



FAU Studien aus der Informatik 9

Heike Leutheuser

Wearable computing applications in eHealth

Heike Leutheuser

Wearable computing applications in eHealth

FAU Studien aus der Informatik

Band 9

Herausgeber der Reihe:

Björn Eskofier, Richard Lenz, Andreas Maier,
Michael Philippsen, Lutz Schröder,
Wolfgang Schröder-Preikschat, Marc Stamminger, Rolf Wanka

Heike Leutheuser

Wearable computing applications in eHealth

Erlangen

FAU University Press

2019

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte bibliografische Daten
sind im Internet über <http://dnb.d-nb.de> abrufbar.

Das Werk, einschließlich seiner Teile, ist urheberrechtlich geschützt.
Die Rechte an allen Inhalten liegen bei ihren jeweiligen Autoren.
Sie sind nutzbar unter der Creative Commons Lizenz BY.

Der vollständige Inhalt des Buchs ist als PDF über den OPUS Server
der Friedrich-Alexander-Universität Erlangen-Nürnberg abrufbar:
<https://opus4.kobv.de/opus4-fau/home>

Bitte zitieren als
Leutheuser, Heike. 2019. *Wearable computing applications in eHealth*.
FAU Studien aus der Informatik Band 9.
Erlangen: FAU University Press. DOI: 10.25593/978-3-96147-261-1

Verlag und Auslieferung:
FAU University Press, Universitätsstraße 4, 91054 Erlangen

Druck: docupoint GmbH

ISBN: 978-3-96147-260-4 (Druckausgabe)
eISBN: 978-3-96147-261-1 (Online-Ausgabe)
ISSN: 2509-9981
DOI: 10.25593/978-3-96147-261-1

Wearable computing applications in eHealth

Wearable Computing Anwendungen im E-Health

Der Technischen Fakultät der
Friedrich-Alexander-Universität Erlangen-Nürnberg

zur

Erlangung des Doktorgrades Dr.-Ing.

vorgelegt von

Heike Leutheuser

aus Erlangen

Als Dissertation genehmigt
von der Technischen Fakultät der
Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung:	25.04.2019
Vorsitzender des Promotionsorgans:	Prof. Dr.-Ing. Reinhard Lerch
Gutachter:	Prof. Dr. Björn M. Eskofier
	Prof. Dr. Maria Teresa Arredondo Waldmeyer

Abstract

Non-communicable diseases are the leading cause of death and disability worldwide. Almost two thirds of them are linked to the following risk factors: physical inactivity, unhealthy diets, tobacco use, and harmful use of alcohol. At the same time, everyone is surrounded increasingly with wearables, like smartphones and smartwatches, making wearable computing an integral part of everyday life. The combination of addressing non-communicable diseases with the help of wearables seems one potential solution. In this thesis, methods and algorithms for three wearable computing applications in eHealth are presented: I) mobile breathing analysis, II) mobile electrocardiogram (ECG) analysis, and III) inertial measurement unit (IMU)-based activity recognition. In the following, these will be discussed individually, before a final paragraph will bring them together.

Respiratory inductance plethysmography (RIP) provides an unobtrusive and mobile method for measuring breathing characteristics, avoiding the measurement with flowmeters (FMs) as with them the natural breathing pattern is altered. The output of RIP devices needs to be adjusted to result in correct ventilatory tidal volumes. State of the art methods for adjusting RIP data can only be applied after the actual measurement as these require the simultaneous data acquisition of RIP and FM. In this thesis, novel adjustment algorithms were created enabling the usage of RIP solely, and by this outside the controlled laboratory environment. The algorithms were evaluated using 193 subjects performing an incremental treadmill running test. In contribution (I), adjustment was achieved by calculating volume motion coefficients (VMCs) from a standing still phase. These were applied afterwards to treadmill running and recovery phases. In contribution (II), generalizable and reference-free adjustment algorithms were investigated. These algorithms include anthropometric data of the subject, time characteristics, and ratios of the two RIP-bands. For practical usability, an adjustment algorithm without any prerecording

should be chosen. This spoke in favor for the reference-free adjustment algorithm with which also better results were obtained. Respiratory diseases and infections are amongst the most leading deaths group. In future work, it has to be investigated how RIP devices can effectively be used for patients suffering from these conditions.

The reasons for mobile ECG analysis were to provide real-time algorithms for two particular scopes. The first scope dealt with identifying arrhythmic beats using only time instances of successive heartbeats – the RR intervals (contribution (III)). In this thesis, ten consecutive RR intervals were used to calculate features for a Naive Bayes classifier. The features came from fitting a sine curve to the consecutive RR intervals. The small number of required RR intervals enabled a real-time and real-life application. In the second scope, three high-accurate single-lead, instantaneous P- and T-wave detection algorithms were compared (contribution (IV)). All three algorithms needed just lead II (according to Einthoven). The best algorithm of them was the fast naive wave event detection algorithm without any filtering. Comparison to algorithms from literature was hindered by the incomplete description of these algorithms and their evaluation metrics. These two applications could help in addressing cardiovascular diseases that are the number one cause of deaths worldwide. An effective and easy to handle arrhythmia classification algorithm, as presented in contribution (III), could send people in risk early to a physician. High-accurate P- and T-wave detection algorithms requiring only a single lead are beneficial for all ECG monitoring fields, and at the same time, enhance the patient's comfort.

IMU-based activity recognition provides an objective method for classifying activities addressing the risk factor physical inactivity. Therefore, a common, publicly available dataset DaLiAc was created to enable the comparison of activity recognition algorithms (contribution (V)). The benchmark dataset DaLiAc comprises data of 19 subjects performing 13 different daily life activities acquired with four IMUs placed at the chest, the right hip, the right wrist, and the left ankle. DaliAc is publicly available for download (<http://www.activitynet.org/>). Using the benchmark dataset DaLiAc, a hierarchical classification system was created that outperformed six state of the art activity recognition algorithms (contribution (VI)). In the hierarchical classification system, activities were grouped and ordered in a hierarchy. This resulted in advantages like the easy integration of new activities without retraining the complete system, and the broad range of options for each single classification system like different features, window sizes, or classifiers. So far, DaLiAc has been used by 13 other research groups for assessing the performance of their

algorithms. Recognizing single activities of daily living might increase the awareness of individuals to increase their physical activity as physical inactivity is one of the four leading risk factors for non-communicable diseases.

In this thesis, three different wearable computing applications addressing different non-communicable diseases were presented. The basis for each application was built on single, distinct biosignals. The fusion of the biosignals could improve the performance of the single application fields and even create new application areas. Such a fusion should happen at both, the algorithmic and acquisition level, like the combination into one principle such as a textile or a smartwatch. Wearables have been continuously used for health monitoring in recent years – with still increasing trend – as they bring certain benefits to the user in daily life. This will be enlarged in the future and fostered by ongoing inventions, accumulating knowledge, technological process, and progressive digitalization.

Zusammenfassung

Nichtübertragbare Krankheiten sind weltweit die Hauptursache für Tod und Invalidität. Fast zwei Drittel der nichtübertragbaren Krankheiten können mit den folgenden Risikofaktoren in Verbindung gebracht werden: körperliche Inaktivität, ungesunde Ernährung, Tabakkonsum und schädlicher Alkoholkonsum. Gleichzeitig ist jeder zunehmend mit Wearables wie Smartphones und Smartwatches umgeben, sodass Wearable Computing ein fester Bestandteil des Alltags ist. Nichtübertragbare Krankheiten mit Hilfe von Wearables anzugehen, scheint eine mögliche Lösung zu sein. In dieser Arbeit wurden Methoden und Algorithmen für drei Wearable Computing Anwendungen im E-Health entwickelt: I) mobile Atemanalyse, II) Mobile Elektrokardiogramm (engl. ECG)-Analyse und III) Aktivitätserkennung mit Inertialsensoren (engl. IMUs).

Mobile Atemwegs-Induktivität-Plethysmographie (engl. RIP) stellt eine nicht-störende Methode dar, um Atmungsmerkmale zu messen. Mit RIP-Geräten könnte auf Durchflussmessgeräte (engl. FMs) verzichtet werden, die das natürliche Atemverhalten beeinflussen. Um korrekte Atemzugvolumina zu erhalten, müssen die mit RIP gemessenen Daten angepasst werden. Dem aktuellen Stand der Technik entsprechende Anpassungsalgorithmen können erst nach den tatsächlichen Messungen verwendet werden, da diese die gleichzeitige Messung von RIP und FM erfordern. In dieser Arbeit wurden neuartige Anpassungsalgorithmen entwickelt, die die Verwendung von RIP auch außerhalb der kontrollierten Laborumgebung ermöglichen. Die Algorithmen wurden mit Daten von 193 Probanden ausgewertet, die einen Stufentest auf dem Laufband durchführten. In Beitrag (I) wurde die Anpassung über die Berechnung von Volumenbewegungskoeffizienten (engl. VMCs) aus einer Ruhephase durchgeführt. Die VMCs wurden danach für die Daten der Laufbandbelastung und der Erholungsphase verwendet. In Beitrag (II) wurden generalisierbare und referenzfreie Anpassungsalgorithmen untersucht. Die Algorithmen beziehen anthropometrische Daten der Probanden, die Zeiten der

Aus- und Einatmung und Verhältnisse der beiden RIP-Bänder mit ein. Für die praktische Anwendbarkeit empfiehlt sich ein Anpassungsalgorithmus, der ohne vorher aufgenommene Daten auskommt. Dies spricht für den referenzfreien Anpassungsalgorithmus, mit dem auch bessere Ergebnisse erzielt wurden. Atemwegserkrankungen und -infektionen gehören zu den häufigsten Todesursachen. Zukünftig sollte untersucht werden, wie RIP-Geräte effektiv für Patienten mit diesen Erkrankungen eingesetzt werden können.

Gründe für die mobile ECG-Analyse waren, echtzeitfähige Algorithmen für zwei bestimmte Bereiche zur Verfügung zu stellen. Im ersten Bereich wurden arrhythmische Herzschläge nur anhand der Zeitpunkte aufeinanderfolgender Herzschläge, der RR-Intervalle, unterschieden (Beitrag (III)). In dieser Arbeit wurden zehn aufeinanderfolgende RR-Intervalle verwendet, um Merkmale für einen Naive Bayes-Klassifikator zu berechnen. Um die Merkmale zu berechnen, wurden die aufeinanderfolgenden RR-Intervalle mit einer Sinuskurve approximiert. Die geringe Zahl an nötigen RR-Intervallen ermöglicht eine Anwendung in Echtzeit. Im zweiten Bereich wurden drei hochgenaue und verzögerungsfreie P- und T-Wellen-Erkennungsalgorithmen miteinander verglichen (Beitrag (IV)). Alle drei Algorithmen basierten nur auf der 2. Ableitung nach Einthoven. Am besten schnitt der schnelle, naive Wellenerkennungsalgorithmus ohne weitere Filterung ab. Der Vergleich mit Algorithmen aus der Literatur war aufgrund der unvollständigen Beschreibung und deren Bewertungskriterien erschwert. Diese beide Bereiche könnten die Reduzierung von Herz-Kreislauf-Erkrankungen, welche weltweit die häufigste Todesursache sind, unterstützen. Ein effektiver und einfach handhabbarer Klassifizierungsalgorithmus für Arrhythmien, wie aus Beitrag (III), könnte Personen mit erhöhtem Risiko frühzeitig zu einem Arzt schicken. Hochpräzise P- und T-Wellen-Erkennungsalgorithmen, die nur eine einzelne Ableitung benötigen, sind für alle Bereiche der EKG-Überwachung von Vorteil und erhöhen gleichzeitig den Tragekomfort der Patienten.

Aktivitätserkennung mit IMUs bietet eine objektive Möglichkeit zur Klassifikation von Aktivitäten um den Risikofaktor körperliche Inaktivität anzugehen. Dafür wurde der öffentlich verfügbare Datensatz DaLiAc aufgenommen um den Vergleich von Aktivitätserkennungsalgorithmen zu ermöglichen (Beitrag (V)). Der Datensatz enthält die IMU-Daten von 19 Probanden, die 13 verschiedene Aktivitäten des täglichen Lebens durchführten. Die IMUs wurden am Oberkörper, an der rechten Hüfte, am rechten Handgelenk und dem linken Fußgelenk platziert. DaLiAc ist als Download verfügbar (<http://www.activitynet.org/>). Unter der Verwendung von DaLiAc wurde ein hierarchisches Klassifikationsystem

entwickelt (Beitrag (VI)). Dieser übertraf sechs dem Stand der Technik entsprechende Aktivitätserkennungsalgorithmen. Im hierarchischen Klassifizierungssystem wurden die aufgenommenen Aktivitäten gruppiert und hierarchisch geordnet. Vorteile hiervon sind die einfache Integration neuer Aktivitäten, ohne dass das gesamte System neu trainiert werden muss, und die Vielfalt an Möglichkeiten für jedes einzelne Klassifizierungssystem, wie z.B. das Verwenden von verschiedenen Merkmalen, Fenstergrößen oder Klassifikatoren. Bisher wurde DaLiAc von 13 weiteren Forschungsgruppen verwendet, um deren Algorithmen zu bewerten. Das Erkennen einzelner Aktivitäten des täglichen Lebens kann das Bewusstsein jedes Einzelnen steigern, die körperliche Aktivität zu erhöhen. Denn, körperliche Inaktivität ist eine der vier Hauptrisikofaktoren für nichtübertragbare Krankheiten.

In dieser Arbeit wurden drei unterschiedliche Wearable Computing Anwendungen, die verschiedene nichtübertragbare Krankheiten angehen, vorgestellt. Einzelne, unterschiedliche Biosignale bildeten die Grundlage jeder Anwendung. Eine Fusion der Biosignale könnte die Leistung der einzelnen Anwendungsgebiete verbessern und sogar neue Anwendungsgebiete schaffen. Eine solche Fusion sollte sowohl auf der algorithmischen Ebene als auch auf der Messebene erfolgen. Dies bedeutet zum Beispiel, die Biosignale mit einem Prinzip wie einem Textil oder einer Smartwatch zu erfassen. Wearables wurden in den letzten Jahren – mit anhaltendem steigenden Trend – vermehrt in der Gesundheitsüberwachung eingesetzt, da sie dem Nutzer im täglichen Leben Vorteile bringen. Dies wird durch das Entstehen neuer Anwendungen, gesammeltes Wissen, technische Fortschritte und der fortschreitenden Digitalisierung weiter vorangetrieben.

Acknowledgment

Quite some years ago, I started working towards my PhD. During this time, I worked on various projects, met interesting people, visited different countries, and had – altogether – a great time. Thanks to everyone who supported me on this journey.

First and foremost, I would like to thank Prof. Dr. Bjoern Eskofier for the great opportunity to work under his supervision and for learning so much about scientific work. I am grateful for his extensive support during all this time, like financing the participation at conferences, sponsoring my research stay in Stanford, and during my part-time parental leave.

I owe thanks to Prof. Maria Teresa Arredondo Waldmeyer, PhD, for accepting being my second reviewer.

I received financial support from various sources. I would like to acknowledge the Bavarian Ministry for Economic Affairs, Infrastructure, Transport, and Technology and the European fund for regional development that funded the research project “miLife” with the partners Astrum IT (Erlangen, Germany), the University Hospital Erlangen, and adidas AG (Herzogenaurach, Germany). In addition, this work was financially supported by adidas Inc. (Portland, US), the Bavarian project “Leistungszentrum Elektroniksysteme (LZE)”, danova GmbH 4.0 (sold), and the Federal Ministry of Education and Research with the research project “Capital4Health”. I would like to thank all research partners that worked with me in these various projects. I want to give special thanks to Ulf Jensen and Dominik Schuldhaus for being remarkable colleagues in all parts of the miLife project, Stefan Gradl for the cooperation on mobile ECG analysis, as well as Christian Heyde for the joint work on breathing analysis, especially for acquiring the huge data set, and Lars Anneken for his enthusiastic way of working and his huge medical knowledge.

I would like to thank all students I worked throughout my research. The outcome of your theses or student assistant works lives on in several publications influencing this thesis.

I would like to acknowledge the whole team of the Pattern Recognition Lab, the Central Institute of Medical Engineering (ZiMT), and the Machine Learning and Data Analytics Lab for an amazing time. I enhanced my knowledge and learned so much from the discussions and the advices I received from you. Thank you in particular for the social events. I want to emphasise Patrick Kugler for introducing me to the world of pattern recognition and biomedical signal analysis and Matthias Ring for the helpful discussions and being a great office colleague.

Finally, I would like to thank my family and friends. I owe thanks to my friends, especially Chrissi and Luisa, for offering conversation topics and sport activities apart from research and PhD studies, providing me with the motivation to continue. Last but not least, I would like to thank my family, my parents, Irene and Robert, my sisters, Katrina and Eva, and Anna, for their encouragement throughout the years, and especially in the last years for also taking care of Noemi. Thank you Andreas for your love and support. I am grateful that Konstantin, Noemi, and you are part of my life.

Contents

I	Research Background	1
1	Introduction	3
1.1	Motivation	3
1.2	Contributions	7
1.3	Structure of this Thesis	12
2	State of the Art: Literature and Patents	15
2.1	Mobile Breathing Analysis	15
2.1.1	Indirect Measurement of Breathing	15
2.1.2	Adjustment Algorithms	20
2.2	Mobile ECG Analysis	22
2.2.1	QRS Complex Detection	22
2.2.2	Clinical Applications of RR Interval Analysis	22
2.2.3	P- and T-wave Detection	28
2.3	IMU-based Activity Recognition	33
2.3.1	Assessment of Physical Activity	33
2.3.2	Activity Classification Algorithms	39
3	Fundamentals	47
3.1	The Electrocardiogram	47
3.1.1	Physiologic Basis	48
3.1.2	Measurement	48
3.2	Regression	50
3.2.1	Simple Linear Regression	51
3.2.2	Multiple Linear Regression	51
3.2.3	Support Vector Regression	52
3.3	Supervised Learning	54
3.3.1	Supervised Learning Pipeline	54

3.3.2	Classifiers	56
3.3.3	Cross-Validation	60
II	Mobile Breathing Analysis	63
4	Individual-Adapted Adjustment of Respiratory Inductance Plethysmography	65
4.1	Methods	65
4.1.1	Data	65
4.1.2	Adjustment Algorithms	70
4.1.3	Calculation of Regression Coefficients	70
4.1.4	Performance Measure and Evaluation	71
4.1.5	Comparison to Algorithms in Literature	71
4.2	Results	72
4.3	Discussion	75
5	Reference-Free Adjustment of Respiratory Inductance Plethysmography	79
5.1	Methods	79
5.1.1	Data	79
5.1.2	Adjustment Algorithms	80
5.1.3	Calculation of Regression Coefficients	83
5.1.4	Performance Measures and Evaluation	84
5.1.5	Comparison to Algorithms in Literature	86
5.2	Results	87
5.3	Discussion	88
III	Mobile ECG Analysis	101
6	Arrhythmia Classification Using RR Intervals	103
6.1	Methods	103
6.1.1	Data	103
6.1.2	Features	105
6.1.3	Classification	108
6.1.4	Performance Measures and Evaluation	108
6.2	Results	110
6.3	Discussion	111

7	ECG Fiducial Points Detection Algorithms	117
7.1	Methods	117
7.1.1	Data	117
7.1.2	Algorithm Overview	120
7.1.3	P- and T-Wave Event Detection Algorithms	123
7.1.4	Performance Measures and Evaluation	125
7.2	Results	126
7.3	Discussion	128
IV	IMU-based Activity Recognition	141
8	Recognition of Single Activities	143
8.1	Methods	143
8.1.1	Data	143
8.1.2	Hierarchical, Multi-Sensor Based Classification System	148
8.1.3	Performance Measures and Evaluation	151
8.1.4	Comparison to Algorithms in Literature	151
8.2	Results	153
8.3	Discussion	153
V	Perspective	163
9	Summary	165
9.1	Mobile Breathing Analysis	165
9.2	Mobile ECG Analysis	166
9.3	IMU-based Activity Recognition	167
10	Discussion	169
10.1	Mobile Breathing Analysis	169
10.2	Mobile ECG Analysis	171
10.3	IMU-based Activity Recognition	173
11	Outlook	177
	Appendices	181
A	Additional Results for Mobile Breathing Analysis	183
B	Additional Results for Mobile ECG Analysis	187

C Subject Information for the DaLiAc Dataset	191
List of Figures	199
List of Tables	201
Glossary	205
Bibliography	207

Part I

Research Background

Chapter 1

Introduction

1.1 Motivation

The growth of non-communicable or chronic diseases together with the rapid population ageing represent important challenges for healthcare systems globally (HelpAge International, 2016; World Health Organization [WHO], 2017). Non-communicable diseases are not transferable among individuals. They are the leading cause of death and disability worldwide, with a still increasing prevalence (NCD Alliance, 2019; WHO, 2017). Most deaths are caused by cardiovascular diseases and malignant neoplasms (cancer) (Figure 1.1). Almost two thirds of them are linked to physical inactivity, unhealthy diets, tobacco use, and harmful use of alcohol (NCD Alliance, 2019). At the same time, everyone is surrounded increasingly with wearables, like smartphones and smartwatches, making wearable computing an integral part of everyday life. In the context of this thesis, wearable computing means the wearable device itself as well as the required algorithms to interpret information from the wearable. The combination – addressing non-communicable diseases with the help of wearables – seems one reasonable solution. However, innovative solutions from the healthcare engineering world are required to meaningfully use wearables and address the high prevalence of non-communicable diseases.

In this thesis, methods and algorithms are presented for three application fields of wearable computing: I) mobile breathing analysis, II) mobile ECG analysis, and III) IMU-based activity recognition. The specific motivation for each of these fields is given on the next pages.

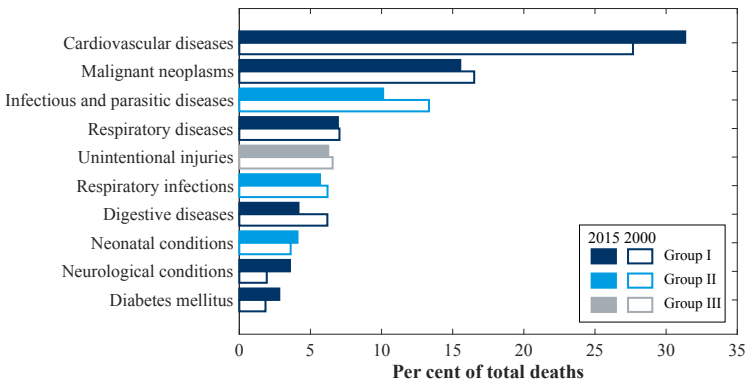


Figure 1.1: Distribution of deaths by leading cause groups comparing the years 2015 and 2000. The three cause groups are given in different colors: Group I: Non-communicable diseases; Group II: Communicable, maternal, perinatal, and nutritional conditions; Group III: Injuries. Reproduced from Figure 4 of WHO (2008, p. 10) using data from WHO (2016).

Mobile Breathing Analysis

Breathing is essential for living. Without adequate breathing, life is in danger (Folke, Cernerud, Ekström, & Hök, 2003). For example, breathing is monitored to identify pathologic minute ventilation or oxygen deficiency (Johansson & Hök, 2004). Minute ventilation, also called pulmonary ventilation or flow of air, means the volume of air that can be inhaled or exhaled within one minute. Oxygen deficiency is caused by apnea or inadequate minute ventilation. Injury or illness can affect respiratory rate and/or tidal volume, which then alter minute ventilation. The respiratory rate states the number of breaths per minute. The tidal volume gives the volume of air that is inhaled or exhaled during breathing within one breath. Hence, it is insufficient to only monitor respiratory rate alone. Instead, the monitoring of both, respiratory rate and tidal volume, is of importance (Johansson & Hök, 2004).

The monitoring of breathing is not only important to diagnose pathologic breathing, but also in the assessment of sports performance where athletes and their coaches intend to enhance their sports performance (O'Donoghue, 2010). Before an improvement is possible, they need to understand, and for that reason to observe, the actual sports performance (O'Donoghue, 2010). Sports performance is related to breathing: correct

breathing in the appropriate aerobic-anaerobic transition region increases the sports performance.

Standard devices for monitoring breathing directly are FMs assessing the airflow directly on the nose and/or mouth. These require mouth pieces and face masks, which change the natural breathing pattern (Askanaži et al., 1980; Gilbert, Auchincloss, Brodsky, & Boden, 1972). They lower the respiratory rate, increase the tidal volume (Gilbert et al., 1972), and are disturbing for the subject. In addition, breathing can only be assessed in non-ambulatory laboratory settings.

Breathing can be monitored directly or indirectly. Several techniques for monitoring breathing indirectly exist (Brimacombe, Macfie, & McCrirrick, 1991; Bronstein & Brocks, 2017; Chen, Wang, & Ko, 1990; Cohen, Zigelbaum, Gross, Roussos, & Macklem, 1982; Dumond et al., 2017; Fekr, Radecka, & Zilic, 2014; Folke et al., 2003; Johansson & Hök, 2004; Lamine et al., 2004; Leutheuser et al., 2011; Leutheuser et al., 2015; Mehra & Strohl, 2009; Tobola, Ershov, Wissendheit, & Couronné, 2009; Tobola & Weigand, 2007; Wilhelm, Roth, & Sackner, 2003) (Section 2.1). These techniques do not measure the actual respiratory flow, but parameters involved in the respiratory muscle activity, rib cage (RC) and abdominal motion, or intrathoracic pressure variations (Johansson & Hök, 2004). Of these, techniques using the RIP method are recommended (Berry et al., 2012; Bronstein & Brocks, 2017), as this method enables the measurement of the entire surface area enclosed by the band.

RIP devices serve as wearable equipments for respiratory assessment, whereas unadjusted values are acquired. These unadjusted values need to be processed to result in correct ventilatory tidal volumes. State of the art methods (Banzett, Mahan, Garner, Brughera, & Loring, 1995; Chadha et al., 1982; Sackner et al., 1989) adjust RIP data by using simultaneously measured FM data, involving the disadvantages of FM.

The first application presented in this thesis is concerned with determining correct ventilatory tidal volumes in the field – outside a controlled laboratory environment and without simultaneous measurement of FM – using novel adjustment algorithms for RIP.

Mobile ECG Analysis

Cardiovascular diseases are the number one cause of death throughout the world (WHO, 2017). The two most common cardiovascular diseases are ischaemic heart disease and stroke (WHO, 2017). Symptoms leading to these diseases are atrial fibrillation and arrhythmias. However, these symptoms are spontaneous and occur rarely when the patient is seeing a

physician. A well documented medical condition is, though, crucial for appropriate patient treatment.

The ECG – measuring the electric activity of the heart – is the most used biosignal in diagnosis and patient monitoring as well as in fitness applications due to simple, risk-free, and inexpensive acquisition modalities. Analysis of the ECG, like the investigation of fiducial points (QRS complex, P- and T-wave) which correspond to different cardiac contraction phases (chapter 3.1), helps to identify cardiovascular diseases and is needed for continuous monitoring of the patient's heart status in intensive care units (ICUs). On the other hand, early detection using a simple indicator like abnormal heart rhythm – determined unobtrusively, in all sorts of environments – could trigger the timely presentation to the emergency department or outpatient unit for minimizing or eliminating resulting damages.

Mobile devices like smartphones and tablets are an integral part of everyone's life (International Data Corporation, 2013; International Telecommunication Union, 2013). Thus, they form the ideal basis for mobile monitoring and real-time analysis of signals related to the human heart, like the ECG (Gradl, Kugler, Lohmüller, & Eskofier, 2012) or the pulse oximeter signal (Scully et al., 2012). Pulse oximeters measure the oxygen saturation non-invasively and from the signal, the heart rate can be predicted. These applications and possible measurement equipment should be simple to handle and should provide comprehensible and easy to follow instructions for the user.

The most used leads in long-term ECG recordings correspond to the three leads of Einthoven (I, II, and III). In recordings of lead II, the different waves and fiducial points are clearly visible (Figure 3.1). Single-lead ECG analysis would reduce the number of necessary electrodes, the measurement equipment, and simultaneously enhance the patient's comfort.

The second application presented in this thesis is concerned with mobile ECG analysis for two particular scopes: Identifying arrhythmic beats using only time instances of individual heart beats – the current heart rate –, and presenting high-accurate single-lead, instantaneous P- and T-wave detection algorithms.

IMU-Based Activity Recognition

Physical activity is one key element in healthy ageing (Daskalopoulou et al., 2017) and healthy living (WHO, 2018). Insufficient physical activity is the 4th leading risk factor for mortality (WHO, n.d.). Approximately 3.2 million people of the world population decrease each year because of insufficient

physical activity (WHO, n.d.). Physically active people have higher levels of health-related fitness and lower levels of various non-communicable diseases (Berlin & Colditz, 1990; Blair, Cheng, & Holder, 2001; I. M. Lee & Paffenbarger, 2000; Oguma, Sesso, Paffenbarger, & Lee, 2002; Powell, Thompson, Caspersen, & Kendrick, 1987; Warburton, Nicol, & Bredin, 2006; WHO, n.d., 2018). The WHO (2018) recommends at least 150 min of moderate physical activities throughout the week. Walking, ascending or descending stairs, or certain household activities are examples for moderate physical activities (Ainsworth et al., 2011; Ainsworth et al., 1993; Ainsworth et al., 2000). The most performed physical activity is walking (Rafferty, Reeves, McGee, & Pivarnik, 2002; Siegel, Brackbill, & Heath, 1995). Walking short distances instead of driving or using stairs instead of taking an elevator are possibilities to enhance one's own activity level day by day (Pate et al., 1995).

Despite knowing to perform regularly physical activity, only 3 in 4 adults meet the recommendations of the WHO. For adolescents it is even worse: more than 80% of the world's adolescents are insufficiently active (WHO, 2018). Assessing the individual daily life activity is of considerable interest in getting insights into the current health status. The conventional assessment of activities with self-reports induces questions regarding reliability, validity, and sensitivity (Shephard (2003), Table 2.3). Assessing activities with small and light-weight wearable IMUs provides a reliable and objective method.

Modern human activity recognition systems differ in the number and kind of sensors, the recognizable activities, and the sampling rate (Manini and Sabatini (2010), Table 2.4). Thus, it is problematic to compare new recognition algorithms to existing algorithms in literature as few commonly used datasets exist.

The third application presented in this thesis is concerned with the objective recognition of activities using wearable IMUs. First, activities were collected to form the benchmark dataset DaLiAc. Second, a hierarchical classification system is presented using DaLiAc.

1.2 Contributions

This thesis contributes to methods and applications in the areas of biosignal processing and machine learning in the context of wearable computing. The main contributions of the thesis are divided in the three abovementioned fields.

Contributions to the Field of Mobile Breathing Analysis

State of the art methods for adjusting RIP data can only be applied after the actual measurement as these require the simultaneous data acquisition of RIP and FM data for the whole measurement time. In this thesis, algorithms for adjustment of RIP are developed to enable the recording of ventilatory data in the field, outside the controlled laboratory environment and without FM measurement. The performance is assessed using a large subject population ($n = 193$) with data acquired in three different phases (standing still, treadmill running, and recovery).

- (I) It is shown that individual adjustment parameters calculated from a standing still phase using low tidal volume data provide correct high tidal volume data. For this, multiple linear regression models with and without offset as well as Support Vector Regression (SVR) were investigated. These algorithms only require the simultaneous data acquisition of RIP and FM during the standing still phase. Methods and results are presented in Chapter 4 and are based on a conference proceeding (Leutheuser, Heyde, Gollhofer, & Eskofier, 2014).
- (II) Novel adjustment algorithms for RIP are presented which are accurate, completely reference-free – without FM measurement –, and ready for use on new subjects. These adjustment algorithms consist of nesting two multiple linear regression models into one multiple linear regression model. An innovative approach for solving the nested multiple linear regression model is suggested. Further, the actual accuracy loss when RIP data is derived using only one band of the standard two bands (either RC or abdomen (AB)) is shown. Methods and results are presented in Chapter 5 and are published in a journal article (Leutheuser, Heyde, Roecker, Gollhofer, & Eskofier, 2017).

Contributions to the Field of Mobile ECG Analysis

The basis of mobile ECG analysis is the detection of needed relevant events in the ECG and the appropriate processing, for example classification. Two aspects of mobile ECG analysis are addressed in this thesis. The first is concerned with the early detection of heart related diseases using a simple, ready and easy to use arrhythmic beat classification algorithm. The second deals with the precise documentation of medical conditions based on the detection of fiducial points.

- (III) An arrhythmia classification algorithm is introduced using only features calculated from the time instances of individual heart beats. The features are based on the phenomenon that consecutive RR intervals are unequal in duration. Methods and results are presented in Chapter 6 and are published in a conference proceeding (Leutheuser et al., 2015). This work was awarded with the “Best Poster Award” of the IEEE International Conference on Wearable and Implantable Body Sensor Networks 2015.
- (IV) Algorithms are presented that enable high-accurate and instantaneous P- and T-wave detection for assessing different cardiac contraction phases. They are single-lead algorithms, appropriate for ECG data based on lead II of Einthoven. Methods and results are presented in Chapter 7 and are published in a conference proceeding (Leutheuser et al., 2016).

Contributions to the Field of IMU-Based Activity Recognition

The comparison of different activity recognition algorithms would benefit from a common used, publicly available dataset. In this thesis, a dataset exactly for this case is created. The performance of existing activity recognition algorithms from literature are compared to a novel algorithm using the same dataset.

- (V) A benchmark dataset called DaLiAc for the comparison of activity recognition algorithms is presented. This dataset encompasses 13 activities of 19 subjects and is publicly available (<http://www.activitynet.org/>). The activities comprise static activities like sitting or lying, locomotion activities like walking and climbing stairs, or physical activities like treadmill running or cycling. This dataset is presented in Chapter 8 and is published in a journal article (Leutheuser, Schuldhaus, & Eskofier, 2013).
- (VI) A novel, hierarchical, multi-sensor based classification system is introduced that outperforms existing algorithms on the basis of the DaLiAc dataset. The hierarchical structure incorporates the easy possibility of integrating additional activities or the splitting up of single activity groups depending on the use case or the functionality. Methods and results are presented in Chapter 8 and are published in a journal article (Leutheuser, Schuldhaus, & Eskofier, 2013).

Influence of Additional Publications

The chapters of this thesis are further influenced by material and methods that had been published in conference proceedings and journals. These are given in the following:

- Adjustment of RIP with data from simultaneous FM measurement (Heyde, Leutheuser, et al., 2014)
The contribution of the author to this study consisted of the complete data analysis.
- Respiratory motion analysis based on range imaging using 4-D shape priors (Wasza, Fischer, Leutheuser, et al., 2016)
The contribution of the author consisted of supervising a student working on his bachelor thesis where the content is part of this publication.
- Presentation of the FitnessSHIRT, a textile integrated wearable computing device (Leutheuser, Lang, et al., 2017)
The contribution of the author consisted of writing most parts of the corresponding publication.
- Underwater monitoring of the ECG signal with an emergency alert system (Cibis, Groh, Leutheuser, & Eskofier, 2015)
The contribution of the author consisted of being the second supervisor of the accompanying student work.
- Mobile measurement of the ECG using a current-based system (Gradl, ..., Leutheuser, et al., 2017)
The contribution of the author consisted of being the second supervisor of the accompanying Master thesis.
- Detection of the R-peak from the ECG signal (Gradl, Leutheuser, Elgendi, et al., 2015; Lang, ..., Leutheuser, et al., 2015).
The contribution of the author consisted of providing feedback and input and reviewing the corresponding paper.
- Analysis of the ECG signal towards arrhythmia classification (Leutheuser, Gottschalk, et al., 2014; Leutheuser, Gradl, Kugler, et al., 2014)
The contribution of the author consisted of having the idea for the study, performing the analysis and writing the corresponding paper.

- Analysis of the ECG signal towards human authentication (Tantinger, ..., Leutheuser, et al., 2015)
The contribution of the author consisted of being the main supervisor of the Bachelor thesis leading to first results that are integrated into this publication.
- Investigation of heart rate variability (HRV) parameters during running (Leutheuser & Eskofier, 2012, 2013)
The contribution of the author consisted of having the idea for the study, performing the analysis and writing the corresponding paper.
- Heart rate estimation using photoplethysmographic and accelerometer data (Mullan, ..., Leutheuser, Eskofier, Pasluosta, 2015)
The contribution of the author consisted of being the supervisor to a project conducted by students.
- Activity recognition using decision level fusion (Schuldhaus, Leutheuser, & Eskofier, 2013)
The contribution of the author consisted of providing feedback and input to the analysis and reviewing the paper.
- A database fusion strategy for activity recognition (Schuldhaus, Leutheuser, & Eskofier, 2014)
The contribution of the author consisted of providing feedback and input to the analysis and reviewing the paper.
- Classification of gait with a single trunk-mounted inertial-magnetic measurement unit (Full, Leutheuser, et al. 2015)
The contribution of the author consisted of being the main supervisor of the accompanying Bachelor thesis.
- Step segmentation using IMU data (Leutheuser, Doelfel, et al., 2014)
The contribution of the author consisted of having the idea for the study, performing the analysis and writing the paper.
- Activity level classification with embedded devices using accelerometer data (Jensen, Leutheuser, et al., 2015)
The contribution of the author consisted of setting up the study, acquiring the data, and reviewing the paper.
- Analysis of free living accelerometer data to differentiate the physical performance phenotypes of patients with and without chronic low back pain (Zheng, Muaremi, Leutheuser, et al., 2017)

The contribution of the author consisted of performing parts of the data analysis.

- Energy expenditure estimation with IMU data (Schuldhaus, Dorn, Leutheuser, et al., 2014)

The contribution of the author consisted of being the second supervisor of the accompanying Bachelor thesis.

- Somnography with accelerometer data and an Android-based mobile phones (Gradl, Leutheuser, et al., 2013)

The contribution of the author consisted of being the main supervisor of the accompanying Bachelor thesis.

- Semiology-independent seizure detection using the electrodermal activity and accelerometer data (Heldberg, Kautz, Leutheuser, et al., 2015)

The contribution of the author consisted of being the second supervisor of the accompanying Bachelor thesis.

- Reduction of electromyogenic artifacts in Electroencephalography data (Gabsteiger, Leutheuser, et al., 2014a, 2014b; Leutheuser, Gabsteiger, et al., 2013)

The contribution of the author consisted of being the main supervisor of the accompanying Bachelor thesis.

- An Ultra-Low-Power Sensor Evaluation Kit (ULPSEK) for evaluation of biomedical sensors and monitoring applications (<http://www.ulpsek.com>) (Tobola, Espig, ..., Leutheuser, et al., 2015; Tobola, Korpok, Leutheuser, et al., 2014; Tobola, Leutheuser, et al., 2018; Tobola, Leutheuser, et al., 2016; Tobola, Streit, ..., Leutheuser, et al., 2015)

The contribution of the author consisted of providing feedback and input to the methodologies and writing parts of the publications.

1.3 Structure of this Thesis

The remaining thesis is structured in five parts (Figure 1.2). The first part gives the research background. The second part refers to mobile breathing analysis. The third part is about mobile ECG analysis. The fourth part deals with IMU-based activity recognition. The fifth part provides the perspective of this thesis. The contributions are handled in the three middle parts. In each of these parts, two different contributions are addressed.

I Research Background		
Introduction Chap. 1	State of the Art Chap. 2	Fundamentals Chap. 3
II Mobile Breathing Analysis	III Mobile ECG Analysis	IV IMU-Based Activity Recognition
Individual Adjustment of RIP Chap. 4	Arrhythmia Classification with RR intervals Chap. 6	Benchmark Dataset DaLiAc Chap. 8
Reference-Free Adjustment of RIP Chap. 5	Fiducial Points Detection Algorithms Chap. 7	Recognition of Single Activities
V Perspective		
Summary Chap. 9	Discussion Chap. 10	Outlook Chap. 11

Figure 1.2: Structure of this thesis. Overview of the analysis steps. RIP: Respiratory inductance plethysmography; ECG: Electrocardiogram; IMU: Inertial Measurement Unit.

The continuing first part, I Research Background, provides a review on literature and patents (chapter 2). This chapter is divided into the three method parts. Additionally, in the first part, the fundamentals for the following chapters are collected (chapter 3).

In the second part, II Mobile Breathing Analysis, novel adjustment algorithms for RIP are created to enable the usage of RIP alone and outside the controlled laboratory environment. In chapter 4, adjustment is achieved by using the corresponding standing still phase. In chapter 5, generalizable and completely reference-free adjustment algorithms are investigated.

The third part considers two scopes of mobile ECG analysis (part III). In chapter 6, arrhythmic beats are identified using only time instances of successive heartbeats. In chapter 7, three high-accurate single-lead P- and T-wave detection algorithms are compared.

In the fourth part, IV IMU-Based Activity Recognition, the benchmark dataset DaLiAc is presented (chapter 8). In the same chapter, a hierarchical classification system is presented and compared with state of the art algorithms from literature using DaLiAc.

Chapter 1. Introduction

In the fifth part, V Perspective, first, the methods and findings of this thesis are summarized (chapter 9). Second, the two single contributions of each part are connected with each other and are taken into context of this thesis (chapter 10). The thesis is completed by providing an outlook considering all contributions together (chapter 11).

Chapter 2

State of the Art: Literature and Patents

In this chapter, literature and patents giving the state of the art for the three parts are collected. First, the indirect measurement of breathing and adjustment algorithms for mobile breathing analysis are considered. Second, for the mobile ECG analysis part, current approaches from literature for QRS complex detection, clinical applications of RR interval analysis, and algorithms for P- and T-wave detection were consolidated. Third, subjective and objective measures for assessing physical activity and activity recognition algorithms were summarized.

2.1 Mobile Breathing Analysis

2.1.1 Indirect Measurement of Breathing

Apart from directly measuring breathing in a stationary setting using face masks, mouth pieces, and nose clips, it is possible to derive breathing pattern and parameters indirectly. Johansson and Hök (2004) and Folke et al. (2003) provide overviews of indirect sensors for breathing monitoring. Based on the work of Johansson and Hök (2004) and the structure provided by Folke et al. (2003), Table 2.1 was created and extended using additional literature (Brimacombe et al., 1991; Bronstein & Brocks, 2017; Chen et al., 1990; Cohen et al., 1982; Dumond et al., 2017; Fekr et al., 2014; Folke et al., 2003; Lamine et al., 2004; Leutheuser et al., 2011; Leutheuser et al., 2015; Mehra & Strohl, 2009; Wilhelm et al., 2003) and patents (Tobola et al., 2009; Tobola & Weigand, 2007). For each method / device, the measuring parameter, the sensor type and placement together with the benefits and the weaknesses are given.

Table 2.1: Ambulatory and indirect determination of tidal volume. This overview is based on the work of Johansson and Hök (2004) and Folke et al. (2003) and extended with additional literature and patents, stated correspondingly and marked in *italic*.

Method / device	Measuring parameter	Sensor type and placement	Benefits (+) and weaknesses (–)
Accelerometer sensor	<i>Anterior-posterior diameter of the chest</i>	<i>Accelerometers on the chest (Fekr et al., 2014)</i>	– <i>Measuring distance and not cross-sectional area</i>
Electrocardiogram (ECG)	Respiratory sinus arrhythmia (RSA) or R-peak modulation calculated from heart activity (<i>Lamine et al., 2004; Leutheuser et al., 2011; Leutheuser et al., 2015</i>)	Two skin surface potential electrodes on the chest	+ Simultaneous measurement of respiration and cardiac monitoring – Respiratory sinus arrhythmia (RSA) not present in all subjects – No breathing assessment during cardiac arrhythmia – Inadequate respiratory rate when respiratory rate is higher than the heart rate (Nyquist theorem) – Not yet possible to derive tidal volume
Electromyography	Muscle activity	Electrodes near the diaphragm either intercostal, in the esophagus, or on the skin of the chest	+ Possibility of early detection of respiratory muscle fatigue (<i>Cohen et al., 1982</i>) – Low signal level – Easy disturbances by electrode motion or other electrical activity of the human body, primarily the ECG – <i>Difficult to remove non-respiratory muscle electromyography (Mehra & Strohl, 2009)</i>

Table 2.1: Ambulatory and indirect determination of tidal volume (continued).

Method / device	Measuring parameter	Sensor type and placement	Benefits (+) and weaknesses (–)
Finger pressure sensors	Peripheral pulse wave	Piezoelectric material (e.g. polyvinylidene fluoride (PVDF)) around the finger (<i>Chen et al., 1990</i>)	+ <i>Possibility of simultaneous measurement of blood pressure, heart rate, and blood flow</i>
Magnetometer	Anteroposterior displacements of the rib cage and abdomen and axial displacements of the chest wall and spine (<i>Dumond et al., 2017; Tobola et al., 2009</i>)	Two coils on the back and two coils on the chest around rib cage and abdomen	– Measuring distance and not cross-sectional area – Sensitive upon distortions with respect to motion, posture variation, and breathing pattern

Table 2.1: Ambulatory and indirect determination of tidal volume (continued).

Method / device	Measuring parameter	Sensor type and placement	Benefits (+) and weaknesses (-)
Photoplethysmography (PPG)	Skin blood volume and flow / venous pulsation	PPG sensor and photodetector	+ Possibility of simultaneous measurement of respiratory rate, heart rate, and arterial oxygen saturation - Loss of signal quality during motion - Risk that low-frequency blood pressure waves may interfere during long-term registrations - Monitoring respiratory rate / tidal volume is still at an early stage, whereas tidal volume can be extracted in healthy subjects - Disturbances in the complex respiratory signal transfer process
Respiratory inductance plethysmography (RIP)	Encircled cross-sectional area of thorax and abdomen	Self-inductive sinusoidal wire coils insulated in elastic bands around rib cage and abdomen (Bronstein & Brocks, 2017; Tobola & Weigand, 2007; Wilhelm et al., 2003)	+ Output more linearly related to lung volume due to measuring cross-sectional area and not distance + Less sensitive to distortion of the two compartments during motion or posture change - Expensive - Covers a large body area

Table 2.1: Ambulatory and indirect determination of tidal volume (continued).

Method / device	Measuring parameter	Sensor type and placement	Benefits (+) and weaknesses (–)
Strain gauges	Thorax and abdominal circumference	<i>Resistive, capacitive, piezoelectric, or fibre-optic strain gauges around thorax and abdomen (Brimacombe et al., 1991; Folke et al., 2003)</i>	– Nonlinear with respect to tidal volume – Difficult procedure of gauge positioning – Dependence on body positions – <i>Tight and uncomfortable fit</i>
Transthoracic impedance plethysmography	Transthoracic impedance	Two skin surface potential electrodes on the chest	-/+ <i>PVDF reacts almost instantaneously to changes in temperature, pressure, strain, and impedance (Koo et al., 2011)</i> – <i>No linear relation to changes in cross-sectional dimensions</i> – Not yet possible to reliably monitor tidal volumes of individual breaths – Difficulties in identifying paradoxical breathing as Konno and Mead principle (<i>Konno & Mead, 1967</i>) not valid – <i>Sensitive to artifacts originating in the skin-electrode interface, cardiac pulsation (cardiogenic artifact), patient motion, and changes in posture (Folke et al., 2003; Mehra & Strohl, 2009)</i>

In Table 2.1, mattress systems (Chow, Nagendra, Abisheganaden, & Wang, 2000), non-contact devices (Wasza et al., 2016), and invasive devices (Meagher, Jensen, Weil, & Shubin, 1966) were excluded as these devices inevitably require stationary laboratory equipment. The focus was the wearable, unobtrusive, and indirect assessment of breathing.

No measurement principle clearly stands out from the nine different possibilities (Table 2.1). Up to now, it is not possible to derive accurate absolute values in tidal volume without an appropriate adjustment for all measurement principles. Solely simple breathing parameters like the individual timings of inspiration and expiration can be extracted without any cumbersome analysis. Appropriate adjustment means that the measured parameters need to be adjusted to the actual tidal volume so that the correct breathing parameters including tidal volume can be measured. For this, adjustment algorithms are needed.

Nevertheless, the advantages of RIP surpass the other methodologies, while the negative aspects resemble the other methodologies. Additionally, RIP was recommended by the American Academy of Sleep Medicine (AASM) for indirect measurement of respiration (Berry et al., 2012; Bronstein & Brocks, 2017). In this thesis, the LifeShirt device (Wilhelm et al., 2003)—based on RIP technology—was chosen for data acquisition (section 4.1.1).

2.1.2 Adjustment Algorithms

The essential breathing parameters, respiratory rate and tidal volume, have to be provided instantaneously—in real-time—and accurately from real-life settings. In the following, state of the art adjustment algorithms for the indirect measurement of breathing are presented. These determine the tidal volume using simultaneous reference measurements at the mouth or nose. Up to now, it is required to have data from both measurement principles to provide accurate tidal volumes.

Konno and Mead (1967) assumed that the chest wall has two moving parts (RC and AB). They defined the following two-degree of freedom model:

$$V_T = V_{T_{RC}} + V_{T_{AB}}, \quad (2.1)$$

where the true tidal volume V_T can be calculated as sum of the adjusted values of the partial tidal volume changes of the RC ($V_{T_{RC}}$) and AB ($V_{T_{AB}}$). They used the isovolume maneuver to obtain adjusted values. In the isovolume maneuver, a known constant lung volume is shifted by the subject from the RC to the AB and vice versa.

Chadha et al. (1982) suggested using least squares approximation (LSQ) to solve equation (2.2):

$$V_T = gain_{RC} \cdot uV_{T_{RC}} + gain_{AB} \cdot uV_{T_{AB}}, \quad (2.2)$$

where $uV_{T_{RC}}$ ($uV_{T_{AB}}$) is the unadjusted value of the partial tidal volume changes of the RC (AB), and $gain_{RC}$ ($gain_{AB}$) is the corresponding VMC. Here, it is assumed that each adjusted value of the partial tidal volume ($V_{T_{RC}}$ or $V_{T_{AB}}$) can be calculated as product of the VMC $gain_{RC,AB}$ and the unadjusted value of the corresponding partial tidal volume change $uV_{T_{RC,AB}}$.

Sackner et al. (1989) presented the Qualitative Diagnostic Calibration (QDC) as a simplified adjustment routine.

$$V_T = M \cdot [K \cdot uV_{T_{RC}} + uV_{T_{AB}}]. \quad (2.3)$$

In this routine, first, a number of known quasi-constant tidal volumes are selected out of a 5 min recording phase under resting conditions. The k-ratio K —reflecting the relative contribution of the $uV_{T_{RC}}$ and $uV_{T_{AB}}$ —is calculated using the following equation by de Groote, Paiva, and Verbandt (2001):

$$K = \frac{SD(uV_{T_{AB}})}{SD(uV_{T_{RC}})}, \quad (2.4)$$

where SD is the standard deviation (SD). In a second step, the scaling factor M is calculated by solving equation (2.3) in the presence of the known tidal volume. As preprocessing step, Sackner et al. (1989) suggested to consider only breaths that lay within ± 1 SD around $[uV_{T_{RC}} + uV_{T_{AB}}]$.

Banzett et al. (1995) suggested another attempt to simplify the adjustment by using a fixed k-ratio:

$$V_T = M[2 \cdot uV_{T_{RC}} + uV_{T_{AB}}]. \quad (2.5)$$

However, both approaches by Sackner et al. (1989) and Banzett et al. (1995) still require measurements at the orifice to calculate the scaling factor M . This might limit certain applications where absolute accuracy is needed.

All state of the art approaches require the simultaneous data acquisition of a standard device measuring breathing directly like FM and an indirect device like RIP. A simple and reliable reference-free method for RIP adjustment in the absence of reference measures at the mouth or nose has not yet been developed. Face masks, mouth pieces, or breathing bags, as they are needed when using FMs, are currently still required. In this part, breathing data using both measurement principles RIP and FM

were acquired. Different algorithms are developed that need at maximum a short prerecording time with both measurement principles. The algorithms are evaluated when applying the obtained VMCs to solely RIP data and comparing the results with the values from FM data. After verifying that accurate breathing data can be acquired with RIP and an appropriate adjustment algorithm, data only with the RIP device can be recorded. These different adjustment algorithms for RIP are described in the chapters 4 and 5 belonging to contributions (I) and (II), respectively.

2.2 Mobile ECG Analysis

2.2.1 QRS Complex Detection

The QRS complex as most prominent wave in the ECG signal, lead II according to Einthoven (Figure 3.1), can be detected with both sensitivity SE and positive predictivity $+P$ above 99% (Elgendi, 2013; Gradl, Leutheuser, Elgendi, Lang, & Eskofier, 2015; Köhler, Hennig, & Orglmeister, 2002) using standard databases. The R-peak is the most prominent peak in the ECG. The algorithms for QRS complex or R-peak detection are based on different underlying principles.

Examples of those together with other principles in literature are signal derivatives and digital filters (Hamilton & Tompkins, 1986; Pan & Tompkins, 1985), Wavelet transforms (J. P. Martínez, Almeida, Olmos, Rocha, & Laguna, 2004), or incorporating existing knowledge (Elgendi, 2013). Out of these, algorithms exist that are capable to run in real-time on wearables (Elgendi, 2013; Gradl et al., 2012; Gradl et al., 2015). In this thesis, the algorithm of Gradl et al. (2015) is used and explained in detail in chapter 7.1.2. This publication includes the work of the author of this thesis.

2.2.2 Clinical Applications of RR Interval Analysis

RR interval analysis is characterized by the phenomenon that consecutive RR intervals are unequal in duration (Leutheuser & Eskofier, 2012, 2013; Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, 1996). This is also called HRV. This effect characterizes a positive health state as HRV changes during pathologic conditions.

The application areas of HRV analysis are diverse. In this work, the focus is on arrhythmia detection and the branch atrial fibrillation detection. Table 2.2 gives an overview of algorithms for atrial fibrillation or arrhythmia detection using RR intervals. The table comprises for each publication

the authors, the key techniques of their algorithms, the training and test dataset, and the performance on the test dataset. The performance is divided into sensitivity SE , specificity SP , positive predictivity $+P$, and accuracy ACC .

As key technique, first, different time series were calculated. Some authors directly used the RR intervals whereas others preferred to solely or additionally use the ΔRR interval series (Huang et al., 2011; Kikillus, Hammer, Lentz, Stockwald, & Bolz, 2007; Y. Li, Tang, Wang, & Tang, 2017; Lian, Wang, & Muessig, 2011; Tateno & Glass, 2001). ΔRR interval series give the time differences between two successive RR intervals. Second, some authors preprocessed the RR intervals, for example with normalizing (Logan & Healey, 2005), eliminating ectopic beats (J. Lee, Nam, McManus, & Chon, 2013), or filtering (Petr nas, Marozas, & S rnmo, 2015; Zhou, Ding, Ung, Pickwell-MacPherson, & Zhang, 2014). Third, features were extracted. Features ranged from calculating parameters like arithmetic means (G. B. Moody & Mark, 1983) or Shannon entropy (Dash, Chon, Lu, & Raeder, 2009; J. Lee, Nam, et al., 2013; McManus et al., 2013; Zhou et al., 2014) to histograms (Huang et al., 2011; Kikillus et al., 2007; Tateno & Glass, 2001) or constructing maps (Y. Li et al., 2017; Lian et al., 2011). The full list of features is given in Table 2.2 in column key techniques. Fourth, for determining the presence of atrial fibrillation or arrhythmias, the features were combined and reduced (Asl, Setarehdan, & Mohebbi, 2008) before compared to thresholds (Ghodrati, Murray, & Marinello, 2008; Islam, Ammour, Alajlan, & Aboalsamh, 2016; Kikillus et al., 2007; J. Lee, Reyes, McManus, Maitas, & Chon, 2013; Lian et al., 2011; Logan & Healey, 2005; G. B. Moody & Mark, 1983; Tateno & Glass, 2001), assessed with logical conditions (J. Lee, Nam, et al., 2013; Tsipouras, Fotiadis, & Sideris, 2005), or a combination of both (Dash et al., 2009). In two works, Support Vector Machine (SVM) was used as final step (Asl et al., 2008; Y. Li et al., 2017).

The algorithms and constraints were created using disjunct training data (Dash et al., 2009; Islam et al., 2016; Lake & Moorman, 2011; J. Lee, Nam, et al., 2013; J. Lee, Reyes, et al., 2013; McManus et al., 2013; Petr nas et al., 2015; Tateno & Glass, 2001; Zhou et al., 2014), by repeatedly dividing the available data in training and testing dataset (Asl et al., 2008), by leave-one-record-out procedure (Y. Li et al., 2017), or by using the same data as the testing data (Dash et al., 2009; Ghodrati et al., 2008; Huang et al., 2011; Islam et al., 2016; Kikillus et al., 2007; J. Lee, Nam, et al., 2013; J. Lee, Reyes, et al., 2013; G. B. Moody & Mark, 1983; Tateno & Glass, 2001). In some works, it was not reported how the algorithm was generated (Lian et al., 2011; Logan & Healey, 2005; Tsipouras et al., 2005).

Table 2.2: Algorithms for atrial fibrillation (AF) or arrhythmia detection using RR intervals. Sensitivity SE , specificity SP , positive predictivity $+P$, and accuracy ACC are given with the key techniques, the required number of RR intervals (N), the training, and the test dataset.

Authors	Key techniques	Training dataset	Test dataset	N	SE	SP	+P	ACC
G. B. Moody and Mark (1983)	Arithmetic mean of scores series, filtering, thresholding	AFDB	AFDB	20	90.7	82.4	nr	nr
Tateno and Glass (2001)	Density histograms of ΔRR , Kolmogorov-Smirnov test, thresholding	AFDB	AFDB	100	94.4	97.2	96.0	nr
			MTDB 100		na	99.0	na	nr
			MTBH 200		88.2	87.6	62.4	nr
Logan and Healey (2005)	Normalizing RR intervals, variance, thresholding, smoothing	nr	AFDB ¹	600	96.0	89.0	nr	nr
Tsipouras et al. (2005)	Knowledge-based system	nr	MTDB ²	3	96.0	95.9	93.2	94.3
Kikillus et al. (2007)	ΔRR , histogram, RR intervals differing at least 200 ms from previous RR interval, thresholding	AFDB, NSR	AFDB, NSR	3501	94.1	93.4	nr	nr

nr: not reported; DB: database; AFDB: MIT-BIH AF DB; MTDB: MIT-BIH Arrhythmia DB; MTDB 100 (200): MITDB 100er (200er) series; NSR: MIT-BIH Normal Sinus Rhythm DB; CUVT: Creighton University Ventricular Tachyarrhythmia DB; DMAF: Draeger Medical AF DB; SCD: Smartphone Camera data; CL: Clinical

RMSSD: root mean square of successive differences

¹Records “00735” and “03665” excluded; ²Results averaged; ³1367 segments; ⁴Records “102”, “104”, “107”, and “217” excluded; ⁵Records “04936” and “05091” excluded

Table 2.2: Algorithms for atrial fibrillation (AF) or arrhythmia detection using RR intervals (continued).

Authors	Key techniques	Training dataset	Test dataset	N	SE	SP	+P	ACC
Asl et al. (2008)	Time & frequency domain & nonlinear features reduced by generalized discriminant analysis, Support Vector Machine (SVM)	Train (2/3) & test (1/3) set; 100 repetitions	MTDB, CUVT ³	32	95.8	99.4	93.6	99.2
Ghodrati et al. (2008)	Normalized absolute difference with interval constraints, thresholding	AFDB, MTDB, DMAF	DMAF AFDB MTDB ⁴	29	87.0 89.0 92.0	nr nr nr	94.0 87.0 73.0	nr nr nr
Dash et al. (2009)	RMSSD, turning point ratio, Shannon entropy, thresholding, logical condition	AFDB	AFDB ⁵ MTDB 100 MTDB 200	128	94.4 nr 96.5	95.1 96.2 69.4	nr nr nr	99.1 na 100.0
Huang et al. (2011)	Filtered peak in Δ RR interval distribution difference curve, histogram analysis, standard deviation analysis, numbering abberant rhythms recognition, Kolmogorov-Smirnov test	AFDB	AFDB	100	96.1	98.1	nr	nr

nr: not reported; DB: database; AFDB: MIT-BIH AF DB; MTDB: MIT-BIH Arrhythmia DB; MTDB 100 (200): MITDB 100er (200er) series; NSR: MIT-BIH Normal Sinus Rhythm DB; CUVT: Creighton University Ventricular Tachyarrhythmia DB; DMAF: Draeger Medical AF DB; SCD: Smartphone Camera data; CL: Clinical

RMSSD: root mean square of successive differences

¹Records “00735” and “03665” excluded; ²Results averaged; ³1367 segments; ⁴Records “102”, “104”, “107”, and “217” excluded; ⁵Records “04936” and “05091” excluded

Table 2.2: Algorithms for atrial fibrillation (AF) or arrhythmia detection using RR intervals (continued).

Authors	Key techniques	Training dataset	Test dataset	N	SE	SP	+P	ACC
Lake and Moorman (2011)	Coefficient of sample entropy	AFDB	U Virginia dataset	11	91.0	98.0	63.0	nr
Lian et al. (2011)	RdR Map: Map of RR intervals vs. Δ RR, 2-D grid, nonempty cells, thresholding	nr	AFDB MTDB	32	94.3 98.1	95.1 77.0	nr nr	nr nr
J. Lee, Reyes, et al. (2013)	Sample entropy, thresholding	AFDB ⁵ , NSR	AFDB ⁵ SCD	63	97.3 92.6	95.9 99.8	nr nr	96.1 95.5
J. Lee, Nam, et al. (2013)	Ectopic beat elimination, frequency variations-time varying coherence functions with Shannon entropy, logical condition	AFDB ⁵	Cl. AF DB AFDB ⁵ MTDB 100 MTDB 200 SCD	13	93.9 94.7 na 92.4 91.4	84.4 90.4 91.8 76.5 93.9	90.5 87.9 na 45.2 nr	90.1 92.2 91.8 79.2 92.5
McManus et al. (2013)	RMSSD/mean + Shannon Entropy	AFDB, NSR	SCD	63	96.2	97.5	nr	96.8

nr: not reported; DB: database; AFDB: MIT-BIH AF DB; MTDB: MIT-BIH Arrhythmia DB; MTDB 100 (200): MITDB 100er (200er) series; NSR: MIT-BIH Normal Sinus Rhythm DB; CUVT: Creighton University Ventricular Tachyarrhythmia DB; DMAF: Draeger Medical AF DB; SCD: Smartphone Camera data; Cl.: Clinical

RMSSD: root mean square of successive differences

¹Records “00735” and “03665” excluded; ²Results averaged; ³1367 segments; ⁴Records “102”, “104”, “107”, and “217” excluded; ⁵Records “04936” and “05091” excluded

Table 2.2: Algorithms for atrial fibrillation (AF) or arrhythmia detection using RR intervals (continued).

Authors	Key techniques	Training dataset	Test dataset	N	SE	SP	+P	ACC
Zhou et al. (2014)	Nonlinear filter, integer filters, symbolic dynamics, Shannon entropy	Long-term AF database	MTDB AFDB	127	97.3 96.9	90.8 98.3	55.3 97.6	91.5 97.7
Petrénas et al. (2015)	Filtering, RR interval irregularity, bigeminy suppression, signal fusion and detection, online / offline	Long-term AF database	AFDB AFDB ¹ AFDB ⁵	15	96.7 96.6 97.5	97.9 97.9 97.8	nr nr nr	nr nr nr
Islam et al. (2016)	Heartbeat duration with affine-normalization, thresholding	AFDB	AFDB MTDB	30	95.8 86.2	95.5 85.9	94.1 41.5	95.6 86.0
Y. Li et al. (2017)	Δ RR are plotted over time, probability density analysis, blank pane ratio, information entropy, SVM	nr and leave-one-record-out for determining SVM parameter	AFDB MTDB	40 s	95.3 94.5	96.3 90.0	94.1 55.0	95.9 90.6

nr: not reported; DB: database; AFDB: MIT-BIH AF DB; MTDB: MIT-BIH Arrhythmia DB; MTDB 100 (200): MITDB 100er (200er) series; NSR: MIT-BIH Normal Sinus Rhythm DB; CUVT: Creighton University Ventricular Tachyarrhythmia DB; DMAF: Draeger Medical AF DB; SCD: Smartphone Camera data; Cl.: Clinical

RMSSD: root mean square of successive differences

¹Records “00735” and “03665” excluded; ²Results averaged; ³1367 segments; ⁴Records “102”, “104”, “107”, and “217” excluded; ⁵Records “04936” and “05091” excluded

In Table 2.2, sensitivity SE , specificity SP , positive predictivity $+P$, and the accuracy ACC are reported. The presented results are reduced to the MIT-BIH Atrial Fibrillation database (Goldberger et al., 2000; G. B. Moody & Mark, 1983), the MIT-BIH Arrhythmia database (Goldberger et al., 2000; G. Moody & Mark, 2001), or own databases. If own databases were used, the results are given with these (Ghodrati et al., 2008; Lake & Moorman, 2011; J. Lee, Nam, et al., 2013; J. Lee, Reyes, et al., 2013; McManus et al., 2013). The performance on additional databases like the MIT-BIH Normal Sinus Rhythm Database (Goldberger et al., 2000) were disregarded in this overview. The results are given with the suggested number of RR intervals or in case of no suggestion, with the minimum number of RR intervals. The required RR intervals ranged from three RR intervals (Tsipouras et al., 2005) to 3501 RR intervals (Kikillus et al., 2007). For a real-time application, an algorithm with minimum RR intervals is required.

None of the found literature integrated features from RSA. RSA gives the connection between heart rhythm and breathing. RSA was investigated during metronomic/paced breathing (Bowers & Murray, 2004; Lamine et al., 2004; Leutheuser et al., 2011). Apart from that, the presence of RSA is a physiological fact even if a person is not performing intentional breathing. In studies related to RSA, RR intervals are modeled with a sine curve having the same frequency as the ordered breathing.

In chapter 6 of the mobile ECG analysis part, it is investigated how features based on RSA or HRV, are beneficial in detecting arrhythmic beats (contribution (III)). The used features are solely based on the RR interval signal. The idea was to model a short number of consecutive RR intervals N ($N = 10, 15$, or 20) using a sine curve. The required frequency of the sine curve was calculated from the corresponding RR intervals. It is assumed that features from the sine curve fit will increase the performance for arrhythmia classification when using only RR intervals. For a real-time application, an algorithm with minimum heartbeats is required. In this thesis, a method is suggested to detect the $(N+1)$ th arrhythmic beat using a short number of preliminary consecutive RR intervals. The short number of consecutive RR intervals would fulfill the prerequisite of a real-time application.

2.2.3 P- and T-wave Detection

A wide variety of ECG fiducial points detection algorithms exists in literature (Akhbari, Shamsollahi, Jutten, Armoundas, & Sayadi, 2016; Akhbari, Shamsollahi, Sayadi, Armoundas, & Jutten, 2016; Bayasi et al., 2014; Boichat, Khaled, Rincón, & Atienza, 2009; Bote, Recas, Rincón, Atienza,

& Hermida, 2018; Cosoli et al., 2015; Di Marco & Chiari, 2011; Dumont, Hernández, & Carrault, 2010; Elgendi, Eskofier, & Abbott, 2015; Elgendi, Meo, & Abbott, 2016; Friganovic, Kukolja, Jovic, Cifrek, & Krstacic, 2018; Ghaffari, Homaeinezhad, Khazraee, & Daevaeiha, 2010; Hu, Liu, Wang, Xiao, & Yao, 2014; Jané, Blasi, García, & Laguna, 1997; Laguna, Jané, & Caminal, 1994; Lenis et al., 2016; C. Li, Zheng, & Tai, 1995; C. Lin et al., 2014; C. Lin, Mailhes, & Tournieret, 2010; A. Martínez, Alcaraz, & Rieta, 2010; J. P. Martínez et al., 2004; Mazomenos et al., 2013; Panigraphy & Sahu, 2018; Rincón, Recas, Khaled, & Atienza, 2011; Sayadi & Shamsollahi, 2009; Soria-Olivas et al., 1998; Sun, Chan, & Krishnan, 2005; Trahanias & Skordalakis, 1990; Vila et al., 2000). They were built on different (mathematical) models and implementation procedures.

Friganovic et al. (2018) identified four processing steps for the detection of characteristic waves in the ECG signal: noise and artefacts reduction, transformations, fiducial marks selection of wave candidates, and decision rules. They combined, compared, and mixed these four steps using five underlying algorithmic concepts from literature to detect QRS complexes, and P- and T-wave peaks (Elgendi, 2013; Elgendi et al., 2016; A. Martínez et al., 2010; J. P. Martínez et al., 2004; Pan & Tompkins, 1985; Sun et al., 2005). The work of Pan and Tompkins (1985) and Elgendi (2013) solely concentrated on QRS complex detection. Elgendi et al. (2015) detected T-wave peaks incorporating clinical knowledge. This resulted in a computationally efficient algorithm suitable for wearables. One year later, Elgendi et al. (2016) revised this algorithm to additionally detect P-wave peaks. The clinical knowledge was substituted with prior information analysis. The algorithm of Sun et al. (2005) consisted of four steps: (1) preprocessing, (2) calculation of the multiscale morphological derivative function, (3) determining local minima and maxima in the multiscale morphological derivative function, and (4) classification of detected waves. A. Martínez et al. (2010) transformed the ECG time domain signal into the phasorial signal in the complex domain using the phasor transform. This had the effect that the ECG waves were enhanced, meaning that the small variations of P- and T-waves were maximized to facilitate their detection and delineation. J. P. Martínez et al. (2004) presented a single-lead ECG delineation algorithm based on the Wavelet transform.

The next paragraph summarizes Wavelet based algorithms for ECG fiducial points detection. The work of J. P. Martínez et al. (2004) is a generalization of the algorithm described by C. Li et al. (1995). The algorithm of J. P. Martínez et al. (2004) is explained in chapter 7.1.3. Rincón et al. (2011) and Boichat et al. (2009) described algorithms based on the work of J. P. Martínez et al. (2004). The work of Rincón et al. (2011) provided an

update to the work of Boichat et al. (2009). Boichat et al. (2009) presented a single-lead algorithm that incorporated a 'genie' selection for performance assessment using the results of two individual leads. Rincón et al. (2011) compared this single-lead algorithm, independently applied to each lead, to two multi-lead versions of this algorithm. In the multi-lead versions, the two single leads are combined into one multi-lead signal using the root mean square value. This combination enabled the overall view of the cardiac phenomena. The two versions had differences in the applied baseline wander corrections – either Cubic Spline Baseline Estimation or Morphological Filtering were applied. Lenis et al. (2016) applied the phase free stationary Wavelet transform to detect the beginning and peak of the P-wave. For the annotation of the training dataset, they used the intra-cardiac signal that was simultaneously measured with the surface ECG (lead II). Di Marco and Chiari (2011) presented a low-computational complexity algorithm with the underlying concept of Wavelet transform. Computational complexity was reduced by using only two scales and few local coefficients instead of the root mean square value. Dumont et al. (2010) used an evolutionary algorithm to adjust the relevant parameters like filter cutoff frequencies, threshold levels, and time windows. Their delineation algorithm was based on a Wavelet-transformed signal using the algorithm from J. P. Martínez et al. (2004). Mazomenos et al. (2013) created a low-complexity hybrid feature extraction algorithm, incorporating the discrete Wavelet transform, and the time-domain morphology.

Bote et al. (2018) suggested a modular fiducial points detection algorithms with low computational costs and low mathematical complexity. The peaks module detected the QRS peak and, optionally, the P- and T-peaks. If switched on, the On-End module detected the beginning and end of the three fiducial events.

Panigraphy and Sahu (2018) used a differential evolution algorithm with extended Kalman smoother framework to detect on, peak, and off positions of P- and T-waves of the ECG signal.

C. Lin et al. (2010) performed a simultaneous estimation of the P- and T-wave fiducial points and waveforms based on a hierarchical Bayesian model. This model incorporated for the prior distribution a modified Bernoulli-Gaussian sequence and a partially collapsed Gibbs sampler as a special Markov chain Monte Carlo method for the posterior distribution. Using the underlying concept, C. Lin et al. (2014) proposed a revised version of this algorithm. C. Lin et al. (2014) suggested two sequential Bayesian models on a beat-to-beat basis by sequentially estimating the waveforms of each beat. Sequentially meant to incorporate information from one or more previous beats as required prior information. The

unknown parameters of these Bayesian models were either calculated using a block Gibbs sampler or a marginalized particle filter as a sequential Monte Carlo method.

Bayasi et al. (2014) suggested an adaptive delineation algorithm for P- and T-waves based on signal filtering, QRS complex detection, backward and forward search windows, and adaptive thresholds.

Sayadi and Shamsollahi (2009) used a model-based Bayesian framework building on a modified extended Kalman filter algorithm where parameters of the ECG dynamical model were integrated. The new extended Kalman filter structure was formulated by a simple autoregressive model for each of the 15 dynamic parameters of the Gaussians. Akhbari, Shamsollahi, Jutten, et al. (2016) improved this concept and suggested a new dynamic model where each ECG beat was a combination of seven Gaussian functions with three adjustable parameters. This nonlinear model was determined using an extended Kalman filter as well.

Akhbari, Shamsollahi, Sayadi, et al. (2016) suggested a multi hidden Markov model. In this model, a separated hidden Markov model is considered for each ECG wave or segment and the parameters of each hidden Markov model are determined separately.

Vila et al. (2000) detected and delineated the T- and U-waves using first mathematical modeling of the relevant ECG segment, followed by a threshold comparison technique. The two waves were modeled one after the other. Input for the modeling was the relevant segment of the ECG. The threshold comparison technique required the modeled signal, its first and second derivative, and appropriate thresholds that were included in forward and backward searches.

Trahanias and Skordalakis (1990) investigated the effect of primitive pattern extraction (e.g. peak, straight line segment, parabolic segment) in the ECG signal. This is also called syntactic pattern recognition.

Apart from these single lead ECG fiducial points detection algorithms, algorithms using two, three, or more leads were also suggested in literature (Ghaffari et al., 2010; Laguna et al., 1994). As the focus of this thesis are single-lead algorithms, only two are mentioned in this paragraph. Laguna et al. (1994) used multilead ECG signals to detect fiducial points in the ECG signal. They used the first derivative, detected zero-crossings and peaks within certain intervals and above distinct thresholds to detect relevant peaks and to delineate the different waves. Jané et al. (1997) used the single-lead version of the algorithm by Laguna et al. (1994) to assess the performance on the QT database. This algorithm is known as low-pass differentiation algorithm. Ghaffari et al. (2010) proposed a three-lead ECG detection-delineation algorithm. First, data of each lead were

preprocessed using discrete Wavelet transform. Second, a statistic called Multiple Higher Order Moments was introduced to calculate features on a sample to sample sliding window. The features for each lead were combined and decision statistics detected and delineated the different ECG events.

In most algorithms presented in literature, search windows for ECG delineation are needed to accurately detect the P- and T-wave peak (Akhbari, Shamsollahi, Sayadi, et al., 2016; Beraza & Romero, 2017; Bote et al., 2018; Di Marco & Chiari, 2011; Ghaffari et al., 2010; Jané et al., 1997; Laguna et al., 1994; C. Li et al., 1995; A. Martínez et al., 2010; J. P. Martínez et al., 2004; Sun et al., 2005). Often, the exact definition of these search windows is not given, making it impossible to reimplement the respective algorithm (Akhbari, Shamsollahi, Sayadi, et al., 2016; Beraza & Romero, 2017; Di Marco & Chiari, 2011; Ghaffari et al., 2010; Jané et al., 1997; Laguna et al., 1994; C. Li et al., 1995; A. Martínez et al., 2010; J. P. Martínez et al., 2004; Sun et al., 2005). For the sake of completeness, not all algorithms presented in literature need search windows (Akhbari, Shamsollahi, Jutten, et al., 2016; Elgendi et al., 2015; Elgendi et al., 2016; C. Lin et al., 2014; Panigraphy & Sahu, 2018).

In long-term ECG recordings, it is common to acquire ECG data from the three leads of Einthoven (I, II, and III). The different waves and fiducial points are clearly delineated in lead II (Figure 3.1). Advantages of single-lead ECG fiducial points detection algorithms are the reduction of necessary electrodes, fewer measurement equipment components, and an enhanced patient comfort. In literature, there exists no ready-to-use, real-time, single-lead ECG fiducial points detection algorithm that was evaluated on a large database.

This is addressed in this thesis with contribution (IV) (chapter 7). In this thesis, three single-lead ECG fiducial points detection algorithms are presented. The reference MATLAB (Mathworks Inc., Natick, MA, US) implementations are available for download (<https://github.com/HeikeLeutheuser/WaveED/>) to facilitate the use for other research groups. All algorithms are suited for real-time wave event detection. The three wave event detection algorithm are: First, a new P- and T-wave detection algorithm is suggested that uses the basic concept of fiducial points detection by Hu et al. (2014) based on line-fitting. Second, a real-time version of the Wavelet based algorithm for P- and T-wave detection by J. P. Martínez et al. (2004) is given. Third, both algorithms are compared to a naive, knowledge-incorporating fiducial points detection algorithm. The performance of the three ECG fiducial points detection algorithms are evaluated using the MIT-BIH Arrhythmia database (Goldberger et al.,

2000; G. Moody & Mark, 2001) with annotations provided by Elgendi et al. (2015), the QT database (Laguna, Mark, Goldberger, & Moody, 1997), and the short-term FAU database – a self-recorded clinical database – with annotations provided by a cardiologist.

2.3 IMU-based Activity Recognition

2.3.1 Assessment of Physical Activity

Physical activity describes all body movements by skeletal muscles that result in energy expenditure (Caspersen, Powell, & Christenson, 1985). It is either quantified by energy expenditure measured in kilocalories or by using the Metabolic Equivalent of Task (MET) of the activity (Preece, Goulermas, Kenney, et al., 2009).

Physical activity can be grouped in four dimensions and four domains (Strath et al., 2013). The four dimensions are (1) type of activity, (2) frequency, (3) duration, and (4) intensity of performed activity (Strath et al., 2013; Warren et al., 2010). The four domains are (1) occupational, (2) domestic, (3) transportation, and (4) leisure time (Strath et al., 2013).

Several methodologies for the assessment of physical activity exist. They can be categorized in either subjective or objective methods. Table 2.3 gives an overview of the methodologies, including associated strengths and weaknesses. The following body-worn sensors or methodologies were not regarded in the table (Preece, Goulermas, Kenney, et al., 2009): inclinometers, goniometers, skin temperature, galvanic skin resistance, humidity, respiratory rate, tilt switches, foot pressure switches, and pressure sensitive insoles. Reason for this is that these sensors or methodologies have a minor impact in literature compared to IMUs (Preece, Goulermas, Kenney, et al., 2009). Summarizing, subjective methodologies induce difficulties regarding reliability, validity, and sensitivity (Table 2.3). There is an ongoing trend to use objective methodologies, like small and light-weight wearable IMUs, for physical activity research (Troiano, McClain, Brychta, & Chen, 2014). In combination with machine learning algorithms, these sensors provide a reliable and objective measurement of physical activity (S. Liu, Gao, & Freedson, 2012), e.g. in the recognition of different activities. In this thesis, IMUs comprising triaxial accelerometers and triaxial gyroscopes were chosen for data acquisition (section 8.1.1).

Table 2.3: Subjective and objective measures for the assessment of physical activity. This overview is based on the work of Strath et al. (2013) and Warren et al. (2010) and extended with additional literature, stated correspondingly and marked in *italic*.

Methods	Strengths (+)	Weaknesses (-)
Subjective Methods		
Physical activity questionnaires (paper-based)	+ Single time point assessment + Ranking into high and low categories	- Unclear validity for physical activity energy expenditure estimation
Physical activity diaries/logs (paper-based)	+ Not subjected to memory or recall as much as other subjective methods + Good subjective measure of physical activity and energy expenditure	- High burden on patients and participants - Complex and time-consuming data reduction and analysis
Electronic versions of questionnaires or diaries / logs	+ <i>Avoidance of secondary data entry errors and easier implementation of skip patterns (Coons et al., 2009)</i> + Easy standardization, cost, and time saving + Immediate scoring, reporting, and interpreting results + More honest reporting owing to greater feeling of anonymity and <i>high patient acceptance (Coons et al., 2009)</i>	- <i>Evidence needed to ensure equivalence with paper-based measure (Coons et al., 2009)</i>

Table 2.3: Subjective and objective measures for the assessment of physical activity (continued).

Methods	Strengths (+)	Weaknesses (–)
Combined remarks for subjective methods	+ Assessing different dimensions and domains + Low cost + Low burden + Quantitative and qualitative information + Simple data collection and analysis + Suitable for all populations	– Requirement of population and culture specific version – Reporting bias (<i>Trost & O'Neil, 2014</i>) – Difficulties in determining the frequency, duration, and intensity of physical activity – Social desirability bias
Objective Methods		
Indirect Calorimetry	+ Accurate + Criterion measure of physical activity and energy expenditure	– Expensive (<i>S. Liu, Gao, & Freedson, 2012</i>) – Technical equipment and trained personnel required (<i>S. Liu, Gao, & Freedson, 2012</i>) – Short time assessment only permissible
The Doubly Labeled Water Method	+ “Gold standard” measure of energy expenditure for total daily energy expenditure in free-living individuals + Low to moderate burden + Suitable for all populations	– Expensive (<i>S. Liu, Gao, & Freedson, 2012</i>) – Technical equipment and trained personnel required (<i>S. Liu, Gao, & Freedson, 2012</i>) – Resting metabolic rate and thermic effect of food required to derive physical activity energy expenditure – Unable to discern dimensions or domains

Table 2.3: Subjective and objective measures for the assessment of physical activity (continued).

Methods	Strengths (+)	Weaknesses (–)
Pedometers	+ Low cost + Low burden + Motivational tool, e.g. when giving a step goal (<i>Trost & O'Neil, 2014</i>) + <i>Most common used assessment tools for objective assessment of physical activity (Trost & O'Neil, 2014)</i> + Suitable for all populations + Easy data collection and analysis	– Simple pedometers cannot measure intensity/duration – No measurement of mode/type – Not accurate for energy expenditure – <i>Validity of the step estimates, distance travelled, or energy expended varies considerably across makes and models of pedometers (Trost & O'Neil, 2014)</i> – Specifically designed to assess walking only, inability to record nonlocomotor movements
Single accelerometer used in activity counts accelerometry / actigraphy	+ Concurrent measure of movement + <i>Provides detailed intensity, frequency, and duration data for weeks at a time</i> + Low burden + Relatively inexpensive + Relationship of activity counts and energy expenditure across many, but not all, activities (<i>Trost & O'Neil, 2014</i>) + Suitable for all populations + Objective indicator of body movement (acceleration) + Relatively easy data collection	– Cannot account for all activities such as cycling, stair use, or activities that require lifting a load (<i>Trost & O'Neil, 2014</i>) – Data reduction, transformation, and processing take time – Cut-point conundrum: activity-dependent relationship of activity counts and energy expenditure (<i>Trost & O'Neil, 2014</i>) – Underestimation of total or physical activity energy expenditure (<i>Trost & O'Neil, 2014</i>) – Requirement to measure over multiple days as habitual activity varies for individual days – Challenge of managing missing data

Table 2.3: Subjective and objective measures for the assessment of physical activity (continued).

Methods	Strengths (+)	Weaknesses (-)
Fitness trackers or smartwatches	+ <i>Most often additional functionality compared to pedometers, e.g. activities categorized by Metabolic Equivalent of Task (MET) values (Jensen, Leutheuser, et al., 2015)</i> + <i>Many of them are commercially available</i> + <i>Motivation tool (Jensen, Leutheuser, et al., 2015)</i> + <i>Different emphasis, e.g. sports tracking or sleep statistics</i>	- <i>Costs dependent on functionality / battery runtime</i> - <i>Black box algorithm for most commercial available trackers</i>
Inertial measurement units (IMUs)	+ <i>Sensors (accelerometers, gyroscopes, optional: magnetometers) placed at one or more body locations</i> + <i>Improvement and additional functionality using pattern recognition or machine learning compared to single accelerometer alone (S. Liu, Gao, & Freedson, 2012)</i>	- <i>Challenge of managing missing data minimized for combining sensor data</i> - <i>Relying on number and placement of sensors</i> - <i>Depending on device, technical expertise is essential</i>

Table 2.3: Subjective and objective measures for the assessment of physical activity (continued).

Methods	Strengths (+)	Weaknesses (–)
Heart Rate Monitors	+ Linear relationship of heart rate and energy expenditure during steady-state exercise (<i>Trost & O'Neil, 2014</i>) + <i>Calculation of heart rate indices (Trost & O'Neil, 2014)</i> + Assessing time from moderate and vigorous activities + Suitable for all populations + Easy and quick data collection + Relatively cheap	– Affected by nonactivity stimuli (emotion, caffeine) – <i>Influenced by age, body size, proportion of muscle mass utilised, cardiorespiratory fitness (Trost & O'Neil, 2014)</i> – Weak relationship at low end of intensity realm – Subject to interference with signal – <i>Burden of electrodes or chest straps (Trost & O'Neil, 2014)</i> – Only useful for aerobic activities
Multisensing units (e.g. IMUs and heart rate monitors)	+ Strengths are combined, thereby negating some of the disadvantages of each method used alone + Non-correlated measurement error	– Data analysis more complex – Monitors relatively expensive
Combined remarks for objective methods	+ Not relying on information provided by the subject + <i>Usage under conditions of free living with minimal inconvenience to the user (Preece, Goulermas, Kenney, et al., 2009)</i>	

2.3.2 Activity Classification Algorithms

Data acquired with IMUs need to be processed and analysed with machine learning algorithms for the recognition of different activities (Table 2.3). After data collection, the data need to be divided in appropriate epochs (windowing). This refers also to the preprocessing step. Then, features are defined, calculated, and selected. As last step, machine learning is applied to recognize activities. Table 2.4 gives an overview of these steps for activity recognition using inertial sensors.

Table 2.4: Machine learning techniques for activity recognition. This overview is based on the work of Preece, Goulermas, Kenney, et al. (2009).

Technique	Characteristics
Windowing	
Sliding windows	No inter-window gaps, with or without overlap of adjacent windows Range of window sizes: 0.25 s - 30 s (S. Liu, Gao, John, et al., 2012) No requirement of pre-processing of the sensor signal Most commonly used windowing approach
Event-defined windows	Events are used to define successive windows Pre-processing required to locate specific events No (compulsory) fixed window size
Activity-defined windows	Each window corresponds to a different activity Dependent on times of activity change
Features	
Heuristic features	Related to specific aspects of individual movements or postures
Time-domain features	Statistical measures like mean, median, variance, skewness, kurtosis, interquartile range Means of low and high frequency components Measures of correlation like cross-correlation coefficient

Table 2.4: Machine learning techniques for activity recognition (continued).

Technique	Characteristics
Frequency-domain features	Preparation: transformation of time-domain signal into frequency domain using Fast Fourier Transform (FFT) Median frequency Subset of the different FFT coefficients Spectral energy (sum of the squared FFT coefficients) Frequency-domain entropy (normalized information entropy of the FFT components)
Wavelet analysis (time-frequency features)	Preparation: transformation of time-domain signal using continuous or discrete Wavelet transform Signal enhancement Identification of activity transition points Statistical measures from the approximation or detail coefficients at specific decomposition levels
Feature selection methods	Visual and statistical analysis Correlation-based: high within-class but low between-class correlation Forward-backward search Clustering
Dimensionality reduction methods	Combination of original features to define a new set of variables Principal component analysis Independent component analysis
Machine learning techniques	
Supervised learning	Algorithm is trained using a significant amount of fully labeled activity data Examples: Threshold-based classification, decision trees, k-Nearest Neighbor, artificial neural networks, Support Vector Machines, Naive Bayes and Gaussian mixture models, fuzzy logics, Markov chains and hidden Markov models
Unsupervised learning	Algorithm automatically identifies data clusters without activity labels provided for the training dataset Data clusters may correspond to single activities
Deep learning (LeCun, Bengio, & Hinton, 2015)	Algorithm is fed with raw data and automatically discovers the representations for classification Layers of features are not designed by engineers

2.3. *IMU-based Activity Recognition*

Each step from Table 2.4 can be combined with another step differently. Hence, a vast amount of possible activity classification algorithms exists (Preece, Goulermas, Kenney, et al., 2009). This number is additionally enlarged with different study setups like sensors, sensor placement, and performed activities. Table 2.5 provides an overview of daily life activity recognition algorithms focusing on the aspect of supervised learning. The table was created to show the vast range of activity recognition system without being complete. Only the number of activities was stated for each publication. The summarized activities are given in Table 2.6.

Table 2.5: Exemplary representation of supervised activity recognition systems considering daily life activities of healthy subjects.

Sensors & Placement	Sampling rate & Epoch size	Used features	Best classifier	Evaluation process
Mäntyjärvi, Himberg, and Seppänen (2001) (4 activities, 6 subjects)				
2 triaxial acc.: left and right hip	256 Hz & 2 s (75% overlap)	Power of Wavelet coefficients of order 5 to 8 (Daubechies mother Wavelet) after independent component analysis	Multilayer perception neural networks	Stratified 10-fold cv.
Bao and Intille (2004) (20 activities, 20 subjects)				
5 biaxial acc.: right hip, right wrist, left arm, right ankle, left thigh	76.25 Hz & 6.7 s (50% overlap)	Mean, energy, frequency domain entropy, correlation of the acc. signals	Decision tree classifier	Leave-one-subject-out cv.
Ravi, Dandekar, Mysore, and Littman (2005) (8 activities, 2 subjects)				
1 triaxial acc.: pelvic region	50 Hz & 5.12 s (50% overlap)	Mean, SD, energy, correlation of the acc. signals	Plurality voting	10-fold cv.
Allen, Ambikairajah, Lovell, and Celler (2006) (8 activities, 6 subjects)				
1 triaxial acc.: waist	45 Hz & 0.02 s (no overlap)	Separated body acc. and gravity components of the signal, first-order delta coefficients for the gravity component, a 3-3-3-5 shifted delta configuration for the body acc. component, signal magnitude area	Gaussian mixture models	None and leave-one-subject-out cv.

acc.: accelerometer/acceleration; cv.: cross-validation; SD: standard deviation; ECG: electrocardiogram

Table 2.5: Exemplary representation of supervised activity recognition systems considering daily life activities of healthy subjects (continued).

Sensors & Placement	Sampling rate & Epoch size	Used features	Best classifier	Evaluation process
Karantonis, Narayanan, Mathie, Lovell, and Celler (2006) (12 activities, 6 subjects)				
1 triaxial acc.: right hip	45 Hz & 1 s (no overlap)	Median filter, low pass filter, normalized signal magnitude area, tilt angle	Hierarchical, threshold classifier	None
Pärkkä et al. (2006) (21 activities, 16 subjects)				
2 acc.: chest, wrist (dominant); additional signals like altitude, ECG, temperature, heart rate	Chest: 200 Hz; wrist: 40 Hz & 1 s (no overlap)	Peak frequency, median, and peak power of up-down chest acc., variance of back-forth chest acc., sum of variances of three-dimensional wrist acc.	Automatically generated decision	Leave-one-subject-out cv.
Pober, Staudenmayer, Raphael, and Freedson (2006) (4 activities, 6 subjects)				
1 MTI Actigraph	1 Hz & 15 s (no overlap)	Second-by-second data	Hidden Markov Model	Leave-one-subject-out cv.
Salarian, Russmann, Vingerhoets, Burkhard, and Aminian (2007) (12 activities, 10 subjects)				
1 biaxial acc.: trunk; 3 uniaxial gyroscope: trunk, left and right shank	200 Hz & n.a.	Transition duration, minimum relative trunk tilt, range of antero-posterior tilt of the trunk, range, minimum, maximum, relative time of minimum and maximum peak of perpendicular acc., inclination, six fuzzy variables	Three-stages: logistic regression, thresholds, Mamdani fuzzy classifier (Mamdani, 1977)	Separate training and test dataset

acc.: accelerometer/acceleration; cv.: cross-validation; SD: standard deviation; ECG: electrocardiogram

Table 2.5: Exemplary representation of supervised activity recognition systems considering daily life activities of healthy subjects (continued).

Sensors & Placement	Sampling rate & Epoch size	Used features	Best classifier	Evaluation process
Preece, Goulermas, Kenney, and Howard (2009) (8 activities, 20 subjects)				
3 triaxial acc.: waist, thigh, ankle	64 Hz & 2 s (50% overlap)	Magnitude of first five components of Fast Fourier Transform (FFT) analysis	k-Nearest Neighbor	Leave-one-subject-out cv.
S. Liu, Gao, John, Staudenmayer, and Freedson (2012) (13 activities, 50 subjects)				
Triaxial acc.: hip (dominant), wrist (dominant); ventilation sensor	30 Hz & 30 s (no overlap)	Hip acc.: x-axis: SD, 25th percentile, spectral entropy; y-axis: spectral entropy; z-axis: SD, 90th percentile; wrist acc.: x-axis: 25th, 50th, 75th, 90th percentiles, SD, spectral energy, spectral entropy; y-axis: all time and frequency domain features, z-axis: all time domain features, spectral energy	Support Vector Machine (SVM) with radial basis function kernel	Leave-one-subject-out cv.
acc.: accelerometer/acceleration; cv.: cross-validation; SD: standard deviation; ECG: electrocardiogram				

Table 2.6: Summarized activities of the previous Table 2.5. The same 10 publications were considered (Allen, Ambikairajah, Lovell, & Celler, 2006; Bao & Intille, 2004; Karantonis, Narayanan, Mathie, Lovell, & Celler, 2006; S. Liu, Gao, John, Staudenmayer, & Freedson, 2012; Mäntyjärvi, Himberg, & Seppänen, 2001; Pärkkä et al., 2006; Pober, Staudenmayer, Raphael, & Freedson, 2006; Preece, Goulermas, Kenney, & Howard, 2009; Ravi, Dandekar, Mysore, & Littman, 2005; Salarian, Russmann, Vingerhoets, Burkhard, & Aminian, 2007). The prevalence of the individual activities is given in the third column (Amount). Either one or more instructions were given to the subject during data acquisition.

Activity	Instruction	Amount
Walking	Carrying items, fast, inclined treadmill, in restaurant, level treadmill, normal, slow, starting and stopping, to a bus stop, to and in a shop	10
Sitting	Drinking, eating, filing papers, quiet, reading, relaxing, simulated computer work, talking, waiting, watching TV, working on computer, writing	6
Sports	Basketball, circuit, hopping on the left and right leg, jogging, jumping, nordic walking, rowing, running, sit-ups, strength training, stretching, tennis, treadmill (4 mph inclined, 6 mph)	6
Stairs	Ascending, descending	5
Lying	Relaxing	5
Standing	Queueing, waiting	4
Daily life	Brushing teeth, combing hair, putting clothes on	4
Household	Folding laundry, moving boxes, scrubbing, vacuuming	4
Bicycling	No additional information, 1-kp and 2-kp resistance	3
Transitions	Lying-to-sit, lying-to-stand, sit-to-lying, sit-to-stand, stand-to-sit	2
Additional locomotion	Riding elevator, riding escalator, traveling in bus	2
Fall	Active, inactive, in chair	1

The different activity classification systems differed in

- the standard of the datasets (number of activities and number of subjects),
- the activities,
- the sensors, the number of sensors, and their placement,

- the sampling rate,
- the epoch duration and their overlap, and
- the evaluation process.

It is problematic to compare new recognition algorithms to existing algorithms in literature without considering the same data basis. In this thesis, the benchmark dataset called DaLiAc is introduced to enable the comparison of activity recognition algorithms (contribution (V), chapter 8). A hierarchical, multi-sensor based classification system is introduced and compared with reimplemented versions of state of the art classification systems (contribution (VI), chapter 8). The algorithms are evaluated on the DaLiAc dataset.

Chapter 3

Fundamentals

In this chapter, the underlying physiologic and algorithmic principles as basis for the following chapters are collected. First, the generation of the biosignal ECG is explained, which is the basis for the chapters 6 and 7. Second, the underlying principles for regression are summarized. In this thesis, regression is relevant for the chapters 4 and 5 in the adjustment algorithms for RIP and the chapters 6 and 7 in the fitting of sine curves or straight lines to RR intervals or ECG segments. Third, relevant aspects of supervised learning are collected. This consisted of the supervised learning pipeline, the underlying principles of five classifiers, and cross-validation for assessing the performance of different algorithms. Elements of supervised learning are relevant for all chapters.

3.1 The Electrocardiogram

The analysis of the ECG signal is the subject of contributions (III) and (IV). Both contributions belong to the part mobile ECG analysis that is divided in chapter 6 and chapter 7. The most known and investigated biosignal is the ECG signal (Rangayyan, 2015). With an ECG, the electric activity of the heart can be measured. An important fact of the ECG is that its morphology is changed by cardiovascular diseases and other abnormalities. Figure 3.1 shows the waves and fiducial points of a schematic, healthy ECG signal.

In the following, the underlying physiologic basis of the ECG is explained as well as how the ECG signal can be measured. The described material in this section comes from the textbook from Rangayyan (2015).

3.1.1 Physiologic Basis

The heart is a four-chambered pump. The two upper chambers, the left and right atrium, collect the blood. The two lower chambers, the left and right ventricle, pump the blood out of the heart during contraction. The heart has its own pacemaker cells that control the relaxation and contraction phases of the heart. The electric system of the heart is a complex system of nodes, bundles, and fibers that is controlled by the autonomous nervous system.

The electric systems starts with the firing of the sinoatrial node (Figure 3.1). This causes a slow-moving contraction, also called depolarization, of the atria and is visible in the P-wave. The P-wave has an amplitude of about 0.1 mV - 0.2 mV and a duration of about 60 ms to 80 ms. The first excitation wave of the sinoatrial node receives a delay at the atrioventricular node. This is visible as PQ segment in the ECG. Then the atrioventricular node fires. The stimulus travels rapidly through the His bundle, the bundle branches, and the Purkinje system of specialized conduction fibers to the ventricles. This causes a fast contraction, resulting in the QRS complex. The QRS complex has an amplitude of about 1 mV and a duration of about 80 ms. The ST segment comes from the plateau of the action potential of the ventricular myocyte. The T-wave is caused by the repolarization of the ventricles. The T-wave has an amplitude of 0.1 mV to 0.3 mV and a duration of 120 ms to 160 ms. After the ventricular relaxation, this process is repeated.

3.1.2 Measurement

The ECG signal can be recorded with surface electrodes. A 12-lead ECG recording is the standard in clinical practice. It requires four limb leads and six chest leads, in total ten electrodes. If just the four limb leads are used, then just three leads are recorded: lead I, lead II, and lead III. For this, one electrode is placed on the right wrist, the second one on the left wrist, and the third one on the left ankle. The reference electrode is placed on the right ankle. The equilateral triangle of the first three electrodes is known as Einthoven's triangle. The positions of the four electrodes on the limbs can also be transferred to placements on the torso. Then, the first three electrodes are placed in a triangle around the heart to form again the Einthoven triangle (Figure 3.2). The ECG of lead II is shown in Figure 3.1.

In chapter 6, arrhythmic beats are detected using the RR interval signal. The R-peak is the highest peak in the ECG signal (Figure 3.1) and the difference between two adjacent R-peaks gives one RR interval. For

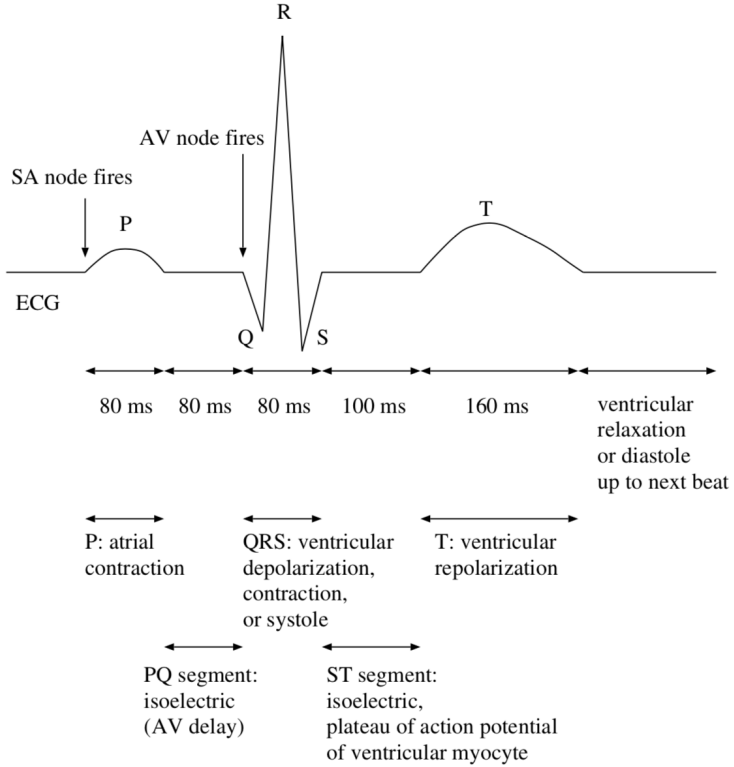


Figure 3.1: Waves and fiducial points of an electrocardiogram (ECG) signal of lead II according to Einthoven. The waves and fiducial points are abbreviated with single upper-case letters starting with the letter P. The duration of the waves and intervals are given with the underlying physiologic basis below the ECG signal. SA: sinoatrial; AV: atrioventricular. Reprinted from Figure 1.26, Rangayyan (2015, p. 28).

this, the R-peak or the QRS complex needs to be detected from the ECG signal. In this thesis, algorithms for QRS complex detection are stated in chapter 2.2.1. In chapter 7, algorithms for high-accurate P- and T-wave detection are presented and compared. These wave event detection algorithms in this thesis just need a single lead. The Short-Term FAU database is presented in chapter 7. The three Einthoven's lead were recorded with electrodes placed at the torso.

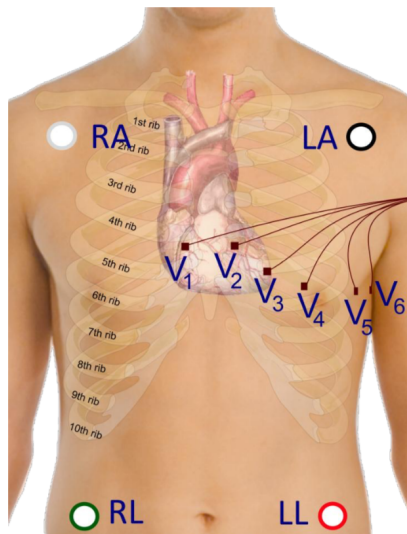


Figure 3.2: Electrode placement for electrocardiogram (ECG) measurement using electrodes on the torso. The four necessary electrodes to measure lead I, lead II, and lead III according to Einthoven are placed at the transferred position of the right arm (RA), the left arm (LA), the right leg (RL), and the left leg (LL). The RL-electrode serves as reference electrode. Placing electrodes at the positions V_1 , V_2 , ..., V_6 is needed, if a 12-lead clinical ECG should be recorded. Reprinted from Figure 3-2, Shimmer Research (2018, p. 9).

3.2 Regression

In regression, the aim is to find a functional description of existing data. Often, the goal is to predict afterwards values for new data. In this thesis, regression is part of the adjustment of RIP data (chapters 4 and 5), the fitting of the sine curve to RR intervals (chapter 6), and the fitting of straight lines to ECG segments for a wave event detection algorithm (chapter 7).

The described material in this section for single und multiple linear regression with LSQ comes from the textbooks from Knabner and Barth (2018) and Weisberg (2005). The material for SVR is based on the work of Smola and Schölkopf (2004) and Schölkopf and Smola (2002).

3.2.1 Simple Linear Regression

Simple linear regression is a linear regression model with a single independent variable, also called predictor, x and one dependent variable y . The simple linear regression model gives a relation between these two variables in $f(x)$:

$$f(x) = y = \beta_0 + \beta_1 x, \quad (3.1)$$

where β_0 and β_1 as parameters to be estimated. These are also called regression coefficients. This equation can be solved by using LSQ that is minimizing the residual sum of squares RSS :

$$RSS = \sum_{i=1}^N (\hat{y}_i - y_i)^2, \quad (3.2)$$

with $\hat{y}_i$ as predicted outcome and N available data samples.

The root mean square error (RMSE) indicates how well the fit is appropriate to the data:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N - k}}, \quad (3.3)$$

with k as number of regression coefficients in the model. In simple linear regression, k would be 2. The RMSE can be used for all kind of regression equations.

3.2.2 Multiple Linear Regression

Often, two or more predictors $x_1, x_2, \dots, x_p$ need to be taken into account for the dependent variable y .

The model for multiple linear regression with more than one predictor is the following:

$$f(x) = y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p. \quad (3.4)$$

This model can be written in matrix notation:

$$\mathbf{y} = \mathbf{X} \cdot \boldsymbol{\beta}. \quad (3.5)$$

A solution to this equation is given by the Moore-Penrose pseudoinverse $\mathbf{X}^+$:

$$\boldsymbol{\beta} = \mathbf{X}^+ \mathbf{y} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (3.6)$$

Calculating the Moore-Penrose pseudoinverse as a solution corresponds to minimizing the residual sum of squares (equation (3.2)). By calculating

the Moore-Penrose pseudoinverse for more than one predictor, this is analogous to LSQ.

Though, it is not recommended to solve multiple linear regression by just calculating the Moore-Penrose pseudoinverse $\mathbf{X}^+$. The results can be highly inaccurate due to errors in the data or the underlying model, or if $\mathbf{X}$ is close to singular or badly scaled. One approach is to calculate $\mathbf{X}^+$ by singular value decomposition (SVD) and considering a regularization method.

The SVD of $\mathbf{X}$, where $\mathbf{X}$ has $M = p + 1$ columns and N rows is:

$$\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{V}^t, \quad (3.7)$$

with $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_N]$, $\mathbf{S}$ as diagonal matrix with the singular values as diagonal elements, and $\mathbf{V}$ as eigenvector matrix. $\mathbf{u}_i$ is given by $\frac{1}{\sqrt{\lambda_i}} \mathbf{X} \mathbf{v}_i$, with λ_i as eigenvalues of $\mathbf{X}$, and $\mathbf{v}_i$ as eigenvectors of $\mathbf{X}$. Singular values are calculated by taking the root of the eigenvalues. Taking into account the SVD of $\mathbf{X}$, the Moore-Penrose pseudoinverse $\mathbf{X}^+$ is:

$$\mathbf{X}^+ = \mathbf{V} \mathbf{S}^+ \mathbf{U}^t. \quad (3.8)$$

If the condition number κ of the matrix $\mathbf{X}$ is too large in relation to the data errors, it is recommended to perform a regularization method. The regularization method considered in this thesis investigates κ and reduces its value. κ is calculated as fraction of the biggest singular value σ_1 to the smallest positive singular value σ_r :

$$\kappa = \frac{\sigma_1}{\sigma_r}. \quad (3.9)$$

Common approach using this regularization method is to eliminate the smallest singular value, that is setting this value to zero in the corresponding matrices, and solve the modified regression equation. Sometimes, it is beneficial to set even more than one singular value to zero.

In chapter 5, the best κ for the corresponding regression equation was determined from a range of values using 10-fold cross-validation (section 3.3.3). In chapters 4 and 6, κ was fixed to the maximum value 10^5 .

3.2.3 Support Vector Regression

The goal in SVR is to find a function $f(x)$ that is as flat as possible and has at most ϵ deviation from the dependent variable y . This means that all

errors up to ϵ are ignored when finding $f(x)$. This version of SVR is called ϵ -SVR. The linear function $f(x)$ with the following form is considered:

$$f(x) = \langle w, x \rangle + b, \quad (3.10)$$

with x as the predictors, and w and b as parameters to be estimated. $f(x)$ should be as flat as possible. This means searching for a small w . Minimizing the norm $\|w\|^2$ ensures this. This leads to the following optimization problem:

$$\text{minimize } \frac{1}{2} \|w\|^2 + c \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (3.11)$$

$$\text{subject to } \begin{cases} y_i - \langle w, x_i \rangle - b & \leq \epsilon - \xi_i \\ b + \langle w, x_i \rangle - y_i & \leq \epsilon - \xi_i^* \\ \xi_i, \xi_i^* & \geq 0. \end{cases} \quad (3.12)$$

The slack variables ξ_i, ξ_i^* are required to realize the optimization problem. The trade-off between the flatness of $f(x)$ and the error tolerance ϵ is given by the constant c . This means to deal with the ϵ -insensitive loss function $|\xi|_\epsilon$:

$$|\xi|_\epsilon := \begin{cases} 0 & \text{if } |\xi| \leq \epsilon \\ |\xi| - \epsilon & \text{otherwise.} \end{cases} \quad (3.13)$$

This is visualized in Figure 3.3 for a linear function $f(x)$. $f(x)$ consists of a tube with radius ϵ that is fitted to the data. All data points being inside the shaded region, do not enter equation (3.11).

The linear case can be transferred to the nonlinear case with the use of kernels. Then, the optimization problem corresponds to finding the flattest function in feature space, not input space.

Examples of four basic kernels K are:

- Linear: $K(x_i, x_j) = \langle x_i, x_j \rangle$
- Radial basis function: $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \gamma > 0$
- Sigmoid: $K(x_i, x_j) = \tanh(\gamma \langle x_i, x_j \rangle + r)$,
- Polynomial: $K(x_i, x_j) = (\gamma \langle x_i, x_j \rangle + r)^d, \gamma > 0$

with γ , r , and d as kernel parameters.

In this thesis, the LIBSVM implementation using version 3.22 (Chang & Lin, 2011) was integrated in the MATLAB package (Mathworks Inc.,

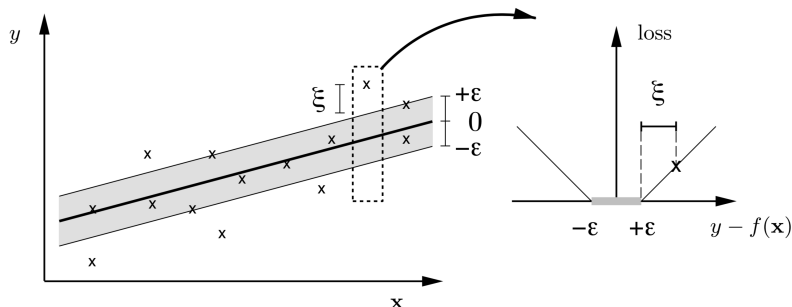


Figure 3.3: Visualization of Support Vector Regression (SVR). It is desired to obtain $f(x)$. This is the shaded region with radius ϵ in the left plot. All data points within this shaded region are not taken into account for the optimization problem (equations (3.11) and (3.12)). The slack variable ξ_i respectively the ϵ -insensitive loss function (right plot) is needed to realize finding $f(x)$. Reprinted from Schölkopf and Smola (2002).

Natick, MA, US) to conduct the SVR computations. All of these four kernels are available in the LIBSVM implementation and were compared when performing the SVR in chapter 4. The parameters c and ϵ were determined in a grid search using 10-fold cross-validation (chapter 3.3.3). The kernel parameters are given in the corresponding chapter.

3.3 Supervised Learning

Elements of supervised learning can be found in all chapters. The next chapters described the general steps of a supervised learning system using the supervised learning pipeline. The underlying principles of the five classifiers used in this thesis are described. Cross-validation provides an approach to assess the performance of the algorithms regarding their generalization capabilities on new data. The described material in this chapter comes from the textbook of Duda, Hart, and Stork (2000).

3.3.1 Supervised Learning Pipeline

The four steps of a supervised learning system are comprised in the supervised learning pipeline – also called pattern recognition or classification pipeline. Figure 3.4 shows a version of this pipeline adapted to the content of this thesis (Duda et al., 2000).



Figure 3.4: Steps of a supervised learning system. Adapted from Figure 1.1, Duda, Hart, and Stork (2000, p. 2).

First, the input of the system is of interest. The input consists of data from databases or own recorded data using various input sources. In this thesis for example, the PhysioNet website (Goldberger et al., 2000) was used to get access to two databases for assessing the performance of the arrhythmia classification with RR intervals (Chapter 6). The own acquired benchmark dataset DaLiAc was used to train and evaluate the performance of the hierarchical classification system (Chapter 8).

Second, data are preprocessed. Preprocessing includes outlier detection and removal, filtering, or labeling of the data. In chapter 6, this consisted of extracting the R-peaks from the dataset. Here, the reference annotations for the R-peaks of the databases were used. Additionally, the specific RR interval data were extracted. In chapter 8, four datasets were excluded and all remaining datasets were labeled. In total, data of 19 subjects performing 13 different activities were included in DaLiAc.

Third, features are extracted. This consists of defining appropriate windows depending on the application, the calculation of features within these windows, and if desired, the selection of features. In chapter 6, features were extracted by fitting a sine curve to segments of consecutive RR intervals. Features were grouped in three feature groups (FGs) and the performance was assessed for all FGs separately. In chapter 8, sliding windows with 50% overlap were used to extract time and frequency domain features. Different features were extracted for the REST classification system – one of the single classification systems. No feature selection was performed for this classification task. Relevant features of the ECG signal are extracted in chapter 7. Despite being no classification task, the algorithms for reference-free adjustment of RIP included a feature selection step (chapter 5). The best set of predictors were determined using a sequential forward feature selection within an inner 10-fold cross-validation. The concept of cross-validation is explained in chapter 3.3.3.

Fourth, classification is performed. The respective classifier uses the features as input and outputs the corresponding class label. Often, different classifiers are compared as their performance depends on the particular situation. This is also summarized in the No Free Lunch Theorem (Duda et al., 2000). The choice of the best features and classifier is done in a training step. The Naive Bayes classifier was used in chapter 6. The four

classifiers, Adaptive Boosting (AdaBoost), Classification and Regression Tree (CART), K-Nearest Neighbor (kNN), and SVM, were compared in chapter 8 and are explained in the following chapters.

3.3.2 Classifiers

In this thesis, five classifiers were used for the different classification tasks. All classifiers were based on supervised learning. Each classifier was first trained with a given training set of the feature vectors and their corresponding class labels. Afterwards, the trained classifiers were tested by predicting the class label of a given new feature vector. This section gives a short overview of the used classifiers in this thesis. The Naive Bayes classifier was used in chapter 6 to detect arrhythmic beats in RR interval segments. The hierarchical classification system (chapter 8) had the advantage that different classifiers could be used in each classification subsystem. For this, the performances of the four classifiers, AdaBoost, CART, kNN, and SVM, were compared.

Naive Bayes

Naive Bayes classifiers are based on the Bayes's theorem. They belong to the family of probabilistic classifiers. For each class, posterior probabilities are computed for each new feature vector. The feature vector is then assigned to the class having the highest posterior probability.

The Bayes's theorem paves the way to compute the posterior probability $P(c|x)$:

$$P(c|x) = \frac{p(x|c) \cdot P(c)}{p(x)}, \quad (3.14)$$

with $p(x|c)$ as class-conditional probability density function, $P(c)$ as prior probability of class c , and $p(x)$ as probability density function of evidence x . $p(x|c)$ is also called likelihood of evidence x with respect to class c .

The Naive Bayes classifier assumes that the features are conditionally independent. Therefore, the Naive Bayes classifier can estimate separate class-conditional probability density functions for each feature. Hence, the classifier is trained through estimating class-conditional probability density functions by means of a given training set of labeled feature vectors (Duda et al., 2000; Theodoridis & Koutroumbas, 2008).

Adaptive Boosting

AdaBoost is the most popular variation of boosting. In general, boosting is used to improve the accuracy of any given learning algorithm. First, a

weak classifier is chosen that has slightly higher accuracy on the training data than chance. Second, new classifiers are generated using the weak classifiers to form an ensemble whose joint decision rule has high accuracy on the training set. The name boosting originates from the fact that the classification performance has been “boosted”.

In AdaBoost, weak classifiers are continuously added until a termination criterion like final number of iterations reached or error below the desired training error. At each iteration, a new classifier is generated by training the weak classifier on different subsets of the entire training data of labeled feature vectors. Weights of the feature vectors determine their probability of being selected for the subset.

All feature vectors have the same weight at the first iteration. At the subsequent iterations, the weight of a feature vector is either increased or decreased. The weight is increased if it was incorrectly classified by the classifier of the previous iteration. The weight is decreased if it was correctly classified. Hence, AdaBoost pays more attention to feature vectors that are difficult to classify. The final classifier is obtained as the weighted average of all classifiers generated before.

Classification and Regression Tree

CART classifiers belong to the family of decision trees. Decision trees ask different questions at each node concerning a particular property of the feature vector. The question on the next node, if any, depends on the answer of the previous question. Successive nodes are visited until a terminal node, where the class label is given, is reached. The results of these questions form a tree-like structure. The ends of this tree are terminal or leaf nodes with no more following questions.

Using the CART approach, a general framework is provided that can be realized in different ways to produce different decision trees. The following six general questions arise:

- Number of splits at a node:
In general, the number of splits at a node is not limited. However, it is common to just have two splits at a node. This forms a binary decision tree.
- Selection of a specific element of the feature vector and the corresponding decision at each node:

Different criteria for choosing a split exists like the Gini's diversity index or also called Gini impurity *Gini*:

$$Gini = \sum_{i \neq j} P(i)P(j), \quad (3.15)$$

with $P(i)$ as the fraction of feature vectors at the relevant node that have the label i . *Gini* will be 0, when all feature vectors belong to one label.

- Declaration of terminal or leaf nodes:
Using cross-validation methods (section 3.3.3) or defining thresholds are two example approaches to declare leaf nodes.
- Pruning:
For pruning, first, the tree is grown completely. Then all branches of a node are considered for elimination. Any branches, whose elimination would bring an appropriate increase in impurity, are eliminated. The preceding node is declared as leaf node, until it might be pruned.
- Class of leaf node:
The leaf will be assigned to the label occurring most often.
- Handling of missing data:
Different options for handling missing data exist. Feature vectors having single missing values could be deleted completely or just the corresponding value could be neglected for this feature vector. Another possibility would be to assign virtual values, its most likely value based on the other feature vectors values.

k-Nearest Neighbor

The kNN classifier is a non-parametric method. The kNN classifier performs its classification decision using the k closest feature vectors. First, the k feature vectors of the training set that are nearest to a given new feature vector are determined. Then, the new feature vector is classified into the class, which occurs most often among the k nearest neighbors. Different metrics or distance measures for the calculation of proximity can be used.

The Minkowski metric is a general class of metrics for the two feature vectors $\mathbf{a}$ and $\mathbf{b}$ being each d -dimensional:

$$L^p(\mathbf{a}, \mathbf{b}) = \left(\sum_{i=1}^d |a_i - b_i|^p \right)^{1/p}. \quad (3.16)$$

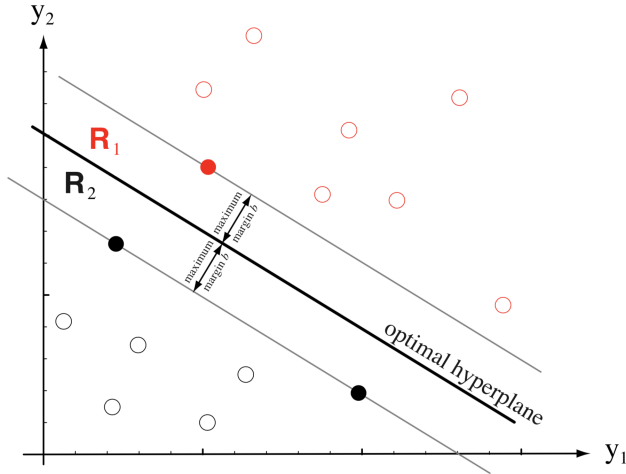


Figure 3.5: Optimal hyperplane as determined by the Support Vector Machine (SVM) classifier. The solid dots are the three support vectors belonging to the class 1 (red circles) or class 2 (black circles). The optimal hyperplane is the one that has the maximum distance from the support vectors. R_1 (R_2) denotes the region for class 1 (class 2). Reprinted from Duda, Hart, and Stork (2000).

This is also referred as the L^p norm. The L^2 norm is also called Euclidean distance. The L^1 norm is named the Manhattan or city block distance.

Support Vector Machine

The training of a SVM classifier is based on finding optimal separating hyperplanes between two classes. The optimal hyperplane is the one that has the maximum distance from the nearest feature vectors of different classes of the training dataset (Figure 3.5). Those nearest feature vectors are also called support vectors. In SVM, the original feature vectors are transformed with kernel functions into a higher dimensional space. The goal is to find a linear separating hyperplane in that space. Four kernel functions are common: linear, radial basis function, sigmoid, and polynomial (chapter 3.2.3). The performance of all four kernels was used in the SVR algorithm (chapter 4). With SVR, a regression task is performed. In regression, a functional description of data is searched to predict new data afterwards. The SVM with radial basis function kernel was used in the hierarchical classification system (chapter 8).

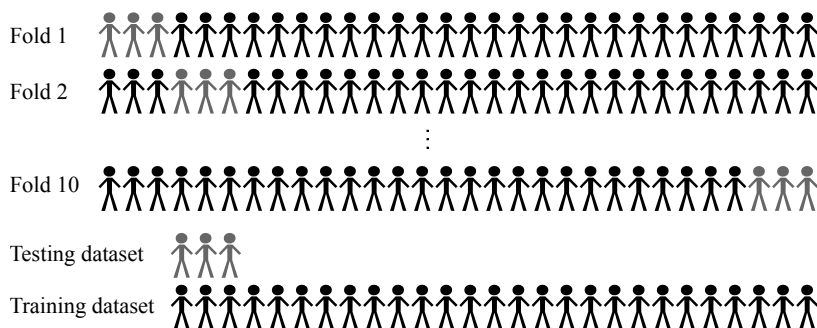


Figure 3.6: Sketch of 10-fold cross-validation. In this example, the dataset consists of data from 30 subjects. In each fold, the data of three subjects are held out as testing dataset. The data of the remaining 27 subjects are used as training dataset. The overall performance is calculated as mean over all ten individual performances obtained in each fold.

3.3.3 Cross-Validation

In all chapters except chapter 7, cross-validation was used to assess the general performance of the algorithms or classifiers. The usage of cross-validation in the respective chapters is given at the end of this chapter.

For validation, the labeled training dataset is split into two parts: the traditional training dataset for adjusting model parameters of the algorithm or classifier and the testing dataset to estimate the performance of the algorithm. In general, the performance of the algorithm should be as high as possible. Hence, this division is performed repeatedly until a maximum in performance is reached. It is fundamental that the testing dataset does not include feature vectors used for training the algorithm.

Generalizing this approach results in m -fold cross-validation. Figure 3.6 illustrates 10-fold cross-validation when the complete dataset consists of 30 subjects. Here, the training dataset is randomly divided into ten sets of equal size. Each of the ten testing datasets consists of data from three subjects. The classifier is trained ten times in total. Each time, a different dataset is held out as testing dataset. The dataset of the remaining data, 27 subjects in this case, is used as training dataset. The overall performance is determined by calculating the mean over all individual performances. If, in each fold, the dataset of exactly one subject is left out, this is called leave-one-subject-out cross-validation.

If the algorithm or classifier to be evaluated requires specific features or parameters to be identified, an approach is to perform two m -fold cross-validations that are nested into one another. In the outer m -fold cross-validation, the performance of the algorithm or classifier is determined. In the inner m -fold cross-validation, the features and parameters are optimized. This is illustrated in Figure 3.7. Sometimes it is required to nest even more m -fold cross-validations.

For the case that 30 subjects are in the complete dataset, first, the outer testing dataset, consisting of the data from three subjects, is left out in the outer fold 1. Second, in the inner fold 1, the inner testing dataset consisting of data from three subjects is left out. The algorithm is trained using the remaining dataset. Third, in the next nine inner folds, inner fold 2 to inner fold 9, the algorithm is trained repeatedly using the remaining training datasets. In the inner folds, all possible parameters are assessed. The best combination of all parameters is then trained using the complete training dataset. Afterwards, the combination is tested on the outer dataset. The same procedure occurs for the remaining outer folds 2 to 9.

The number of subjects in the inner testing datasets varies in this example. For robust results, no data from single subjects should be used for testing when used for training. After comprising three subjects in the outer testing datasets, 27 subjects remain for the inner 10-fold cross-validation. This results in seven inner testing datasets comprising three subjects each and three inner testing datasets having two subjects each.

Single 10-fold cross-validation was used in chapter 4. Two SVR parameters were determined in a grid search within a 10-fold cross-validation using the standing still phase of the training dataset. The treadmill running phase and the recovery phase of the training dataset served as testing dataset. Two nested 10-fold cross-validations using the training dataset were used in chapter 5. The best parameter for applying LSQ was determined in the outer 10-fold cross-validation. The best set of predictors was calculated in the inner 10-fold cross-validation. In both chapters, the reliability and the validation dataset being independent from the training dataset were used for testing the algorithms. The reliability dataset consisted of the same subjects, whereas coming from a different run of the experimental protocol. The validation dataset consisted of completely new, unseen subjects. The performance of the Naive Bayes classifier (chapter 6) was assessed using 10-fold cross-validation. Leave-one-subject-out cross-validation was used for the single classification systems of the hierarchical system to assess the performances (chapter 8).

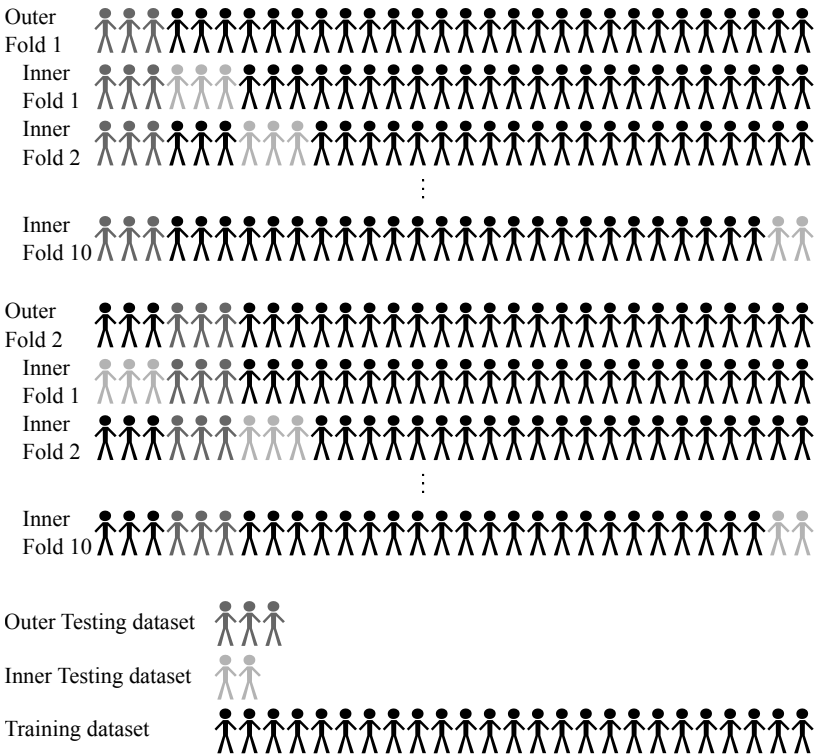


Figure 3.7: Sketch of two nested 10-fold cross-validations. First, in the outer fold 1, the outer testing dataset consisting of data from three subjects is left out. Second, in the inner fold 1, the inner testing dataset consisting of data from three subjects is left out. The algorithm is trained using the remaining training dataset. Third, in the inner fold 2 to 9, the algorithm is trained repeatedly using the remaining datasets. Depending on the number of total subjects in the dataset and the fact that data from single subjects should not be used for testing when used for training, the last three inner testing datasets consists of only two subjects. In the inner folds, all possible parameters are assessed. The best combination of parameters is then used for testing on the outer dataset. The same procedure occurs for the remaining outer folds 2 to 10.

Part II

Mobile Breathing Analysis

Chapter 4

Individual-Adapted Adjustment of Respiratory Inductance Plethysmography

In this chapter, the performance of different individual adjustment algorithms during the treadmill running phase and the recovery phase based on adjustment factors calculated in the corresponding standing still phase were investigated. By this, just an individual standing still phase with the simultaneous measurement of RIP and FM is required to acquire afterwards solely RIP data for an actual application in the field. This addresses contribution (I). Parts of this chapter have been published previously (Leutheuser, Heyde, Gollhofer, & Eskofier, 2014; Leutheuser, Heyde, Roecker, et al., 2017).

4.1 Methods

This part describes the used data and presents the adjustment algorithms. Results are compared to two algorithms from literature. Figure 4.1 shows the analysis steps for individual-adapted adjustment of RIP.

4.1.1 Data

Hardware Equipment

RIP data were acquired using the LifeShirt garment (Vivonoetics, San Diego, CA, USA) (Figure 4.2). In the LifeShirt garment, two parallel elastic bands with insulated sinusoidal wires are embedded. One of these bands surrounds the RC and the other one the AB. The VivoSense software

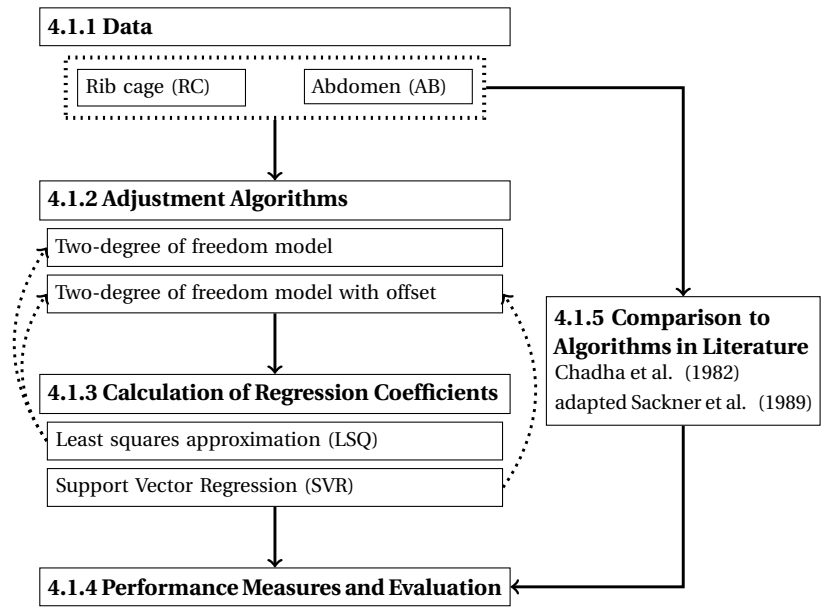


Figure 4.1: Analysis steps for individual-adapted adjustment of respiratory inductance plethysmography (RIP).

(Vivonoetics, San Diego, CA, USA) was used to process the recorded data. The outputted unadjusted values of the partial tidal volumes breath-per-breath data for RC $u V_{T_{RC}}$ and AB $u V_{T_{AB}}$ were used. A FM (Oxycon Pro Care Fusion, San Diego, CA, USA) was utilized for ground truth measurement regarding ventilatory timing and volume. Before each measurement, the FM was calibrated using a 3 l syringe. The FM required the subject to wear a face mask that was connected to a stationary system.

Experimental Protocol

The experiments were performed on a treadmill (Quasar, H/P/Cosmos Sports and Medical GmbH, Nussdorf-Traunstein, Germany). Each experiment consisted of three phases: standing still, treadmill running, and recovery. In the first phase, the standing still phase, the subjects stood in an upright position for about 5 min on the treadmill. In the second phase, the treadmill running phase, the subjects performed an incremental running test on the treadmill. The starting speed was set to 6 km/h. The speed was increased in steps of 2 km/h every 3 min until exhaustion.

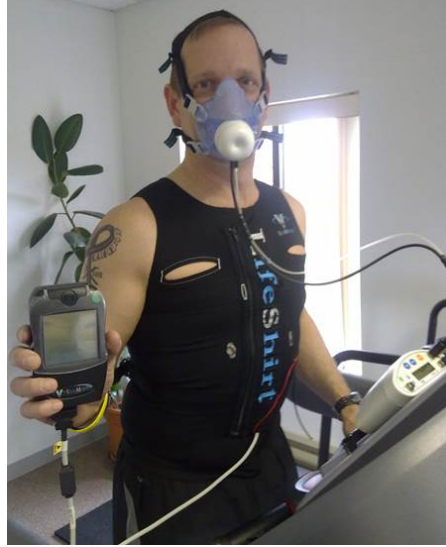


Figure 4.2: LifeShirt garment. The subject is wearing the LifeShirt garment (Vivo-Metrics, Ventura, CA, USA) and the face mask of a flowmeter (FM) (Oxycon Pro Care Fusion, San Diego, CA, USA). Picture adapted from Leutheuser et al. (2014) © 2014 IEEE.

In the third phase, the recovery phase, the subjects stood in an upright position for 10 min. In between two different speeds in the running phase, a resting period of 30 s took place to collect blood lactate samples. These pauses were excluded from the analysis.

140 of 193 subjects performed the experimental protocol twice with a pause of seven days. The available data were divided into three different datasets: training dataset, reliability dataset, and validation dataset. The training dataset contained the first run of the experimental protocol of the 140 subjects who performed the experimental protocol twice. The reliability dataset was composed of the second run of the experimental protocol for the same 140 subjects. The validation dataset was composed of the remaining 53 subjects who performed the experimental protocol once. Table 4.1, Table 4.2, and Table 4.3 give the number of breaths and the breathing frequencies of each individual phase: standing still, treadmill running, and recovery, respectively. These are shown for the three different datasets. For analysis, the unadjusted values of the partial tidal volumes of each band $uV_{T_{RC,AB}}$ for each phase after preprocessing were used.

Table 4.1: Characteristics of the standing still phase: Number of breaths and breathing frequency for the three different datasets.

Dataset	Number of breaths			Breathing frequency [Hz]		
	Mean $\pm$ SD	Min	Max	Mean $\pm$ SD	Min	Max
Training	94 $\pm$ 22	34	149	0.26 $\pm$ 0.06	0.11	0.41
Reliability	98 $\pm$ 22	39	146	0.25 $\pm$ 0.06	0.13	0.39
Validation	89 $\pm$ 22	45	138	0.27 $\pm$ 0.06	0.11	0.40

SD: standard deviation; Min: minimum value; Max: maximum value

Table 4.2: Characteristics of the treadmill running phase: Number of breaths and breathing frequency for the three different datasets.

Dataset	Number of breaths			Breathing frequency [Hz]		
	Mean $\pm$ SD	Min	Max	Mean $\pm$ SD	Min	Max
Training	542 $\pm$ 144	254	931	0.55 $\pm$ 0.10	0.32	0.79
Reliability	559 $\pm$ 155	195	963	0.56 $\pm$ 0.11	0.34	0.85
Validation	557 $\pm$ 146	255	854	0.56 $\pm$ 0.11	0.34	0.81

SD: standard deviation; Min: minimum value; Max: maximum value

Table 4.3: Characteristics of the recovery phase: Number of breaths and breathing frequency for the three different datasets.

Dataset	Number of breaths			Breathing frequency [Hz]		
	Mean $\pm$ SD	Min	Max	Mean $\pm$ SD	Min	Max
Training	246 $\pm$ 54	61	386	0.42 $\pm$ 0.09	0.25	0.73
Reliability	253 $\pm$ 69	19	617	0.42 $\pm$ 0.09	0.27	0.66
Validation	228 $\pm$ 78	14	380	0.43 $\pm$ 0.09	0.23	0.70

SD: standard deviation; Min: minimum value; Max: maximum value

Subjects

193 healthy subjects participated in the study. Table 4.4 shows detailed information about the subjects. Of them, 140 subjects performed the experimental protocol twice. All subjects had experience in treadmill running.

The RC circumference and the AB circumference of each subject was measured to guarantee an appropriate fit of the LifeShirt garment. These circumferences were measured while each subject stood in an upright

Table 4.4: Subject characteristics for the three available datasets. The data of subjects performing the experimental protocol twice on different days were divided into training and reliability dataset. The training dataset consisted of the first run of the experimental protocol.

	Training & reliability dataset			Validation dataset		
	140 subjects (64 female)			53 subjects (28 female)		
	Mean $\pm$ SD	Min	Max	Mean $\pm$ SD	Min	Max
Age (years)	28 $\pm$ 9	16	60	24 $\pm$ 7	16	52
Height (cm)	176 $\pm$ 9	154	196	175 $\pm$ 8	161	192
Weight (kg)	69 $\pm$ 12	45	100	67 $\pm$ 9	50	93
RC circ. (cm)	95 $\pm$ 7	80	114	95 $\pm$ 6	84	116
AB circ. (cm)	81 $\pm$ 7	65	98	80 $\pm$ 7	66	104

SD: standard deviation; Min: minimum value; Max: maximum value; RC: rib cage; AB: abdomen; circ.: circumference

position and breathed normally. The Ethics Committee of the University of Freiburg, Re.-No. 208/11, approved the study. All subjects gave written informed consent to participate in the study.

Preprocessing

The synchronization of FM and RIP data was done using a Visual Basic for Applications (VBA) routine in Microsoft Excel (Microsoft Office 2010, Microsoft Corporation, Redmond, WA, USA). The synchronization was based on the total breathing time t_{TOT} (Heyde, Leutheuser, Eskofier, Roecker, & Gollhofer, 2014). On a breath-by-breath basis, a segment of 15 breaths was shifted through the complete dataset while calculating the cross-correlation. The segment stopped automatically when differences between data streams exceeded the limits of $\pm 10\%$ (in relation to the corresponding face mask t_{TOT}). Incomparable values (i.e. when one valid breath was divided in two or two valid breaths were combined into one) were excluded from further analysis. In average, the relative portion of incomparable breaths was below 5% in each individual dataset. Further analysis was performed using the MATLAB package (Mathworks Inc., Natick, MA, USA).

4.1.2 Adjustment Algorithms

The basic idea was to use data from the standing still phase ($u V_{T_{RC}}$ and $u V_{T_{AB}}$ of RIP and $V_{T_{FM}}$ of FM data) to determine individual VMCs $gain_{RC,AB}$ for each subject. These individual VMCs were further utilized for a priori adjustment of the tidal volume data during the treadmill running and the recovery phase.

Two models were chosen to calculate accurate VMCs from the resting phase:

1. The two-degree of freedom model by Chadha et al. (1982) based on the two-degree of freedom model of Konno and Mead (1967) (equation (2.2)):

$$V_T = gain_{RC} \cdot u V_{T_{RC}} + gain_{AB} \cdot u V_{T_{AB}} \quad (4.1)$$

2. The above model extended with an offset c_{offset} as constant:

$$V_T = gain_{RC} \cdot u V_{T_{RC}} + gain_{AB} \cdot u V_{T_{AB}} + c_{offset}. \quad (4.2)$$

4.1.3 Calculation of Regression Coefficients

Each of the two models (equations (4.1) and (4.2)) were solved using two different algorithms. The first algorithm consisted of LSQ. The second algorithm was SVR (Schölkopf & Smola, 2002; Smola & Schölkopf, 2004) with four different kernels (linear, radial basis function, sigmoid, and polynomial) (chapter 3.2.3). The LIBSVM implementation using version 3.22 (Chang & Lin, 2011) was integrated in the MATLAB package (Mathworks Inc., Natick, MA, US) to conduct the SVR computations. The radial basis function kernel consisted of $\exp(\frac{1}{2} \|u V_{T_{RC}} - u V_{T_{AB}}\|^2)$. The sigmoid kernel consisted of $\tanh(\frac{1}{2} \langle u V_{T_{RC}}, u V_{T_{AB}} \rangle)$. The polynomial kernel consisted of $(\frac{1}{2} \langle u V_{T_{RC}}, u V_{T_{AB}} \rangle)^3$.

The cost parameter c and ϵ in the loss function of the SVR were determined using only the training data. Possible parameters for c were $\{0.001; 0.01; 0.1; 1; 10; 100; 1000; 10,000; 100,000; 1,000,000\}$ and for ϵ were $\{10^{-7}; 10^{-6}; 10^{-5}; 10^{-4}; 0.001; 0.01; 0.1\}$.

4.1.4 Performance Measure and Evaluation

Performance Measure

The RMSE for multiple regression (Snedecor & Cochran, 1989) was calculated:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (V_{T,i} - V_{T_{FM},i})^2}{N - k}}, \quad (4.3)$$

with $V_{T_{FM}}$ as the reference volume of the FM, N as the total number of breaths, and k as the number of regression coefficients in the model. In regression, a perfect fit is represented by small residuals $V_T - V_{T_{FM}}$ and thus a small *RMSE* value.

Evaluation

The training dataset was used to determine the best SVR parameters c and ϵ . Afterwards, the performance was determined using the validation and the reliability dataset. Using the reliability dataset, RMSE values were calculated with either VMCs from the standing still phase of the reliability dataset or the training dataset.

The SVR parameters were determined in a grid search within a 10-fold cross-validation using the training dataset. For each subject individually, the standing still phase of the training dataset was used to train the current SVR-model with all pairs of c and ϵ . This model was then evaluated for each subject individually on the treadmill running phase and the recovery phase using the training dataset. For each pair of c and ϵ , an averaged RMSE value was determined. The best pair of c and ϵ were defined belonging to the minimum RMSE value. If improvements in RMSE were only marginal (smaller than 10^{-6}), the smaller c and the larger ϵ were used.

The performance of the adjustment algorithms was assessed using two phases:

1. Running with normal and enhanced breathing depth (treadmill running phase)
2. Standing still with enhanced breathing depth (recovery phase)

4.1.5 Comparison to Algorithms in Literature

The individual adjustment algorithms were compared to two state of the art algorithms in the literature: Chadha et al. (1982) and Sackner et

Table 4.5: Support Vector Regression (SVR) parameters c and ϵ determined using the training dataset.

Kernel type	Treadmill Running		Recovery	
	c	ϵ	c	ϵ
Linear	100000	0.00001	1000	0.01
Radial basis function	0.1	0.01	0.1	0.1
Sigmoid	0.1	0.01	0.1	0.1
Polynomial	0.1	0.1	0.01	0.00001

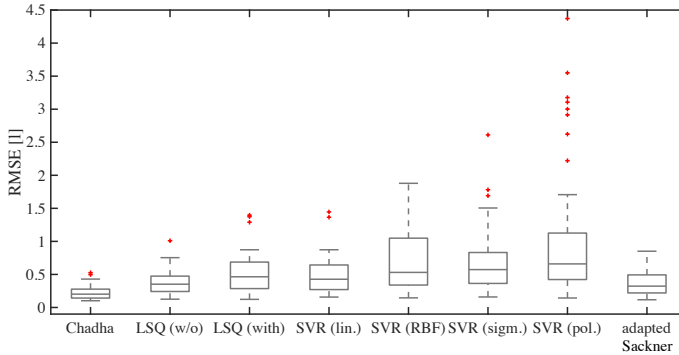
al. (1989). The LSQ by Chadha et al. (1982) served as best known gold standard model. This algorithm was designed requiring simultaneous measurement of RIP and FM data. The QDC by Sackner et al. (1989) was applied to the data in the same way as the suggested individual adjustment algorithms presented in this chapter. This means that the standing still phase was used to calculate the k-ratio K and the scaling factor M (equation (2.3)). These VMCs were then applied to the treadmill running and the recovery phase. This algorithm is called adapted Sackner in the remaining part of this chapter.

4.2 Results

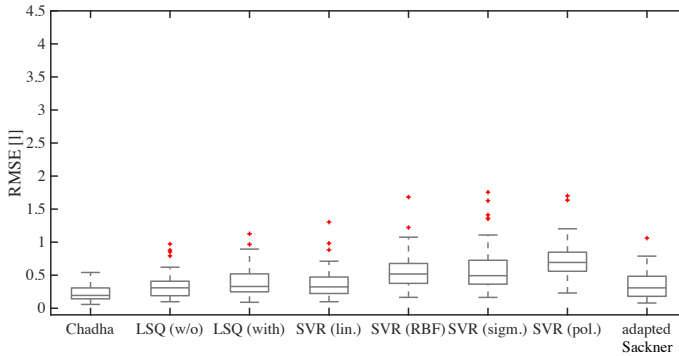
Table 4.5 shows the final SVR parameters c and ϵ for the different kernels divided for the treadmill running and the recovery phase.

The individual RMSE values are represented in a boxplot for each individual adjustment algorithm for the validation dataset (Figure 4.3). The RMSE values were calculated for each subject separately. In each box, the central mark shows the median RMSE value of the subjects. The top and bottom edges denote the 25th and 75th percentiles, respectively. The plus signs indicate outliers. Adjustment algorithms that are applicable to a broad subject population are characterized by a small median, few or ideally none outliers, and grouped closely around the median. Outliers indicate subjects for whom the adjustment algorithm was not appropriate.

Comparing the RMSE values of the training data and the reliability data adjusted with the same VMCs revealed similar results (Figure 4.4). The VMCs were calculated using the training data. Results are shown for the best four adjustment algorithms.



(a) Treadmill Running



(b) Recovery

Figure 4.3: Root mean square error (RMSE) values illustrated as boxplots for all eight different individual adjustment algorithms for the validation dataset. The results are shown for the two phases **(a)** treadmill running and **(b)** recovery independently. All algorithms except from Chadha et al. (1982) used the principle of volume motion coefficient (VMC) calculation in the standing still phase and afterwards application to the corresponding phases. The VMCs were calculated using the standing still phase of the validation dataset. In each box, the central mark shows the median RMSE value of the subjects. The top and bottom edges denote the 25th and 75th percentiles, respectively. The plus signs indicate outliers. Chadha: adjustment algorithm by Chadha et al. (1982) using simultaneous data of respiratory inductance plethysmography (RIP) and flowmeter (FM); LSQ: Least squares approximation; LSQ (w/o): LSQ without offset; LSQ (with): LSQ with offset; SVR: support vector regression; lin.: linear; RBF: radial basis function; sigm.: sigmoid; pol.: polynomial; adapted Sackner: adjustment algorithm by Sackner et al. (1989) where the VMCs were determined in the standing still phase and applied to either the treadmill running phase or recovery phase.

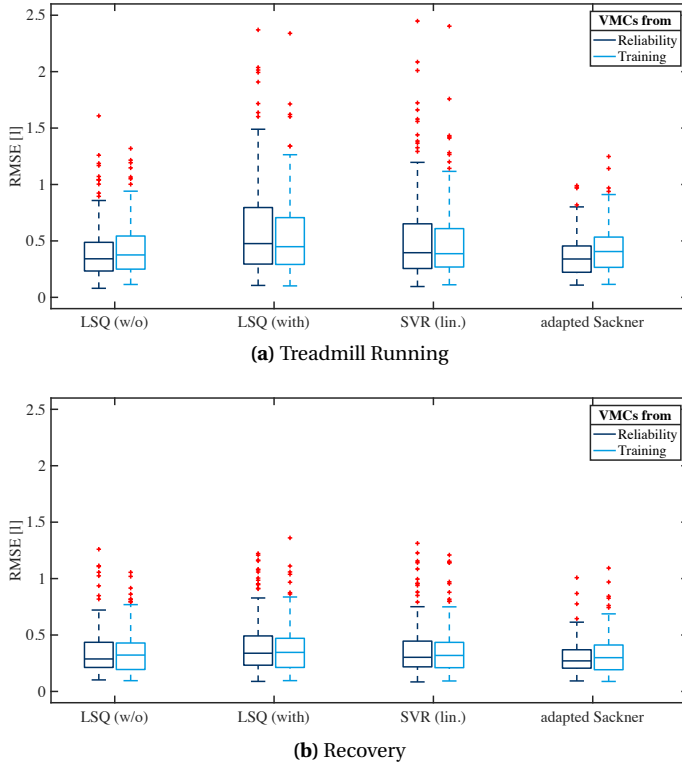


Figure 4.4: Comparison of root mean square error (RMSE) values for individual adjustment using the reliability dataset. The volume motion coefficients (VMCs) were either calculated using the standing still phase of the reliability dataset (dark blue) or the training dataset (light blue). The VMCs were applied to the reliability data, either (a) the treadmill running phase or (b) the recovery phase. The RMSE values are illustrated as boxplots for the best four different individual adjustment algorithms for the reliability dataset. In each box, the central mark shows the median RMSE value of the subjects. The top and bottom edges denote the 25th and 75th percentiles, respectively. The plus signs indicate outliers. LSQ: Least squares approximation; LSQ (w/o): LSQ without offset; LSQ (with): LSQ with offset; SVR: support vector regression; lin.: linear; adapted Sackner: adjustment algorithm by Sackner et al. (1989) where the VMCs were determined in the standing still phase and applied to either the treadmill running phase or recovery phase.

4.3 Discussion

In this chapter, algorithms were developed to enable the adjustment of RIP data without simultaneous measurement of RIP and FM data. The standing still phase of each subject was used to calculate individual VMCs. Then, these were applied to conditions of treadmill running and recovery.

For assessing the performance of this method, ventilatory data were recorded with both measurement devices, RIP and FM. For an actual application in the field, this would require to only record with both modalities during a standing still phase to determine individual VMCs. Afterwards, these would be stored for direct application of incoming ventilatory data during field tests.

General discussion

The adapted Sackner algorithm (Sackner et al., 1989) achieved the lowest RMSE values throughout all results, when considering a real-life scenario where simultaneous data acquisition is not feasible (Figure 4.3). This is therefore suggested as the best adjustment algorithm when it is possible to acquire data before the actual assessment in the field.

Overall, smaller RMSE values were computed for the recovery phase than the treadmill running phase (Figure 4.3 and Figure 4.4). For the best algorithm, the results were comparable. There are distinct differences between the phases: Breathing frequencies in the standing still phase were smallest, followed by breathing frequencies in the recovery phase (Table 4.1 and Table 4.3). The treadmill running phase had the highest breathing frequencies (Table 4.2). Additionally, there was an enhanced movement in the upper body due to running and increased breathing depth. In the other two phases, the subjects stood still. In general, the performance of the best algorithm (adapted Sackner algorithm) is crucial. Hence, as similar results were obtained, it is concluded that the suggested adjustment algorithm works as intended.

Performance in Relation to Gold Standard

The results were shown in comparison to the gold standard adjustment. For gold standard adjustment, the best known adjustment algorithm presented in literature was chosen. This was the LSQ without offset algorithm (Chadha et al., 1982; Heyde et al., 2014). The results of the gold standard adjustment demonstrated the best possible outcome, but required the simultaneous measurement of RIP and FM data. The determined RMSE

values of the gold standard adjustment indicated that even the results for the gold standard are not perfect.

The adapted Sackner algorithm (Sackner et al., 1989) that is a contribution of this thesis scored best amongst all a priori adjustment algorithms and slightly worse compared to the gold standard adjustment. However, it has the advantage that no simultaneous measurement of RIP and FM data was needed. In this work, the LSQ without offset algorithm scored second amongst all a priori adjustment algorithms, when using the VMCs from the standing still phase and applied to the treadmill running phase or the recovery phase (Figure 4.3).

Usability

The (re)usability of the determined VMCs is given when comparing the performances on the reliability dataset (Figure 4.4). Comparable results were obtained for the best four adjustment algorithms. This shows the practical ability of the suggested adjustment algorithm. In this work, seven days lay between the measurements for the training and the reliability dataset. Hence, if the time interval between two measurements is small enough (e.g. one week), no new standing still phase has to be recorded with both measurement modalities. The actual measurement in the field can start without any delay. The effect of longer durations between two measurements has to be further investigated.

Comparison to Literature

An adapted version of the QDC algorithm (Sackner et al., 1989) was implemented in the same way as the other proposed adjustment algorithms. This algorithm outperformed the performance of the LSQ algorithms (Figure 4.3 and Figure 4.4), resulting in the suggestion of using this algorithm in the future.

Adaptations Compared to Original Publication

Some things were changed compared to the work by Leutheuser, Heyde, Gollhofer, and Eskofier (2014). First, the QDC algorithm was implemented as suggested by de Groote et al. (2001). This resulted in improved RMSE values. Second, additional data were included. This resulted in additional results and findings. The performances of the adjustment algorithms were assessed using the treadmill running and the recovery phase. The parameter optimization for the SVR algorithm was revised. The parameter were optimized regarding the performance on the treadmill running phase

or the recovery phase. Previously, the parameters were optimized within the standing still phase as only one dataset was available. The results for VMCs calculated using different datasets, the training and the reliability dataset, applied to the reliability dataset were shown. The expiratory tidal volume was used instead of the inspiratory tidal volume (Leutheuser, Heyde, Roecker, et al., 2017).

Limitations

The adjustment algorithms were calculated using only data from the standing still phase. Additionally, tidal volume data from the standing still phase were lower than data from the treadmill running phase or the recovery phase. The standing still phase had a minimum duration of 4.8 min with a minimum number of breaths of 39 (Table 4.1). This means that data of subjects with low breathing frequencies were used to calculate VMCs. VMCs of such few values might be inaccurate. In the future, it might be beneficial in using a predefined number of breaths rather a predefined time.

The LSQ models might possibly be improved by solving equation (4.1) using least absolute deviations or L^p -norms with $p > 2$ instead of LSQ (equation (3.16)). The algorithmic possibilities of SVR like the usage of different kernels could not reach the performance of the LSQ algorithms.

Chapter highlights

- The adjustment of respiratory inductance plethysmography (RIP) data without simultaneous measurement of RIP and flowmeter (FM) data is possible. For this the prerecorded standing still phase of each subject was used. This resulted in individual adjustment parameters for each subject. The adjustment parameters were applied to conditions of treadmill running and recovery.
- The best adjustment algorithm was the adapted Sackner algorithm outperforming Support Vector Machine (SVM) with four different kernels and least squares approximation (LSQ) with and without offset. In the Sackner algorithm or Qualitative Diagnostic Calibration (QDC) method, first, the relative contribution of both bands is calculated to get the k-ratio K . Second, this is set in relation to the true tidal volume to obtain the scaling factor M . In the adapted Sackner algorithm, only the standing still phase was used to calculate the k-ratio K and the scaling factor M .
- Comparable results were obtained when applying the determined adjustment parameters to the data of the same subjects acquired with a difference of one week. This shows the repeatability and usability of the suggested adjustment parameters from the standing still phase in different situations.

Chapter 5

Reference-Free Adjustment of Respiratory Inductance Plethysmography

In this chapter, the performance of generalizable and reference-free adjustment algorithms during the treadmill running and the recovery phase was considered. The proposed adjustment algorithms were designed during measurement with RIP. Generalizable means that no prerecording of any data is needed. Reference-free stands for the fact that the adjustment works having solely RIP data. This means that the simultaneous data acquisition of RIP and FM data is not needed. This addresses contribution (II). Parts of this chapter have been published previously (Leutheuser, Heyde, Gollhofer, & Eskofier, 2014; Leutheuser, Heyde, Roecker, et al., 2017).

5.1 Methods

This part of the chapter introduces the data and presents the adjustment algorithms together with the calculation of regression coefficients. Results are compared to algorithms in literature with distinct performance measures. Figure 5.1 shows the different analysis steps.

5.1.1 Data

The same data as in section 4.1.1 were used (Table 4.1, Table 4.2, and Table 4.3). Table 4.4 gives detailed information about the subjects. The performance of different adjustment algorithms during the treadmill running and the recovery phase was investigated.

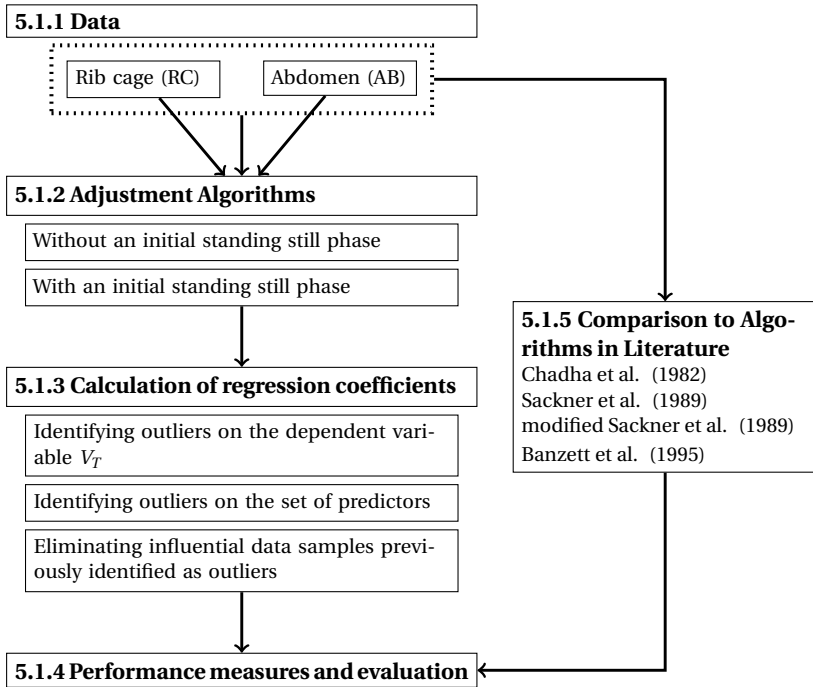


Figure 5.1: Analysis steps for reference-free adjustment of respiratory inductance plethysmography (RIP).

5.1.2 Adjustment Algorithms

The adjustment algorithms were based on the two-degree of freedom model (Konno & Mead, 1967). Additionally, it was investigated how the performance changed when only one band instead of two was used. The dependent variable (expiratory volume of the FM device) was predicted from the unadjusted partial expiratory tidal volumes of the RIP device based on anthropometric data and other derived variables from the AB device.

Adjustment model – without an initial standing still phase

The tidal volume V_T was calculated as the linear combination of the unadjusted values of the partial tidal volumes of each band $uV_{T_{RC,AB}}$:

$$V_T = gain_{RC} \cdot uV_{T_{RC}} + gain_{AB} \cdot uV_{T_{AB}} + c_{offset}, \quad (5.1)$$

where $gain_{RC,AB}$ comprises the VMC for the RC or AB band, respectively, and c_{offset} is the constant.

Anthropometric data were included to improve the accuracy of the adjustment. Each VMC was set as a linear combination of individual anthropometric data values:

$$gain = \sum_{i=1}^8 b_i \cdot p_i + b_{offset} \quad (5.2)$$

where b_i are the regression coefficients and p_i are the individual anthropometric data values. The possible p_i were from the following set of anthropometric data {sex, height, weight, age, RC circumference, AB circumference, AB/RC ratio, RC/AB ratio}.

Four additional parameters were included: two time characteristics (inspiratory time t_{in} and expiratory time t_{exp}), and the current ratio of the unadjusted partial tidal volumes $\frac{u V_{TAB}}{u V_{TRC}}$ and its inverse $\frac{u V_{TRC}}{u V_{TAB}}$.

Combined, the final regression equation using the RC and the AB band can be summarized as

$$\begin{aligned} V_T = & \left(\sum_{i=1}^8 b_i p_i + b_9 + b_{10} t_{in} + b_{11} t_{exp} \right) \cdot u V_{TRC} + \\ & + \left(\sum_{i=1}^8 b_{i+11} p_i + b_{20} + b_{21} t_{in} + b_{22} t_{exp} \right) \cdot u V_{TAB} + \\ & + b_{23} \frac{u V_{TAB}}{u V_{TRC}} + b_{24} \frac{u V_{TRC}}{u V_{TAB}} + b_{25}, \end{aligned} \quad (5.3)$$

where b_i are the regression coefficients.

In the case of using only one band, equation (5.3) was modified. When only the RC band was considered, the regression coefficients b_{12} to b_{24} were set to zero before calculating the remaining regression coefficients. When only the AB band was considered, the parameters b_1 to b_{11} , b_{23} , and b_{24} were set to zero before calculating the remaining regression coefficients.

Adjustment model – with an initial standing still phase

The adjustment model with an initial standing still phase is based on the adjustment model without an initial standing still phase with four additional parameters. These parameters were calculated from the standing still phase using only the RIP device. The four parameters were

1. The average value of the ratio of the unadjusted partial tidal volumes $S_{avg,1}$:

$$S_{avg,1} = \frac{1}{N_r} \sum_{i_r=1}^{N_r} \frac{u V_{T_{AB},i_r}}{u V_{T_{RC},i_r}}, \quad (5.4)$$

where N_r are the total number of breaths in the standing still phase.

2. The average value of the inverse ratio of the unadjusted partial tidal volumes $S_{avg,2}$:

$$S_{avg,2} = \frac{1}{N_r} \sum_{i_r=1}^{N_r} \frac{u V_{T_{RC},i_r}}{u V_{T_{AB},i_r}} \quad (5.5)$$

3. The mean of the unadjusted tidal volume of the RC band:

$$M_{RC} = \frac{1}{N_r} \sum_{i_r=1}^{N_r} u V_{T_{RC},i_r} \quad (5.6)$$

4. The mean of the unadjusted tidal volume of the AB band

$$M_{AB} = \frac{1}{N_r} \sum_{i_r=1}^{N_r} u V_{T_{AB},i_r} \quad (5.7)$$

Combined, the final regression equation using the RC and the AB band can be summarized as

$$\begin{aligned} V_T = & \left(\sum_{i=1}^8 b_i p_i + b_9 + b_{10} t_{in} + b_{11} t_{exp} \right) \cdot u V_{T_{RC}} + \\ & + \left(\sum_{i=1}^8 b_{i+11} p_i + b_{20} + b_{21} t_{in} + b_{22} t_{exp} \right) \cdot u V_{T_{AB}} + \\ & + b_{23} \frac{u V_{T_{AB}}}{u V_{T_{RC}}} + b_{24} \frac{u V_{T_{RC}}}{u V_{T_{AB}}} + b_{25} + \\ & + (b_{26} S_{avg,1} + b_{27} S_{avg,2} + b_{28} M_{RC} + b_{29} M_{AB}) \cdot u V_{T_{RC}} + \\ & + (b_{30} S_{avg,1} + b_{31} S_{avg,2} + b_{32} M_{RC} + b_{33} M_{AB}) \cdot u V_{T_{AB}}, \end{aligned} \quad (5.8)$$

where b_i are the regression coefficients.

In the case of using only one band, equation (5.8) was modified. When only the RC band was considered, the regression coefficients b_{12} to b_{24} , b_{26} , b_{27} , and b_{29} to b_{33} were set to zero before calculating the remaining regression coefficients. When only the AB band was considered, the parameters b_1 to b_{11} , b_{23} , b_{24} , and b_{26} to b_{32} were set to zero before calculating the remaining regression coefficients.

5.1.3 Calculation of Regression Coefficients

The multiple linear regression equations were solved using LSQ including an outlier detection. To calculate the regression coefficients using LSQ, the pseudoinverse was determined as described in the following and in more detail in chapter 3.2.2. A SVD of the matrix encompassing all possible predictor values was performed. It was ensured that the condition number of the matrix whose diagonal elements were the singular values was below κ . As the condition number is the fraction of the biggest to the smallest singular value, the smallest singular value was set one after another to zero unless the new fraction was below κ . The pseudoinverse was then calculated using the truncated singular values matrix. For each phase, κ was chosen from one of the following set {10; 100; 1000; 10,000; 100,000} determined using 10-fold cross-validation (see section 5.1.4).

The outlier detection consisted of three steps (Stevens, 2009):

- **Identifying outliers on the dependent variable V_T**
- **Identifying outliers on the set of predictors**
- **Eliminating influential data samples previously identified as outliers**

The individual steps are described in the following.

Identifying outliers on the dependent variable V_T

Data samples were identified as outliers on the dependent variable V_T if the calculated standardized residuals r_i of these data samples exceeded the value 3 (Hoaglin & Welsch, 1978). The value 3 originated from the fact that, assuming a normal distribution, about 99% of all data samples should lie within the range of 3 standard deviations of the mean:

$$r_i = \frac{V_{T,i} - V_{T_{FM},i}}{\hat{\sigma} \sqrt{1 - h_{ii}}} > 3, \quad (5.9)$$

with $V_{T,i} - V_{T_{FM},i}$ as raw residuals that are standardized with the $\hat{\sigma}^2$ being the variance not predictable from the regression and h_{ii} being the i th diagonal element of the hat matrix (Hoaglin & Welsch, 1978).

Identifying outliers on the set of predictors

Data samples were identified as outliers on the set of predictors if the Mahalanobis distance D_i^2 (Mahalanobis, 1936) exceeded the critical value

M_{crit} calculated using an α -value of 0.05. The Mahalanobis distance D_i^2 was calculated using following equation:

$$D_i^2 = \mathbf{x}_i^t S^{-1} \mathbf{x}_i, \quad (5.10)$$

with S as the covariance matrix and $\mathbf{x}_i$ as the vector difference between the vector of the data for case i and the vector of means for the predictors.

Eliminating influential data samples previously identified as outliers

In the last step, influential data samples were eliminated that were identified as outliers both on the dependent variable and the set of predictors. To do so, the Cook's Distance CD_i was calculated and all data samples with a CD_i larger than 1 were eliminated:

$$CD_i = \frac{1}{k} r_i^2 \frac{h_{ii}}{1 - h_{ii}}, \quad (5.11)$$

with k as the number of regression coefficients in the model, r_i as the standardized residual, and h_{ii} as the i th diagonal element of the hat matrix.

5.1.4 Performance Measures and Evaluation

Performance Measures

Apart from calculating the RMSE for multiple regression (Snedecor & Cochran, 1989) (equation (4.3)), four additional performance measures were calculated. To assess nonsystematic differences, two additional performance measures were calculated for each subject individually and after the adjustment procedure. Firstly, the percentage of breaths within the limits of Equivalence (LoE) $\pm 20\%$ was determined. Percent differences were calculated using $V_T - V_{T_{FM}}$ with regard to $V_{T_{FM}}$. Secondly, the slope between the reference volume of the FM as the dependent variable $V_{T_{FM}}$ was determined and the calculated tidal volume V_T as the independent variable using the following linear regression equation:

$$V_{T_{FM}} = m \cdot V_T + d_{offset}, \quad (5.12)$$

with m as slope and d_{offset} as the constant.

Additionally, for the adjustment algorithms incorporating both bands, two accuracy measures were calculated that compare each adjusted breath with the respective adjusted breath using the algorithm described

by Chadha et al. (1982). The first accuracy measure $N_{outliers}$ gives the number of adjusted breaths declared as outliers:

$$N_{outliers} = N_{upper} + N_{lower}, \quad (5.13)$$

with $N_{upper} > q_{75} + 1.5(q_{75} - q_{25})$ and $N_{lower} > q_{25} - 1.5(q_{75} - q_{25})$, where q_{25} and q_{75} are the 25th and 75th percentiles, respectively. The results are presented using the percentage of breaths considered as outliers. The second accuracy measure ΔV_i indicates the tidal volume difference of the related adjusted breaths:

$$\Delta V_i = V_{T_{Chadha,i}} - V_{T_{both\ bands,i}}, \quad (5.14)$$

with $V_{T_{Chadha,i}}$ as adjusted tidal volume of breath i using the adjustment algorithm by Chadha et al. (1982) and $V_{T_{both\ bands,i}}$ as adjusted tidal volume of breath i using the adjustment algorithms incorporating both bands, either without or with the standing still phase. Both accuracy measures were averaged over all subjects after determining the mean value for each subject.

The Bland and Altman technique (Bland & Altman, 1986) to investigate the agreement between different methods was not applied as $V_{T_{FM}}$ has been used as dependent criterion variable in all adjustment algorithms (O'Connor, Mahar, Lauglihn, & Jackson, 2011).

Evaluation

The training dataset was used to determine the best κ value and the best set of predictors. Afterwards, the performance was assessed using the reliability dataset and the validation dataset. The reliability dataset consisted of the same subjects as in the training dataset measured while performing the experimental protocol for a second time.

First, for each phase individually and using only the training data, the best κ value was determined from the following set $\{10; 100; 1000; 10,000; 100,000\}$. For each κ , 10-fold cross-validation was used to estimate the performance for new data. A sequential forward feature selection was performed to select the best set of predictors. The feature selection was done within the κ 10-fold cross-validation using an inner 10-fold cross-validation. After κ determination and for each phase individually, the best set of features was calculated in a 10-fold cross-validation using only the training data. The offset coefficient b_{24} was excluded from the selection process and was always included in the regression equations (5.3) and (5.8). Using the RMSE value (equation (4.3)), κ and the best set

of predictors were optimized. A termination tolerance of 10^{-6} was used. In all 10-fold cross-validation runs, the same folds for each run were used. Additionally, it was ensured that breath data for individual subjects were not in the training and testing dataset simultaneously.

After calculating the best set of predictors individually for each of the six adjustment algorithms for the two phases (see below), the results were obtained using the reliability dataset and the validation dataset with best and fixed κ . The results were calculated for each subject individually. Here the results are presented as mean values with SDs over all subjects for each phase.

The performance of the two adjustment algorithms was evaluated when using breath data from the combination of the RC and AB band, when using only the RC band, and when using only the AB band. The performance was compared using two phases:

1. Running with normal and enhanced breathing depth (treadmill running phase)
2. Standing still with enhanced breathing depth (recovery phase)

5.1.5 Comparison to Algorithms in Literature

The two reference-free adjustment algorithms presented were compared to four state of the art algorithms in the literature: Chadha et al. (1982), Sackner et al. (1989), modified Sackner et al. (1989) incorporating the standing still phase, and Banzett et al. (1995). Three algorithms from the literature were designed for the application using the combination of RC and AB bands requiring simultaneous measurement of RIP and FM data. Furthermore, in three algorithms presented from literature the RIP data were adjusted against the FM data for each person individually, and every breath from both sets of measurement equipment was needed.

In the fourth algorithm, a modified version of the Sackner et al. (1989) algorithm, the standing still phase was used to calculate the scaling factor M (Leutheuser, Heyde, Gollhofer, & Eskofier, 2014), similar as in chapter 4. The k-ratio K was determined for the incoming breaths of the corresponding phase. This version of the Sackner et al. (1989) algorithm only required the simultaneous measurement of RIP and FM data during the standing still phase. In this modified version, all breaths of the corresponding phase were used and no preprocessing was applied to determine the breaths with quasi-constant tidal volume. In the two new reference-free adjustment algorithms and the modified Sackner et al. (1989) algorithm, the adjustment can be done without the simultaneous measurement of FM

Table 5.1: Determined condition number κ for the different phases and adjustment algorithms.

		Treadmill Running	Recovery
Adjustment without standing still phase	only RC	1000	10
	only AB	10	1000
	both bands	10,000	10,000
Adjustment with standing still phase	only RC	1000	1000
	only AB	10,000	100,000
	both bands	100,000	100
Comparison to literature	Chadha et al. (1982)	100	100
	Sackner et al. (1989)	na	na
	mod. Sackner et al. (1989)	na	na
	Banzett et al. (1995)	na	na

na: not applicable

and RIP data in the final application. The modified Sackner et al. (1989) algorithm required the simultaneous measurement of FM and RIP during a standing still phase. The suggested adjustment algorithm with standing still phase required the prerecording of solely RIP data in a standing still phase.

5.2 Results

Different κ values were determined for the different phases and algorithms (Table 5.1). The κ were determined in a 10-fold cross-validation using the training dataset for the different phases. In three algorithms (both versions of the algorithm by Sackner et al. (1989) and Banzett et al. (1995)), no κ had to be determined as only one dependent variable was available. Therefore, κ was always 1.

The best anthropometric predictors were determined for each condition (Table 5.2). For each condition, the best predictors were determined using the training dataset in a 10-fold cross-validation using the previously determined κ value. In the case of only one band (RC or AB), the relevant anthropometric parameters were tied with either $uV_{T_{RC}}$ or $uV_{T_{AB}}$, respectively. In the case of two bands (RC and AB), particular relevant anthropometric parameters were tied with $uV_{T_{RC}}$ and particular ones with $uV_{T_{AB}}$. This means that individual anthropometric parameters were

multiplied with the corresponding unadjusted partial tidal volume before entering them in the regression equation. The offset b_{25} was always included.

The individual RMSE values were illustrated in a boxplot, individually for the validation and the reliability dataset, respectively (Figures 5.2 and 5.3). The results were obtained for each subject individually. In each box, the central mark shows the median RMSE value of the subjects. The top and bottom edges mark the 75th and 25th percentiles, respectively. The plus signs indicate outliers.

For the validation dataset, Table 5.3 presents the $\pm 20\%$ LoE using mean $\pm$ SD over all subjects for the two phases treadmill running and recovery. The slopes were calculated using FM as the dependent variable and the calculated tidal volume as the independent variable. The same information for the reliability dataset is given in the appendix (Table A.1).

Frequency histograms are shown for the treadmill running phase and the recovery phase using the validation dataset (Figures 5.4 and 5.5). The histograms give the percentage of breaths over the difference in tidal volume ratio (% Difference in tidal volume). % Difference in tidal volume corresponds to $\frac{V_t - V_{TFM}}{V_{TFM}}$, with V_T as the calculated tidal volume and V_{TFM} as the reference volume of the FM. In each figure, three histograms are presented. The first two ones are the frequency histograms for the adjustment without and with the initial standing still phase using both bands. The third ones are the frequency histograms for the modified adjustment algorithm by Sackner et al. (1989). The same figures for the reliability dataset are given in the appendix (Figures A.1 and A.2).

The two accuracy measures are shown for both reference-free adjustment algorithms incorporating both bands compared to the adjustment algorithm by Chadha et al. (1982) for the validation dataset (Table 5.4). The percentage of breaths considered as outliers ranged from 0.03% to 0.05%. The average difference in tidal volume leveled out around 0.02 l with SDs ranging from 0.16 l (recovery phase) to 0.22 l (treadmill running phase). The same information for the reliability dataset is given in the appendix (Table A.2).

5.3 Discussion

In this chapter, algorithms for reference-free RIP adjustment were developed and evaluated. Two adjustment algorithms were presented: one adjustment with and one without an initial standing still phase. Both were applied to treadmill running and recovery phases. The performance was

Table 5.2: Determined anthropometric parameters for each of the twelve conditions. The specific anthropometric parameters were either tied with the unadjusted values of the partial tidal volume of the rib cage (RC) band $u V_{RC}$ or the abdomen (AB) band $u V_{AB}$. This means that individual anthropometric parameters were multiplied with the corresponding unadjusted partial tidal volume before they were entered as predictor in the regression equation.

	Tied with	Treadmill Running	Recovery
Adjustment without standing still phase	only RC	$u V_{RC}$	sex, height, t_{in}
	only AB	$u V_{TAB}$	age, t_{exp}
	both bands	$u V_{RC}$	height, weight, age, t_{in}
		$u V_{TAB}$	sex, AB circ., RC/AB ratio, t_{in}
Adjustment with standing still phase	only RC	$u V_{RC}$	sex, height, age, t_{in} , t_{exp} , M_{RC}
	only AB	$u V_{TAB}$	sex, weight, age, RC circ., RC/AB ratio, AB/RC ratio, t_{in} , t_{exp} , M_{AB}
	both bands	$u V_{RC}$	sex, height, $S_{avg,1}$
		$u V_{TAB}$	height, RC circ., AB circ., $S_{avg,1}$

t_{in} : inspiratory time; t_{exp} : expiratory time; circ.: circumference; $M_{RC,AB}$: mean of the unadjusted tidal volume of the RC/AB band from the standing still phase (equations (5.6) and (5.7)); $S_{avg,1}$: average value of the ratio of the unadjusted partial tidal volumes of the standing still phase (equation (5.4)); $S_{avg,2}$: average value of the inverse ratio of the unadjusted partial tidal volumes of the standing still phase (equation (5.5))

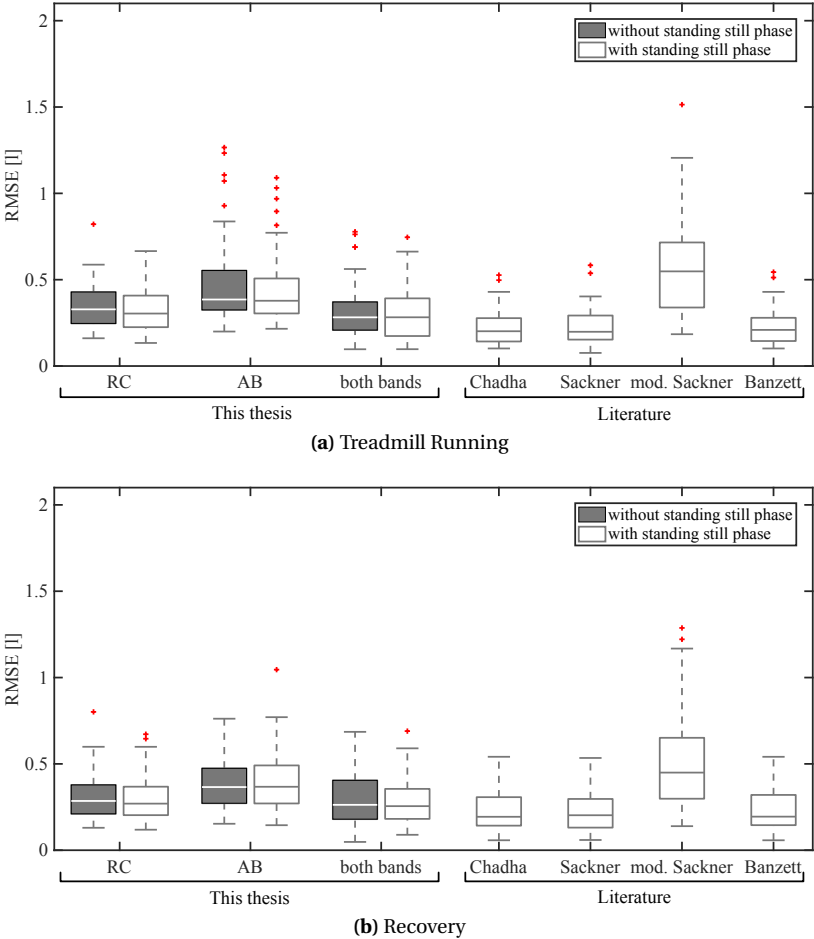


Figure 5.2: Root mean square error (RMSE) values illustrated as a boxplot for all ten different adjustment algorithms for the validation dataset. The results are shown for the two phases **(a)** treadmill running and **(b)** recovery independently. In each box, the central mark shows the median RMSE value of the subjects. The top and bottom edges denote the 25th and 75th percentiles, respectively. The plus signs indicate outliers. RC: adjustment only with the rib cage (RC) band; AB: adjustment only with the abdomen (AB) band; Chadha / Sackner / Banzett: adjustment algorithm by Chadha et al. (1982) / Sackner et al. (1989) / Banzett et al. (1995); mod. Sackner: modified (reference-free) adjustment algorithm by Sackner et al. (1989).

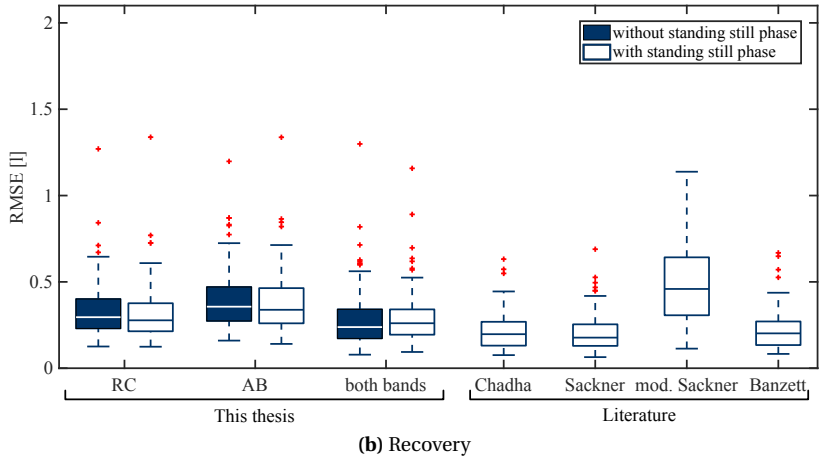
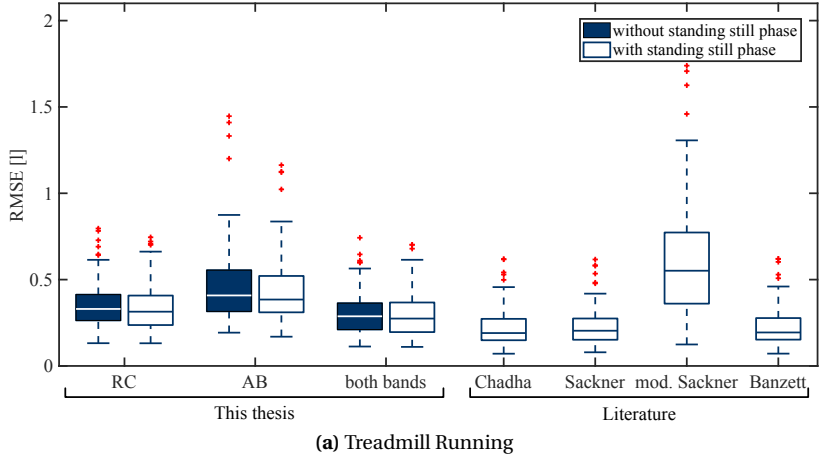


Figure 5.3: Root mean square error (RMSE) values illustrated as a boxplot for all ten different adjustment algorithms for the reliability dataset. The results are shown for the two phases (a) treadmill running and (b) recovery independently. In each box, the central mark shows the median RMSE value of the subjects. The top and bottom edges denote the 25th and 75th percentiles, respectively. The plus signs indicate outliers. RC: adjustment only with the rib cage (RC) band; AB: adjustment only with the abdomen (AB) band; Chadha / Sackner / Banzett: adjustment algorithm by Chadha et al. (1982) / Sackner et al. (1989) / Banzett et al. (1995); mod. Sackner: modified (reference-free) adjustment algorithm by Sackner et al. (1989).

Table 5.3: Limits of Equivalence (LoE) in percent (%) and slopes over all ten conditions for the validation dataset.

		Treadmill Running				Recovery			
		LoE		Slopes		LoE		Slopes	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Adjustment without standing still phase	only RC	75.23±21.76		1.42± 0.31		64.20±22.48		1.05± 0.33	
	only AB	59.78±21.98		0.83± 0.16		58.04±18.74		1.12± 0.42	
	both bands	82.85±19.21		1.17± 0.20		76.08±21.58		1.07± 0.18	
Adjustment with standing still phase	only RC	79.44±21.18		1.11± 0.19		68.92±21.05		1.10± 0.22	
	only AB	62.16±19.99		1.07± 0.37		60.11±20.54		0.98± 0.33	
	both bands	84.69±19.87		1.02± 0.17		76.20±21.15		1.00± 0.19	
Comparison to literature	Chadha et al. (1982)	94.99± 4.40		0.89± 0.09		88.12±11.37		0.90± 0.11	
	Sackner et al. (1989)	93.42± 7.10		0.79± 0.08		83.78±13.53		0.61± 0.13	
	mod. Sackner et al. (1989)	43.63±35.89		0.88± 0.18		35.01±34.69		0.69± 0.20	
	Banzett et al. (1995)	94.82± 4.57		0.80± 0.08		86.72±12.63		0.67± 0.12	

SD: standard deviation

Table 5.4: Accuracy measures for the comparison of the adjustment algorithms of both bands to the algorithm from Chadha et al. (1982) using the validation dataset.

	Treadmill Running				Recovery			
	$N_{outliers}$ [%]		ΔV [l]		$N_{outliers}$ [%]		ΔV [l]	
	Mean±	SD	Mean±	SD	Mean±	SD	Mean±	SD
Adjustment without standing still phase	0.03±	0.03	0.01±	0.25	0.04±	0.03	0.01±	0.17
Adjustment with standing still phase	0.03±	0.02	0.02±	0.24	0.05±	0.04	0.01±	0.16

SD: standard deviation

Chapter 5. Reference-Free Adjustment of Respiratory Inductance Plethysmography

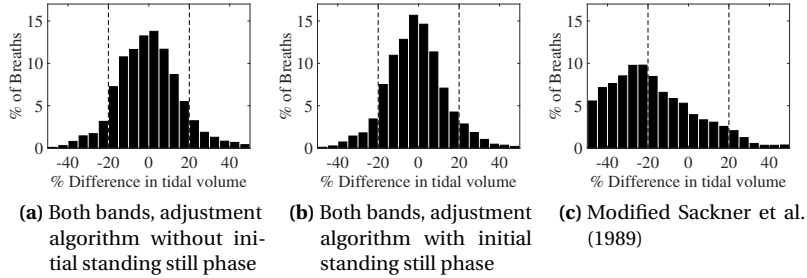


Figure 5.4: Frequency histograms during treadmill running using the validation dataset. % Difference in tidal volume was calculated using $V_T - V_{T_{FM}}$ with regard to $V_{T_{FM}}$, with V_T as the calculated tidal volume and $V_{T_{FM}}$ as the reference volume of the flowmeter (FM).

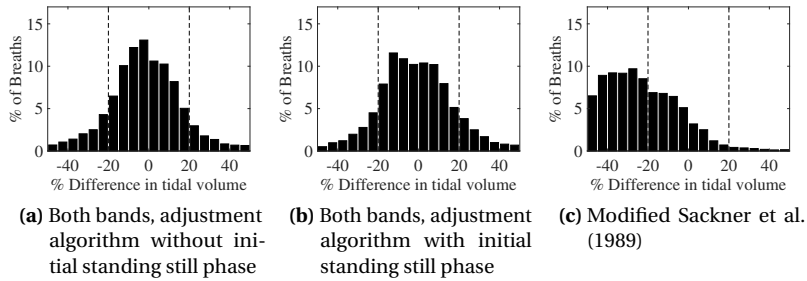


Figure 5.5: Frequency histograms during recovery using the validation dataset. % Difference in tidal volume was calculated using $V_T - V_{T_{FM}}$ with regard to $V_{T_{FM}}$, with V_T as the calculated tidal volume and $V_{T_{FM}}$ as the reference volume of the flowmeter (FM).

assessed when using a two-degree of freedom system (RC and AB) and a one-degree of freedom system (RC or AB).

Ventilatory data were recorded with both measurement devices (RIP and FM) to assess the performance of these algorithms. For an actual application in the field, this would require to collect anthropometric data of the subject, and maximal record both modalities during a standing still phase. Afterwards, these characteristics would be stored for direct application of incoming ventilatory data during field tests.

General discussion

The best results using both bands (RC and AB) for the algorithms were slightly better using the adjustment with the standing still phase, directly followed by the adjustment without the standing still phase (Figure 5.2, Figure 5.3, and Table 5.3). The adjustment worked better for the treadmill running phase than for the recovery phase (Table 5.3). This is probably due to the breath characteristics and the number of breaths that were used for the evaluation (Table 4.1, Table 4.2, and Table 4.3). The number of larger tidal volumes under conditions of treadmill running was higher compared to the recovery conditions where tidal volumes flattened out after the first two minutes.

The applicability of the determined adjustment parameters is given when comparing the performances between the validation and the reliability dataset (Figure 5.2 and Figure 5.3). Comparable results were obtained for the adjustment with and without the standing still phase. All subjects in the reliability dataset were not considered in any part of training the adjustment algorithm.

Comparing adjusted tidal volumes using both bands, the average difference was around 0.01 l, whereas only few breaths had to be excluded as they were considered outliers (Table 5.4). The range of the SDs were higher and lay between 0.16 l to 0.25 l. Up to now, to the best knowledge of the author, no adjustment algorithm for an outside setting has been suggested. Hence, as only a small amount of breaths were considered as outliers, an error of up to 0.25 l seemed acceptable for the first versions of reference-free adjustment algorithms. There is no doubt that the development of adjustment algorithms should continue.

The subject population consisted of 193 subjects. This is, to the best knowledge of the author, the study with the largest study population.

Potential of one band

The best results using only one band were obtained using the RC band alone compared to the AB band alone (Figures 5.2 and 5.3). The results using only the RC band were slightly worse compared to using the combination of both bands. Similarly to the adjustment using both bands, the adjustment algorithm with the standing still phase and the treadmill running phase revealed the best results. The fact that better results were obtained with the RC band is further supported by the adjustment algorithm that Banzett et al. (1995) suggested. Banzett et al. (1995) gave the RC

band double weighting compared to the AB band. Hence, this suggests that a higher impact is associated with the RC band.

The results obtained using the adjustment with the standing still phase outperformed the results without the standing still phase marginally. However, the differences between the results from both bands vs. only one band were more pronounced. Hence, one could argue that with regard to applicability, future work should be focused on improving the adjustment algorithm for a single band system in general and on improving the anatomical position of the single band.

Usability

Reference-free adjustment meant that this algorithm could be applied directly to the data acquired using the RIP device without any simultaneous data acquisition using the FM device. The differences between the adjustment with and without standing still phase were marginal (Figure 5.2), also when looking at the reliability dataset where no data of these subjects had been used in training the algorithm (Figure 5.3). The algorithm with standing still phase required the recording of solely RIP data in a 5 min standing still phase. Apart from RIP data and data also directly acquired with this device such as time characteristics, only anthropometric data such as age, sex, height, weight, RC, and AB circumference had to be available. This anthropometric data can be obtained using a bathroom scale and a measuring tape. After determining this anthropometric data for the first measurement, this information can be stored for further analysis. No FM device is needed during any part of the data acquisition. This enables the usability of RIP outside the controlled laboratory environment to gather field data in real-world situations.

Relation to implemented algorithms from literature

Three of the algorithms from literature needed the simultaneous data acquisition of RIP and FM data (Banzett et al., 1995; Chadha et al., 1982; Sackner et al., 1989). As fourth algorithm, a modified version of the Sackner et al. (1989) algorithm was implemented to compare with a reference-free algorithm.

All algorithms with simultaneous data acquisition of RIP and FM were mostly used for verification that accurate respiratory data can be acquired using the RIP device but were not applicable for data acquired using only the RIP device (Heyde et al., 2014). As RIP data were mapped directly to the FM data for each person individually using the direct versions of

the algorithms from the literature, it was not surprising that the three algorithms proposed by Chadha et al. (1982), Sackner et al. (1989), and Banzett et al. (1995) outperformed the proposed adjustment algorithms in this chapter. If comparing the results obtained with the proposed algorithms in this chapter to the modified version of the Sackner et al. (1989) algorithm, the proposed algorithms outperform this algorithm from literature. Here, even with the usage of only one band (RC or AB), better results were derived.

Comparing the direct implementations of the three algorithms from literature, the algorithm by Chadha et al. (1982) was postulated as the best performing adjustment algorithm for two-degrees of freedom RIP adjustment (Heyde et al., 2014; Leutheuser, Heyde, Gollhofer, & Eskofier, 2014). Hence, it was expected that this algorithm outperforms the other two algorithms presented in the literature (Banzett et al., 1995; Sackner et al., 1989), which was the case in this study. The algorithm proposed by Banzett et al. (1995) outperformed the algorithm by Sackner et al. (1989) slightly. In the algorithm suggested by Sackner et al. (1989) k -ratios K were calculated individually for each person. Banzett et al. (1995) assumed a fixed K ($K = 2$). The high results by Banzett et al. (1995) might be due to the fact that during exercise thoraco-abdominal contribution changes in favor of RC (Teramoto, Fukuchi, Nagase, Matsuse, & Orimo, 1995). This resulted in predicting only one degree of freedom (M in equation (2.5)). In future work, research could focus on predicting M with anthropometric data.

Regression slopes were calculated between derived values from the proposed reference-free adjustment algorithms as well as four algorithms from the literature (Banzett et al., 1995; Chadha et al., 1982; Sackner et al., 1989) and the reference measurement from the FM for the phases treadmill running and recovery (Table 5.3). The slopes were calculated incorporating an offset (equation (5.12)), whereas the adjustment algorithms described in the literature (Banzett et al., 1995; Chadha et al., 1982; Sackner et al., 1989) were designed without an offset. The general linear regression procedure was designed with an offset. Thus it was decided to incorporate the offset.

The slopes calculated using the algorithms from literature (Banzett et al., 1995; Chadha et al., 1982; Sackner et al., 1989) underestimated the relationship between RIP and FM (Table 5.3). The adjustment algorithm with standing still phase incorporating both bands predicted most accurately the slope, whereas a higher SD was obtained than the original versions of the algorithms from literature (Banzett et al., 1995; Chadha et al., 1982; Sackner et al., 1989). However, the range of this SD was in the same range

than the modified version of the algorithm by Sackner et al. (1989). In addition, the frequency histograms (Figure 5.4c and Figure 5.5c) show a clear left shift. Both the lower slopes and the shift are probably due to the additional degrees of freedom in axial direction which are not considered at higher tidal volumes. This is consistent with previous studies which showed that the measurement of the body surface displacement in axial direction slightly improves the accuracy (McCool, Kelly, Loring, Greaves, & Mead, 1986).

In contrast, the adjustment algorithm with a standing still phase using both bands applied to the recovery phase predicted the slope accurately (Table 5.3). Nine of the proposed algorithms overestimated the relationship between RIP and FM up to 1.17 and one up to 1.42. Accordingly, the left shift in the frequency histograms (Figures 5.4a, 5.4b, 5.5b, and 5.5a) was reduced. In fact, the frequency histogram using both bands and the suggested adjustment algorithms show a good fit in general. It was assumed that the integration of multiple predictors yields the optimized slopes ("1") and the reduced shift in the mean over all subjects.

Comparison to additional approaches in literature

In 2012, S. Liu, Gao, He, Staudenmayer, and Freedson (2012) presented an updated version of the standard two-degree of freedom model. They compared five models with different compositions of additional features such as volume features, frequency features, and body size features in the adjustment model based on the raw ventilatory data. The best model included four additional volume features and the breathing frequency. A comparison with the proposed adjustment algorithms to this approach was not feasible as only preprocessed instead of raw data was available.

Limitations

The suggested adjustment algorithms were solved using least squares approximation including a three-step outlier removal. This approach is only one of many possibilities. Extensions and modifications to this approach could include a varied outlier detection, solving the regression equation with least absolute deviations or L^p norms with $p > 2$ (equation (3.16)), or changing the feature selection procedure.

Chapter highlights

- The proposed adjustment algorithms are generalizable and reference-free during application for respiratory inductance plethysmography (RIP). Generalizable means that no prerecording of any data is needed. Reference-free stands for the fact that the adjustment works having solely RIP data.
- The adjustment algorithms include anthropometric data of the subject, time characteristics, and ratios of rib cage (RC) to abdomen (AB), whereas the latter two are acquired and instantaneously incorporated in the adjustment. All characteristics have to be known for the first measurement and are stored for all further measurements.
- The adjustment algorithms were assessed in the phases treadmill running and recovery. Better results were obtained for the treadmill running phase. The results from the validation and the reliability dataset suggest the (re-)usability of the adjustment for unseen subjects and the robustness for future data.
- The two-degree of freedom system with both bands (RC and AB) was compared to the two one-degree of freedom system (either RC or AB). Adjusting with both bands achieved slightly better results compared to considering only one band. The RC band outperformed the AB band.

Part III

Mobile ECG Analysis

Chapter 6

Arrhythmia Classification Using RR Intervals

In this chapter, arrhythmia classification is performed using only features from the time instances of individual heart beats using RR intervals. This is based on the phenomenon that consecutive RR intervals are unequal in duration. This addresses contribution (III). Parts of this chapter have been published previously (Leutheuser et al., 2015).

6.1 Methods

6.1.1 Data

The website PhysioNet (Goldberger et al., 2000) offers free access to a large collection of recorded physiological signals. All databases available from the PhysioNet website are placed into one category: class 1, class 2, or class 3 (PhysioNet, 2018). Reviewed and completely reference databases belong to class 1. Copies of raw data that support published research that are contributed by authors or journals are included in class 2. Other contributed collections of data like works in progress are part of class 3.

From there, two databases, the MIT-BIH Arrhythmia database (G. Moody & Mark, 2001) and the European ST-T database (Taddei et al., 1992), were used. These two databases are class 1 databases which were carefully analyzed and thoroughly annotated (Goldberger et al., 2000).

MIT-BIH Arrhythmia database

The MIT-BIH Arrhythmia database (G. Moody & Mark, 2001) consisted in total of 48 half-hour ECG recordings of 47 subjects (22 female and 25

male, age 64 ± 18 years, mean $\pm$ SD). Each recording was digitized at 360 Hz per channel. The 100er series include 23 recordings and the 200er series the remaining 25 recordings. The datasets 102, 104, 107, and 217 were excluded as they originated from patients having an implantable cardioverter-defibrillator.

European ST-T database

The European ST-T database (Taddei et al., 1992) consists of 90 two-hour ECG recordings of 79 subjects (8 female and 70 male, age 55 ± 11 years, mean $\pm$ SD). The information for one subject is missing. Each recording was digitized at 250 Hz.

Heartbeat Classes

The aim was to provide the user with a simple recommendation – seeing a physician. Hence, all beats with pathologic appearance were fused into the class pathologic and all beats with normal appearance were fused into the class normal. The pathologic class consisted of

- (left / right) bundle branch block beats,
- atrial / aberrated atrial / nodal (junctional) / supraventricular (atrial or nodal) premature beats,
- ectopic (atrial or nodal) beats,
- (R-on-T) premature ventricular contractions,
- atrial / nodal (junctional)/supraventricular (atrial or nodal) / ventricular escape beats, and
- unclassifiable/unclassified beats.

Preprocessing

In this chapter, only the performance of arrhythmia classification using RR intervals is of interest. Hence, instead of applying a QRS complex detection or R-peak detection algorithm, the provided reference annotation files were used. This is common in literature (Islam et al., 2016; Y. Li et al., 2017). Additionally, as stated in chapter 2.2.1, accurate R-peak detection algorithms exist that are at the same time suited for real-time implementation (Gradl et al., 2015).

For the classification task, specific RR interval data were needed. For calculating the sine function, a short number (10, 15, or 20, or generic N) of consecutive RR intervals was required. Hence, segments of RR intervals were extracted where N consecutive RR intervals were labeled as normal and the $(N + 1)$ th RR interval was either labeled as pathologic or normal. As far more normal labeled RR intervals than pathologic labeled RR intervals occurred, the segments with the $(N + 1)$ th RR interval labeled as pathologic were extracted first. Before extracting segments with the $(N + 1)$ th RR interval labeled as normal, all previously found segments were excluded to prevent overlapping.

As class imbalances might negatively affect the classification performance, the same amount of segments were used for both classes. The segments were chosen randomly in case of imbalance in the number of segments. This resulted in 5474 segments for the calculation with 10 RR intervals, in 4232 segments for the calculation with 15 RR intervals, and in 3534 segments for the calculation with 20 RR intervals.

6.1.2 Features

Sinusoidal Regression Feature

The N RR intervals were modeled using a sine function g_{RR} :

$$g_{RR}(t_n) = b_0 + A \cdot \sin(2\pi f t_n + \varphi) \quad (6.1)$$

with t_n as the discrete time values, b_0 as the constant offset, A as the amplitude, f as the frequency, and φ as the phase shift (Figure 6.1).

Equation (6.1) can be reformulated to

$$g_{RR}(t_n) = b_0 + A \cdot \sin(2\pi f t_n + \varphi) \quad (6.2)$$

$$= b_0 + b_1 \sin(2\pi f t_n) + b_2 \cos(2\pi f t_n) \quad (6.3)$$

with $A = \sqrt{b_1^2 + b_2^2}$ and $\varphi = \arctan(b_2/b_1)$.

The regression coefficients b_0 , b_1 , and b_2 were obtained using LSQ. The Moore-Penrose pseudoinverse using SVD was calculated under the assumption of a condition number smaller than 10^5 . Outliers on the independent and dependent variables were removed using standardized residuals, the Mahalanobis distance, and the Cook's distance (Stevens, 2009) as described in section 5.1.3 and by Leutheuser, Heyde, Roecker, et al. (2017).

The sinusoidal regression feature is the residual from the $(N + 1)$ th RR interval to the sinusoidal model.

Frequency Calculation for Sinusoidal Model

The sine function required a frequency f underlying the RR interval oscillation. Due to the phenomenon of RSA, this frequency is typical in the range of the breathing frequency (0.2 Hz) (Bowers & Murray, 2004; Lamine et al., 2004; Leutheuser et al., 2011). In this chapter, the frequency f was calculated using four different approaches. Three of them used the principle of Fourier transform. This means that signals in the time domain are transformed into the frequency domain using the Fourier transform. This conversion is reversible. This is used to determine the relevant frequencies of non-periodic time series. In signal processing, when the signal has discrete values and is finite, the FFT is used. Different algorithms for FFT computation exist. The most common one is the Cooley-Tukey algorithm, which was also used in this thesis.

The fourth approach was built on the Lomb-Scargle periodogram P_{LS} (Laguna, Moody, & Mark, 1998; Lomb, 1976; Scargle, 1982; VanderPlas, 2018). The normalized Lomb-Scargle periodogram P_{LS} has the form of a generalized periodogram and can be expressed with this equation:

$$P_{LS}(f) = \frac{1}{2\sigma^2} \left\{ \frac{\left(\sum_n (h_n - \bar{h}) \cos(2\pi f[t_n - \tau]) \right)^2}{\sum_n \cos^2(2\pi f[t_n - \tau])} + \frac{\left(\sum_n (h_n - \bar{h}) \sin(2\pi f[t_n - \tau]) \right)^2}{\sum_n \sin^2(2\pi f[t_n - \tau])} \right\}, \quad (6.4)$$

with h_n as the RR intervals, $\bar{h}$ and σ^2 as mean and variance of the RR intervals, and t_n as the discrete time values. To ensure time-shift invariance, τ was specified for each f as follows:

$$\tau = \frac{1}{4\pi f} \arctan \left(\frac{\sum_n \sin(4\pi f t_n)}{\sum_n \cos(4\pi f t_n)} \right). \quad (6.5)$$

The different approaches for frequency f calculation are detailed below:

1. Frequency f corresponded to the maximum of the magnitude FFT.
2. Frequency f corresponded, equal to method 1), to the maximum of the magnitude FFT. A Hamming window function was applied to the time-domain signal before FFT estimation (Clifford & Tarassenko, 2005).

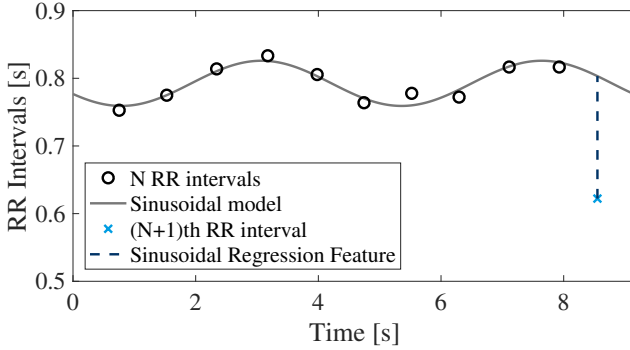


Figure 6.1: Visualization of the sinusoidal model and the sinusoidal regression feature. Ten RR intervals (o) are plotted over time. The sinusoidal model is fitted using least squares approximation (solid line). The sinusoidal regression feature (dashed line) is the amplitude between the eleventh RR interval (x) and the sine curve (solid line). Picture adapted from Leutheuser et al. (2015) © 2015 IEEE.

3. Frequency f was calculated as central frequency after magnitude FFT with Hamming window:

$$f = \frac{\sum_{j=-1}^1 X_{f_{k+j}} \cdot f_{k+j}}{\sum_{j=-1}^1 X_{f_{k+j}}} \quad (6.6)$$

with X as magnitude FFT of the RR intervals, where X_{f_k} corresponds to the maximum of the magnitude FFT located at the frequency f_k .

4. Frequency f was calculated as corresponding to the maximum peak in the Lomb-Scargle periodogram P_{LS} (Laguna et al., 1998; Lomb, 1976; Scargle, 1982; VanderPlas, 2018).

In all four methods, the segment was excluded when the maximum of the power spectra was centered at the boundaries (smallest or largest). Methods 1), 2), and 3) required evenly sampled signals. Hence, when using these approaches, the RR intervals were linearly interpolated with a sampling frequency of 7 Hz (Clifford & Tarassenko, 2005). Furthermore, the DC component was eliminated in method 1), 2), and 3) by subtracting the average RR interval from the considered window.

Feature Groups

Four feature types were extracted: 1) the sinusoidal regression feature as described above together with 2) the amplitude A and 3) the phase shift φ , and 4) the N RR intervals as well as the $(N + 1)$ th RR interval. As A and φ are distinct attributes of the sine model, these attributes were additionally included as features. To assess the influence of the sinusoidal regression feature, three FGs were defined: FG I consisted of all features, FG II consisted only of the sinusoidal regression feature, and FG III consisted only of the RR intervals (without the sinusoidal regression feature, A , and φ).

6.1.3 Classification

A Naive Bayes classifier (section 3.3.2) was used to classify the two types of heartbeat classes: normal and pathologic. Naive Bayes classifiers are based on supervised learning (section 3.3). First, the classifier is trained with a given training set of feature vectors and their corresponding class labels. Second, the classifier is tested by predicting the class label of a given new feature vector.

Naive Bayes classifiers belong to the family of probabilistic classifiers. For each new feature vector, posterior probabilities are computed for each class and the feature vector is assigned to the class corresponding to the higher posterior probability. The posterior probabilities are computed using the Bayes's theorem. Assuming that all features are conditionally independent given the class, the Naive Bayes classifier separately estimates class-conditional probabilities density functions for each class and each feature. Thus, the classifier is trained through estimating class-conditional probability density functions by means of a given training set of labeled feature vectors (Duda et al., 2000; Theodoridis & Koutroumbas, 2008). In this chapter, the class-conditional probability density functions were chosen to describe a normal or Gaussian distribution. The parameters of the normal distribution, namely mean and variance, were directly derived from the feature vectors of the training set. More detailed descriptions can be found in the textbooks by Duda et al. (2000) and Theodoridis and Koutroumbas (2008).

6.1.4 Performance Measures and Evaluation

In total, 36 different classification systems with different number N of considered RR intervals in creating the sinusoidal model, combinations of frequency calculations, and three different FGs were assessed (Table 6.1).

Table 6.1: Overview of the 36 classification systems.

N in sinusoidal model	1. 10 RR intervals 2. 15 RR intervals 3. 20 RR intervals
Frequency calculation	1. Fast Fourier Transform (FFT) 2. FFT with Hamming window 3. Central FFT with Hamming window 4. Lomb-Scargle periodogram
Feature groups (FGs)	1. All features 2. Sinusoidal regression feature 3. RR intervals

The predicted heartbeat classes were compared with all true heartbeat classes. Heartbeats predicted as pathologic and being pathologic were counted as true positives ($\#TP$). Heartbeats predicted as pathologic whereas being normal were counted as false positives ($\#FP$). Heartbeats predicted as pathologic whereas being normal were counted as false negatives ($\#FN$). Heartbeats predicted as pathologic by being normal were counted as false negatives ($\#FP$). This assignment is also shown in Table 6.2, where a general confusion matrix with the two classes pathologic and normal is given.

The classification system were compared regarding five evaluation metrics: the sensitivity SE , the specificity SP , the balanced accuracy BA , the positive predictivity $+P$, and the accuracy ACC . The sensitivity SE is also known as recall or true positive rate:

$$SE = \frac{\#TP}{\#TP + \#FN} \quad (6.7)$$

The sensitivity SE gives the proportion of pathologic beats, which were correctly classified as pathologic. The specificity SP is also known as true negative rate and is calculated with following equation:

$$SP = \frac{\#TN}{\#TN + \#FP} \quad (6.8)$$

The specificity SP represents the proportion of normal beats that were correctly classified as normal. The balanced accuracy BA is the mean of the sensitivity SE and the specificity SP :

$$BA = \frac{1}{2}(SE + SP) \quad (6.9)$$

Table 6.2: Confusion matrix for the two classes pathologic and normal.

		Predicted condition	
		Pathologic	Normal
True condition	Pathologic	$\#TP$	$\#FP$
	Normal	$\#FN$	$\#TN$

$\#TP$: true positives; $\#FP$: false positives; $\#FN$: false negatives; $\#TN$: true negatives

The positive predictivity $+P$ is also known as precision:

$$+P = \frac{\#TP}{\#TP + \#FP} \quad (6.10)$$

The accuracy ACC represents the overall performance of the algorithm and is calculated by

$$ACC = \frac{\#TP + \#TN}{\#TP + \#TN + \#FP + \#FN} \quad (6.11)$$

The performances were evaluated using 10-fold cross-validation (chapter 3.3.3). The evaluation metrics were calculated for each fold. Afterwards, each evaluation metric was averaged individually and stated with the corresponding SD.

6.2 Results

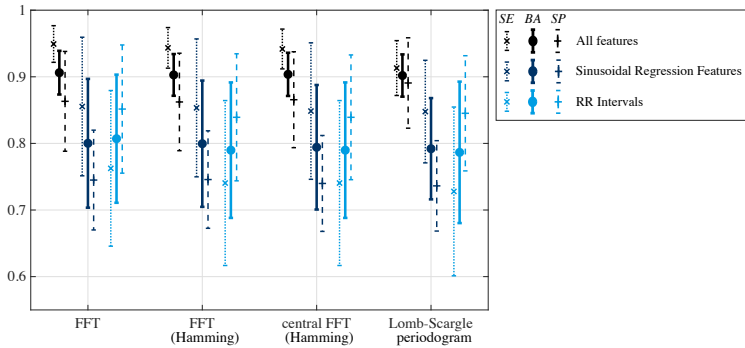
Depending on the frequency calculation, some segments had to be excluded as the frequency calculation failed (Table 6.3). It was assumed that neither the first nor the last position in the frequency domain represented the exact frequency component if the maximum frequency occurred at these values.

The effect of the different FGs on sensitivity SE , balanced accuracy BA , and specificity SP is shown in Figure 6.2. The results are shown for 10 RR intervals. In the appendix, Figure B.1 and Figure B.2 show the effect of different FGs using 15 and 20 RR intervals, respectively.

Figure 6.3 shows the sensitivity SE , the balanced accuracy BA , and the specificity SP for the four different frequency calculation methods considering all features. Table 6.4 gives the positive predictivity $+P$ and accuracy ACC using the same conditions as in Figure 6.3. The positive predictivity $+P$ and the accuracy ACC for all three RR interval options

Table 6.3: Number of segments with invalid frequency calculation. These segments were excluded in the following calculations.

Frequency calculation	RR intervals		
	10	15	20
Fast Fourier Transform (FFT)	0 (0%)	0 (0%)	0 (0%)
FFT with Hamming window	102 (1.9%)	56 (1.3%)	43 (1.2%)
central FFT with Hamming window	102 (1.9%)	56 (1.3%)	43 (1.2%)
Lomb-Scargle periodogram	313 (5.7%)	240 (5.7%)	193 (5.5%)
Total number of segments	5474	4232	3534

**Figure 6.2:** Effect of different feature group (FG) on sensitivity SE , balanced accuracy BA , and specificity SP for RR interval classification using the four different frequency calculation methods. Results are shown using 10 RR intervals. The evaluation metrics are averaged over the 10-folds and are shown with the standard deviations as error bars. FFT: Fast Fourier Transform.

and the four frequency options are given in Table B.1 and Table B.2 in the appendix, when considering only the sinusoidal regression feature and the RR intervals, respectively.

6.3 Discussion

The aim of this chapter was to build an arrhythmia classification algorithm only with RR intervals. It was assumed that only a small number of RR intervals was needed to classify into the classes pathologic and normal. Further, it was suggested that a sinusoidal regression feature

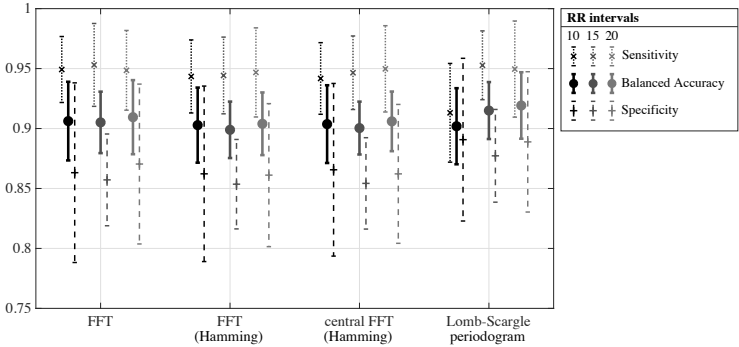


Figure 6.3: Sensitivity, balanced accuracy, and specificity for RR interval classification using the four different frequency calculation methods shown when considering all features. The evaluation metrics are averaged over the 10-folds and are shown with the standard deviations as error bars. FFT: Fast Fourier Transform.

which resulted from fitting a sine curve to the RR intervals improved the performance.

General discussion

The combination of using all features (sinusoidal regression feature, RR intervals, A and φ) achieved the best results (Figure 6.2, Figure B.1, and Figure B.2). When looking at single FGs, the sinusoidal regression feature resulted in great performances for the sensitivity and lower performances for the specificity, whereas it was exactly the opposite for the RR intervals for a segment duration of 10 RR intervals (Figure 6.2). In order to reduce false alarms, a high sensitivity and specificity should be aimed for, recommending the combination of both features. The feature set all features comprised these two features together with two parameters of the sine curve as additional features. The feature set RR intervals was independent from the frequency calculation (Figure 6.2). The small differences resulted from the exclusion of segments due to invalid frequency calculation. The exclusion led either to higher performances or lower performances (Table B.2).

The four frequency calculation methods and the three RR interval segment lengths had few influences on the best feature set (Figure 6.3 and Table 6.4). In order to fit a sine curve to the RR intervals, the frequency of the underlying oscillation had to be determined. Four different frequency

Table 6.4: Evaluation metrics in percent (%) over the different frequency calculation methods when considering all features (mean $\pm$ standard deviation). Segments with invalid frequency calculation were excluded (Table 6.3).

	Frequency options			
	FFT	FFT (Hamming)	central FFT (Hamming)	Lomb- Scargle periodogram
Positive Predictivity				
10 RR intervals	83.95 $\pm$ 10.51	83.95 $\pm$ 10.35	84.47 $\pm$ 9.99	87.89 $\pm$ 9.36
15 RR intervals	83.83 $\pm$ 5.03	83.55 $\pm$ 4.87	83.58 $\pm$ 4.98	86.09 $\pm$ 5.01
20 RR intervals	85.04 $\pm$ 10.12	84.04 $\pm$ 8.81	84.16 $\pm$ 8.52	87.12 $\pm$ 8.75
Accuracy				
10 RR intervals	89.62 $\pm$ 4.49	89.33 $\pm$ 4.30	89.51 $\pm$ 4.26	89.65 $\pm$ 3.89
15 RR intervals	89.79 $\pm$ 2.70	89.23 $\pm$ 2.54	89.35 $\pm$ 2.42	90.97 $\pm$ 2.58
20 RR intervals	90.09 $\pm$ 4.50	89.50 $\pm$ 3.65	89.70 $\pm$ 3.50	91.24 $\pm$ 3.86

FFT: Fast Fourier Transform

calculations were suggested and the differences in the results are negligible. Best sensitivities were achieved with the simple FFT. Best specificities were obtained with the Lomb-Scargle periodogram, whereas also most segments had to be excluded with this frequency calculation method (Table 6.3). Hence, it is suggested to use FFT as frequency calculation method as the smallest number of segments had to be eliminated. In future work, other algorithms for determining the underlying frequency could be assessed like using the ECG-derived respiration technique or determining the breathing frequency with an additional sensor. All RR interval segment lengths led to best performances in some cases. For a real-time implementation, as few heartbeats as possible are desired proposing to use the minimum amount of RR intervals.

Usability

In this chapter, the arrhythmia classification algorithm is based only on 10 RR intervals. This is a small enough number to be used in a real-time and real-life setting. The proposed algorithm needed, in the current version, normal heartbeats respectively RR intervals preceding an pathologic heart beat. The high performances, even with this boundary constraint, indicate that real-time algorithms with simple features are feasible and achievable.

Comparison to literature

The algorithm presented in this chapter did not focus on atrial fibrillation detection alone as most of the algorithms presented in chapter 2.2.2. Atrial fibrillation is a type of abnormal heart rhythm (arrhythmia). In this chapter, all beat labels indicating a pathologic appearance were merged into the class pathologic. All other beats were fused into the class normal. This resulted in a whole different task than the detection of atrial fibrillation. To the best knowledge of the author, only two publications were found that concentrated on arrhythmia detection using RR intervals (Asl et al., 2008; Tsipouras et al., 2005).

Nevertheless, this algorithm needs to be compared with reimplemented versions of existing algorithms in literature to the same data to compare the performances.

Adaptations compared to original publication

Due to a change in the MATLAB package (Mathworks Inc., Natick, MA, US) version compared to the original publication (Leutheuser et al., 2015), the function for training the Naive Bayes classifier was updated to `fitcnb`. In the original publication, the `NaiveBayes.fit` function was used. The performance measures were expanded with the sensitivity SE , the specificity SP , the positive predictivity $+P$, and the accuracy ACC . In the original publication, the results for the mean-class dependent classification rate were solely presented. The balanced accuracy is the same as the true positive rate when defining the same conditions for the $\#TP$, $\#FP$, $\#FN$, and $\#TN$. The confusion matrix (Table 6.2) gave best credit to pathologic beats.

Extensions to the classification systems

In this chapter, the Naive Bayes classifier was used to investigate the performance for detecting pathologic beats. In general, no single classifier is suited for all classification tasks (No Free Lunch Theorem) (Duda et al., 2000) (chapter 3.3). Hence it might be that a different classifier would improve the performance.

Additional features could be added to the FGs, like HRV parameters (Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, 1996), Shannon entropy (Dash et al., 2009; J. Lee, Nam, et al., 2013; McManus et al., 2013; Zhou et al., 2014), or any mentioned feature from Table 2.2. Apart from using distinct FGs, a feature selection procedure like forward feature selection could be

investigated. Additionally, one should have in mind the computational costs for each classification step, especially when desiring a real-time implementation on wearables (Jensen, Kugler, Ring, & Eskofier, 2015; Ring, Jensen, Kugler, & Eskofier, 2012).

Limitations

By calculating the sinusoidal regression feature (and the amplitude A and the phase φ), a linear regression task was solved. The criteria for a good fit are indicated by small root mean squared error ($RMSE$, equation (4.3)). In this chapter, the validity of the sinusoidal model was not regarded as the performance of the classification systems including these features were of sole interest.

The data were preprocessed to obtain RR interval segments where N consecutive heartbeats/RR intervals were labeled as normal. This resulted in excluding all segments not fulfilling this condition. Having only one pathologic heart beat is no sufficient indicator for seeing a physician according to the Lown classification (Lown & Wolf, 1971). The sinusoidal regression feature, A , and φ are dependent on the current RR interval as RR intervals are plotted over time. Further investigation is needed how to address this issue. A possible solution could be to use decision level fusion based on the results of different classifiers.

Chapter highlights

- One RR interval is the difference of the time instances of two successive heartbeats. With N (10, 15, or 20) consecutive RR intervals, features were extracted and fed into a Naive Bayes classifier to distinct between pathologic and normal beats.
- With the phenomenon of respiratory sinus arrhythmia (RSA) or heart rate variability (HRV), a sine curve was fitted to RR interval segments. Features generated by this procedure were the amplitude A , the phase φ , the sinusoidal regression feature – the distance between the $(N + 1)$ th RR interval and the sine curve –, and the RR intervals by themselves. The best feature set consisted of using all of these features.
- Best real-time performances were obtained with the smallest segment length of 10 RR intervals and the Fast Fourier Transform (FFT) as frequency calculation method.

Chapter 7

ECG Fiducial Points Detection Algorithms

In this chapter, single-lead ECG delineation algorithms for real-time analysis are presented. The algorithms are compared regarding their performance for P- and T-peak detection. This addresses contribution (IV). Parts of this chapter have been published previously (Leutheuser et al., 2016).

7.1 Methods

7.1.1 Data

In total, three different databases were used: two publicly available databases from the Physionet website (Goldberger et al., 2000) and one self-recorded database. From the Physionet website (Goldberger et al., 2000), the MIT-BIH Arrhythmia database (G. Moody & Mark, 2001) and the QT database (Laguna et al., 1997) were used. From these two databases, all records were used that contained ECG data from lead II according to Einthoven and that had annotations for the P- and T-wave available. For the self-recorded database called Short-Term FAU database, ECG data from lead II were acquired using a wearable sensor in clinical setting. Table 7.1 gives an overview of the used records for the three databases.

MIT-BIH Arrhythmia database

This database consists of 48 half-hour ECG recordings that were digitized at 360 Hz per channel (see chapter 6.1.1). In this chapter, the datasets 102 and 104 were excluded due to missing lead II ECG data (Table 7.1). The

Table 7.1: Overview of databases (DB), the used records, and the number of annotated beats (# Beats) for electrocardiogram (ECG) fiducial points detection.

DB	Record names	# Beats
MIT-BIH Arrhythmia DB (G. Moody & Mark, 2001)	100, 101, 103, 105, 106, 107, 108, 109, 111, 112, 113, 114, 115, 116, 117, 118, 119, 121, 122, 123, 124, 200, 201, 202, 203, 205, 207, 208, 209, 210, 212, 213, 214, 215, 217, 219, 220, 221, 222, 223, 228, 230, 231, 232, 233, 234	102,218 (P-wave); 104,873 (T-wave)
QT DB (Laguna et al., 1997)	sel100, sel103, sel114, sel116, sel117, sel123, sel213, sel223, sel230, sel231, sel233	337
Short-Term FAU DB	01, 02, 04, 05, 06, 08, 09, 10	427

P- and T-peak annotations were provided by Elgendi et al. (2015) in addition to the already existing R-peak annotations available on the PhysioNet website (Goldberger et al., 2000).

QT database

This database consists of 105 fifteen-minute ECG recordings with a sampling rate of 250 Hz. These 105 records originated from seven different databases. Several records had to be excluded due to incomplete information regarding the ECG data. In each record, at least 30 normal heartbeats were manually annotated by cardiologists. The annotation files provided by one cardiologist with ending “.1qc” were used. The heartbeats could be annotated with the beginning, peak, and end of the P-wave, the QRS complex, and the beginning, peak, and end of the T-wave. Though, the annotated heartbeats did not have to be consecutive. For determining the accurate performance of individual delineation algorithms only groups of contiguous annotated heartbeats were used where the peak of the P-wave and the peak of the T-wave were annotated. Otherwise, it was not possible to grade a detected wave event as true positive, false positive, or false negative. The exact groups of heartbeats can be found in Table 7.2.

Short-Term FAU database

This database consists of 33 ECG recordings that were sampled at 512 Hz. Mean record duration was 46 s. Of these recordings, the first 10 were

Table 7.2: Overview of used annotations from the QT database (Laguna et al., 1997).

Record names	Groups of used contiguous heartbeat annotations
sel100, sel103, sel117, sel123, sel223	{1–30}
sel114	{1–30}, {31–35}, {36–50}
sel116	{8–30}, {31–37}
sel213	{1–19}, {20–35}
sel230	{1–34}, {45–50}
sel231	{1–36}, {37–50}
sel233	{17–21}

indented for P- and T-wave detection. The datasets 03 and 07 of these had to be excluded: One dataset originated from a patient having an assisted device. In the other dataset, P- and T-wave labeling was not feasible due to too much movement in the upper electrodes.

The ECG data were collected using one Shimmer 2R (Shimmer Research, Dublin, Ireland) sensor node (Burns et al., 2010) with ECG module. Surface Ag/AgCl electrodes were placed at the torso according to the Einthoven triangle (chapter 3.1.2). These were wired to the Shimmer sensor node. The raw ECG data were transmitted via Bluetooth from the sensor node to an Android device and shown there in real-time. The data were simultaneously stored on the Android device. The acquired data consisted of all three leads according to Einthoven.

All subjects rested in supine position during the ECG recording. A 12-channel clinical ECG was acquired simultaneously while measuring.

Subjects aged 58 ± 17 years (mean $\pm$ SD) gave written informed consent for participation in the clinical study. The clinical study was approved by the local Ethics Committee of the University.

The different annotations were generated in a two-step process. First, the R-peaks were detected automatically using only lead II ECG data and manually realigned by a researcher like adding or deleting the position of the found R-peaks in case of misdetection of the automatic detection. After R-peak detection, the beginning, peak, and end of the P- and T-wave were annotated by a researcher after training by a cardiologist. For the P- and T-wave annotations, all three leads were considered. Second, a cardiologist verified the position of the R-peaks and the beginning, peak, and end of the P- and T-waves, and provided the class label to each corresponding heartbeat. For the verification done by the cardiologist, a

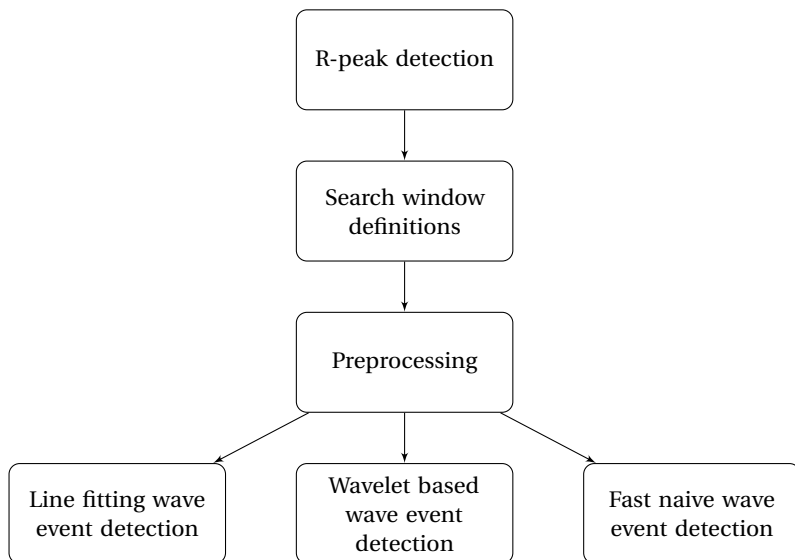


Figure 7.1: Analysis steps for electrocardiogram (ECG) fiducial points detection.

Graphical User Interface (GUI) in the MATLAB package (Mathworks Inc., Natick, MA, US) was created considering all three leads.

7.1.2 Algorithm Overview

The procedure for detecting the P- and T-peak consisted of four steps (Figure 7.1). First, R-peaks were detected. Second, the search window for the corresponding wave event was determined. Third, the signal was either preprocessed (band-pass filtered) or not. Fourth, the corresponding wave event was detected with all of the three different wave event detection algorithms. Afterwards, the performance was assessed.

R-peak detection

For the automated detection of QRS-complexes, an adapted version of the method suggested by Elgendi (2013) with an additional R-peak position refinement post-processing step (Gradl et al., 2015) was used. This method had been published previously in a conference proceeding (Gradl et al., 2015) where the author of this thesis was a co-author. First, the ECG signal is passed through a third-order Butterworth filter with a pass-band

between the frequencies 8 Hz and 20 Hz. The resulting signal is squared and used as input for the calculation of three moving average values. Using the moving average values, a region of interest was generated. Inside this region, the actual R-peak was determined by searching for the largest absolute value in the bandfiltered signal. This position was mapped to the original signal. Considering lead II according to Einthoven, the R-peak position was refined by searching for the positive maximum of the signal (Gradl et al., 2015).

Search window definitions

In literature, search windows for ECG delination are suggested that depend on the current R-peak or QRS-position and the current heart rate (Bote et al., 2018; Di Marco & Chiari, 2011; J. P. Martínez et al., 2004). Whereas, sometimes, no additional information was given about the exact definition of the search windows (Akhbari, Shamsollahi, Sayadi, et al., 2016; Beraza & Romero, 2017; Di Marco & Chiari, 2011; Ghaffari et al., 2010; Jané et al., 1997; Laguna et al., 1994; C. Li et al., 1995; A. Martínez et al., 2010; J. P. Martínez et al., 2004; Sun et al., 2005). For the sake of completeness, there exist algorithms in literature, that do not need specific search windows (Akhbari, Shamsollahi, Jutten, et al., 2016; Elgendi et al., 2015; Elgendi et al., 2016; C. Lin et al., 2014; Panigraphy & Sahu, 2018). Hence, the following rules were used for the search windows (Figure 7.2):

Search windows for P-wave detection (Figure 7.2): The start of the search window was chosen out of three options: {0.3, 0.35, 0.4} of the current RR-interval (RR_C) before the current R-peak. The duration of the Q-peak (left to the R-peak) to the R-peak was estimated to 0.06 s. This constituted the end of the search interval.

Search windows for T-wave detection (Figure 7.2) The start of the search window was set to the S-peak (right to the R-peak). The duration of R-peak to S-peak was estimated to 0.06 s. The end of the search window was chosen out of three options: {0.5, 0.55, 0.6} of the subsequent RR-interval (RR_S).

These rules were designed based on input from an expert cardiologist. Using these definitions, three beats were needed to determine P- and T-peak. The P-peak of the first heartbeat was not computed. The T-peaks of the first and last heartbeat were not computed. These search window definitions were suited for real-time implementation.

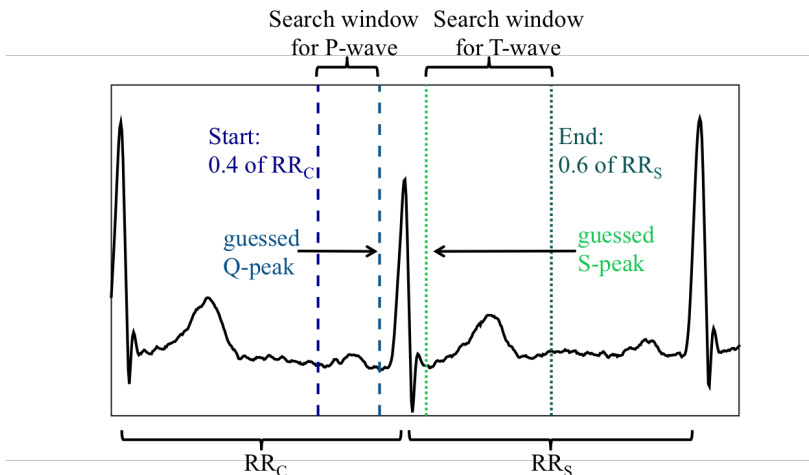


Figure 7.2: Illustration of the defined search windows within the wave event detection algorithms for the P- and T-wave. The Q-peak was guessed to 0.06 s. The S-peak was guessed to 0.06 s. RR_C : current RR-interval; RR_S : subsequent RR-interval.

Preprocessing

Preprocessing steps like band-pass filtering of the signal can enhance the performance in detecting individual wave events. Often, no specific or recommended preprocessing step was mentioned in previous work. Hence, the performance of a second-order Butterworth filter with a band-pass from 0.5 Hz to 15 Hz was compared to the application of no preprocessing step.

Wave event detection algorithms

Three different wave event detection algorithms were used. Two of them were real-time adaptations from published algorithms (Hu et al., 2014; J. P. Martínez et al., 2004). All three algorithms and the adaptations are described in the following sections. The reference MATLAB (Mathworks Inc., Natick, MA, US) implementations are available for download (<https://github.com/HeikeLeutheuser/WaveED/>).

7.1.3 P- and T-Wave Event Detection Algorithms

Line fitting wave event detection

Hu et al. (2014) suggested an R-peak detection algorithm based on line fitting. Cosoli et al. (2015) showed that this algorithm was suited for the detection of various wave events in four different biosignals. The algorithm from Hu et al. (2014) was adapted to allow the detection of wave events like P- or T-peaks. The algorithm incorporated the morphology of the ECG signal in the detection of individual waves and peaks. Two straight lines were fitted to consecutive ECG signal values. The two lines intersected in the ECG signal value in the middle. A peak was detected, when the angle between these two straight lines fulfilled certain conditions. Figure 7.3 illustrates this algorithm.

In the following, a detailed description of the adaptations and the algorithm is given. For this, it was assumed that the intersection point of the two straight lines was at ECG sample n . The first line was fitted from signal values $n - k$ to n . The second line was fitted from signal values n to $n + k$. Line fitting was realized by solving a linear regression equation (polynomial of degree 1) with respect to the corresponding time intervals using LSQ. The slopes of the two lines could be transferred to degrees (or radians) using the inverse tangent. In the following the angle θ_1 referred to the angle from the first line to the horizontal axis, and θ_2 to the angle from the second line to the horizontal axis (Figure 7.3). This was the same procedure as described by Hu et al. (2014).

First, within the search window for the corresponding wave event (P- or T-peak), two angles were calculated in the range of $[-90^\circ; 90^\circ]$. Next, θ , defined as the acute angle between the two fitted lines, was calculated. In the calculation of θ , it was taken into consideration that an upwards or downwards oriented wave event could occur. Afterwards, the smallest (positive) θ -angle below 180° in the search window was determined. This and the constraints for θ_1 and θ_2 were modified compared to the original publication (Hu et al., 2014). One of the angles (θ_1 or θ_2) had to be smaller than 0° and the other had to be larger than 0° .

Hu et al. (2014) used fixed k values for the detection of large and small waves. They used $k = 5$ for large waves and $k = 3$ for small waves with a sampling rate of 500 Hz. $k = 5$ equaled the duration of 10 ms. The k value was adapted to the actual sampling rate of the data so that the signal durations of 10 ms were incorporated in one line fitting. To not use a non-integer number of samples, the value was mathematically rounded.

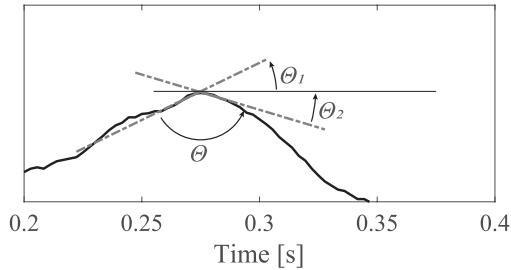


Figure 7.3: Illustration of line fitting wave event detection algorithm. Two straight lines (dashed) were fitted to the electrocardiogram (ECG) signal. θ_1 and θ_2 were angles representing the slopes of the two lines to the horizontal axis. θ was the acute angle between these lines and was calculated using θ_1 and θ_2 . The smallest positive θ below 180° in the corresponding search window corresponded to the highest peak that was desired. The image was adapted from Hu et al. (2014) and Leutheuser et al. (2016). © 2016 IEEE.

Wavelet based wave event detection

J. P. Martínez et al. (2004) suggested a Wavelet based ECG delineator algorithm. This algorithm was able to detect the QRS-complex, P- and T-wave, and, additionally, to perform QRS-complex, P-, and T-wave delineation. The algorithm was adapted so that P- and T-wave peaks could be detected in the given search windows.

J. P. Martínez et al. (2004) applied the *algorithme à trous* to the complete ECG signal and investigated the high-pass signal values at different scales. Here, the different scales were only calculated for the signal between two adjacent R-peaks where the corresponding search window was located. The FIR filters for the decomposition of the signal $H(z)$ and $G(z)$

$$\begin{aligned} H(z) &= \frac{1}{8} \cdot (z^2 + 3z + 3 + z^{-1}) \\ G(z) &= 2 \cdot (z - 1) \end{aligned} \quad (7.1)$$

were provided for a sampling rate of 250 Hz of the incoming signal. In case of different sampling rates, instead of changing the filter coefficients, the sampling rate of the incoming signal (e.g. downsampling or upsampling) was adapted.

As described by J. P. Martínez et al. (2004), it was first looked for local maxima in the absolute of the high-pass filtered signal at scale k $|W_{2^k} x[n]|$. $x[n]$ was the input signal. Figure 7.4 shows the flowchart of the calculation

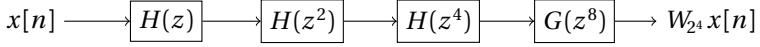


Figure 7.4: Flowchart of the Wavelet decomposition. The Wavelet based wave event detection used for Wavelet decomposition the *algorithme à trous*. Here, this is shown up to scale $k = 4$.

of the signal values at different scales. If at least two local maxima were above the threshold ϵ_{wave} , a wave event was considered to be present. Afterwards, all local maxima with amplitude above γ_{wave} were identified. The zero-crossings in scale $k = 3$ were considered as the corresponding wave peaks. If no zero-crossings could be found in scale $k = 3$, the scale k in which the local maxima were detected were chosen. In the first round of local maxima search, k was set to 4. If no local maxima / zero-crossings were identified, the procedure was repeated using $k = 5$. The same thresholds as suggested by J. P. Martínez et al. (2004) were used.

Fast naive wave event detection algorithm

The naive wave event detection algorithm detected the maximum peak in each search window. This maximum peak was assigned as the corresponding peak of the wave event.

7.1.4 Performance Measures and Evaluation

All detected wave events were compared with all labeled wave events. Matched wave events within a range of $\pm$ error interval were counted as true positives ($\#TP$). The performance was calculated for several error intervals to compare the performance with other algorithms in literature (Elgendi et al., 2015; Rincón et al., 2011). The error intervals were chosen from the following values {25 ms, 30 ms, 40 ms, 50 ms, 80 ms, 100 ms, 125 ms, 150 ms}. Annotated waves not detected by the algorithm were counted as false negatives ($\#FN$). Detected wave events that did not match any annotated wave were counted as false positives ($\#FP$). For performance assessment, the sensitivity SE (equation (6.7)) and the positive predictivity $+P$ (equation (6.10)) were calculated.

The mean absolute error MAE was calculated for all true positive detected wave events. First, the mean absolute error within each record MAE_{record} was determined as positive time difference between the provided gold standard wave event annotations and the detected wave events. Second, MAE was then calculated as the average of these individual

Table 7.3: Summarized overview of data, preprocessing, and algorithms for electrocardiogram (ECG) fiducial points detection.

Databases	1. MIT-BIH Arrhythmia database (Goldberger et al., 2000; G. Moody & Mark, 2001) with annotations provided by Elgendi et al. (2015) 2. QT database (Laguna et al., 1997) 3. Short-Term FAU database	
Search Windows	P-Wave	T-Wave
	0.30 of RR_C	0.50 of RR_S
	0.35 of RR_C	0.55 of RR_S
	0.40 of RR_C	0.60 of RR_S
Filtering	1. none 2. Second-order band-pass Butterworth filter from 0.5 Hz to 15 Hz	
Algorithms	1. Line fitting wave event detection (revised (Hu et al., 2014)) 2. Wavelet based wave event detection (adapted (J. P. Martínez et al., 2004)) 3. Fast naive wave event detection algorithm	

RR_C : current RR-interval; RR_S : subsequent RR-interval

MAE_{record} . Additionally, the mean of the intrarecording standard deviations SD was computed. After calculating the standard deviation within each record, the average was calculated to obtain SD . If the algorithm detected no true positives $\#TP$ in a specific record, this record had the $SE = 0$ and $+P = 0$, and was excluded in the calculation of MAE and SD . The number of records excluded is stated in the corresponding results.

The performance was assessed using different options for each step. These options are summarized in Table 7.3.

7.2 Results

For P-wave detection, record 04 of the Short-Term FAU database was excluded from the results. In this record, the P-wave was not existent and hence not annotated. Though, all algorithms detected for almost every heart beat or RR interval a P-wave peak.

Best performances were most often achieved with the smallest search window. Exceptions are stated in Table 7.4. All following results were given when using the search window with which best performances were obtained.

Table 7.4: Sensitivities SE in percent (%) for P- and T-wave detection comparing results of the best search window to the smallest search window. The smallest possible search window for the P-wave was 0.3 of the current RR-interval (RR_C) and for the T-wave 0.50 of the subsequent RR-interval (RR_S).

						Sensitivities SE	
		Error interval	Filtering	Algorithm	Best search window	Best search window	Smallest search window
QT DB	P-wave	40 ms	-	WT	0.35 of RR_C	88.10	73.68
				WT	0.35 of RR_C	77.69	74.53
	T-wave	80 ms	BT	LF	0.55 of RR_S	63.77	62.19
MIT-BIH Arrhythmia DB	T-wave	80 ms	-	WT	0.60 of RR_S	67.61	63.64
			-	NA	0.55 of RR_S	89.69	89.30
			BT	LF	0.60 of RR_S	45.53	42.32
			BT	NA	0.60 of RR_S	79.43	76.89

DB: database; BT: Butterworth filter; WT: Wavelet based wave event detection; LF: Line fitting wave event detection; NA: Fast naive wave event detection

Figure 7.5 shows the sensitivity SE of the different wave event detection algorithms for the P- and T-wave over different error intervals. The sensitivities SE for the three databases are shown in distinct plots. Table 7.5 and Table 7.6 give the results for positive predictivity $+P$, mean absolute error MAE , and SD for the P- and T-wave, respectively.

Most reported algorithms from literature reported their results using the QT database. Table 7.7 compares the best algorithms for P- and T-wave detection using 17 algorithms from literature (Akhbari, Shamsollahi, Jutten, et al., 2016; Akhbari, Shamsollahi, Sayadi, et al., 2016; Bayasi et al., 2014; Bote et al., 2018; Di Marco & Chiari, 2011; Dumont et al., 2010; Elgendi et al., 2015; Friganovic et al., 2018; Jané et al., 1997; Lenis et al., 2016; C. Lin et al., 2014; A. Martínez et al., 2010; J. P. Martínez et al., 2004; Mazomenos et al., 2013; Panigraphy & Sahu, 2018; Rincón et al., 2011; Vila et al., 2000). For P-peak detection, the Wavelet based wave event detection algorithm was used with an error interval of 40 ms. For T-peak detection, the fast naive wave event detection algorithm was used with an error interval of 80 ms.

7.3 Discussion

In this chapter, three P- and T-wave detection algorithms were presented and compared. These were based on either line-fitting, discrete Wavelet transform, or incorporating morphological knowledge. All three algorithms were able to run in real-time for instantaneous detection using single-lead ECG data (lead II of Einthoven). The performance was assessed using the MIT-BIH arrhythmia database (G. Moody & Mark, 2001), the QT database (Laguna et al., 1997), and a self-recorded database – the Short-Term FAU database.

General discussion

The fast naive wave event detection algorithm without filtering performed best in all databases for P- and T-wave detection (Figure 7.5, Tables 7.5 and 7.6). The second best scoring algorithm was the Wavelet based event detection algorithm. As best results were obtained for this algorithm when detecting P-peaks within the error interval of 40 ms, this algorithm was compared to algorithms from literature. However, this algorithm also revealed the worst performance for T-wave detection using the QT database. In this matter, the Wavelet based algorithm detected no T-wave in five records, within the respective error interval. When no $\#TP$ beats were detected, SE and $+P$ of these records were set to zero.

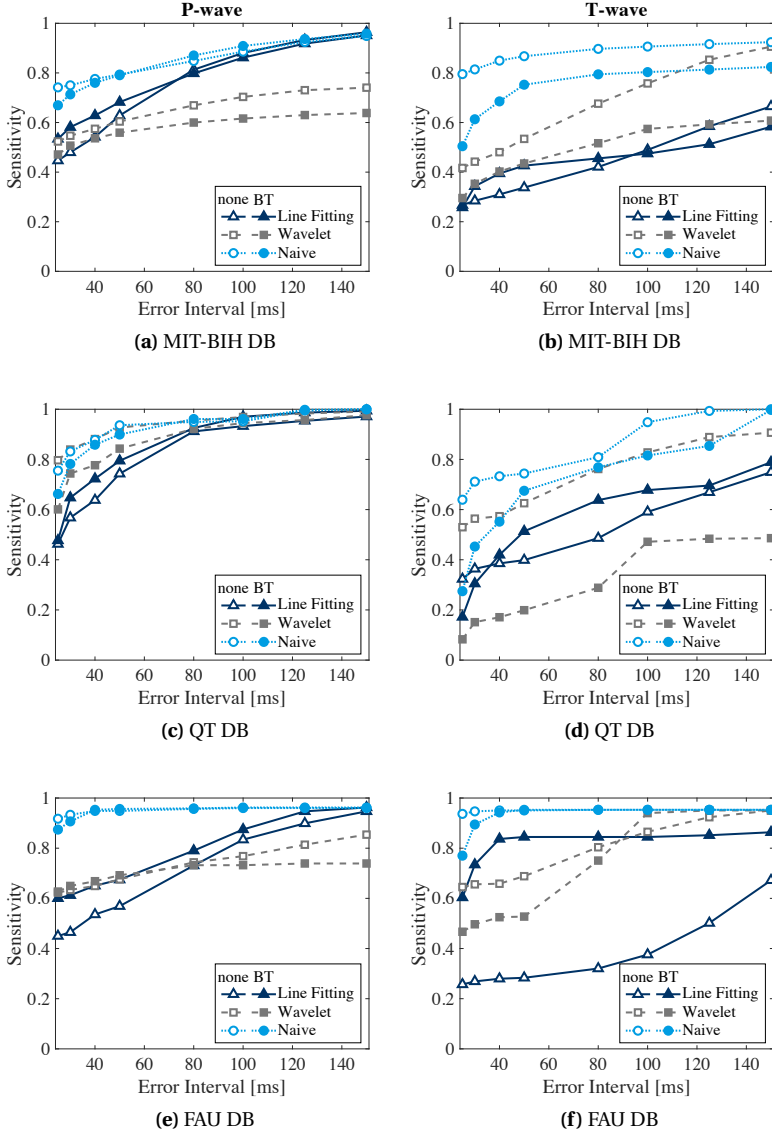


Figure 7.5: Sensitivities for the P-wave (left) and T-wave (right) applying the different wave event detection algorithms. The sensitivities are plotted over different error intervals for the different databases individually. BT: Butterworth filter; DB: database; MIT-BIH DB: MIT-BIH Arrhythmia DB; FAU DB: Short-Term FAU DB.

Table 7.5: Positive predictivity $+P$ in percent (%), mean absolute error MAE in milliseconds (ms), and mean standard deviation (SD) in (ms) for P-wave detection. Results are shown using an error interval of ± 40 ms. This error interval corresponds to half of the normal P-wave duration (Rangayyan, 2015).

		P-Wave								
		MIT-BIH DB			QT DB			FAU DB		
		$+P$	MAE	SD	$+P$	MAE	SD	$+P$	MAE	SD
Line Fitting	-	53.16	11.16	13.84	64.49	15.14	11.03	55.44	11.52	15.73
	BT	62.70	15.99	11.50	72.19	18.70	9.16	66.64	14.67	8.66
Wavelet	-	72.89	9.91	11.47	88.10	9.99	9.95	73.52	6.50*	7.21*
	BT	75.48	15.54	11.48	77.62	16.76	9.56	78.39	14.30*	7.11*
Naive	-	75.85	3.22	6.30	87.92	11.87	11.20	97.34	5.44	8.23
	BT	74.29	13.79	9.26	85.87	16.12	9.08	97.79	13.59	6.06

*One record excluded;

DB: database; MIT-BIH DB: MIT-BIH Arrhythmia DB; FAU DB: Short-Term FAU DB; BT: Butterworth filter

Table 7.6: Positive predictivity $+P$ in percent (%), mean absolute error MAE in milliseconds (ms), and mean standard deviation (SD) in (ms) for T-wave detection. Results are shown using an error interval of ± 80 ms for the T-wave evaluation. This error interval corresponds to half of the T-wave duration (Rangayyan, 2015).

		T-Wave								
		MIT-BIH DB			QT DB			FAU DB		
		$+P$	MAE	SD	$+P$	MAE	SD	$+P$	MAE	SD
Line Fitting	-	41.96	23.08	31.13	48.63	21.95	22.85	33.33	15.14	22.81
	BT	45.44	26.54	19.12	62.22	34.62	18.36	88.71	21.09	6.98
Wavelet	-	68.96	25.99	27.98	79.66	23.79	20.09	84.17	15.05	20.93
	BT	78.80	28.81*	22.32*	29.07	40.81***	20.95***	78.68	34.71	17.43
Naive	-	89.33	6.33	11.42	80.94	21.68*	11.15*	100.0	6.41	8.54
	BT	79.14	22.52	15.85	77.03	31.69**	11.39**	100.0	19.16	6.67

*One record excluded; **Two records excluded; ***Five records excluded;

DB: database; MIT-BIH DB: MIT-BIH Arrhythmia DB; FAU DB: Short-Term FAU DB; BT: Butterworth filter

Table 7.7: Comparison of best P- and T-wave detection algorithms to algorithms from literature. For each publication, the key technique, the error interval (int.), the used number of beats (N), the sensitivity *SE* in percent (%), the positive predictivity *+P* in percent (%), the mean error (mean) in milliseconds (ms) and the mean standard deviation (SD) in (ms) is given for the P- and T-wave, respectively.

Authors	Key techniques	Error int.	P-wave					T-wave				
			N	<i>SE</i>	<i>+P</i>	mean	SD	N	<i>SE</i>	<i>+P</i>	mean	SD
This thesis ¹	Wavelet (P-wave)	40 ms	337	88.1	88.1	10.0	10.0					
	Naive (T-wave)	80 ms						337	80.9	80.9	21.7	11.2
Bote et al. (2018)	High-accuracy mode	200 ms	nr	99.3	95.4	13.5	7.3	nr	99.99	99.2	8.4	14.3
Friganovic et al. (2018) ²	Combinations of algorithms	75 ms	3194	89.5	89.3	nr	nr	3542	87.9	89.1	nr	nr
Panigraphy and Sahu (2018)	Differential evolution	nr	3194	99.3	99.5	3.6	4.9	3542	99.2	99.5	1.6	5.2
Akhbari, Shamsollahi, Sayadi, et al. (2016) ³	Multi Hidden Markov Model	nr	nr	nr	nr	0.2	3.5	nr	nr	nr	−0.4	5.6

nr: not reported; na: not applicable

¹Only annotated beats of lead II; mean absolute error (*MAE*) instead of mean; ²P-wave: Lead 1 (A. Martínez et al., 2010); T-wave: Lead 2 (J. P. Martínez et al., 2004) and templates; ³2-fold cross-validation; ⁴P-wave: nonlinear, T-wave: linear; ⁵Lead 1; ⁶13-fold cross-validation, values read from boxplot

Table 7.7: Comparison of best P- and T-wave detection algorithms to algorithms from literature (continued).

Authors	Key techniques	Error int.	P-wave					T-wave				
			N	SE	+P	mean	SD	N	SE	+P	mean	SD
Akhbari, Shamsollahi, Jutten, et al. (2016) ⁴	Extended Kalman filter	nr	197	nr	nr	9.0	13.0	197	nr	nr	10.0	17.0
Elgendi et al. (2015)	Clinical knowledge	30 ms	na	na	na	na	na	3542	98.9	98.8	nr	nr
Lenis et al. (2016)	Phase free stationary Wavelet transform	nr	3194	100.0	88.2	5.8	9.1	na	na	na	na	na
Bayasi et al. (2014)	Adaptive delineation	nr	nr	nr	nr	7.3	10.3	nr	nr	nr	9.8	4.8
C. Lin et al. (2014)	Beat-to-beat marginalized particle filter	nr	3176	100.0	99.2	1.2	5.3	3403	100.0	99.2	−0.4	4.8
Mazomenos et al. (2013)	Hybrid feature extraction	nr	1620	nr	nr	5.0	9.5	1620	nr	nr	−15.3	29.3

nr: not reported; na: not applicable

¹Only annotated beats of lead II; mean absolute error (*MAE*) instead of mean; ²P-wave: Lead 1 (A. Martínez et al., 2010); T-wave: Lead 2 (J. P. Martínez et al., 2004) and templates; ³2-fold cross-validation; ⁴P-wave: nonlinear, T-wave: linear; ⁵Lead 1; ⁶13-fold cross-validation, values read from boxplot

Table 7.7: Comparison of best P- and T-wave detection algorithms to algorithms from literature (continued).

Authors	Key techniques	Error int.	P-wave					T-wave				
			N	SE	+P	mean	SD	N	SE	+P	mean	SD
Di Marco and Chiari (2011)	Wavelet, 32-bit integer, and linear algebra	150 ms	3194	98.2	91.0	−4.7	9.7	3542	99.7	97.8	−0.3	12.8
Rincón et al. (2011) ⁵	Single lead, Wavelet	320 ms	nr	98.8	92.7	13.4	14.8	nr	99.4	99.5	5.3	18.4
Dumont et al. (2010) ⁶	Evolutionary algorithm	nr	nr	nr	nr	0.8	9.2	nr	nr	nr	−0.8	11.5
A. Martínez et al. (2010)	Phasor transform	nr	nr	98.7	97.5	32.0	25.7	nr	99.2	99.0	5.3	5.8
J. P. Martínez et al. (2004)	Wavelet	nr	3194	98.9	91.0	3.6	13.2	3542	99.8	97.8	0.2	13.9
Vila et al. (2000)	Mathematical modeling	nr	na	na	na	na	na	3528	92.6	nr	−12.0	23.4

nr: not reported; na: not applicable

¹Only annotated beats of lead II; mean absolute error (*MAE*) instead of mean; ²P-wave: Lead 1 (A. Martínez et al., 2010); T-wave: Lead 2 (J. P. Martínez et al., 2004) and templates; ³2-fold cross-validation; ⁴P-wave: nonlinear, T-wave: linear; ⁵Lead 1; ⁶13-fold cross-validation, values read from boxplot

Table 7.7: Comparison of best P- and T-wave detection algorithms to algorithms from literature (continued).

Authors	Key techniques	Error int.	P-wave					T-wave				
			N	SE	+P	mean	SD	N	SE	+P	mean	SD
Jané et al. (1997)	Threshold based detector / Low-pass differentiator	nr	2626	97.0	99.5	−0.5	11.0	2932	98.8	99.8	23.3	28.3

nr: not reported; na: not applicable

¹Only annotated beats of lead II; mean absolute error (*MAE*) instead of mean; ²P-wave: Lead 1 (A. Martínez et al., 2010); T-wave: Lead 2 (J. P. Martínez et al., 2004) and templates; ³2-fold cross-validation; ⁴P-wave: nonlinear, T-wave: linear; ⁵Lead 1; ⁶13-fold cross-validation, values read from boxplot

For calculating *MAE* and *SD*, these records were excluded. The third algorithm, line fitting wave event detection, achieved better results in detecting the P-wave than the T-wave. Though, this algorithm could not keep up with the performances of the other two algorithms.

In this thesis, the line fitting wave event detection algorithm (Hu et al., 2014) was revised as 1) this approach could easily be adapted to additional wave events (Cosoli et al., 2015), and 2) this approach required the knowledge of ECG signal whereas was not dependent on thresholds from the incoming ECG signal. The Wavelet based ECG delineator of J. P. Martínez et al. (2004) was adapted due to their high citation rate and their state of the art research impact. Rincón et al. (2011) and Boichat et al. (2009) also presented modifications to the original algorithm. Their modifications, disregarding one thresholding and suggesting specific search window definitions, were not incorporated to keep closely to the original publication.

Usability

The naive wave event detection algorithm was suitable for instantaneous, real-time P- and T-peak detection. This algorithm was dependent on the search window definition and detected in every given search window a wave event. All algorithms detected several P-waves in record 04 of the Short-Term FAU database, whereas no P-wave was existent and hence not labeled. It has to be investigated if the performance of the naive wave event detection algorithm is appropriate in these cases. It might also be that the corresponding obtained features are so different that this might even be beneficial for the applied arrhythmia classification algorithm.

Comparison to literature

The comparison of different algorithms is hindered by the incomplete description of how the individual evaluation metrics were assessed. This involved the unspecified number of beats, the undisclosed usage of the lead, or the unknown error interval. Some approaches from literature did not mention the used lead or used the lead with the best performances after applying the algorithm (Boichat et al., 2009; Jané et al., 1997; J. P. Martínez et al., 2004). This led to improved results. In this thesis, the performances were assessed using solely lead II of Einthoven. Hence, several records had to be excluded – especially for the QT database – leaving just a small portion of the possible beats.

In event detection, all detected wave events need to be compared to all labeled wave events. Normally, the different algorithms do not detect the wave event at the exact same location as the label. Hence, it is common to use a certain error interval around the labeled wave event to differentiate between matched wave events and not detected wave events. If no error interval is stated, it is difficult to judge the validity of the performances. Additionally, in literature, error intervals were mentioned that ranged from 10 ms (Hu et al., 2014), 30 ms (Elgendi et al., 2015) to 150 ms (Di Marco & Chiari, 2011), 200 ms (Bote et al., 2018), or 320 ms (Boichat et al., 2009; Rincón et al., 2011). A normal P-wave duration lasts about 80 ms (Rangayyan, 2015). The T-wave duration is deeply correlated with the current heart rate, lasting about 160 ms. Having an error interval of 320 ms meant that the detected wave event could occur almost anywhere in one heartbeat.

For P-wave detection, the SD of 10 ms is comparable to 9.1 s (Lenis et al., 2016), 10.3 ms (Bayasi et al., 2014), 9.5 ms (Mazomenos et al., 2013), 9.7 ms (Di Marco & Chiari, 2011), and 9.2 ms (Dumont et al., 2010). Better results were obtained from Akhbari, Shamsollahi, Sayadi, et al. (2016), Bote et al. (2018), C. Lin et al. (2014), Panigraphy and Sahu (2018), whereas of these, only Bote et al. (2018) gave their error interval of 200 ms exceeding clearly the used error interval in this thesis of 40 ms.

For T-wave detection, the SD of five algorithms had smaller and hence, better values than this work (Akhbari, Shamsollahi, Sayadi, et al., 2016; Bayasi et al., 2014; C. Lin et al., 2014; A. Martínez et al., 2010; Panigraphy & Sahu, 2018). Of these, none reported the corresponding error interval. Whereas all approaches from literature had higher sensitivities SEs and positive predictivities $+Ps$ than the algorithms of this thesis (Table 7.7).

Wrapping up, for the QT database, the Wavelet based P-wave event detection algorithm is comparable to the suggested algorithms from literature whereas the naive T-wave event detection algorithms was not able to keep up.

Adaptations compared to original publication

Three adaptations were integrated to improve the original work (Leutheuser et al., 2016). First, additional data were included. The Short-Term FAU database was a self-recorded data acquired in clinical setting. This led to additional results and findings. Second, bugs in the code were corrected so that the best search window was chosen instead of using the smallest search window. Third, the results for mean absolute error MAE

and SD are given in milliseconds (ms) instead of samples to enable a fair comparison to the algorithms from literature.

Limitations

In this thesis, ECG fiducial points detection algorithms using the single lead II of Einthoven were investigated. Hence, the criteria of using the lead II was posed to the ECG data, resulting in using only selected records of the MIT-BIH Arrhythmia database and the QT database (Table 7.1).

In this chapter, there were two options as preprocessing step: no filter and a second-order Butterworth filter. The effect of other filter approaches could enhance the performance of the algorithms. The respective search windows for P- and T-wave detection were based on the current or subsequent RR-interval in relation to the current R-peak. A misdetected R-peak consequently affected the creation of a suitable search window. In this chapter, the detected R-peaks were not verified to be correct R-peaks within an appropriate range. This was done to determine the performance in a real environment where the correct R-peak annotations are not known beforehand. This might led to weak results.

Chapter highlights

- Three single-lead electrocardiogram (ECG) algorithms suitable for real-time instantaneous P- and T-wave detection were compared. These steps were used for all algorithms: R-peak detection, definition of search windows, preprocessing, and then one of the three wave event detection algorithms.
- Best algorithm was the fast naive wave event detection algorithm without any filtering. This algorithm is dependent on the search window definition and due to its nature, will detect a wave event in every given search window.
- Comparison to algorithms from literature is hindered by the incomplete description of these algorithms and their evaluation metrics.
- Two publicly available databases from the Physionet website, the MIT-BIH Arrhythmia database and the QT database, and one self-recorded database, the Short-Term FAU database were used to assess the performance of the different algorithms.

Part IV

IMU-based Activity Recognition

Chapter 8

Recognition of Single Activities

In this chapter, a hierarchical, multi-sensor based classification system is presented. The performance is compared to six state of the art algorithms from literature using the newly created benchmark dataset DaLiAc. The study setup behind DaLiAc is explained. This chapter addresses contributions (V) and (VI). Parts of this chapter have been published previously (Leutheuser, Schuldhaus, & Eskofier, 2013). In this publication, the first authorship is shared. The hierarchical classification system was designed together and mainly implemented by the other first author. The state of the art algorithms from literature were jointly selected and the implementations were divided. The author of this thesis wrote the main parts of the corresponding publication.

8.1 Methods

This part of the chapter introduces the study and hardware setup behind the DaLiAc dataset and presents the hierarchical classification algorithm. This algorithm is compared to six state of the art algorithms from literature.

8.1.1 Data

Hardware Equipment and Sensor Setup

The hardware equipment for the data collection consisted of four IMUs (Shimmer2R, Shimmer, Dublin, Ireland), one shirt, and one pair of shoes. Each IMU consisted of one triaxial accelerometer and one triaxial gyroscope. The range of the accelerometers was set to ± 6 g. The range of the gyroscopes was set to ± 500 deg/s for all sensors except the sensor placed at the ankle. The range of the gyroscope for the sensor at the ankle was

set to ± 2000 deg/s. A sampling rate of 204.8 Hz was chosen and the data were stored on SD card.

The four IMUs were placed at the chest, the right hip, the right wrist, and the left ankle (Figure 8.1). The first two positions were chosen to have IMU data closely attached to the body's center of gravity (Trost, McIver, & Pate, 2005). Placing IMUs on the chest, the trunk, or the hip satisfy this condition. Sensors on the hip were utilized in a variety of different studies (Bao & Intille, 2004; Ermes, Pärkkä, Mäntyjärvi, & Korhonen, 2008; S. Liu, Gao, John, et al., 2012; Preece, Goulermas, Kenney, & Howard, 2009; Ravi et al., 2005; Trost et al., 2005; Ward, Evenson, Vaughn, Rodgers, & Troiano, 2005). IMUs on the trunk or chest were also common in the literature (Najafi et al., 2003; Pärkkä et al., 2006). The last two positions were chosen to cover the extremities. IMUs on the wrist enable a correct classification of activities mainly dominated by the upper body (Bao & Intille, 2004; Ermes et al., 2008; S. Liu, Gao, John, et al., 2012; Pärkkä et al., 2006). IMUs positioned on the ankle are common in gait analysis (Barth et al., 2015; Mariani et al., 2010) and activity recognition studies (Bao & Intille, 2004; Preece, Goulermas, Kenney, & Howard, 2009). Schuldhaus et al. (2012) showed that sensors on the ankle support the recognition of ascending or descending stairs.

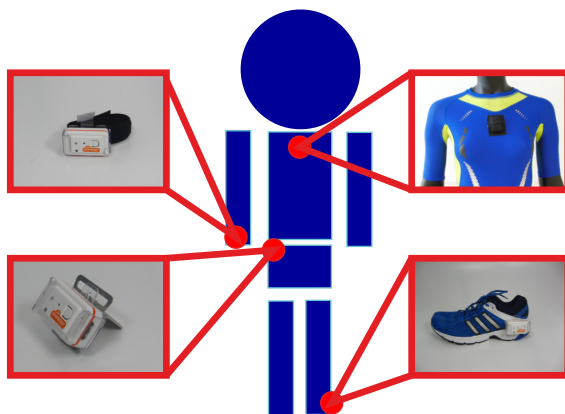


Figure 8.1: Placement of the inertial measurement units (IMUs) used for the DaLiAc dataset. Four IMUs were placed on the right wrist, the chest, the right hip, and the left ankle. Reprinted with permission from Leutheuser, Schuldhaus, and Eskofier (2013).

Table 8.1: Activities in the DaLiAc dataset. List of studied activities, durations, and intensities expressed in Metabolic Equivalent of Task (MET).

Activity	Duration [min]	Intensity [MET]
Sitting	1	1.3
Lying	1	1.0
Standing	1	1.3
Washing dishes	2	2.5
Vacuuming	1	3.3
Sweeping	1	3.3
Walking	na*	3.5
Ascending stairs	na**	5.0
Descending stairs	na**	3.5
Treadmill running	2	9.0
Bicycling on ergometer (50 watt)	2	3.5
Bicycling on ergometer (100 watt)	2	6.8
Rope jumping	na***	8.8

na: not applicable

*All subjects had to walk on the university campus from one building to another building.

**All subjects had to climb stairs to the third floor and then back again.

***All subjects had to perform five trials with at least five jumps each.

The types of shirt and shoes were the same for all participants. Four different shirt sizes were used (S, M, L, XL) to ensure a tight fit around the chest and to have similar measurement conditions. The chest width of each volunteer was measured to identify the best shirt size. Each subject wore the same shoe models (adidas Duramo, adidas AG, Herzogenaurach, Germany). The shoes were equipped with a special mounting on the outside of each ankle. This guaranteed the same sensor positions for each subject.

Experimental Protocol

A total of 13 activities were included in the study (Table 8.1). The activities were taken from the “Compendium of Physical Activity” (Ainsworth et al., 2011; Ainsworth et al., 1993; Ainsworth et al., 2000). In these compendiums, over 800 activities were assigned to four categories depending on the MET values: sedentary (1.0 - 1.5 METs), light-intensity (1.6 - 2.9 METs), moderate-intensity (3.0 - 5.9 METs), and vigorous-intensity (≥ 6 METs). Table 8.1 lists the durations and the intensities of the executed activities.



Figure 8.2: Screenshot of the data labeling application for Android. A researcher used this application to label the start and end of each activity. The researcher accompanied the subject for the whole data acquisition. Reprinted with permission from Leutheuser, Schuldhaus, and Eskofier (2013).

First, the static activities (sitting, lying, standing) as well as the household activities were performed. The subject was told to use the vacuum cleaner with the right hand. Afterwards, the subject walked on the university campus to another building. In this building, walking upstairs (until the third floor) and walking downstairs (back to the main floor) were recorded. The subject walked again on the university campus before performing the physically demanding exercises indoor. Of these, one exercise consisted of running on a treadmill (h/p/cosmos quasar, h/p/cosmos & medical gmbh, Nussdorf-Traunstein, Germany). The treadmill speed was set to 8.3 km/h. Another exercise required the subjects to bicycle on an ergometer (sanabike 250 F, MESA Medizintechnik GmbH, Benediktbeuern, Germany) with two different resistance levels (50 watt and 100 watt). The treadmill speed and the resistance level were chosen to obtain activities with different MET values (Table 8.1). The subjects were told to keep the revolutions per minute constant at 70 during the two different resistance levels. Thus, the differences of the two levels were not due to different revolutions per minute. The last physically demanding exercise was rope jumping. For this activity, the subject had to perform five trials with at least five jumps each.

A researcher accompanied the subjects during the whole data acquisition and labeled the start and end of each activity using a mobile phone (Samsung Galaxy S2, Samsung, Seoul, South Korea). An Android-based labeling app (Figure 8.2) was created to mark the start and the end time of single activities concurrently to the data collection.

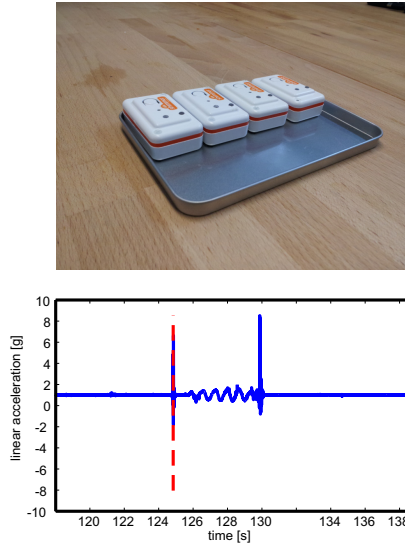


Figure 8.3: Inertial measurement unit (IMU) synchronization. A plate with four IMUs was used for synchronization (top). The pattern of dropping and up- and down-movement resulted in a sinusoidal synchronization signal (bottom). The dashed red line depicts the synchronization start point. Reprinted with permission from Leutheuser, Schuldhaus, and Eskofier (2013).

Before the start of the first activity and after the end of the last activity, a synchronization movement was performed with the four IMUs. For this synchronization movement, all four IMUs were placed on a plate (Figure 8.3). This plate was then dropped twice and, in between, the plate was moved up and down. The synchronization movement was repeated at the end of the measurement to check if problems with the sensors occurred during data acquisition.

Subjects

23 healthy subjects (10 female and 13 male, age 27 ± 7 years, Body Mass Index (BMI) $24.0 \text{ kg/m}^2 \pm 3.5 \text{ kg/m}^2$, mean $\pm$ SD) were recruited for the study. Of these 23 subjects, 21 were right handed and two were left handed.

The Research Ethics Committee of the Friedrich-Alexander University Erlangen-Nürnberg (FAU) confirmed that there is no need to obtain the approval of the local Ethics Committee. Ethics approval was deemed

unnecessary because only healthy subjects participated in the study. The subjects were in good physical state and did not suffer from a disease. All subjects gave written informed consent about their participation. All volunteers filled in the Physical Activity Readiness Questionnaire (PAR-Q) (Adams, 1999)). The PAR-Q provides a self-administered screening before performing physical activity. The aim of the PAR-Q is to identify those subjects who should consult a physician before performing physical activity. In the study, only subjects who passed the PAR-Q were included. The subject information including the PAR-Q are given in appendix C.

Preprocessing

Four datasets had to be excluded from further processing. Three of them were excluded because of problems during gyroscope initialization. The remaining one was excluded as no data of the ankle sensor were acquired. In total, data of 19 subjects were included in the DaLiAc dataset.

The four Shimmer sensor nodes were synchronized offline. For this, the first up-down movement of the sensor signal was manually selected in the linear acceleration of the vertical direction in all sensor nodes (Figure 8.3, bottom, vertical line). This point constituted the common start point of all sensors. The labeling was done automatically due to the saved start and end times of the Android app (Figure 8.2). For each labeled activity, two seconds at the beginning and at the end were cut, in order to eliminate measuring errors during labeling.

8.1.2 Hierarchical, Multi-Sensor Based Classification System

The general idea was to set up a hierarchy of classification systems, where each system solved a different classification problem. For this, the individual activities were divided in separate classification problems.

Figure 8.4 shows an overview of the proposed classification system. The BASE system forms the first hierarchy – the basis – of the classification system. It differentiates in the four subsystems REST, HOUSE, WALK, and BICYCLE and two single activities. Each of the subsystems consists of an activity group. The single activities are rope jumping and washing dishes that did not fit in any of the groups. Each subsystem incorporates one classifier, taking into account that different classifiers could be chosen for single subsystems. Such a hierarchical system was chosen to enable the easy adding of new activities without retraining all classifiers. The system is therefore flexible in its application to different activities.

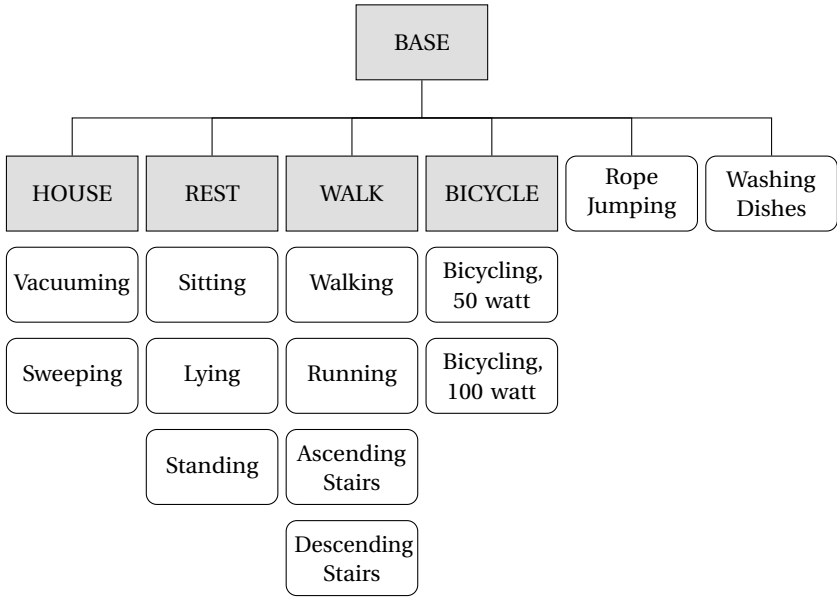


Figure 8.4: Hierarchical classification system. Filled rectangles indicate single classification systems. Unfilled rectangles indicate single activities.

Preprocessing

Data processing was performed in sliding windows with 50% overlap (Bao & Intille, 2004; Preece, Goulermas, Kenney, & Howard, 2009; Ravi et al., 2005). The width of the window was set to 5 s. This is comparable to the work of Ravi et al. (2005) and Bao and Intille (2004), which used 5.12 s and 6.7 s, respectively.

Feature Extraction

A generic feature set was defined for the classification systems BASE, HOUSE, WALK, and BICYCLE. This set was computed for each sliding window. The generic feature set consisted of six features that were computed for every sensor axis and one feature that was computed for either all accelerometer axes or all gyroscope axes of each IMU. The six features for every sensor axis included four time domain and two frequency domain features.

The four time domain features were:

- Minimum amplitude,
- Maximum amplitude,
- Mean amplitude, and
- Variance of amplitude.

The minimum and the maximum amplitude extracted range information of the amplitude. The mean and variance of the amplitude gave important knowledge about the statistics of the signal.

The two frequency domain features were:

- Spectral centroid and
- Bandwidth.

Spectral centroid and bandwidth delivered important information about the frequency distribution of the activities (Agostini, Longari, & Pollastri, 2003).

The single feature that was computed for each sensor type (accelerometer or gyroscope) of one IMU was the energy. The energy was calculated in three steps. First, the sum of the squared values for each axis was calculated. Second, the three sums were added together and divided by three. Third, the sum was divided by the number of samples. The energy gave important information about the activity level of a person. In total, this resulted in 152 features for each sliding window.

A different feature set was defined for the classification system REST. For this system, the gravitational component of the acceleration signal was extracted in applying a third-order elliptic low pass filter with an infinite impulse response and a cut-off frequency at 0.25 Hz (Karantonis et al., 2006). This means that all three gravitational acceleration components of all four sensors were used as features. This was done because mainly the orientation of the body is important for the discrimination of the activities sitting, lying, and standing. In total, this resulted in 12 features for each sliding window.

Classification

No single classifier is suitable for all classification tasks (Duda et al., 2000). Hence, the performance of different classification systems was assessed (Duda et al., 2000; Theodoridis & Koutroumbas, 2008): AdaBoost, CART,

kNN, and SVM with a radial basis function kernel. Section 3.3.2 gives more information about the principles of the different classifiers.

When using AdaBoost, decision stumps were chosen as weak learners. A decision stump classifier forms a one-level decision tree. As termination criterion, the training stopped when reaching 100 iterations. In the case of CART, a binary decision tree was grown. The Gini impurity *Gini* (equation (3.15)) was used for obtaining the corresponding decision at a node. Nodes must be split, when having 10 or more feature vectors with different labels. Feature vectors having just single missing values were not eliminated completely. Instead, just the valid values were used for this feature vector. For the kNN classifier, k was set to five. The k nearest neighbors were determined with the Euclidean distance measure. The cost parameter of the SVM classifier was set to one and the gamma parameter to one divided by the number of features.

8.1.3 Performance Measures and Evaluation

For performance assessment, the mean class dependent classification rate and the overall mean classification rate were computed with a leave-one-subject-out cross-validation for all five classification systems (Figure 8.4). In each leave-one-subject-out fold, all epochs of one subject were removed from the training set. In order to evaluate the whole hierarchical classification system, the classifier with the best performance was chosen for the subsystems BASE, HOUSE, REST, WALK, and BICYCLE.

8.1.4 Comparison to Algorithms in Literature

The hierarchical classification model was compared to six state of the art algorithms in literature (Bao & Intille, 2004; Karantonis et al., 2006; S. Liu, Gao, John, et al., 2012; Pärkkä et al., 2006; Preece, Goulermas, Kenney, & Howard, 2009; Ravi et al., 2005). These were chosen due to their citation rate and hence, their state of the art research impact. All six algorithms had intersections with the described study setup regarding the used IMUs (accelerometers and gyroscopes), their placement, and the performed activities (Table 2.5).

Table 8.2 gives an overview of the different sensor placements in literature compared to DaLiAc. Only identical positions and sensor data were used resulting in using at maximum four positions of IMU data (Figure 8.1) and other signals like the heart rate were disregarded. For example, the IMU placements and axes alignment suggested by Bao and Intille (2004) were modified. Bao and Intille (2004) used five biaxial accelerometers and

Table 8.2: Sensors placement in the six state of the art algorithms in the literature that were implemented and compared in the present study. The different sensor placement are given compared to the DaLiAc dataset.

Authors (Year)	Original sensor placement	Substitute in DaLiAc
Bao and Intille (2004)	right hip right wrist left arm right ankle left thigh	right hip right wrist right ankle
Ravi et al. (2005)	pelvic region	right hip
Karantonis et al. (2006)	right hip	right hip
Pärkkä et al. (2006)	chest dominant wrist <i>additional signals like altitude, ECG, temperature, and heart rate</i>	chest right wrist
Preece, Goulermas, Kenney, and Howard (2009)	waist thigh ankle	right hip right ankle
S. Liu, Gao, John, et al. (2012)	dominant hip dominant wrist ventilation sensor	right hip right wrist

placed them on the right hip, the right wrist, the left arm, the right ankle, and the left thigh. In the DaLiAc dataset, data of triaxial accelerometers were acquired. For the comparison, data of only two axes and only consistent placement – the right hip, the right wrist, and the left ankle – were used.

All six state of the art algorithms used a lower sampling frequency than the suggested sampling frequency of 204.8 Hz (Table 2.5). For an appropriate comparison, the DaLiAc dataset was down-sampled to the sampling frequencies used in the corresponding algorithms. The down-sampling was performed by a linear interpolation method. Furthermore, the window size and overlap were set according to the description of each publication (Table 2.5). The features and classifiers were implemented as described in the different approaches. The hierarchical classification

method was only compared to the suggested final feature set and the classifier with whom the best classification results were obtained. In order to use the same evaluation process in each approach, a leave-one-subject-out cross-validation was performed in each of the six state of the art approaches, except in the algorithm of Karantonis et al. (2006). Karantonis et al. (2006) did not apply a training step and used predefined and fixed thresholds.

Each algorithm in literature used different activities (Tables 2.5 and 2.6). All six algorithms were evaluated on all recorded activities in the DaLiAc dataset, except for the algorithm of Karantonis et al. (2006). The threshold-based classifier used in Karantonis et al. (2006) was optimized for fall detection and therefore not applicable to all acquired activities. This means, for the other five algorithms, that activities were investigated that were not considered in the original works.

8.2 Results

Table 8.3 gives the overall mean classification rates after leave-one-subject-out cross-validation for the individual subsystems. SVM scored best for the BASE, the REST, and the BICYCLE system. AdaBoost scored best for the HOUSE system. kNN scored best for the WALK system. Table 8.4 shows the confusion matrix of the suggested hierarchical classification system. Table 8.5 gives the mean class dependent classification rates and the overall mean classification rates of the suggested hierarchical classification system in comparison to the state of the art algorithms from literature (Bao & Intille, 2004; Karantonis et al., 2006; S. Liu, Gao, John, et al., 2012; Pärkkä et al., 2006; Preece, Goulermas, Kenney, & Howard, 2009; Ravi et al., 2005). The algorithm of Karantonis et al. (2006) was not applied to all activities, as activity optimized features were used. Table 8.6 shows the usage of the benchmark dataset DaLiAc in literature.

8.3 Discussion

In this chapter, the benchmark dataset DaLiAc was introduced covering 13 activities of 19 subjects. Using DaLiAc, a hierarchical classification system was developed being capable of distinguishing these 13 activities with an overall mean classification rate of 89.6% (Table 8.5).

Table 8.3: Mean classification rates in percent (%) of the five subsystems (BASE, REST, HOUSE, WALK, BICYCLE). For each subsystem, the performance of all four classifiers was assessed: Adaptive Boosting (AdaBoost), Classification and Regression Tree (CART), K-Nearest Neighbor (kNN), and Support Vector Machine (SVM). Best results are printed bold.

	AdaBoost	CART	kNN	SVM
BASE	64.8	96.1	97.7	97.9
HOUSE	89.9	84.0	86.5	85.5
REST	95.1	92.7	96.1	97.4
WALK	94.7	93.3	97.7	94.3
BICYCLE	60.8	53.7	49.4	61.6

General discussion

Different classifiers achieved the best overall mean classification rates regarding the five systems BASE, HOUSE, REST, WALK, and BICYCLE (Table 8.3). SVM was chosen three out of five times as best classifier. AdaBoost and kNN were selected as best classifier for the remaining two systems. The selection of different classifiers supports the hypothesis that no single classifier is suitable for all classification tasks (Duda et al., 2000). The application of different classifiers was beneficial for distinguishing different activities or activity groups. SVM is for example known as a classifier with a good generalization performance (Theodoridis & Koutroumbas, 2008).

The high mean classification rates for the different systems showed the applicability of each system to distinguish between activity groups and single activities (Table 8.3). The performance and flexibility of the hierarchical classification system is enhanced with the usage of activity groups. For each activity group, a different classifier could be used and three different ones were chosen.

The hierarchical classification showed that the classification of activities can benefit from sensor data fusion of accelerometer and gyroscope (Table 8.5). It is assumed that especially the gyroscope improves the classification of activities which include rotational movements like washing dishes, ascending stairs, or rope jumping. It is assumed that the gyroscope of the ankle provides useful information about the inclination (Schuldhuis et al., 2012). Most approaches found in literature only used accelerometers. One approach (Salarian et al., 2007) acquired data with gyroscopes.

Table 8.4: Confusion matrix of the proposed hierarchical classification system. Each entry represents the number of classified epochs. The confusion matrices of each leave-one-subject-out trial were summed up.

		Predicted class												
		SI	LY	ST	WD	VC	SW	WK	AS	DS	RU	BC 50	BC 100	RJ
Actual class	SI	383	0	22	14	1	2	0	0	2	0	7	0	0
	LY	0	435	0	0	0	0	0	0	0	0	0	0	0
	ST	0	0	386	27	9	0	0	0	0	0	8	0	0
	WD	0	0	10	896	5	1	0	0	0	0	1	0	0
	VC	0	0	0	0	369	55	0	0	0	0	6	2	0
	SW	0	0	1	6	43	643	4	7	5	0	4	2	0
	WK	0	0	0	0	4	4	1995	5	7	0	0	0	0
	AS	0	0	0	0	0	3	10	279	0	0	0	0	0
	DS	0	0	0	0	0	1	11	0	237	0	0	0	0
	RU	0	0	0	0	0	0	0	0	0	886	0	0	0
	BC50	0	0	0	0	4	22	0	1	0	0	620	250	0
	BC100	0	0	0	1	0	6	0	0	0	0	410	480	0
	RJ	0	0	0	0	0	0	0	0	0	0	0	0	236

SI: Sitting; LY: Lying; ST: Standing; WD: Washing dishes; VC: Vacuuming; SW: Sweeping; WK: Walking; AS: Ascending stairs; DS: Descending stairs; RU: Treadmill running; BC50: Bicycling on ergometer (50 watt); BC100: Bicycling on ergometer (100 watt); RJ: Rope jumping

Table 8.5: Mean class dependent classification rates in percent (%) for all 13 activities and overall mean classification rates of proposed system and state of the art systems (Bao & Intille, 2004; Karantonis et al., 2006; S. Liu, Gao, John, et al., 2012; Pärkkä et al., 2006; Preece, Goulermas, Kenney, & Howard, 2009; Ravi et al., 2005). Best results are printed bold.

	Bao and Intille (2004)	Ravi et al. (2005)	Karantonis et al. (2006)	Pärkkä et al. (2006)	Preece et al. (2009)	Liu et al. (2012)	Hierarchi- cal classi- fication
Sitting	83.1	83.7	81.8	65.5	67.7	38.1	88.9
Lying	94.5	87.6	94.7	98.2	100.0	57.4	100.0
Standing	80.7	60.2	64.0	65.6	48.6	44.4	89.8
Washing dishes	88.9	79.0	—*	75.4	56.0	89.6	98.1
Vacuuming	66.9	17.7	—*	36.0	39.2	42.9	85.4
Sweeping	81.2	57.8	—*	60.7	62.6	54.3	89.9
Walking	96.2	93.0	98.7	74.3	97.6	88.1	99.0
Ascending stairs	79.5	18.6	—*	28.5	70.6	29.3	95.5
Descending stairs	73.1	16.4	—*	44.2	60.2	35.3	95.2
Running	100.0	98.3	—*	92.7	97.2	94.4	100.0
Bicycling, 50 watt	48.1	50.4	—*	40.2	64.3	47.8	69.1
Bicycling, 100 watt	48.5	11.7	—*	46.2	41.7	48.3	53.5
Rope jumping	99.4	93.4	—*	76.6	86.7	33.3	100.0
Mean	80.0	59.1	84.8	61.8	68.7	54.1	89.6

*The algorithm of Karantonis et al. (2006) was not applied to all activities, as activity optimized features were used.

Table 8.6: Usage of DaLiAc in literature. All publications are listed that either used DaLiAc and cited the corresponding publication about the benchmark dataset DaLiAc and the hierarchical classification system or just cited the publication by Leutheuser, Schuldhaus, and Eskofier (2013) (Google Scholar, 2019; Scopus, 2019). N gives the number of publications.

Usage of DaLiAc and citation of corresponding publication (N = 20)

Ali et al. (2017), Awais, Palmerini, and Chiari (2016), Casale et al. (2015), Chen et al. (2016), Hur et al. (2018), Jurca et al. (2018), Khandelwal and Wickström (2018), Khandelwal (2018), Khojasteh et al. (2018), Moldovan et al. (2019), Moldovan et al. (2018), Munoz-Organero and Lotfi (2016), Nazábal (2017), Nazábal et al. (2016), Nguyen et al. (2018), Nweke et al. (2018), Raza et al. (2019), Schuldhaus, Leutheuser, and Eskofier (2014), Zdravevski (2017), Zdravevski, Lameski, et al. (2017)

Citation of corresponding publication (N = 75)

Abdul Rahman et al. (2017), Anand et al. (2018), Awais (2016), Awais, Chiari, et al. (2019), Awais et al. (2015), Awais, Palmerini, Bourke, et al. (2016), Awais et al. (2019), Baloch et al. (2019), Billiet (2018), Billiet et al. (2018), Blank et al. (2015), de la Cal (2018), Casilari et al. (2017a,2017b), Casilari-Pérez and García-Lagos (2019), Cerbulescu and Cerbulescu (2017), Chatzaki et al. (2017), Chen and Wang (2017), Clark (2017), Clark et al. (2017), Clark et al. (2018), Cleland et al. (2018), Dauriac (2018), Dimitrievski et al. (2016), Eskofier et al. (2014), Full et al. (2015), Fuster-Garcia et al. (2015), Ghobadi et al. (2015), Ginelli et al. (2018), Guo and Wang (2018), Heldberg et al. (2015), Hessen and Tessem (2016), Hunt (2017), Huppert et al. (2016), Ćinanić et al. (2018), Jensen (2016), Jensen, Leutheuser, et al. (2015), Kautz (2017), Kautz, Eskofier, and Pasluosta (2017), Kautz, Groh, et al. (2017), Kautz et al. (2015), Koskimäki and Siirtola (2014), Lasek and Gagolewski (2015), Lemmens et al. (2015), Leutheuser, Doelfel, et al. (2014), J. Liu et al. (2017), Malliotakis (2017), Mangiarotti et al. (2019), Martindale et al. (2016), Martindale et al. (2019), McCarthy et al. (2015), McCarthy et al. (2016), Mehmood et al. (2016), Merchán-Baeza et al. (2014,2015), Merry (2017), Micucci et al. (2017), Moldovan, Anghel, et al. (2019), Palumbo (2016), Palumbo et al. (2016), Pannurat et al. (2017), Pucci (2014), Rahman (2017), Schuldhaus, Dorn, et al. (2014), Taborri et al. (2014), Tobola, Espig, et al. (2015), Tobola, Streit, et al. (2015), Urwyler et al. (2015), Vavoulas et al. (2016), Wang and Chen (2017), White (2018), Wundersitz (2015), Wundersitz, Gastin, et al. (2015), Wundersitz, Josman, et al. (2015), Zdravevski, Stojkoska, et al. (2017)

Benchmark dataset DaLiAc

In this chapter, the dataset DaLiAc was described and acquired. This dataset is publicly available to enable and facilitate the comparison of state

of the art activity recognition algorithms (<http://www.activitynet.org/>). The recorded activities correspond to the relevant activities that other research groups have used in their studies (Table 2.6). The corresponding publication (Leutheuser, Schuldhaus, & Eskofier, 2013) was cited 95 times considering the Scopus (Scopus, 2019) and the Google Scholar (Google Scholar, 2019) document search. In 20 publications of them, other research groups used the DaLiAc dataset to assess the performance of their algorithms.

Advantages of hierarchical classification

The hierarchical structure has several advantages:

- Different classifiers can be used for the classification of different activity groups / activities.
- Additional activities are easily integrated without retraining the complete system.
- If further classification of the activities in the activity groups is not necessary, the HOUSE, REST, WALK, and BICYCLE system can easily be neglected to improve the performance of the classification system. Also, activities in one system could be grouped. The activities walking and running as well as the activities ascending and descending stairs could be grouped. This would result in two systems below the WALK system.
- Different window sizes can be chosen for different system. This might increase the classification rate. Additionally, depending on the use case for the final classification system, resources can be saved.

Comparison to implemented algorithms from literature

The hierarchical classification system was compared to six state of the art algorithms in literature (Bao & Intille, 2004; Karantonis et al., 2006; S. Liu, Gao, John, et al., 2012; Pärkkä et al., 2006; Preece, Goulermas, Kenney, & Howard, 2009; Ravi et al., 2005). They reached an overall mean classification rate between 54.1% (S. Liu, Gao, John, et al., 2012) and 84.8% (Karantonis et al., 2006) (Table 8.5). Best results from literature when considering the complete DaLiAc dataset, were obtained with the algorithm by Bao and Intille (2004) with 80.0%. The reasons might be the following: The study design had three sensor positions in common. In

other approaches, at maximum two sensor positions were identical. The approach by Bao and Intille (2004) was applied to 20 activities under both controlled conditions in the laboratory and semi-naturalistic conditions outside the laboratory. Thus, the approach by Bao and Intille (2004) seemed to be applicable for a large set of daily life activities, especially under semi-naturalistic conditions. The differences in the classification accuracies compared to the original publication can be explained with the different study setups like the sensors, their placement, and the activities and the suggested classification systems by itself.

Differences relating to study setup

In the DaLiAc dataset, data was acquired with four IMUs. Each IMU consisted of a triaxial accelerometer and a triaxial gyroscope. As different sensors were used, the range of sensors could be different. It might be that due to saturation worse results were obtained.

The four sensor placements of the IMUs were compared to one sensor position (Karantonis et al., 2006; Ravi et al., 2005), two sensor positions (Pärkkä et al., 2006; Preece, Goulermas, Kenney, & Howard, 2009), and three sensor positions (Bao & Intille, 2004; S. Liu, Gao, John, et al., 2012). Additionally, some sensor positions were different or neglected. Bao and Intille (2004) used additional sensors on the thigh and arm. Preece, Goulermas, Kenney, and Howard (2009) used an additional sensor on the thigh. Pärkkä et al. (2006) and S. Liu, Gao, John, et al. (2012) incorporated additional sensors apart from IMUs like a temperature or a ventilation sensor. Even when using the same sensor positions (e.g. right hip), it might be that the placement in this study was different to the original publication relating to different results. Sensor alignments were adopted where possible.

DaLiAc comprised the most relevant activities that other research groups have used (Tables 8.1 and 2.6). The acquired activities in DaLiAc were still different from the activities used in the literature. Additional activities were acquired in literature, whereas DaLiAc comprised activities not existent in all approaches. Activities not considered in the original study setup were classified with poor classification rates. With the algorithm from S. Liu, Gao, John, et al. (2012), ascending and descending stairs were classified with only about 30%. Both activities were not considered in the original study setup. Also, S. Liu, Gao, John, et al. (2012) merged the activities sitting and standing to one class.

Differences relating to classification systems

The algorithms from literature were reimplemented without any adaptations or modifications. From each author group, solely the best algorithm was considered. The focus was not to improve the suggested results from literature, rather to compare the hierarchical classification system in the best possible way with algorithms from literature. Better results might have been obtained when the algorithms were not only reimplemented and instead their broad idea was implemented.

The suggested epoch sizes from literature were not always compatible with the duration of the recorded activities in the DaLiAc dataset. S. Liu, Gao, John, et al. (2012) suggested an epoch size of 30 s. In DaLiAc, several activities were performed for 1 min. To eliminate labeling errors, each labeled activity was shortened with 2 s at the beginning and end. This resulted in only one epoch for these activities considering the algorithm by S. Liu, Gao, John, et al. (2012). The epoch duration of 5 s was also chosen to grasp high-intensity, fast-changing activities, and to provide a recognition in quasi real-time.

The high classification rates demonstrate the usability of the generic features for four of the systems. Also the reduced feature set for the REST system consisting of only 12 features in each window was able to classify the static activities. The feature set suggested by Karantonis et al. (2006) was optimized for walking. They used the postural orientation for walking (tilt angle feature). Hence, due to this optimization, their algorithm was only applied to a subset of activities. S. Liu, Gao, John, et al. (2012) incorporated a two-step features selection. They performed a statistical analysis followed by the minimal-redundancy-maximal-relevance heuristic (Peng, Long, & Ding, 2005). This resulted in a specialized feature set which might not be applicable to all activities. The best feature set determined for a specific dataset comprising a subset of activities might not be best for the DaLiAc dataset.

Relevance of sensor placement

The performance is further dependent on the sensor positions and the performed activities. Sensor positions near the body's center of mass (hip and chest) are used to cover a wide range of basic activities such as sitting, standing, lying, and walking. The mean class dependent classification rates of sitting and standing were smaller compared to lying. This trend coincides with the results for the other approaches (Table 8.5). Ravi et al.

(2005) used one hip sensor with a sampling rate of 50 Hz. Using the suggested algorithms by Ravi et al. (2005), activities involving upper and lower extremity movements were classified with low classification rates. Pärkkä et al. (2006) and S. Liu, Gao, John, et al. (2012) used one wrist sensor and either chest or hip sensor. Using this sensor combination together with the suggested algorithms, activities including lower extremity motions were classified with low classification rates. Activities include ascending or descending stairs and bicycling.

In this chapter, IMUs on lower and upper extremities (wrist and ankle) were used to distinguish between more complex activities like ascending stairs, descending stairs, vacuuming, and sweeping. Nevertheless, vacuuming was often misclassified as sweeping and vice versa (Table 8.4). In this case, an additional sensor on the other wrist might incorporate additional information into the classification system. Descending and ascending stairs were often misclassified as walking (Table 8.4). In order to reduce the misclassification, specialized gait features might improve the results for the WALK system. Nevertheless, given the flexibility of the proposed classification system, the incorporation of these ideas is straightforward.

Limitations

The overall mean classification rates of all classifiers in the case of the BICYCLE system were rather low compared to the classification rates of the other four subsystems (Table 8.3). For bicycling activities, the revolutions per minute were kept constant while only changing the resistance level. The high overall mean classification rate of the BASE system (Table 8.5) showed that it is possible to classify the merged activity group of the two resistance levels of bicycling. The rather low mean class dependent classification rates of the two resistance levels of bicycling demonstrate the challenge to distinguish between the single resistance levels: the low resistance level was misclassified as the high resistance level and vice versa (Table 8.4). Results might be improved by incorporating different features or an additional type of sensors to acquire the physical exertion.

In this chapter, 152 features (except for the REST system) for each sliding window were used by each classification system. A high number of features leads to high computational complexity in real-time applications or in embedded systems. Hence, an automatic reduction of this feature set using appropriate feature selection procedures like sequential forward selection should be applied.

In DaLiAc, only data under controlled conditions were collected. Data under more realistic conditions would be desirable, including transitions between single activities.

Chapter highlights

- In the benchmark dataset DaLiAc, data of 13 different activities of 19 subjects were recorded. This publicly available dataset comprises the relevant activities that other researchers have used. This dataset is suited for comparing state of the art activity recognition algorithms.
- In the hierarchical classification system, activities were grouped and ordered in a hierarchy. This provided advantages like the easy integration of new activities without retraining the complete system, and the broad range of options for each single classification system: different features, window sizes, or classifiers.
- The hierarchical classification system outperformed six state of the art activity recognition algorithms from literature considering the complete DaLiAc dataset. The overall mean classification rate of the hierarchical classification system was 89.6%.

Part V

Perspective

Chapter 9

Summary

In this thesis, three wearable computing applications in eHealth were considered. The first application referred to mobile breathing analysis. In the second application, mobile ECG analysis algorithms were investigated. The third application dealt with activity recognition using IMU data. For each application, a short summary is given in the next sections.

9.1 Mobile Breathing Analysis

The mobile breathing analysis part was divided in two sections. This part had the aim to show that with RIP devices accurate tidal volumes can be acquired even without simultaneous measurement of FMs. In the two sections, different kinds of adjustment algorithms were investigated. Both sections used the same data acquired with RIP and FM. 193 subjects performed an incremental running test on a treadmill. 140 of them performed the study protocol twice.

In chapter 4 addressing contribution (I), individual adjustment parameters were calculated using a previously recorded phase. The prerecorded phase consisted of the standing still phase of each subject. In this phase, data with both measurement devices – RIP and FM – were acquired to calculate appropriate individual VMCs. Afterwards, the VMCs were applied to conditions of treadmill running and recovery. The best adjustment algorithm was an adapted algorithm from literature: the adapted Sackner algorithm or adapted QDC method (Sackner et al., 1989). This algorithm outperformed SVR with four different kernels and LSQ with and without offset. In the Sackner algorithm (Sackner et al., 1989), first, the k -ratio K is calculated to get the relative contribution of both bands. Second, K is set in relation to the true tidal volume – acquired with FM – to obtain

the scaling factor M . In the adapted Sackner algorithm, only the standing still phase was used to calculate the k-ratio K and the scaling factor M . As data of three different datasets were available, it was investigated if the same adjustment parameters could be applied to data acquired one week later. With this, it was investigated if the VMCs could be stored for following data acquisitions. Comparable results were obtained showing the usability of the suggested adjustment parameters.

In chapter 5 addressing contribution (II), the aim was to create generalizable and reference-free adjustment algorithms. Generalizable and reference-free meant that no prerecording of any data is needed and that the adjustment works solely with RIP data. The adjustment algorithms included anthropometric data of the subject, time characteristics, and ratios of RC to AB, whereas the latter two are acquired and instantaneously incorporated in the adjustment. All characteristics have to be known for the first measurement and are stored for all further measurements. When considering both bands without any prerecording phase – the best option –, the subjects have to provide their age, height, weight, sex, and AB circumference. The adjustment worked better for treadmill running conditions than for recovery conditions. The results from the different datasets suggested the (re-)usability of the adjustment for unseen subjects and the robustness for future data. The two-degree of freedom system with both bands (RC and AB) was compared to the two one-degree of freedom system (either RC or AB). Adjusting with both bands achieved slightly better results compared to the consideration of only one band. The RC band outperformed the AB band.

9.2 Mobile ECG Analysis

The mobile ECG analysis part was divided in two sections. This part had the aim to provide real-time ECG analysis algorithms. In the first section, the signal originating from the time instances of successive heartbeats was used to perform arrhythmia classification – to differentiate between pathologic and normal beats. In the second section, algorithms suitable for real-time P- and T-wave detection were compared. The algorithms used single lead ECG data. The single lead corresponded to lead II of Einthoven. This lead can be measured in wearable and real-time settings.

In chapter 6 addressing contribution (III), arrhythmia classification was performed using the RR intervals. One RR interval is the difference of the time instances of two successive heartbeats. With N (10, 15, or 20) consecutive RR intervals, the following features were extracted: the

amplitude A , the phase φ , the sinusoidal regression feature – the distance between the $(N + 1)$ th RR interval and the sine curve –, and the RR intervals by themselves. For feature extraction, the phenomenon of RSA or HRV was used to fit a sine curve to the RR interval segments. The fitting required the determination of the underlying frequency. In this chapter, the performance of four variations was assessed. The features were then fed into a Naive Bayes classifier to differentiate between pathologic and normal beats. The best feature set consisted of using all of these features. Best real-time performances were obtained with the smallest segment length of 10 RR intervals.

In chapter 7 addressing contribution (IV), three single-lead ECG algorithms suitable for real-time P- and T-wave detection were compared. The three investigated algorithms were the line fitting wave event detection (revised (Hu et al., 2014)), the Wavelet based wave event detection (adapted (J. P. Martínez et al., 2004)), and the fast naive wave event detection. The following steps were applied to the single-lead (lead II of Einthoven) ECG data. First, R-peaks were detected. Second, search windows were defined incorporating the R-peak positions. For each wave, three different options for search windows were defined taking into account either the current or the subsequent RR-interval. Third, the option of filtering vs. no filtering was assessed in a preprocessing step. Fourth and final, each of the three wave event detection algorithms was applied. All options were used for all three wave event detection algorithms. The performances were assessed using two publicly available databases from the Physionet website, the MIT-BIH Arrhythmia database and the QT database, and one self-recorded clinical database – the Short-Term FAU database. The best algorithm was the fast naive wave event detection algorithm without any filtering. This algorithm is dependent on the search window definition and, due to its nature, will detect a wave event in every given search window. Whether this impedes with algorithms applied afterwards, or is even beneficial due to the completely different results, has to be investigated in future work.

9.3 IMU-based Activity Recognition

The IMU-based activity recognition was concerned with the presentation of the benchmark dataset DaLiAc and a hierarchical classification system. The DaLiAc dataset enabled the comparison of state of the art activity recognition algorithms. Both aspects were presented in Chapter 8 in the contributions (V) and (VI).

The dataset DaLiAc is publicly available for download (<http://www.activitynet.org/>). Data of 19 subjects were acquired with four IMUs. These IMUs were placed at the chest, the right hip, the right wrist, and the left ankle. All 19 subjects performed 13 different daily life activities. The recorded daily life activities were relevant activities that other researchers have used as well. So far, 13 other research groups have used this dataset to assess the performance of their algorithms. This dataset is suited for comparing state of the art activity recognition algorithms.

In the hierarchical classification system, activities were grouped and ordered in a hierarchy. This resulted in advantages like the easy integration of new activities without retraining the complete system, and the broad range of options for each single classification system: different features, window sizes, or classifiers. The hierarchical classification system outperformed six state of the art activity recognition algorithms from literature considering the complete DaLiAc dataset. The hierarchical classification system had an overall mean classification rate of 89.6%.

Chapter 10

Discussion

In this thesis, three wearable computing applications were presented. In this chapter, these are discussed individually. This is done in relation to address non-communicable diseases.

10.1 Mobile Breathing Analysis

In this part, the idea was to substitute FMs, which are the current standard in acquiring breathing parameters, with RIP. One disadvantage associated to FM measurement is that FMs change the natural breathing pattern. Subjects have a smaller breathing rate and a higher tidal volume to keep up with the demand which should remain constant. If a measurement device like RIP is used, which does not change the breathing pattern, this would result in acquiring the normal breathing pattern including breathing frequency and tidal volume. In this thesis, both devices – RIP and FM – were used simultaneously to verify that same tidal volumes can be measured with RIP. If in the future RIP will be used alone, the breathing pattern should change and more normal values should be acquired.

The two different adjustment algorithms presented in this thesis were the individual approach with VMCs from a corresponding still standing phase (Chapter 4) and the generalizable, reference-free approach (Chapter 5). Each of them represent one contribution (contributions (I) and (II), respectively). The performance for both algorithms was assessed during the treadmill running and the recovery phase. For practical usability, an approach without any prerecording should be favored as this reduces the needed study equipment and the preparatory time to a minimum. This speaks in favor of the reference-free approach (contribution (II)). Just the RIP device, a bathroom scale, and a measuring tape are then necessary.

This enables the usability of RIP outside the controlled laboratory environment to gather field data in real-world situations. Besides, better results were obtained with the generalizable and reference-free algorithms.

For this part, all measurements were taken in upright position: the different subjects were either standing or running on the treadmill. In literature, it is reported that different postures like sitting and lying influence the VMCs (Brüllmann, Fritsch, Thurnheer, & Bloch, 2010; Grossman, Wilhelm, & Brutsche, 2010; Stradling, Chadwick, Quirk, & Phillips, 1985). Grossman et al. (2010) proposed to perform separate adjustment procedures for different postures and activities. Their reported performance was comparable with the algorithms of Sackner et al. (1989) or Chadha et al. (1982) using data of 20 men. In this thesis, data from the upright position were used, whereas the data also originated from different phases – standing still and treadmill running – with low and high tidal volume data. It has to be investigated if and how the adjustment algorithms need to be adapted when a change in orientation occurs. Data from 193 subjects performing an incremental running test were used. This is the study with the largest study population reported in literature. Additionally, different tidal volume lengths and breathing frequencies exist in the available data. Thus, a certain variety in the data is available that is unique compared to reported datasets from literature. This aspect and the integrated concepts of machine learning like 10-fold cross-validation form robust algorithms for unseen data, with the explicit consideration of posture changes in future work.

Two additional aspects have to be considered in further studies. RIP devices are influenced by the correct fitting and fixation of the two bands (Brüllmann et al., 2010; Hollier et al., 2014). Additionally, Brown et al. (1998) reported that data acquired with RIP devices can drift over time. The correct fitting and fixation was guaranteed as the same and experienced sport scientist performed every measurement. A possible drift was considered and compensated for by the VBA routine to synchronize FM to RIP. Nevertheless, these aspects have to be kept in mind when the subject uses just the RIP device alone without supervision.

The suggested algorithms from this part might not be appropriate to all possible application areas of RIP breathing measurement. Healthy subjects underwent an incremental treadmill running protocol. This resulted in larger tidal volumes for all subjects with increasing running speed. Best results were obtained using a combination of both bands assuming that the chest wall has two moving parts. For certain fields of application like sleep evaluation or airway obstruction, the assumption

that the chest wall has two moving parts might not be valid. In all algorithms, the two degree of freedom model by Konno and Mead (1967) was considered where each band acts independently of each other. Recent research indicated that there is a more complex interaction between RC and AB, with multiple degrees of freedom present during respiration (Hollier et al., 2014; Raoufy et al., 2013). This would result in easy violations of the assumptions of RIP algorithms and linear adjustment approaches even in healthy individuals, and presumably with even higher influence during alterations of breathing pattern or asynchronous breathing (Poole, Thompson, Hallinan, & Beardsmore, 2000; Raoufy et al., 2013). Raoufy et al. (2013) suggested to use non-linear adjustment models to improve the accuracy of RIP. The suggested adjustment algorithms in this thesis were built on the two degree of freedom model (Konno & Mead, 1967), whereas extended extensively and validated using a large database and with machine learning concepts. This suggests that the adjustment algorithms are yet quite robust. Nevertheless, non-linear models for the adjustment of RIP data could be part of future studies.

For certain diseases, no data with higher tidal volumes might be available. The major interests in this part were healthy subjects with running as form of sport. Running was here realized by treadmill running. Even for other forms of sports like cycling when the position of the upper body changes compared to running, the determined adjustment algorithms might not be applicable. One reason might be the reallocation of the two bands (RC and AB). Still, the presented algorithms can be adapted to a variety of application areas. Then, the algorithms including the determination of the best set of predictors have to be retrained.

Respiratory diseases and respiratory infections are amongst the most leading deaths group. RIP devices with appropriate adjustment algorithms enable the possibility of measuring breathing parameters unobtrusively and accurately. Before, this was not possible to this extent. In this thesis, data from RIP devices were acquired from healthy athletes on an incremental treadmill running test. It has to be investigated how RIP devices can be effectively used for patients having respiratory diseases or respiratory infections.

10.2 Mobile ECG Analysis

In this part, the idea was to provide real-time algorithms for ECG analysis. Real-time arrhythmia classification just with RR instances has the advantage that single ECG leads are sufficient as appropriate single-lead R-peak

detection algorithms exist. This was comprised in contribution (III) (chapter 6). These detection algorithms even provide high accuracies for noisy and/or complicated recordings (Asl et al., 2008). Noisy ECG recordings happen easily when the subject is moving, touching (accidentally) the electrodes, cables, or other measurement equipment sensitive to noise. Classification algorithms working with a high performance even with these constraints are advantageous.

Additionally, other mobile methodologies like video or microphone recording, photo-plethysmographic signals, and others have been proven to detect the heart rate accurately. The recording of photo-plethysmographic signals is integrated in smartwatches and smartphones (McManus et al., 2013; Scully et al., 2012) to measure the pulse oximeter signal. Newly, more and more fitness applications measure the heart rate and provide feedback to the user. If the currently available algorithms based on the RR interval signal – as in this thesis for arrhythmia classification – are extended to various settings (e.g. in- and outside of clinical environments), this could have a tremendous impact. One possible example would be that collecting a standard 12-lead ECG signal is not required anymore. Simple approaches for measuring the RR interval signal, without expensive equipment are already available and need to be further improved in the near future. Additionally, these HRV algorithms might be integrated in devices that were originally intended to provide training feedback, but could further be used to prevent avoidable heart conditions through early screening and warning.

Also, in this part, three single-lead algorithms for real-time P- and T-wave detection were compared (contribution (IV) presented in chapter 7). In this thesis, the fast naive wave event detection algorithm performed best. All presented algorithms were limited in detecting the peak of the corresponding P- and T-waves. The delineation of the complete waves – the detection of the beginning and end of the corresponding waves – was not considered. The delineation of the complete waves lacks the required annotations. Only selected records of the QT database (Laguna et al., 1997) could be used as lead II of Einthoven was a prerequisite. In this database, almost no annotations for the beginning of the T-wave were provided. No annotations for the beginning and end of the P- and T-peak for the MIT-BIH Arrhythmia database (Elgendi et al., 2015; Goldberger et al., 2000; G. Moody & Mark, 2001) were available. In the Short-Term FAU database, the annotations were available, whereas in total from just 427 heartbeats. There is a clear need to provide to these commonly used databases valid annotations for all fiducial points of the ECG signal to enable the evaluation of different fiducial points detection algorithms. These

annotations should include the beginning and end of the P- and T-peak as well as the annotations for the complete QRS complex. The detection and delineation of these single waves form the basis for accurate monitoring of the heart, and for high performing arrhythmia classification algorithms. With those, subjects could be sent timely to the emergency department or an outpatient unit to minimize or eliminate resulting damage.

Cardiovascular diseases are the number one cause of deaths worldwide. Having a simple and effective arrhythmia classification algorithm could send people in risk early to a physician. There, thorough tests could further clarify the underlying disease or condition. Such a simple algorithm could require only the time instances of individual heart beats – the RR intervals. In this thesis, such an algorithm was presented. High-accurate P- and T-wave detection algorithms requiring only a single lead enhance the patient's comfort in long-term monitoring: the number of necessary electrodes and thus the measurement equipment is reduced. The fast naive wave event detection algorithm works with a single lead. These algorithms might be additionally beneficial in detecting specific diseases or conditions.

10.3 IMU-based Activity Recognition

In this part, the idea was to use IMUs for activity recognition (chapter 8). The dataset DaLiAc was acquired and made publicly available to enable and facilitate the comparison of state of the art activity recognition algorithms (contribution (V)). Up to now, the corresponding publication (Leutheuser, Schuldhaus, & Eskofier, 2013) was cited 95 times considering the Scopus (Scopus, 2019) and the Google Scholar (Google Scholar, 2019) document search. In 13 publications of them, other research groups used the DaLiAc dataset to assess the performance of their algorithms.

In DaLiAc, data were collected under controlled conditions. Data under more realistic conditions, e.g. including transitions between single activities or subjects performing the activities without a supervisor, would be desirable. Besides, Warren et al. (2010) suggested to consider the day-to-day and seasonal variability when assessing physical activity. When acquiring DaLiAc, the data acquisition lasted for about one hour for each subject during the same season. In this part, the focus was on recognizing the different activities instead of assessing the overall physical activity of individual subjects. Depending on the kind of data that should be acquired and the accompanying use case, e.g. measuring the course of amount of physical activity over various days, the day-to-day and seasonal variability

should be considered. Additionally, in these use cases, the algorithm could adapt to common activities of the corresponding season. In the current case, activity recognition of single activities was the aim. Hence, the day-to-day and seasonal variability could be neglected.

In DaLiAc, data were acquired with four IMUs placed at different body positions. The hierarchical classification system was based on the raw accelerometer and gyroscope data of all four IMUs (contribution (VI)). Fewer or single IMUs could be more pleasant to wear for the subject when acquiring data over multiple days or weeks. It is often assumed that by using more IMUs or sensors, complex activities or tasks can be classified more precisely, enhancing the accuracy of the recognition system. Certainly, this depends on the chosen sensors and their positions. It was not the focus to investigate the best position in terms of accuracy, placement, or sampling rate. The intended use case of this study was the recognition of daily life activities.

The hierarchical classification system has some benefits compared to other approaches from literature. New activities could be easily integrated. The broad range of options ensured that the activities could be classified with high accuracy. The subsystem BICYCLE and the two bicycling activities with different resistance levels scored worst among the five single subsystems and the single activities. In this thesis, single activities should be classified and the two bicycling activities were recorded as two single activities. A different arrangement of the activities, e.g. in activity levels or broad activity groups, would change the setup of the hierarchical classification system and constitute a different classification task. In future work, it has to be investigated what is most beneficial in addressing the risk factor of physical inactivity.

The ongoing research for activity recognition develops into two directions. First, the number of IMUs are reduced and they are integrated into smart devices which are commonly used in daily life. Even without integrating them in smart devices, miniaturized versions could just be placed at voluntary positions. Then, an automatic assessment of the current IMU position is needed to adapt and refine the algorithm correspondingly (Mannini, Sabatini, & Intille, 2015). Second, IMUs are integrated into clothing. Then, the number of IMUs is not limited and multiple sensors from different methodologies could be integrated. For this, the connection within the different sensors must function. The distinction in one of these options is made depending on the use case.

Physical inactivity is one of the four leading risk factor for non-communicable diseases. The deaths of approximately 3.2 million people is accounted by physical inactivity. Non-communicable diseases are one of

the important challenges for healthcare systems globally. Recognizing activities of daily living might increase the awareness of the individuals to increase their physical activity.

Chapter 11

Outlook

Non-communicable diseases are the leading cause of death and disability worldwide. In this thesis, different wearable computing applications were introduced. These wearable computing applications focused on single or multiple non-communicable diseases. In the mobile breathing analysis part, accurate and unobtrusive breathing measurements in the field are enabled using the RIP device and appropriate reference-free adjustment algorithms. In the mobile ECG analysis part, two approaches were presented. The RR arrhythmia classification could facilitate the early detection of arrhythmic beats. The P- and T-wave detection improved the detection of fiducial points in the ECG signal. In the IMU-based activity recognition part, the DaLiAc dataset and a hierarchical classification system were proposed to address one of the leading risk factors for non-communicable diseases – physical inactivity.

In this thesis, different biosignals were used as a basis for each part. These biosignals were either unadjusted tidal volume values of the RC and AB band, the ECG signal, or the accelerometer and gyroscope data of the IMUs. The fusion of the biosignals could improve the performance of the single application fields and create new application areas.

Including the ECG signal or perhaps just the heart rate as additional parameters in the regression equations could further improve the results of the adjustment algorithms for RIP. IMU data could further give insights into the current position or activity of the patient or athlete. If needed, the adjustment algorithm could then switch between modes to ensure appropriate breathing parameters.

Being afraid or under stress increase the current heart rate. The simultaneous measurement of IMUs and the ECG could provide beneficial insights to the current activity and enhance both, the results for activity

recognition and the results for mobile ECG analysis. Also, certain activities are influenced by breathing. These activities belong for example to relaxation techniques, sport activities like yoga, singing or playing musical instruments.

The fusion of these biosignals could also be realized by combining them into one measurement principle. The measurement of the three biosignals could for example be integrated into one shirt or textile. The FitnessSHIRT (Fraunhofer IIS, Erlangen, Germany) is one example of a textile with which the three biosignals, ECG, breathing, and acceleration, can be measured (Leutheuser, Lang, et al., 2017). The three modalities could further be integrated into a smartwatch or smartphone. Apple presented recently the Apple Watch Series 4 (Apple Inc., Cupertino, California, US). With this smartwatch, the user is able to acquire an ECG and single IMU data. Using a smartphone, similar signals can be measured: the heart rate using the integrated camera, the breathing rate through the microphone, and movement data from the integrated IMU.

Using wearables and wearable computing in daily life could bring certain benefits to the user. Apart from the possibilities of measuring these biosignals and their analysis, it is still an open field of research to investigate why the adherence to these wearables declines after some time (Hermesen, Moons, Kerkhof, Wiekens, & De Groot, 2017). Future work should investigate how people can be motivated to adhere to wearables. Also, it should be evaluated if patients with prescriptions for wearables adhere longer to these. For the others, devices as unobtrusive and comfortable in wearing could be one aspect for sustaining longer to them. In the end, a lifestyle change is associated with using wearables that the user should put into practice.

In recent years, wearables have been increasingly used for health monitoring (Banaee, Ahmed, & Loutfi, 2013) to improve health and healthcare systems and to change the lifestyle in the long term. This is associated to challenges which should and will be addressed by future research. Algorithms should be based on reliable data that came from the actual (clinical) setting. For building up on this thesis, this would mean to acquire RIP data from patients having non-communicable diseases or to get real-life activity data from subjects being physically inactive. This leaves the research question how a realistic surrounding for data acquisition is created where the subjects act normally and are not influenced by the measurement equipment. Having unobtrusive wearables integrated in textiles or in objects worn in everyday life should be taken for granted.

Another challenge is concerned with the labeling of the real-world data. If clinical datasets were acquired, appropriate experts, e.g. specialists or clinicians must label them. Often, someone supervises the data acquisition to be able to set afterwards the correct labels. It is unclear if authentic data can be recorded under the eyes of a supervisor. Furthermore, labeling of (long-term) data is time consuming, expensive and non-trivial (Banaee et al., 2013). In this thesis, all self-recorded data were supervised by one or more scientists. This ensured that the data were labeled correctly whereas might limit the authenticity of the data. An additional answer to this challenge might be to apply deep learning and unsupervised learning methods to the raw data. Then, no labels would be needed.

Three wearable computing applications were presented in this thesis. Similar to them, various wearable computing applications already exist on the market. More and more new wearable computing applications will enter the market to support and address the current health challenges. This will not only apply for the non-communicable diseases, whereas for all diseases and the complete healthcare systems. The new applications will be fostered by new inventions, new knowledge, technological process, and progressive digitalization.

Appendices

Chapter A

Additional Results for Mobile Breathing Analysis

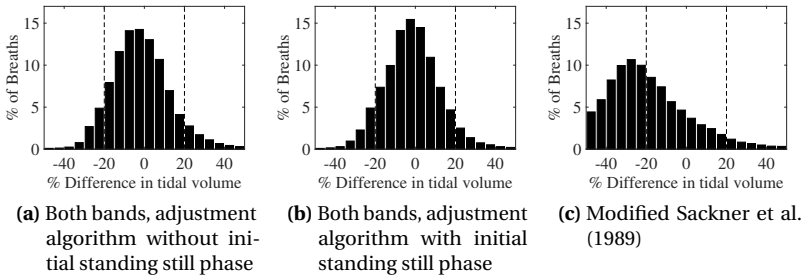


Figure A.1: Frequency histograms during treadmill running using the reliability dataset. % Difference in tidal volume was calculated using $V_T - V_{T_{FM}}$ with regard to $V_{T_{FM}}$, with V_T as the calculated tidal volume and $V_{T_{FM}}$ as the reference volume of the flowmeter (FM).

Table A.1: Limits of Equivalence (LoE) in percent (%) and slopes over all ten conditions for the reliability dataset.

		Treadmill Running				Recovery			
		LoE		Slopes		LoE		Slopes	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Adjustment without standing still phase	only RC	75.07 ± 19.89		1.50 ± 0.40		65.82 ± 19.99		0.67 ± 0.15	
	only AB	63.03 ± 22.97		0.82 ± 0.14		56.83 ± 17.27		1.28 ± 0.41	
	both bands	83.42 ± 16.64		1.22 ± 0.20		77.00 ± 20.04		1.13 ± 0.22	
Adjustment with standing still phase	only RC	76.72 ± 22.60		1.16 ± 0.25		68.67 ± 21.66		1.12 ± 0.26	
	only AB	66.72 ± 20.86		1.25 ± 0.37		60.73 ± 17.43		1.12 ± 0.33	
	both bands	84.68 ± 17.62		1.07 ± 0.17		72.54 ± 22.83		1.06 ± 0.22	
Comparison to literature	Chadha et al. (1982)	95.05 ± 4.83		0.93 ± 0.10		89.71 ± 8.57		0.93 ± 0.12	
	Sackner et al. (1989)	94.76 ± 5.17		0.80 ± 0.08		86.49 ± 10.74		0.59 ± 0.13	
	mod. Sackner et al. (1989)	39.91 ± 36.76		0.91 ± 0.18		33.40 ± 32.97		0.68 ± 0.18	
	Banzett et al. (1995)	94.83 ± 5.00		0.81 ± 0.08		88.93 ± 8.80		0.66 ± 0.12	

SD: standard deviation

Table A.2: Accuracy measures for the comparison of the adjustment algorithms of both bands to the algorithm from Chadha et al. (1982) using the reliability dataset.

	Treadmill Running				Recovery			
	$N_{outliers}$ [%]		ΔV [I]		$N_{outliers}$ [%]		ΔV [I]	
	Mean±	SD	Mean±	SD	Mean±	SD	Mean±	SD
Adjustment without standing still phase	0.03±	0.02	0.03±	0.21	0.05±	0.04	0.02±	0.16
Adjustment with standing still phase	0.03±	0.03	0.01±	0.22	0.05±	0.04	0.02±	0.19

SD: standard deviation

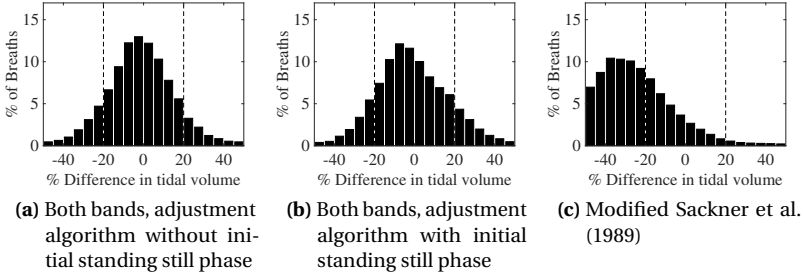


Figure A.2: Frequency histograms during recovery using the reliability dataset. % Difference in tidal volume was calculated using $V_T - V_{T_{FM}}$ with regard to $V_{T_{FM}}$, with V_T as the calculated tidal volume and $V_{T_{FM}}$ as the reference volume of the flowmeter (FM).

Chapter B

Additional Results for Mobile ECG Analysis

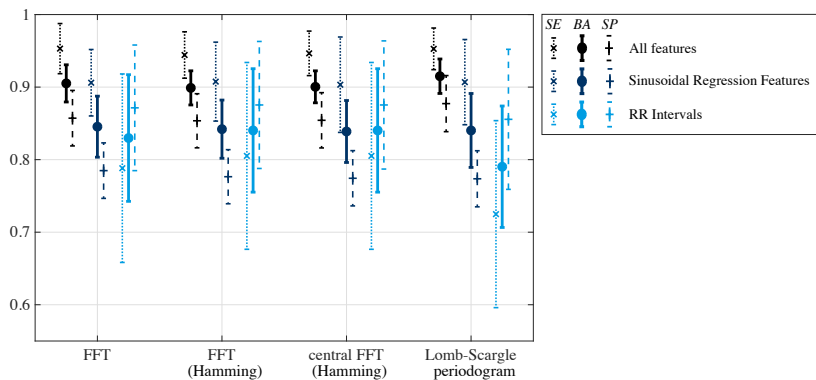


Figure B.1: Effect of different feature groups (FGs) on sensitivity SE , balanced accuracy BA , and specificity SP for RR interval classification using the four different frequency calculation methods. Results are shown using 15 RR intervals. The evaluation metrics are averaged over the 10-folds and are shown with the standard deviations as error bars. FFT: Fast Fourier Transform

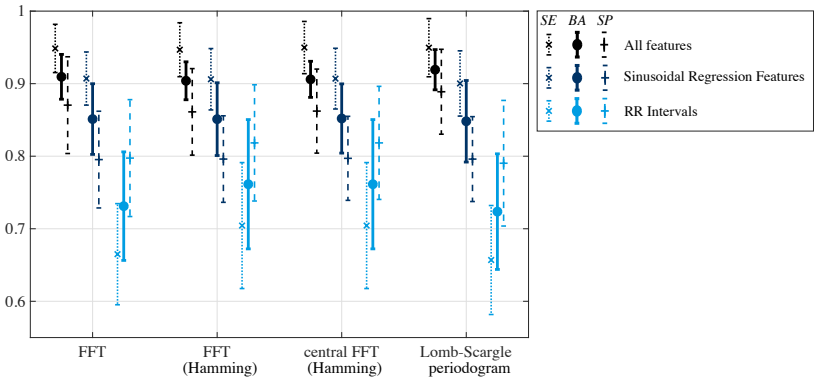


Figure B.2: Effect of different feature groups (FGs) on sensitivity SE , balanced accuracy BA , and specificity SP for RR interval classification using the four different frequency calculation methods. Results are shown using 20 RR intervals. The evaluation metrics are averaged over the 10-folds and are shown with the standard deviations as error bars. FFT: Fast Fourier Transform

Table B.1: Evaluation metrics in percent (%) over the different frequency calculation methods when considering only the sinusoidal regression feature (mean $\pm$ standard deviation). Segments with invalid frequency calculation were excluded (Table 6.3).

	Frequency options			
	FFT	FFT (Hamming)	central FFT (Hamming)	Lomb-Scargle periodogram
Positive Predictivity				
10 RR intervals	68.58 $\pm$ 14.07	69.05 $\pm$ 13.83	68.27 $\pm$ 13.65	66.54 $\pm$ 14.48
15 RR intervals	73.17 $\pm$ 14.00	71.83 $\pm$ 14.74	71.59 $\pm$ 14.91	70.46 $\pm$ 16.73
20 RR intervals	75.02 $\pm$ 13.45	75.13 $\pm$ 13.67	75.34 $\pm$ 13.21	74.21 $\pm$ 14.13
Accuracy				
10 RR intervals	78.74 $\pm$ 9.90	78.71 $\pm$ 9.70	78.15 $\pm$ 9.57	77.68 $\pm$ 8.30
15 RR intervals	82.56 $\pm$ 5.87	81.87 $\pm$ 5.93	81.44 $\pm$ 5.98	81.46 $\pm$ 7.31
20 RR intervals	83.64 $\pm$ 6.59	83.64 $\pm$ 6.75	83.75 $\pm$ 6.44	83.28 $\pm$ 7.18

FFT: Fast Fourier Transform

Table B.2: Evaluation metrics in percent (%) over the different frequency calculation methods when considering only the RR intervals (mean $\pm$ standard deviation). Segments with invalid frequency calculation were excluded (Table 6.3).

	Frequency options			
	FFT	FFT (Hamming)	central FFT (Hamming)	Lomb- Scargle periodogram
Positive Predictivity				
10 RR intervals	85.62 $\pm$ 16.54	85.61 $\pm$ 16.24	85.61 $\pm$ 16.24	86.05 $\pm$ 16.81
15 RR intervals	87.69 $\pm$ 16.17	87.60 $\pm$ 16.24	87.60 $\pm$ 16.24	87.36 $\pm$ 17.13
20 RR intervals	84.10 $\pm$ 12.53	85.02 $\pm$ 12.34	85.02 $\pm$ 12.34	83.11 $\pm$ 13.49
Accuracy				
10 RR intervals	78.13 $\pm$ 11.99	76.01 $\pm$ 12.81	76.01 $\pm$ 12.81	75.19 $\pm$ 13.06
15 RR intervals	80.08 $\pm$ 11.64	81.37 $\pm$ 11.31	81.37 $\pm$ 11.31	74.71 $\pm$ 11.31
20 RR intervals	70.29 $\pm$ 7.78	73.92 $\pm$ 9.39	73.92 $\pm$ 9.39	69.55 $\pm$ 8.37

FFT: Fast Fourier Transform

Chapter C

Subject Information for the DaLiAc Dataset

Information

Dear volunteer,

**thank you for participating in
this study about**

**non-invasive biosignal
analysis for
sports and movement.**



Details about the study

The Digital Sports Group (Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg) deals with the sensor-based application of pattern recognition algorithms in the area of sports and movement. The used sensors are small, lightweight and non-invasive and include e.g. motion sensors (accelerometer, gyroscope, magnetometer) and physiological sensors (electrocardiogram (ECG), electromyogram (EMG), electroencephalogram (EEG)). The results of this analysis are an important aspect for the creation of athletes-support-systems.

The generation and analysis of new data is essential for the implementation and evaluation of new algorithms. The sensors are for example integrated in clothes or are placed with the help of different holders on the human body. It is planned to place the sensors for example to the following positions:

- upper and lower limbs
- torso
- hip or
- head.

The sensors are placed in such a way that they do not influence the participants. There is no danger due to the sensors.

All activities are performed in a laboratory or under real conditions outside the laboratory. Possible activities are for example:

- walking
- ascending and descending stairs
- jogging
- running
- swimming
- rope jumping
- bicycling
- isolated movements of single muscle groups or
- other typical body movements.

Profile of Subject

Surname	_____
First name	_____
Street	_____
ZIP / City	_____
Telephone	_____
E-mail	_____

Date of birth	_____
Sex	☐ male ☐ female
Body height [cm]	_____
Body weight [kg]	_____
Size of shoe [EUR]	_____
Dress size [S, M, L, XL]	_____
Chest width [cm]	_____
Handedness	☐ left ☐ right

Physical Activity Readiness Questionnaire (PAR-Q)

For most people, physical activity should not pose any problem or hazard. The PAR-Q is designed to identify the small number of adults for whom physical activity might be inappropriate or those who should seek medical advice concerning the type of activity most suitable for them.

1. Do you have a bone or joint problem such as arthritis, which has been aggravated exercise or might be made worse with exercise?	Y / N
2. Do you have or ever suffered a heart condition?	Y / N
3. Have you ever felt pain in your chest when you do physical exercise?	Y / N
4. Is your doctor currently prescribing you drugs or medication?	Y / N
5. Have you ever suffered from shortness of breath at rest or with mild exercise?	Y / N
6. Do you ever feel faint, have spells of dizziness or have ever lost consciousness?	Y / N
7. Do you know of any other reason why you should not participate in a program of physical activity?	Y / N

If you answered NO to all questions:

If you answered the PAR-Q honestly and accurately, you have reasonable assurance of your present suitability for participating in the study about activity recognition.

Assumption of Risk

I hereby state that I have read, understood and answered honestly the questions above. I also state that I wish to participate in all the activities mentioned in *details about study*. I realize that my participation in these activities involve the risk of injury. I hereby confirm that I am voluntarily engaging in an acceptable level of activity, which has been recommended to me.

In addition

All of my additional questions were answered to my complete satisfaction.

On the part of the university no insurance protection exists for the participant.

Erlangen, the	_____
Signature of participant	_____
Erlangen, the	_____
Signature of advisor	_____

Agreement for Participation in the Study

With your signature you confirm the following statements:

- The content of the study was completely explained to me.
- I read and understood the section *details about study*.
- All of my additional questions were answered so that I was completely satisfied.
- I agree to participate in the mentioned study.
- The participation in the study is voluntarily.
- I know that I can cancel the agreement for participation at any time.

Data protection statement:

I hereby agree that the acquired data and the personal data can be used for scientific research. The acquired data will be treated in an anonymous manner (i.e. the mapping to your name is only possible with utilities like a reference list). The acquired data can be used for

- this study
- further studies or publications.

The acquired data can be disclosed to third parties in an anonymous manner. The agreement can be canceled at any time and without the specification of reasons.

Withdrawal of the agreement for data usage:

I know that I can withdraw my agreement for data usage at any time and without any reason.

In case of withdrawal, I agree that the data will be saved for control purposes.

I have the right to demand deletion.

I know that, in case of anonymous storage, the requested deletion of the data is not possible.

Erlangen, the	_____
Signature of participant	_____
Erlangen, the	_____
Signature of advisor	_____

Contact Person

If you have questions please contact:

Prof. Dr. Bjoern Eskofier

Digital Sports Group

Pattern Recognition Lab

Martensstr. 3

91058 Erlangen

Phone: 09131 85 27297

e-mail: bjoern.eskofier@cs.fau.de

List of Figures

1.1	Distribution of deaths by leading cause groups	4
1.2	Structure of this thesis	13
3.1	Waves and fiducial points of an ECG signal	49
3.2	Electrode placement for ECG measurement	50
3.3	Visualization of SVR	54
3.4	Steps of a supervised learning system	55
3.5	Optimal hyperplane as determined by the SVM classifier	59
3.6	Sketch of 10-fold cross-validation	60
3.7	Sketch of two nested 10-fold cross-validations	62
4.1	Analysis steps for individual-adapted adjustment of RIP	66
4.2	LifeShirt garment	67
4.3	RMSE values for individual adjustment using the validation dataset	73
4.4	RMSE values for individual adjustment using the reliability dataset	74
5.1	Analysis steps for reference-free adjustment of RIP	80
5.2	RMSE values for reference-free adjustment using the validation dataset	90
5.3	RMSE values for reference-free adjustment using the reliability dataset	91
5.4	Frequency histograms during treadmill running using the validation dataset	94
5.5	Frequency histograms during recovery using the validation dataset	94
6.1	Visualization of the sinusoidal model and the sinusoidal regression feature	107

List of Figures

6.2	Effect of different FGs using RR interval classification	111
6.3	Impact of segment length for RR interval classification . . .	112
7.1	Analysis steps for ECG fiducial points detection	120
7.2	Search window definitions for P- and T-wave detection . .	122
7.3	Illustration of the line fitting wave event detection algorithm	124
7.4	Flowchart of the Wavelet decomposition	125
7.5	Sensitivities for P- and T-wave detection for the three wave event detection algorithms applied to three differ- ent databases	129
8.1	Placement of the IMUs used for the DaLiAc dataset	144
8.2	Screenshot of the data labeling application	146
8.3	IMU synchronization	147
8.4	Hierarchical classification system	149
A.1	Frequency histograms during treadmill running using the reliability dataset	183
A.2	Frequency histograms during recovery using the reliabil- ity dataset	186
B.1	Effect of different FGs using RR interval classification with 15 RR intervals	187
B.2	Effect of different FGs using RR interval classification with 20 RR intervals	188

List of Tables

2.1	Ambulatory and indirect determination of tidal volume . .	16
2.2	Algorithms for atrial fibrillation or arrhythmia detection using RR intervals	24
2.3	Subjective and objective measures for the assessment of physical activity	34
2.4	Machine learning techniques for activity recognition	39
2.5	Exemplary representation of supervised activity recog- nition systems considering daily life activities of healthy subjects	42
2.6	Summarized activities of the previous Table 2.5	45
4.1	Characteristics of the standing still phase	68
4.2	Characteristics of the treadmill running phase	68
4.3	Characteristics of the recovery phase	68
4.4	Subject characteristics for the three available datasets . . .	69
4.5	SVR parameters c and ϵ determined using the training dataset	72
5.1	Determined condition number κ for the different phases and adjustment algorithms	87
5.2	Determined anthropometric parameters for each of the twelve conditions	89
5.3	LoE in percent (%) and slopes over all ten conditions for the validation dataset.	92
5.4	Accuracy measures using the validation dataset	93
6.1	Overview of the 36 classification systems	109
6.2	Confusion matrix for the two classes pathologic and normal	110
6.3	Number of segments with invalid frequency calculation . .	111

List of Tables

6.4	Evaluation metrics in percent (%) over the different frequency calculation methods when considering all features	113
7.1	Overview of databases, the used records, and the number of annotated beats for ECG fiducial points detection	118
7.2	Overview of used annotations from the QT database	119
7.3	Summarized overview of data, preprocessing, and algorithms for ECG fiducial points detection	126
7.4	Sensitivities SE in percent (%) for P- and T-wave detection comparing results of the best search window to the smallest search window	127
7.5	Positive predictivity $+P$ in percent (%), mean absolute error MAE in milliseconds (ms), and mean SD in (ms) for P-wave detection	130
7.6	Positive predictivity $+P$ in percent (%), mean absolute error MAE in milliseconds (ms), and mean SD in (ms) for T-wave detection	131
7.7	Comparison of best P- and T-wave detection algorithms to algorithms from literature	132
8.1	Activities in the DaLiAc dataset	145
8.2	Sensors placement in the six state of the art algorithms in the literature that were implemented and compared in the present study	152
8.3	Mean classification in percent (%) of the five subsystems .	154
8.4	Confusion matrix of the proposed hierarchical classification system	155
8.5	Mean class dependent classification rates in percent (%) for all 13 activities and overall mean classification rates of proposed system and state of the art systems	156
8.6	Usage of DaLiAc in literature	157
A.1	LoE in percent (%) and slopes over all ten conditions for the reliability dataset	184
A.2	Accuracy measures using the reliability dataset	185
B.1	Evaluation metrics in percent (%) over the different frequency calculation methods when considering only the sinusoidal regression feature	188

B.2	Evaluation metrics in percent (%) over the different frequency calculation methods when considering only the RR intervals	189
-----	---	-----

Glossary

AASM American Academy of Sleep Medicine. 20

AB Abdomen. 8, 20, 21, 65, 68, 79, 80, 81, 82, 86, 87, 88, 94, 95, 96, 99, 166, 170, 171, 177

AdaBoost Adaptive Boosting. 55, 56, 57, 150, 153, 154

BMI Body Mass Index. 147

CART Classification and Regression Tree. 55, 56, 57, 150, 153

ECG Electrocardiogram. iii, iv, vii, viii, 3, 6, 8, 9, 10, 12, 13, 16, 22, 28, 29, 30, 31, 32, 47, 48, 49, 55, 103, 104, 112, 117, 118, 119, 120, 121, 123, 124, 126, 128, 136, 138, 139, 165, 166, 167, 171, 172, 177, 178, 199, 200, 202

FFT Fast Fourier Transform. 39, 42, 105, 106, 107, 108, 110, 111, 112, 116, 187

FG Feature group. 55, 107, 108, 110, 112, 114, 187, 199, 200

FM Flowmeter. iii, vii, 5, 7, 8, 10, 21, 65, 66, 69, 70, 71, 72, 75, 76, 78, 79, 84, 86, 88, 94, 96, 97, 98, 165, 169, 170, 183

GUI Graphical User Interface. 119

HRV Heart Rate Variability. 11, 22, 28, 114, 116, 166, 172

ICU Intensive care unit. 6

IMU Inertial Measurement Unit. iii, iv, vii, viii, 3, 7, 11, 12, 13, 33, 34, 39, 143, 144, 147, 146, 149, 150, 151, 159, 161, 165, 167, 173, 174, 177, 178, 200

kNN k-Nearest Neighbor. 55, 56, 58, 150, 153, 154

LoE Limits of Equivalence. 84, 88, 183, 201, 202

LSQ Least squares approximation. 20, 50, 51, 61, 65, 70, 71, 72, 75, 76, 77, 78, 82, 105, 123, 165

MET Metabolic Equivalent of Task. 33, 34, 145

PAR-Q Physical Activity Readiness Questionnaire. 147

QDC Qualitative Diagnostic Calibration. 21, 71, 76, 78

RC Rib cage. 5, 8, 20, 21, 65, 68, 79, 80, 81, 82, 86, 87, 88, 94, 95, 96, 97, 99, 166, 170, 171, 177

RIP Respiratory inductance plethysmography. iii, vii, 5, 7, 8, 10, 12, 13, 16, 20, 21, 47, 49, 55, 65, 69, 71, 72, 75, 76, 78, 79, 81, 86, 88, 94, 96, 97, 98, 99, 165, 166, 169, 170, 171, 177, 178, 199

RMSE Root mean square error. 51, 70, 71, 72, 75, 76, 84, 85, 88, 199

RSA Respiratory sinus arrhythmia. 16, 28, 105, 116, 166

SD Standard deviation. 21, 86, 88, 95, 97, 103, 104, 110, 119, 126, 128, 132, 137, 147, 202

SVD Singular value decomposition. 51, 52, 82, 105

SVM Support Vector Machine. 23, 24, 42, 55, 56, 59, 78, 150, 153, 199

SVR Support Vector Regression. 8, 50, 52, 53, 59, 61, 65, 70, 71, 72, 76, 77, 165, 199, 201

VBA Visual Basic for Applications. 69, 170

VMC Volume motion coefficient. iii, vii, 21, 70, 71, 72, 75, 76, 77, 80, 81, 165, 169, 170

WHO World Health Organization. 7

Bibliography

- Abdul Rahman, H., Ge, D., Le Faucheur, A., Prioux, J., & Carrault, G. (2017). Advanced classification of ambulatory activities using spectral density distances and heart rate. *Biomed Signal Process Control*, 34, 9–15. (Cit. on p. 157).
- Adams, R. (1999). Revised physical activity readiness questionnaire. *Can Fam Physician*. 45, 992–995. (Cit. on p. 148).
- Agostini, G., Longari, M., & Pollastri, E. (2003). Musical instrument timbres classification with spectral features. *EURASIP J Adv Sig Pr*. 2003, 5–14. (Cit. on p. 150).
- Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. J. J., Tudor-Locke, C., Greer, J. L., Vezina, J., Whitt-Glover, M. C., & Leon, A. S. (2011). 2011 compendium of physical activities: A second update of codes and MET values. *Med Sci Sports Exerc*. 43(8), 1575–1581. (Cit. on pp. 7, 145).
- Ainsworth, B. E., Haskell, W. L., Leon, A. S., Jacobs, D. R., Montoye, H. J., Sallis, J. F., & Paffenbarger, R. S. (1993). Compendium of physical activities: Classification of energy costs of human physical activities. *Med Sci Sports Exerc*. 25(1), 71–80. (Cit. on pp. 7, 145).
- Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., O'Brien, W. L., Bassett, D. R. J., Schmitz, K. H., Emplaincourt, P. O., Jacobs, D. R., & Leon, A. S. (2000). Compendium of physical activities: An update of activity codes and MET intensities. *Med Sci Sports Exerc*. 32 (supplement 9), 498–516. (Cit. on pp. 7, 145).
- Akhbari, M., Shamsollahi, M. B., Jutten, C., Armoundas, A. A., & Sayadi, O. (2016). ECG denoising and fiducial point extraction using an extended kalman filtering framework with linear and nonlinear phase observations. *Physiol Meas*. 37(2), 203–226. (Cit. on pp. 28, 31, 32, 121, 128, 133).
- Akhbari, M., Shamsollahi, M. B., Sayadi, O., Armoundas, A. A., & Jutten, C. (2016). ECG segmentation and fiducial point extraction using multi

- hidden markov model. *Comput Biol Med.* 79, 21–29. (Cit. on pp. 28, 31, 32, 121, 128, 132, 137).
- Ali, A. K. M., King, R., Janko, B., Sack, E., Ashburn, A., Burnett, M., Craddock, I., & Harwin, W. (2017). Activity recognition from body worn accelerometers - toward real-time event detection. *In Proc: Expert Update.* (Cit. on p. 157).
- Allen, F. R., Ambikairajah, E., Lovell, N. H., & Celler, B. G. (2006). Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models. *Physiol Meas.* 27(10), 935–951. (Cit. on pp. 42, 45).
- Anand, A., Sharma, M., Srivastava, R., Kaligounder, L., & Prakash, D. (2018). Wearable motion sensor based analysis of swing sports. *In Proc.: 16th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018*, 261–267. (Cit. on p. 157).
- Askanazi, J., Silverberg, P. A., Foster, R. J., Hyman, A. I., Milic-Emili, J., & Kinney, J. M. (1980). Effects of respiratory apparatus on breathing pattern. *J Appl Physiol Respir Environ Exerc Physiol.* 48(4), 577–580. (Cit. on p. 5).
- Asl, B. M., Setarehdan, S. K., & Mohebbi, M. (2008). Support Vector Machine-based arrhythmia classification using reduced features of Heart Rate Variability signal. *ArtifIntell Med.* 44, 51–64. (Cit. on pp. 23, 25, 114, 172).
- Awais, M. (2016). Physical activity classification meeting daily life conditions. *In Proc: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 393–398. (Cit. on p. 157).
- Awais, M., Chiari, L., Ihlen, E. A. F., Helbostad, J. L., & Palmerini, L. (2019). Physical activity classification for elderly people in free-living conditions. *IEEE J Biomed Health Inform.* 23(1), 197–207. (Cit. on p. 157).
- Awais, M., Mellone, S., & Chiari, L. (2015). Physical activity classification meets daily life: Review on existing methodologies and open challenges. *In Proc: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015*, 5050–5053. (Cit. on p. 157).
- Awais, M., Palmerini, L., Bourke, A. K., Ihlen, E. A. F., Helbostad, J., & Chiari, L. (2016). Performance evaluation of state of the art systems for physical activity classification of older subjects using inertial sensors in a real life scenario: A benchmark study. *Sensors (Basel),* 16(12), 2105. (Cit. on p. 157).
- Awais, M., Palmerini, L., & Chiari, L. (2016). Physical activity classification using body-worn inertial sensors in a multi-sensor setup. *In Proc: IEEE 2nd International Forum on Research and Technologies for*

- Society and Industry Leveraging a Better Tomorrow (RTSI)*. (Cit. on p. 157).
- Awais, M., Raza, M., Ali, K., Ali, Z., Irfan, M., Chughtai, O., Khan, I., Kim, S., & Ur Rehman, M. (2019). An internet of things based bed-egress altering paradigm using wearable sensors in elderly care environment. *Sensors (Basel)*, 19(11), 2498. (Cit. on p. 157).
- Baloch, Z., Shaikh, F. K., & Unar, M. A. (2019). Deep architectures for human activity recognition using sensors. *3C Tecnología. Glosas de innovación aplicadas a la pyme. Special Issue*, 14–35. (Cit. on p. 157).
- Banaee, H., Ahmed, M. U., & Loutfi, A. (2013). Data mining for wearable sensors in health monitoring systems: A review of recent trends and challenges. *Sensors (Basel)*, 13(12), 17472–17500. (Cit. on pp. 178, 179).
- Banzett, R. B., Mahan, S. T., Garner, D. M., Brughera, A., & Loring, S. H. (1995). A simple and reliable method to calibrate respiratory magnetometers and respiration. *J Appl Physiol*. 79(6), 2169–2176. (Cit. on pp. 5, 21, 80, 86, 87, 90–92, 95–97, 184).
- Bao, L. & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In A. Ferscha & F. Mattern (Eds.), *Pervasive computing* (Vol. 3001, pp. 1–17). Lecture Notes in Computer Science. Springer Berlin Heidelberg. (Cit. on pp. 42, 45, 144, 149, 151–153, 156, 158, 159).
- Barth, J., Oberndorfer, C., Pasluosta, C., Schüle, S., Gassner, H., Reinfelder, S., Kugler, P., Schuldhuis, D., Winkler, J., Klucken, J., & Eskofier, B. M. (2015). Stride segmentation during free walk movements using multi-dimensional subsequence dynamic time warping on inertial sensor data. *Sensors*, 15(3), 6419–6440. (Cit. on p. 144).
- Bayasi, N., Tekeste, T., Saleh, H., Khandoker, A., Mohammad, B., & Ismail, M. (2014). Adaptive technique for P and T wave delineation in electrocardiogram signals. In *Proc: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 90–93. (Cit. on pp. 28, 31, 128, 133, 137).
- Beraza, I. & Romero, I. (2017). Comparative study of algorithms for ECG segmentation. *Biomed Signal Process Control*, 34, 166–173. (Cit. on pp. 32, 121).
- Berlin, J. A. & Colditz, G. A. (1990). A meta-analysis of physical activity in the prevention of coronary heart disease. *Am J Epidemiol*. 132(4), 612–628. (Cit. on p. 7).
- Berry, R. B., Budhiraja, R., Gottlieb, D. J., Gozal, D., Iber, C., Kapur, V. K., Marcus, C. L., Mehra, R., Parthasarathy, S., Quan, S. F., Redline, S., Strohl, K. P., Davidson Ward, S. L., & Tangredi, M. M. (2012). Rules

- for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events: Deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine. *J Clin Sleep Med.* 8(5), 597–619. (Cit. on pp. 5, 20).
- Billiet, L. (2018). *Clinical decision support* (Doctoral dissertation, KU Leuven). (Cit. on p. 157).
- Billiet, L., Swinnen, T., de Vlam, K., Westhovens, R., & Van Huffel, S. (2018). Recognition of physical activities from a single arm-worn accelerometer: A multiway approach. *Informatics*, 5(2), 20. (Cit. on p. 157).
- Blair, S. N., Cheng, Y., & Holder, J. S. (2001). Is physical activity or physical fitness more important in defining health benefits? *Med Sci Sports Exerc.* 33 (supplement 6), 379–399. (Cit. on p. 7).
- Bland, J. M. & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1(8476), 307–310. (Cit. on p. 85).
- Blank, P., Hoßbach, J., Schuldhaus, D., & Eskofier, B. (2015). Sensor-based stroke detection and stroke type classification in table tennis. *In Proc: 2015 ACM International Symposium on Wearable Computers (ISWC)*, 93–100. (Cit. on p. 157).
- Boichat, N., Khaled, N., Rincón, F., & Atienza, D. (2009). Wavelet-based ECG delineation on a wearable embedded sensor platform. *In Proc: IEEE-EMBS 6th Internat. Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 256–261. (Cit. on pp. 28–30, 136, 137).
- Bote, J. M., Recas, J., Rincón, F., Atienza, D., & Hermida, R. (2018). A modular low-complexity ECG delineation algorithm for real-time embedded systems. *IEEE J Biomed Health Inform.* 22(2), 429–441. (Cit. on pp. 28, 30, 32, 121, 128, 132, 137).
- Bowers, E. J. & Murray, A. (2004). Interaction between cardiac beat-to-beat interval changes and systolic blood pressure changes. *Clin Auton Res.* 14, 92–98. (Cit. on pp. 28, 106).
- Brimacombe, J. R., Macfie, A. G., & McCrirrick, A. (1991). The extensometer: Potential applications in anaesthesia and intensive care. *Anaesthesia*, 46(9), 756–761. (Cit. on pp. 5, 15, 19).
- Bronstein, J. Z. & Brocks, L. J. (2017). A potential alternative to respiratory inductance plethysmography for children? *J Clin Sleep Med.* 13(2), 159–160. (Cit. on pp. 5, 15, 18, 20).
- Brown, K., Aun, C., Jackson, E., Mackersie, A., Hatch, D., & Stocks, J. (1998). Validation of respiratory inductive plethysmography using the qualitative diagnostic calibration method in anaesthetized infants. *Eur Respir J.* 12(4), 935–943. (Cit. on p. 170).

- Brüllmann, G., Fritsch, K., Thurnheer, R., & Bloch, K. E. (2010). Respiratory monitoring by inductive plethysmography in unrestrained subjects using position sensor-adjusted calibration. *Respiration*, 79(2), 112–120. (Cit. on p. 170).
- Burns, A., Greene, B. R., McGrath, M. J., O'Shea, T. J., Kuris, B., Ayer, S. M., Stroiescu, F., & Cionca, V. (2010). Shimmer – a wireless sensor platform for noninvasive biomedical research. *IEEE Sensors J.* 10(9), 1527–1534. (Cit. on p. 119).
- Casale, P., Altini, M., & Amft, O. (2015). Transfer learning in body sensor networks using ensembles of randomized trees. *IEEE Internet Things J.* 2(1), 33–40. (Cit. on p. 157).
- Casilari-Pérez, E. & García-Lagos, F. (2019). A comprehensive study on the use of artificial neural networks in wearable fall detection systems. *Expert Syst Appl.* 138, 112811. (Cit. on p. 157).
- Casilari, E., Santoyo-Ramón, J. A., & Cano-García, J. M. (2017a). Analysis of public datasets for wearable fall detection systems. *Sensors (Basel)*, 17(7). (Cit. on p. 157).
- Casilari, E., Santoyo-Ramón, J. A., & Cano-García, J. M. (2017b). UMAFall: A multisensor dataset for the research on automatic fall detection. *Procedia Comput Sci.* 110, 32–39. (Cit. on p. 157).
- Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Rep.* 100(2), 126–131. (Cit. on p. 33).
- Cerbulescu, C. & Cerbulescu, C. M. (2017). Proposed architecture to manage critical states predictions in IoT applications. *ACM International Conference Proceeding Series, Part F130953*. (Cit. on p. 157).
- Chadha, T. S., Watson, H., Birch, S., Jenouri, G. A., Schneider, A. W., Cohn, M. A., & Sackner, M. A. (1982). Validation of respiratory inductive plethysmography using different calibration procedures. *Am Rev Respir Dis.* 125(6), 644–649. (Cit. on pp. 5, 21, 66, 70–73, 75, 80, 85–88, 90–93, 96, 97, 170, 184, 185).
- Chang, C.-C. & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol.* 2(3), 27:1–27:27. (Cit. on pp. 53, 70).
- Chatzaki, C., Pediaditis, M., Vavoulas, G., & Tsiknakis, M. (2017). Human daily activity and fall recognition using a smartphone's acceleration sensor. *Communications in Computer and Information Science*, 736, 100–118. (Cit. on p. 157).

- Chen, Y., Guo, M., & Wang, Z. (2016). An improved algorithm for human activity recognition using wearable sensors. *In Proc: 8th International Conference on Advanced Computational Intelligence (ICACI)*, 248–252. (Cit. on p. 157).
- Chen, Y., Wang, L., & Ko, W. H. (1990). A piezopolymer finger pulse and breathing wave sensor. *Sens Actuators, A*, 23(1-3), 879–882. (Cit. on pp. 5, 15, 17).
- Chen, Y. & Wang, Z. (2017). A hierarchical method for human concurrent activity recognition using miniature inertial sensors. *Sensor Rev.* 37(1), 101–109. (Cit. on p. 157).
- Chow, P., Nagendra, G., Abisheganaden, J., & Wang, Y. T. (2000). Respiratory monitoring using an air-mattress system. *Physiol Meas.* 21(3), 345–354. (Cit. on p. 20).
- Cibis, T., Groh, B., Leutheuser, H., & Eskofier, B. M. (2015). Wearable real-time ECG monitoring with emergency alert system for scuba diving. *In Proc: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6074–6077. (Cit. on p. 10).
- Clark, C. C. T. (2017). *Profiling movement and gait quality in children's physical activity* (Doctoral dissertation, Swansea University). (Cit. on p. 157).
- Clark, C. C. T., Barnes, C. M., Stratton, G., McNarry, M. A., Mackintosh, K. A., & Summers, H. D. (2017). A review of emerging analytical techniques for objective physical activity measurement in humans. *Sports Med.* 47(3), 439–447. (Cit. on p. 157).
- Clark, C. C. T., Nobre, G. C., Fernandes, J. F. T., Moran, J., Drury, B., Mannini, A., Gronek, P., & Podstawski, R. (2018). Physical activity characterization: Does one site fit all? *Physiol Meas.* 39(9), 09TR02. (Cit. on p. 157).
- Cleland, I., Donnelly, M. P., Nugent, C., Hallberg, J., Espinilla, M., & Garcia-Constantino, M. (2018). Collection of a diverse, realistic and annotated dataset for wearable activity recognition. *In Proc: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 555–560. (Cit. on p. 157).
- Clifford, G. D. & Tarassenko, L. (2005). Quantifying errors in spectral estimates of HRV due to beat replacement and resampling. *IEEE Trans Biomed Eng.* 52(4), 630–638. (Cit. on pp. 106, 107).
- Cohen, C., Zigelbaum, G., Gross, D., Roussos, C., & Macklem, P. (1982). Clinical manifestations of inspiratory muscle fatigue. *Am J Med.* 73(3), 308–316. (Cit. on pp. 5, 15, 16).

- Coons, S. J., Gwaltney, C. J., Hays, R. D., Lundy, J. J., Sloan, J. A., Revicki, D. A., Lenderking, W. R., Cella, D., & Basch, E. (2009). Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (pro) measures: ISPOR ePRO good research practices task force report. *Value Health*, 12(4), 419–429. (Cit. on p. 34).
- Cosoli, G., Casacanditella, L., Pietroni, F., Calvaresi, A., Revel, G. M., & Scalise, L. (2015). A novel approach for features extraction in physiological signals. In *Proc: 10th Annual IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 380–385. (Cit. on pp. 29, 123, 136).
- Dash, S., Chon, K. H., Lu, S., & Raeder, E. A. (2009). Automatic real time detection of atrial fibrillation. *Ann Biomed Eng.* 37(9), 1701–1709. (Cit. on pp. 23, 25, 114).
- Daskalopoulou, C., Stubbs, B., Kralj, C., Koukounari, A., Prince, M., & Prina, A. M. (2017). Physical activity and healthy ageing: A systematic review and meta-analysis of longitudinal cohort studies. *Ageing Res Rev.* 38, 6–17. (Cit. on p. 6).
- Dauriac, B. (2018). *Contribution à la mise en oeuvre et l'évaluation de technologies embarquées pour l'appareillage de personnes amputées du membre inférieur* (Doctoral dissertation, Arts et Métiers ParisTech). (Cit. on p. 157).
- de la Cal, E., Fañez, M., Villar, J. R., & Suárez, V. M. (2018). Plataforma para el estudio de caídas y desvanecimientos en grupos de personas mayores. In *Proc: Cognitive Area Networks*, 5(1), 17–23. (Cit. on p. 157).
- de Groote, A., Paiva, M., & Verbandt, Y. (2001). Mathematical assessment of qualitative diagnostic calibration for respiratory inductance plethysmography. *J Appl Physiol.* 90(3), 1025–1030. (Cit. on pp. 21, 76).
- Di Marco, L. Y. & Chiari, L. (2011). A Wavelet-based ECG delineation algorithm for 32-bit integer online processing. *Biomed Eng Online*, 10–23. (Cit. on pp. 29, 30, 32, 121, 128, 134, 137).
- Dimitrievski, A., Zdravevski, E., Lameski, P., & Trajkovik, V. (2016). A survey of ambient assisted living systems: Challenges and opportunities. In *Proc: IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP)*, 49–53. (Cit. on p. 157).
- Duda, R., Hart, P., & Stork, D. (2000). *Pattern classification* (2nd). New York, NY: Wiley-Interscience. (Cit. on pp. 54–56, 59, 108, 114, 150, 154).
- Dumond, R., Gastinger, S., Rahman, H. A., Le Faucheur, A., Quinton, P., Kang, H., & Prioux, J. (2017). Estimation of respiratory volume from thoracoabdominal breathing distances: Comparison of two models

- of machine learning. *Eur J Appl Physiol.* 117(8), 1533–1555. (Cit. on pp. 5, 15, 17).
- Dumont, J., Hernández, A. I., & Carrault, G. (2010). Improving ECG beats delineation with an evolutionary optimization process. *IEEE Trans Biomed Eng.* 57(3), 607–615. (Cit. on pp. 29, 30, 128, 134, 137).
- Elgendi, M. (2013). Fast QRS detection with an optimized knowledge-based method: Evaluation on 11 standard ECG databases. *PLOS ONE*, 8(9), 73557. (Cit. on pp. 22, 29, 120).
- Elgendi, M., Eskofier, B. M., & Abbott, D. (2015). Fast T wave detection calibrated by clinical knowledge with annotation of P and T waves. *Sensors (Basel)*, 15(7), 17693–17714. (Cit. on pp. 29, 32, 33, 118, 121, 125, 126, 128, 133, 137, 172).
- Elgendi, M., Meo, M., & Abbott, D. (2016). A proof-of-concept study: Simple and effective detection of P and T waves in arrhythmic ECG signals. *Bioengineering (Basel)*, 3(4), E26. (Cit. on pp. 29, 32, 121).
- Ermes, M., Pärkkä, J., Mäntyjärvi, J., & Korhonen, I. (2008). Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *Trans Info Tech Biomed.* 12(1), 20–26. (Cit. on p. 144).
- Eskofier, B. M., Paulus, J., Paulsen, U., Burkart, M., Wullich, B., & Huppert, V. (2014). An ambulatory sensor-based system for quantification of nighttime micturition for accurate nocturia assessment. *In Proc: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 566–569. (Cit. on p. 157).
- Fekr, A. R., Radecka, K., & Zilic, Z. (2014). Tidal volume variability and respiration rate estimation using a wearable accelerometer sensor. *In Proc: 4th International Conference on Wireless Mobile Communication and Healthcare (MobiHealth)*, 1–6. (Cit. on pp. 5, 15, 16).
- Folke, M., Cernerud, L., Ekström, M., & Hök, B. (2003). Critical review of non-invasive respiratory monitoring in medical care. *Med Biol Eng Comput.* 41, 377–383. (Cit. on pp. 4, 5, 15, 16, 19).
- Friganovic, K., Kukulja, D., Jovic, A., Cifrek, M., & Krstacic, G. (2018). Optimizing the detection of characteristic waves in ECG based on processing methods combinations. *IEEE Access*, 6, 50609–50626. (Cit. on pp. 29, 128, 132).
- Full, K., Leutheuser, H., Schlessman, J., Armitage, R., & Eskofier, B. M. (2015). Comparative study on classifying gait with a single trunk-mounted inertial-magnetic measurement unit. *In Proc: IEEE-EMBS 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 1–6. (Cit. on pp. 11, 157).

- Fuster-Garcia, E., Bresó, A., Martínez-Miranda, J., Rosell-Ferrer, J., Matheson, C., & García-Gómez, J. M. (2015). Fusing actigraphy signals for outpatient monitoring. *Inform Fusion*, 23, 69–80. (Cit. on p. 157).
- Gabsteiger, F., Leutheuser, H., Reis, P., Lochmann, M., & Eskofier, B. M. (2014a). ICA-based reduction of electromyogenic artifacts in EEG data: Comparison with and without EMG data. *In Proc: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3861–3864. (Cit. on p. 12).
- Gabsteiger, F., Leutheuser, H., Reis, P., Lochmann, M., & Eskofier, B. M. (2014b). SVM for semi-automatic selection of ICA components of electromyogenic artifacts in EEG data. *In Proc: 15th International Conference on Biomedical Engineering (ICBME)*, 132–135. (Cit. on p. 12).
- Ghaffari, A., Homaeinezhad, M. R., Khazraee, M., & Daevaeiha, M. M. (2010). Segmentation of Holter ECG waves via analysis of a discrete Wavelet-derived multiple skewness-kurtosis based metric. *Ann Biomed Eng.* 38(4), 1497–1510. (Cit. on pp. 29, 31, 32, 121).
- Ghobadi, M., Sosnoff, J., Kesavadas, T., & Esfahani, E. T. (2015). Using mini minimum jerk model for human activity classification in home-based monitoring. *In Proc: IEEE 14th International Conference on Rehabilitation Robotics (ICORR)*, 2015-September, 909–912. (Cit. on p. 157).
- Ghodrati, A., Murray, B., & Marinello, S. (2008). RR interval analysis for detection of atrial fibrillation in ECG monitors. *In Proc: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 601–604. (Cit. on pp. 23, 25, 28).
- Gilbert, R., Auchincloss, J. H., Brodsky, J., & Boden, W. (1972). Changes in tidal volume, frequency, and ventilation induced by their measurement. *J Appl Physiol.* 33(2), 252–254. (Cit. on p. 5).
- Ginelli, D., Micucci, D., Mobilio, M., & Napoletano, P. (2018). UniMiB AAL: An Android sensor data acquisition and labeling suite. *Appl Sci. (Basel)*, 8(8). (Cit. on p. 157).
- Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23), e215–e220. (Cit. on pp. 28, 32, 55, 103, 117, 118, 126, 172).
- Google Scholar. (2019). Document search. Accessed on August 2nd, 2019. Retrieved from <https://scholar.google.com>. (Cit. on pp. 157, 158, 173)

- Gradl, S., Cibis, T., Lauber, J., Richer, R., Rybalko, R., Pfeiffer, N., Leutheuser, H., Wirth, M., von Tscharnier, V., & Eskofier, B. M. (2017). Wearable current-based ECG monitoring system with non-insulated electrodes for underwater application. *Appl Sci.* 7(12), 1277. (Cit. on p. 10).
- Gradl, S., Kugler, P., Lohmüller, C., & Eskofier, B. (2012). Real-time ECG monitoring and arrhythmia detection using Android-based mobile devices. *In Proc: 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2452–2455. (Cit. on pp. 6, 22).
- Gradl, S., Leutheuser, H., Elgendi, M., Lang, N., & Eskofier, B. M. (2015). Temporal correction of detected R-peaks in the ECG signals: A crucial step to improve QRS detection algorithms. *In Proc: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 522–525. (Cit. on pp. 10, 22, 104, 120, 121).
- Gradl, S., Leutheuser, H., Kugler, P., Biermann, T., Kreil, S., Kornhuber, J., Bergner, M., & Eskofier, B. M. (2013). Somnography using unobtrusive motion sensors and Android-based mobile phones. *In Proc: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. (Cit. on p. 12).
- Grossman, P., Wilhelm, F. H., & Brutsche, M. (2010). Accuracy of ventilatory measurement employing ambulatory inductive plethysmography during tasks of everyday life. *Biol Psychol.* 84(1), 121–128. (Cit. on p. 170).
- Guo, M. & Wang, Z. (2018). Segmentation and recognition of human motion sequences using wearable inertial sensors. *Multimed Tools Appl.* 77(16), 21201–21220. (Cit. on p. 157).
- Hamilton, P. S. & Tompkins, W. J. (1986). Quantitative investigation of the QRS detection rules using the MIT/BIH arrhythmia database. *IEEE Trans Biomed Eng. BME-33*(12), 1157–1165. (Cit. on p. 22).
- Heldberg, B. E., Kautz, T., Leutheuser, H., Hopfengärtner, R., Kasper, B., & Eskofier, B. M. (2015). Using wearable sensors for semiology-independent seizure detection - towards ambulatory monitoring of epilepsy. *In Proc: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5593–5596. (Cit. on pp. 12, 157).
- HelpAge International. (2016). Ageing and non-communicable diseases: Opportunities for the EU to respond. (Cit. on p. 3).
- Hermesen, S., Moons, J., Kerkhof, P., Wiekens, C., & De Groot, M. (2017). Determinants for sustained use of an activity tracker: Observational study. *JMIR Mhealth Uhealth*, 5(10), e164. (Cit. on p. 178).

- Hessen, H.-O. & Tessem, A. J. (2016). *Human activity recognition with two body-worn accelerometer sensors* (Master's thesis, Norwegian University of Science and Technology). (Cit. on p. 157).
- Heyde, C., Leutheuser, H., Eskofier, B. M., Roecker, K., & Gollhofer, A. (2014). Respiratory inductance plethysmography - A rationale for validity during exercise. *Med Sci Sports Exerc.* 46(3), 488–495. (Cit. on pp. 10, 69, 75, 96, 97).
- Hoaglin, D. & Welsch, R. (1978). The hat matrix in regression and ANOVA. *Am. Stat.* 32, 17–22. (Cit. on p. 83).
- Hollier, C. A., Harmer, A. R., Maxwell, L. J., Menadue, C., Willson, G. N., Black, D. A., & Piper, A. J. (2014). Validation of respiratory inductive plethysmography (LifeShirt) in obesity hypoventilation syndrome. *Respir Physiol Neurobiol.* 194(2014), 15–22. (Cit. on pp. 170, 171).
- Hu, X., Liu, J., Wang, J., Xiao, Z., & Yao, J. (2014). Automatic detection of onset and offset of QRS complexes independent of isoelectric segments. *Measurement*, 51(1), 53–62. (Cit. on pp. 29, 32, 122–124, 126, 136, 137, 167).
- Huang, C., Ye, S., Chen, H., Li, D., He, F., & Tu, Y. (2011). A novel method for detection of the transition between atrial fibrillation and sinus rhythm. *IEEE Trans Biomed Eng.* 58(4), 1113–1119. (Cit. on pp. 23, 25).
- Hunt, D. (2017). *Punctual activity classification within constrained sporting domains using reservoir computing* (Doctoral dissertation, Auckland University of Technology). (Cit. on p. 157).
- Huppert, V., Paulus, J., Paulsen, U., Burkart, M., Wullich, B., & Eskofier, B. (2016). Quantification of nighttime micturition with an ambulatory sensor-based system. *IEEE J Biomed Health Inform.* 20(3), 865–872. (Cit. on p. 157).
- Hur, T., Bang, J., Huynh-The, T., Lee, J., Kim, J.-I., & Lee, S. (2018). Iss2Image: A novel signal-encoding technique for CNN-based human activity recognition. *Sensors (Basel)*, 18(11), E3910. (Cit. on p. 157).
- İnanç, N., Kayri, M., & Erjtuğrul, Ö. F. (2018). Recognition of daily and sports activities. In *Proc: 2018 IEEE International Conference on Big Data (Big Data)*, 2216–2220. (Cit. on p. 157).
- International Data Corporation. (2013). Press Release: Smartphones Expected to Grow 32.7% in 2013 Fueled by Declining Prices and Strong Emerging Market Demand. Accessed on April 10, 2018. Retrieved from <http://www.idc.com/getdoc.jsp?containerId=prUS24143513>. (Cit. on p. 6)
- International Telecommunication Union. (2013). *Measuring the information society*. ITU. (Cit. on p. 6).

- Islam, M. S., Ammour, N., Alajlan, N., & Aboalsamh, H. (2016). Rhythm-based heartbeat duration normalization for atrial fibrillation detection. *Comput Biol Med.* 72, 160–169. (Cit. on pp. 23, 27, 104).
- Jané, R., Blasi, A., García, J., & Laguna, P. (1997). Evaluation of an automatic threshold based detector of waveform limits in Holter ECG with the QT database. *Computers in Cardiology*, 24, 295–298. (Cit. on pp. 29, 31, 32, 121, 128, 135, 136).
- Jensen, U. (2016). *Design considerations and application examples for embedded classification systems* (Doctoral dissertation, Friedrich-Alexander University Erlangen-Nürnberg (FAU)). (Cit. on p. 157).
- Jensen, U., Kugler, P., Ring, M., & Eskofier, B. M. (2015). Approaching the accuracy-cost conflict in embedded classification system design. *Pattern Anal Applic.* 19(3), 839–855. (Cit. on p. 115).
- Jensen, U., Leutheuser, H., Hofmann, S., Schuepferling, B., Suttner, G., Seiler, K., Kornhuber, J., & Eskofier, B. M. (2015). A wearable real-time activity tracker. *Biomed Eng Lett.* 5(2), 147–157. (Cit. on pp. 11, 37, 157).
- Johansson, A. & Hök, B. (2004). Sensors in medicine and health care. In P. A. Öberg, T. Togawa, & F. A. Spelman (Eds.), (Chap. Sensors for Respiratory Monitoring, pp. 161–186). Wiley-VCH Verlag GmbH & Co KGaA, Weinheim. (Cit. on pp. 4, 5, 15, 16).
- Jurca, R., Cioara, T., Anghel, I., Antal, M., Pop, C., & Moldovan, D. (2018). Activities of daily living classification using recurrent neural networks. *In Proc: 17th RoEduNet IEEE International Conference: Networking in Education and Research.* (Cit. on p. 157).
- Karantonis, D. M., Narayanan, M. R., Mathie, M., Lovell, N. H., & Celler, B. G. (2006). Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Trans Inf Technol Biomed.* 10(1), 156–167. (Cit. on pp. 43, 45, 150–153, 156, 158–160).
- Kautz, T. (2017). *Acquisition, filtering and analysis of positional and inertial data in sports* (Doctoral dissertation, Friedrich-Alexander University Erlangen-Nürnberg (FAU)). (Cit. on p. 157).
- Kautz, T., Eskofier, B. M., & Pasluosta, C. F. (2017). Generic performance measure for multiclass-classifiers. *Pattern Recognit.* 68, 111–125. (Cit. on p. 157).
- Kautz, T., Groh, B. H., & Eskofier, B. M. (2015). Sensor fusion for multi-player activity recognition in game sports. *In Proc: 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Workshop on Large-Scale Sports Analytics*, 1–4. (Cit. on p. 157).

- Kautz, T., Groh, B., Hannink, J., Jensen, U., Strubberg, H., & Eskofier, B. (2017). Activity recognition in beach volleyball using a deep convolutional neural network: Leveraging the potential of deep learning in sports. *Data Min Knowl Discov.* 31(6), 1678–1705. (Cit. on p. 157).
- Khandelwal, S. (2018). *Gait event detection in the real world* (Doctoral dissertation, Halmstad University). (Cit. on p. 157).
- Khandelwal, S. & Wickström, N. (2018). Novel methodology for estimating initial contact events from accelerometers positioned at different body locations. *Gait Posture*, 59, 278–285. (Cit. on p. 157).
- Khojasteh, S., Villar, J., Chira, C., González, V., & de la Cal, E. (2018). Improving fall detection using an on-wrist wearable accelerometer. *Sensors (Basel)*, 18(5). (Cit. on p. 157).
- Kikillus, N., Hammer, G., Lentz, N., Stockwald, F., & Bolz, A. (2007). Three different algorithms for identifying patients suffering from atrial fibrillation during atrial fibrillation free phases of the ECG. *Computers in Cardiology*, 34, 801–804. (Cit. on pp. 23, 24, 28).
- Knabner, P. & Barth, W. (2018). *Lineare Algebra Grundlagen und Anwendungen*. Springer-Verlag GmbH Deutschland. (Cit. on p. 50).
- Köhler, B. U., Hennig, C., & Orglmeister, R. (2002). The principles of software QRS detection. *IEEE Eng Med Biol Mag.* 21(1), 42–57. (Cit. on p. 22).
- Konno, K. & Mead, J. (1967). Measurement of the separate volume changes of rib cage and abdomen during breathing. *J Appl Physiol.* 22(3), 407–422. (Cit. on pp. 19, 20, 70, 80, 171).
- Koo, B. B., Drummond, C., Surovec, S., Johnson, N., Marvin, S. A., & Redline, S. (2011). Validation of a polyvinylidene fluoride impedance sensor for respiratory event classification during polysomnography. *J Clin Sleep Med.* 7(5), 479–485. (Cit. on p. 19).
- Koskimäki, H. & Siirtola, P. (2014). Recognizing gym exercises using acceleration data from wearable sensors. In *Proc: IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, 321–328. (Cit. on p. 157).
- Laguna, P., Jané, R., & Caminal, P. (1994). Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database. *Comput Biomed Res.* 27(1), 45–60. (Cit. on pp. 29, 31, 32, 121).
- Laguna, P., Mark, R. G., Goldberger, A. L., & Moody, G. (1997). A database for evaluation and algorithms for measurement of QT and other waveform intervals in the ECG. *Computers in Cardiology*, 24, 673–676. (Cit. on pp. 33, 117–119, 126, 128, 172).
- Laguna, P., Moody, G. B., & Mark, R. G. (1998). Power spectral density of unevenly sampled data by least-squares analysis: Performance and

- application to heart rate signals. *IEEE Trans Biomed Eng.* 45(6), 698–715. (Cit. on pp. 106, 107).
- Lake, D. E. & Moorman, J. R. (2011). Accurate estimation of entropy in very short physiological time series: The problem of atrial fibrillation detection in implanted ventricular devices. *Am J Physiol Heart Circ Physiol.* 300(1), H319–H325. (Cit. on pp. 23, 26, 28).
- Lamine, S. B., Calabrese, P., Perrault, H., Dinh, T. P., Eberhard, A., & Benchetrit, G. (2004). Individual differences in respiratory sinus arrhythmia. *Am J Physiol Heart Circ Physiol*, 286, H2305–H2313. (Cit. on pp. 5, 15, 16, 28, 106).
- Lang, N., Brischwein, M., Haßlmeyer, E., Tantinger, D., Feilner, S., Heinrich, A., Leutheuser, H., Gradl, S., Weigand, C., Eskofier, B., & Struck, M. (2015). Novel filter technique to improve R-peak detection for ECG data with motion artefacts from wearable systems. *In Proc: Computing in Cardiology (CinC)*. (Cit. on p. 10).
- Lasek, J. & Gagolewski, M. (2015). The winning solution to the AAILA'15 data mining competition: Tagging firefighter activities at a fire scene. *In Proc: Federated Conference on Computer Science and Information Systems (FedCSIS)*, 375–380. (Cit. on p. 157).
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. (Cit. on p. 40).
- Lee, I. M. & Paffenbarger, R. S. J. (2000). Associations of light, moderate, and vigorous intensity physical activity with longevity. The Harvard alumni health study. *Am J Epidemiol.* 151(3), 293–299. (Cit. on p. 7).
- Lee, J., Nam, Y., McManus, D. D., & Chon, K. H. (2013). Time-varying coherence function for atrial fibrillation detection. *IEEE Trans Biomed Eng.* 60(10), 2783–2793. (Cit. on pp. 23, 26, 28, 114).
- Lee, J., Reyes, B. A., McManus, D. D., Maitas, O., & Chon, K. H. (2013). Atrial fibrillation detection using an iPhone 4S. *IEEE Trans Biomed Eng.* 60(1), 203–206. (Cit. on pp. 23, 26, 28).
- Lemmens, R. J. M., Janssen-Potten, Y. J. M., Timmermans, A. A. A., Smeets, R. J. E. M., & Seelen, H. A. M. (2015). Recognizing complex upper extremity activities using body worn sensors. *PLoS ONE*, 10(3). (Cit. on p. 157).
- Lenis, G., Pilia, N., Oesterlein, T., Luik, A., Schmitt, C., & Dössel, O. (2016). P wave detection and delineation in the ECG based on the phase free stationary Wavelet transform and using intracardiac atrial electrograms as reference. *Biomed Eng Biomed Tech.* 61(1), 37–56. (Cit. on pp. 29, 30, 128, 133, 137).

- Leutheuser, H., Doelfel, S., Schuldhaus, D., Reinfelder, S., & Eskofier, B. M. (2014). Performance comparison of two step segmentation algorithms using different step activities. *In Proc: IEEE-EMBS 11th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 143–148. (Cit. on pp. 11, 157).
- Leutheuser, H. & Eskofier, B. M. (2012). Heart rate variability during physical exercise. *In Proc: 9. Symposium der Sektion Sportinformatik der Deutschen Vereinigung für Sportwissenschaft*, 12–15. (Cit. on pp. 11, 22).
- Leutheuser, H. & Eskofier, B. M. (2013). Heart rate variability during physical exercise. *Int J Comp Sci Sport*, 12(1), 61–68. (Cit. on pp. 11, 22).
- Leutheuser, H., Gabsteiger, F., Hebenstreit, F., Reis, P., Lochmann, M., & Eskofier, B. M. (2013). Comparison of the AMICA and the InfoMax algorithm for the reduction of electromyogenic artifacts in EEG data. *In Proc: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6804–6807. (Cit. on p. 12).
- Leutheuser, H., Gottschalk, T., Anneken, L., Struck, M., Heuberger, A., Arnold, M., Achenbach, S., & Eskofier, B. M. (2014). Automatic ECG arrhythmia detection in real-time on Android-based mobile devices. *In Proc: Conference on Mobile and Information Technologies in Medicine (MobileMed)*. (Cit. on p. 10).
- Leutheuser, H., Gradl, S., Anneken, L., Arnold, M., Lang, N., Achenbach, S., & Eskofier, B. M. (2016). Instantaneous P- and T-wave detection: Assessment of three ECG fiducial points detection algorithms. *In Proc: IEEE-EMBS 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 329–334. (Cit. on pp. 9, 117, 124, 137).
- Leutheuser, H., Gradl, S., Kugler, P., Anneken, L., Arnold, M., Achenbach, S., & Eskofier, B. M. (2014). Comparison of real-time classification systems for arrhythmia detection on Android-based mobile devices. *In Proc: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2690–2693. (Cit. on p. 10).
- Leutheuser, H., Heyde, C., Gollhofer, A., & Eskofier, B. M. (2014). Comparison of a priori calibration models for respiratory inductance plethysmography during running. *In Proc: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6393–6396. (Cit. on pp. 8, 65, 67, 76, 79, 86, 97).

- Leutheuser, H., Heyde, C., Roecker, K., Gollhofer, A., & Eskofier, B. M. (2017). Reference-free adjustment of respiratory inductance plethysmography for measurements during physical exercise. *IEEE Trans Biomed Eng.* 64(12), 2836–2846. (Cit. on pp. 8, 65, 77, 79, 105).
- Leutheuser, H., Jeleazcov, C., Schaffer, T., Weigand, C., Schüttler, J., & Hensel, B. (2011). Quantitative analysis of the cardiorespiratory system during paced respiration. In *Proc: 7th Russian-Bavarian Conference on Biomedical Engineering (RBC)*, 7, 9–10. (Cit. on pp. 5, 15, 16, 28, 106).
- Leutheuser, H., Lang, N., Gradl, S., Struck, M., Tobola, A., Hofmann, C., Ankenen, L., & Eskofier, B. M. (2017). Smart textiles. In S. Schneegass & O. Amft (Eds.), *Smart textiles: Fundamentals, design, and interaction* (Chap. Textile Integrated Wearable Technologies for Sports and Medical Applications, pp. 359–382). Springer International Publishing. (Cit. on pp. 10, 178).
- Leutheuser, H., Schuldhaus, D., & Eskofier, B. M. (2013). Hierarchical, multi-sensor based classification of daily life activities: Comparison with state-of-the-art algorithms using a benchmark dataset. *PLOS ONE*, 8(10), e75196. (Cit. on pp. 9, 143, 144, 146, 147, 157, 158, 173).
- Leutheuser, H., Tobola, A., Anneken, L., Arnold, M., Lang, N., Achenbach, S., & Eskofier, B. M. (2015). Arrhythmia classification using RR intervals: Improvement with sinusoidal regression feature. In *Proc: IEEE-EMBS 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 1–5. (Cit. on pp. 5, 9, 15, 16, 103, 107, 114).
- Li, C., Zheng, C., & Tai, C. (1995). Detection of ECG characteristic points using Wavelet transforms. *IEEE Trans Biomed Eng.* 42(1), 21–28. (Cit. on pp. 29, 32, 121).
- Li, Y., Tang, X., Wang, A., & Tang, H. (2017). Probability density distribution of delta RR intervals: A novel method for the detection of atrial fibrillation. *Australas Phys Eng Sci Med.* 40, 707–716. (Cit. on pp. 23, 27, 104).
- Lian, J., Wang, L., & Muessig, D. (2011). A simple method to detect atrial fibrillation using RR intervals. *J Am Coll Cardiol.* 107(10), 1494–1497. (Cit. on pp. 23, 26).
- Lin, C., Kail, G., Giremus, A., Mailhes, C., Tournieret, J.-Y., & Hlawatsch, F. (2014). Sequential beat-to-beat P and T wave delineation and waveform estimation in ECG signals: Block Gibbs sampler and marginalized particle filter. *Signal Processing*, 104, 174–187. (Cit. on pp. 29, 30, 32, 121, 128, 133, 137).

- Lin, C., Mailhes, C., & Tourneret, J.-Y. (2010). P- and T-wave delineation in ECG signals using a Bayesian approach and a partially collapsed Gibbs sampler. *IEEE Trans Biomed Eng.* 57(12), 2840–2849. (Cit. on pp. 29, 30).
- Liu, J., Sohn, J., & Kim, S. (2017). Classification of daily activities for the elderly using wearable sensors. *J Healthc Eng*, 2017. (Cit. on p. 157).
- Liu, S., Gao, R. X., & Freedson, P. S. (2012). Computational methods for estimating energy expenditure in human physical activities. *Med Sci Sports Exerc.* 44(11), 2138–2146. (Cit. on pp. 33, 35, 37).
- Liu, S., Gao, R. X., John, D., Staudenmayer, J. W., & Freedson, P. S. (2012). Multisensor data fusion for physical activity assessment. *IEEE Trans Biomed Eng.* 59(3), 687–696. (Cit. on pp. 39, 44, 45, 144, 151–153, 156, 158–161).
- Liu, S., Gao, R., He, Q., Staudenmayer, J., & Freedson, P. (2012). Improved regression models for ventilation estimation based on chest and abdomen movements. *Physiol Meas.* 33(1)(1), 79–93. (Cit. on p. 98).
- Logan, B. & Healey, J. (2005). Robust detection of atrial fibrillation for a long term telemonitoring system. *Computers in Cardiology*, 32, 619–622. (Cit. on pp. 23, 24).
- Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophys Space Sci.* 39, 447–462. (Cit. on pp. 106, 107).
- Lown, B. & Wolf, M. (1971). Approaches to sudden death from coronary heart disease. *Circulation*, 44(1), 130–142. (Cit. on p. 115).
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India*, 12, 49–55. (Cit. on p. 83).
- Malliotakis, T. (2017). *Human activity recognition using smartphones*. (Master's thesis, Technological Educational Institute of Crete). (Cit. on p. 157).
- Mamdani, E. (1977). Application of fuzzy logic to approximate reasoning using linguistic synthesis. *IEEE Trans Comput.* 26(12), 1182–1191. (Cit. on p. 43).
- Mangiarotti, M., Ferrise, F., Graziosi, S., Tamburrino, F., & Bordegoni, M. (2019). A wearable device to detect in real-time bimanual gestures of basketball players during training sessions. *J Comput Inf Sci Eng.* 19(1), 011004. (Cit. on p. 157).
- Mannini, A., Sabatini, A. M., & Intille, S. S. (2015). Accelerometry-based recognition of the placement sites of a wearable sensor. *Pervasive Mob Comput.* 21, 62–74. (Cit. on p. 174).

- Mannini, A. & Sabatini, M. (2010). Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, 10(2), 1154–1175. (Cit. on p. 7).
- Mäntyjärvi, J., Himberg, J., & Seppänen, T. (2001). Recognizing human motion with multiple acceleration sensors. In *Proc: IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 747–752. (Cit. on pp. 42, 45).
- Mariani, B., Hoskevec, C., Rochat, S., Büla, C., Penders, J., & Aminian, K. (2010). 3D gait assessment in young and elderly subjects using foot-worn inertial sensors. *J Biomech.* 43(15), 2999–3006. (Cit. on p. 144).
- Martindale, C. F., Sprager, S., & Eskofier, B. M. (2019). Hidden markov model-based smart annotation for benchmark cyclic activity recognition database using wearables. *Sensors (Basel)*, 19(8), E1820. (Cit. on p. 157).
- Martindale, C., Wirth, M., Schneegas, S., Zrenner, M., Groh, B., Blank, P., Schuldhaus, D., Kautz, T., & Eskofier, B. M. (2016). Workshop on wearables for sports. In *Proc: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 877–880. (Cit. on p. 157).
- Martínez, A., Alcaraz, R., & Rieta, J. J. (2010). Application of the phasor transform for automatic delineation of single-lead ECG fiducial points. *Physiol Meas.* 31(11), 1467–85. (Cit. on pp. 29, 32, 121, 128, 132–135, 137).
- Martínez, J. P., Almeida, R., Olmos, S., Rocha, A. P., & Laguna, P. (2004). A Wavelet-based ECG delineator: Evaluation on standard databases. *IEEE Trans Biomed Eng.* 51(4), 570–581. (Cit. on pp. 22, 29, 30, 32, 121, 122, 124–126, 128, 132–136, 167).
- Mazomenos, E. B., Biswas, D., Acharyya, A., Chen, T., Maharatna, K., Rosen Garten, J., Morgan, J., & Curzen, N. (2013). A low-complexity ECG feature extraction algorithm for mobile healthcare applications. *IEEE J Biomed Health Inform.* 17(3), 459–469. (Cit. on pp. 29, 30, 128, 133, 137).
- McCarthy, M. W., James, D. A., Lee, J. B., & Rowlands, D. D. (2015). Decision-tree-based human activity classification algorithm using single-channel foot-mounted gyroscope. *Electron Lett.* 51(9), 675–676. (Cit. on p. 157).
- McCarthy, M., James, D. A., Lee, J. B., Wada, T., & Rowlands, D. (2016). Effect of machine learning techniques upon wearable devices. In *Proc: Symposium on sports and human dynamics*. (Cit. on p. 157).

- McCool, F. D., Kelly, K. B., Loring, S., Greaves, I., & Mead, J. (1986). Estimates of ventilation from body surface measurements in unrestrained subjects. *J Appl Physiol*. 61(3), 1114–1119. (Cit. on p. 98).
- McManus, D. D., Lee, J., Maitas, O., Esa, N., Pidikiti, R., Carlucci, A., Harrington, J., Mick, E., & Chon, K. H. (2013). A novel application for the detection of an irregular pulse using an iPhone 4S in patients with atrial fibrillation. *Heart Rhythm*. 10(3), 3. (Cit. on pp. 23, 26, 28, 114, 172).
- Meagher, P. F., Jensen, R. E., Weil, M. H., & Shubin, H. (1966). Measurement of respiration rate from central venous pressure in the critically ill patient. *IEEE Trans Biomed Eng. BME-13*(2), 54–57. (Cit. on p. 20).
- Mehmood, A., Raza, A., Nadeen, A., & Saeed, U. (2016). Study of multi-classification of advanced daily life activities on shimmer sensor dataset. *International Journal of Communication Networks and Information Security*, 8(2), 86–92. (Cit. on p. 157).
- Mehra, R. & Strohl, K. P. (2009). Sleep disorders medicine - basic science, technical considerations, and clinical aspects. In S. Chokroverty (Ed.), (3rd, Chap. Evaluation and Monitoring of respiratory function, pp. 188–197). Elsevier Inc. (Cit. on pp. 5, 15, 16, 19).
- Merchán-Baeza, J. A., González-Sánchez, M., & Cuesta-Vargas, A. I. (2014). Reliability in the parameterization of the functional reach test in elderly stroke patients: A pilot study. *Biomed Res Int*. 2014. (Cit. on p. 157).
- Merchán-Baeza, J. A., González-Sánchez, M., & Cuesta-Vargas, A. I. (2015). Comparison of kinematic variables obtained by inertial sensors among stroke survivors and healthy older adults in the functional reach test: Cross-sectional study. *Biomed Eng Online*, 14(1). (Cit. on p. 157).
- Merry, K. J. (2017). *Differentiating common workplace postures through plantar pressure/ laying the groundwork for a low-cost instrumented insole* (Master's thesis, Simon Fraser University). (Cit. on p. 157).
- Micucci, D., Mobilio, M., & Napoletano, P. (2017). UniMiB SHAR: A dataset for human activity recognition using acceleration data from smart-phones. *Appl Sci. (Basel)*, 7(10). (Cit. on p. 157).
- Moldovan, D., Anghel, I., Cioara, T., Salomie, I., Chifu, V. R., & Pop, C. B. (2019). Kangaroo mob heuristic for optimizing features selection in learning the daily living activities of people with alzheimer's. In *Proc: 22nd International Conference on Control Systems and Computer Science (CSCS)*. (Cit. on p. 157).

- Moldovan, D., Antal, M., Pop, C., Olosutean, A., Cioara, T., Anghel, I., & Salomie, I. (2019). Spark-based classification algorithms for daily living activities. *Adv Intell Syst Comput.* 764, 69–78. (Cit. on p. 157).
- Moldovan, D., Olosutean, A., Chifu, V., Pop, C., Cioara, T., Anghel, I., & Salomie, I. (2018). Big data analytics for the daily living activities of the people with dementia. In *Proc: IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP)*, 175–181. (Cit. on p. 157).
- Moody, G. B. & Mark, R. G. (1983). A new method for detecting atrial fibrillation using R-R intervals. *Computers in Cardiology*, 227–230. (Cit. on pp. 23, 24, 28).
- Moody, G. & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE Eng Med Biol.* 20(3), 45–50. (Cit. on pp. 28, 33, 103, 117, 118, 126, 128, 172).
- Mullan, P. J., Kanzler, C. M., Lorch, B., Schröder, L., Winkler, L., Laich, L. H., Riedel, F., Richer, R., Luckner, C., Leutheuser, H., Eskofier, B. M., & Pasluosta, C. F. (2015). Unobtrusive heart rate estimation during physical exercise using photoplethysmographic and acceleration data. In *Proc: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6114–6117. (Cit. on p. 11).
- Munoz-Organero, M. & Lotfi, A. (2016). Human movement recognition based on the stochastic characterisation of acceleration data. *Sensors (Basel)*, 16(9). (Cit. on p. 157).
- Najafi, B., Aminian, K., Paraschiv-Ionescu, A., Loew, F., Büla, C. J., & Robert, P. (2003). Ambulatory system for human motion analysis using a kinematic sensor: Monitoring of daily physical activity in the elderly. *IEEE Trans Biomed Eng.* 50(6), 711–723. (Cit. on p. 144).
- Nazábal, A. (2017). *Markov modelling on human activity recogniti* (Doctoral dissertation, Universidad Carlos III de Madrid). (Cit. on p. 157).
- Nazábal, A., García-Moreno, P., Artes-Rodriguez, A., & Ghahramani, Z. (2016). Human activity recognition by combining a small number of classifiers. *IEEE J Biomed Health Inform.* 20(5), 1342–1351. (Cit. on p. 157).
- NCD Alliance. (2019). Making NCD prevention and control a priority, everywhere. Accessed on January 20th, 2019. Retrieved from <https://ncdalliance.org>. (Cit. on p. 3)
- Nguyen, N. D., Bui, D. T., Truong, P. H., & Jeong, G.-M. (2018). Position-based feature selection for body sensors regarding daily living activity recognition. *Journal of Sensors*, 2018(Article ID 9762098), 13 pages. (Cit. on p. 157).

- Nweke, H. F., Teh, Y. W., Alo, U. R., & Mujtaba, G. (2018). Analysis of multi-sensor fusion for mobile and wearable sensor based human activity recognition. *ACM International Conference Proceeding Series*, 22–26. (Cit. on p. 157).
- O'Connor, D. P., Mahar, M. T., Lauglihn, M. S., & Jackson, A. S. (2011). The Bland-Altman method should not be used in regression cross-validation studies. *Res Q Exerc Sport*. 82(4), 610–616. (Cit. on p. 85).
- O'Donoghue, P. (2010). *Research methods for sports performance analysis*. Routledge. (Cit. on p. 4).
- Oguma, Y., Sesso, H., Paffenbarger, R., & Lee, I. (2002). Physical activity and all cause mortality in women: A review of the evidence. *Br J Sports Med*. 36(3), 162–172. (Cit. on p. 7).
- Palumbo, F. (2016). *Ambient intelligence in assisted living environments* (Doctoral dissertation, Università degli Studi di Pisa). (Cit. on p. 157).
- Palumbo, F., Gallicchio, C., Pucci, R., & Micheli, A. (2016). Human activity recognition using multisensor data fusion based on reservoir computing. *J Ambient Intell Smart Environ*. 8(2), 87–107. (Cit. on p. 157).
- Pan, J. & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Trans Biomed Eng*. 32(3), 230–236. (Cit. on pp. 22, 29).
- Panigraphy, D. & Sahu, P. K. (2018). P and T wave detection and delineation of ECG signal using differential evolution (DE) optimization strategy. *Australas Phys Eng Sci Med*. 41(1), 225–241. (Cit. on pp. 29, 30, 32, 121, 128, 132, 137).
- Pannurat, N., Thiernjarus, S., Nantajeewarawat, E., & Anantavasilp, I. (2017). Analysis of optimal sensor positions for activity classification and application on a different data collection scenario. *Sensors (Basel)*, 17(4). (Cit. on p. 157).
- Pärkkä, J., Ermes, M., Korpipää, P., Mäntyjärvi, J., Peltola, J., & Korhonen, I. (2006). Activity classification using realistic data from wearable sensors. *IEEE Trans Inf Technol Biomed*. 10(1), 119–128. (Cit. on pp. 43, 45, 144, 151–153, 156, 158, 159, 161).
- Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., Buchner, D., Ettinger, W., Heath, G. W., King, A. C., Kriska, A., Leon, A. S., Marcus, B. H., Morris, J., Paffenbarger, R. S., Patrick, K., Pollock, M. L., Rippe, J. M., Sallis, J., & Wilmore, J. H. (1995). Physical activity and public health: A recommendation from the centers for disease control and prevention and the american college of sports medicine. *J Am Med Assoc*. 273(5), 402–407. (Cit. on p. 7).

- Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell.* 27(8), 1226–1238. (Cit. on p. 160).
- Petrénas, A., Marozas, V., & Sörnmo, L. (2015). Low-complexity detection of atrial fibrillation in continuous long-term monitoring. *Comput Biol Med.* 65, 1. (Cit. on pp. 23, 27).
- PhysioNet. (2018). PhysioBank Databases. Accessed on February 3rd 2018. Retrieved from <https://www.physionet.org/physiobank/database/>. (Cit. on p. 103)
- Pober, D. M., Staudenmayer, J., Raphael, C., & Freedson, P. S. (2006). Development of novel techniques to classify physical activity mode using accelerometers. *Med Sci Sports Exerc.* 38(9), 1626–1634. (Cit. on pp. 43, 45).
- Poole, K. A., Thompson, J. R., Hallinan, H. M., & Beardsmore, C. S. (2000). Respiratory inductance plethysmography in healthy infants: A comparison of three calibration methods. *Eur Respir J.* 16(6), 1084–1090. (Cit. on p. 171).
- Powell, K. E., Thompson, P. D., Caspersen, C. J., & Kendrick, J. S. (1987). Physical activity and the incidence of coronary heart disease. *Annu Rev Public Health.* 8, 253–287. (Cit. on p. 7).
- Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., & Howard, D. (2009). A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Trans Biomed Eng.* 56(3), 871–879. (Cit. on pp. 44, 45, 144, 149, 151–153, 156, 158, 159).
- Preece, S. J., Goulermas, J. Y., Kenney, L. P., Howard, D., Meijer, K., & Crompton, R. (2009). Activity identification using body-mounted sensors—a review of classification techniques. *Physiol Meas.* 30(4), R1–R33. (Cit. on pp. 33, 38, 39, 41).
- Pucci, R. (2014). *Phd thesis proposal: Evaluation and deployment of models for activity recognition* (Doctoral dissertation, Università di Pisa). (Cit. on p. 157).
- Rafferty, A. P., Reeves, M. J., McGee, H. B., & Pivarnik, J. M. (2002). Physical activity patterns among walkers and compliance with public health recommendations. *Med Sci Sports Exerc.* 34(8), 1255–1261. (Cit. on p. 7).
- Rahman, H. A. (2017). *Multi-sensor based activity recognition: Development and validation in real-life context* (Doctoral dissertation, École normale supérieure de Rennes). (Cit. on p. 157).
- Rangayyan, R. M. (2015). *Biomedical signal analysis* (I. Press, Ed.). Wiley. (Cit. on pp. 47, 49, 130, 131, 137).

- Raoufy, M. R., Hajizadeh, S., Gharibzadah, S., Mani, A. R., Eftekhari, P., & Masjedi, M. R. (2013). Nonlinear model for estimating respiratory volume based on thoracoabdominal breathing movements. *Respirology*, 18(1), 102–116. (Cit. on p. 171).
- Ravi, N., Dandekar, N., Mysore, P., & Littman, M. L. (2005). Activity recognition from accelerometer data. *In Proc: 17th Conference on Innovative Applications of Artificial Intelligence (IAAI)*, 1541–1546. (Cit. on pp. 42, 45, 144, 149, 151–153, 156, 158–161).
- Raza, M., Awais, M., Ellahi, W., Aslam, N., Nguyen, H. X., & Le-Minh, H. (2019). Diagnosis and monitoring of alzheimer's patients using classical and deep learning techniques. *Expert Syst Appl.* 136, 353–364. (Cit. on p. 157).
- Rincón, F., Recas, J., Khaled, N., & Atienza, D. (2011). Development and evaluation of multilead Wavelet-based ECG delineation algorithms for embedded wireless sensor nodes. *IEEE Trans Inf Technol Biomed.* 15(6), 854–863. (Cit. on pp. 29, 30, 125, 128, 134, 136, 137).
- Ring, M., Jensen, U., Kugler, P., & Eskofier, B. M. (2012). Software-based performance and complexity analysis for the design of embedded classification systems. *In Proc: 21st International Conference on Pattern Recognition (ICPR)*, 2266–2269. (Cit. on p. 115).
- Sackner, M. A., Watson, H., Belsito, A. S., Feinerman, D., Suarez, M., Gonzalez, G., Bizousky, F., & Krieger, B. (1989). Calibration of respiratory inductive plethysmography during natural breathing. *J Appl Physiol.* 66(1), 410–420. (Cit. on pp. 5, 21, 66, 71–76, 80, 86–88, 90–92, 94, 96–98, 165, 170, 183, 184, 186).
- Salarian, A., Russmann, H., Vingerhoets, F. J., Burkhard, P. R., & Aminian, K. (2007). Ambulatory monitoring of physical activities in patients with Parkinson's disease. *IEEE Trans Biomed Eng.* 54(12), 2296–2299. (Cit. on pp. 43, 45, 154).
- Sayadi, O. & Shamsollahi, M. B. (2009). A model-based bayesian framework for ECG beat segmentation. *Physiol Meas.* 30(3), 335–352. (Cit. on pp. 29, 31).
- Scargle, J. (1982). Studies in astronomical time series analysis. ii. Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys J*, 263, 835–853. (Cit. on pp. 106, 107).
- Schölkopf, B. & Smola, A. J. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press, Cambridge, MA, USA. (Cit. on pp. 50, 54, 70).
- Schuldhuis, D., Dorn, S., Leutheuser, H., Tallner, A., Klucken, J., & Eskofier, B. M. (2014). An adaptable inertial sensor fusion-based approach

- for energy expenditure estimation. In *Proc: 15th International Conference on Biomedical Engineering (ICBME)*, 124–127. (Cit. on pp. 12, 157).
- Schuldhuis, D., Kugler, P., Leible, M., Jensen, U., Schlarb, H., & Eskofier, B. M. (2012). Classification of surfaces and inclinations during outdoor running using shoe-mounted inertial sensors. In *Proc: 21st International Conference on Pattern Recognition (ICPR)*, 2258–2261. (Cit. on pp. 144, 154).
- Schuldhuis, D., Leutheuser, H., & Eskofier, B. M. (2013). Classification of daily life activities by decision level fusion of inertial sensor data. In *Proc: 8th International Conference on Body Area Networks (BodyNets)*, 77–82. (Cit. on p. 11).
- Schuldhuis, D., Leutheuser, H., & Eskofier, B. M. (2014). Towards big data for activity recognition: A novel database fusion strategy. In *Proc: 9th International Conference on Body Area Networks (BodyNets)*, 97–103. (Cit. on pp. 11, 157).
- Scopus. (2019). Document search. Accessed on August 2nd, 2019. Retrieved from <https://www.scopus.com/search/form.uri?zone=TopNavBar&origin=NO%20ORIGIN%20DEFINED&display=basic>. (Cit. on pp. 157, 158, 173)
- Scully, C. G., Lee, J., Meyer, J., Gorbach, A., Granquist-Fraser, D., Mendelson, Y., & Chon, K. H. (2012). Physiological parameter monitoring from optical recordings with a mobile phone. *IEEE Trans Biomed Eng.* 59 (2), 303–306. (Cit. on pp. 6, 172).
- Shephard, R. J. (2003). Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med.* 37(3), 197–206. (Cit. on p. 7).
- Shimmer Research. (2018). ECG user guide, revision 1.12. (Cit. on p. 50).
- Siegel, P. Z., Brackbill, R. M., & Heath, G. W. (1995). The epidemiology of walking for exercise: Implications for promoting activity among sedentary groups. *Am J Public Health.* 85(5), 706–710. (Cit. on p. 7).
- Smola, A. J. & Schölkopf, B. (2004). A tutorial on support vector regression. *Stat Comput.* 14(3), 199–222. (Cit. on pp. 50, 70).
- Snedecor, G. W. & Cochran, W. G. (1989). *Statistical methods*. Iowa State University Press, Ames, Iowa. (Cit. on pp. 71, 84).
- Soria-Olivas, E., Martínez-Sober, M., Calpe-Maravilla, J., Guerrero-Martínez, J. F., Chorro-Gascó, J., & Espí-López, J. (1998). Application of adaptive signal processing for determining the limits of P and T waves in an ECG. *IEEE Trans Biomed Eng.* 45(8), 1077–1080. (Cit. on p. 29).
- Stevens, J. (2009). *Applied multivariate statistics for the social sciences*. Taylor & Francis Group, LLC. (Cit. on pp. 83, 105).

- Stradling, J. R., Chadwick, G. A., Quirk, C., & Phillips, T. (1985). Respiratory inductance plethysmography: Calibration techniques, their validation and the effects of posture. *Bull Eur Physiopathol Respir*. 21(4), 317–324. (Cit. on p. 170).
- Strath, S. J., Kaminsky, L. A., Ainsworth, B. E., Ekelund, U., Freedson, P. S., Gary, R. A., Richardson, C. R., Smith, D. T., & Swartz, A. M. (2013). Guide to the assessment of physical activity: Clinical and research applications: A scientific statement from the american heart association. *Circulation*, 128(12), 2259–2279. (Cit. on pp. 33, 34).
- Sun, Y., Chan, K. L., & Krishnan, S. M. (2005). Characteristic wave detection in ECG signal using morphological transform. *BMC Cardiovasc Disord*. (Cit. on pp. 29, 32, 121).
- Taborri, J., Rossi, S., Palermo, E., Patanè, F., & Cappa, P. (2014). A novel HMM distributed classifier for the detection of gait phases by means of a wearable inertial sensor network. *Sensors (Basel)*, 14(9), 16212–16234. (Cit. on p. 157).
- Taddei, A., Distanto, G., Emdin, M., Pisani, P., Moody, G. B., Zeelenberg, C., & Marchesi, C. (1992). The European ST-T Database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *Eur Heart J*. 13, 1164–1172. (Cit. on pp. 103, 104).
- Tantinger, D., Zrenner, M., Lang, N., Leutheuser, H., Eskofier, B., Weigand, C., & Struck, M. (2015). Human authentication implemented for mobile applications based on ECG-data acquired from sensorized garments. *In Proc: Computing in Cardiology (CinC)*, 417–420. (Cit. on p. 11).
- Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation*. 93, 1043–1065. (Cit. on pp. 22, 114).
- Tateno, K. & Glass, L. (2001). Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and Δ RR intervals. *Med Biol Eng Comput*. 39, 664–671. (Cit. on pp. 23, 24).
- Teramoto, S., Fukuchi, Y., Nagase, T., Matsuse, T., & Orimo, H. (1995). A comparison of ventilation components in young and elderly men during exercise. *J Gerontol A Biol Sci Med Sci*. 50A(1), B34–B39. (Cit. on p. 97).
- Theodoridis, S. & Koutroumbas, K. (2008). *Pattern recognition* (4th). San Diego, CA: Academic Press. (Cit. on pp. 56, 108, 150, 154).
- Tobola, A., Ershov, S., Wissendheit, U., & Couronné, R. (2009). European Patent No. EP 2 174 590 B1. Vorrichtung und Verfahren zum Erfassen

- einer Atmung eines Lebewesens. Europäisches Patentamt. (Cit. on pp. 5, 15, 17).
- Tobola, A., Espig, C., Streit, F. J., Korpok, O., Leutheuser, H., Schmitz, B., Hofmann, C., Struck, M., Weigand, C., Eskofier, B. M., & Fischer, G. (2015). Scalable ECG hardware and algorithms for extended runtime of wearable sensors. *In Proc: 10th Annual IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 255–260. (Cit. on pp. 12, 157).
- Tobola, A., Korpok, O., Leutheuser, H., Schmitz, B., Hofmann, C., Struck, M., Weigand, C., Eskofier, B. M., Heuberger, A., & Fischer, G. (2014). System design impacts on battery runtime of wearable medical sensors. *In Proc: Conference on Mobile and Information Technologies in Medicine (MobileMed)*, 1–4. (Cit. on p. 12).
- Tobola, A., Leutheuser, H., Pollak, M., Spies, P., Hofmann, C., Weigand, C., Eskofier, B., & Fischer, G. (2018). Self-powered multiparameter health sensor. *IEEE J Biomed Health Inform.* 22(1), 15–22. (Cit. on p. 12).
- Tobola, A., Leutheuser, H., Schmitz, B., Hofmann, C., Struck, M., Weigand, C., Eskofier, B. M., & Fischer, G. (2016). Battery runtime optimization toolbox for wearable biomedical sensors. *In Proc: IEEE-EMBS 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 199–204. (Cit. on p. 12).
- Tobola, A., Streit, F., Espig, C., Korpok, O., Leutheuser, H., Sauter, C., Lang, N., Schmitz, B., Hofmann, C., Struck, M., Weigand, C., Eskofier, B. M., & G., F. (2015). Sampling rate impact on energy consumption of biomedical signal processing systems. *In Proc: IEEE-EMBS 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 1–6. (Cit. on pp. 12, 157).
- Tobola, A. & Weigand, C. (2007). German Patent No. DE 10 2007 053 843 A1 2009.05.20. Kleidungsstück zur Erfassung einer Atembewegung. Bundesrepublik Deutschland Deutsches Patent- und Markenamt. (Cit. on pp. 5, 15, 18).
- Trahanias, P. & Skordalakis, E. (1990). Syntactic pattern recognition of the ECG. *IEEE Trans Pattern Anal Mach Intell.* 12(7), 648–657. (Cit. on pp. 29, 31).
- Troiano, R. P., McClain, J. J., Brychta, R. J., & Chen, K. Y. (2014). Evolution of accelerometer methods for physical activity research. *Br J Sports Med.* 48(13), 1019–1023. (Cit. on p. 33).
- Trost, S. G., McIver, K. L., & Pate, R. R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc.* 37 (supplement 11), 531–543. (Cit. on p. 144).

- Trost, S. G. & O'Neil, M. (2014). Clinical use of objective measures of physical activity. *Br J Sports Med.* 48(3), 178–181. (Cit. on pp. 35, 36, 38).
- Tsipouras, M. G., Fotiadis, D., & Sideris, D. (2005). An arrhythmia classification system based on the RR-interval signal. *Artif Intell Med.* 33(3), 237–250. (Cit. on pp. 23, 24, 28, 114).
- Urwyler, P., Rampa, L., Stucki, R., Büchler, M., Müri, R., Mosimann, U., & Nef, T. (2015). Recognition of activities of daily living in healthy subjects using two ad-hoc classifiers. *Biomed Eng Online*, 14(1). (Cit. on p. 157).
- VanderPlas, J. T. (2018). Understanding the Lomb-Scargle periodogram. *Astrophys J Suppl Ser.* 236(1), 16. (Cit. on pp. 106, 107).
- Vavoulas, G., Chatzaki, C., Malliotakis, T., Pediaditis, M., & Tsiknakis, M. (2016). The MobiAct dataset: Recognition of activities of daily living using smartphones. In *Proc: 2nd International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE)*, 143–151. (Cit. on p. 157).
- Vila, J. A., Gang, Y., Rodriguez Presedo, J. M., Fernández-Delgado, M., Barro, S., & Malik, M. (2000). A new approach for TU complex characterization. *IEEE Trans Biomed Eng.* 47(6), 764–472. (Cit. on pp. 29, 31, 128, 134).
- Wang, Z. & Chen, Y. (2017). Recognizing human concurrent activities using wearable sensors: A statistical modeling approach based on parallel HMM. *Sensor Rev.* 37(3), 330–337. (Cit. on p. 157).
- Warburton, D. E. R., Nicol, C. W., & Bredin, S. S. D. (2006). Health benefits of physical activity: The evidence. *CMAJ*, 174(6), 801–809. (Cit. on p. 7).
- Ward, D. S., Evenson, K. R., Vaughn, A., Rodgers, A. B., & Troiano, R. P. (2005). Accelerometer use in physical activity: Best practices and research recommendations. *Med Sci Sports Exerc.* 37 (supplement 11), 582–588. (Cit. on p. 144).
- Warren, J. M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N., & Vanhees, L. (2010). Assessment of physical activity - a review of methodologies with reference to epidemiological research: A report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *Eur J Cardiov Prev R.* 17(2), 127–139. (Cit. on pp. 33, 34, 173).
- Wasza, J., Fischer, P., Leutheuser, H., Oefner, T., Bert, C., Maier, A., & Hornegger, J. (2016). Real-time respiratory motion analysis using 4-D shape priors. *IEEE Trans Biomed Eng.* 63(3), 485–495. (Cit. on pp. 10, 20).

Bibliography

- Weisberg, S. (2005). *Applied linear regression*. John Wiley & Sons, Inc. (Cit. on p. 50).
- White, J. R. (2018). *Using topic models to detect behaviour patterns for healthcare monitoring* (Doctoral dissertation, University of Reading). (Cit. on p. 157).
- Wilhelm, F. H., Roth, W. T., & Sackner, M. A. (2003). The LifeShirt. An advanced system for ambulatory measurement of respiratory and cardiac function. *Behav Modif.* 27(5), 671–691. (Cit. on pp. 5, 15, 18, 20).
- World Health Organization. (n.d.). Health topics: Physical activity. Accessed on April 10, 2018. Retrieved from http://www.who.int/topics/physical_activity/en/. (Cit. on pp. 6, 7)
- World Health Organization. (2008). The global burden of disease: 2004 update. (Cit. on p. 4).
- World Health Organization. (2016). Global health estimates 2015: Deaths by cause, age, sex, by country and by region, 2000-2015. Accessed on April 10, 2018. Retrieved from http://www.who.int/healthinfo/global_burden_disease/. (Cit. on p. 4)
- World Health Organization. (2017). The top 10 causes of death [Fact sheet]. Accessed on April 10, 2018. Retrieved from <http://www.who.int/mediacentre/factsheets/fs310/en/>. (Cit. on pp. 3, 5)
- World Health Organization. (2018). Physical activity [Fact sheet]. Accessed on April 10, 2018. Retrieved from <http://www.who.int/mediacentre/factsheets/fs385/en/>. (Cit. on pp. 6, 7)
- Wundersitz, D. (2015). *Accelerometer validity to measure and classify movement in team sports* (Doctoral dissertation, Deakin Univeristy). (Cit. on p. 157).
- Wundersitz, D. W. T., Gastin, P. B., Robertson, S., Davey, P. C., & Netto, K. J. (2015). Validation of a trunk-mounted accelerometer to measure peak impacts during team sport movements. *Int J Sports Med.* 36(9), 742–746. (Cit. on p. 157).
- Wundersitz, D. W. T., Josman, C., Gupta, R., Netto, K. J., Gastin, P. B., & Robertson, S. (2015). Classification of team sport activities using a single wearable tracking device. *J Biomech.* 48(15), 3975–3981. (Cit. on p. 157).
- Zdravevski, E. (2017). *Parallelization of algorithms for data processing and classification* (Doctoral dissertation, Ss. Cyril and Methodius University of Skopje). (Cit. on p. 157).
- Zdravevski, E., Lameski, P., Trajkovik, V., Kulakov, A., Chorbev, I., Goleva, R., Pombo, N., & Garcia, N. (2017). Improving activity recognition

- accuracy in ambient-assisted living systems by automated feature engineering. *IEEE Access*, 5, 5262–5280. (Cit. on p. 157).
- Zdravevski, E., Stojkoska, B., Standl, M., & Schulz, H. (2017). Automatic machine-learning based identification of jogging periods from accelerometer measurements of adolescents under field conditions. *PLoS ONE*, 12(9). (Cit. on p. 157).
- Zheng, P. Z., Muaremi, A., Leutheuser, H., Norden, J., Sinha, A., Eskofier, B. M., & Smuck, M. (2017). Poster 109: Discriminating physical performance phenotypes of patients with chronic low back pain. *In Proc: Annual Assembly of the American Academy of Physical Medicine and Rehabilitation (PM&R)*, 9(9 Supplement 1), S169. (Cit. on p. 11).
- Zhou, X., Ding, H., Ung, B., Pickwell-MacPherson, E., & Zhang, Y. (2014). Automatic online detection of atrial fibrillation based on symbolic dynamics and Shannon entropy. *Biomed Eng Online*, 13, 18. (Cit. on pp. 23, 27, 114).

Non-communicable diseases are the leading cause of death and disability worldwide. Nowadays, wearable computing is an integral part of everyday life. Everyone is surrounded increasingly with wearables, like smartphones and smartwatches. The combination of addressing non-communicable diseases with the help of wearables seems one potential solution. In this thesis, methods and algorithms for three wearable computing applications in eHealth are presented: I) mobile breathing analysis, II) mobile electrocardiogram (ECG) analysis, and III) inertial measurement unit (IMU)-based activity recognition.

In the mobile breathing analysis part, accurate and unobtrusive breathing measurements in the field are enabled using the RIP device and appropriate reference-free adjustment algorithms. In the mobile ECG analysis part, two approaches were presented. The RR arrhythmia classification could facilitate the early detection of arrhythmic beats. The P- and T-wave detection improved the detection of fiducial points in the ECG signal. In the IMU-based activity recognition part, the DaLiAc dataset and a hierarchical classification system were proposed to address one of the leading risk factors for non-communicable diseases - physical inactivity.

ISBN 978-3-96147-260-4

