sleep apnea and the prefrontal cortex: towards a comprehensive model linking nocturnal upper airway obstruction to daytime cognitive and behavioral deficits,“ *Journal of Sleep Research*, Jg. 11, Nr. 1, S. 1–16, 2002.

- [153] E. Shahar, C. W. Whitney, S. Redline, E. T. Lee, A. B. Newman, F. Javier Nieto, G. T. O'connor, L. L. Boland, J. E. Schwartz und J. M. Samet, „Sleep-disordered breathing and cardiovascular disease: cross-sectional results of the Sleep Heart Health Study,“ *American Journal of Respiratory and Critical Care Medicine*, Jg. 163, Nr. 1, S. 19–25, 2001.
- [154] W. W. Flemons, D. Buysse, S. Redline, K. Pack, K. Strohl, J. Wheatley, T. Young, N. Douglas, P. Levy, W. McNicholas, J. Fleetham, D. White, W. Schmidt-Nowarra, D. Carley und J. Romaniuk, „Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research,“ *Sleep*, Jg. 22, S. 667–689, 1999.
- [155] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger und J. H. Peter, „The apnea-ECG database,“ in *Computers in cardiology 2000*, IEEE, 2000, S. 255–258.
- [156] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng und H. E. Stanley, „PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals,“ *Circulation*, Jg. 101, Nr. 23, e215–e220, 2000.
- [157] C. Zywiets, V. Von Einem, B. Widiger und G. Joseph, „ECG analysis for sleep apnea detection,“ *Methods of Information in Medicine*, Jg. 43, Nr. 01, S. 56–59, 2004.
- [158] M. Jarvis und P. Mitra, „Apnea patients characterized by 0.02 Hz peak in the multitaper spectrogram of electrocardiogram signals,“ in *Computers in Cardiology 2000*, IEEE, 2000, S. 769–772.
- [159] J. E. Mietus, C.-K. Peng, P. C. Ivanov und A. L. Goldberger, „Detection of obstructive sleep apnea from cardiac interbeat interval time series,“ in *Computers in Cardiology 2000*, IEEE, 2000, S. 753–756.
- [160] Z. Shinar, A. Baharav und S. Akselrod, „Obstructive sleep apnea detection based on electrocardiogram analysis,“ in *Computers in Cardiology 2000*, IEEE, 2000, S. 757–760.
- [161] S. Boriah, V. Chandola und V. Kumar, „Similarity measures for categorical data: A comparative evaluation,“ in *Proceedings of the 2008 SIAM International Conference on Data Mining*, SIAM, 2008, S. 243–254.
- [162] A. Ziberna, N. Kejzar und P. Golob, „A comparison of different approaches to hierarchical clustering of ordinal data,“ *Metodoloski zvezki*, Jg. 1, Nr. 1, S. 57, 2004.

- [163] M. Giordan und G. Diana, „A clustering method for categorical ordinal data,“ *Communications in Statistics—Theory and Methods*, Jg. 40, Nr. 7, S. 1315–1334, 2011.
- [164] O. M. San, V.-N. Huynh und Y. Nakamori, „An alternative extension of the k-means algorithm for clustering categorical data,“ *International Journal of Applied Mathematics and Computer Science*, Jg. 14, S. 241–247, 2004.
- [165] Z. Huang, „Extensions to the k-means algorithm for clustering large data sets with categorical values,“ *Data Mining and Knowledge Discovery*, Jg. 2, Nr. 3, S. 283–304, 1998.
- [166] P. A. Gagniuc, *Markov Chains: From Theory to Implementation and Experimentation*. John Wiley & Sons, 2017.
- [167] P. Barton, S. Bryan und S. Robinson, „Modelling in the economic evaluation of health care: selecting the appropriate approach,“ *Journal of Health Services Research & Policy*, Jg. 9, Nr. 2, S. 110–118, 2004.
- [168] M. Fackrell, „Modelling healthcare systems with phase-type distributions,“ *Health Care Management Science*, Jg. 12, Nr. 1, S. 11, 2009.
- [169] Á. Rebuge und D. R. Ferreira, „Business process analysis in healthcare environments: A methodology based on process mining,“ *Information Systems*, Jg. 37, Nr. 2, S. 99–116, 2012.
- [170] F. A. Sonnenberg und J. R. Beck, „Markov models in medical decision making: a practical guide,“ *Medical Decision Making*, Jg. 13, Nr. 4, S. 322–338, 1993.
- [171] A. Ades und S. Cliffe, „Markov chain Monte Carlo estimation of a multiparameter decision model: consistency of evidence and the accurate assessment of uncertainty,“ *Medical Decision Making*, Jg. 22, Nr. 4, S. 359–371, 2002.
- [172] J. A. Cruz und D. S. Wishart, „Applications of machine learning in cancer prediction and prognosis,“ *Cancer Informatics*, Jg. 2, S. 117 693 510 600 200 030, 2006.
- [173] H. B. Burke, P. H. Goodman, D. B. Rosen, D. E. Henson, J. N. Weinstein, F. E. Harrell Jr, J. R. Marks, D. P. Winchester und D. G. Bostwick, „Artificial neural networks improve the accuracy of cancer survival prediction,“ *Cancer*, Jg. 79, Nr. 4, S. 857–862, 1997.
- [174] B. Zupan, J. Demšar, M. W. Kattan, J. R. Beck und I. Bratko, „Machine learning for survival analysis: a case study on recurrence of prostate cancer,“ *Artificial Intelligence in Medicine*, Jg. 20, Nr. 1, S. 59–75, 2000.



- [175] A. R. Brannon, A. Reddy, M. Seiler, A. Arreola, D. T. Moore, R. S. Pruthi, E. M. Wallen, M. E. Nielsen, H. Liu, K. L. Nathanson u. a., „Molecular stratification of clear cell renal cell carcinoma by consensus clustering reveals distinct subtypes and survival patterns,“ *Genes & Cancer*, Jg. 1, Nr. 2, S. 152–163, 2010.
- [176] R. G. Verhaak, K. A. Hoadley, E. Purdom, V. Wang, Y. Qi, M. D. Wilkerson, C. R. Miller, L. Ding, T. Golub, J. P. Mesirov u. a., „Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1,“ *Cancer Cell*, Jg. 17, Nr. 1, S. 98–110, 2010.
- [177] X. Xing und M. Sun, „Optical blood pressure estimation with photoplethysmography and FFT-based neural networks,“ *Biomedical optics express*, Jg. 7, Nr. 8, S. 3007–3020, 2016.
- [178] M. FUCHS, P. WAGNER, G. BAUSCH und P. FRENZEL, „Kamerabasierte Erfassung von Skelettdaten und Vitalparametern,“ *Technologien im Leistungssport 3: Tagungsband zur 19. Frühjahrsschule am 14./15. Mai 2018 in Leipzig*, S. 62, 2018.
- [179] C. Zhang und Y. Tian, „RGB-D camera-based daily living activity recognition,“ *Journal of Computer Vision and Image Processing*, Jg. 2, Nr. 4, S. 12, 2012.
- [180] S. Fleck und W. Straßer, „Smart camera based monitoring system and its application to assisted living,“ *Proceedings of the IEEE*, Jg. 96, Nr. 10, S. 1698–1714, 2008.
- [181] R. Higginson, „Respiratory assessment in critically ill Patients: airway and breathing,“ *British Journal of Nursing*, Jg. 18, Nr. 8, S. 456–461, Apr. 2009.
- [182] C. Lieber und V. Mohsenin, „Cheyne-Stokes respiration in congestive heart failure,“ *The Yale Journal of Biology and Medicine*, Jg. 65, Nr. 1, S. 39–50, 1992.
- [183] A. K. Gupta, „Respiration Rate Measurement Based on Impedance Pneumography,“ Texas Instruments, Techn. Ber., 2011, S. 1–11.
- [184] A. Grenvik, S. Ballou, E. McGinley, J. E. Millen, W. L. Cooley und P. Safar, „Impedance pneumography: Comparison between chest impedance changes and respiratory volumes in 11 healthy volunteers,“ *Chest Journal*, Jg. 62, Nr. 4, S. 439–443, 1972.
- [185] P. a. Leonard, J. G. Douglas, N. R. Grubb, D. Clifton, P. S. Addison und J. N. Watson, „A fully automated algorithm for the determination of respiratory rate from the photoplethysmogram,“ *Journal of Clinical Monitoring and Computing*, Jg. 20, Nr. 1, S. 33–6, 2006.

- [186] L. Sohrt-Petersen, „Evaluation of algorithms for ECG derived respiration in the context of heart rate variability studies,“ Master Thesis, Aalborg University, 2013.
- [187] C. Orphanidou, S. Fleming, S. A. Shah und L. Tarassenko, „Data fusion for estimating respiratory rate from a single-lead ECG,“ *Biomedical Signal Processing and Control*, Jg. 8, Nr. 1, S. 98–105, 2013.

## **Anhang A**

### **Evaluierung der EKG-basierten Atemfrequenzbestimmung**

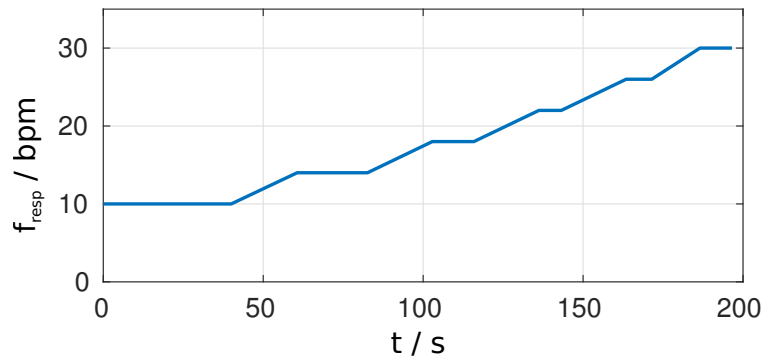
#### **A.1 Einleitung**

Die Atemfrequenz spielt eine wichtige Rolle bei der Einschätzung des Zustandes eines Patienten [181]. Eine abnormale Atemfrequenz oder abnormales Atemverhalten ist oft typisch für verschiedene Arten von Herzerkrankungen [182]. Die am häufigsten angewendete Methode zur Messung der Atmung ist die Messung mittels Brustgurt oder Flow-Sensor im Bereich von Mund und Nase [183, 184]. Diese Art der Messung ist jedoch mit erheblichem messtechnischen Aufwand verbunden. Weiterhin schränkt sie den Patienten durch die zusätzlichen Sensoren stark ein. Neben diesen direkten Messverfahren existieren noch weitere Verfahren, die die Atemfrequenz auf indirekte Weise ermitteln. Diese basieren meist auf einer geeigneten Auswertung des PPG oder EKG [118, 185]. Da diese Biosignale im klinischen Umfeld ohnehin gemessen werden, bietet es sich an sie zur Schätzung der Atemfrequenz zu verwenden.

Im folgenden Kapitel soll die in Kapitel 4 verwendete Methode zur indirekten Ableitung der Atemfrequenz evaluiert werden. Zu diesem Zweck wird ein Experiment durchgeführt, bei dem den Teilnehmern über einen festgelegten Zeitraum hinweg fest definierte Atemfrequenzen vorgegeben werden. Anhand des parallel dazu aufgenommenen EKG soll die Atemfrequenz mittels der neu vorgestellten Autokorrelationsmethode abgeleitet werden. Die gewonnenen Ergebnisse sollen mit den Resultaten etablierter Verfahren aus dem Zeit- und Frequenzbereich verglichen werden.

#### **A.2 Methode**

Zur Validierung und Evaluierung der Methode wurde eine Studie durchgeführt, bei der elf Probanden aus den Studenten und Mitarbeitern der Arbeitsgruppe rekrutiert wurden. Mit Hilfe eines Metronoms wurde den Teilnehmern eine Atemfrequenz vorgegeben, an die sich diese während des Versuchs anpassen sollten. Die vorgegebene Atemfrequenz wurde zwischen 10 und 30 Atemzügen pro Minute (Beats per Minute (bpm)) systematisch variiert. Dieser Ablauf ist in Abbildung A.1 dargestellt. Die Probanden führten diesen Versuch



**Abbildung A.1:** Variation der durch das Metronom vorgegebenen Atemfrequenz während des Versuchs. Die Atemfrequenz wurde ausgehend von 10 bpm in fünf Schritten auf 30 bpm angehoben.

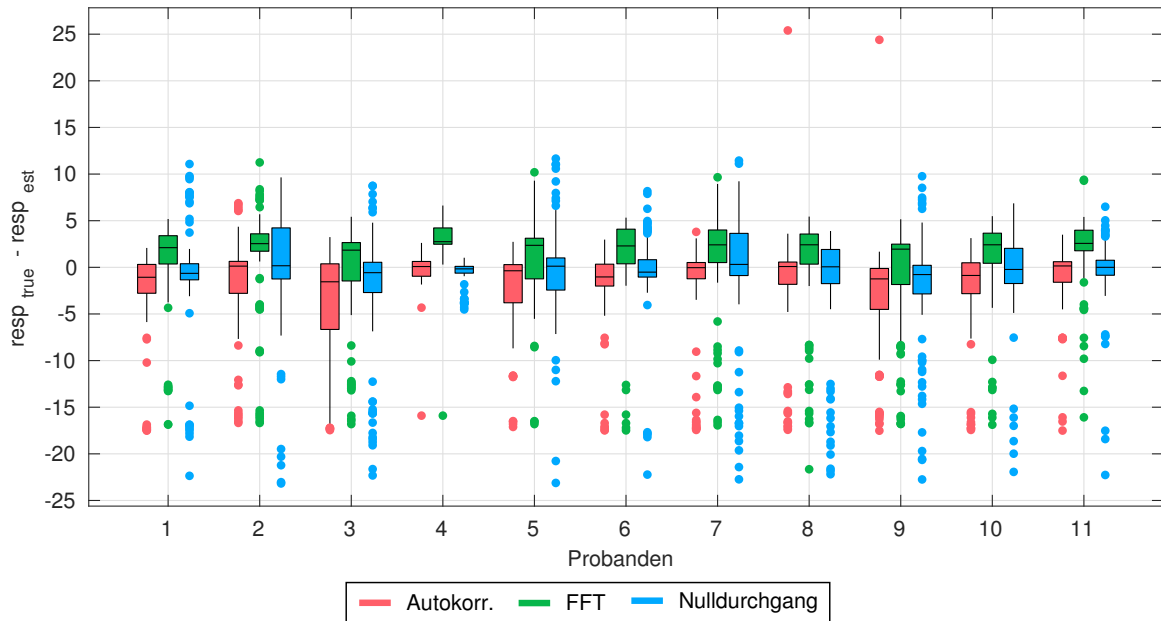
liegend aus. Während des Versuchs wurde sowohl das PPG mittels Finger-Clip am rechten Zeigefinger des Probanden als auch das EKG mit einer Abtastfrequenz von  $f_s = 1000\text{Hz}$  aufgezeichnet. Hierfür wurde das mehrkanalige Messsystem „biosignals Researcher“<sup>1</sup> verwendet, welches die Messdaten in einem internen Speicher zur späteren Auslesung über Bluetooth ablegt. Die ausgelesenen Daten wurden in Messfenster unterschiedlicher Länge geteilt und analysiert.

Um die Ergebnisse der Autokorrelationsmethode einordnen zu können, werden zwei weitere Verfahren hinzugezogen [186, 187]. Bei der Fast Fourier Transformation (FFT)-Methode wird zunächst das Spektrum des Atemsignals berechnet. Es wird dann die Frequenzkomponente  $f_{max}$  mit der stärksten Amplitude in einem Bereich von 0,1 Hz bis 0,6 Hz als Atemfrequenz interpretiert. Diese Frequenz wird mit Gleichung A.1 in die Atemfrequenz (bpm) umgerechnet.

$$f_{resp} = \frac{60}{T_{resp}} \quad (\text{A.1})$$

Weiterhin soll die Atemfrequenz mit Hilfe der Nulldurchgangsmethode geschätzt werden. Es wird zunächst der Gleichanteil des Atemsignals durch Subtraktion des Mittelwertes eliminiert. Dann werden die Abstände  $t_i$  aller Nulldurchgänge detektiert. Die Atemfrequenz nach der Nulldurchgangsmethode berechnet sich anhand der mittleren Abstände aller  $N$  Nulldurchgänge  $t_i$  mit Gleichung A.2.

$$f_{resp} = \frac{60}{\sum_{i=1}^N t_i} \quad (\text{A.2})$$



**Abbildung A.2:** Probandenspezifische Fehlerverteilungen bei der Extraktion der Atemfrequenz aus dem EKG.

### A.3 Ergebnisse

Abbildung A.2 zeigt die Ergebnisse der Extraktion der Atemfrequenz aus dem EKG mittels der neu vorgestellten Autokorrelationsmethode und den beiden Vergleichsverfahren. Dargestellt ist die Abweichung der geschätzten Atemfrequenz von der vorgegebenen wahren Atemfrequenz. Es ist sofort ersichtlich, dass die FFT-Methode die tatsächliche Atemfrequenz bei allen Probanden unterschätzt. Dies liegt in der Tatsache begründet, dass dieses Verfahren aufgrund der FFT eine Frequenzauflösung besitzt, die von der Länge des gewählten Betrachtungszeitraums abhängt. Ist dieses zu klein gewählt, wird die Frequenzauflösung zu grob und somit eine korrekte Schätzung der Atemfrequenz erschwert.

Die Nulldurchgangsmethode und die vorgeschlagene Autokorrelationsmethode schneiden bei den meisten Probanden ähnlich gut ab. Wie aus Tabelle A.1 zeigen beide Methoden einen ähnlichen mittleren Fehler. Bei einigen Probanden in Abbildung A.2 zeigt die Nulldurchgangsmethode jedoch eine größere Varianz des Fehlers bei der Atemfrequenzbestimmung. Auch bei der Betrachtung über alle Probanden hinweg weist die Nulldurchgangsmethode zwar eine etwas geringeren Fehler-Median auf als die Autokorrelationsmethode, zeigt aber die größte Fehlervarianz aller verglichenen Methoden.

<sup>1</sup>PLUX wireless biosignals S.A. [www.plux.info](http://www.plux.info)

	$\bar{x}_{err}$ / bpm	$\sigma_{err}^2$ / bpm
Autokorrelation	-0,62	27,71
FFT	2,41	28,16
Nulldurchgang	-0,12	35,09

**Tabelle A.1:** Median und Varianz der Fehler der betrachteten Methoden für alle Probanden.

Es wurde ein Mann-Whitney-U-Test auf Gleichheit der Fehlerverteilungen der Autokorrelationsmethode und den beiden Vergleichsmethoden durchgeführt. Es zeigt sich hierbei ein signifikanter Unterschied ( $p < 0.01$ ).

#### A.4 Diskussion

Es wurde ein autokorrelationsbasiertes Verfahren zur Schätzung der Atemfrequenz anhand von aus EKG-Daten gewonnen Atemsignalen vorgestellt. Zur Evaluierung dieses Verfahrens wurde eine Messstudie durchgeführt, bei der die Probanden ihre Atmung an eine vorgegebene Frequenz anpassen mussten, welche systematisch variiert wurde. Aus dem währenddessen aufgenommenen EKG wurde die Atemfrequenz mit Hilfe der Autokorrelationsmethode extrahiert. Diese Ergebnisse wurden mit den Ergebnissen von zwei Referenzmethoden (FFT-basierte Methode und eine Methode auf Basis der Nulldurchgangsanalyse) verglichen.

Dabei zeigten die Autokorrelationsmethode sowie die Nulldurchgangsanalyse den kleinsten Fehler-Median. Die FFT-Methode wird für den Zweck der Atemfrequenzschätzung als ungeeignet angesehen, da die Genauigkeit der Atemfrequenzbestimmung stark von der Frequenzauflösung der zugrundeliegenden FFT abhängt. Da der Frequenzbereich der Atemfrequenz sehr klein ist, kann eine hohe Genauigkeit bei der Atemfrequenzbestimmung bei kurzen Signalabschnitten mit geringer Abtastfrequenz nicht gewährleistet werden. Beim Vergleich von Autokorrelations- und FFT-Methode wies letztere die höchste Fehlervarianz auf.

Es konnte gezeigt werden, dass die Autokorrelationsmethode für die Extraktion der Atemfrequenz aus dem EKG geeignet ist. Die für die Validierung verwendeten Daten sind jedoch unter optimalen Bedingungen aufgenommen worden. Aufgrund der Tatsache, dass die Probanden den Versuch liegend durchgeführt haben, finden sich nur sehr wenige Artefakte im EKG-Signal. Weiterhin wurden Probanden zur Brustatmung instruiert um das aufmodulierte Atemsignal im EKG besonders hervorzuheben. Bei Datensätzen, die unter realistischeren Bedingungen im klinischen Umfeld zustande gekommen sind, können diese beiden Eigenschaften jedoch nicht zwingen vorausgesetzt werden. So können durch Bewegungsartefakte

und andere Störeinkopplungen höherfrequente Signalanteile im extrahierten Atemsignal auftreten, welche zu zusätzlichen Maxima in der Autokorrelationsfunktion führen. Durch die in Abschnitt 4.1.4 beschriebene schwellwertbasierte Auswahl der relevanten Maxima aus der Autokorrelationsfunktion konnte jedoch eine gewisse Robustheit gegenüber solchen Störungen erreicht werden. Um dies zu verifizieren müsste eine neue Studie unter weniger kontrollierten Bedingungen durchgeführt werden.





## Anhang B

### Tabellen und Abbildungen

Proband	Cluster-Anzahl K								
	2	3	4	5	6	7	8	9	10
ptt002	<b>0,78</b>	0,67	0,76	0,68	0,66	0,72	0,73	0,73	0,72
ptt003	0,81	0,61	<b>0,85</b>	0,79	0,77	0,79	0,79	0,76	0,75
ptt004	<b>0,85</b>	0,69	0,75	0,76	0,75	0,75	0,74	0,74	0,73
ptt005	<b>0,81</b>	0,78	0,79	0,79	0,75	0,74	0,73	0,68	0,64
ptt006	<b>0,87</b>	0,83	0,78	0,81	0,78	0,78	0,79	0,79	0,78
ptt007	<b>0,84</b>	0,82	0,76	0,72	0,73	0,76	0,75	0,67	0,67
ptt009	0,84	<b>0,85</b>	0,82	0,78	0,78	0,69	0,71	0,71	0,73
ptt011	<b>0,84</b>	0,75	0,75	0,83	0,81	0,81	0,79	0,81	0,8
ptt012	<b>0,89</b>	0,87	0,88	0,84	0,79	0,77	0,71	0,7	0,68
ptt013	0,68	0,61	0,65	<b>0,69</b>	0,66	0,62	0,64	0,67	0,65
ptt014	0,74	0,66	<b>0,78</b>	0,77	0,73	0,69	0,74	0,73	0,73
ptt015	<b>0,92</b>	0,83	0,86	0,86	0,86	0,68	0,58	0,57	0,56
ptt016	<b>0,82</b>	0,8	0,76	0,75	0,68	0,7	0,76	0,74	0,73
ptt017	0,77	<b>0,82</b>	0,81	0,72	0,68	0,66	0,75	0,73	0,72
ptt018	<b>0,87</b>	0,83	0,8	0,75	0,75	0,76	0,74	0,73	0,74
ptt020	<b>0,91</b>	0,81	0,88	0,85	0,85	0,76	0,74	0,7	0,71
ptt021	<b>0,9</b>	0,7	0,82	0,75	0,72	0,75	0,75	0,76	0,74

Proband	Cluster-Anzahl K								
	2	3	4	5	6	7	8	9	10
ptt023	<b>0,85</b>	0,83	0,77	0,78	0,72	0,72	0,7	0,7	0,64
ptt025	<b>0,87</b>	0,84	0,85	0,84	0,83	0,85	0,85	0,86	0,85
ptt026	0,76	<b>0,77</b>	0,69	0,7	0,66	0,63	0,71	0,71	0,7
ptt027	<b>0,91</b>	0,88	0,82	0,78	0,74	0,71	0,71	0,7	0,75
ptt028	<b>0,89</b>	0,81	0,79	0,81	0,79	0,82	0,81	0,79	0,84
ptt029	<b>0,86</b>	0,78	0,78	0,77	0,75	0,76	0,75	0,69	0,67
ptt031	<b>0,92</b>	0,89	0,88	0,86	0,85	0,82	0,78	0,7	0,71
ptt032	<b>0,94</b>	0,91	0,87	0,86	0,86	0,84	0,85	0,76	0,73
ptt033	<b>0,87</b>	0,85	0,81	0,83	0,83	0,81	0,81	0,84	0,84

**Tabelle B.1:** SI-Koeffizienten der Cluster-Lösungen ( $K = 2 \dots 10$ ) aller ausgewerteten Probanden. Der höchste SI-Koeffizient jedes Probanden ist fett dargestellt und repräsentiert die optimale Cluster-Lösung.

Proband	pNN20	$P_{VLF}/P_{HF}$	ANI	EDA	Resp.	BP
VD004	✓	✓	✓	✗	✓	✓
VD005	✓	✓	✓	✗	✓	✓
VD007	✓	✓	✓	✗	✓	✗
VD010	✓	✓	✓	✗	✓	✗
VD011	✓	✓	✓	✗	✓	✗
VD012	✓	✓	✓	✗	✓	✗
VD013	✓	✓	✓	✗	✓	✓
VD014	✓	✓	✓	✗	✓	✓
VD015	✓	✓	✓	✗	✗	✗
VD016	✓	✓	✓	✗	✓	✗
VD018	✓	✓	✓	✗	✓	✗
VD019	✓	✓	✓	✓	✓	✗
VD020	✓	✓	✓	✓	✓	✓
VD021	✓	✓	✓	✓	✓	✓
VD022	✓	✓	✓	✗	✓	✗
VD023	✓	✓	✓	✗	✓	✗
VD024	✓	✓	✓	✗	✗	✗
VD025	✓	✓	✓	✗	✓	✓

**Tabelle B.2:** Auflistung der zur Klassifikation von Gesundheitszuständen verwendeten Vitalparameter und Biosignale für jeden Probanden mit protokollierter Schmerzmittelgabe.

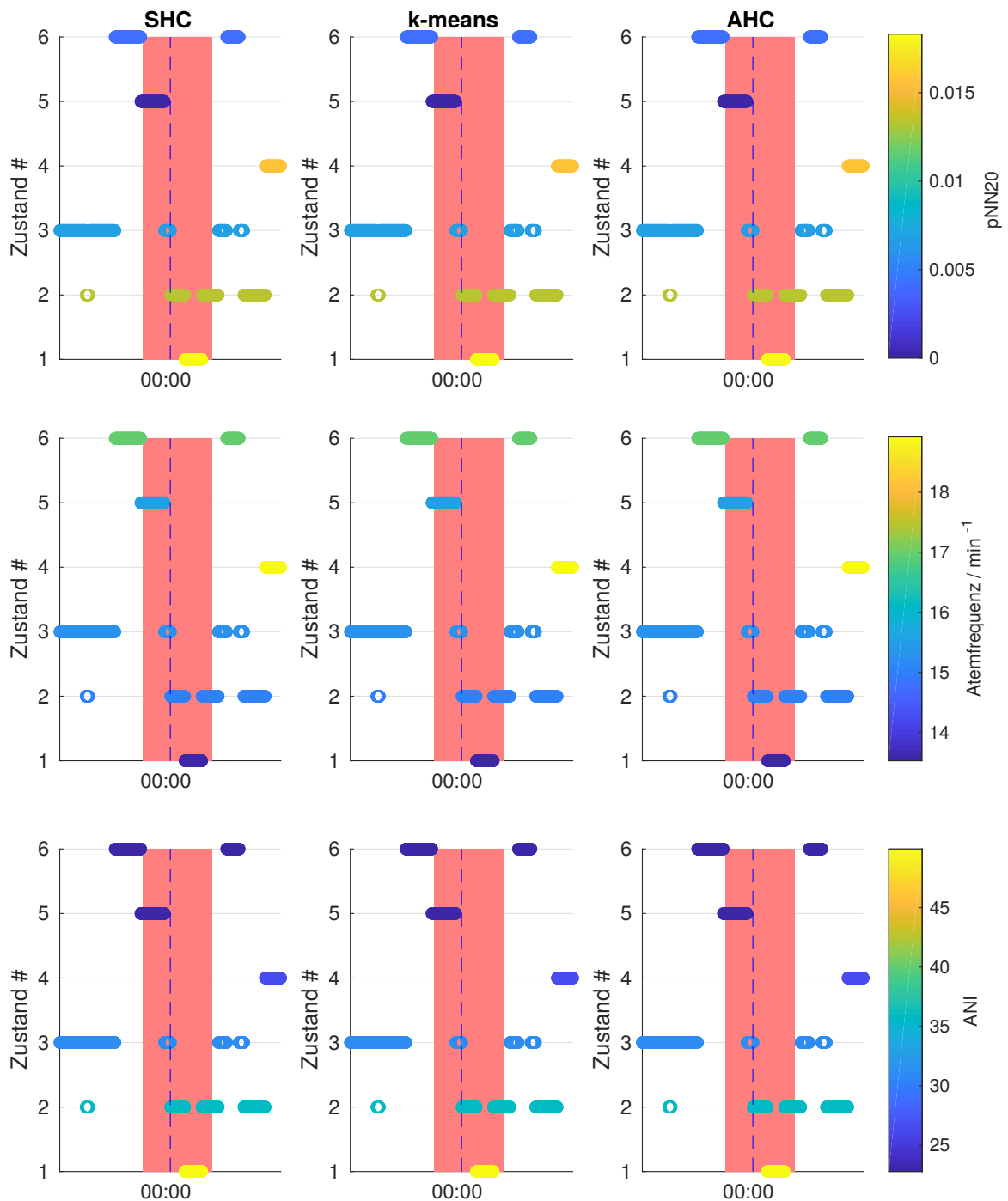
Proband	Uhrzeit	SHC			k-means			AHC		
		K	SI	Q	K	SI	Q	K	SI	Q
VD004	19:00	9	0,63	0,80	11	0,65	0,79	2	0,58	0,00
	10:00	11	0,65	0,60	5	0,65	0,73	5	0,61	0,63
VD005	11:00	4	0,64	0,68	11	0,66	0,65	2	0,59	0,00
	18:00	7	0,80	0,77	6	0,80	0,78	4	0,78	0,65
	22:00	4	0,86	0,79	5	0,87	0,84	2	0,86	0,86
VD007	06:00	11	0,62	0,62	9	0,68	0,59	4	0,67	0,68
	07:00	11	0,72	0,00	5	0,75	0,00	8	0,64	0,00
	08:00	6	0,82	0,77	6	0,83	0,83	7	0,71	0,74
	09:00	5	0,81	0,00	3	0,80	0,00	4	0,80	0,00
VD010	00:00	3	0,68	0,63	3	0,67	0,68	7	0,52	0,00
	06:00	2	0,89	0,00	3	0,87	0,81	2	0,89	0,00
VD011	00:00	6	0,64	0,86	9	0,69	0,70	9	0,62	0,75
	06:00	11	0,65	0,00	11	0,69	0,00	4	0,56	0,00
	18:00	2	0,72	0,00	4	0,69	0,71	9	0,62	0,63
	19:00	7	0,63	0,72	7	0,69	0,73	7	0,59	0,00
	20:00	4	0,73	0,68	4	0,74	0,68	4	0,72	0,71
VD012	06:00	5	0,69	0,59	3	0,70	0,00	2	0,64	0,00
	12:00	7	0,65	0,69	5	0,69	0,73	9	0,54	0,00
	18:00	12	0,73	0,71	11	0,76	0,74	12	0,64	0,69
	21:00	2	0,61	0,54	10	0,59	0,00	3	0,60	0,00
VD013	01:00	10	0,78	0,59	9	0,79	0,60	5	0,74	0,64
	06:00	5	0,71	0,78	6	0,71	0,77	5	0,68	0,77

Proband	Uhrzeit	SHC			k-means			AHC		
		K	SI	Q	K	SI	Q	K	SI	Q
VD013	20:00	7	0,77	0,74	9	0,79	0,72	4	0,67	0,00
	06:00	2	0,80	0,00	10	0,78	0,73	5	0,73	0,77
VD014	18:00	7	0,72	0,62	10	0,76	0,65	8	0,68	0,66
	23:00	9	0,66	0,81	10	0,70	0,81	10	0,58	0,00
VD015	12:00	5	0,62	0,00	4	0,69	0,00	8	0,58	0,00
	06:00	6	0,64	0,67	5	0,66	0,77	2	0,72	0,70
	00:00	10	0,55	0,00	4	0,60	0,64	3	0,47	0,00
	18:00	2	0,71	0,90	2	0,76	0,90	3	0,70	0,87
VD016	09:30	5	0,61	0,62	4	0,67	0,65	4	0,59	0,00
	19:00	4	0,67	0,89	3	0,69	0,88	3	0,52	0,00
VD018	05:00	5	0,68	0,78	3	0,68	0,80	5	0,57	0,00
	06:00	2	0,64	0,00	11	0,71	0,00	10	0,67	0,00
	01:00	3	0,58	0,00	10	0,65	0,68	2	0,63	0,00
	19:00	3	0,83	0,87	3	0,84	0,86	3	0,80	0,85
VD019	06:00	3	0,77	0,65	3	0,77	0,67	3	0,77	0,65
	11:00	10	0,66	0,74	11	0,71	0,76	10	0,65	0,73
	17:00	5	0,80	0,00	5	0,80	0,00	5	0,78	0,00
VD020	00:00	4	0,77	0,70	5	0,75	0,69	4	0,76	0,71
	12:00	3	0,79	0,65	3	0,82	0,64	3	0,79	0,59
	17:00	10	0,87	0,72	7	0,88	0,68	9	0,87	0,74
	23:00	4	0,77	0,00	4	0,78	0,00	4	0,76	0,00
VD021	02:00	11	0,62	0,54	11	0,70	0,59	2	0,62	0,00

Proband	Uhrzeit	SHC			k-means			AHC		
		K	SI	Q	K	SI	Q	K	SI	Q
VD021	06:00	3	0,68	0,00	2	0,66	0,00	3	0,66	0,00
	18:00	6	0,75	0,86	7	0,78	0,81	4	0,75	0,83
	22:00	2	0,67	0,81	4	0,64	0,00	2	0,61	0,00
	23:00	2	0,67	0,60	5	0,67	0,86	2	0,61	0,74
VD022	06:00	3	0,61	0,00	10	0,67	0,00	3	0,68	0,00
	07:00	3	0,61	0,00	3	0,69	0,62	3	0,68	0,62
	09:00	8	0,63	0,61	8	0,65	0,62	10	0,49	0,00
VD023	06:00	9	0,63	0,74	13	0,66	0,65	11	0,55	0,00
	22:00	5	0,62	0,72	4	0,64	0,78	2	0,47	0,00
VD024	10:00	3	0,66	0,69	3	0,67	0,69	3	0,64	0,66
	11:00	10	0,70	0,62	8	0,71	0,64	4	0,62	0,00
	12:00	5	0,72	0,77	6	0,76	0,80	4	0,71	0,78
	13:00	4	0,76	0,00	4	0,77	0,00	4	0,69	0,00
	14:00	11	0,62	0,63	2	0,65	0,64	2	0,62	0,63
	15:00	11	0,69	0,70	10	0,70	0,61	6	0,63	0,75
	20:00	3	0,76	0,00	9	0,76	0,65	3	0,76	0,00
	21:00	5	0,65	0,75	3	0,70	0,79	4	0,64	0,72
	22:00	5	0,65	0,56	3	0,70	0,57	4	0,64	0,58
	23:00	11	0,68	0,67	10	0,70	0,62	3	0,67	0,80
	00:00	10	0,69	0,66	9	0,72	0,68	3	0,61	0,00
	01:00	10	0,68	0,62	11	0,71	0,65	9	0,68	0,67
02:00	10	0,66	0,65	11	0,72	0,59	11	0,62	0,56	

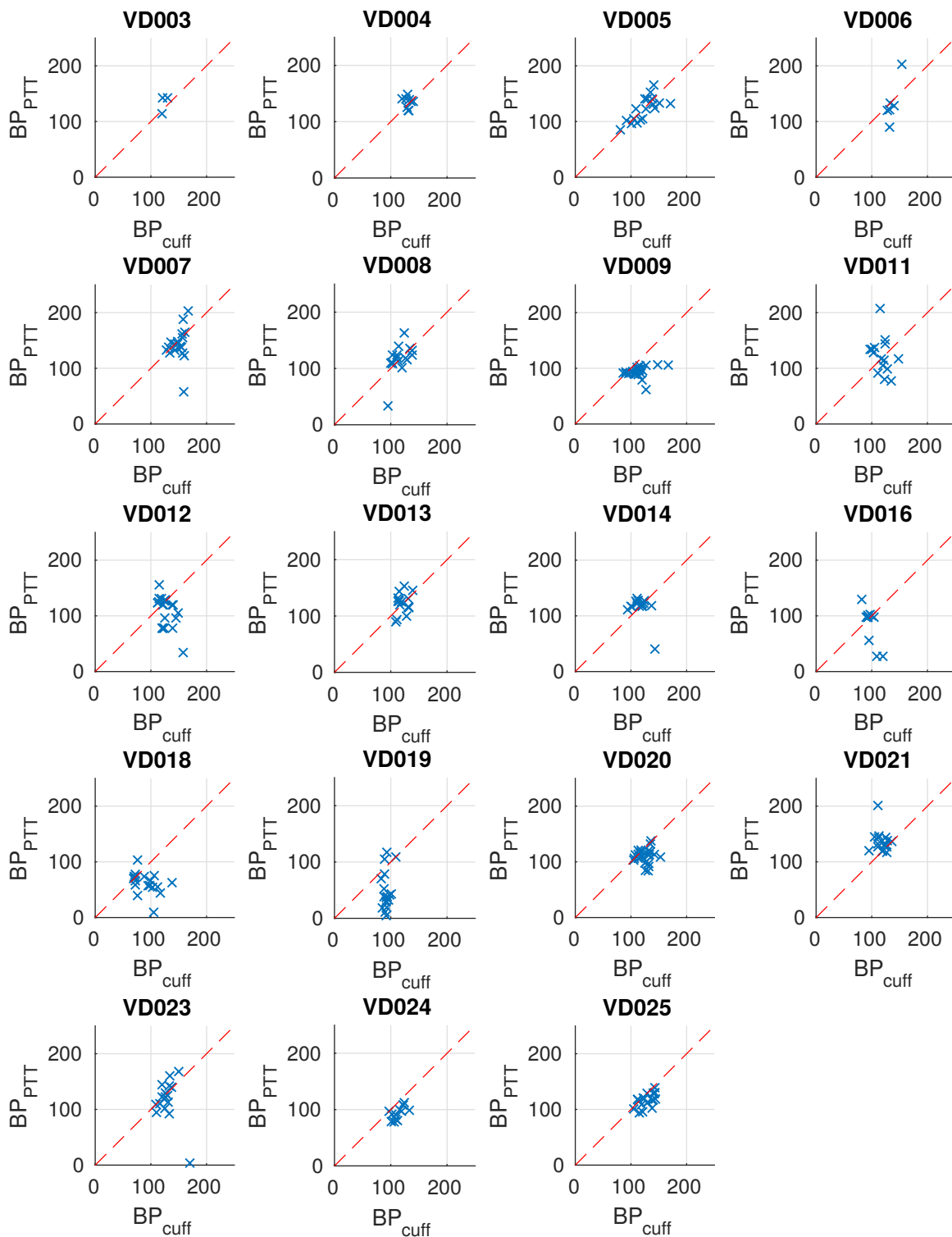
Proband	Uhrzeit	SHC			k-means			AHC		
		K	SI	Q	K	SI	Q	K	SI	Q
VD024	03:00	11	0,71	0,61	10	0,73	0,63	9	0,66	0,62
	04:00	11	0,64	0,57	11	0,67	0,59	11	0,56	0,00
	05:00	2	0,80	0,00	2	0,80	0,00	2	0,78	0,00
	06:00	2	0,72	0,50	2	0,66	0,50	2	0,73	0,51
VD025	02:00	10	0,69	0,79	6	0,69	0,84	7	0,65	0,83
	04:00	6	0,63	0,79	11	0,65	0,74	5	0,62	0,66
	23:00	2	0,68	0,90	3	0,67	0,88	3	0,54	0,00

**Tabelle B.3:** Ergebnisse der Cluster-Analyse zu Zeitpunkten von Medikamentengaben. Analysiert wurde jeweils ein Zeitraum zwischen zwei Stunden vor und nach einer Schmerzmittelgabe mit den Verfahren SHC, k-means und AHC. Abgebildet sind die Cluster-Lösungen mit der Cluster-Anzahl K, für die der SI-Koeffizient maximal wird. Weiterhin wird für jede gefundene Lösung der ermittelte Q-Parameter zur Einschätzung der Rekonstruktionsqualität der Schmerzmittelgabe aufgelistet.



**Abbildung B.1:** Verlauf der Zustandsübergänge aus Kapitel 4, welche von den Verfahren SHC sowie den Referenzverfahren detektiert wurden. Die Färbung repräsentiert die Mediane der Vitalparameter pNN20, Atemfrequenz und ANI in den erkannten Gesundheitszuständen.





**Abbildung B.2:** Vergleich zwischen manschettensbasierter Blutdruckmessung und PTT-basierter Blutdruckschätzung (systolischer Wert) für Probanden mit vollständigen EKG- und PPG-Signalen.



## **Eigenständigkeitserklärung**

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig, ohne Hilfe Dritter und nur unter Verwendung der angegebenen Hilfsmittel und Quellen angefertigt habe. Alle den benutzten Quellen wörtlich oder sinngemäß entnommenen Stellen sind als solche einzeln kenntlich gemacht.

Diese Arbeit ist bislang keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht worden.

Ort, Datum

Unterschrift

# Lebenslauf

## **Persönliche Daten**

Name: Fabian Schrupf  
Geburtsdatum: 23. August 1985  
Geburtsort: Gera

## **Schulbildung**

09/1997 – 06/2004 Georg-Samuel-Dörffel-Gymnasium Weida  
Abschluss: Abitur

## **Studium und wissenschaftliche Tätigkeit**

seit 08/2018 Wissenschaftlicher Mitarbeiter im Projekt „Kamerabasiertes System zur kontaktlosen Erfassung von Vitaldaten, Mimik und Körperbewegungen“ am Forschungs- und Transferzentrum Leipzig e.V.

01/2016 – 07/2020 Doktorand an der HTWK Leipzig und Universität Rostock

05/2015 – 07/2018 Wissenschaftlicher Mitarbeiter im Projekt „Nichtinvasive Erfassung und Analyse von Biosignalen zur unterstützenden Bewertung kritischer Vitalzustände“ an der HTWK Leipzig

10/2013 – 04/2014 Wissenschaftlicher Mitarbeiter an der HTWK Leipzig und am Forschungs- und Transferzentrum Leipzig e.V.

10/2011 – 01/2014 Studium der Elektro- und Informationstechnik an der HTWK Leipzig, Fachrichtung Kommunikationstechnik und Biosignalsverarbeitung, Abschluss: Master of Science

03/2009 – 03/2011 Studium der Elektro- und Informationstechnik an der HTWK Leipzig, Fachrichtung Kommunikationstechnik, Abschluss: Bachelor of Engineering

Leipzig, 22. Juli 2020