Diss. ETH No. 25371

Monitoring outpatients in palliative care through wearable devices

A thesis submitted to attain the degree of

Doctor of Sciences of ETH Zurich

(Dr. sc. ETH Zurich)

presented by

Vanessa Christina Klaas

Dipl.-Inform, LMU München, Germany
born on 21.04.1979
citizen of Germany

accepted on the recommendation of

Prof. Dr. Gerhard Tröster, examiner
Prof. Dr. med. Matthias Guckenberger, co-examiner
Prof. Dr. Klaas Enno Stephan, co-examiner

2018

Vanessa Christina Klaas
Monitoring outpatients in palliative care through wearable devices
Diss. ETH No. 25371

First edition 2018
Published by ETH Zürich, Switzerland

Printed by Reprozentrale ETH

Acknowledgements

Especially, I would like to thank Prof. Gerhard Tröster for giving me the opportunity to work on this research project in his group. Throughout my research, I very much appreciated his guidance on topics off all kind. I am grateful for pushing me out of my comfort zone which accelerated my personal development as much as the deepening of my knowledge. His intense supervision and the excellent atmosphere in his group contributed much to the realisation of this project and success of this work.

For bringing the palliative care study into being as principal investigator, fruitful discussions during the project and co-examining my thesis, I would like to thank Prof. Matthias Guckenberger.

Also for co-examining my work and taking time for support, I would like to thank Prof. Klaas Enno Stephan.

The studies would not have been possible without the engagement of my collaborators Dr. Matea Pavic, Dr. Gudrun Theile and Johannes Kraft. For the reliable assistance during data collection, I would like to thank especially Debora Mittner. Likewise I thank Sonja Schwenne for her support during data collection. For collaboration and discussion in the fatigue study, I would like to thank Prof. Josef Jenewein, Dr. Natalie Büel-Drabe and Prof. Heinrich Walt.

Not only for great collaboration, but also for many intense cross-atlantic late evening skype calls, I would like to thank sincerely Dr. Annina Seiler.

I would like to express my deepest gratitude and compassion to all patients and their relatives for supporting my research through their participation and for very special and valuable moments.

I would like to thank my past and current lab members, especially Dr. Michael Hardegger for supporting the app implementation and discussing my research plan, Dr. Amir Muaremi for support during project initiation and later discussions, Dr. Alberto Calatroni and Dr. David Bannach for letting me experiment with the Bonsai Sensor, for many fruitful discussions and for being always open to answer my questions. For providing an iPhone app and discussions, I would like to thank Dr. Ulf Blanke. Likewise, I thank Stefan Knobelspiess and Alwin Daus as best office mates, Matjia Varga, Dr. Sebastian Feese, Tobias Grosshauser, Dr. Christian Vogt, Andreas Mehman, Matthias Flückiger and Amir Sarabadani Tafreshi, Lars Büthe, Rolf Adelsberger, Daniel

Waltisberg, Zack Zhou, Sinziana Mazilu, Long-Van Nguyen-Dinh and Dr. Giovanni Salvatore. For administrative support, I would like to thank especially Ruth Zähringer and Paul Holz.

For fast and flexible solutions for the IT infrastructure, I would like thank Simon Alder.

For fruitful discussion, exchange of experience and reflection, I would like to thank Gertrud Lindner, Dr. Sara Karlina-Barber, Dr. Marie Francine Lagadec and Dr. Franca Contratto.

For the weekly BodyPump class and PhD related discussions, I would like to thank Sonja Peteranderl, Katja Gfeller and Joram Schito.

For the great infrastructure and great view over the lake, I would like to thank ETH Zürich.

For valuable work during their master and semester thesis, I would like to thank my students Simon Laube, Andreas Brüniger, Guillermo Ramos and Jagruti Patel.

For proof-reading and being great friends, I would like to thank Natalie Savaricas and Agnieszka Bidas.

I would like to thank deeply my coaches Pia Neubauer without whom I would not be at this place right now and Tulia Lopes for support in the preparation of my defense as well as Dr. Shivani Sood and Dr. Marco Righetti for taking care of me.

Finally, I would like to thank deeply my mother Heidi and sisters Dani and Annika for always being there and for their understanding and support, my father Dieter for believing in me and liking also my rejected papers, my best friends Babsi and Artemis for deep discussions and their unconditional friendship and especially I thank my girlfriend and love of my live Lucie who showed up in a tough moment, supported me and who is making my life brighter every day.

Zürich, July 2018 Vanessa Christina Klaas

Contents

Abstract

Patients in palliative care suffer from a life-threatening disease. Holistic treatment includes control of symptoms (e. g., pain, nausea, sleeplessness) as well as psychosocial and spiritual help which is also extended to the relatives of a patient. For advanced cancer patients in palliative care, a crucial phase is the transition from palliative care in the hospital to the home setting, where care around the clock is not guaranteed any more, leads to an increased number of unplanned hospital re-admissions and emergency visits. Physicians aim to fill this care gap by monitoring physical and social activities as well as vital signs. Daily monitoring data, provided to caregivers, could enable caregivers to timely intervene when symptoms of a patient deteriorate.

Besides patients in palliative care, also cancer survivors suffering from cancer-related fatigue could benefit from activity monitoring. Up to now, the remedies and effective treatments for cancer-related fatigue are limited. Research still has to unveil the underlying mechanisms that lead to a state of chronic exhaustedness. Measures that help healthy people like regenerative sleep show no or little effect in fatigued patients. Besides psycho-stimulants that come with the risk of addiction, cognitive behavioural therapy and moderate physical exercise have been shown to be effective. However, research still has to investigate timing, frequency and intensity of physical activity and researchers need a better understanding how the fatigue evolves during the day and in long-term.

This thesis investigates the possibilities and limitations of activity monitoring using wearable devices such as smartphones and an arm-worn devices that is capable of measuring vital signs such as heart rate. Three studies involving cancer patients are conducted:

- An interview study including 12 cancer patients enabled a patient-centric design for an Android activity monitoring app for smartphones.
- Only using the smartphone as monitoring device, a study with 7 cancer survivors suffering from cancer-related fatigue was conducted as a pre-study in order to gain first experiences and to explore the possible knowledge gain about cancer-related fatigue through activity monitoring.

- During a planned study period of 12 weeks per patient, 30 patients in ambulatory palliative care were wearing a smartphone and the arm-worn sensor as monitoring devices. The age range of the study participants was 39 to 85 years. In weekly interviews, patients were asked about their experiences with the devices and their quality of life. The aim of the study was to evaluate feasibility and acceptance of activity monitoring in this patient group. Furthermore, exploratory data analysis investigated the possibilities and limitations of unsupervised methods on this real-world data set.

The two data sets, collected during the fatigue study and during the palliative care study, were pre-processed including cleaning steps, classification and clustering methods to add higher level information such as visited locations (anonymized). From these prepared data sets, features were extracted such as number of places visited per day. On the resulting datasets of features, statistical methods were applied to explore relations between sensor data, self-reports and, in case of the palliative care study, emergency visits to the hospital. For the latter analysis, patients who experienced an emergency room visit and those who did not were compared by means of hypothesis testing. For each feature, the underlying alternative hypothesis was that the change of a feature between the first week of study participation at home and the week before an emergency visit (or the last week of study participation for the patients without an emergency visit), differs in the two patient groups. The rate of change was defined by the ratio of the medians of the two weeks. Changes of three features, namely resting heart rate, resting heart rate variability and step speed were identified to have significant group differences:

- The resting heart rate had an increasing trend in the group with emergency visits (median=1.01, interquartile range [0.96, 1.12]) and a decreasing trend in the group without an emergency visit (median=0.9, interquartile range [0.89, 0.99]) with a nominal significance of p=.021 and a medium effect size r=.46.

- The resting heart rate variability had a decreasing trend in the group with emergency visits (mean=0.81, standard deviation=0.14) and an increasing trend in the group without an emergency visit(mean=1.17, standard deviation=0.46) with a nominal significance of p=.011 and a large effect size r=.53.

- The step speed had an increasing trend in the group with emergency visits (median=1.1, interquartile range [1.08, 1.13]) and a decreasing trend in the group without an emergency visit (median=0.99, interquartile range [0.96, 1.04]) with a nominal significance of p=.003 and a large effect size r=.61.

In contrast, hypothesis testing for features based on patients' subjective self-reports for pain, distres and global quality of life did not reveil any significant differences. Hence, activity monitoring of vital signs and physical activity outperformed patients' self-reports. However, a power analysis based on the three nominally significant results would recommend an independent study with 84 patients to confirm the results of this study.

Furthermore, a set of recommendations for future research was concluded from the experiences gained through conducting these studies.

Zusammenfassung

Patienten in Palliativbetreuung leiden an einer lebensbedrohlichen Krankheit. Die Behandlung folgt einem holistischem Ansatz, der Symptommanagement (z.B. von Schmerzen, Übelkeit, Schlaflosigkeit usw.) sowie psycho-soziale und spirituelle Hilfe, die auch die Angehörigen des Patienten einschliesst, umfasst. Für Krebspatienten im fortgeschrittenen Stadium ist die Übergangsphase nach der Entlassung aus dem Spital nach Hause besonders kritisch, da im häuslichen Umfeld die rund-um-die Uhr Betreuung aus der Klinik nicht mehr gewährleistet ist. Diese Umstellung führt zu vermehrten ungeplanten Wiederaufnahmen und ambulanten Besuchen im Notfall. Ärzte möchten diese Versorgungslücke durch Überwachung von physischer und sozialer Aktivität sowie von Vitalparametern der Patienten füllen: täglich bereitgestellte Monitoringdaten könnten es Betreuern ermöglichen, zur richtigen Zeit einzugreifen, wenn sich die Symptome eines Patienten verschlechtern.

Neben Patienten in Palliativpflege könnten auch Krebsüberlebende, die an sogenannter tumorassoziierter Müdigkeit (Fatigue) leiden, von Aktivitätsmonitoring auf folgende Weise profitieren: bis jetzt sind die Mittel und effektive Behandlungsmethoden gegen krebsassoziierte Müdigkeit eingeschränkt. Die Forschung muss noch die zugrundeliegenden Mechanismen, die zu dieser Art von chronischer Erschöpfung führen, erklären. Massnahmen wie beispielsweise erholsamer Schlaf, die gesunden Menschen gegen Müdigkeit helfen, zeigen keinen oder nur geringen Effekt bei chronisch müden Patienten. Neben Psychopharmaka, die abhängig machen können, wurde für kognitive Verhaltenstherapie und modarates physisches Training Effektivität gegen Fatigue durch Studien gezeigt. Jedoch muss die Forschung noch eingehender untersuchen, wann, in welcher Häufigkeit und mit welcher Intensität optimalerweise trainiert werden sollte. Dazu ist es auch notwendig, besser zu verstehen, wie die Fatigue sich im Verlaufe eines Tages und über einen längeren Zeitraum hinweg entwickelt.

Diese Arbeit untersucht die Möglichkeiten und Grenzen von Aktivitätsmonitoring mithilfe von tragbaren Geräten wie Smartphones und ein am Oberarm getragener Sensor namens Everion®. Drei Studien wurden mit Krebspatienten durchgeführt:

- Eine Interviewstudie mit 12 Krebspatienten hat ein patientengerechtes Design einer Aktivitätsmonitoring-App für Android

Smartphones ermöglicht.

- In einer Vorstudie wurden 7 Krebsüberlebende, die an tumorassoziierter Fatigue leiden, nur mit einem Smartphone als Monitoringgerät ausgestattet. Diese Vorstudie diente dem Sammeln erster Erfahrungen und der Erforschung, welcher Informationsgewinn bzgl. tumorassoziierter Fatigue durch Aktivitätsmonitoring möglich ist.

- Während einer geplanten Studiendauer von 12 Wochen pro Patient trugen 30 Krebspatienten in ambulanter Palliativepflege ein Smartphone und den Everion® als Monitoringgeräte. In wöchentlichen Interviews wurden sie bzgl. ihrer Erfahrungen mit den Geräten und ihrer Lebensqualität befragt. Das Ziel der Studie war es, die Machbarkeit und Akzeptanz von Aktivitätsmonitoring in dieser Patientengruppe in einem Alter bis zu 85 Jahren zu erforschen. Ausserdem wurde eine explorative Datenanalyse durchgeführt und die Möglichkeiten und Grenzen von unüberwachten Methoden auf diesem Datensatz aus der "wirklichen Welt" untersucht.

Die beiden Datensätze, die während der Fatigue-Studie und der Palliativ-Studie erhoben wurden, wurden aufbereitet, was Datensäuberungsschritte, Klassifizierungen, Clusteringmethoden zum Anreichern mit aus den Sensordaten gewonnener Informationen wie Anzahl besuchter Orte pro Tag und die Extrahierung von Merkmalen beinhaltete. Auf die daraus gewonnenen Datensätze, bestehend aus den extrahierten Merkmalen, wurden statistische Methoden angewandt, um Zusammenhänge zwischen den Sensordaten, Selbsteinschätzungen zu Symptomen und, im Falle der Palliativ-Studie, ambulante und stationäre Notfallbesuche, aufzuzeigen. Als Ergebnis konnten drei Merkmale, nämlich der Ruhepuls, Herzratenvariabilität in Ruhe und die vom Smartphone gemessene Schrittgeschwindigkeit identifiziert werden. Diese drei Parameter zeigten signifikante Unterschiede zwischen den beiden Patientengruppen:

- Der Ruhepuls zeigte einen steigenden Trend in der Gruppe mit Notfallbesuchen (median=1.01, Interquartilsabstand [0.96, 1.12]) und einen fallenden Trend in der Gruppe ohne Notfallbesuch (median=0.9, Interquartilsabstand [0.89, 0.99]) mit einer nominellen Signifikanz von p=.021 und einer mittleren Effektgrösse r=.46.

- Die Herzratenvariabilität in Ruhe zeigte einen fallenden Trend in der Gruppe mit Notfallbesuchen (mean=0.81, Standardabweichung=0.14) und einen steigenden Trend in der Gruppe ohne Notfallbesuche (mean=1.17, Standardabweichung=0.46) mit einer nominellen Signifikanz von p=.011 und einer grossen Effektgrösse von r=.53.

- Die Schrittgeschwindigkeit zeigte einen leicht steigenden Trend in der Gruppe mit Notfallbesuchen (median=1.1, Interquartilsabstand [1.08, 1.13]) und einen leicht fallenden Trend in der Gruppe ohne Notfallbesuche (median=0.99, Interquartilsabstand [0.96, 1.04]) mit einer nominellen Signifikanz von p=.003 und einer grossen Effektgrösse r=.61.

Im Gegensatz dazu, Hypothesentests für Merkmale basierend auf den Selbstberichten von Patienten für Schmerz, Belastung und globaler Gesundheitsstatus haben keine signifikanten Unterschiede zwischen den beiden Gruppen aufgedeckt. Eine Poweranalyse basierend auf den drei nominell signifikanten Ergebnissen empfiehlt eine unabhängige Studie mit 84 Patienten, um die Ergebnisse dieser Studie zu bestätigen.

Empfehlungen für zukünftige Forschung wurden aus den Erfahrungen, die in den während dieser Arbeit durchgeführten Studien gewonnen wurden, gesschlossen.

1

Introduction

The introduction motivates observational studies to examine the feasibility and usefulness of monitoring severely ill patients by means of wearables.

1.1 Background, Motivation and Aim

1.1.1 Palliative Care

Background. The World Health Organisation's (WHO) 2014 cancer report estimated in 2012 that 8.7 million people over 15 years old had been diagnosed with a cancer disease the year before, 22 million in the previous three years and more than 32 million people in the previous five. Furthermore, the WHO estimates that over 19 million adults are in need of palliative care (PC) at the end of life with 34 % having cancer [1]. Other life-threatening diseases are cardiovascular diseases and chronic obstructive pulmonary disease.

Unlike curative care, palliative care does not aim to cure a disease, but instead to provide the best possible quality of life by following a holistic approach: control of symptoms (e.g., pain, nausea, sleeplessness) as well as psychosocial and spiritual help which is also extended to the relatives of a patient [2].

Motivation. Whereas hospitals, hospices and nursing homes provide PC around the clock, patients usually don't have this 24/7 professional support in the home environment. The transition from hospital to home is an especially crucial phase for patients and their relatives, leading to a relevant number of unplanned re-hospitalisations and emergency visits [3]. Patients possibly have to adjust to symptoms (e.g. pain) or a changed body schema (e. g., loss of weight) resulting from the progress of the disease or stressful therapy. In addition, they have to face an extended comprehension (understanding) of their disease. At the same time, with the discharge from hospital they lose the support of an interdisciplinary professional team that has been available around the clock. Re-hospitalisation through emergency visits happens also due to suffering from symptoms that could have been effectively treated in the ambulatory setting – if detected in time. Delgado-Guay et al. give changes in the mental status, dyspnea, fever, infections, etc. as main reasons for admittance to the emergency department. They rated 23 % of emergency visits by patients with advanced cancer receiving palliative care as potentially avoidable with constipation and running out of pain medications as main causes [4].

Basch et al. showed that symptom monitoring during chemotherapy treatment by the means of electronic patient-reported outcomes

improve health-related quality of life, reduces emergency visits and hospitalisations and prolongs survival [5, 6].

However, Pakhomov et al. found disagreement between patient-reported symptoms and their documentation in medical records [7]. Hence, this thesis examined if more reliable symptom-related information can be collected through continuous monitoring of physical activity and of vital signs. The feasibility and information gain are examined in this work and results are presented in Chapter 5.

1.1.2 Cancer related fatigue

Background. Cancer survivors have to deal with the side effects of the disease even after having defeated it. One of the most prevalent problems is cancer related fatigue (CRF): almost all cancer patients suffer from CRF during treatment [8] and about 25 % of cancer survivors have to deal with CRF after successful treatment for up to 10 years [9]: the exhausting treatment and stressful situation of the severe disease are finished, yet CRF patients still feel constantly exhausted on a physical and psychological level. Furthermore, usual behaviours like regenerative sleeping do not improve the tiredness of CRF patients.

The etiology of CRF is complex and includes physiological causes related to cancer treatment and medication as well as behavioural factors such as fear, depression or insomnia and reduced physical activity [9].

A moderate to severe CRF reduces physical and mental capabilities. This can lead to an inability to work, impairment of social life, daily activities and hobbies, and can also negatively impact relationships. Overall, this huge negative interference with daily life reduces quality of life [10, 11]. During FATIGUE2 study [10], 376 patients were interviewed about the physical, emotional, behavioural and economical impact of fatigue after being treated with chemotherapy. 90 % of the interviewed patients suffered from fatigue (n = 301) and reported that CRF prevents them from leading a normal life. Curt concludes that CRF reduces quality of life (QoL) more than other symptoms [10].

Motivation. Psychological interventions and psycho-stimulant drugs have been used as therapies although evidence of efficacy is still low for both categories of intervention. both categories of intervention, evidence is still low. Drugs have negative side effects and a risk of addiction, leading to increased research for alternatives. Studies have

shown evidence for cognitive behavioural therapy, mindfulness-based meditation [12], physical interventions including moderate physical exercise, walking or yoga [13]. However, more research is required to reveal deeper knowledge and understanding of the underlying mechanisms of the disease. More detailed and individualised recommendations are needed on when, how often, how much and what type of exercise is most effective in the context of other diseases as well, such as depression [14].

After a systematic review of studies investigating CRF, Prue et al. concluded that studies should have a longitudinal design including frequent assessments of the fatigue and a healthy control group [15]. Kelley et al. confirm Prue with their conclusion that "additional well-designed randomised controlled trials and meta-analyses appear warranted" [16].

This work examines how activity monitoring and digital questionnaires provided by the smartphone (that is used by patients anyway) could be applied in longitudinal studies, realising a fatigue assessment with a frequency of several times per day. The study goal is to gain information concerning the course of fatigue, patients' behaviour and physical activity. Results are presented in chapter 4.

1.2 Goals of this work

From the motivation of section 1.1, the following questions are derived:

- How should a monitoring system look like in order to reduce non-elective re-admissions to hospital for outpatients in palliative care?
- Is activity monitoring accepted by patients that have to deal with a severe disease and face their end of life?
- Which sensor modalities are useful in order to predict non-elective re-admissions to hospital?

In order to answer these research questions, this work investigated the usage of two wearable devices that are mobile and can be worn on the body (not hand-held) in the setting of ambulatory palliative care.

Cancer related fatigue. As part of a pre-study with reduced setup and shorter study period, 8 patients suffering from CRF received a

smartphone with a pre-installed activity monitoring app. Beside collecting data from various phone sensors as described in chapter 2, patients had to fill out digital questionnaires several times per day. Study aims were to

- evaluate the acceptance of activity tracking with smartphones in this patient group
- test the developed monitoring system as described in chapter 2 and test the design of digital questionnaires
- evaluate the information gained from the collected data

Palliative Care. During an observational study including 31 patients (one patient passed away few days after discharge to home) in ambulatory palliative care, the study participants received an arm-mounted sensor and a smartphone with a pre-installed activity monitoring app. The study participants had to fill out digital questionnaires once a day. Study aims were to

- evaluate the feasibility of remotely monitoring palliative patients' vital signs, physical and social activity using the arm-mounted sensor and a smartphone
- evaluate the developed monitoring system as described in chapter 2
- analyse behavioural pattern in patients with deteriorating symptoms

1.3 Wearable technology for measuring vital signs and physical activity

Wearable devices are devices that can be worn on the body without the using hands to carry them. They come in the form of accessories like bracelets, anklets, armbands, contact lenses, necklaces, glasses, rings and watches, but also as part of clothes, e. g., shirts, shoes and jackets [17] and the smartphone worn in a pocket or on a belt.

1.3.1 Photoplethysmography

Photoplethysmography (optical method to measure heart rate) (PPG) is used to measure pulse-related parameters such as heart rate. [18] gives an overview over the measurement principle, the current technology and clinical applications. The technique is implemented in devices such as pulse oximeters and smartwatches. The measurement principle is based on light transmission or reflection: measurement devices are equipped with LEDs that send light of different wavelengths into the human body. The blood flow influences the transmitted or reflected light which is received by a photodiode. From this measurement, several vital signs such as heart rate (HR), respiration rate (RR), energy expenditure (EE), peripheral capillary oxygen saturation (SpO_2), etc. can be derived. Measurement quality is limited by tattoos at the measurement spot and artifacts introduced by movement. Other crucial factors that influence the measurement quality are contact to skin and blood perfusion. The performance of wrist-worn sensors is better during resting and worsens with increasing impact with a deviation in heart rate measurement below 10 % [19]. Georgiou et al. stated that wearable PPG-based measurements of heart rate variability (HRV) can be an alternative to standard techniques for resting or mild exercise conditions.

1.3.2 Accelerometer

Smartphones are equipped with a tri-axis accelerometer which is used for many smartphone functions such as turning on the screen when grabbing the phone as well as for activity recognition, e. g., counting steps.

1.3.3 GPS

Smartphones normally are equipped with a GPS module for localisation purposes. The currently used sensors are basic modules with lower accuracy than more advanced modules. Smartphones use also the information of Wi-Fi (Wi-Fi) signals and provider network information to improve localisation accuracy leading to errors up to 6 % for distance measuring [20]. When outdoors in nature, the deviance is on average 10 m [21]. However, the accuracy of Global Positioning System (GPS) measurement in cities or while being indoors decreases due to the surroundings that block the direct sight to GPS satellites [22].

1.3.4 Wi-Fi

Wi-Fi is the commonly used name for the IEEE 802.11 standard for wireless networking. An access point (AP) is a device providing a Wi-Fi network. Client devices like smartphones are scanning regularly for APs. Smartphones use the scan results to improve GPS based localisation. Furthermore, it is possible to record the receive signal strength indicator (RSSI). The RSSI depends on the signal propagation. Different material such as metal, brick walls, etc., attenuate differently the signal sent by APs which leads to characteristics of, e. g., rooms, so-called fingerprints. Wi-Fi fingerprinting is a common technique for indoor-localisation. The smartphones used in this work detected signals with a minimum RSSI of −103 dBm.

1.4 State of the art

State of the art used throughout this thesis comprises several fields. The first section focuses on solutions for remote monitoring as well as the application of information and communication technology (ICT) for disease and symptoms management in several medical fields. The second section gives an overview over the usage of wearables for health care and well-being. The following sections focus on the extraction of higher-level information out of raw sensor signals, where more complex methods are needed. The last section gives an overview over currently available methods for prediction.

1.4.1 Remote monitoring and ICT for disease management

Cardiovascular diseases

Especially with the treatment by means of implantable devices, e. g., pacemaker and implantable cardioverter defibrillators (ICDs), remote monitoring is already well established in the field of cardiovascular diseases. Pacemakers and ICDs collect vital sign measurements, information about occurred heart rhythm anomalies and the device status during their regular functioning. On a daily basis, they send the collected data to manufacturers who provide summaries to health carers. These summaries include among others early indicators for a deterioration of the disease, e. g., worsening of the heart function. Device manufacturers routinely perform clinical trials to show the efficacy of new devices and device functionalities, including monitoring services.

[23, 24, 25, 26] showed reduced all-cause mortality and hospitalisation for heart failure patients with remote monitoring.

In addition to implant based-solutions, research has also investigated feasibility and efficacy of non-invasive monitoring systems. Common approaches are based on structured telephone support, regular manual self-assessments transmitted via ICT or measurement devices. These devices measure body signs such as weight, blood pressure, heart rate and heart rhythm. The collected data is automatically transmitted to a telemonitoring centre either by means of a built-in modem or through a mobile phone that is connected via Bluetooth [27, 28]. Bashi et al. give an overview of systematic reviews. Typical study outcomes are, e. g., mortality (all-cause and specific), hospitalisation and quality of life. In general, studies agree on the positive impact of monitoring and show significant outcomes improvement. Bashi et al. conclude that physiological monitoring should be part of the daily care for HF patients [29].

Mental and neurological diseases

Giunti et al. give an overview over commercially available smartphone applications for multiple sclerosis (Multiple Sclerosis (MS)) patients [30]. They state that although MS is one of the dominant neurological diseases, the available apps are underrepresented in comparison to diseases like cancer, diabetes or HIV.

Mental diseases

For depression, various studies have evaluated the efficacy of interventions, e. g., based on cognitive behavioural therapy and mindfulness meditation [31, 32]. Firth et al. conclude that smartphones are promising tools for managing depression and supporting interventions like cognitive behavioural therapy (cognitive behavioural therapy (CBT)).

Palliative care and oncology

Although remote monitoring is already established and has been proven to be effective in reducing re-hospitalisations and mortality in several fields as described above in sections 1.4.1, 1.4.1, 1.4.1, there are not any established solutions for palliative care patients. Various research on ICT systems has already shown some evidence on usage of web-based and mobile phone applications, e. g., to support symptoms

management, to improve contact between patients and care givers over distance [33, 34] or intervention (e.g., nutrition, social network, physical exercises) apps to improve quality of life [35]. Kawsar et al. investigate feasibility and challenges for the digitalisation of the paper-based questionnaires ESAS for breast cancer patients in Bangladesh. They focus on research with rural populations in developing countries. They concluded that the digital questionnaire on a smartphone has advantages (e. g., reducing long, exhausting commuting to physicians) and plan to extend their research on automatic physical tracking using smartphones [36].

[37] reviews actual research concerning web-based tools for self-management in patients with cancer related fatigue. Studies evaluating interventions for fatigued cancer survivors showed small to moderate effects to improve quality of life and depression [38]. The usage of mobile technology for collection of patient subjective reported outcome is already established in oncology [39, 40, 41].

Andebe et al. studied the usefulness of a mobile phone assessment in 15 palliative patients and caregivers in an African population by performing qualitative interviews after an usage period of 6 weeks [34]. Andebe et al. concluded that the mobile phone assessment tool improved communication between patients and caregivers as well as symptoms management. However, the authors did not list detailed inclusion criteria concerning patient characteristics and none of the work included in this literature research involves wearable sensing technologies.

1.4.2 Wearables for well-being, healthy behaviour and disease monitoring and management

Wearable technology

Smartphone apps for a healthier lifestyle and wrist-worn fitness trackers as well as smartwatches are already a commercial hype with over 100 million fitness tracker devices sold in 2016 [42]. However, the measurement accuracy of these devices is shown to be low, underestimating the heart rate (HR) up to 30 bpm (beats per minute), especially during exercising [42, 19]. An overview of the wrist-worn fitness trackers that were available for consumers in 2017, evaluates the monitoring of heart rate, energy expenditure (EE) and physical activity (step count). Bunn et al. confirm the tendency to underestimate the measured variables [43]. The authors summarise, that in general, the devices underes-

timates the heart rate and the energy expenditure, although perform better during sedentary activities and worse during exercising. In addition, the errors augment when the arm wearing the device increases its movement. The reason behind this observation lies in the measurement method: increased movement alters the blood flow inducing artifacts that hinder accurate measurements. In contrast, the devices perform better in step counting with increased speeds (3.22 km/h to 8.05 km/h) and underestimate the step count with speeds < 3.22 km/h. The researchers give as possible reasons instability and wrong positioning at the wrist. Gillinov et al. evaluated the accuracy of a chest strap, one monitor worn at the fore-arm and four wrist-worn heart rate monitors from different brands against a standard ECG measurement in healthy subjects while exercising. They conclude that electrode-based monitors (e. g., chest strap) outperform optical sensors and the accuracy of optical sensors depend on the type of activity a user is doing [44].

[45, 46, 47, 48] present reviews on monitoring systems for health care, activity recognition and stress detection. The accuracy and granularity in the activity recognition depend on the number and placement of sensors, which in return affects the system obtrusiveness to users [49].

However, when being used for research, wearables have to fulfil specific requirements, e. g., full access on collected raw signals and full control over data accessibility – depending on national laws concerning data security, privacy and clinical trials that have to been approved by a national or local ethical committee. A market research revealed only three available unobtrusive devices: Empatica E4 [1], the Angel Sensor that was developed by a startup and whose development was shut down by the end of 2016[2] and the Biovotion Everion® that is CE certified as a medical device in compliance with directive EC 765/2008 and also benefits from FDA approval (Food and Drug Administration[3]) as a class 2 device.

[1] https://store.empatica.com/products/e4-wristband?variant=39588207747

[2] http://www.mobihealthnews.com/content/open-source-wearable-angel-shuts-down

[3] federal agency of the United States Department of Health and Human Services that is responsible to approve food, medication, medical devices etc. for the U.S. market

Relationship between wearable monitoring and symptoms and diseases

In various studies, activities were shown to be related to pain level [50] and mental health [51].

The group of Tanzeem Choudhury investigated the usage of the smartphone itself to assess the user's well-being. [52, 53, 54] propose systems that range from monitoring and visualising the user's behaviour in terms of physical and social activity as well as sleep hygiene to systems providing interventions like timely suggestions of alternative healthy behaviours for the app user.

Research on the management of mental diseases with the support of wearables focuses on schizophrenia [55, 56, 57], depression [58], bipolar disorder [59, 60, 61] and stress [62, 63]. Also the usage of wearables in the field of neurological diseases such as Multiple Sclerosis [64, 65] and Parkinson [66] is investigated. Smartphone-based symptom assessment for Parkinson was evaluated in [67] showing high correlations with several standardised disease scales (r = .72 up to r = .91, all <= 0.002).

Timmerman et al. examined the relationship between cancer related fatigue and daily physical activity using acceleration sensors. The study showed no significant difference between the fatigued group and a control group, but a significant decrease in the activity from morning to evening as well as a positive correlation between the self-reported fatigue level and the magnitude of the decrease from afternoon to evening. Timmerman et al. concluded that monitoring reveals patterns of fatigue and physical activity and therefore could be used to improve the treatment of fatigue [68].

1.4.3 Heart rate variability analysis

Medical significance of heart rate variability

HRV was extensively examined since 1965 [69] leading to standardised nomenclature and measurement methods in 1996 [70]. Researchers have shown that the HRV provides information about the state of the autonomic nervous system [71], blood pressure, renal function failure, diabetes and other diseases [72]. [73] gives an overview on HRV in epilepsy and confirms the significance of several HRV features for epilepsy. Frequency analysis focusing on low frequency band (LF) and high frequency band (HF) frequency bands were shown to relate

with stress. Lower vagal and higher sympathetic tone estimated by LF/HF ratio have been shown to predict morbidity and mortality in cardiovascular diseases [74]. A relation between HRV features in the frequency domain and (experimentally induced) pain was shown in healthy subjects [72]. However, recent research also shows that HRV needs to be studied more [75, 76], e. g., including high mental stress and low physical stress situations, in order to develop HRV metrics that discriminate better between mental and physical stress and are related with sympathetic and vagal activities.

Significance of HRV in cancer patients

Studies examining HRV in cancer patients and advanced cancer patients concluded that HRV predicts overall survival [77, 78]. [79] examines the HRV response to opioid treatment against breakthrough pain and identifies a frequency domain HRV feature as potential indicator for pain. Study results in predicting the admission outcome are contradictory and often based on small sample sizes [80]. However, the behaviour of HRV in advanced cancer patients is still relatively unknown. In contrast to a healthy population, there does not exist any normal ranges of HRV metrics for cancer patients.

Methods for HRV analysis

Besides research on the application and clinical meaning of HRV, authors also developed methods for HRV analysis, e. g., signal preprocessing and effects of bad data quality. The effects of missing data on time and frequency domain HRV analysis were studied in depth by K. Kim et al.[81, 82]. Several authors also proposed methods to detect and correct ectopic beats as an essential part of the signal preprocessing [83]. However, the studies do not consider a continuous measurement of HRV in an ambulatory setting.

1.4.4 Clustering of GPS data

The GPS signal of smartphones is used by various applications such as navigation apps and provides useful information in various studies such as examining recreational movement [21]. The studies described in this thesis investigate behavioural patterns such as leaving home, daily covered distance or number of visited places. Therefore, a set of GPS points has to be partitioned into groups that represent meaningful

places, so-called points of interest (Point of interest (POI)). Not every location is also a place, e. g., if you are stuck in traffic at a highly frequented crossing close to a beer garden for an hour, this sample of 20 GPS points (in case of requesting a GPS sample every 3 min) builds a cluster, but not a meaningful place in the context of the studies of this thesis, whereas if you go for a beer and actually sit in the beer garden for an hour, the sample of 20 GPS points that is just a hundreds of meters away from the traffic jam cluster, could be a meaningful place. Data collected during the studies of this work does not contain any ground truth concerning patient location. Hence, the following paragraph summarises the state of the art of unsupervised methods.

Extracting places from GPS data in an unsupervised setting is based on clustering the recorded set of GPS points. In case of spatio-temporal data, a data sample contains also a timestamp, in addition to the location information. A special case of spatio-temporal data are trajectories that are recorded by moving instances. [84, 85] give overviews on the commonly used approaches for clustering spatio-temporal data. Concerning trajectories, there are two questions of interest. Firstly, for a given trajectory, extract semantic information such as stops and moves (intra-trajectory clustering). Secondly, for a given set of trajectories, identify groups of similar trajectories (inter-trajectory clustering) [86].

In general, clustering approaches comprise distance-based methods, e. g., k-means (k-means), and density-based clustering methods among others. The most commonly used density-based method is DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [87]. The advantage of DBSCAN over k-means is that one does not need to specify the number of clusters. However, temporal-spatial datasets usually do not have a uniform density. The algorithm OPTICS (Ordering Points To Identify the Clustering Structure) addresses this specific characteristic of varying densities by finding clusters with varying density [88]. Birant et al. extended DBSCAN to st-DBSCAN (spatio-temporal DBSCAN) by adding a new arbitrary dimension, e. g., the day and night average temperature measured at a location. Similar points then have to be close enough in both dimensions [89]. Finally, they also request that points are "temporal neighbours", i. e., the samples are measured in "consecutive time units" [89]. For example, two samples with the same location and with same 2^{nd} dimension measured at, e. g., at 9 o'clock in the morning on two different days would be part of two different clusters if the threshold for the time difference is a day. However, this is not useful for the application in this thesis,

where, e. g., the patient's home should be one place independent of the time.

Trajectories pose the additional challenge of connecting different stays (a place where a person is staying for some amount of time). Proposed methods to extract places out of trajectories are, e. g., SMoT (Stops and Moves of Trajectories) [90], CB-SMoT (a clustering-based algorithm to identify stops and moves of trajectories) [91] and DB-SMoT (Direction-Based Stops and Moves of Trajectories) [92]. However, SMoT is based on providing additional information such as maps and public POI. Therefore it is not suitable for applications with high privacy concerns. CB-SMoT does not consider the possibility of revisiting the same places: the home in the morning is classified as a different stop than (the same) home in the evening.

1.4.5 Wi-Fi clustering

A map containing RSSI fingerprints is the basis for the commonly used method to localise a device based on the RSSIs of detected APs. Without this information, unsupervised clustering is still possible in order to detect localisation changes at room level. However, aside from walls, fluids, and hence the human body can influence Wi-Fi signals. Therefore, the quality of a localisation method lies in its robustness against such influences.

Research focuses on indoor-localisation with a known Wi-Fi environment. The two most common approaches are based on tri-lateration of RSSI values or on fingerprinting techniques. The first approach uses the signal strengths of APs measured by the device that shall be located. Using tri-lateration, the position can be estimated. This method is deployable in public buildings, where the infrastructure is well known. However, the method's sensitivity to noise can introduce errors of more than 20 m [93]. The second approach is based on creating Wi-Fi maps, i. e., a database containing RSSI fingerprints. The position of a device is then estimated by comparing the current fingerprint with the database. This approach needs an initial effort of creating the map and also demands maintenance in terms of updating the map since even small environmental changes, e. g., adding fire doors to segment large corridors, alternate Wi-Fi characteristics significantly [93]. Additional limitations in accuracy are due to the physics of signal propagation, i. e., the human body can decrease the signal strength by around 10 dBm

[94]. The variety of the built-in Wi-Fi receivers further induce inaccuracies [95].

Once there is a reliable map available, Kang et al. suggest to treat a dataset of locations as a trajectory by involving also the time component in order to improve the extraction of places. They suggest an algorithm consisting of two steps: first, they use a mere distance threshold between points that are adjacent in time in order to decide if they belong to the same cluster or not. A cluster is considered as a place if its size in terms of time exceeds a specified threshold. Before adding a cluster to the list of places, their algorithm checks – again by comparing its distance to existing places with a threshold – if the cluster is added as a new place or merged with an existing place [96].

However, there is only little research for unsupervised scenarios. Lau et al. propose DCCLA (Density-based Clustering Combined Localisation Algorithm) to cluster Wi-Fi scans from smartphones for a room-level localisation [97]. The approach is split in two phases. Initially, DCCLA generates a database of fingerprints for rooms by combining clusters of the same time frame using DBSCAN. They consider a place as meaningful if a person stays for longer than 10 min. After this learning period, DCCLA identifies the places based on the stored fingerprints when a person is re-visiting. Xu et al. investigate in depth DCCLA and state that the performance increases with a higher amount of access points to up to 100 % accuracy [98]. However, during their experiment, the smartphones were always placed on a table and therefore with a fixed position in the room. Furthermore, they do not provide an unsupervised solution for detecting new rooms: after the initial creation of the fingerprint database, the algorithm needs user input for rooms that are not yet in the database. Nevertheless, their method could be extended by novelty detection or by using a distance threshold up to which a point is considered to be sufficiently close to a fingerprint.

1.4.6 Voice analysis

In this work, voice is analysed using voice features, not the content of conversations. Research on this topic focused so far on artificially produced datasets, e. g., EMO-DB [99] with German speech and RML [100, 101] with multiple languages. The wizard-of-oz dataset is closest to natural emotions because the study participants did not know that they were recorded [102]. A review over emotional speech databases is

given in [103] and over approaches is given in [104]. Usually, artificial databases are created by recording actors saying defined utterances playing different emotions. The quality of the recordings is verified by having an audience rate the emotions. Recordings with insufficient consistency are discarded. Typical emotions are, e. g.of EMO-DB, neutral, anger, fear, joy, sadness, disgust and boredom. These emotions can be ordered in two dimensions, namely arousal and valence. Some approaches aim to classify only according to this two-dimensional scheme.

Emotion analysis can be done online, e. g., EmoVoice [105, 106], or offline, e. g., [107]. Both approaches are evaluated on artificial databases. However, Vogt et al., the creators of EmoVoice, evaluated commonly used feature sets for "acted and spontaneous speech". They concluded that features differ for acted and natural emotions. For the first setting, pitch and pauses are dominating, whereas in the latter one, Mel Frequency Cepstral Coefficients (MFCC) and their derivates are more important [108]. Offline recognition performs better than online recognition [105]. Binary classifiers for arousal and valence achieve accuracies of up to 85 % [104]. The recognition of seven different emotions achieves similar accuracies [109].

For applications like recorded phone calls, the recorded audio has to be segmented in order to eliminate phases of silence and for speaker diarization. Giannakopoulos provides a complete chain from feature extraction, segmentation and classification [107].

There are also apps available that claim to detect emotions from voice, e. g., GOOD VIBRATIONS sell an application programming interface (API) and a software development kit (SDK) for emotion recognition from audio, but they don't provide any validation results to show the performance of their classifier [4].

1.4.7 Overview over predictive modelling methods

Feature Selection

Commonly, feature selection increases model performance. Currently used methods comprise wrappers that select a subset of features based on the prediction performance an algorithm achieves with it. Prediction performance usually is evaluated with cross-validation. Research was done on search strategies since an exhaustive search is not feasible for

[4] http://www.good-vibrations.nl

large sets of features. Greedy search strategies are efficient and have the additional advantage to not overfit and can be designed as *forward selection* and *backward selection*. Wrappers can be combined directly with the training procedure, leading to embedded methods [110]. In contrast to wrappers, filters do not involve a predicting model, but they use metrics with less computational effort, e. g., mutual information or based on correlation [110].

Classification

Common techniques for classification comprise, amongst others, Support Vector Machines (SVMs) and Random Forests (RFs)). They are popular because they do not make any assumptions about the distribution of features and normally find the important ones by themselves. In addition, they are able to find non-linear decision boundaries. They are used in a broad range of fields, e. g., activity recognition [111], the above mentioned emotion extraction from voice and provide even support for medical decisions [112].

Novelty and outlier detection

Novelty and outlier detection is of interest in a broad range of fields, e. g., from supervision of manufacturing [113] to fraud detection[114], e. g., for credit cards [115]. Common approaches estimate an *anomaly score* for each data point. The anomaly score measures the degree of being an outlier. This estimation can be based on principal component analysis (PCA), the angle of vectors in the feature space (angle based outlier detection (LOF)) or density based, i. e., considering the *reachability distance* of points (local outlier factor (LOF)) [113].

Model-based approaches adapt classification methods to detect novelty, e. g., the one-class-SVM of Schölkopf et al. and the IF [116]. Both methods can be trained with datasets containing no outliers. An overview over one-class classifiers is given in [117].

If the underlying data are time series, breakout or change-point analysis can be performed. Gurarie et al. developed a method to detect behavioural changes based on change points in one-dimensional autocorrelated Gaussian time series with missing data [118]. Anomalies in data provide an additional challenge for change point analysis. One can detect them first [119] and remove them or replace them using an imputation method [120]. James et al. provide a breakout detection

that is robust against anomalies and can be applied on the time series directly [121]. Their approach is based on E-Divisive with Medians (EDM) and works also for multivariate series [122].

Linear Mixed Models

For longitudinal studies that repeat measurements in individuals, linear mixed models are well-established. They have the advantage that they deal well with missing or irregularly sampled data and take differences in individuals into account by estimating fixed and random effects [123].

1.5 Contributions

Aiming for the goals of section 1.2, this thesis makes the following contributions:

Patient studies: Methods how to conduct studies with severely-ill patients involving wearables.

Monitoring system for severely-ill patients: A monitoring system for physical and social activity tracking including monitoring of vital signs that is developed following a patient-centric design.

Feature extraction pipeline: A data processing pipeline consisting of necessary steps for preprocessing and data cleaning as well as of methods based on machine learning to add higher level information and to extract features for several sensor modalities. The last two steps provide information about patients' behaviour.

Datasets of CRF and PC patients: Comprehensive datasets containing activity data of CRF patients and activity and vital data of palliative patients plus daily self-reports and weekly answers to validated questionnaires. Evaluation of feasibility and acceptance of monitoring these patient groups.

Methods for analysis of the the datasets: Methods how to generate knowledge out of the collected datasets.

Insights to the daily life of CRF and palliative patients: Discovery, visualisation and statistical analysis of behavioural changes, subjective feelings of patients and features related to unplanned re-admissions.

1.6 Thesis outline

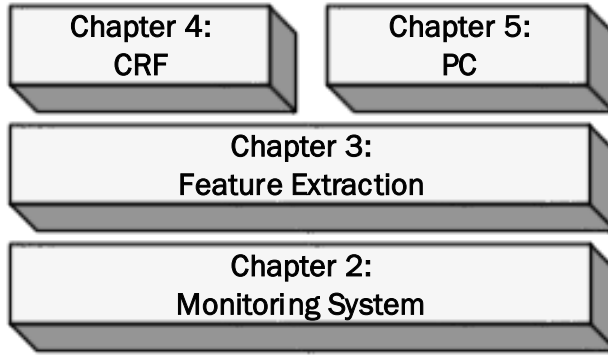


Figure 1.1: Thesis structure overview.

This thesis is structured as depicted in Figure 1.1. First, the methods that are used in both studies are described, then the studies with their outcome are presented.

Chapter 2 presents the monitoring system that was designed and implemented with the focus of being deployed in the palliative care and cancer related fatigue studies.

Chapter 3 describes the feature extraction for all collected sensor data. It includes basic re-sampling as well as machine learning methods such as clustering and supervised learning which is needed to either enrich the dataset with additional information or to clean the dataset from noise or invalid data.

Chapter 4 and 5 present the performed studies with cancer survivors suffering from cancer related fatigue and palliative patients in ambulatory care. Each chapter comprises a description of the study protocol, how the study was conducted, methods for study specific data analysis as well as the study outcomes and results. Each chapter closes with a study specific discussion including conclusions from the results, limitations, recommendations and outlook for future research on these fields.

Finally, chapter 6 summarises the thesis with overall limitations and conclusions and gives an outlook on future work in the field of using wearables in studies with severely-ill patients.

1.7 Publications

This thesis is based on the following publications:

- V. C. Klaas, A. Calatroni, and M. Hardegger, "Monitoring Patients in Ambulatory Palliative Care: A Design for an Observational Study", Wirel. Mob. Commun. Healthc., vol. 192, pp. 207–214, 2017
- V. C. Klaas, G. Tröster, N. Buel, H. Walt, and J. Jenewein, "Smartphone based monitoring of cancer related fatigue", 2017 IEEE 13th Int. Conf. Wirel. Mob. Comput. Netw. Commun. Rome, Italy: IEEE, oct 2017, pp. 249–256
- submitted:
  V. C. Klaas, A. Calatroni, M. Pavic, M. Guckenberger, G. Theile, Gerhard Tröster, "Preprint Research Article/Editorial Monitoring Patients in Ambulatory Palliative Care: a Design for an Observational Study", EAI Endorsed Transactions
- submitted to MDPI information, special issue "e-Health Pervasive Wireless Applications and Services (e-HPWAS'17)":
  V. C. Klaas, G. Tröster, H. Walt, and J. Jenewein, "Remotely Monitoring Cancer-Related Fatigue Using the Smart-Phone: Results of an Observational Study"
- in preparation for JAMA Oncology:
  M.Pavic*, V.Klaas*, J.Kraft, M.Guckenberger, G.Tröster, G.Theile, "Mobile health technologies for continuous monitoring of palliative care patients at the interface of in-patient to outpatient care: A feasibility study to predict health status deterioration"

Additionally, the following publications are co-authored:

- A. Seiler, V. Klaas, G. Tröster, and C. P. Fagundes, "eHealth and mHealth interventions in the treatment of fatigued cancer survivors: A systematic review and meta-analysis," Psychooncology., vol. 26, no. 9, pp. 1239–1253, sep 2017.
- G. Theile, V. Klaas, G. Tröster, and M. Guckenberger, "mHealth Technologies for Palliative Care Patients at the Interface of In-Patient to Outpatient Care: Protocol of Feasibility Study Aiming to Early Predict Deterioration of Patient's Health Status," JMIR Res. Protoc., vol. 6, no. 8, p. e142, aug 2017

2

The activity monitoring system and the conducted studies

The chapter describes an activity monitoring system that was deployed in observational studies. Patient devices which were developed following a patient-centric design, are described in detail. Finally, an overview on the conducted studies is given.

2.1 System overview

As described in section 1.1, this research is motivated by the idea of a remote monitoring of palliative outpatients in order to reduce non-elective re-admissions to hospital and emergency visits in the crucial phase of transition from hospital to home. Figure 2.1 shows a monitoring system that could support patients in this transition phase. This work focuses on the patient interface and data processing. In future, a web-based interface could automatically provide information about patients' health status to care givers, e. g., to physicians. The monitoring system as shown in 2.1, was designed to be deployed in two observational studies (ref. section 2.3 and section 2.4): for these studies, commercially available wearable devices have been integrated by means of an Android app. Since the observational studies involve digital questionnaires to be filled out by patients on a regular basis, the designed app provides a user interface that has been developed following a patient-centric approach.

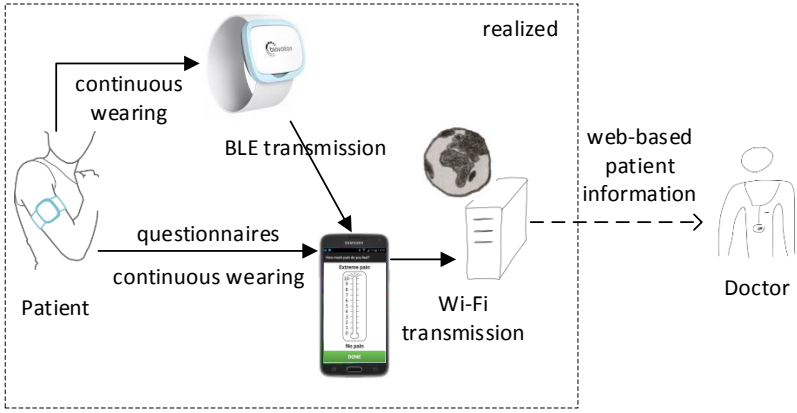


Figure 2.1: Overview over monitoring system: the Everion® has been deployed only in the palliative care study, the apps used in both studies differ only in the user interface (digital questionnaires for patients and configuration menu for staff). BLE: Bluetooth low energy

2.1.1 System components

The activity monitoring system consists of two wearable devices, namely a smartphone (typically a Samsung Galaxy S5) and an arm-worn sensor (Everion®) which provide the interface to the patients, and of a data receiving server at ETH Zürich in Switzerland. Similar systems were used in both studies. The differences have been marked in the text.

2.1.2 Secure and reliable data transmission

The smartphone app sends encrypted data to a secured server at ETH using the Wi-Fi infrastructure available at patients' homes. If needed, patients are provided with a mobile hotspot. By using a Wi-Fi network, patients are not charged by their carrier for mobile data usage. Furthermore, Wi-Fi provides a more stable and usually faster connection. Successfully uploaded data is automatically deleted from the smartphone.

2.2 Patient-centric design

This section is based on [124][1].

2.2.1 Usability considerations and interviews

The patients who participated in the studies (ref. section 2.3 and section 2.4) were described as mainly elderly people who were not necessarily acquainted with technical devices and could have mental or physical limitations. Device handling included general control and usage of the smartphone, at least the usage of the Activity Monitoring app presented here, putting on and off a wristband (Fitbit Charge HR, since at that time, the Everion® was not yet available) and charging both devices once per day.

To evaluate and finalise the app user interface for the palliative care study (ref. section 2.4), interviews were conducted with 12 patients in the age of 49 to 80 years (mean: 63.5, sd:9.7), four of them were female [124]. The interviewees could choose between various designs

[1] V. C. Klaas, A. Calatroni, and M. Hardegger, "Monitoring Patients in Ambulatory Palliative Care: A Design for an Observational Study," Wirel. Mob. Commun. Healthc., vol. 192, pp. 207–214, 2017

(variations in size and colour of the interface elements), test different control elements (a button and a slider) and they were also asked about their opinion if they would like to receive feedback about their activity through the app. The outcome has been grouped in four categories and considered for the finalisation of the app.

App design and usability

In general, patients preferred more design solutions which reduce the risk of misuse, e. g., bigger numbers and a confirmation dialogue before saving the input. The question of smileys in the app design was quite controversial and was therefore emotionally discussed, using words like "hate" and "sweet". All patients who liked smileys also preferred a more colourful design. All interviewees were able to use the smartphone app and handle the wristband. Special needs were related to co-morbidities, e. g., limited vision capabilities due to advanced age. The interviews did not reveal any special needs related to the palliative situation of the patients.

Feedback through the smartphone app

Eight out of ten patients understood the concept of feedback. Three patients – having already experience with fitness trackers – stated that they were interested in their physical activity data once per day. However, they had different attitudes, considering the feedback as a source of information or rather as an entertainment.

Motivation to use a monitoring system

Most of the interviewees were motivated to use a monitoring system and to support research by participating in a study. Two patients suffered from fatigue and found the idea of carrying the smartphone with them also at home too cumbersome. Concerns about privacy and data security were discussed twice. One of the fatigued patients stated that she did not want to take an additional burden for the little time that is left for her.

Vulnerability and sensitivity of patients in palliative care

Some of the interviewed patients were not aware of their health condition and optimistic to recover from their disease. The exceptional

situation of the interviewees in addition to a new encounter with the previously unknown interviewer yielded in dynamic and intense interviews that required mindfulness and an open attitude from the interviewer. Some patients used the opportunity for longer conversations. About 25 min of each interview (50 % of the total interview duration) were about topics irrelevant to the interviews and were considered as a gift by the interviewer.

2.2.2 Smartphone and app

The activity monitoring app was designed based on the interview results of section 2.2. The app consists of two parts which have been described in detail in the following subsections:

- The patient interface provides digital questionnaires to allow patients to rate their subjective perception of specific symptoms.
- The sensor logging module is responsible for recording and transmitting signals from built-in smartphone sensors as well as for the interface to the Everion®, including management of the BLE connection, requesting, receiving and recording Everion® sensor signals and transmitting these signals to a server.

Patient interface

The activity monitoring app used in the observational studies (ref. section 2.3.2, 2.4) was transparent for the user:

- The app is started by the study staff before handing out the smartphone.
- Once the app is started, it will restart automatically, e. g., after rebooting the smartphone.
- The recording and transmission of sensor data is handled automatically in the background.

The only interaction between the patient and the smartphone consisted in answering the digital questionnaires. In both studies, patients were asked to answer the digital questionnaires by a vibration alarm in combination with showing the first questionnaire on the screen.

The digital symptom questionnaires were designed based on the National Comprehensive Cancer Network® (NCCN) distress thermometer (ref. section A.2.7) as visual analogue scales (VAS) with values from 0 (no symptom) to 10 (maximal symptom). In case the questionnaires were not answered, the alarm was repeated up to 5 times at 5 minutes intervals.

Palliative Care App Once per day, the app shows two questionnaires in a row concerning the current subjective pain and the current subjective distress as illustrated by Figure 2.2. The reasoning for the frequency of the questionnaires is given in section 2.4.2.

Fatigue App 4 times per day, the app shows three questionnaires in a row concerning the current subjective tiredness, the perceived impact of the tiredness on patient's daily life and about the currently performed activity. The VAS are designed as shown in Figure , latter questionnaire lets choose from a list of 8 activities as illustrated by Figure 2.3.

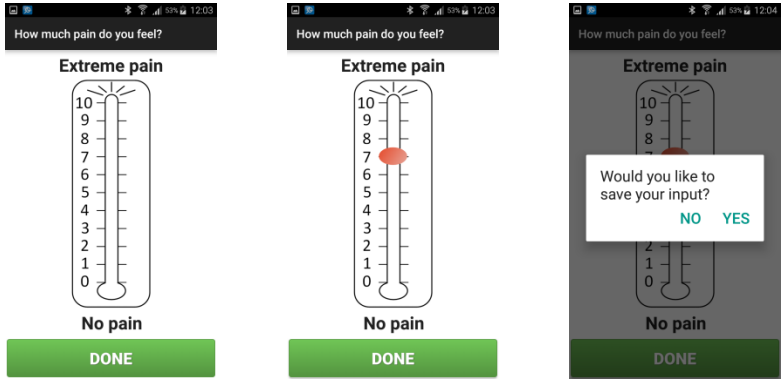
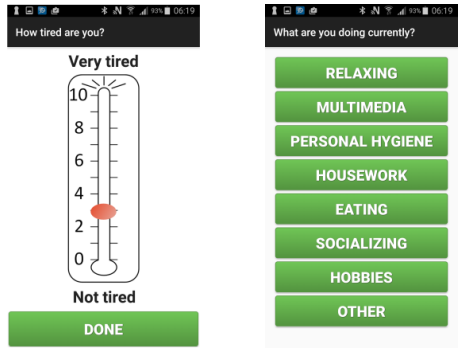


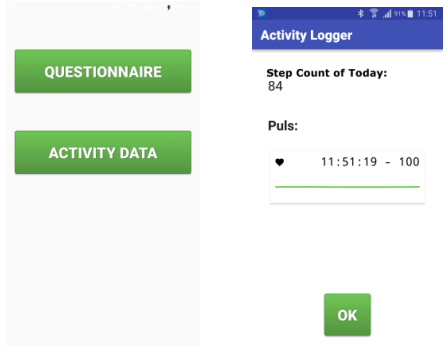
Figure 2.2: English translation of the digital pain questionnaire as deployed in German in the palliative care study. a) the empty thermometer is shown in order to avoid a bias of the answers, b) the value can be selected by touching and sliding, c) a confirmation dialogue minimises misuse.



a) fatigue scale b) activity selection

Figure 2.3: English translation of the digital questionnaire as deployed in the Fatigue study. a) VAS scale, b) question concerning the currently performed activity.

The patients in palliative care can view their current heart rate and current step count of the current day as shown in Figure 2.4.



a) start screen b) current data

Figure 2.4: Information about current heart rate and daily step count for the user shown by the app: a) menu to enter the screen, b) view on the current vital and activity data.

Sensor logging module

The recording of sensor data through the activity monitoring app has been designed so that the smartphone can be used for at least one day before it has to be re-charged. The smartphone logs sensor measurements as described in Table 2.1. The given sampling rates are a trade-off between battery consumption and data richness.

2.2.3 Everion®

In the palliative care study, the Everion® developed by the start-up company Biovotion[2], which is located in Zurich, has been deployed as the wearable sensor for measuring vital signs. The only interaction of the patients with the Everion® is to wear it and to charge it. Patients receive the devices ready to use, i. e., the devices are already paired via BLE and fully charged. The sensor is streaming measurement data to the tracking app automatically. The Everion® fulfils the following requirements:

- automatic logging and data transmission to smartphone (no user interaction needed): the Everion® provides streaming to the smartphone and the option for an automatic connection/reconnection with the smartphone
- data storage to avoid data loss during lost connection: an internal ring buffer is capable of storing the data up to 4 days
- control over access to recorded data (i. e., no transmission into any manufacturer cloud): a developer SDK that allows to fully control the sensor through a smartphone app.
- accurate measurement of the heart rate: It performs measurements as specified in Table 2.2. Biovotion has evaluated and documented the measurement accuracies.
- at least consumer CE mark: approval as medical device (class 2) in Europe by a notified body and in the United States of America (USA) by the FDA.

[2] http://www.biovotion.com/

Table 2.1: Sensor modalities recorded by the smartphone. The sensor type is given in the description column of the table. (This table has been adapted from [125].)

Modality	Description	Sampling Rate
Physical Activity	Changes in physical behaviour can be related to progression of disease or worsening of symptoms. The sampling rate is controlled by Android and varies depending on the usage.	
	Accelerometer	40 Hz
	Barometer	2 Hz
	Magnetometer	10 Hz
Location	Location provides frequency and duration of visited places and unveils changes in user's behaviour, in particular his social interactions [126], which are also known to be a health indicator [127].	
	GPS	every 3 min
	Wi-Fi	every 20 s
Social Activity	Social behaviour can be related to changes in the disease [127].	
	Phone call statistics	once per day
	App usage statistics	once per day
	Battery level, indicator for phone usage	at change
Emotions	Phone calls are recorded and encrypted for later voice analysis.	

Table 2.2: Sensor modalities recorded by the Everion®

Modality	Unit	Sampling Rate	Source	Description
Heart Rate (HR)	bpm	1 Hz	PPG	heart rate
Blood Oxygen Saturation (SpO_2)	%	1 Hz	PPG	level of oxygen saturation in blood
Respiration rate (RR)	bpm	1 Hz	PPG	number of inhale-exhale cycles per minute
Energy expenditure (EE)	$kcal\,s^{-1}$	1 Hz	PPG	energy expenditure
Blood pulse wave (BLPW)	––	1 Hz	PPG	indicator for rhythm and shape of the blood pulse wave which is related to HRV and stress
Blood perfusion (BLPE)	––	1 Hz	PPG	skin blood perfusion depends on skin temperature and a good level is necessary for accurate measurements
Inter pulse interval (IPI)	ms	depending on heart function	PPG	basis for various metrics that serve as indicators for health and fitness (ref. section 1.4.3)
Activity (act)	––	1 Hz	3-axis accelerometer	activity index in the range of 0–255 based on acceleration
Class (class)	––	1 Hz	3-axis accelerometer	activity class 0–8 denoting activities, e.g., resting, walking etc.
Step count (steps)	$steps\,s^{-1}$	1 Hz	accelerometer	step count per second
Skin temperature (Temp)	°C	1 Hz	Temperature sensor	skin temperature, normal range: 32 °C–35 °C
Galvanic Skin Response (GSR)	kΩ	1 Hz	GSR sensor	The skin resistance (impedance) is affected by emotions.
Air pressure (bar)	mbar	1 Hz	barometer	Atmospheric air pressure

2.2.4 Server

The back end consists of a web server that receives the data and stores them in a storage accessible via Secure File Transfer Protocol (SFTP). As marked in Figure 2.1, the web interface for physicians can be designed based on the outcome of this work in the future.

2.3 Pre-study: Activity monitoring of fatigued cancer survivors

A pre-study with seven fatigued cancer survivors was conducted from May 2016 to January 2017. The feasibility study targeted participants of a completed study conducted by the psychiatry ward of the University Hospital Zürich [128].

The study protocol with all necessary documents including informed consent, forms and paper questionnaires for the patients has been approved by the local ethics committee (Kantonale Ethikkommission Zürich) as an amendment to the study with the number KEK-ZH-Nr. 2012-0563.

2.3.1 Patient recruitment

Participants had to fulfil the following inclusion criteria:

- Aged > 18 years
- mid to severe fatigue, i. e., FACIT-F $\leq$ 30, ref. section A.2.8
- willingness to use the provided smartphone
- successful briefing how to use the smartphone and activity monitoring app
- signed informed consent

Excluded were patients with a relevant cognitive impairment or severe depression assessed by M.I.N.I, a validated measure to assess psychiatric clinical pictures.

Patients were recruited by phone calls during which they were:

- explained the study concept and goal
- asked to participate in the study

- tested for eligibility, i. e., tested for fatigue FACIT-F and depression (M.I.N.I)

The results of patient recruitment are described in detail in section 4.2.

2.3.2 Study Procedure

Baseline

Interested patients were either visited at home or met at the university hospital or another place convenient for them. They had to fill out and sign a general and informed consent. Participants have been provided with a smartphone (Samsung Galaxy S5) with the pre-installed activity monitoring app and equipped with a prepaid SIM-card, a smartphone charger and a smartphone belt or pocket depending on their needs. After introducing them to the app and, if necessary, to the smartphone, they had to show that they are able to handle the device. If they had no Wi-Fi at home, they receive a Wi-Fi hotspot as well.

Study period

Naturally, the participants had the possibility to contact the study team at any point of time.

The first week. On the third day after start, any upcoming questions concerning the devices or the study were clarified during a home visit or phone call.

Digital questionnaires. Four times per day (random in fixed time windows), the participants were notified via smartphone vibration alarm to answer the digital questionnaires. The frequency of four times per day was chosen as a trade-off between a comprehensive dataset using the experience-based sampling method (ESM) [129] and a reasonable burden put on the group of fatigued patients who already had a reduced energy level insufficient to perform the activities of daily living (ADL) [130].

Wearing the smartphone. The participants should wear the smartphone the all day long and charge it over night.

Weekly interviews. Weekly interviews comprised the FACIT-F questionnaire and the usage of the smartphone.

Study end

At the end of the follow-up interval, patients filled out again the FACIT-F and were interviewed about their experiences during the study and their usage the smartphone and app. The interviews were conducted in a personal meeting using the technique of a guided conversation [131].

2.4 Activity monitoring of patients in ambulatory palliative care

The study protocol is described in detail in [132]. The study was designed as an observational study, i. e., there were no planned interventions. The study protocol with all necessary documents including informed consent, forms and paper questionnaires for the patients, was approved by the local ethics committee (Kantonale Ethikkommission Zürich) under the number PB_2016-00895. The study was funded by the "Research in Palliative Care" program of the Swiss Academy of Medical Sciences (SAMW) and the Gottfried and Julia Bangerter-Rhyner Foundation, Bern, Switzerland.

2.4.1 Patient recruitment

Between 1. March 2017 and 31. March 2018, 31 patients were recruited at the university hospital Zurich (USZ) on two wards: radiation-oncology and palliative care. One patient passed away a few days after being discharged from hospital and did not participate actively in the study. The last patient finished the study on 24. Mai 2018. Inclusion criteria were

- established diagnosis of metastatic cancer or other severe illness
- limited life-expectancy with a physicians' guess of > 8 weeks and < 12 months
- Karnofsky Index $\geq 50\,\%$ (ref. section A.1.2)
- ECOG ≤ 2 (ref. section A.1.1)

- age > 18 years

Patients without sufficient knowledge of German or with a relevant cognitive impairment were excluded.

Detailed results of patient recruitment have been described in section 5.2.1.

2.4.2 Study procedure

Screening and baseline

Interested patients had to fill out an informed consent, were introduced to the devices and tested if they were able to handle the devices. If they showed their capability to handle the devices, they received a kit as shown in Figure 2.5 consisting of

(a) a smartphone with the app, which had been installed and started before, and equipped with a prepaid SIM-card and a charger for the phone

(b) the Everion® with a custom-sized band and a charger for the sensor

(c) a smartphone belt (optional)

(d) a mobile Wi-Fi hotspot including a charger, if there was not Wi-Fi available at home

Patients who had been already using Android smartphones before, could choose between using their own smartphone or switching to a study smartphone (Samsung Galaxy S5 or Samsung Galaxy S5Neo). All patients had the option to keep the study smartphone after their study participation – also in case of an early drop-out for any reason.

Baseline. At discharge from hospital, patients filled out the Edmonton Symptom Assessment Scale (ESAS), ref. section A.2.1, and the quality of life questionnaire (QLQ-C30) of the European Organisation for Research and Treatment of Cancer (EORTC), ref. section A.2.2.

Study period

Figure 2.6 illustrates the study procedure for included patients. Patients were asked to participate for 12 weeks. The period of 12 weeks

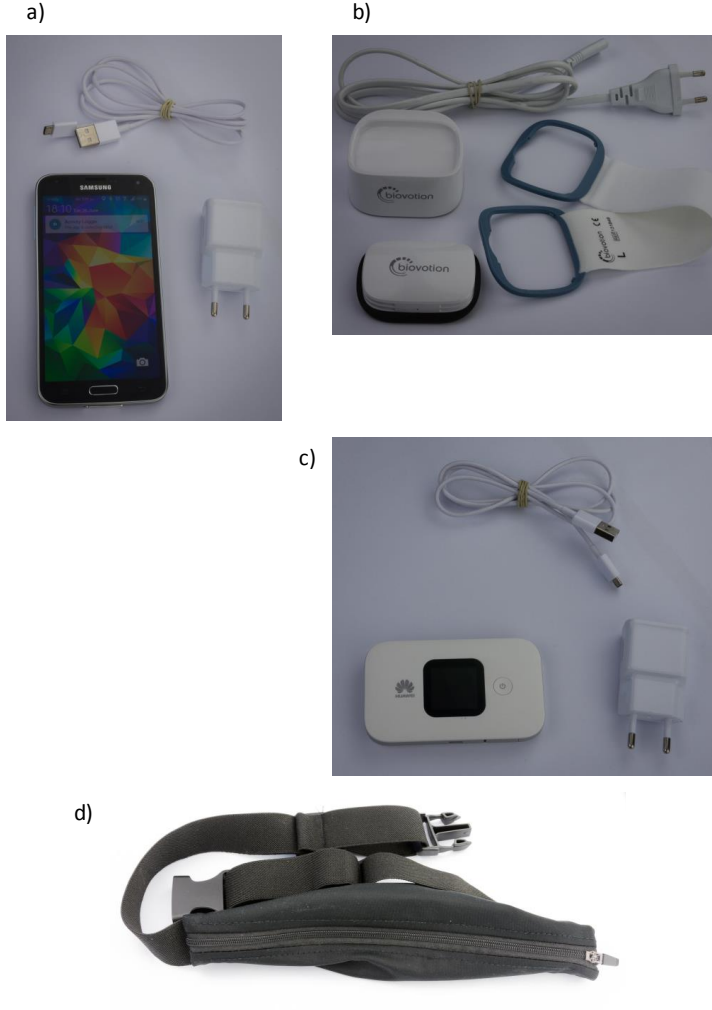


Figure 2.5: A patient's tracking kit comprises: (a) smartphone with charger, (b) arm-sensor with elastic band and charger, (c) smartphone belt, (d) mobile hotspot with charger

was fixed as a trade-off between a comprehensive dataset and a reasonable burden put on this particularly vulnerable group of patients

who were confronted with the fact of being terminally ill.

Exceptions were early-drop outs as described in section 5.2.1 and 5.3.5. In addition, some patients travelled for holidays during their study participation. For the holiday period, the tracking was paused and resumed afterwards. In these cases, the duration of 12 weeks was extended accordingly if the patient agreed.

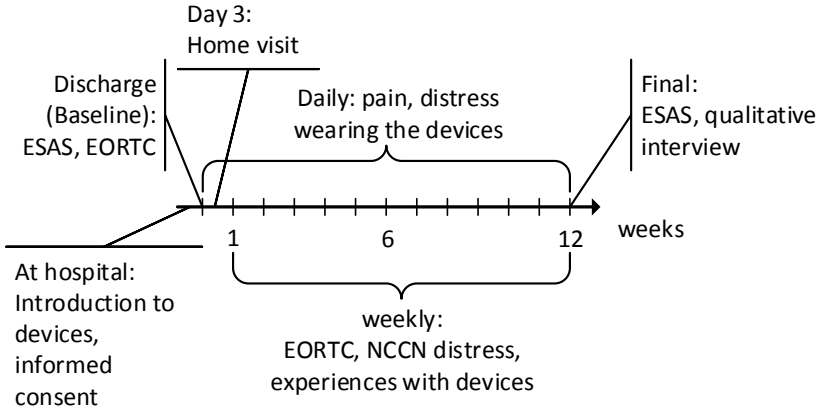


Figure 2.6: Time line for the study participation of a patient.

The first week. In the first week, a home visit was planned to ensure the correct deployment at home (mainly configuration of Wi-Fi for data transmission) and to provide additional support for the patient if needed. Several patients did not return to home directly from hospital, but stayed at a rehabilitation centre for two weeks or more. In those cases, the home visit was made after their return from the rehabilitation centre. Especially with younger patients, home visits were not always necessary. In those cases, they were replaced by a phone call.

Digital questionnaires. Every day, the smartphone asked the patients via vibration alarm to rate their pain and distress on a VAS as described in section 2.2.2. The time of the day was randomised between 8:00 and 20:00 o'clock. The time window could be adapted to patients' needs. Patients could also start the VAS on their own if they wished so.

Wearing the devices. Patients were wearing the devices all day long. During night time, they charged their smartphone and the Everion® if not worn. Patients decided themselves if they were wearing the Everion® during night time or not. If they were wearing it during night hours, they were recommended to charge the sensor twice per day: after getting up and during the evening.

Weekly interviews. Once per week, patients were contacted by the study team. They answered the EORTC QLQ-C30 questionnaire, a questionnaire concerning their experiences with the devices in the previous week as shown in section A.2.5 and, if they rated the distress on the digital questionnaire above 4 during the previous week, the second part of the NCCN distress thermometer which has been described in section A.2.7.

Study end

Thank you card. One to two weeks before the study end, patients received a card thanking them for their participation together with two questionnaires as preparation for the final interviews.

Final interviews. After the last study day, patients were called or visited for a final interview comprising the ESAS questionnaire and a qualitative, semi-structured interview following the guideline as provided in section A.2.9).

Devices. The patients were offered to keep the smartphones except for the SIM-cards. The Everion® sensors were collected either during an organised appointment or home visit or the patients sent them back by means of an already prepared box. They only needed to put the sensor, charger and SIM-card in the box, use the stamps and adhesive address label, close the box and bring it to the next post office.

Cases of death. Close relatives who were registered as contact person received a condolence card. The mode and timing of returning the devices (excluding the smartphone) by relatives was organised depending on the needs of the bereaved.

3

Feature Extraction

This chapter describes the feature extraction pipelines for both studies. The pipelines take as input the raw text files as received by the web servers and give regularly sampled series of features as output. Feature extraction is necessary for the later performed analysis.

3.1 Goals and overview

Before classifying or clustering unknown data, it is favourable to extract features from raw signals as input for the subsequent machine learning and statistical methods. Deriving features from raw values on sliding windows has the following benefits:

- Possibility to reduce the data amount (samples) by reducing the sampling rate with sliding window technique: since many algorithms are worse than linear in computational complexity for training, e. g., the computational complexity of a kernel SVM is quadratic in the number of samples
- Adding information about context: by deriving values using a sliding window, features contain information about the temporal context. Not the absolute value of a signal is used, but statistics over several samples. For example, zero crossing rate, variance and frequency based features give information about the dynamics of the signal, the mean on a sliding windows leads to a moving average that is a finite impulse response filter (FIR) filter to smooth the signal by eliminating fluctuations within a window.

For the data analysis of the two studies which were introduced in Chapter 2, a feature extraction pipeline was established. Differences are specified accordingly in the subsections. The context for the feature extraction pipeline is illustrated in Figure 3.1. For data from both devices, the smartphone and the Everion®, the same preprocessing steps have to be applied. Naturally, the implementation of signal preprocessing is device and sensor modality specific and therefore described in the subsequent sections as annotated in Figure 3.1.

The feature extraction pipeline comprises several steps as illustrated in Figure 3.2. First, the raw data from the sensor signals as received by the web severs (one for each study) is preprocessed to obtain data with constant sampling frequencies. The output data can contain missing values. For example, if the smartphone is in flight mode, it does not measure and store GPS and Wi-Fi signals.

This chapter is structured according to this pipeline: the following two sections describe the data preparation consisting of preprocessing, cleaning and basic feature extraction steps for the signals of both devices followed by the final feature extraction. The extracted features

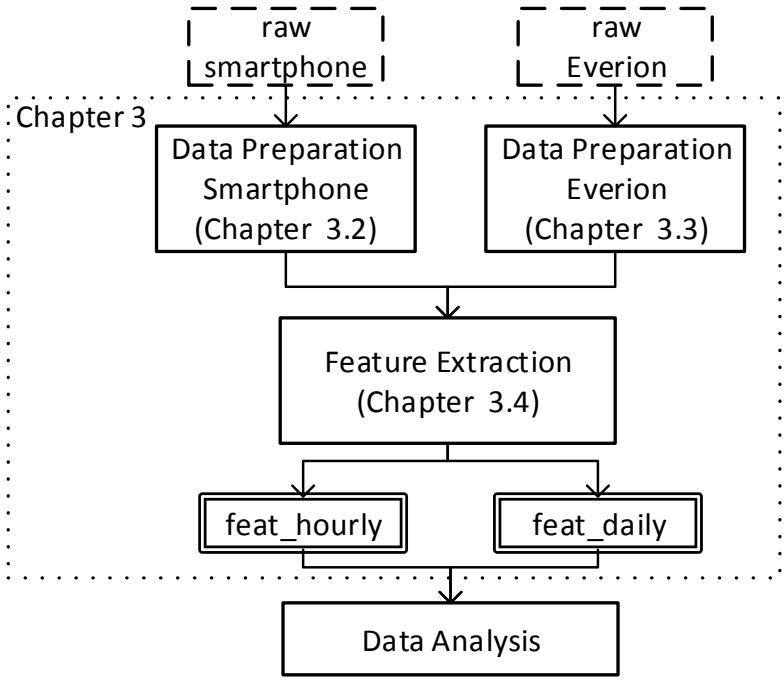


Figure 3.1: Feature extraction pipeline – from raw sensor signals to features ready for data analysis: This chapter describes the performed steps to transform raw sensor signals to input data for analysis methods. Double lines identify output of feature extraction.

are analysed in the subsequent chapters. Chapter 4 presents the analysis of the fatigue study data and uses only the smartphone features, whereas Chapter 5 presents the analysis of the palliative care study data and uses features of both wearable devices.

3.2 Preparation of the smartphone data

The smartphone data was processed for both studies (ref. sections 2.3, 2.4) in the same way: the smartphone data comprises two categories of data, namely pseudo-continuously sampled data and data recorded only when triggered by events, e. g., phone calls and self-reports. Some

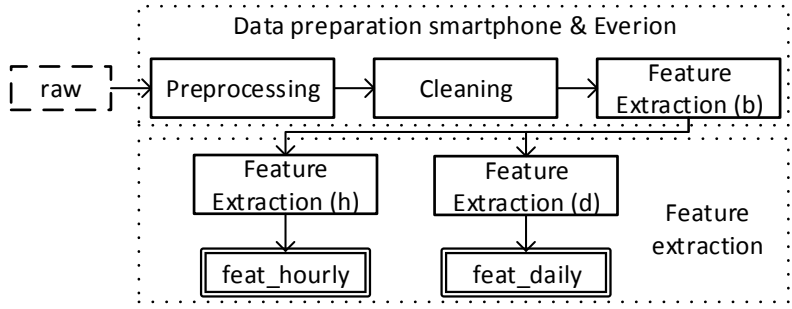


Figure 3.2: Feature extraction pipeline for both devices with three levels of features: (b) basic with sampling rate depending on feature, (h) hourly sampled, (d) daily sampled.

sensor modalities (acceleration, GPS, Wi-Fi) need an additional information extraction step, e. g., adding information like on-/off-body labels or clustering labels for location signals. These are described in detail in the following subsections as indicated in Figure 3.3.

In case of recorded phone calls (voice), the analysis could not be continued in order to respect data privacy of the patients. Nevertheless, section A.3.1 shows the possibilities and challenges for unsupervised learning given by the collected (completely unlabelled) dataset.

The Tables 3.1, 3.2 specify the recorded smartphone sensor signals in detail.

3.2.1 Preprocessing of smartphone data: re-sampling and cleaning of continuous signals

Since Android is not a real-time operating systems, the sampling rate is not constant and due to high processor load or low battery state, samples (or measurements) may be dropped. For further analysis, e. g., in frequency domain, many methods require a uniform sampling rate and gap-less data. To achieve a uniform sampling rate, the data is re-sampled and gaps up to a certain length are interpolated. Longer gaps, e. g., an hour of missing acceleration data, are treated as missing data. The re-sampling frequency and the tolerated gap length depends on the sensor modality. Table 3.3 shows the distribution of the different observed sampling rates, to which frequency they are re-sampled dur-

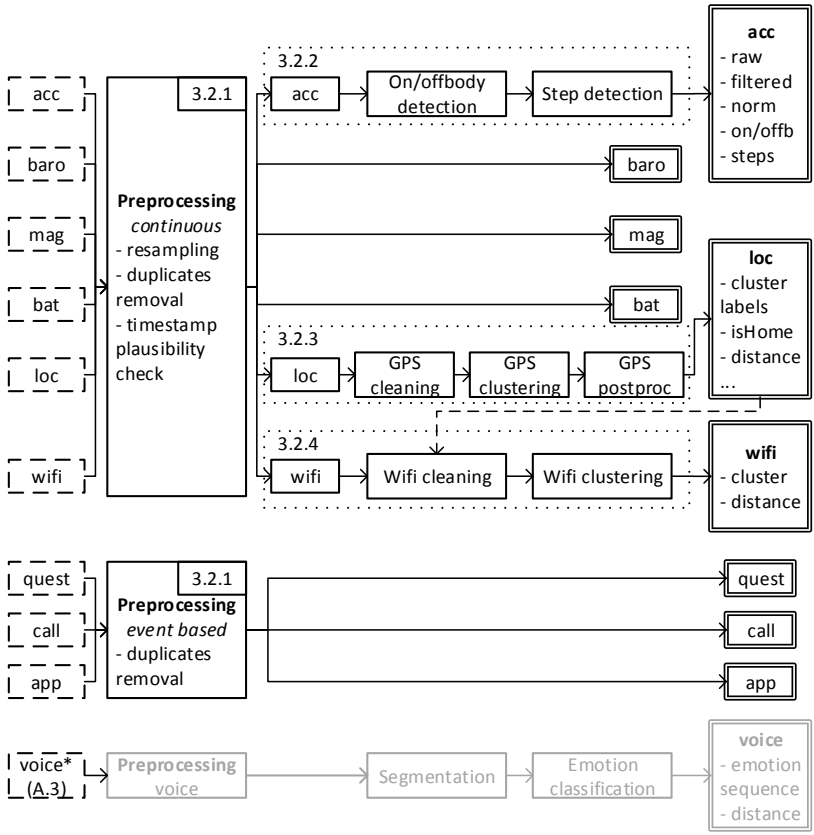


Figure 3.3: Data preparation of smartphone signals. Dashed lines identify input data, double lines identify the output of the data preparation, grey parts (voice features) were not performed for all patients (ref. section A.3.1).

ing preprocessing and up to which length gaps are filled by Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) interpolation. The distribution of the sampling intervals was determined experimentally based on one day of test data. The re-sampling frequencies were chosen according to the specification of Android. In general, down-sampling was preferred to up-sampling in order to not add synthetic information (i. e., values that were not measured). Interpolating gaps up to a

Table 3.1: Overview over the smartphone sensor modalities that are recorded continuously.

Value	Unit	Description
Accelerometer (acc)		
datetime	ms	Coordinated Universal Time (UTC) time in ms from 1970-01-01
x	$m\,s^{-2}$	acceleration in x-axis
y	$m\,s^{-2}$	acceleration in y-axis
z	$m\,s^{-2}$	acceleration in z-axis
Barometer (baro)		
datetime	ms	UTC time in ms from 1970-01-01
p	mbar	ambient air pressure
Magnetometer (mag)		
datetime	ms	UTC time in ms from 1970-01-01
x	µT	strength of geomagnetic field along x-axis
y	µT	strength of geomagnetic field along y-axis
z	µT	strength of geomagnetic field along z-axis
Battery (bat)		
datetime	ms	UTC time in ms from 1970-01-01
batlevel	%	battery charging level
GPS (loc)		
datetime	ms	UTC time in ms from 1970-01-01
long	°	Longitude coordinate
lat	°	Latitude coordinate
q_loc	m	Accuracy: radius denoting the circle having a probability of 68 % that the real coordinate is lying in it

Table 3.1: Overview over the smartphone sensor modalities that are recorded continuously (continued).

Value	Unit	Description
Wi-Fi (wifi)		
datetime	ms	UTC time in ms from 1970-01-01
ap_1	–	Media-Access-Control address (MAC) address of access point *i* found during Wi-Fi scan
$rssi_1$	dBm	detected signal level (RSSI) of access point *i*
...		
ap_n	–	MAC address of access point *N* found during Wi-Fi scan
$rssi_n$	dBm	detected signal level (RSSI) of access point *N*

certain size has the advantage that the input data for subsequent steps is less segmented and consists of more regular time series. The gap size threshold is determined as a trade-off between the aim of regular time series and the aim of not adding artificial information.

Furthermore, the raw data files contain also duplicates due to repeated data upload. Duplicates are identified by the timestamps and only the first values of such sets with identical timestamps are kept.

3.2.2 Processing of the acceleration signal

On-/off-body detection from smartphone data

The on-/off-body detection for the smartphone is adapted from Choi et al. which is based on non-zero counts (nczs) [133]. A non-zero count is a sample that is sufficiently different from the window mean. The algorithm of Choi was designed to detect off-body for wearables that are attached to the body with the single purpose of logging. However, a smartphone usually is not just attached to the body, but also used with the hands. To take this difference into account, the algorithm parameters are adapted to the values as listed in Table 3.4.

Table 3.2: Overview over the smartphone sensor modalities that are recorded when triggered by an event. A "–" denotes that a measurement variable is unit-less.

Value	Unit	Description
Questionnaires (quest)		
datetime	ms	UTC time in ms from 1970-01-01
Pain	–	VAS 0 to 10
Distress	–	VAS 0 to 10
Fatigue	–	VAS 0 to 10
Interference	–	VAS 0 to 10
Activity	–	a value out of: relaxing, multimedia, personal hygiene, housework, eating, socialising, hobbies, other
Phone call statistics (call)		
datetime	ms	UTC time in ms from 1970-01-01
duration	s	duration of call
partner	–	encrypted number of communication partner
type	–	flag denoting if the call is incoming or outgoing
success	–	flag if the call was answered or not
App usage statistics (app)		
datetime	ms	UTC time in ms from 1970-01-01
name_1	–	process name of running app 1
...		
name_N	–	process name of running app N
Phone calls audio (voice)		
datetime	ms	UTC time in ms from 1970-01-01
audio	–	encrypted audio file

Acceleration data is processed per patient and day. Figure 3.4 illustrates the off-body detection steps. The feature ncz is calculated on rolling windows of length of 1 min by applying the formula `count`. If

Table 3.3: Examination of Android sampling intervals based on one day of data, re-sampling frequency to which data is re-sampled and length of gaps up to which gaps are interpolated. "–": no re-sampling/interpolation

Sensor	Intervals [second]	Amount [%]	Re-sampling frequency [Hz]	Gap size [second]
Accelerometer 40 Hz	0.025	76.6	40	0.5
Accelerometer 50 Hz	0.02	23.4	40	0.5
Magnetometer 10 Hz	0.0998	100	10	0.5
Barometer 5 Hz	0.1747	80.6	5	2
Barometer 5 Hz	0.180	19.4	5	2
Wi-Fi	20	100	–	–
GPS	180.5	100	–	–

Table 3.4: Parameters for on-/off-body detection.

Parameter	Unit	Value	Purpose
Time Interval TI	min	1	basic window size
Sensor Threshold T_g	g	0.06	not count noise as non-zero
Non-Zero Count Threshold T_{nzc}	$\frac{counts}{\text{min}}$	100	consider small phone interaction (check time) as off-body
Consecutive Non-Zero Counts W_{nzc}	–	10	reduce false off-body detection and allow frequent short usages of the phone
Artifactual Movement Interval I_{AM}	min	0	allow up to I_{AM} consecutive on-body TI in off-body period
Up/Downstream Window W_{UpDown}	min	10	keep on-body label if there is more interaction with the phone within a window of size W_{UpDown}

$ncz > T_{nzc}$ a window is marked with 0 which stands for *off-body*. To avoid that a short interaction with the smartphone, e. g., grabbing it from the table to check the time or an incoming message, is considered as *on-body*, the non-zero count threshold T_{nzc} is set to a larger value than in the original implementation. Similarly, to avoid that a single time interval with little movement is marked as *off-body*, there have to be W_{nzc} off-body time intervals in a row in order to keep the label off-body. However, the value of 10 min takes into account that one tends to look at the phone for short periods of time, e. g., to check messages. Setting the artifactual movement interval $I_{AM} = 0$, on-body time intervals are not ignored. The size of the Up/Downstream window of 10 min allows that the time between regular interactions with the phone (e. g., check of messages every 15 minutes) is counted as *off-body*, whereas interactions longer than a minute are counted as *on-body* time.

Evaluation of smartphone on-/off-body detection. On a test dataset consisting of two test subjects and in total over 5 days of continuous data, the on-/off-body detection yields a F1-score of 0.96. Hence, the on-/off-body detection is applied to the collected smartphone data as a cleaning step.

Step detection from smartphone acceleration data

The step detection based on the acceleration signal recorded by the activity monitoring app on the smartphone is based on peak detection and realised as follows: the euclidean norm of the 3-axis signal is filtered using a low-pass FIR-filter with a cutoff frequency of 4 Hz. The filtering step prevents from detecting noise as peaks. With the chosen cutoff frequency, step frequencies can pass the filter and noise is filtered out [134].

On the result, periods with walking are identified based on the zero-crossing rate on a window of 256 samples: if the detected gait cycle frequency lies between 0.5 Hz and 2 Hz, the window is labelled as "walking". The filtered signal is input for the peak detection algorithm of Duarte [135]. Detected peaks that lie outside of "walking"-windows are ignored and the "peak" label is removed. The number of steps per second is then defined as the number of peaks in a window of one second. An example is given in Figure 3.5.

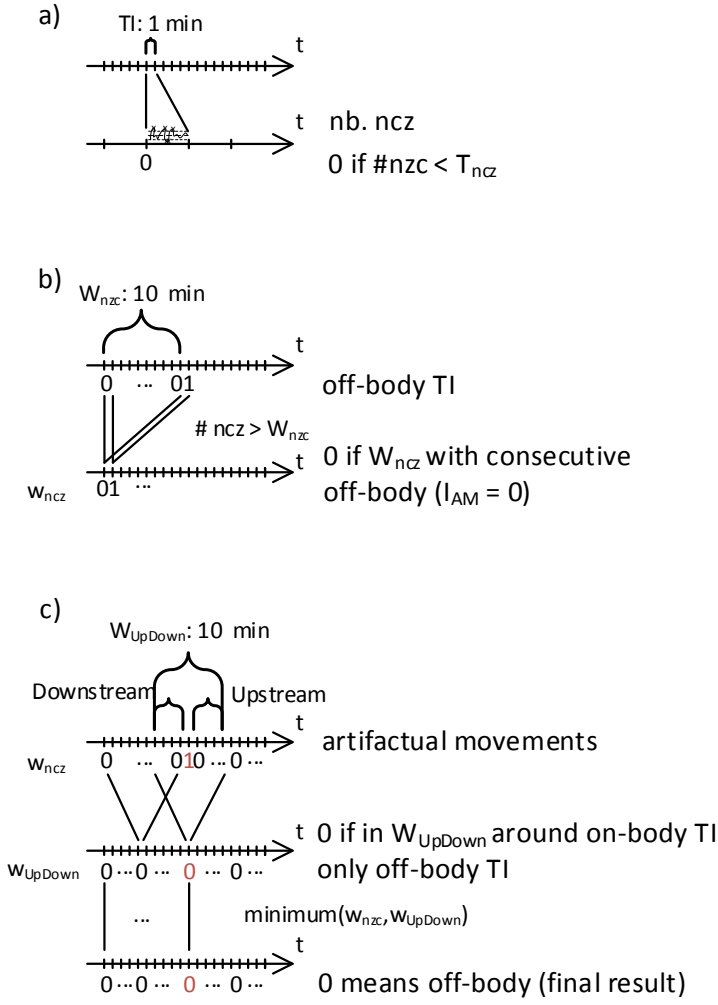


Figure 3.4: Smartphone on-/off-body detection. a) A Time Interval *TI* is marked with 0 if its non-zero count (nzc) is lower the threshold T_{nzc}; b) a window W_{nzc} is marked with 0 if all *TI* in this window are marked with 0; c) a window W_{nzc} marked with 1 is set to if the surrounding windows in the Up/Downstream window are all 0.

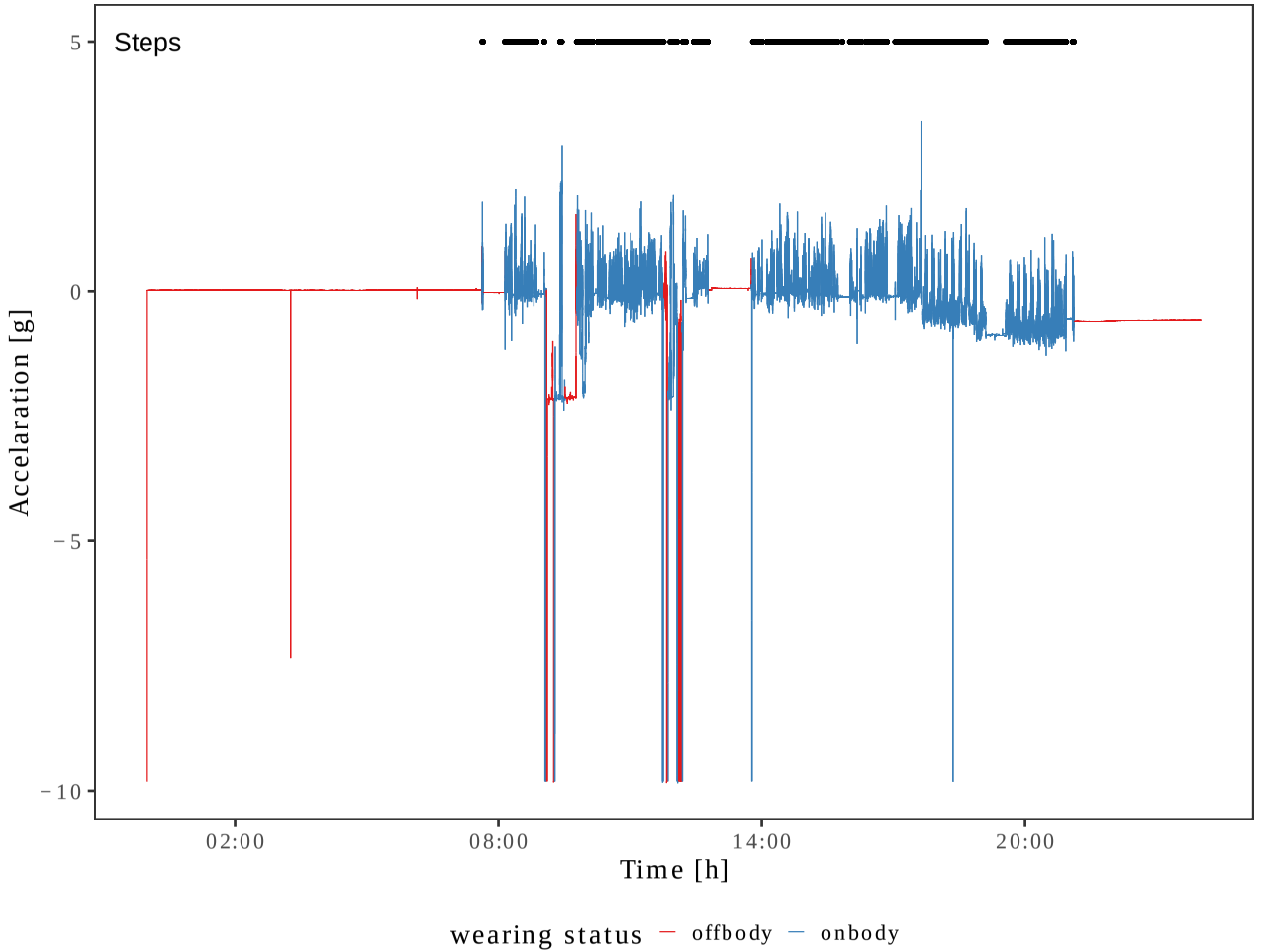


Figure 3.5: Smartphone step detection. In this plot, 10816 steps are counted.

3.2.3 GPS

As illustrated in Figure 3.3, the GPS data is processed in several steps consisting of additional, GPS-specific cleaning, clustering and clustering post-processing (extraction of places). The goal is to extract high level features based on locations such as time spent at home, number of places visited per day, etc., since these features can unveil changes in the smartphone user's behaviour [126, 127].

GPS cleaning

Especially indoors, the accuracy of measured GPS coordinates is varying widely. Phone GPS sensors deliver with the coordinates a radius to specify a circle in which the real location lies with a probability of 68 % (ref. Table 3.1). To filter the data for accurate localisation, points with a radius $> 50\,\mathrm{m}$ are discarded. In addition, since the studies focus on features based on places, samples that were recorded while the subject is moving, are filtered out. Movement is detected with a speed $> 10\,\mathrm{km\,h^{-1}}$. This threshold is a speed that is achieved by running. The value is chosen so that movement indoors is not filtered out. Most patients were using other means of transport such as cars or public transport so that moving outside is filtered out. For the patients that walked outside, the threshold was adapted empirically in order to keep the samples at home and eliminate moving outside. For the palliative care study, samples recorded while patients were in vacations is ignored (as is done with all sensor modalities). The other studies do not comprise vacations or the information is not available.

GPS clustering

After cleaning, the clustering method OPTICS (ref. section 1.4.4) is applied with the parameters `minPoints` = 10 and `eps` = 10 on the remaining set of GPS samples. As described in section 1.4.4, OPTICS is a state-of-the-art method for spatio-temporal clustering. OPTICS delivers an order of the samples according to their reachability distance, i. e., points that are lying close together will follow each other in the ordering. To extract clusters from the OPTICS result, the `ExtractDBSCAN` algorithm as implemented by Hahsleret al. and described in [136] is used. The parameter `eps_cl` is determined manually per patient. The clustering result is evaluated by visual inspection. Note that in order to

keep the privacy of patients, data is plotted anonymised, i. e., without labelled axis and without map.

Clustering post-processing

A cluster where a patient has stayed less then 30 min over the whole study period is not considered as an important place. Therefore, clusters that contain less than 10 samples are ignored since it means that over the whole study period, a person was less then 30 min at this place. Then, the information as listed in Table 3.5 is added to each sample of the GPS dataset.

3.2.4 Wi-Fi

This section describes the processing of Wi-Fi data.

Wi-Fi cleaning

Data structure. The recorded Wi-Fi signals are stored in a table where a row is resulting from one Wi-Fi scan (i. e., one sample) and each column provides the space for the RSSI values of one access point.

Filtering for home signals. The Wi-Fi signals were analysed with the goal to reveal information about patients' activity at home. Therefore, in addition to the previously described pre-processing step for continuous smartphone sensor signals, the result of GPS analysis (`isHome`) is used to keep only Wi-Fi samples that are recorded at a patient's home. This filtering leads to columns containing only missing values (`NA`). Such columns are deleted.

Filtering out erroneous signals. Examination of the Wi-Fi signals recorded by Android smartphones show periods where the Wi-Fi scan freezes, i. e., it does not update the scan result. This observation is illustrated in Figure 3.6.

These periods with invalid data are filtered out by applying the rule

$$\forall i, 1 \leq i \leq M : RSSI_j^{AP_i} - RSSI_{j+1}^{AP_i} = 0 \implies \texttt{delete}RSSI_{j+1}$$

with M the number of AP.

Table 3.5: Information added to the GPS dataset.

Name	Symbol	Unit	Calculation
Cluster label	cl_i	–	Unique number identifying the cluster a sample ($i = 1, \ldots, M$, with M the number of remaining clusters)
Cluster centroids	c_i	(°,°)	for each cluster cl_i containing points $p_j = (lat_j, long_j)$, $j = 1, \ldots, N_i$, with N_i the number of samples of cl_i, $c_i = (lat_i, long_i)$ with $lat_i = \frac{\sum_{j=1}^{N_i} lat_j}{N_i}$ and $long_i = \frac{\sum_{j=1}^{N_i} long_j}{N_i}$
Home	cl_H, c_H	–, (°,°)	For the palliative care study, the home addresses are available and therefore used to identify the home cluster. For all patients, the home cluster is the largest cluster in terms of number of samples it is containing. Hence, the heuristic "largest cluster" can be used to reliably identify the home cluster cl_H and the home centroid c_H
distHome	D_H	m, km	For each sample p, the distance from home is calculated using the haversine distance [137]: $\mathtt{D}_H = \mathtt{haversine}(p, c_h)$
isHome	`isHome`	–	Each sample is flagged by `TRUE`/`FALSE` if it is at home or not using the rule: $\mathtt{isHome} = \mathtt{D}_H < 50\,\mathrm{m}$. Note that the threshold is the same as the accuracy limit chosen in section 3.2.3.

Figure 3.6: Stuck Wi-Fi signal. The straight lines consist of identical RSSI values for each AP.

Replacing missing values. Since the used methods cannot handle missing values, `NA` values are replaced by the minimum value subtracted by 1, i. e., $\texttt{min}(RSSI_{i,j}) - 1$.

Wi-Fi clustering

The goal of Wi-Fi clustering is to extract situations out of the recorded Wi-Fi data. These situations are influenced by the room a smartphone user is staying in, the position of the smartphone and the position of present persons. Hence, these situations unveil behaviour or behavioural changes of a smartphone user at home.

Toy dataset. For the design of the Wi-Fi clustering algorithm, a test dataset was recorded comprising two places (home and university) with 4 rooms, 6 rooms respectively. The second dataset was recorded on two days with the purpose of having a realistic scenario. In addition to the recording of the phone, the ground truth (in which room the phone is located) was logged manually. Figure 3.7 gives a schematic map for the second data set.

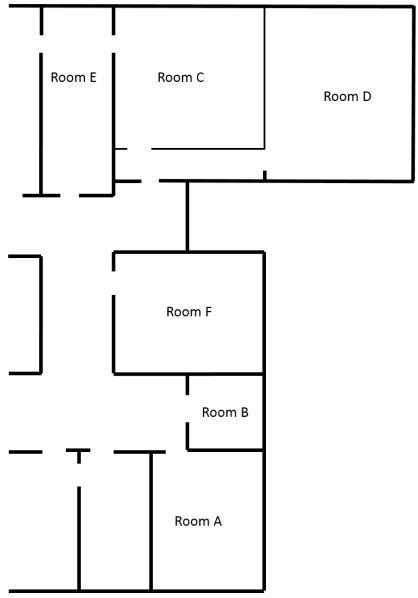


Figure 3.7: Schematic of the rooms that are contained in the datasets.

Feature extraction. For the clustering, on the development dataset as well as on the patient dataset, mean and standard deviation are calculated as features on sliding windows of 6 samples (i. e., 2 min) with an overlap of 5 samples. Note that sliding windows respect data gaps greater than 40 s (2 samples), i. e., sliding windows are calculated on segments between such data gaps.

Clustering. Two clustering algorithms are compared for the situation detection task: GMM (Gaussian Mixed Model) and k-means. The GMM does not require to know the number of clusters in advance. It is run with different parameters and the best model is selected using the BIC (Bayesian Information Criterion).

Result on development dataset. To evaluate the clustering result of the development dataset, the adjusted Rand index [138] is used to measure the overlap of predicted and actual label sets. In addition, two graphical representations show in more details the clustering performance. Figure 3.8 shows the predictions of different clustering methods

and the actual labels over the time for the dataset that contains four rooms. It reveals that part of the mismatch between actual labels and clustering result is at the transition between rooms. One room (Room A), is separated in two rooms, which is explainable by the fact that the phone was for some time in the pocket and for some time on the desk. Furthermore, both, k-means and GMM mix rooms that are adjacent.

Using multi-dimensional scaling (MDS), it is possible to project the high-dimensional room of AP signal strengths on a two-dimensional surface while keeping the euclidean distance, i. e., points with a large euclidean distance in the original vector space will also have a large euclidean distance in the projection. Figure 3.9 shows the MDS of the cluster centroids of the ground truth, GMM and k-means confirms the conclusion drawn from Figure 3.8, showing that the centroids of k-means are closer to the ground truth than the centroids of GMM. Although the clustering is not sufficiently precise in order to distinguish accurately between different rooms, the centroids are close to the ground truth. However, the meaning of the clustering can be extended to situations and the number of clusters can be used as a measure for indoor activity.

3.2.5 Preprocessing of event triggered modalities

The collected files containing the data of questionnaires, call statistics and app statistics are parsed, duplicates are removed and the resulting data structures efficiently stored as data frames.

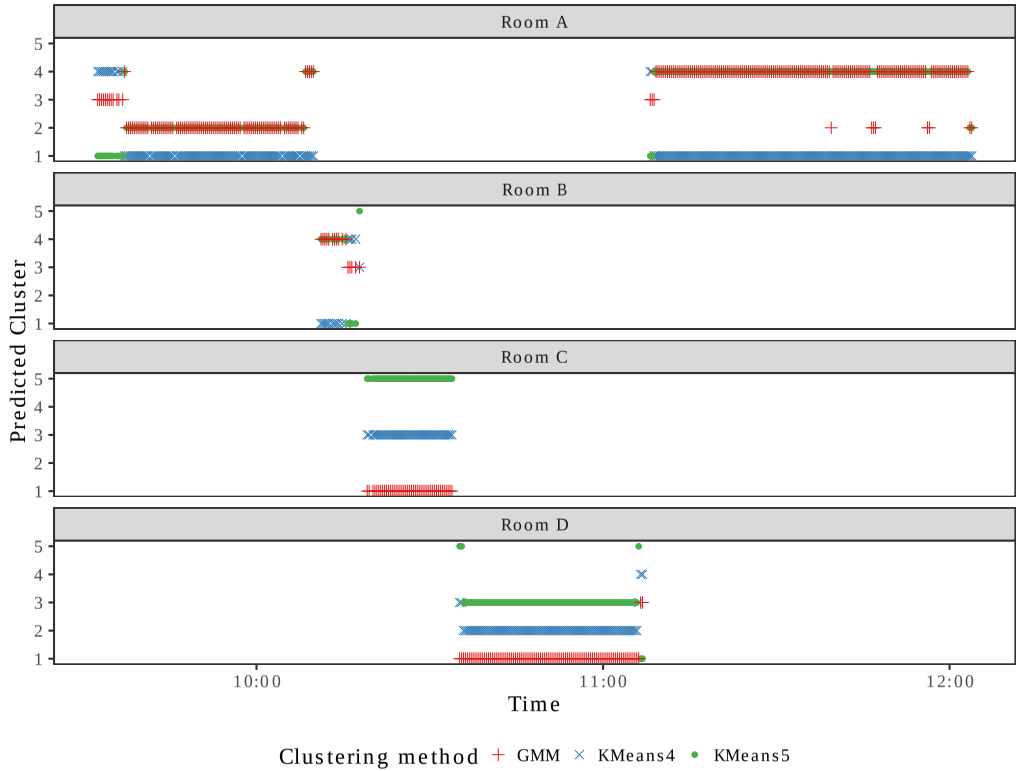


Figure 3.8: Faceted diagram for the time series of cluster labels. Each facet (Room A, Room B, Room C, Room D) symbolises a room according to the ground truth. The dots show the cluster memberships as determined by k-means with k=4 (adjusted RandIndex = .84), with k=5 (adjusted RandIndex = .52) and GMM (adjusted RandIndex = .43).

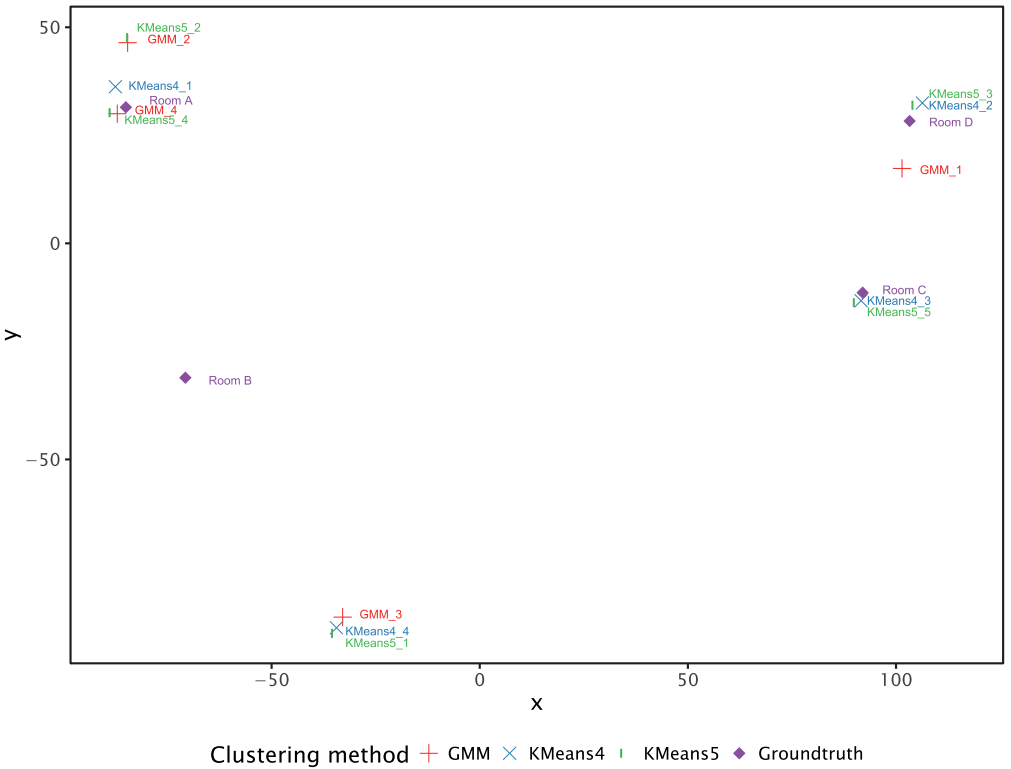


Figure 3.9: MDS of cluster centroids and ground truth.

3.3 Preparation of the Everion® data

As described in Chapter 2, the Everion® sensor was applied only in the palliative care study. The Everion® sensor transmits the measured data in several messages. The content of these messages is described by the following tables. For some measurements, the Everion® delivers also an estimation of the measurement accuracy. Table 3.6 describes the first part of vital measurements based on the PPG sensor as well as values derived from the 3-axis accelerometer. Table 3.7 gives the second part of vital measurements based on the PPG sensor. The measurements listed in Table 3.8 come from physical sensors (temperature, GSR and barometer). Table 3.13 shows the information sent for IPI. Since due to the nature of the signal, IPI is not sampled at a fixed frequency, but the timestamps of the measurements depend on the measurement value, the message format differs from the other messages and the data processing is described separately in section 3.3.4.

3.3.1 Preprocessing: re-sampling and cleaning of Everion®signals

With each connection and reconnection of the Everion®with the smartphone, the internal clock of the Everion® is synchronised. This is necessary because the internal quartz based clock has a limited accuracy and is a bit too fast compared to UTC: from a recorded test data set, a sampling frequency of 1.04 Hz was extracted. The sampling frequency that is higher than the resolution of measurement timestamps of 1 s, lost messages and a wearing time that is less than 24 h per day lead to four types of timestamp-counter anomalies in the data:

1. **duplicates.** There are timestamp duplicates from different measurements, because the actual sampling frequency is about 0.96 Hz.

2. **negative gaps.** Having an Everion® signal sorted by the message counter, there appear negative time gaps due to time synchronisation triggered by the smartphone (changing the sensor time back-wards).

3. **positive gaps.** A continuation (i. e., increment by 1) of the counter with a positive gap in the time > 2 s indicates a discontinuation in the measurement, e. g., because the sensor was not worn.

4. **counter gap.** A gap in the counter values indicates lost messages.

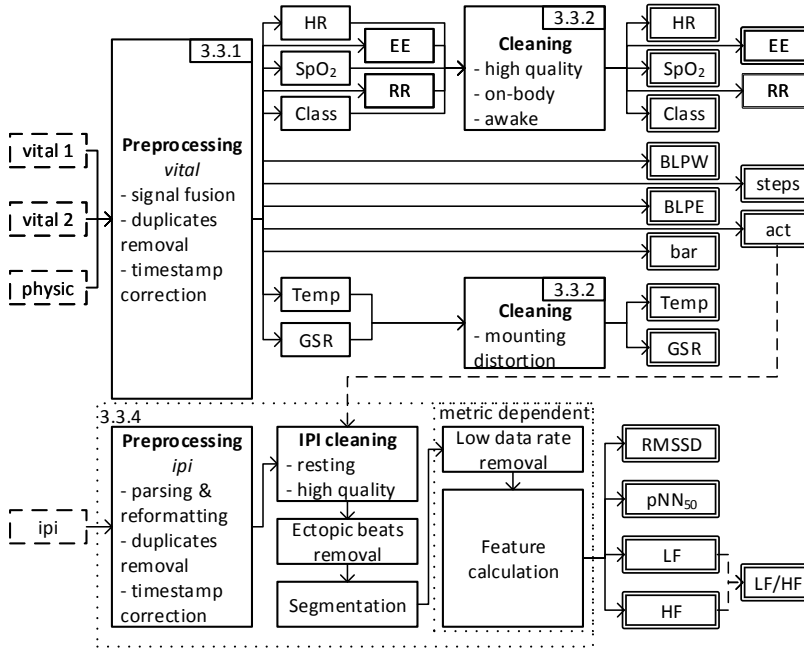


Figure 3.10: Preprocessing of Everion® signals. Dashed lines identify input data, double lines identify the output of the preprocessing.

Table 3.6: Data of vital 1 message as received by the Everion®; Q-Value gives the range of the measurement quality as well as the threshold to accept a measurement as valid. "–": no unit/no Q-value

Value	Unit	Q-value	Description
datetime	s	–	UTC time in ms from 1970-01-01
cnt	–	–	message counter that identifies the true order of the measurements.
HR	bpm	0 to 100 (50)	heart rate
SpO_2	%	0 to 100 (50)	the level of oxygen saturation in the blood
BLPW	–	–	indicator for rhythmic and shape of the blood pulse wave which is related to HRV and stress
BLPE	–	–	skin blood perfusion depends on skin temperature and a good level is necessary for accurate measurements
act	–	–	activity index in the range of 0–255 based on acceleration
class	–	0 to 100 (50)	activity class 0–8 denoting activities, e.g., resting, walking etc.
steps	steps s^{-1}	–	the step count per second

Figure 3.11 illustrates all four anomalies. During preprocessing, duplicate samples are identified by identical message counter and only first values are kept. Note that duplicate timestamps do not necessarily imply duplicate samples. A re-sampling algorithm corrects negative time gaps. A detailed description of the algorithm can be found in section A.3.2.

3.3.2 Cleaning for high measurement quality, on-body data and awake data

This section describes the cleaning procedures for Everion® data.

Table 3.7: Data of vital 2 message as received by the Everion®; Q-Value gives the range of the measurement quality as well as the threshold to accept a measurement as valid.

Value	Unit	Q-value	Description
datetime	s	–	UTC time in ms from 1970-01-01
cnt	–	–	message counter that identifies the true order of the measurements.
HRV	ms	0 to 100 (50)	the RMSSD derived from the HR
RR	bpm	0 to 100 (50)	the number of inhale-exhale cycles per minute
EE	$\mathrm{kcal\,s^{-1}}$	0 to 100 (50)	energy expenditure

Table 3.8: Data of physical message as received by the Everion®; Q-Value gives the range of the measurement quality as well as the threshold to accept a measurement as valid.

Value	Unit	Q-value	Description
datetime	s	–	UTC time in ms from 1970-01-01
cnt	–	–	message counter that identifies the true order of the measurements.
(skin) temperature (temp)	°C	–	skin temperature, after transformation $\frac{Temp}{1000}$
galvanic skin response (GSR)	kΩ	–	Galvanic skin resistance (impedance), after transformation $\frac{Phase \times 256 + Amplitude}{3000}$
bar	mbar	–	Atmospheric air pressure

Validity of samples

HR, SpO_2, Activity class, EE and RR are measurements for which the Everion® itself delivers an estimation of the measurement quality. Based on these Q-values, the samples are tagged with “valid” or

Table 3.9: Data of IPI message as received by the Everion®; Q-Value gives the range of the measurement quality as well as the threshold to accept a measurement as valid.

Value	Unit	Q-value	Description
datetime	s	–	UTC time in ms from 1970-01-01
cnt	–	–	message counter that identifies the true order of the measurements.
inter pulse interval $(IPI)_1$	ms	0 to 15 (8)	first of 39 (in one frame transmitted) interval durations between two pulses
IPI_{39}	ms	0 to 15 (8)	last of max 39 interval durations between two pulses

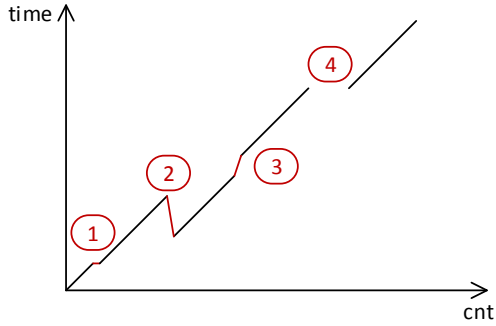


Figure 3.11: Anomalies of counter and timestamps appearing in the recorded Everion® data. 1) duplicate timestamp, 2) negative time gap, 3) positive time gap, 4) positive counter gap

"invalid" according to the rules specified in Table 3.6, Table 3.7 and Table 3.9.

On-/off-body detection from Everion® data

Unfortunately, under certain circumstances (shadow), the Everion® does not recognise when being off-body and continues measuring ("measuring air"). Figure 3.12 gives an example of the Everion® measuring air. The Everion® estimates a part of these measurement values

to be of sufficient quality. Hence, these samples should be filtered out and only measurements coming from a patient should be kept. In such a situation of "measuring air", the Everion® estimates the quality more often to be below the quality threshold than in the situation of being on-body. This ratio of Q-values can be exploited for on-/off-body detection.

Labelled dataset. To clean the dataset from this noise, a SVM is trained on a manually labelled dataset consisting of 4 patients.

Table 3.10: Overview over class distributions: number of samples per patient and label (detailed information to the patients' participation is given by Table 5.1)

	Patient 162	Patient 183	Patient 194	Patient 216
on-body	91129	295977	164535	172806
off-body	10269	902	3192	928
awake	57262	201489	141694	162691
sleep	33839	89437	22820	10101

Feature extraction for on-/off-body detection. For on-/off-body detection, the following statistics are extracted from a sliding window on the measurement signals transmitted by the Everion®: mean, variance, minimum, maximum, first quantile and third quantile as well as energy. The the sliding window has a size of 33 samples with an overlap of 11 samples. Vital signs like heart rate and skin temperature don't change abrupt. Therefore, a high variance in a short window of about 30 s indicates a not physiological signal behaviour and therefore was probably measured off-body. The overlap of 11 samples was chosen in order to reduce the data amount. For the quality Q-values, the ratio of valid values vs. total number samples in the window is used as feature. All features are standardised to mean=0 and sd=1.

On-/off-body classification. Training of different classifiers such as linear SVM, SVM with rbf kernel, adaboost and random forest are

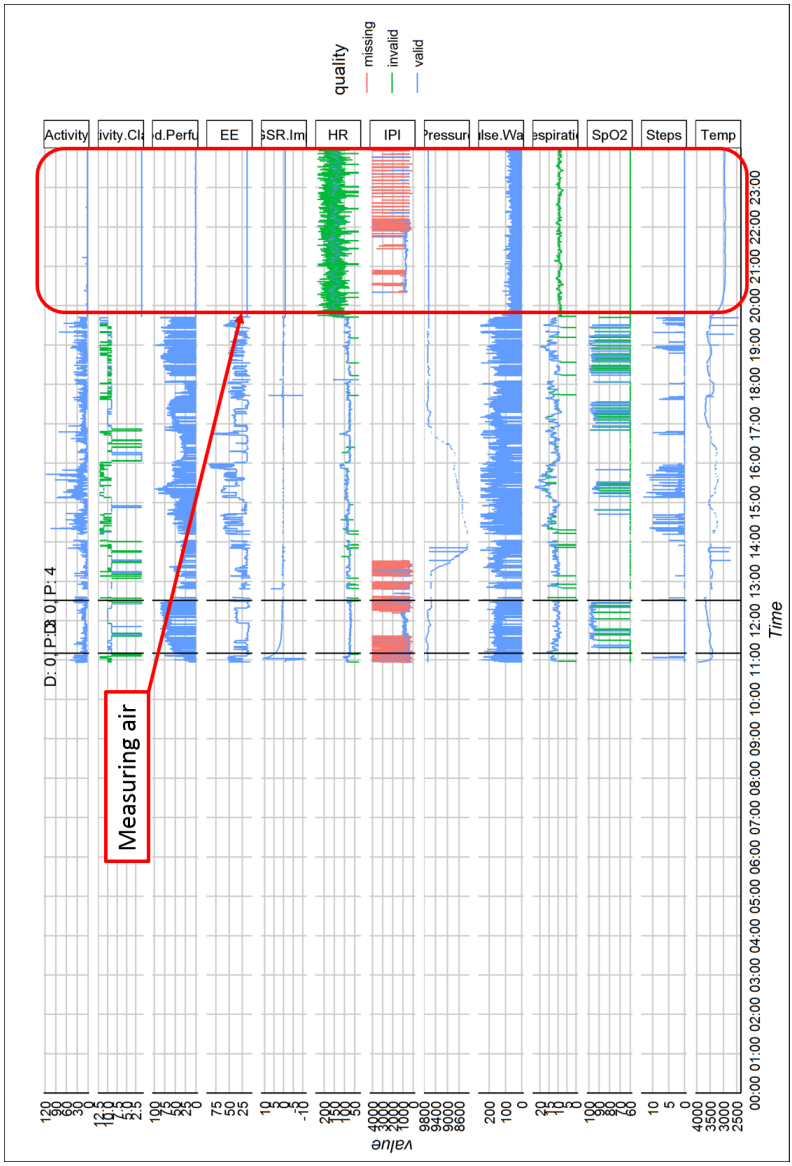


Figure 3.12: One day of test data showing different pattern in measurement values and q-values when measuring air.

compared following a leave-one-patient-out approach. In general, unbalanced classes impede the classification task. Table 3.10 shows that the classes on-/off-body are unbalanced. For each patient, the number of samples labelled as on-body is much higher than the number of samples labelled as off-body. The best performing classifier that generalises is the SVM with rbf kernel with performance metrics as listed in Table 3.11. The classifier achieves a high accuracy. However, the precision is low in three out of four patients.

Table 3.11: Classification result for on-/off-body detection using a SVM with rbf kernel and parameters C=10, gamma=0.01; the metrics precision, recall and F1-score are calculated for the off-body class

	Patient 162	Patient 183	Patient 194	Patient 216
Accuracy	0.94	0.99	0.99	0.99
Precision	0.65	0.62	0.72	0.96
Recall	0.94	0.96	0.86	0.65
F1-score	0.77	0.75	0.78	0.78

Methods for unbalanced class distributions such as Oversampling (SMOTE), random under-sampling, feature selection using Random Forest and manual feature selection based on density distribution plots did not improve the classification performance. Considering the example data and that the difference is visible at a glance, the following reasons could cause the classification result:

- unbalanced classes
- inexact labels

Alternative approaches for on-/off-body detection. Since the labelled dataset show that the off-body data is less than 10 % of the dataset, an alternative method to clean the data set from potential off-body data was applied: a window of 1 min was considered to be invalid if the percentage of invalid samples is greater than 50 %. The reasoning for this approach is the following: the examination of the labelled data set unveiled that the probability is high that the sensor is off-body when the number of invalid samples is greater than 50 %. If not, a maximum of 30 s of valid data is discarded wrongly. Introducing

small data gaps can be tolerated since the analysis methods used in Chapter 5 do not require gap-less time series. Hence, the damage of removing some valid data by mistake is negligible.

Another approach would be to tolerate the presence of off-body data, since the amount is insignificant.

Sleep/awake detection

This section describes an approach to detect periods of sleep during night. The individual daytime of a patient is defined as follows: day starts for an individual patient when he wakes up from night sleep and ends for an individual when going to bed for night sleep. To know which data is recorded during sleep has several advantages:

- features like step count can be set in relation to the time being awake
- vital signs are influenced by sleeping and therefore yield in separate features for recordings during sleep and recordings during awake time
- the awake/sleep time itself is a feature candidate

Dataset. The same labelled dataset as for on-/off-body detection is used for sleep/awake detection. Table 3.10 gives the class distributions.

Feature extraction for sleep/awake detection. The same method is used as for on-/off-body detection. However, feature analysis unveiled that features should individually be selected for each patient.

Classification of sleep/awake status. A SVM with radial basis function (rbf) kernel is trained for each patient separately by taking the first days as training data and the last days as test data. The dataset split for each patient is listed in Table 3.12.

As for on-/off-body detection, the unbalanced classes impede the classification task. Again, accuracy is high, but Recall is not sufficient. Furthermore, this approach requires a partly labelled dataset for each patient. Possible reasons are

- unbalanced classes sleep/awake
- inexact labelling (include sleep during the day)

Table 3.12: Dataset split (patient 4 with only 6 training days with sleep data) and classification result for sleep/awake detection using a SVM with rbf kernel and parameters C=10, gamma=0.01

	Patient 162	Patient 183	Patient 194	Patient 216
Study Days	31	96	35	72
Training Days	7	24	9	56
Accuracy	0.799	0.822	0.82	0.911
Precision	0.75	0.65	0.31	0.61
Recall	0.73	0.90	0.43	0.99
F1-score	0.74	0.75	0.36	0.76

Alternative approaches for sleep/awake. Manual labelling is a tedious and time consuming task, taking up to five minutes per patient day, and re-usable solutions are preferable. Therefore, all samples are tagged with "daytime" or "night" based on the local time of the sample timestamp in this work. For feasibility analysis, a time from 8:00 to 20:00 is considered as day and the time outside of that window as night. For other features, there is no differentiation between day and night since the labelling has shown that patients seldom wear the device while sleeping.

3.3.3 Cleaning of GSR and skin temperature

The GSR sensor and the temperature sensor have an adaption phase: when the Everion® is put on the skin, the signals need some time until they are reliable as shown in Figure 3.14. For example, a daily feature calculated by the minimum value would be always too low since the temperature values start with room temperature when the Everion® sensor is put on. Figure 3.13 illustrates schematically the behaviour of the temperature signal after putting on the sensor.

The cleaning of both signals is done on basis of the temperature signal: the temperature signal is processed segment-wise as defined by time gaps greater than 5 minutes since the temperature signal follows slowly abrupt changes. On each segment, a low-pass Butterworth filter of third order with critical frequency of 0.01 Hz is applied. This filtering step smooths the signal in order to avoid a false detection of a switched sign of the first derivate of the signal due to noise. The temperature

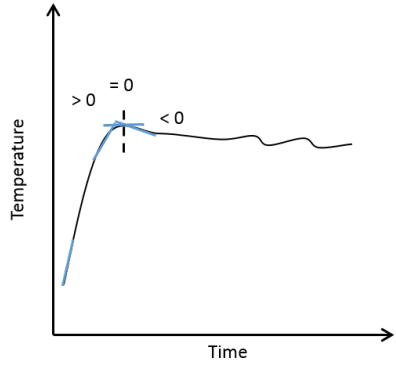


Figure 3.13: Schematic of temperature signal after putting on the sensor. The temperature signal is considered to be valid when the sign of the first derivate of the signal changes. The blue lines denote the first derivate of the temperature signal.

signal is considered as being adapted to the skin temperature, if the first derivate of the filtered signal changes the sign to the opposite of the beginning ramp. A sporadically starting peak as in Figure 3.14 is ignored. The first derivate is illustrated in Figure 3.14 by the red line (diff).

As result, all temperature and GSR values between dashed vertical line and solid vertical line are flagged as “invalid”.

3.3.4 HRV analysis

Cleaning of the inter pulse interval signal IPI

Physical, mental and emotional activity as well as the respiration influences the heart rate variability. As defined by [139], standardised measurements are performed during resting [140]. To follow this recommendation, the activity measurement of the Everion® is used to identify resting periods.

Ectopic beats removal

The presence of abnormal beats influences HRV metrics and therefore have to be removed prior to metrics calculation [83]. Ectopic beats are

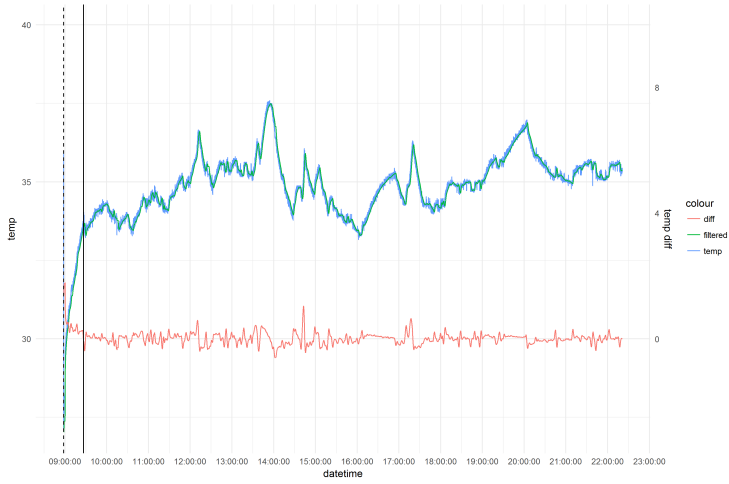


Figure 3.14: Skin temperature measurement showing the time of putting on the sensor (dashed vertical line) and detected end of sensor adaption phase (solid vertical line). Example of a typical day of one patient. temp: temperature, diff: difference (lag 1) of the filtered signal

removed using the algorithm as described in [141], which is able to identify ectopic beats based only on the RR-interval signal with an accuracy of 98% [141].

Segmentation

Standardised window lengths for calculation of HRV metrics are 5 minutes and 24 hours [140]. Due to the limited battery capacity of the Everion® the focus lies on 5 min windows.

Low data rate removal

Since missing values induce errors in the feature calculation, windows with too large data gaps are discarded. The length of tolerated data gaps is defined by a a target error for time domain features ≤ 10%. For frequency domain features, we respect the Nyquist-Shannon sampling theorem: since we analyse the power spectrum for frequencies up to $f_{HF,upper} = 0.4\,\text{Hz}$ (upper limit of HF frequency band), we determine the Nyquist frequency as

$$f_c \geq \frac{1}{\frac{1}{2 * f_{HF,upper}}} = 0.8\,\text{Hz}$$

Therefore, we set the thresholds when to discard a window as follows:

μ, SDNN, RMSSD more than 35 consecutive seconds missing [82]

pNN_{50} more than 5 consecutive seconds missing [82]

LF, HF, LF/HF number of samples $N < 240$

Feature calculation

On the remaining windows, the features as listed in Table 3.13 are calculated.

3.4 Feature extraction

From the above described sensor modalities and processing results, hourly and daily features are extracted. Due to different data quality for the three studies, the feature sets vary. Table 3.14 and Table 3.15

Table 3.13: HRV metrics extracted from the IPI signal.

Feature	Unit	Description
SDNN	ms	Standard deviation of intervals between normal heart beats
RMSSD	ms	Root Mean Sum of Squared Distance of intervals between normal heart beats: indicator for ability of recovery [142]
pNN_{50}	%	% normal intervals that differ more than 50 ms. Indicator for first deterioration.
LF	ms^2	Signal power in the frequency band 0.04 Hz to 0.15 Hz
HF	ms^2	Signal power in the frequency band 0.04 Hz to 0.15 Hz
LF/HF ratio	-	The ratio is commonly used as an indicator for stress, but controversially discussed [140]

give an overview over the extracted features and in which study they are used.

Table 3.14: List of extracted features from the Everion® as used in the following analysis; h/d: hourly or daily feature

Name	h/d	Statistic	Description
RHR	h	mean	resting heart rate
RRR	h	mean	resting respiration rate
RHRV	h	mean	resting HRV (RMSSD)
SpO2.mean	h	mean	SpO2
SpO2.sd	h	sd	SpO2
SpO2.min	h	min	SpO2
SpO2.max	h	max	SpO2
blpw.mean	h	mean	blood pulse wave
blpw.sd	h	sd	blood pulse wave
blpe.mean	h	mean	blood perfusion
blpe.sd	h	sd	blood perfusion
cnt	h	count	nb of samples
cnt.hr	h	count	nb of valid HR
cnt.rr	h	count	nb of valid RR
cnt.spo2	h	count	nb of valid SpO2
step.cnt	h	count	nb of steps
step.speed	h	mean	speed (steps/sec)
step.speed.sd	h	sd	speed (steps/sec)
step.speed.var	h	var	speed (steps/sec)
act.zero	h	count	nb of activity==0
act.mean	h	mean	activity
act.sd	h	sd	activity
act.var	h	var	activity
act.class	h	mode	most frequent activity class
cor.p	h	cor	Pearson correlation between HR and act
cor.k	h	cor	Kendall correlation between HR and act
cor.s	h	cor	Spearman correlation between HR and act
t.max	h	max	skin temperature
t.min	h	min	skin temperature
t.range	h	max-min	skin temperature
t.mean	h	mean	skin temperature
t.sd	h	sd	skin temperature
gsr.mean	h	mean	GSR impedance
gsr.sd	h	sd	GSR impedance
gsr.min	h	min	GSR impedance
gsr.max	h	max	GSR impedance
t.min.day	d	min	skin temperature
t.max.day	d	max	skin temperature
t.range.day	d	max-min	skin temperature
ipi.mean_mean	h	mean	ipi mean (HR)
ipi.std_mean	h	mean	ipi sd on 5 min. window
ipi.pnn50_mean	h	mean	PNN50
ipi.lf_mean	h	mean	LF
ipi.hf_mean	h	mean	HF
ectopic_mean	h	mean	nb of ectopic beats
lf_hf_ratio_mean	h	mean	LF/HF
ipi.mean_sd	h	sd	ipi mean (HR)
ipi.std_sd	h	sd	ipi sd on 5 min. window
ipi.pnn50_sd	h	sd	PNN50
ipi.lf_sd	h	sd	LF
ipi.hf_sd	h	sd	HF
ectopic_sd	h	sd	nb of ectopic beats
lf_hf_ratio_sd	h	sd	LF/HF
ipi.mean_cnt	h	count	ipi mean (HR)
ipi.std_cnt	h	count	ipi sd on 5 min. window
ipi.pnn50_cnt	h	count	PNN50
ipi.lf_cnt	h	count	LF
ipi.hf_cnt	h	count	HF
ectopic_cnt	h	count	nb of ectopic beats
lf_hf_ratio_cnt	h	count	LF/HF

Table 3.15: List of extracted features from the smartphone as used in the following analysis; h/d: hourly or daily feature; y: used in the study

Name	h/d	Statistic	Description
maxDistFromHome	h	max	maximal distance from home
sumDistOutside	h	sum	sum of distance covered outside
sumDistHome	h	sum	sum of distance covered at home
cntHome	h	count	nb samples at home
cntOutside	h	count	nb samples outside
cntCluster	h	count	nb cluster
cntLeavingHome	h	count	nb of times leaving home
cntComingHome	h	count	nb of times coming home
firstLeavingHome	d	min	time of first leaving home
lastComingHome	d	max	time of last coming home
cntHomeDay	d	count	nb samples at home
cntLeavingHomeDay	d	count	nb of times leaving home
cntComingHomeDay	d	count	nb of times coming home
firstLeavingHomeDay	d	min	time of first leaving home
lastComingHomeDay	d	max	time of last coming home
maxDistFromHomeDay	d	max	maximal distance from home
sumDistOutsideDay	d	sum	sum of distance covered outside
cntOutsideDay	d	count	nb samples outside
cntClusterDay	d	count	nb cluster
acc.cnt	h	count	nb samples acceleration
cntOnbodyValid	h	count	nb samples on-body
onbodypercent	h	perc	percentage wearing the smartphone on-body
cntWalkingValid	h	count	nb samples walking
walkingpercent	h	perc	percentage of walking
acc.mean	h	mean	norm of acceleration
acc.sd	h	sd	norm of acceleration
acc.var	h	var	norm of acceleration
acc.q10	h	10% quantile	norm of acceleration
acc.q25	h	25% quantile	norm of acceleration
acc.q75	h	75% quantile	norm of acceleration
acc.q90	h	90% quantile	norm of acceleration
moving	h	count	nb samples while moving
movingPercent	h	perc	percentage of moving
acc.steps.speed	h	mean	smartphone steps speed (steps/sec)
acc12	d	count	nb samples acceleration
cntOnbodyValidDay	d	count	nb samples on-body
onbodypercentDay	d	perc	percentage wearing the smartphone on-body
cntWalkingValidDay	d	count	nb samples walking
walkingpercentDay	d	perc	percentage of walking
acc.meanDay	d	mean	norm of acceleration
acc.sdDay	d	sd	norm of acceleration
acc.varDay	d	var	norm of acceleration
acc.q10Day	d	10% quantile	norm of acceleration
acc.q25Day	d	25% quantile	norm of acceleration
acc.q75Day	d	75% quantile	norm of acceleration
acc.q90Day	d	90% quantile	norm of acceleration
movingDay	d	count	nb samples while moving
movingPercentDay	d	perc	percentage of moving
acc.steps.speedDay	d	mean	smartphone steps speed (steps/sec)
cnt.total.in	d	count	nb of incoming calls
cnt.total.out	d	count	nb of outgoing calls
cnt.succ.in	d	count	nb of answered incoming calls
cnt.succ.out	d	count	nb of answered outgoing calls
sum.succ.in	d	sum	duration of incoming calls
sum.succ.out	d	sum	duration of outgoing calls
min.succ.in	d	min	duration of shortest incoming call
min.succ.out	d	min	duration of shortest outgoing call
max.succ.in	d	max	duration of longest incoming call
max.succ.out	d	max	duration of longest outgoing call
avg.succ.in	d	mean	duration of longest incoming call
avg.succ.out	d	mean	duration of longest outgoing call

4

Pre-study: Monitoring cancer related fatigue

This chapter presents the results of an observational study, conducted in Zurich, of cancer survivors suffering from cancer related fatigue.

4.1 Fatigue study overview

4.1.1 Goals of fatigue study

As a pre-study, the fatigue study was conducted with the following goals:

- to test the usability of the activity monitoring app
- to gain experience in activity monitoring severely-ill patients
- to explore the collected data to gain insights into the daily life of the patient group

4.1.2 Summary of study procedure

As already described in section 2.3, a study of 7 cancer survivors suffering from CRF was conducted in collaboration with the University Hospital Zürich. Adults with a mid-to-severe fatigue and who were not suffering from depression were asked if they would carry and use a smartphone for 2 to 3 weeks to record the sensor data, as described in Table 2.1. Furthermore, they were asked to answer digital questionnaires four times per day. They completed two VAS scales about their currently perceived fatigue, and how much this fatigue currently affected their daily life and the currently performed activity (ref. Figure 2.3). At the beginning (baseline) and after each week of study participation, patients were called to assess their level of fatigue using the FACIT-F questionnaire (ref section A.2.8). They were also interviewed about their experiences with the smartphone.

4.2 Study cohort

In total, 7 cancer survivors between the ages of 32 and 73 years (median: 54, sd: 15.77, 6 female) were included in the study. The study participants are listed in Table 4.1.

4.3 Collected data and feasibility

As listed in Table 4.2, the cancer survivors participated for 2 to 3 weeks. Patient 66 dropped out as a consequence of unstable health. She is excluded from further data analysis. During in total 124 days, 2772 hours of data and 523 answered questionnaires were collected.

Table 4.1: Overview on the included study participants. BL: baseline **: value after week 1; *: drop out

Pat. Id	Age [y]	Gender	FACIT-F (BL)
103	32	f	22**
64	58	m	35
219	73	f	23
62	37	f	24
111	33	f	23
232	52	f	29
66*	56	f	24

Table 4.2: Overview on data completeness. Days: number of study days; DwQ: percentage of days with answered digital questionnaires; QpD: average number of questionnaires per day; DwS: percentage of days with smartphone on-body; HSO: average percentage of time when the smartphone was on-body during daytime

Pat. Id	Days	No. questionnaires	DwQ	QpD	HwS	DwS	HSO (WI)
103	32	127	90.6	4	739	90.6	25
64	16	53	87.5	3.3	342	100	33
219	31	121	96.8	3.9	735	100	43
62	15	61	93.3	4.1	324	100	42
111	15	62	86.7	4.1	301	100	38
232	15	99	93.3	6.6	331	100	48

4.3.1 Wearing the smartphone

The smartphone's acceleration signal is analysed to determine for how long a patient was wearing the smartphone (on the body). As described in section 3.2.2, during the pre-processing of the sensor signals, the smartphone's acceleration data was tagged with a label indicating weather the smartphone was worn on-body during the recording. The wear-index (WI) is calculated as the ratio between on-body samples (i. e., measurement points) and the total number of measurement samples:

$$WI = \frac{\#\ on\text{-}body\ samples}{\#\ on\text{-}body\ samples + \#\ off\text{-}body\ samples}$$

Table 4.2 shows that all of the patients were wearing the smartphone every day. However, the percentage of time per day that patients wore the smartphone on-body is lower than 50 %. The lowest values are explained as follows:

- Patient 64 did not follow the instruction to wear the phone also at home.
- Patients 103 and 111 both reported that they were very tired all of the time and were lying down for large parts of the day. When they were resting, they did not wear the smartphone.

A WI of 100 % means that the smartphone was worn on-body for 24 h. Naturally, this is hard to achieve outside of a laboratory, as it means the smartphone should be on the body during activities such as sleeping and showering. Estimating 8 h of sleep and additional time spent on personal hygiene leads to a WI of about 60 % to 70 %.

4.3.2 Answering the questionnaires

All of the patients except two filled out the questionnaires (ref. section 2.3.2) at least 4 times per day for around 90 % of the days. Patient 66 who dropped out stated that she did not want to reflect on her fatigue four times per day or being constantly reminded of her situation. Since fatigue is a sensitive topic for her, she found it difficult to cope with her situation when she was asked about it several times per day.

4.4 Analysis of self-reports

4.4.1 FACIT-F

Figure 4.1 shows that four out of six patients had improved FACIT-F values at the end of the study.

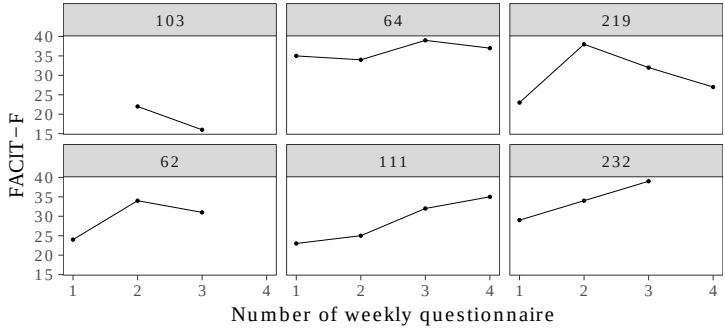


Figure 4.1: The course of FACIT-F reports for all patients.

4.4.2 Digital questionnaires: fatigue and interference

Distribution of VAS values

Figure 4.2 shows the histogram of the reported fatigue on a VAS scale. Peaks are visible for the even numbers (2, 4, 6, 8); in each case, the adjacent odd numbers were selected less often compared to the even numbers. In this pre-study, the VAS scales were labelled only with even numbers, as shown in Figure 2.3 a). Omitting odd numbers possibly created a bias towards even numbers.

Relation between fatigue and interference of fatigue

The two VAS scales for fatigue and interference are positively correlated ($r = .77$, $n = 523$, $p < .0001$). Two possible explanations are

1. The higher the fatigue, the more it impacts on daily activities. If one is less tired, it is probably easier to ignore one's lack of energy during activities of daily life.

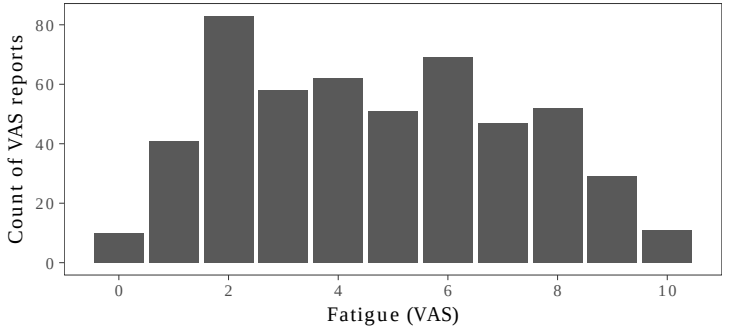


Figure 4.2: Histogram of Fatigue (VAS) shows the count of fatigue values reported by all patients through the smartphone VAS. Peaks for the even numbers 2, 4, 6 and 8 indicate a bias through the app design (ref. Figure 2.3).

2. The patients could not distinguish between the two questionnaires.

Figure 4.3 shows the reported fatigue with the corresponding reported interference per patient. For each reported pair, the dot size indicates the number of occurrences of this pair in the dataset. Only for values 1 and 2, there is an accumulation of same answers. For the other values, the diagonal is not over-represented except for patient 64, who had low fatigue and interference (2) or high fatigue and interference (9) and only a few values in between. Hence, Figure 4.3 supports explanation no. 1 (i. e., the higher the fatigue, the more it is perceived as having an impact on daily life).

Intra-day course of fatigue

The ESM was used for collecting the digital self-reports. ESM unveiled that fatigue is not constant throughout the day. All self-reports were grouped according to their timestamp in order to identify patterns in the daily course of fatigue. The day was divided into six time frames:

late night 2-6 o'clock

morning 6-10 o'clock

midday 10-14 o'clock

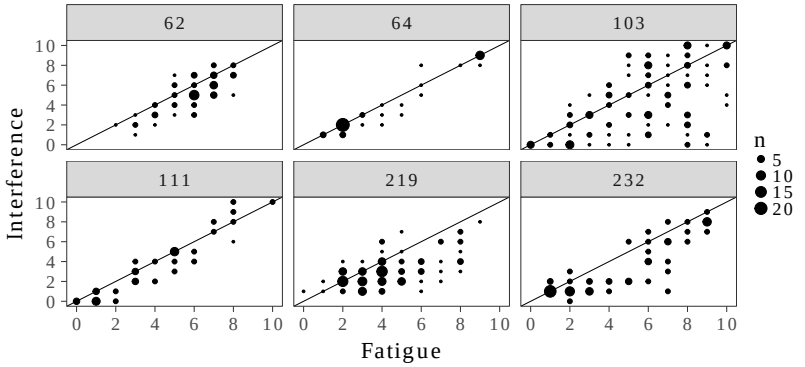


Figure 4.3: Scatter plot of all digitial questionnaires per patient concerning fatigue vs. interference: The dot size represents the count of a VAS value pair. The diagonal denotes where fatigue is equal to interference.

afternoon 14-18 o'clock

evening 18-22 o'clock

night 22-2 o'clock

Naturally, there are few values available for night time (seven reports in total) and late night time (one report). Figure 4.4 shows that the perceived fatigue varies during the day. In this study, there are different patterns present: In two patients, the fatigue increases during the day; one patient has a decrease towards midday, followed by an increase towards the evening that is close to the morning value, and three patients have a constant median throughout the day, but with a varying inter quartile range (IQR).

4.4.3 Correlation between VAS and FACIT-F

To compare the paper-based questionnaires with the VAS reports for fatigue and interference, the median of all the answered VAS for one week is compared to the FACIT-F value at the end of the same week.

Figure 4.5 shows that for most patients, both measurement instruments show similar trends. Only for patient 62 did the VAS remain constant while the FACIT decreased (i. e., increasing fatigue).

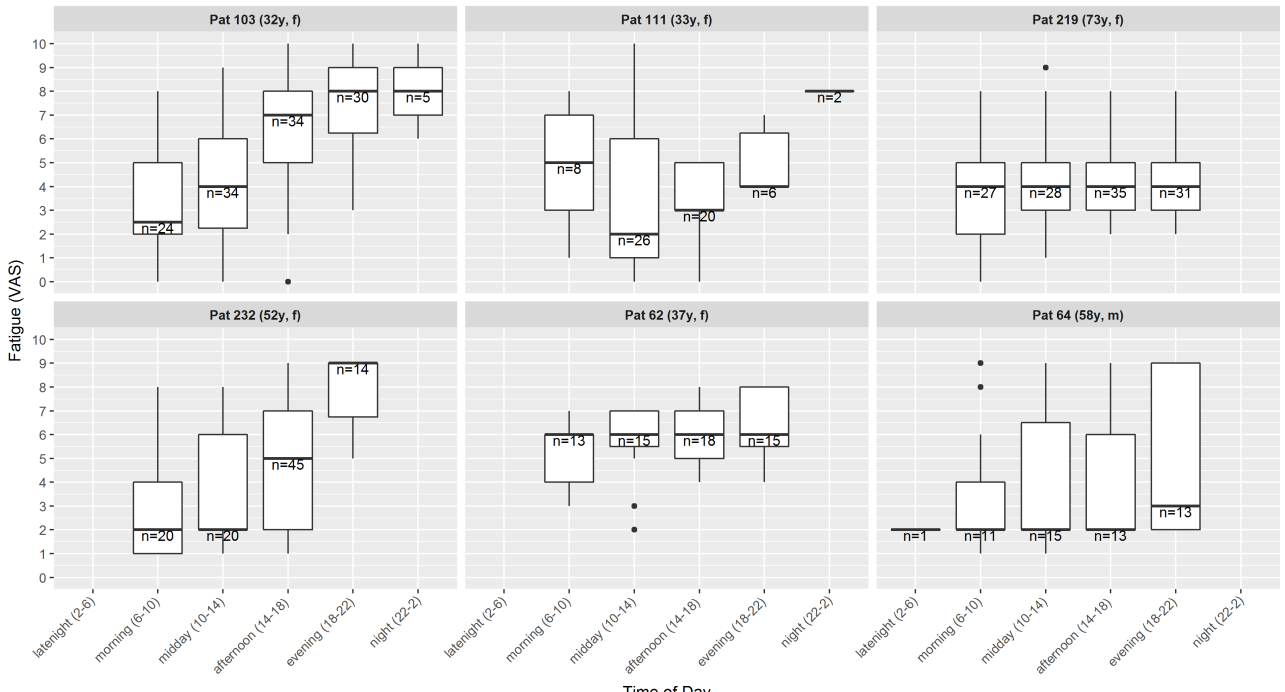


Figure 4.4: Box plot of perceived fatigue values per time at day shows intra-day variance. The sample size is denoted in each box by *n*. (Figure taken from [125])

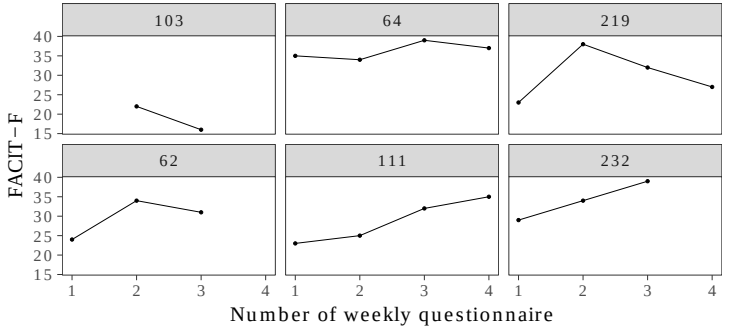


Figure 4.5: The course of all patients' FACIT-F reports over the study period.

4.5 Correlations between sensor modalities and self-reports

Different features, based on the recorded acceleration signal and GPS signal of the smartphone, were tested for correlations with the VAS values concerning fatigue and the interference of fatigue. However, there was no feature that correlated similarly in several patients. Possible reasons for low correlations include:

- small amount of valid activity data at home
- the use of an additional smartphone is different to how one uses one's own smartphone
- cause and effect remain unknown: a patient could be active, and they could feel tired afterwards because of this activity; in contrast to this, a patient might not be active because they feel tired
- depending on external requirements, some patients may have to manage their daily lives, regardless of their degree of fatigue

4.6 Discussion

4.6.1 Summary of results

This pre-study showed that adult patients suffering from medium-to-severe CRF are eligible for activity monitoring through a smartphone.

The patients' self-reporting via their smartphones was successful; they reached reporting rates of roughly 90 %. However, wearing the phone on-body at home was noted as being inconvenient or meaning too much effort for some patients.

In assessing perceived fatigue, experience-based sampling unveiled information about the general course of fatigue: The rate of perceived fatiuge varies throughout the course of the day. The patients of this study had three different patterns. Furthermore, the VAS scales showed similar trends to the weekly questionnaires.

4.6.2 Limitations

The pool from which to recruit patients was limited for this study. The sample size of seven patients was too small as a result and general conclusions could not be drawn. The study period of 2-3 weeks is not sufficient to evaluate retention for longer periods. A longer time frame is needed.

4.6.3 Lessons applied to the palliative care study

Experience gained through the fatigue study influenced the palliative care study in the following ways:

- The VAS scales on the smartphone were changed to show every possible whole number in the range from 0 to 10, in order to avoid a bias towards even numbers.
- The smartphone belt is considered a crucial element to increase smartphone wearability, and is, therefore, included in each patient's kit by default.
- Patient meetings are planned with time-slots of one hour plus an additional buffer of one hour for any unforeseen delays.

5

Monitoring cancer patients in palliative care

This chapter presents the results of an observational study performed on 31 cancer patients in ambulatory palliative care.

5.1 Overview of the palliative care study

5.1.1 Goals of the palliative care study

The palliative care study was designed as an observational study with the following goals:

- to evaluate feasibility of activity monitoring in this specific patient group
- to evaluate patients' acceptance and usability of the wearable devices
- to conduct a descriptive data analysis of the patient-reported outcomes
- to conduct an exploratory data analysis of recorded sensor signals and non-elective (i. e., unplanned) hospital visits and readmissions

5.1.2 Summary of study procedure

As already described in section 2.4, a study with 31 outpatients in palliative care was conducted in collaboration with the University Hospital Zürich. Patients selected for this study were adults with metastatic cancer who were given between 8 weeks and 12 months to live and who were mobile for at least 50 % of their wake-time. They were asked if they would carry and use a smartphone and wear the Everion® for 12 weeks in order to record the sensor data as described in Table 2.1 and 2.2. Furthermore, they were asked to answer digital questionnaires once per day. These questionnaires comprised two VAS scales, where patients could record their current pain and distress (ref. Figure 2.2). At the beginning (baseline) and after each week of study participation, patients were called to assess their quality of life using the validated questionnaire QLQ-C30 (ref section A.2.2). They were also interviewed about their experiences with the smartphone.

5.1.3 Performed analysis

This chapter presents the results of several analyses. Figure 5.1 outlines the topics in this chapter. Section 5.2.1 describes the patients' demographics. Section 5.3 presents the analysis and results concerning the

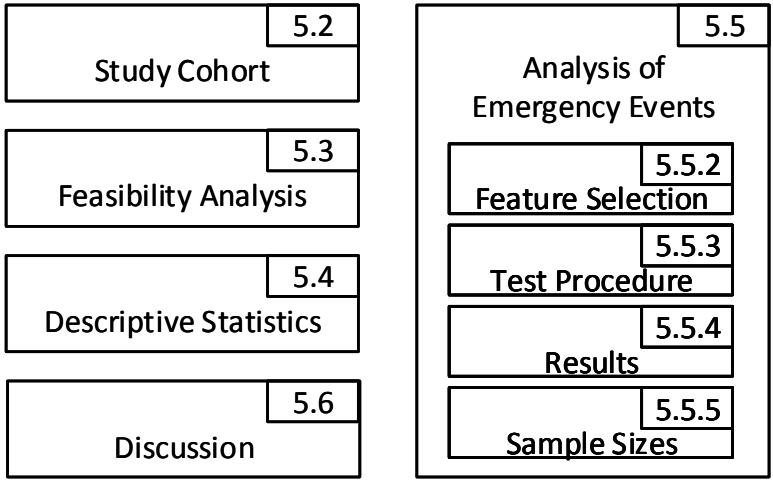


Figure 5.1: Graphical overview on the analysis of the palliative care study.

general feasibiltiy of the activity monitoring as well as the patients' acceptance of it. Section 5.4 presents the outcomes of the patients' self-reports. Furthermore, an approach to predict emergency events using anomaly detection is presented in the appendix in section A.6.

5.2 Study cohort

5.2.1 Study cohort

In total, 31 patients, between the ages of 39 and 85 years (mean: 61.9, sd: 11.4, 9 female), were included in the study. Table 5.1 gives an overview of the patients. In total, 11 patients had at least one non-elective (i. e., unplanned) visit to hospital (ambulatory or stationary), 5 patients passed away, either during their study period or up to one month after finishing the study., and 8 patients were using a walking aid, i.e., a pole, crutches, a walking frame or sitting in a wheel chair.

Patient 61's wife stated that her husband was fearful when discharged from hospital and apathetic when at home. He did not wear

Table 5.1: Overview of included patients. Gender: m – male, f – female; Age: in years; Karnofsky: in %; SDN: number of study days; EV: day of first emergency visit (0: no emergency visit); Death: 1 – patient passed away; Aid: 1 – patient used a walking aid; †: passed away during participation or shortly after.

ID	Gender	Age	ECOG	Karnofsky	SDN	EV	Death	Aid
23	m	61	2	70	106	0	0	0
55	m	66	1	80	104	2	0	0
61	m	61	1	90	NA	0	1	0
69	m	73	1	80	128	0	0	0
128	f	49	2	70	140	0	0	0
142	m	77	1	80	80	0	0	0
162†	f	39	1	NA	94	5	1	0
168	f	69	1	80	108	0	0	0
192	f	57	1	80	105	0	0	0
183	m	63	2	60	97	12	0	1
187	m	52	1	90	88	0	0	0
194†	m	71	1	80	148	6	1	0
193	m	78	1	80	90	0	0	0
165	m	50	2	60	38	0	0	0
216†	m	81	2	60	75	8	1	1
246	m	50	1	NA	94	11	0	0
271†	m	68	2	NA	85	0	0	1
277	m	51	0	NA	94	0	0	0
319	m	65	2	60	105	0	0	1
324	m	76	2	60	34	0	0	1
351	m	53	2	60	85	4	0	0
340	f	65	2	60	29	2	0	1
359	m	55	0	90	97	0	0	0
355	f	66	2	60	95	0	0	0
377	f	41	1	80	97	0	0	0
405	m	60	1	70	85	6	0	1
423	f	85	1	80	85	75	0	0
424	m	62	1	80	35	0	0	0
429	m	60	0	NA	85	0	0	0
430	f	50	2	NA	85	8	0	0
438	m	64	2	50	83	0	1	1

the study devices and passed away a few days after discharge. Nevertheless, his wife stated that he was happy and looking forward to participating in the study. He is excluded from further analysis.

5.3 Collected data and feasibility

After preprocessing the data, as described in chapter 3, the data is analysed with respect to its completeness. The collected amount of valid data is different for the devices and is summarised in Table 5.2: the analysis evaluated three tasks wearing the smartphone, wearing the Everion® and answering the digital questionnaires:

answering the digital questionnaires: the percentage of days at which the questionnaires were answered (DwQ) was calculated as well as the average number of questionnaires answered per day of the days with at least one questionnaire answer (QpD).

wearing the Everion®: the percentage of days at which the Everion® was worn (DwE) was calculated as well as the average percentages of wearing time of the Everion® during the day (HwED) and night (HwEN).

wearing the smartphone: the percentage of days at which the smartphone was worn (DwS) was calculated as well as the average percentage of wearing time of the smartphone during the day.

The following sections investigate he extent to which the patients wore the devices. The completeness of the data from the questionnaires is also investigated.

5.3.1 Wearing the Everion®

Measurement quality of the Everion®

Figure 5.2 shows the quality of the data collected by the Everion® for a patient's typical day. The y-axis denotes the time from midnight to midnight. The x-axis denotes the available measurements of the sensor. Figure 5.2 shows that patient 23 either woke up during night or went to bed around 1 o'clock. Indeed, in the weekly interview, the patient stated that when he was wearing the sensor at night-time, he was woken up by a smartphone notification, which asked him to recharge the sensor. (The patients were advised to recharge the sensor before

Table 5.2: Overview of data completeness as percentages. Days: number of days being either at home or in rehabilitation; DwQ: percentage of days with answered digital questionnaire; QpD: average number of questionnaires per day; DwE: percentage of days with Everion® data; HwED: percentage of Everion® wearing time in hours during day; HwEN: Everion® wearing time in hours during night; StS: 1 – study smartphone, *: exclusively; DwS: percentage of days with smartphone on-body; HSO: avg. percentage of smartphone on-body during daytime (wear-index); †: passed away.

ID	Days	DwQ	QpD	DwE	HwED	HwEN	StS	DwS	HSO
23	106	83.96	1.71	56.6	70.9	1.33	1;0*	89.62	48.39
55	94	72.34	1.69	44.69	72.3	2.04	1	89.36	74.46
69	127	98.43	3.74	94.49	87.39	1.35	1*	100	74.21
128	125	86.4	1.2	69.6	54.68	0.76	1*	88.8	72.44
142	79	96.2	5.68	81.01	51.12	1.59	1*	98.73	41.89
162†	58	25.86	1.07	25.86	62.23	1.58	1	20.69	3.3
168	102	65.69	2.55	14.71	58.76	0.3	1	69.61	6.47
192	104	80.77	1.08	5.77	76.66	0	1*	93.27	17.53
183	96	90.62	1.52	62.5	78.37	2.08	1;0*	93.75	46.01
187	84	92.86	1.64	93.75	84.56	0.16	1*	100	78.02
194†	132	36.36	1.29	21.21	76.3	0	1	45.45	58.19
193	91	83.52	2.05	55.56	60.85	1.06	1*	90	36.34
165	36	80.56	1.48	77.78	49.49	0.78	1*	94.44	41.24
216†	67	43.28	1.34	19.12	45.6	0	1*	67.65	64.25
246	76	96.05	1.11	64.47	69.97	0.22	1	94.74	9.48
271†	85	96.47	1.37	68.24	79.19	0.27	1*	97.65	68.29
277	96	85.42	1.54	92.55	80.7	0.05	1	100	84.68
319	104	69.23	1.1	50.96	37.57	1.79	1*	89.42	83.94
324	33	45.45	1.67	33.33	65.05	0.95	1	54.55	35.65
351	61	96.72	1.08	67.21	58.3	1.52	0*	100	38.69
340	23	69.57	2.56	41.18	61.72	0.82	1	100	37.47
359	96	70.83	2.01	67.71	81.54	0	1	79.17	76.37
355	94	86.17	1.48	92.55	54.51	2.37	0*	95.74	20.4
377	80	83.75	1.21	85	55.93	2.12	1	92.5	28.32
405	71	49.3	1.03	43.94	58.16	0	0* ;1*	93.94	40.22
423	76	11.84	1.22	10.53	52.2	0.32	1	51.32	73.35
424	20	90	1.17	15	23.88	0	0*	100	29.52
429	84	92.86	1.53	63.1	83.93	0	1*	98.81	74.55
430	46	82.61	1.16	15.22	20.99	1.14	1	86.96	58.88
438†	87	56.32	1.08	78.31	62	2.23	1	84.34	63.06

going to bed.) He put on the sensor in the morning around 9 o'clock. Before 20:00 o'clock, he took the sensor off.

Patients cannot actively influence the measurement quality of the Everion® (except for correct placement), since the measurement quality depends on many factors, such as skin colour, arm movement and blood perfusion. In general, the data quality is high throughout the day, except for the measurement of SpO_2, which is more sensitive to artifacts introduced by movements of the arm. Only data labelled as valid is used for the analysis in section 5.5.

Retention (Everion®)

In total, 16 587.43 h data were collected by the arm-sensors. The bars in Figure 5.3 show the number of patients per day that were not hospitalised or in vacation. The black line indicates the number of patients in the study, and the red line indicates the percentage of patients that have wore the sensor at a day. In total, 12 patients wore the device at less than 50 % of the days. The patients' main reasons for not wearing the device included forgetting to put it on, their bad health, the fact that they were in rehabilitation and some technical issues (mainly instable Bluetooth connection with some devices, ref. section 5.3.5). The other 18 patients wore the device for 73.41 % of the days of the study.

The blue and green lines show the percentages of data collected during the day (8:00 to 20:00), with a 62.5 % average for all patients, and night hours (20:00 to 8:00) with 1.7 % average for all patients. The majority of patients did not wear the device during night. The blue and green lines show that the wearing time was stable over the entire study period.

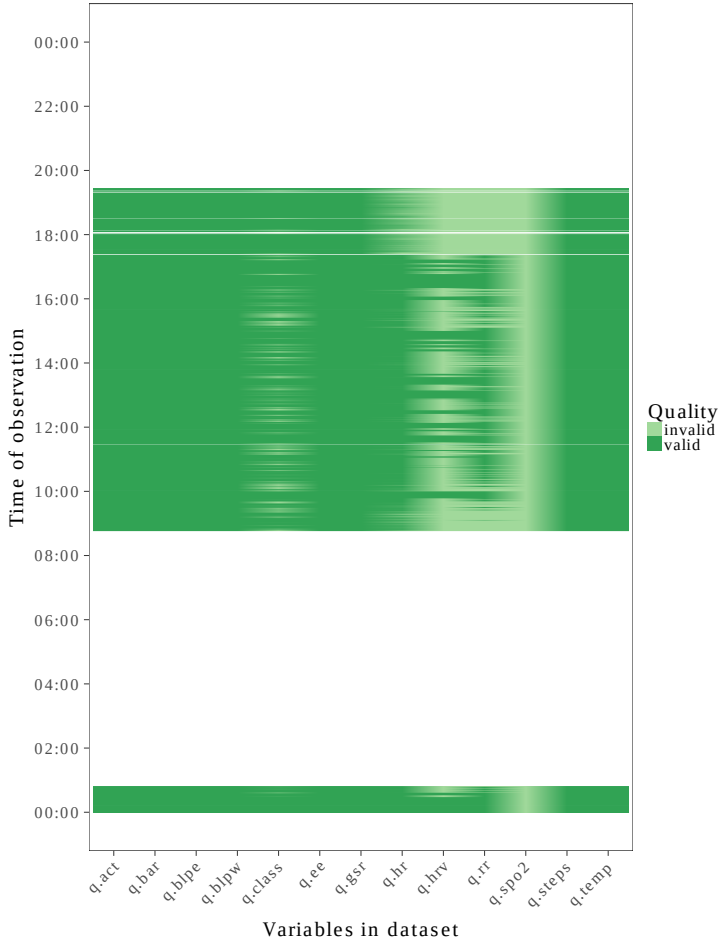


Figure 5.2: Data completeness plot for a typical day of patient 23. Valid data is shown in dark green and invalid (i.e., of low quality) data is shown in light green. White denotes periods without any data (missing values). The prefix denotes the quality value, not the measurement value. The abbreviated names of the measurement values are explained in Tables 3.6, 3.7 and 3.8.

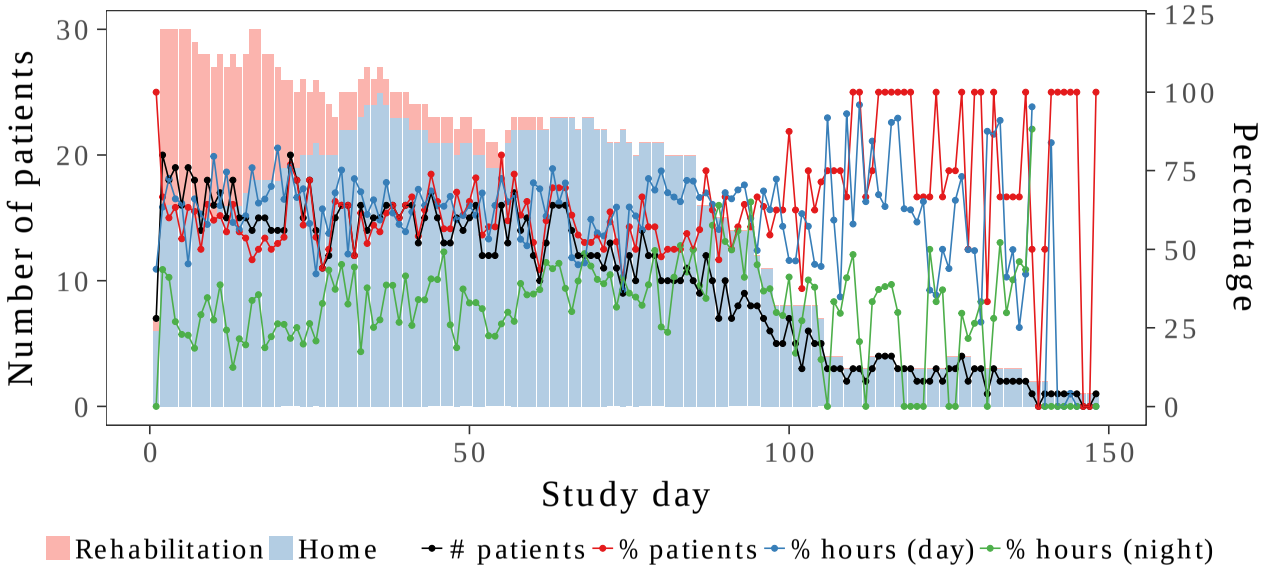


Figure 5.3: Rate of wearing the Everion®. The bars display the number of patients at home or in rehabilitation for each day. The black line denotes the number of patients wearing the Everion® at a given day, the red line denotes the percentage of patients wearing the Everion® at a given day. The two other coloured lines (for day and night) show that the wearing percentage was stable over the study period. (Note: Due to breaks during study participation, the study days count up to 148.)

5.3.2 Wearing the smartphone

For a reliable analysis of the physical activity measured by the smartphone, one has to take into account that the smartphone is not on-body all the time (e. g., when indoors, one does not usually keep the smartphone on-body). Off-body time does not provide information, apart of GPS (under the assumption that the patients take the smartphone with them whenever they leave the house). In this case, an off-body smartphone still provides information about where the patient was staying during the off-body time. Only one patient (424) admitted that she is forgetful and often left the devices (smartphone and Everion®) at home, because she was not used to having a smartphone. In order to make the wearing of a smartphone at home more convenient, patients received a smartphone belt that is popular in running sports because it is small, unobtrusive and does not move easily, even with high - sports. Nevertheless, for some patients, it was too cumbersome to wear the smartphone while staying at home. As swown in Table 5.2, column HSO, patients wore the smartphone during the day (8:00 to 20:00) on average 49.52 % of the time at on average 85.35 % of the days (column DwS).

Figure 5.4 shows that there is no significant difference in the WI between patients who used the smartphone as their primary smartphone or as their secondary smartphone during their participation in the study. Thus, an additional (secondary) smartphone fulfils the purpose of logging sensor data just as efficiently as a primary smartphone does.

Retention (smartphone)

Similar to the retention analysis for the Everion®, Figure 5.5 shows that patients wore the smartphone on a constant level around 50 % between 8:00 and 20:00 over the whole study period.

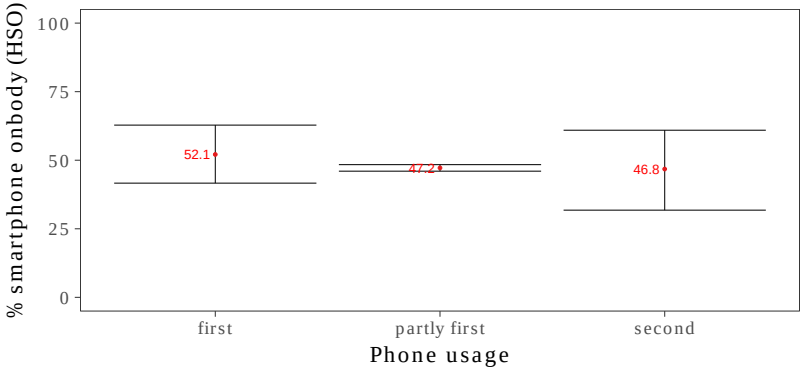


Figure 5.4: Comparison of the mean on-body percentage during the day of patients who had just one smartphone (first), switched between phones (partly first) or used the study smartphone as a second smartphone next to their private one (second). The error bars denote the 95 % confidence intervals, showing that the difference of 5 % between the use of the study smartphone as the main phone or as a secondary phone is insignificant.

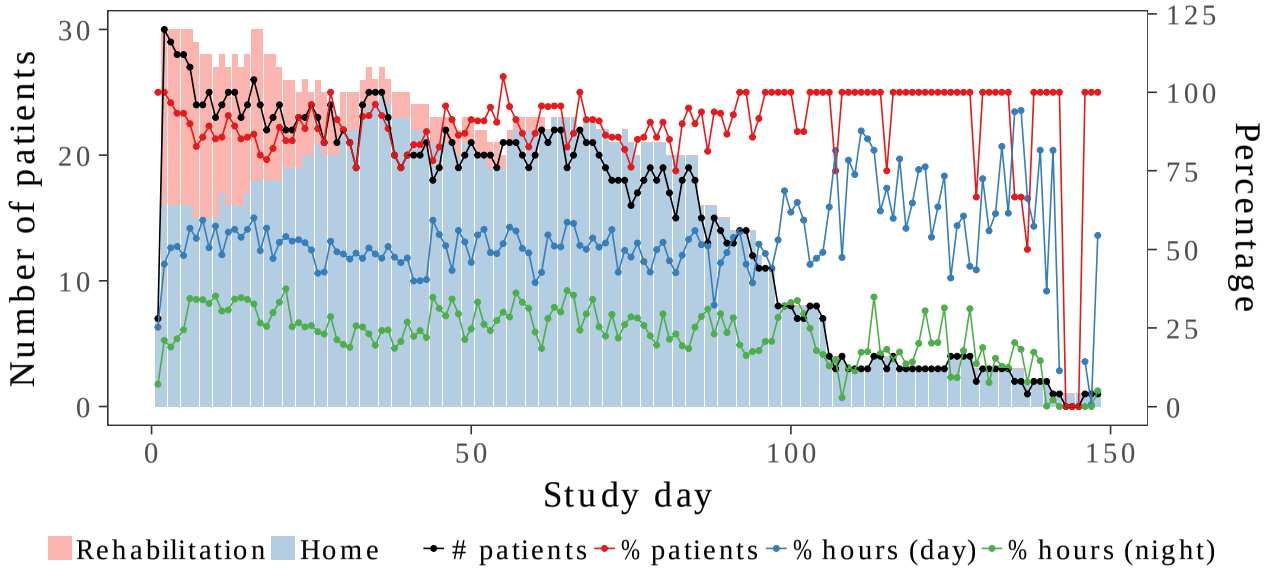


Figure 5.5: Rate of wearing the smartphone on-body. The bars display the number of patients at home or in rehabilitation for each day. The black line denotes the number of patients wearing the smartphone at a given day, the red line denotes the percentage of patients wearing the smartphone at a given day. The two other coloured lines (for day and night) show that the wearing percentage is constant over the study period. (Note: Due to breaks during study participation, the study days count up to 148.))

5.3.3 Answering the questionnaires

All patients could use the smartphone app to answer the digital questionnaires, as Table 5.2 shows. In total, 3348 answers for the digital questionnaires were collected. The patients filled out the questionnaires for, on average, 72.75 % of the days they were at home or in rehabilitation. In total, 3 patients answered for less than 40 % of the days. Table 5.3 shows possible reasons for the low answering rates of these patients. Excluding these 3 special cases leads to a response rate of 78.91 % for 90 % of the patients. Figure 5.6 shows that the retention over time is on a constant level over the whole study period.

Table 5.3: Reasons for low VAS answering rates of three patients; Days: percentage of days with VAS answers; †: passed away

ID	Days	Situation
162†	25	This patient underwent several emergency visits and the tracking was 'too much' in combination with her rapidly worsening health condition.
192	36.84	After an unplanned stay in a hospice, her study participation was restarted.
423	11.54	The oldest participant described herself as being very forgetful.

Remarkably, 13 out of 30 patients filled out the questionnaire on average more than 1.5 times per day. Patient 142 stated that he wanted to fill it out every time he feels a change and did so on average 5.68 times per day.

5.3.4 Weekly interviews

Completeness of weekly interviews

As described, the patients were called once per week. It was the responsibility of the interviewers to call, so that the patients were not obliged to remember the interview appointments. In general, the patients were easy to contact. In total, 251 out of 301 weeks are complete (i. e., 83 %). Patient 424 is excluded from the analysis of the weekly interviews because his health deteriorated quickly, and he was unable

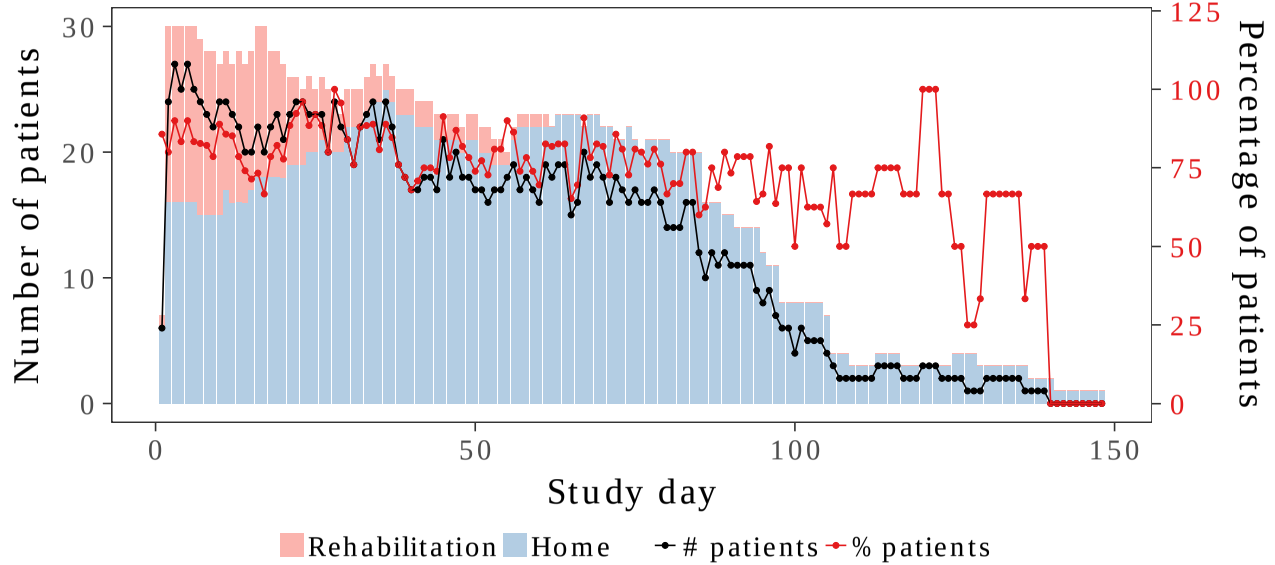


Figure 5.6: Rate of answered questionnaires per day. The bars display the number of patients being at home or in rehabilitation for each day, the lines denote the number of patients that answered a questionnaire. Excluded in both (bars and lines) are absences due to holidays and hospital visits. The red curve shows a constant level of retention over the study period (Note: Due to breaks during study participation, the study days count up to 148.).

to speak on the phone. Details about the completeness of EORTC are provided in Table 5.4.

Table 5.4: Completeness of QLQ-C30 scores for baseline (i.e., at discharge from hospital), final and weekly interviews of all patients. The absolute numbers of complete, incomplete and missing questionnaires are provided in three columns.

type	complete	incomplete	missing
baseline (BL)	23	4	3
final	18	0	12
interview	251	19	31

Reliability of EORTC QLQ-C30

The reliability of the scores is tested with Cronbach's alpha [143]. This metric provides the internal consistency of scores that are measured by multiple Likert-type scales and items. In general, a threshold of $\alpha = .8$ is accepted. The QLQ-C30, as shown in section A.2.2, comprises 30 questions that build 15 scores. In total, nine of these scores are composed of more than one item and are, therefore, tested for reliability. The reliability of each score is estimated on the basis of all available questionnaire answers of all study participants. The instructions concerning how to calculate the scores are given in [144].

Table 5.5 shows the calculated Cronbach's alpha values for the QLQ-C30 questionnaire results that were collected during the performed study. The scores for concentration/focus (CF), family and friends (S) and nausea (NV) do not fulfil the threshold. The analysis also revealed that dropping various items would not increase the reliability of these three scores. As a result, these three scores should not be used in future analysis of this data set. However, the global health status (QL2) can be used for further analysis of this data set.

Table 5.5: Reliability of QLQ-C30 scores using Cronbach's alpha. *: low reliability

Score	Alpha	Score name
QL2	0.89	global health status/ QoL
PF	0.88	physical functioning
RF	0.92	role functioning (work and leisure)
FE	0.81	emotional functioning
CF*	0.79	cognitive functioning (concentration/focus)
S*	0.62	social functioning (family and friends)
FA	0.82	tiredness
NV*	0.75	nausea and vomiting
PA	0.92	pain

5.3.5 Evaluation of the feasibility of activity monitoring in the study cohort

Duration of study participation

Most of the patients participated for the entire study period (12 weeks) and agreed to compensate for breaks due to vacation, rehabilitation or re-hospitalisation by continuing after week 12. Patients who dropped out had to do so mostly due to health reasons.

Only two patients (324, 351) aborted the study due to reasons other than severe worsening health. Patient 324 had to deal with an unstable Bluetooth connection. The reasons for the drop-out of patient 351 are a combination of bad health condition (2 unplanned re-hospitalisations in 4 weeks) with difficult device handling due to unstable Bluetooth connection.

Feedback concerning the Everion®

In general, the feedback concerning the Everion® was positive: The patients stated that it did not require a lot of effort to wear it and that it did not disturb them. Many patients would have participated for a longer study period. Reactions from the environment of the patients were mostly curiosity. One patient discussed the topic of data security with others.

Only two patients complained about the Everion®. The first patient said that the band was uncomfortable and restricted his arm. This patient was obese and needed a very large band. The issue was reported to Biovotion, who reported that they were working on an improved design of the elastic. Another patient experienced difficulties in putting on the sensor as he had to use crutches when standing and moving. One patient experienced an allergic reaction, diagnosed as a contact allergy by his dermatologist (probably on the GSR sensor field). Despite this, the patient continued to wear the Everion®. He replaced the elastic with a bigger elastic. The allergic reaction healed and did not show up again.

Smartphone usage

All patients – apart from two patients who experienced difficulties to use a smartphone in general – were able to use our app and and the Everion®.

Patient 355 received her first smartphone when she began to participate in the study. Using it for general things, such as making calls and sending texts, was difficult for her. She stated that she would have given up on the study if not for several home visits.

The oldest participant, patient 423, was an 85 years old woman with hyperthyroidism. She wanted to abort the study because of her worsening health and because of the difficulty she experienced in using the smartphone.

Reasons for not wearing the devices

Common reasons for not wearing the devices included personal hygiene, sleep, the need to charge the devices or not having the smartphone on-body while at home. As well as this, worsening health conditions among some patients led to them not wearing the sensor, especially in cases of re-hospitalisation. In the case of patient 423, she forget to take her device with her on certain days, so some days are missing.

5.3.6 Which patients are eligible for activity monitoring?

In general, all of our included patients could handle the devices. In terms of data completeness, new smartphone users were highly capable. However, in combination with forgetfulness, data completeness dropped significantly (e. g., patient 423). The second obstacle was the

worsening of symptoms: To a certain extent, patients could handle the additional tasks of the stufy. However, when their health conditions worsened, the patients had to deal with its associated issues. This meant that they stopped wearing the device or stopped tracking entirely.

5.4 Descriptive statistics of self-reports

5.4.1 Edmonton Symptom Assessment Scale

At the beginning of the study (baseline) and at the end of the study (final), the patients reported the ESAS (section A.2.1). Here, 'baseline' refers to the day of discharge from hospital and 'final' refers to the last day of study participation (i. e., when they the Everion® and the final interview was conducted; Figure 2.6). Note that in the following analysis, only complete cases (i. e., baseline and final values present) are considered. Figure 5.7 shows that the medians of almost all symptoms (over the study cohort) have not changed over the study period, except for tiredness, which worsened. However, the IQR is shifted towards a positive difference, indicating a worsening of symptoms, including depression, minimal and current pain, tiredness and general wellbeing.

It is worth looking at the baseline itself, as presented in Figure 5.8. Upon discharge from hosptial, a general lack of appetite was the symptom with the largest IQR, together with poor wellbeing and tiredness, followed by pain and fear. However, most patients were at a level of 5 or lower for most symptoms and were, therefore, stable.

5.4.2 Quality of life

The EORTC QLQ-C30 global quality of life (QL2) score is considered as an early indicator of changes in quality of life. Physicians consider decreases of the score of $\geq 10\,\%$ as relevant drops in quality of life. As illustrated in Figure 2.6, patients had to fill out the questionnaire before being discharged from hospital, referred to as 'baseline' (BL),and were then assessed in a weekly interview. Table 5.6 provides an overview of the relevant drops in the study cohort. For 4 patients out of 29 (excluding patient 61 and 424 for aforementioned reasons), the baseline QLQ-C30 values are missing. In this case, the first QLQ-C30 questionnaire from the weekly interviews is taken as a reference in order to calculate relevant drops in quality of life.

Table 5.6: Overview of patients with relevant drops in the global QoL score (QL2). patid: patient id; no. interviews: number of interviews; no. QL2 drops: number of relevant drops in score QL2; BL: baseline value available

patid	no. interviews	no. QL2 drops	BL
23	13	3	no
55	12	0	yes
69	8	0	yes
128	13	0	yes
142	10	3	yes
162	1	0	yes
165	14	1	yes
168	13	1	yes
183	12	1	yes
187	9	1	yes
192	10	7	no
193	11	1	yes
194	5	0	yes
216	7	6	yes
246	11	0	yes
271	11	0	yes
277	12	0	yes
319	10	1	yes
324	3	1	yes
340	9	2	yes
351	1	0	yes
355	9	0	no
359	11	1	yes
377	9	0	yes
405	10	4	yes
423	3	0	no
429	12	0	yes
430	7	0	yes
438	10	0	yes

Figure 5.9 shows in detail the development of the QL2 score for all patients. The red horizontal line indicates the baseline value if not missing. The figure shows multiple patterns for the course of QL2:

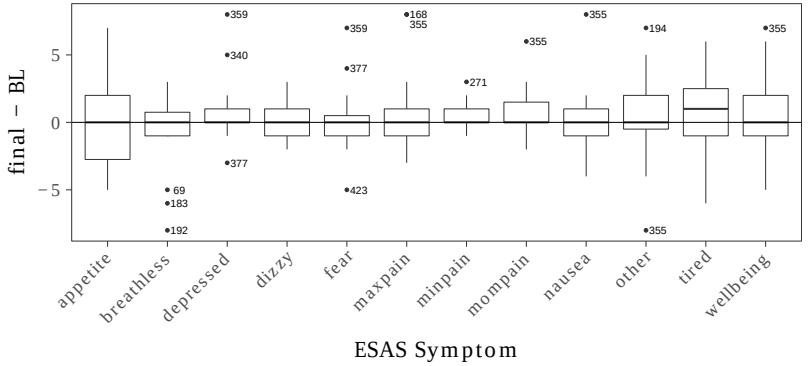


Figure 5.7: Box plot for the change of ESAS symptoms from baseline (BL) to final over 23 patients. The x-axis denotes the symptoms, the y-axis denotes the difference `final − BL` negative differences mean improvement, positive differences mean worsening of symptoms. The zero-line denotes no change in a symptom. The outliers are named using the patients' IDs.

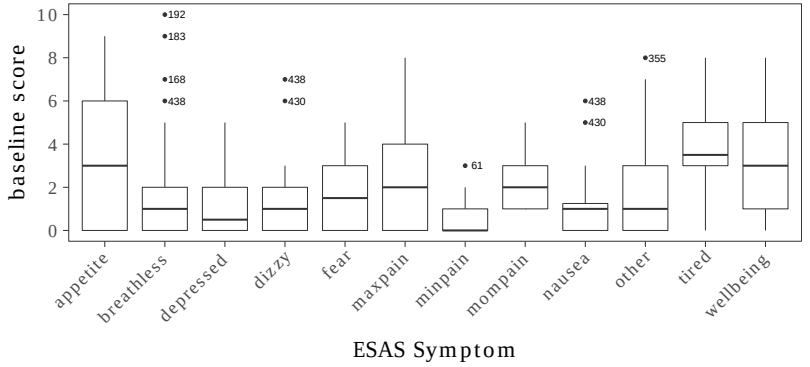


Figure 5.8: Box plot for the baseline scores of ESAS symptoms for 28 patients. Outliers are named with patient id.

constant or improving Patients 55, 69, 128, 193, 194† 246, 271, 277, 355, 377, 429, 430 and 438 did not experience a relevant drop. Instead, they were stable or recovering.

relevant drop and improving Patients 142, 168, 183, 319 and 340 experienced a relevant drop during the first month (405 in week 6) and could recover to the baseline level or improve more.

improving, relevant drop, improving Patients 23, 165, 187, 359, 405 were stable or improving for the first weeks of the study before experiencing a relevant drop from which they recovered.

relevant drop with little improving Patient 192's health was mostly worsening throughout her study participation with an improvement during the last month. The original baseline value and the first weeks are missing, because after her initial beginning, she had to take a break from the study due to her delayed discharge to home.

relevant drop with decreasing trend Patient 216† experienced a relevant drop from the beginning and did not recover (downward trend).

To summarise, the QL2 score dropped in the first month of study participation rather than in later months. Most patients recovered and were, by the end of the study, close to the baseline value or even improved. Both, patient 194 and patient 216, passed away. Their reports show different trends: Patient 194 reported an improved quality of life, whereas patient 216 experienced a constant reduction in quality of life. These different courses of the QL2 score belong to two patients who passed away after experiencing different courses of their diseases. Patient 194 had an acute emergency due to an epileptic seizure, which happened without forewarning. Patient 216 experienced a continuous worsening of his health.

5.4.3 Pain and distress

Figure 5.10 shows box plots for the answers to the digital questionnaires concerning pain and distress. These answers were recorded by the smartphone app throughout the whole study period. 22 out of 30 patients had an increase of at least 2 points compared to the baseline value at some point in time (relevant increase). The baseline value is

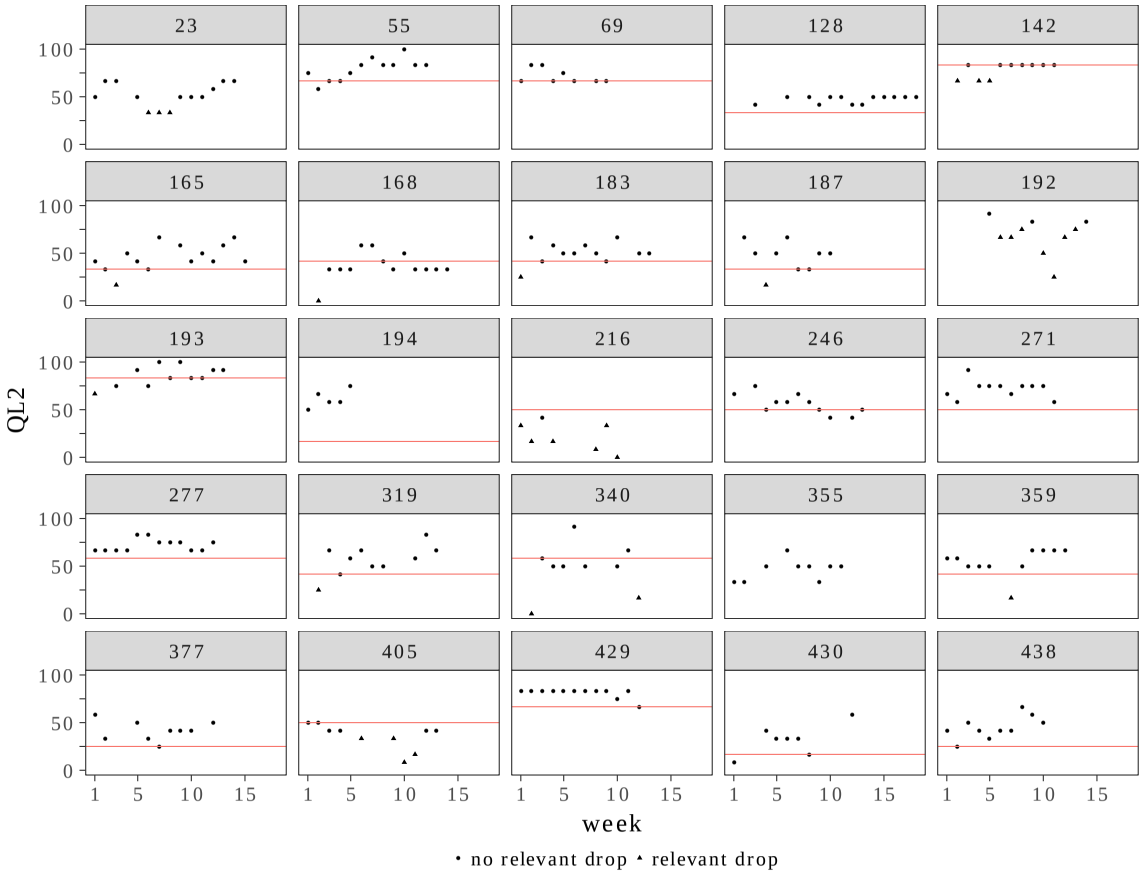


Figure 5.9: Overview on the course of the global health score (QL2) for patients with at least 4 data points. The red line indicates the baseline value at discharge if existent.

defined as the first value. However, the small inter quartile ranges of only up to 1 point for 22 patients (pain), 18 patients (distress) resp., and up to 2 points for 28 patients (both scales), illustrate the low variance in the digital self-reports and a strong stability of the subjective perception of pain and distress over the whole study period.

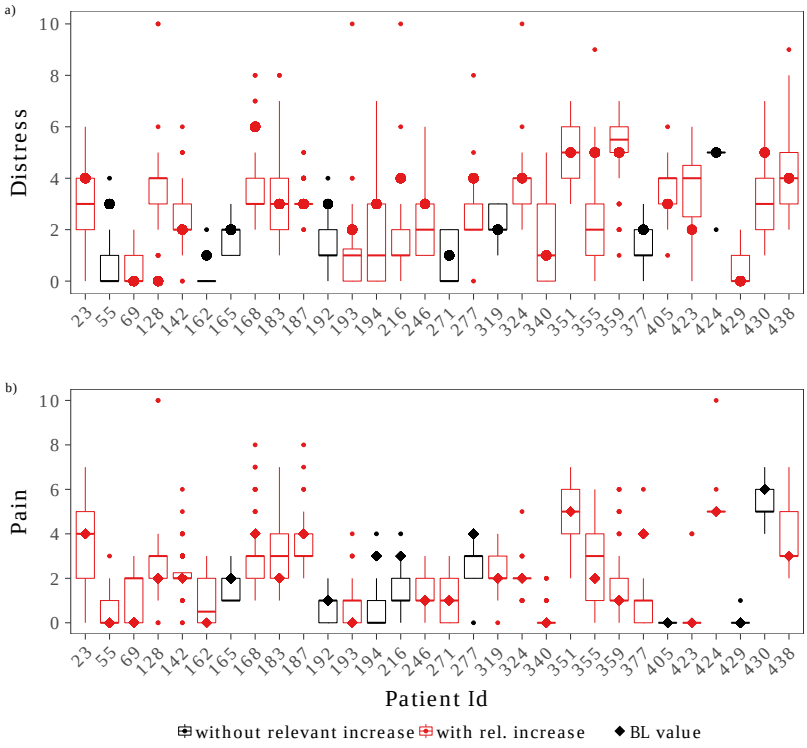


Figure 5.10: Variability of patients' self-reports of (a) Distress and (b) Pain. 22 out of 30 patients had a relevant increase (i.e., at least one answer > 2 points compared to the BL value).

Table 5.7 shows that over study period, the median for both scales did not move from the BL values for roughly half of the patients. The first line of the table shows that for many patients (Pain: 8, Distress: 14), these symptoms improved compared to the baseline. Only for a minority of the patients (Pain: 6, Distress: 4), the median over the study period increased compared to the baseline values.

Table 5.7: Number of patients that had an increased, decreased or stable median (Pain, Distress) compared to the baseline.

Trend	Pain	Distress
median < BL	8	14
median == BL	16	12
median > BL	6	4

5.5 Analysis of emergency visits

In addition to evaluating the feasibility of the activity monitoring of ambulatory patients in palliative care, the explanatory power of the collected sensor data was investigated in the context of emergency visits (EVs) to the hospital.

5.5.1 Overview

First, redundant features were identified and eliminated. The reduced feature set was explored in order to discover features that showed differences between the group of patients who experienced a non-elective hospital visit (PEV) and the group without such an experience (PnoEV).

The general idea comprises the following: For each patient, use data from the beginning of their study participation as the baseline behaviour or values (BL) (e.g., take the data from the first week spent at home after discharge and compare the course of the patient's data to that original baseline).

As listed in Table 5.8, 11 patients experienced a non-elective hospital visit during their study participation. In total, two of them (55, 340) were (non-electively) re-hospitalised while staying at a rehabilitation unit and are, therefore, excluded from the analysis of this section. Patient 351 experienced re-hospitalisation in the first week after discharge, so not enough baseline data could be collected. Therefore, he is excluded from the analysis below, too.

Two patients experienced two and four resp., non-elective hospital admissions. Since after the discharge from hospital, a patient's situation might have changed drastically, a new baseline value would be needed. Since the amount of data between consecutive emergency visits is not

Table 5.8: Reasons for recorded non-elective visits in hospital during study participation. sdn: number of study participation day; *: included in data analysis

patid	sdn	stationary	reason for emergency visit
55	11	yes	new placement of jejunal tube after accidental pull
162*	12	no	bronchitis (coughing and thoracic pain)
162	29	no	swelling and pain under right clavicle (DD port infection, DD metastasis)
162	41	yes	strong nausea and general weakness (asthenia)
162	84	yes	dyspnea, pain and general weakness (asthenia)
183*	79	no	low SpO_2 of 80 %, abort of planned treatment; clarification in emergency unit did not reveal any changes
194*	38	yes	seizure
216*	51	yes	no information; reported earlier about increasing pain
246*	72	yes	kidney stone
351	7	yes	nausea and vomiting, stomach pain
351	22	yes	diarrhoea, nausea and vomiting, stomach pain
340	8	yes	progressive hemisyndrom left side as consequence of growing cerebral oedema
405*	37	yes	pneumonia (dyspnea and coughing)
423*	75	yes	pyelonephritis
430*	48	yes	general weakness (asthenia)

sufficient to create new baseline values, only the first emergency visit of each patient is included in the analysis.

From the multi-variate time series of features, as described in Chapter 3.4, the change of features between baseline (BL) measurements, i. e., the measurements during the first week, and measurements close to emergency events (EV) was extracted. Figure 5.11 illustrates which

measurements were included in the following analysis.

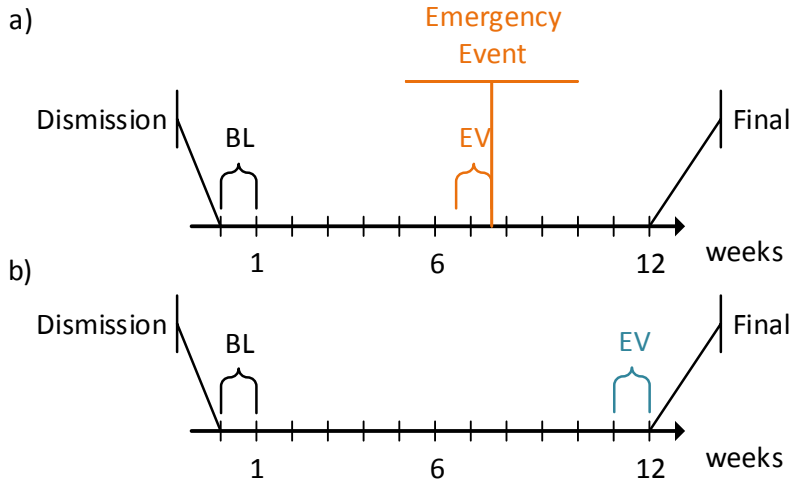


Figure 5.11: Considered periods for t-test: (a) shows the case of a patient with emergency visit, (b) shows the case of a patient without emergency visits. BL: the first week as baseline; EV: the event week comprising all data 7 days before the day of the emergency visits in situation (a) and all data of the last 7 days of the study participation in situation (b)

For all features f^i, $i \in \{1, .., m\}$ (m is the number of features) the baseline value f^i_{BL} and the event value f^i_{EV} are calculated by taking the median of all available measurement values (already aggregated to hourly or daily values) and ignoring the missing values. The change $\mathfrak{C}(f^i)$ of feature i is then calculated as the ratio between BL and EV:

$$\mathfrak{C}(f^i) = \frac{f^i_{EV}}{f^i_{BL}} \tag{5.1}$$

The calculations according to Equation (5.1) leads to a data set of one value per patient and feature, i. e., a $n \times m$-matrix, where $n = 30$ is the number of patients and $m = 121$ is the number of features as listed in Table 3.14 and Table 3.15.

5.5.2 Feature reduction

After removing features with zero variance (i. e., constant over patients), $m = 115$ features remained. With a pairwise correlation analysis of feature pairs $< f_i, f_j, 0 \leq i, j \leq 114, i < j \leq 115 j$ (leading to $\sum_{i=1}^{n-1} i = \frac{n(n-1)}{2} = 6555$ correlations according to Gauss [145]), highly correlated features ($|r| > 0.99, p < 0.05$) were identified. Feature f_j was removed only if f_i was not already removed (added to the list `indexDeletedFeatures` in algorithm 5.1. If f_i already has been removed, feature f_j is kept, because the highly correlated counterpart is already eliminated, and one of the two has to be kept. This principle is illustrated in Figure 5.12.

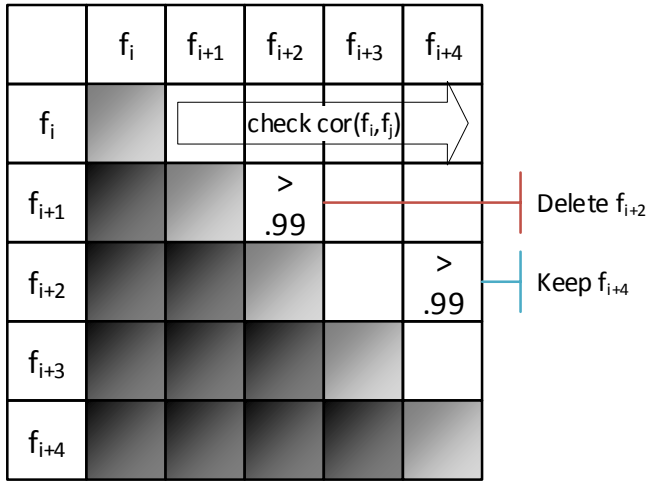


Figure 5.12: Illustration of the algorithm to reduce the feature space, based on correlation analysis: The correlation matrix is calculated. Due to its symmetry, only the upper triangle is processed. Pairs with $r > .99, p < 0.05$, are checked: f_{i+4} has to be kept, because f_{i+2} is already marked to be deleted.

This dimensionality reduction yielded a set of 80 remaining features and is considered a cleaning step rather than a feature selection step.

These remaining 80 features were inspected visually using a parallel plot [147]. Figure 5.13 provides an example of a parallel plot for ten

Algorithm 5.1: Correlation-based feature elimination.

input : set of m features
output: list of indexes to be removed from feature set

```
indexDeletedFeatures ← [];
for i ← 1 to m − 1 do
    for j ← i + 1 to m do
        if abs(cor(f_i, f_j)) > 0.9 ∧ p < 0.05 then
            if i ∉ indexDeletedFeatures then
                remove f_j;
                indexDeletedFeatures.append(j)
return indexDeletedFeatures
```

features. Each line represents the feature ratios $\mathfrak{C}(f^i)$ of one patient. The colour denotes the group. It is clear that none of the displayed feature ratios are able to separate the two groups by itself. However, the resting heart rate (RHR) and the resting heart rate variability (RHRV) show the best separation of the two groups. This is confirmed by the density plot shown in Figure 5.14. For plotting the densities, the features were scaled to the range 0 to 1, using the transformation

$$scaled(x_i) = \frac{x_i - min(\mathbf{X})}{max(\mathbf{X}) - min(\mathbf{X})}$$

for $\mathbf{X} = \{x_i | i = 1, \ldots, 30\}$.

The plotting function `geom_density` of the R package `ggplot` plots a density estimate based on the data samples using a Gaussian kernel density estimator [148]. The overlapping areas show which value ranges of a feature occur in both patient groups. These overlapping areas would lead to misclassifications if the two groups were separated based on these features.

5.5.3 Test procedure for group comparisons

Following an exploratory approach, statistical tests are conducted to compare the described change of features $\mathfrak{C}(f^i)$ in the group of patients who have experienced at least one emergency visit during their study participation (PEV) and in the group of patients that have not experienced any emergency visit during their study participation (PnoEV).

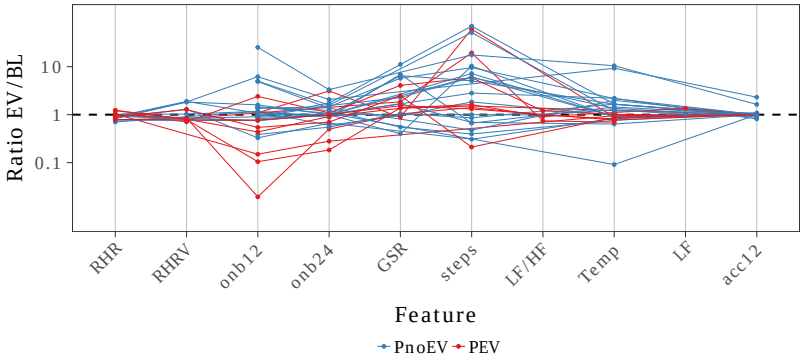


Figure 5.13: Parallel plot for ten features. The x-axis presents the features, the y-axis (log-scale) contains the values of these features. Each line represents one patient, coloured by the patients group (PEV/PnoEV). The dashed line represents the value 1 indicating no change between baseline and the event.

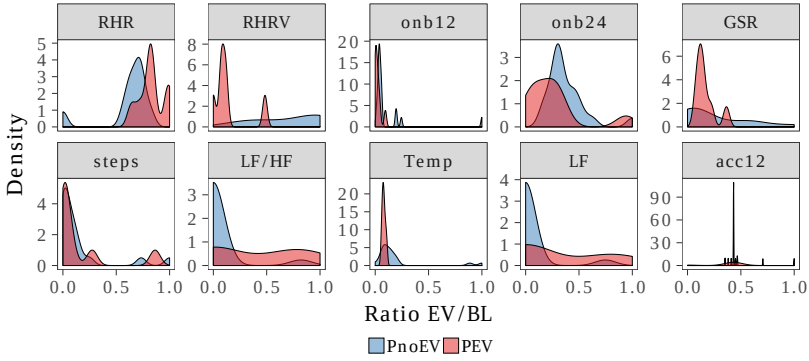


Figure 5.14: Density plots for ten features (scaled range $[0, 1]$) per patients group (PEV: 7 patients, PnoEV: 20 patients).

These group comparisons serve to discover promising features, not to show statistical evidence [149]. In total, three patients were excluded from the PEV group, because they experienced the emergency visit during their rehabilitation period or too soon (7 days) after their discharge.

Test selection

Since the group sizes are small ($n_{PEV} = 8$, $n_{PnoEV} = 20$, the statistical test has to be chosen carefully. The test selection depends on the data distribution in both groups, as illustrated in Figure 5.15. In this work, the randomisation test was performed with 1000 permutations.

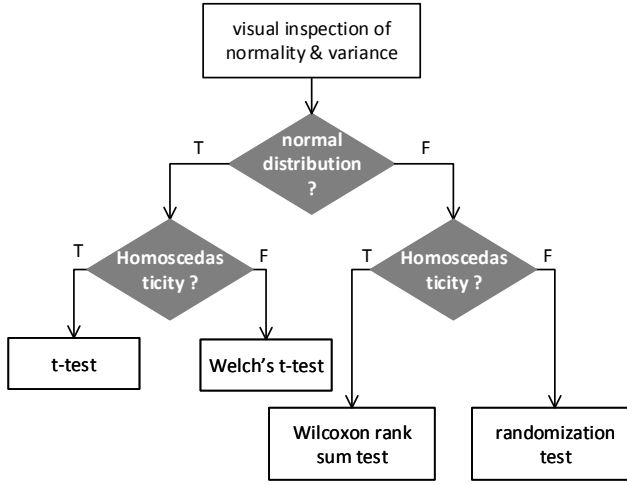


Figure 5.15: Test structure for comparison of groups: for all feature ratios, the distribution of each group is inspected in terms of normality and homoscedasticity, and the relevant test is chosen.
T – TRUE, F – FALSE

Examination of data distribution

The data distribution of each feature is examined group-wise using box plots and quantile-quantile plots (q-q plots). The latter plots the quantiles of the data versus the quantiles of a theoretical normal distribution. If the plotted points lie on or close to the plotted diagonal line, the examined data is normally distributed. In addition to evaluating the symmetry of the data distribution in each group, the box plot also investigates wether the variances are sufficiently similar in both groups.

Interpretation of test results

As part of the feature discovery method, nominal significances are reported for the features that are below the threshold of $p < .05$ [149]. For the Wilcoxon rank-sum test, the effect size r is calculated by

$$r = \frac{z}{\sqrt{N}}$$

where z is the z-score and N is the total sample size (i. e., the sum of the two group sizes).

The randomisation test delivers the difference in means as test statistics and a p-value.

5.5.4 Results: identified group differences

The comparison of groups delivered nominal p-values ($p < .05$) for three features, namely RHR, RHRV and Vsteps (feature name: `acc.steps.speed` in ref. Table 3.15: The daily mean of the speed of steps is calculated, as measured by the smartphone in steps s^{-1}). The test results and further analysis of these three features are presented in the following sections.

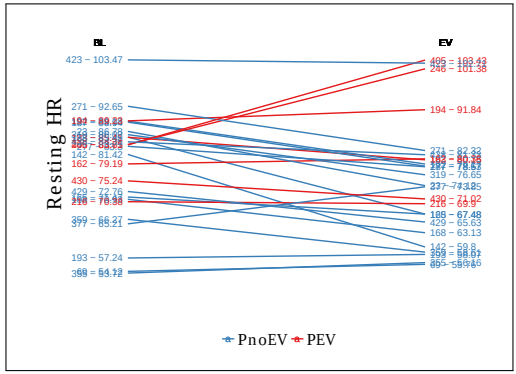
Resting heart rate (RHR)

Based on the data distribution of RHR in the groups PEV and PnoEV (ref. section A.4.1), the Wilcoxon rank-sum test was applied, and a randomisation test was performed as confirmation.

Conducting the Wilcoxon rank-sum test delivered the following results: on average, patients of the PEV group had a larger ratio of the RHR $\mathfrak{C}(f^{RHR})$(mdn = 1.01) than the PnoEV group (mdn = 0.79). This difference was nominally significant $W = 25, p = .021$ with a medium-sized effect $r = .46$. The randomisation test confirms the significance of difference in means with $p = 0.005$.

The slope graph in Figure 5.16 shows that, for 4 out of 7 patients with emergency visits, the median RHR was higher in the week before the event than in the baseline week ($f_{EV}^{RHR} > f_{BL}^{RHR}$), two patients had a decreased median RHR and one patient saw no change. The RHR of the PnoEV group changed for 13 out of 15 patients in the opposite direction.

(a)



(b)

Slope	PEV	PnoEV
$f_{EV}^{RHR} > f_{BL}^{RHR}$	4	3
$f_{EV}^{RHR} < f_{BL}^{RHR}$	2	13
$\|f_{EV}^{RHR} - f_{BL}^{RHR}\| < 1$	1	1

Figure 5.16: a) Slope graph and b) slope summary for RHR. The lines show the change of the resting heart rate for each patient from baseline (BL) to the week before an event or at study end (EV). The table gives a summary of the rising and falling slopes per patient group.

Figure 5.17 shows a scatter plot where the baseline RHR values are on the x-axis, and the final RHR values are on the y-axis. The majority of PEV cases lie above the black line, denoting a ratio of 1 (no change); all samples of the PEV group lie below the black line, i. e., these patients experienced an increased RHR. Only three patients of the PnoEV group (two of them very close to the border) experienced an increase, too. According to the literature review, no studies have examined the relationship between resting heart rate and stage of cancer development.

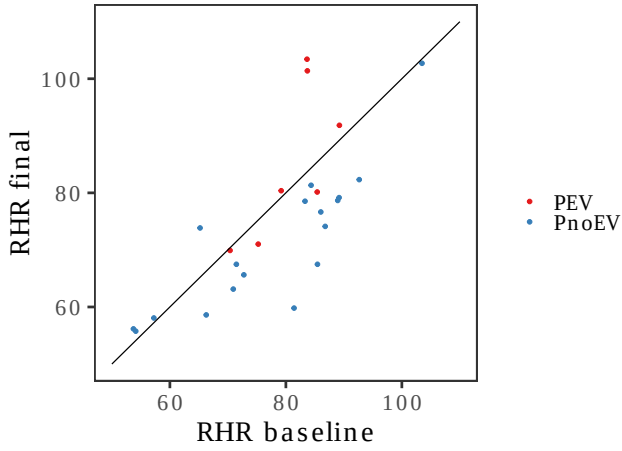


Figure 5.17: Scatter plot showing the baseline RHR vs. final RHR. The black line indicates no change.

According to experts, the normal range of RHR for adults is 60 bpm to 100 bpm [150]. Common causes for tachycardia (higher heart rate) include hyperthyroidism, fever and heart insufficiency as well as drugs and caffeine. As an example, patient 423 had diagnosed hyperthyroidism, with radio iodine therapy taking place in the month after her study participation. Her RHR at the beginning of the study was the highest of all patients and stayed almost constant over the study period. Common causes for bradycardia (lower heart rate) include hypothyroidism, opiates, barbiturates, tranquilisers, inflammations (like meningitis and heart muscle inflammation) and increased physical fitness. The RHR is influenced by many co-founders, e. g., excitement, emotions, caffeine, stress, the weather and also depends on the time of

the day. As a general health indicator, an increased RHR can signal a worsened general health condition. The found difference in the groups (increase in PEV vs. decrease in PnoEV) accords with the literature.

Resting heart rate variability (RHRV)

Based on the data distribution of RHRV in the groups PEV and PnoEV (ref. section A.4.2), the Welch's t-test was applied, and a randomisation test was performed as confirmation.

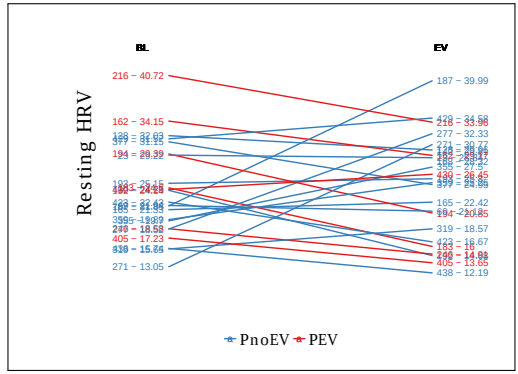
The Welch's t-test delivered the following results: on average, patients of the PEV group had a smaller ratio of the HRV metric (RMSSD) $\mathfrak{C}(f^{RMSSD})$ (mn=0.81) than the PnoEV group (mn=1.17). This difference was nominally significant $t(df = 19.7) = 2.79, p = .011$ with a large-sized effect $r = .53$. When removing the outlier of the PEV group with $\mathfrak{C}(f^{RMSSD}) = 1.09$ (patient 430), the randomisation test confirms the result with $p = 0.43$. A specificity of patient 430 is the following: after discharge from hospital, the patient lost weight (several kg) and the elastic of the Everion® was too wide for her arm. After consulting the study team and after clarification from the manufacturer, she was wearing the device on her calf. However, this new position can have influenced the measurement.

The slope graph in Figure 5.18 shows that in fact, for 6 out of 7 patients with emergency visits, the median RHRV was lower in the week before the event than in the baseline week ($f_{EV}^{RHRV} < f_{BL}^{RHRV}$), and only one patient had an increased median. The RHRV of the PnoEV group changed for 8 out of 16 patients in the opposite direction, with five decreasing and three constant medians.

Figure 5.19 shows a scatter plot where the baseline RHRV values are on the x-axis, and the final RHRV values are on the y-axis. Except for one case, the black line denoting a ratio of 1 (no change), all samples of the PEV group lie below the black line, i. e., these patients experienced a decrease in the RHRV. However, a considerable number of PnoEV patients also experienced a decrease and lie in similar ranges for baseline and final value. A possible explanation for this is due to the fact that the RHRV is influenced by many factors such as stress and inflammation, which in many cases, does not lead to emergency visits.

The normal range of resting HRV (RMSSD) for healthy adults is 42 ms with a range of 19 ms to 75 ms [151]. However, a study focusing on cancer patients found a significantly reduced (RMSSD) of 24 ms ± 20 ms [152]. De Couck et al. argument that both causal direc-

(a)



(b)

Slope	PEV	PnoEV
$f_{EV}^{RHRV} > f_{BL}^{RHRV}$	1	8
$f_{EV}^{RHRV} < f_{BL}^{RHRV}$	6	5
$\lvert f_{EV}^{RHRV} - f_{BL}^{RHRV} \rvert < 1$	0	3

Figure 5.18: a) Slope graph and b) slope summary for RHRV. The lines show the change of the resting HRV for each patient from baseline (BL) to the week before an event or at study end (EV). The table gives a summary of rising and falling slopes per patient group.

tions are plausible. The progress of tumours and reduced activity of the vagal nerve have three common causes: the presence of free radicals, excessive inflammation and activity of the sympathicus, which is increased by stress. However, one study did show that low HRV predicted an increased risk for all cause (especially cancer) mortality [153]. However, the study considered only men in the age of 40 to 60 years. Dekker et al. conclude that a reduced HRV indicates a compromised health [153].

Step speed (Vsteps)

Based on the data distribution of Vsteps in the groups PEV and PnoEV (ref. section A.4.3), the Wilcoxon rank-sum test was applied, and a randomisation test was performed as confirmation.

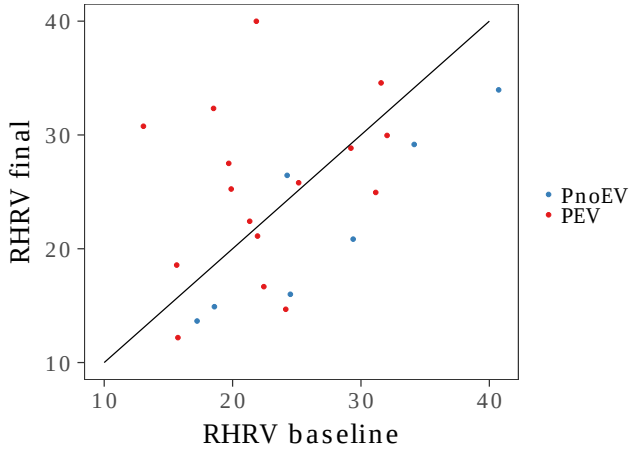


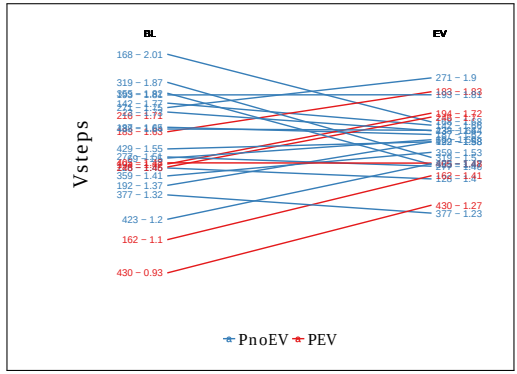
Figure 5.19: Scatter plot showing the baseline RHRV vs final RHRV. The black line indicates no change.

On average, patients of the PEV group had a larger ratio of Vsteps (mdn=1.1) than the PnoEV group (mdn=0.99). This difference was nominally significant $W = 11, p = .003$ with a large-sized effect $r = .61$. The randomisation test confirmed the significance with $p = 0.003$.

The slope graph in Figure 5.20 shows that in fact, for 5 out of 6 patients with emergency visits, the median Vsteps was higher in the week before the event than in the baseline week ($f_{EV}^{Vsteps} > f_{BL}^{Vsteps}$) and one patient had a change smaller than 0.1. The Vsteps of the PnoEV group remained constant (i. e., changed less than 0.1) for 9 out of 17 patients; decreases and increases occurred just as often (4 times).

Figure 5.21 shows a scatter plot where the baseline Vsteps values are on the x-axis, and the final Vsteps values are on the y-axis. Except for one case, which lies on the black line, denoting a ratio of 1 (no change), all samples of the PEV group lie above the black line, i. e., these patients experienced an increase in the Vsteps. However, a considerable number of PnoEV patients also experienced an increase, and they lie in similar ranges for baseline and final value. A possible reason for this is due to the fact that the Vsteps of the study cohort was diverse in terms of walking capabilities, ranging from a farmer taking care of his cattle to patients depend on a walking aid or a wheelchair. These

(a)



(b)

Slope	PEV	PnoEV
$f_{EV}^{Vsteps} > f_{BL}^{Vsteps}$	5	4
$f_{EV}^{Vsteps} < f_{BL}^{Vsteps}$	0	4
$\lvert f_{EV}^{Vsteps} - f_{BL}^{Vsteps} \rvert < 0.1$	1	9

Figure 5.20: a) Slope graph and b) slope summary for Vsteps. The lines show the change of the Vsteps for each patient, from baseline (BL) to the week before an event or at study end (EV). The table gives a summary of the rising and falling slopes per patient group.

limitations of walking capabilities were not always a consequence of the cancer, but originated from reasons independent of the cancer.

Subjective self-reports

Analogous testing of differences in the daily (pain, distress) and weekly self-reports (QL2 – global quality of life and health score of QLQ-C30) showed no significant differences between the PEV and PnoEV groups, as listed in Table 5.9. Physicians consider the following changes in the analysed scales (pain, distress, QL2) as relevant:

QL2: a drop of at least 10 % of the baseline value (at discharge)

Pain: an increase of at least 2 points compared to the baseline value (first value)

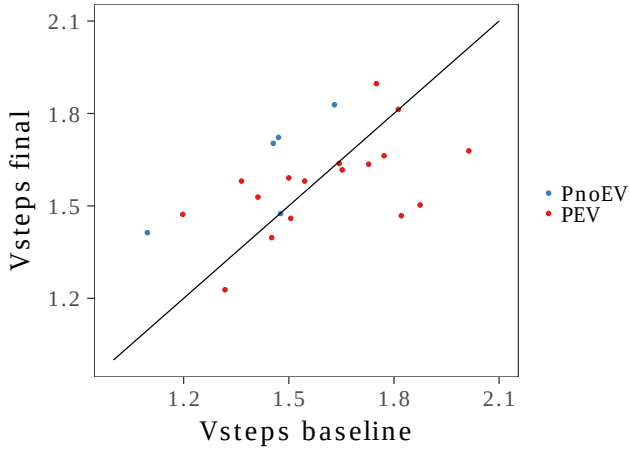


Figure 5.21: Scatter plot showing the baseline Vsteps vs. final Vsteps. The black line indicates no change.

Distress: an increase of at least 2 points compared to the baseline value (first value)

For all tests, only data before an emergency event was considered. Note that for the digital scales, the ratios cannot be calculated for many patients due to reported values of 0 in the baseline.

5.5.5 Recommended sample size based on power analysis results

Table 5.10 summarises recommended sample sizes that were estimated by a power analysis (details of the power analysis are given in section A.5).

Assuming a similar re-hospitalisation rate (10 out of 30; i. e., ⅓), a total sample size of $n = 60$ as minimum should be targeted in future studies. Considering a similar rate of re-hospitalisations that actually could be included in the analysis (7 out of 28; i. e., ¼), a total sample size of $n = 21 \times 4 = 84$ is recommended.

Table 5.9: List of tested self-report features with p-values; none of the tested features extracted from self-reports were significant.

Feature	p-value (t-test)	p-value (Wilcoxon)
QL2 ratio EV/BL	0.74	0.9
QL2 minimum	0.18	0.25
QL2 maximum	0.2	0.22
$\frac{\texttt{nb.ofrelevantQL2drops}}{\texttt{nb.ofreportedvalues}}$	0.81	0.74
Pain ratio EV/BL	–	0.5
Distress ratio EV/BL	–	–
$\frac{\texttt{nb.ofrelevantPainchanges}}{\texttt{nb.ofreportedvalues}}$ (whole period before EV)	0.27	0.1
$\frac{\texttt{nb.ofrelevantDistresschanges}}{\texttt{nb.ofreportedvalues}}$ (whole period before EV)	0.33	0.6
$\frac{\texttt{nb.ofrelevantPainchanges}}{\texttt{nb.ofreportedvalues}}$ (week before EV)	0.09	0.88
$\frac{\texttt{nb.ofrelevantDistresschanges}}{\texttt{nb.ofreportedvalues}}$ (week period before EV)	0.33	0.34

Table 5.10: Minimal sample sizes for the RHR and RHRV hypothesis testing for *each* group in case of nominal $p = .05$ and adjusted p with Bonferroni correction, as well as the actual sample size n of PEV.

feature	nominal p	adj. p (2)	adj. p (2)	n
RHR	10	14	15	7
RHRV	16	19	21	7
Vsteps	5	14	17	6

5.6 Discussion

5.6.1 Summary of results

Feasibility

Thischapter presented results on data completeness, showing that both devices can be used for monitoring patients in ambulatory palliative care. The devices were accepted by all patients, and most patients had no difficulties with the devices or with the usage of the activity monitoring app. The combination of forgetfulness and lack of experience with smartphones reduces data completeness.

Differences in groups

Hypothesis testing revealed the following results:

- The exploratory analysis revealed, that the vital sign parameters resting heart rate and resting HRV (RMSSD) as well as step speed change differently in the two groups PEV and PnoEV with nominal p-values < 0.05 (not corrected for multiple testing).
- The analogous analysis of features extracted from patients' self-reports did not unveil any significant differences between PEV and PnoEV.
- Sensor signals, especially the vital signs resting heart rate and resting HRV as well as the walking speed outperformed patients' self-reports.

Predictive power of extracted features

Based on anomaly detection, emergency visits could not be predicted (details are given in section A.6). Possible reasons for this include the following:

- The study cohort varies in terms of cancer type, and comorbidities and life circumstances (e. g., city vs countryside, active in work life vs bound to wheel chair, etc.). Therefore, the extracted features are not only influenced by the worsening of the disease, but also by other uncontrollable factors.
- During the study, different types of emergencies were observed that may need different window sizes.

- An in-subject analysis of the feature changes (details are given in section A.7) revealed that patients are diverse and that individual prediction models should be trained based on a patient-specific feature selection.

5.6.2 Limitations

Sample size

In order to achieve statistically valid results, a study with a larger sample size has to be performed.

Observational study

Since this study was observational with an exploratory study design, not all parameters necessary to control important co-founding variables such as medication were tracked systematically. Furthermore, the exploratory data analysis does not allow the stating of p-values.

Used devices

The used sensor is still under development. During the study period, improved firmware became available, which was not used in order to prevent artificial shifts in measurements. However, considering the strict ethical requirements of study devices and the devices available on the market, the Everion® is still the best available option. Besides the firmware and software under development, the battery capacity of the Everion® limited data completeness. Runtime and, therefore, data completeness can be increased (up to 30 hours of runtime according to the manufacturer) by sacrificing the SpO_2 measurement. Another solution is to equip each patient with two sensors worn alternately.

In this study, only Android phones were used. For a more natural setting, also iPhones should be included.

Sparse ground truth

The only reliable ground truth of the collected data was that patients went to hospital for emergency reasons or were re-hospitalised (although there is no guarantee that all such events were registered) and provided subjective self-reports. However, it is not guaranteed that all

emergency visits (e. g., in a different hospital than the university hospital in Zürich) were reported by the patients, and the circumstances of emergency visits were not recorded in detail. Developed classifiers could benefit from being trained with separate datasets with richer and more precise labelling.

6

Conclusion and outlook

This chapter summarises the thesis with focus on achievements and limitations, leading to the final conclusions. The chapter ends with giving an outlook for future work.

6.1 Summary and achievements

This thesis investigates the possibilities of activity monitoring of outpatients in palliative care by means of wearable devices such as smartphones and an arm-band. The topic comprises of three questions:

- How should a monitoring system look like in order to reduce non-elective re-admissions to hospital for outpatients in palliative care?
- Is activity monitoring accepted by patients that have to deal with a severe disease and face their end of life?
- Which sensor modalities are useful in order to predict non-elective re-admissions to hospital?

To answer these questions, a monitoring system was developed following a patient-centric design involving interviews with palliative cancer patients. The monitoring system consists of an activity monitoring app running on Android smartphones and an arm-sensor. The activity monitoring app was evaluated in a pre-study in patients suffering from cancer-related fatigue. The Everion® was chosen as the arm-band since it provides multiple measurements such as heart rate, heart rate variability and galvanic skin response. At the same time, it enables full data access and privacy control by means of an Android SDK. Finally, the whole system comprising both wearable devices was evaluated in an observational study involving 31 palliative cancer patients. This thesis features technical and methodological achievements as well as medical achievements and findings.

6.1.1 Technical and methodological achievements

- An activity monitoring app was developed under consideration of requirements concerning usability (power consumption, questionnaire design) and data security (storage and transmission) and robustness. 37 patients with an age of up to 85 years used the app over several weeks with positive feedback such as “easy to use”.
- Conducting the studies unveiled a set of best practises and recommendations for future studies. For example:

  - A regular (time interval depends on buffer size of devices) check of the data transmitted by the wearable devices is necessary to improve data completeness.
  - Weekly calls for interviews were appreciated by most patients as an opportunity to talk.
  - Device-related questionnaires should be formulated in a patient-focused way that focuses on the patient ("How are you dealing with the devices?" instead of "How good are the devices?").
  - A recommended cohort size n for future studies to confirm the results of the palliative care study described in this thesis was estimated by a power analysis. A future study should target a minimum cohort size of 84 patients.
  - The diversity of the study cohort ranged from active patients still going to work (e. g., taking care of their farm) to patients bound to a wheel-chair and resting mainly at home. To control this diversity, a future study can increase the cohort size and systematically log possible confounders such as daily life circumstances and changes of medications or restrict the inclusion criteria resulting in a more homogeneous study cohort.

- A feature extraction pipeline for efficient storage of data and of processing results was developed. This pipeline comprises several pre-processing steps like re-sampling and timestamp correction, GPS clustering, HRV analysis and feature extraction on sliding windows. Emotion analysis based on voice and Wi-Fi clustering was investigated.

- The possibilities and limitations to apply unsupervised methods on a real-world dataset with sparse labelling and restrictions to preserve the privacy of patients, were explored:

  - Clustering of GPS data results in extraction of places. Home clusters can reliably be identified as the cluster that is occurring most often in the dataset of a patient.
  - Clustering of Wi-Fi results in extraction of Wi-Fi situations. However, the lack of additional labels or information about rooms at the patients' homes increased fine-grained location uncertainty, e. g., in associating the extracted clusters

to rooms of a patient's home. Additional information such as the number of rooms at a patient's home can be used as input parameter for a clustering method (e. g., number of clusters) in order to facilitate the clustering and to increase the interpretability of Wi-Fi clustering result.

- For emotion analysis of recorded phone calls, a classifier was trained on an artificial dataset labelled with seven classes of emotions. The classifier performance is comparable to results reported in literature. However, the application on the recorded phone calls faced two challenges: the recorded audio may contain the voice of other speakers in the foreground (e. g., the phone call partner) and background (e. g., television). Methods for speaker diarization and identification of the relevant speaker as well as filtering of background noise have to be developed on a labelled dataset that does not underly restrictions concerning privacy. Literature second-guesses that a classifier for emotions trained on an artificial dataset will work with the same classification accuracy. The lack of emotion labels does not allow to evaluate the classification results on the real-world phone calls.
- The analysis of vital signs worked well without additional labelling. Additional information such as a patient's hyperthyroidism was supported by the measured vital signs.

6.1.2 Medical achievements and findings

- For the first time, cancer patients in ambulatory palliative care were observed over a period of 12 weeks by means of a weekly assessment of QoL and a daily assessment of pain and distress. In addition, a comprehensive dataset including physical activity and social activity as well as vital signs was collected.
- Data analysis involving machine learning and statistics were applied to the data sets collected during the CRF study and the PC study, unveiling new knowledge:
  - Analysis of retention showed that activity monitoring with smartphones and the Everion® is accepted and feasible in these patient groups.

- The subjectively perceived CRF varies with multiple patterns throughout the day.
- The self-reported daily pain and distress in palliative patients showed less variance than the weekly reported global quality of life. Furthermore, a comparison of self-reports between the groups PEV and PnoEV did not reveal any significant differences.
- Palliative patients with emergency visits differed most significantly from those without emergency visits in these parameters: resting heart rate and resting heart rate variability as well as the speed of steps measured by the smartphone.
- The sensor data showed more differences between the two groups than the self-reports.

6.2 Limitations

The studies conducted in this thesis followed an exploratory approach with the goal to gain knowledge about feasibility and to identify features of behavioural change related to emergency visits and quality of life. The exploratory approach contains to the following limitations:

- Considering the number of diverse sensor modalities and features analysed, the cohort sizes n of the studies are too small ($n \leq 30$) for parametric statistics and multiple testing.
- The diversity of the patient groups also hinders the aim to find distinct and discriminative features. In a future study, inclusion criteria should be more restrictive or more patient related data should be logged systematically. This additional, patient related data could be used by a predictive model or classifier to control for such other parameters like changing treatments.

The observational study design allowed to gain first experiences with activity monitoring of palliative outpatients. However, the observational design leads to the following limitations:

- When designing the study, it was unknown how many tasks would the palliative patients tolerate. In order to not excessively burden the patients, labelling tasks for the patients were limited to the digital questionnaires once per day and to the weekly interviews. As a consequence, labelling in the data set is sparse.

- Since the palliative care study was designed as a feasibility study, patients treatment and course of disease were not recorded systematically. As a consequence, changes in sensor data, e. g., reduced or elevated vital signs, can be due to a changed treatment plan.

The used arm-sensor is still under development. The early development state leads to the following limitations:

- The BLE connection was not reliable. This lead to additional interaction with the patients resulting in a possible influence of study results.

6.3 Conclusions

Based on this thesis, the following conclusions can be drawn:

- Remote monitoring as realised in this thesis is feasible and accepted in the examined patient groups.
- Predictive models have to be personalised because of the diversity of the patients.
- Sensor-based activity monitoring outperformed patients' subjective self-reports in this study cohort.
- Parameters with significant change before a non-elective visit to the hospital were: resting heart rate, resting heart rate variability and speed of steps.

6.4 Outlook

To advance research in this field, a larger study should be conducted. The preparation need some additional development.

Activity recognition

More information can be extracted by using the Google API for Android and Apple's iOS ARKit or by developing a classifier on a labelled dataset. However, when selecting or creating a dataset, one has to take into account that the phone can be worn in any arbitrary position on the body.

Design of an Intervention

Following the recommendations of section A.8.2, a future study can add an intervention in the form of a notification or call through a professional care provider based on the identified feature candidates. The design of the intervention should be explored by following a patient-centric approach involving interviews. These interviews should investigate what kind of intervention patients would appreciate and when it would be effective. The interviewees should be recruited from non-elective re-admissions and non-elective ambulatory hospital visits. The interviews should shed light on the course of events that led to the emergency event. Through conversation with a patient, the interviewer should try to identify the points in time where it would be useful to intervene.

Continuous observation

During a future study, the data has to be observed and analysed continuously in order to start an intervention based on criteria that remain to be defined. Therefore, a processing tool chain has to be developed that processes automatically the incoming data on a daily basis and delivers reports to the study staff. For the design of the reports, physicians and other care givers could be interviewed in order to gather ideas about optical form and content.

Voice analysis

Voice analysis could help to estimate a patient's emotional status. There is the need for further development in a natural setting with labelled data to solve the issue of speaker diarization and use a natural emotion dataset such as Wizard of Oz for building a classifier. There are also pre-trained classifiers available [154]. However, none of them is trained on a natural dataset. Furthermore, to reduce the amount of transferred data, emotion recognition could be implemented on the smartphone. More details on the topic of voice analysis is given in section A.3.1.

A

Appendix

A.1 Performance scales

A.1.1 ECOG Performance Status

The Eastern Cooperative Oncology Group (ECOG) defined a performance status that evaluates the capabilities of a patient from 0 (fully active, without any restrictions compared to the pre-disease status) to 5 (death).

A.1.2 Karnofsky performance status scale

The Karnofsky performance status scale is another commonly used status scale and can be mapped to the ECOG as shown in Table A.1.

Table A.1: Mapping between ECOG and Karnofsky.

ECOG	Karnofsky
0	90 - 100
1	70 - 80
2	50 - 60
3	30 - 40
4	10 - 20
5	-

A.2 Used paper-based questionnaires

A.2.1 ESAS

The Edmonton Symptom Assessment Scale ia a validated questionnaire to assess crucial symptoms in palliative care patients [155]. It consists of nine VAS assessing the following symptoms as perceived by the patient during the last 24 hours:

- current pain
- maximal pain
- minimal pain
- tiredness

- nausea
- depression
- anxiety
- drowsiness
- lack of appetite
- wellbeing
- other

For all scales, 0 means that symptom not existent (or best wellbeing), 10 means that a symptom is very present.

A.2.2 EORTC QLQ-C30

The hereafter shown specimen of the QLQ-C30 [156] is protected by copyright and authorization for any use of a measure of the EORTC group has to be requested by directly contacting the QoL-Department[1] prior to usage.

ENGLISH

EORTC QLQ-C30 (version 3)

We are interested in some things about you and your health. Please answer all of the questions yourself by circling the number that best applies to you. There are no "right" or "wrong" answers. The information that you provide will remain strictly confidential.

Please fill in your initials:
Your birthdate (Day, Month, Year):
Today's date (Day, Month, Year): 31

	Not at All	A Little	Quite a Bit	Very Much
1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?	1	2	3	4
2. Do you have any trouble taking a long walk?	1	2	3	4
3. Do you have any trouble taking a short walk outside of the house?	1	2	3	4
4. Do you need to stay in bed or a chair during the day?	1	2	3	4
5. Do you need help with eating, dressing, washing yourself or using the toilet?	1	2	3	4

During the past week:

	Not at All	A Little	Quite a Bit	Very Much
6. Were you limited in doing either your work or other daily activities?	1	2	3	4
7. Were you limited in pursuing your hobbies or other leisure time activities?	1	2	3	4
8. Were you short of breath?	1	2	3	4
9. Have you had pain?	1	2	3	4
10. Did you need to rest?	1	2	3	4
11. Have you had trouble sleeping?	1	2	3	4
12. Have you felt weak?	1	2	3	4
13. Have you lacked appetite?	1	2	3	4
14. Have you felt nauseated?	1	2	3	4
15. Have you vomited?	1	2	3	4
16. Have you been constipated?	1	2	3	4

Please go on to the next page

[1] http://groups.eortc.be/qol/

During the past week:	**Not at All**	**A Little**	**Quite a Bit**	**Very Much**
17. Have you had diarrhea?	1	2	3	4
18. Were you tired?	1	2	3	4
19. Did pain interfere with your daily activities?	1	2	3	4
20. Have you had difficulty in concentrating on things, like reading a newspaper or watching television?	1	2	3	4
21. Did you feel tense?	1	2	3	4
22. Did you worry?	1	2	3	4
23. Did you feel irritable?	1	2	3	4
24. Did you feel depressed?	1	2	3	4
25. Have you had difficulty remembering things?	1	2	3	4
26. Has your physical condition or medical treatment interfered with your family life?	1	2	3	4
27. Has your physical condition or medical treatment interfered with your social activities?	1	2	3	4
28. Has your physical condition or medical treatment caused you financial difficulties?	1	2	3	4

For the following questions please circle the number between 1 and 7 that best applies to you

29. How would you rate your overall health during the past week?

1	2	3	4	5	6	7
Very poor						Excellent

30. How would you rate your overall quality of life during the past week?

1	2	3	4	5	6	7
Very poor						Excellent

A.2.3 Previous experiences with electronic devices

Wearable Computing Lab. ETH Zürich

UniversitätsSpital Zürich Klinik für Radio-Onkologie

Fragebogen

zu Ihren Erfahrungen mit elektronischen Geräten

Name, Vorname:_____Geburtsdatum:_____

1. Besitzen Sie eines oder mehrere der folgenden Geräte und wenn ja, wie lange schon benutzen Sie es?

O Mobiltelefon seit _____ Jahren/ Monaten (Zutreffendes bitte unterstreichen)

O Smartphone seit _____ Jahren/ Monaten

O Tablet-Computer seit _____Jahren/Monaten

O Laptop/Computer seit _____ Jahren/Monaten

O Tracking-Armband oder Pulsuhr (z.B. Polar) seit _____ Jahren/Monaten

O Blutdruckmessgerät seit _____ Jahren/Monaten

O Blutzuckermessgerät seit _____ Jahren/Monaten

O _____ seit _____ Jahren/Monaten

O Besitze kein elektronisches Gerät

2. Hatten Sie schon Umgang mit einem Schlafmessgerät?

O Nein O Ja, und zwar _____

3. Wenn Sie keines der Geräte besitzen, wären Sie dennoch an der Benutzung eines solchen Gerätes interessiert?

O Nein, sicher nicht O Vielleicht/Kommt drauf an O Ja, interessiert mich

4. Mögen Sie uns einen Grund für Ihre Antwort nennen?

_____

_____

_____

A.2.4 Previous experiences with electronic devices (englisch)

Wearable Computing Lab. ETH Zürich

UniversitätsSpital Zürich Klinik für Radio-Onkologie

Questionnaire

on your experience with electronic devices

Last name, first name:_____ Birthdate:_____

1. **Do you have one or several of the following devices and if yes, for how long have you been using it?**

○ Mobile phone for _____ years / months (please underline what applies)

○ Smartphone for _____ years / months

○ Tablet computer for _____ years / months

○ Laptop / Computer for _____ years / months

○ Activity-tracking bracelet or watch (e.g. Polar) for _____ years / months

○ Blood pressure meter for _____ years / months

○ Blood sugar meter for _____ years / months

○ _____ for _____ years / months

○ I don't have any electronic device

2. **Have you ever used a sleep monitoring device?**

○ No ○ Yes, namely _____

3. **If you don't have any of the abovementioned devices, would you still be interested in using one of such devices?**

○ No, definitely not ○ Maybe / it depends ○ Yes, I'm interested

A.2.5 How are you doing with the devices

Wearable Computing Lab. ETH Zürich

UniversitätsSpital Zürich Klinik für Radio-Onkologie

Bewertungsfragebogen

Wie geht es Ihnen mit den Geräten?

1 Gab es in dieser Woche Probleme im Umgang mit den Geräten und wenn ja, welche?

O Mit dem Aufladen Smartphone

O Mit der Bedienung des Smartphones

O Mit dem Aufladen des Trackingarmbands

O Mit der Bedienung des Trackingarmbands

O Mit der Anwendung der Apps

_____

_____

_____

_____

2a In den vergangenen 7 Tagen: An wie vielen Tagen haben Sie das Armband getragen?

O 1 O 2 O 3 O4 O5 O6 O7

2b In den vergangenen 7 Tagen: In wie vielen Nächten haben Sie das Armband getragen?

O 1 O 2 O 3 O4 O5 O6 O7

2b Was waren Gründe dafür, das Armband nicht zu tragen (tags oder nachts)?

_____

_____

2c In den vergangenen 7 Tagen: An wie vielen Tagen haben sie das Smartphone immer bei sich gehabt?

O 1 O 2 O 3 O4 O5 O6 O7

2d Was waren Gründe dafür, dass Sie das Smartphone nicht bei sich trugen?

_____

_____

3a Die Bedienung der Apps (Schmerzskala und Distressthermometer)

○ fällt mir leicht ○ braucht etwas Gewöhnung ○ ist mühsam

3c Was finden Sie an der Bedienung der Apps mühsam, was gefällt Ihnen nicht?

_____

_____

4 Gibt es noch etwas, dass Sie uns in Bezug auf den Gebrauch der Geräte mitteilen möchten?

_____

_____

_____

_____

A.2.6 How are you doing with the devices (english)

Wearable Computing Lab.
ETH Zürich

UniversitätsSpital
Zürich

Klinik für
Radio-Onkologie

Assessment questionnaire
How are you doing with the devices?

1 In this week, did you have any troubles handling the devices and if yes, which?

- ○ Charging the phone
- ○ Operating on the phone
- ○ Charging the tracking bracelet
- ○ Operating on the tracking bracelet
- ○ Using applications on the phone

_____

_____

_____

_____

2a In the past 7 days: On how many days did you wear the bracelet?

○ 1 ○ 2 ○ 3 ○4 ○5 ○6 ○7

2b In the past 7 days: In how many nights did you wear the bracelet?

○ 1 ○ 2 ○ 3 ○4 ○5 ○6 ○7

2c Which were the reasons for not wearing the bracelet (in the daytime or at night)?

_____

_____

2d In the past 7 days: On how many days did you always take the phone with you?

○ 1 ○ 2 ○ 3 ○4 ○5 ○6 ○7

2d Which were the reasons for not carrying the phone with you?

_____

_____

3a Operating in the applications (pain scale and distress thermometer)

O is easy for me O requires some adaption O is difficult

3b What do you find difficult about using the applications, what do you dislike?

_____

_____

4 Is there anything else you would like to tell us regarding the handling of the devices?

_____

_____

_____

_____

A.2.7 NCCN Distress Thermometer

The NCCN published guidelines and a questionnaire to assess patients' distress and problems [157]. The NCCN Distress Thermometer cannot be included in this thesis due to copyright. The German version as used in the palliative care study was validated by Mehnert et al. [158].

A.2.8 FACIT-F

In this work, fatigue is assessed using the FACIT-F fatigue scale (version 4) (Functional Assessment of Chronic Illness Therapy) [159]. It is a validated instrument to measure fatigue in cancer patients as well as patients suffering from other chronic diseases.

A.2.9 Final interview

Wearable Computing Lab.
ETH Zürich

UniversitätsSpital Zürich
Klinik für Radio-Onkologie

Abschlussbefragung

1. Wie ist jetzt nach Abschluss der Studie Ihr Gesamteindruck? Was ist gut gelaufen, was war schwierig?

2. Wie sind sie mit den Geräten zu Recht gekommen?

3. Wie regelmässig haben Sie sie getragen?

4. Wenn Sie sie nicht getragen haben, woran lag das jeweils?

5. Gab es etwas, dass Sie als besonders lästig/unangenehm empfunden haben?

6. Gab es etwas, dass Ihnen gut gefallen hat?

7. Gab es Reaktionen aus Ihrer Umwelt/ von anderen Menschen und wenn ja, welche?

8. Gibt es Empfehlungen, die Sie uns als Forschungsteam geben, wenn sie jetzt daran denken, dass die Geräte regelmässig bei Patienten eingesetzt werden sollen?

9. Was würden Sie anderen Patienten über Ihre Erfahrungen sagen?

A.2.10 Final interview (english)

Wearable Computing Lab.
ETH Zürich

UniversitätsSpital Zürich
Klinik für Radio-Onkologie

Final Interview

ID, Date

1. Now, after finishing the study, what's your general impression? What went well, what was difficult?

2. How did you manage handling the devices?

3. How regularly did you wear them?

4. If you didn't wear them, what was the reason for it?

5. Is there anything you found in particular annoying/uncomfortable?

6. Is there anything you liked?

7. Did you get any feedback from your surroundings / other people and if yes, which was it?

8. Would you give us any recommendations as a research team if you imagine the devices to be used regularly with patients?

9. What would you tell other patients about your experiences?

A.3 Details of pre-processing methods

A.3.1 Voice analysis

For voice analysis, the python library [107] was used, since it provides methods for all required tasks.

Voice Preprocessing

Since the audio files are encrypted and the privacy of the patients has to be protected, the following steps are performed in a chain:

- decryption
- trans-code to wav format and delete decrypted file
- audio analysis task
- delete wav file

Segmentation

Usually, during phone calls, the speakers alternate. After analysis of different phones used in the study, it turned out that not all data can be used while preserving the privacy of the patients: some phones record only the user of the smartphone, some phones record also the other speaker. This leads to two segmentation tasks.

Silence removal. Using a SVM on features (MFCC, Spectral, Harmonic, Chroma, zero crossing rate, energy, entropy, etc., more details in [107]) extracted from 50 ms windows with 50 % overlap, the periods of silence can be identified and removed. It remains segments of speaking. The number of segments is equivalent to the number of turns in case of phones that record only the smartphone user and not the call partner. Figure A.1 gives an example

Speaker Diarization. Figure A.2 shows a conversation where probably both speakers are recorded since there is almost no silence. Since this pattern shows in every call of patients using a phone generation more recent than the Samsung Galaxy S5 and in none of the patients using the Samsung Galaxy S5, a test with two such phones confirms

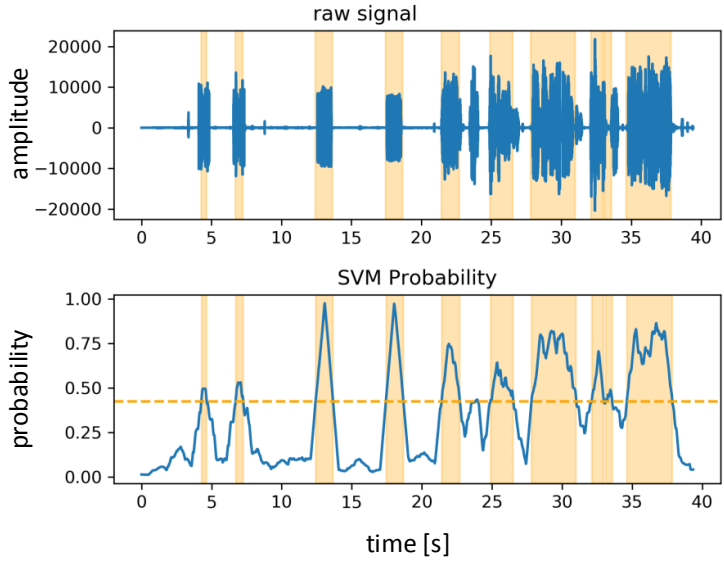


Figure A.1: Silence detection using a SVM with probability output. The in yellow highlighted areas are the periods of a conversation with voice recognised. It is also visible, that only the smartphone microphone is recorded.

that only the Samsung Galaxy S5 model records only the smartphone user.

Having both voices recorded leads to the problem of speaker diarization which is usually tackled by clustering [107]. However, none of the methods found in literature provides sufficient performance in order to use it in a completely unsupervised manner. Furthermore, after identification of different speakers, the patient has to be recognised out of all other speakers. methods for speaker diarization and identification have to be developed and evaluated on a labelled dataset before applying them to a dataset where the privacy of study participants has to be protected. In a future study, a voice sample of each patient could be taken in order to facilitate the task of voice identification.

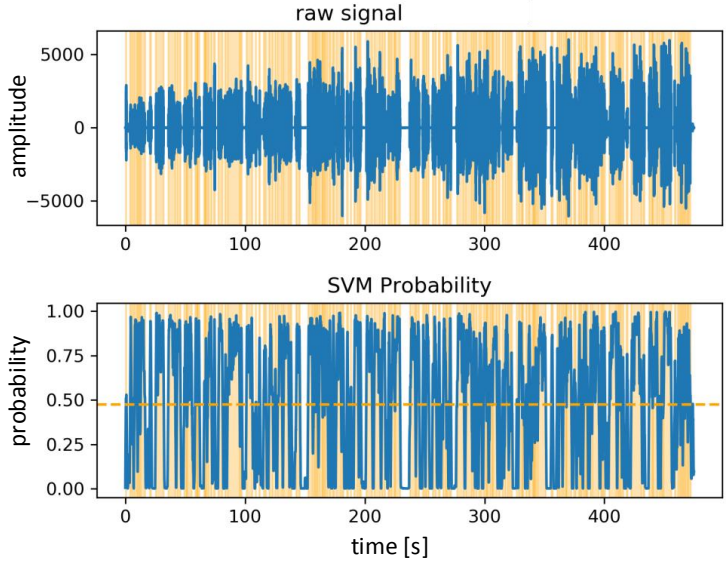


Figure A.2: Recorded phone call without silence periods and both call partners recorded: silence removal marked with yellow the periods with recognised voice.

Cosine similarity

It was investigated if voice features that are commonly used for emotion classification change over the time. For each patient, all calls are compared with each other. for each phone call, a feature vector is created by the following method: after silence removal, a phone call is segmented by the detected periods of silence. For each segment, features as described in [107] are extracted and aggregated using the statistics mean, min, max, range, standard deviation, kurtosis and skewness. This aggregation leads to a vector of 238 features per segment. These feature vectors are further aggregated by using the mean, yielding in one feature vector per phone call. A phone call is reprepresented as feature vector in a 238-dimensional vector space. For comparison, the cosine similarity is calculated pairwise (i. e., for each pair of two feature vectors $F_i, F_j, i, j = 1, \ldots, m; m$ number of phone calls). The cosine simi-

larity is defined for two feature vectors $\mathbf{F} = (f_1, \ldots, f_n), \mathbf{G} = (g_1, \ldots, g_n)$ as

$$\cos(\mathbf{F}, \mathbf{G}) = \frac{\mathbf{FG}}{||\mathbf{F}||||\mathbf{e}||} = \frac{\sum_{i=1}^{n} f_i g_i}{\sqrt{\sum_{i=1}^{n} f_i^2} \sqrt{\sum_{i=1}^{n} g_i^2}}$$

Figure A.3 shows that from the extracted audio features, it is distinguishable when the patient was dismissed from rehabilitation. Some calls show different similarities to the adjacent calls. The example shows that it is important to consider also the environment as well as contextual information for interpretation.

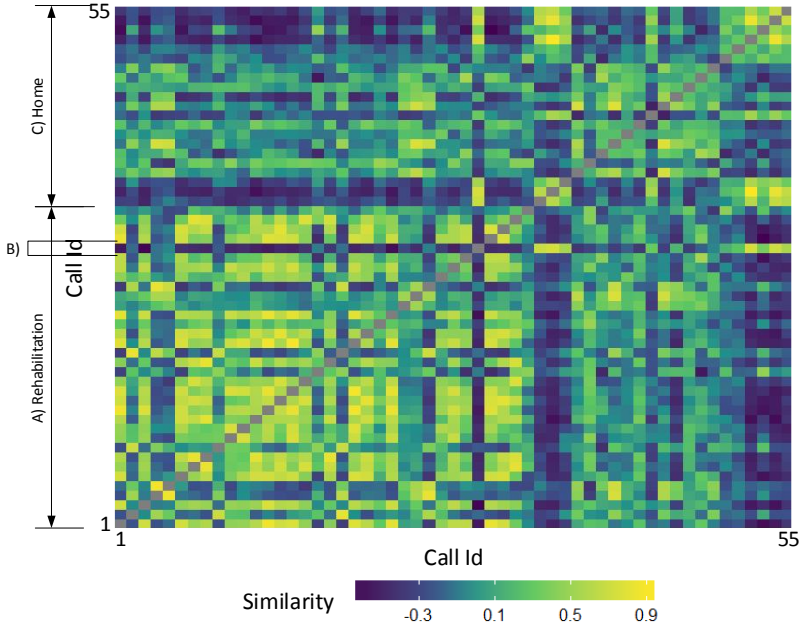


Figure A.3: Pairwise cosine distances between phone calls. A) Calls made during rehabilitation are very similar. The patient used the smartphone loudspeaker and background noise is hearable during weekly interview (television); B) this day is different from other calls during rehabilitation, but similar to calls made at home; C) Calls at home are more similar to each other than to rehabilitation calls – especially the last 6 calls

Emotion recognition

To classify emotions, a SVM was trained on the EMO dataset [99] using the same features as extracted during silence removal, with a performance as listed in Table A.2. The classifier performs best for sadness, anger and disgust.

Table A.2: Precision, recall and F1-score for 7 emotions.

Class	Precision	Recall	F1-Score	Support
A (fear)	0.69	0.6	0.64	15
E (disgust)	0.77	0.83	0.8	12
F (happiness)	0.55	0.61	0.58	18
L (boredom)	0.83	0.65	0.73	23
N (neutral)	0.56	0.78	0.65	18
T (sadness)	1	0.92	0.96	13
W (anger)	0.94	0.89	0.91	35
average/total	0.78	0.76	0.77	134

The confusion matrix of Figure A.4 shows that boredom and fear are mainly mixed with happiness. However, [108] show that features can differ between natural and acted emotions.

Conclusion for voice features

To conclude this section, it can be stated that voice features in terms of similarity or classified emotions can be used when taking precautionary measures to ensure that the input data are clean, i. e., only the patients' voice is recorded and background noise like television is filtered out. Since the dataset of the palliative care study cannot be cleaned sufficiently with preserving the patients' privacy, voice features are not used in later analysis.

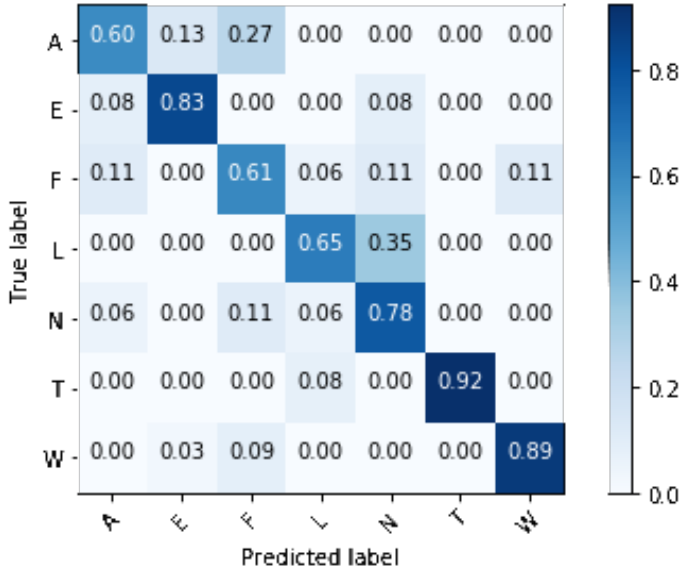


Figure A.4: Confusion matrix for emotion classification.

A.3.2 Everion® preprocessing: resampling algorithm

Algorithm A.1 was used to correct the timestamps of the Everion® signals.

Algorithm A.1: Correction of timestamp anomalies.

input : dataframe of vital signs with counter and timestamps
output: input dataframe with corrected timestamps

df ← data to resample (input);
tDiff ← -1, `Diff(`df.*timestamp*`)` ;
tDiff [last] ← -1;
negGaps ← indices where tDiff < 0;
segments ← tuples (df [negGaps [i]:negGaps [i+1]],negGaps [i],negGaps [i+1]), $i \in 0, ..., last-2$;

for $s \in$ segments **do**
  tDiffSegment ← `Diff(`*s[0].timestamp*`)`;
  artCnt [0] ← 0;
  for $i \leftarrow 1$ **to** `length(`s[0]) − 1 **do**
    if tDiffSegment *[i] == 0* **then**
      artCnt [i] ← artCnt [i-1] + 1
    else
      artCnt [i] ← artCnt [i-1] + tDiffSegment [i]

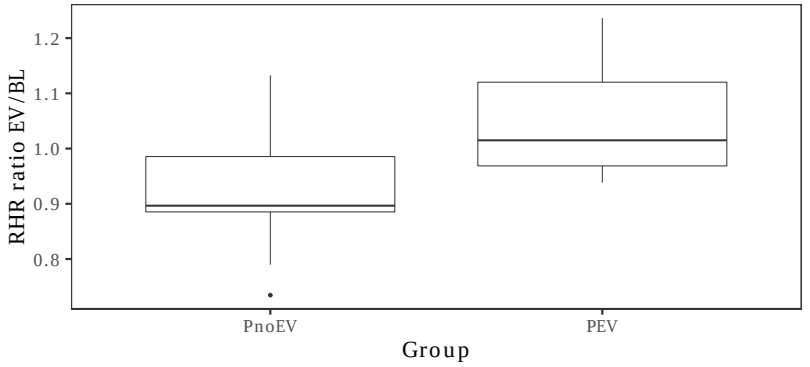
  resampleFactor ← $\frac{s[0].timestamp[last]-s[0].timestamp[first]}{\texttt{max(artCnt)+1}}$;
  newTime ← s[0].*timestamp*[0] + resampleFactor ∗ artCnt;
  `updateTime(`df *[s[1]:s[2]],* newTime`)`
return df

A.4 Data distributions of feature ratios

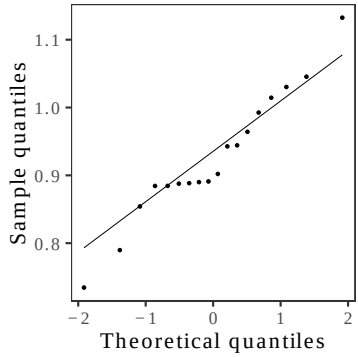
A.4.1 Distribution of RHR

As Figure A.5 shows, the data deviates from a normal distribution, especially in the PnoEV group. The box plot shows a right skewness in both groups. The variances in the groups are not equal, but the difference between the variances is small (0.006). The q-q plots in b) and c) show the deviations from the normal distribution for both groups.

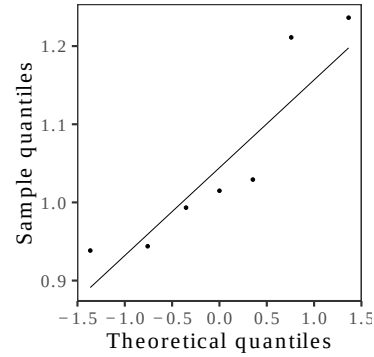
Hence, a Wilcoxon rank-sum test is preferred to a Welch's t-test. Since the variances are not equal, also a randomisation test was conducted to confirm the result of the Wilcoxon rank-sum test.



(a) Box plot of RHR



(b) q-q plot for RHR (PnoEV)



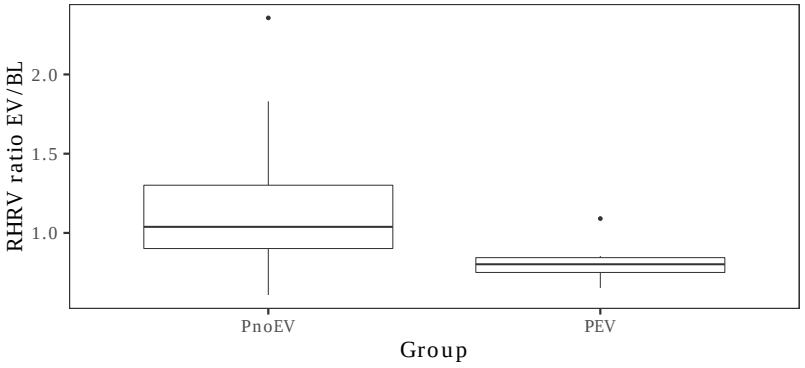
(c) q-q plot for RHR (PEV)

Figure A.5: Analysis of distribution of RHR in the groups PEV and PnoEV reveals that a Wilcoxon rank-sum test is preferred to a Welch's t-test because of the right skewness of the data, especially visible in a) and b).

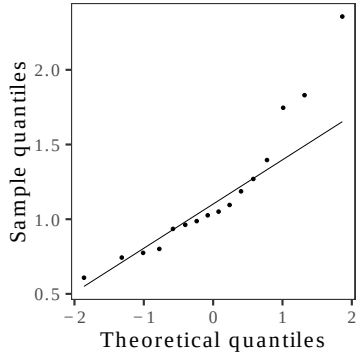
A.4.2 Distribution of RHRV

Figure A.6, subfigure a), shows that a Wilcoxon rank-sum test is not advised due to different variances in the two groups (0.2). The box

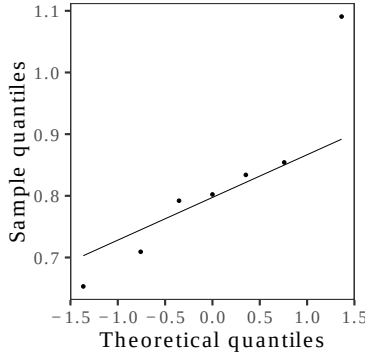
plot shows a right skewness in the PnoEV group. The q-q plot of the PnoEV group (subfigure b)) shows sufficiently normally distributed data points, when considering three largest points as outlier. Subfigure c) allows to assume that the data of the PEV group are sufficiently normally distributed for a t-test when considering the smallest and the largest points as outlier. Therefore, a Welch's t-test was conducted and the result confirmed with a randomisation test.



(a) Boxplot of RHRV



(b) q-q plot for RHRV (PnoEV)

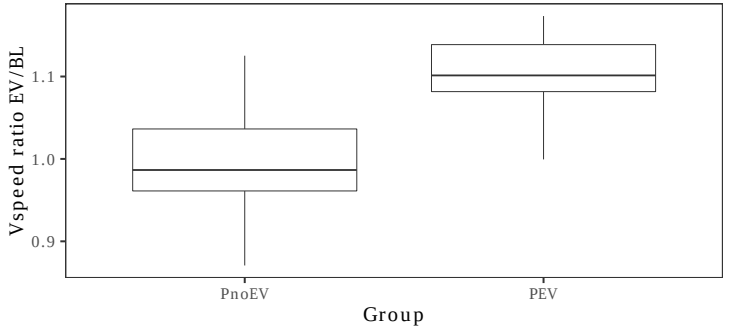


(c) q-q plot for RHRV (PEV)

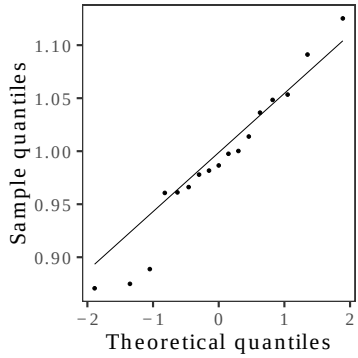
Figure A.6: Analysis of distribution of RHRV in the groups PEV and PnoEV shows different variances in the groups (a)) and sufficiently normally distributed data (b), c)) to allow a Welch's t-test.

A.4.3 Distribution of Vsteps

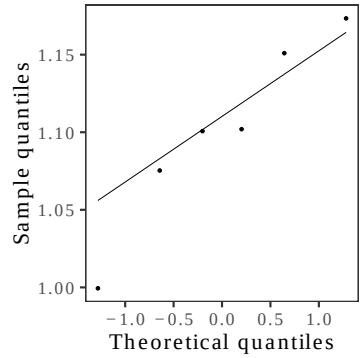
As Figure A.7 shows, the data in both groups are not normally distributed. The box plot in subfigure a) shows a right skewness in both groups. The variances in the groups are almost equal with a difference of 0.001. The q-q plots in b) and c) imply that both tests, Welch's t-test or a a Wilcoxon rank-sum test, are possible. To contribute to the right skewness, the Wilcoxon rank-sum test is preferred. A randomisation test was conducted to confirm the result of the Wilcoxon rank-sum test.



(a) Box plot of Vsteps



(b) q-q plot for Vsteps (PnoEV)



(c) q-q plot for Vsteps (PEV)

Figure A.7: Analysis of distribution of Vsteps in the groups PEV and PnoEV

A.5 Power analysis for group comparisons

A power analysis for the three features RHR, RHRV and Vsteps was performed. The power of a test describes the test's capability to confirm a hypothesis.

A.5.1 Methods for power analysis

For the t-test and Wilcoxon rank-sum test, different methods have to be used for a power analysis. The term power is illustrated by Table A.3. A statistical test estimates the probability of a null hypothesis H_0, given the data sample. If the test results in $p < .05$ (the commonly chosen threshold), the null hypothesis is unlikely with a probability of less than 5 %. Hence, α denotes the tolerated error rate that we accept H_A while H_0 is true.

Table A.3: Confusion matrix for statistical tests, given a null hypothesis H_0 and an alternative hypothesis H_A: the power is given by $1 - \beta$

		truth	
		H_0	H_A
predicted	H_0	$1 - \alpha$	β
	H_A	α	$1 - \beta$

Power analysis for a t-test. The power analysis for t-tests is performed using the R package `pwr` [160]. The `pwr` package provides an implementation of the analysis methods formulated by Jacob Cohen in [161]. As input, the method requires the effect size d (Cohen's d, given by Equation (A.1)), the requested significance level and power and the test type.

$$d = \frac{m_1 - M_2}{\frac{\sqrt{s_1^2 + s_2^2}}{2}} \tag{A.1}$$

where M_i denotes the mean of group i and s_i denotes the standard deviation of group i. Cohen's d is specific for each tested feature. Table A.4 lists the constant parameters as used in the following analysis. The method returns the required minimum sample size for both groups.

Table A.4: Fixed parameters for an independent, two-sided t-test for testing of 1, 2 or 3 hypothesis with Bonferroni-corrected significance levels

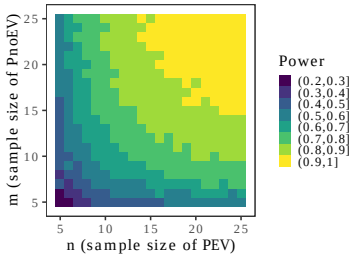
Parameter	Value (1)	Value (2)	Value (3)
sig. level	**0.05**	**0.025**	**0.0167**
power	0.8	0.8	0.8
type	two.sample	two.sample	two.sample
alternative	two.sided	two.sided	two.sided

Power analysis for a Wilcoxon rank-sum test. To estimate the sample size when using the Wilcoxon rank-sum test, a simulation is performed. In general, a power of 0.8 is targeted. To estimate the necessary sample size, a Wilcoxon rank-sum test is applied on randomly generated data of a specific sample size and tested if p is smaller the required significance level. This is repeated 1000 times yielding in a success rate (i. e., power) for the specific group sample sizes. To estimate the required sample sizes, this repeated simulation is performed on a grid of sample sizes with n denoting the size of group PEV and m denoting the size of group PnoEV. As input, the method requires the group means and a group standard deviation to generate the random group data as well as the required significance level.

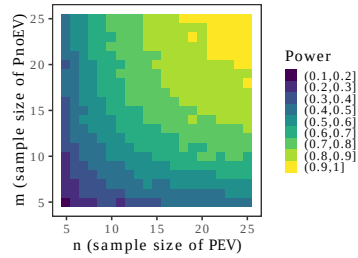
A.5.2 Power analysis for RHR

Since a Wilcoxon rank-sum test was performed to test the feature RHR, the power is estimated through simulation. The random data is generated as normal distributions with mean 1.01 for PEV and mean 0.89 for PnoEV and a common standard deviation of 0.11 (i. e., the mean of the standard deviations of the two groups).

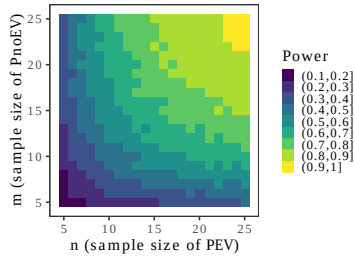
From the performed power analysis, a heat map is generated showing the power for given samples sizes n, m. Figure A.8 shows the simulation results for different significance levels. For the PEV group a minimal size of 10 is needed (given a minimal sample size of 22 of the PnoEV group) – up to 15 (given a minimal sample size of 25 of the PnoEV group) in case of multiple hypothesis testing.



(a) Power for $p < 0.05$



(b) Power for $p < 0.025$



(c) Power for $p < 0.0167$

Figure A.8: Power analysis for RHR and a Wilcoxon rank-sum test on simulated data with a standard deviation of $sd = 0.11$.

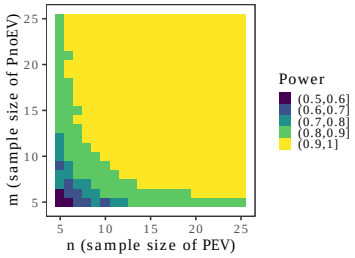
A.5.3 Power analysis for RHRV

Since an independent, two-sided t-test was performed to test the feature RHRV, the power analysis using the R package `pwr` was used with Cohen's $d = 1.037$. For the three different significance levels, a sample size per group of 16 to 21 patients is required. Compared to the sample size of 7 in the PEV group, the power analysis shows that the result should be confirmed as by an independent study with a larger study cohort.

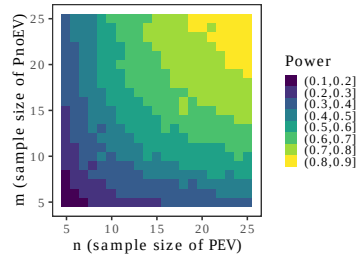
A.5.4 Power analysis for Vsteps

Since a Wilcoxon rank-sum test was performed to test the feature Vsteps, the power is estimated through simulation. The random data is generated as normal distributions with mean 0.99 for PEV and mean 1.1 for PnoEV and a common standard deviation of 0.066 (i. e., the mean of the standard deviations of the two groups).

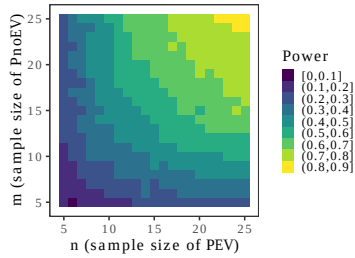
From the performed power analysis, a heat map is generated showing the power for given samples sizes n, m. Figure A.9 shows the simulation results for different significance levels. For the PEV group a minimal size of 5 is needed (given a minimal sample size of 13 in the PnoEV group) – up to 17 (given a minimal sample size of 22 in the PnoEV group) in case of multiple hypothesis testing.



(a) Power for $p < 0.05$



(b) Power for $p < 0.025$



(c) Power for $p < 0.0167$

Figure A.9: Power analysis for RHR and a Wilcoxon rank-sum test on simulated data with a standard deviation of $sd = 0.066$.

A.6 Prediction of emergency visits

In this section, the question is examined if emergency visits can be predicted based on recorded sensor signals.

A.6.1 Dataset for prediction

Out of 30 patients, 10 visited the hospital as an emergency. Two occurred during rehabilitation stay and therefore are excluded from the dataset. Therefore, 28 patients remained for the analysis. For the approach of anomaly detection by means of a one-class SVM, the dataset is divided in a training set that contains only samples without emergency samples and a test dataset that contains both, samples with emergency (EV) and samples without emergency (noEV). The labelling of samples is illustrated in Figure A.10. The dataset is split in two different manners:

General dataset includes noEV samples from all patients in the training dataset and noEV and EV samples from the 8 patients of the PEV group in the test dataset. This leads to a general classifier.

Patient-specific dataset generates 28 pairs of train/test datasets, where the first ⅔ of the days are used as training dataset and the last ⅓ of the data is used as test dataset.

A.6.2 Feature extraction for prediction

As described in section 3.4, 121 features were extracted. These features are used for the following analysis. This feature extraction yielded in hourly or daily sampled multi-variate time series per patient. Features on a sliding window of 5 days were extracted as illustrated by Figure A.10 by calculating mean and variance on the features described in section 3.4 leading to 279 features. The last sample of the dataset of a patient is either a) the one that is followed by a day with an emergency visit or b) followed by the last day of study participation.

All features are standardised to mean=0 and sd=1. Features were then selected by recursive feature elimination resulting in a set of 11 features as listed in Table A.5.

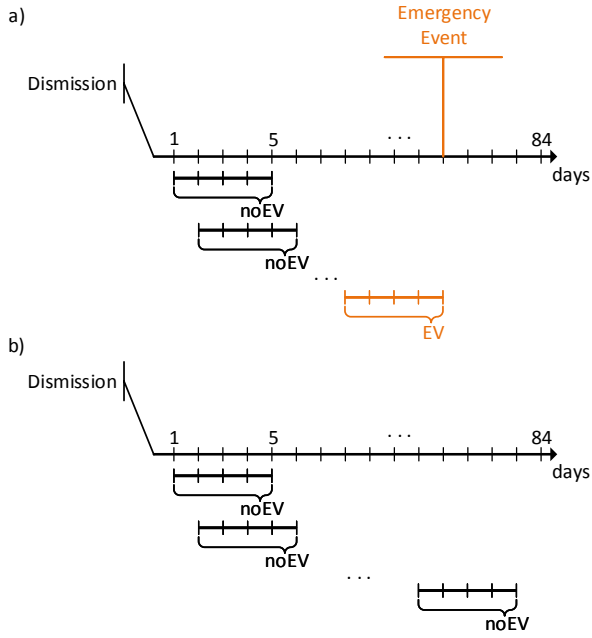


Figure A.10: Rolling windows with labelling for emergency prediction. a) depicts the case of a patient who as experienced an emergency visit, b) depicts the case of a patient who had no emergency visit

A.6.3 Classification method to predict emergency visits

Since the dataset is unbalanced with at most one EV sample per patient at 8 EV samples vs. 1703 samples without emergency visit, a binary classifier is inadequate for a patient specific classifier. Hence, anomaly detection by means of a one-class SVM with rbf kernel was applied.

A.6.4 Emergency prediction result

For the general dataset, the prediction result is low with an accuracy of 0.7. As shown by the confusion matrix in Table A.6, only 3 out of 8 emergency visits are detected, whereas 25 % of the noEV samples were wrongly detected as emergency visits. The detected emergency visits were of the patients 183, 194 and 351. Further metrics for both classes

Table A.5: Best feature set for one-class SVM.

feature	mean	variance
RHR	✓	✗
RRR	✓	✓
RHRV	✓	✓
BLPW.mean	✓	✓
BLPW.sd	✓	✓
BLPE.mean	✓	✗
BLPE.sd	✗	✓

are listed in Table A.7 and show that the detection of emergency visits using a general classifier is not feasible. Possible reasons are:

- The study cohort is varying too much for the sample size.
- The circumstances of each patient and reasons leading to an emergency visit or re-hospitalisation are different. Therefore, each case should be considered separately.
- The fixed window size of 5 days does not reflect the different causes for re-hospitalisation (e. g., epileptic seizure vs. deterioration of symptoms)
- If not an acute emergency (e. g., epileptic seizure) is present, the decision to go to the hospital or try to stay at home is personal. One patient might feel that she/he has to go to see a doctor, whereas another would decide to try to stay at home. Therefore, in addition to the small amount of available labels, the labels also include subjective judgement that is not captured by any feature and therefore cannot be taught to a classifier.

For the patient specific datasets, none of the 8 emergency visits could be predicted. Patient individual feature selection can probably improve this. The behaviour of features in individual patients is given in section A.7.

A.7 In-subject analysis of features

Using the change of features $\mathfrak{C}(f^i)$, the complete feature set is analysed for each patient. Table A.8 gives the frequencies how often a feature

Table A.6: Confusion Matrix for EV prediction

True Class \ Predicted Class	noEV	EV
noEV	39	13
EV	5	3

Table A.7: Performance metrics of EV prediction

	noEV	EV
Accuracy	0.7	
Precision	0.88	0.18
Recall	0.75	0.36
F1-score	0.81	0.27

has the highest ratio event to baseline. Table A.9 lists the frequencies of how often a feature is the top feature having the biggest decrease with respect to baseline. Interestingly, the increasing ratios differ more between the two patient groups than the decreasing features. In patients without emergency visit, the features measuring physical activity indicate that single patients had their biggest change in becoming physically more active, e. g., they walked more or faster, had a higher variance in the activity signal and covered a larger distance outside from home. This indicates that over the study period, they could recover.

In contrast, similar features decreas from baseline to event period in both patients groups. The set of features with biggest decreases is a mix of all sensor modality categories such as call statistics, GPS based, physical activity and vital signs.

Figure A.11 shows a summary of the last week before an emergency visit. The lighter a day, the more features were changed to an outlier range (i. e., $1.5 \times IQR$ (inter quartile range)) at least for one hourly value of that day. As comparison, the baseline week (BL) itself is shown. It is the first week at a patient's home. The week before the event (EV) shows more variation than the baseline week, especially of patients 187 and 423. Most patients show in the EV week more increases than decreases except for patient 430. For patients 183 and 246, there are more changes during the last three days before the event, whereas for the other patients, there is not a specific day at which a period with more changes starts.

For comparison, Figure A.12 represents in the same way the last week of study participation for PnoEV patients. Except for patients 23, 324, the baseline week (BL) and last week (EV) look similar.

Table A.8: Number of biggest increases in patients ordered by frequencies in EV patients.

Feature name	noEV	EV	total
acc.var	2	2	4
cntComingHome	0	1	1
cntLeavingHome	0	1	1
cor.p	0	1	1
firstLeavingHomeDay	1	1	2
gsr.max	0	1	1
gsr.sd	0	1	1
maxDistFromHomeDay	0	1	1
step.cnt	4	1	5
acc.varDay	1	0	1
act.var	1	0	1
cnt.spo2	2	0	2
cntWalkingDay	2	0	2
gsr.mean	1	0	1
gsr.min	1	0	1
movingDay	1	0	1
movingPercent	1	0	1
movingPercentDay	1	0	1
step.speed.var	1	0	1
sumDistOutside	2	0	2
walkingpercent	1	0	1
walkingpercentDay	1	0	1

Table A.9: Number of biggest decreases in patients ordered by frequencies in EV patients.

Feature name	noEV	EV	total
maxDistFromHomeDay	2	2	4
acc.varDay	2	1	3
cntWalking	1	1	2
firstLeavingHomeDay	0	1	1
ipi.lf_mean	0	1	1
maxDistFromHome	6	1	7
onbodypercentDay	0	1	1
sum.total.out	0	1	1
walkingpercent	1	1	2
cnt.spo2	2	0	2
gsr.sd	4	0	4
moving	1	0	1
movingPercent	1	0	1
sumDistOutside	1	0	1
t.range.day	1	0	1

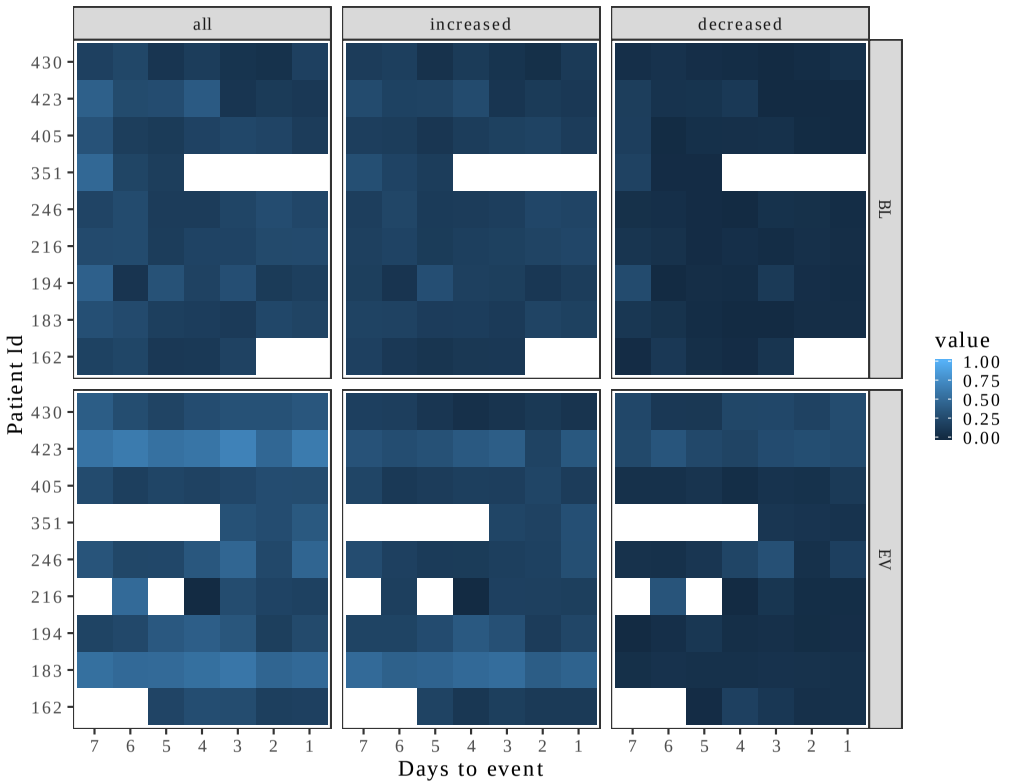


Figure A.11: Heat map for patients with emergency visit (PEV), showing the last week before the emergency visit (EV). A value of 1 (light colour) means that all available features have an hourly value with a distance of at least $1.5 \times IQR$ (inter quartile range) from baseline week (BL). The three columns show all changes, only increases, i.e., only $> 1.5 \times IQR$, and only decreases, i.e., only $< 1.5 \times IQR$.

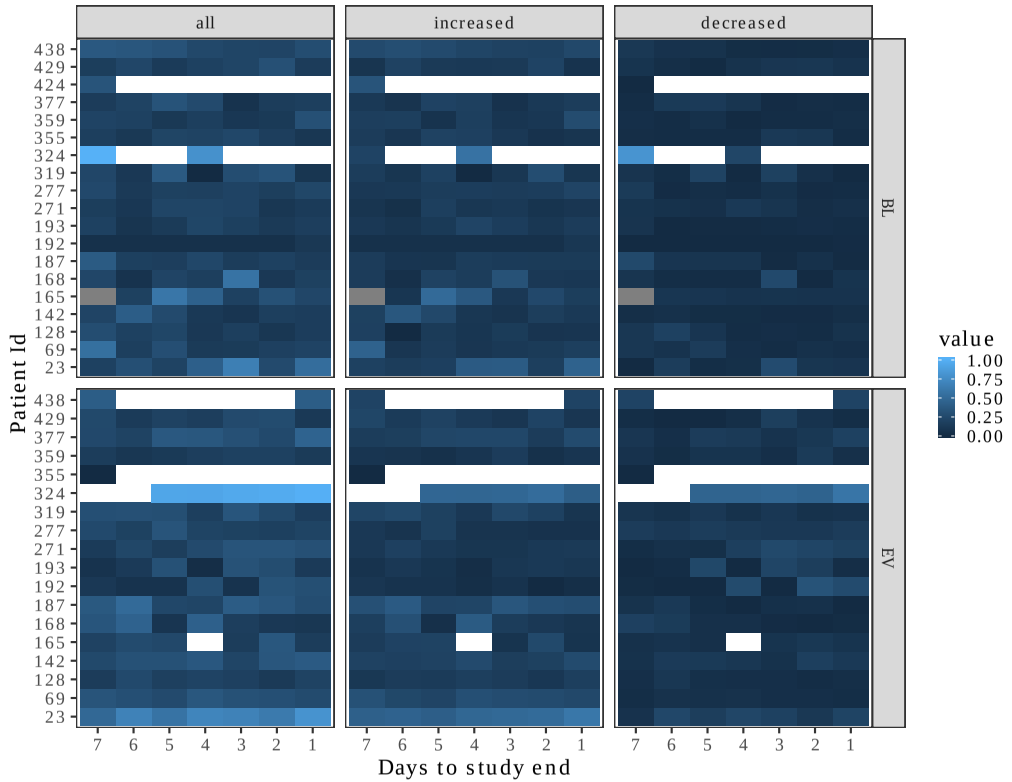


Figure A.12: Heat map for patients without emergency visit (PnoEV), showing the last week before the emergency visit (EV). A value of 1 (light colour) means that all available features have an hourly value with a distance of at least $1.5 \times IQR$ (inter quartile range) from baseline week (BL). The three columns split the change in all, increase, i.e., only $> 1.5 \times IQR$, and decrease, i.e., only $< 1.5 \times IQR$.

Figure A.13 shows the baseline values vs the final values (also EV values). Points below the line indicate a decrease from baseline to final. For all patients except 128 (active mother of 2 children, working in part-time and only 49 years old), 187 (a farmer working on his farm during the whole study period) and 23 (whose last check was negative and also one year after study participation cancer-free), most of the features decreased towards the final time. For the three aforementioned patients, most features increased. In general, many feature ratios are close to the black line, meaning that they were almost constant. Analysing the distribution of the modality categories, there is no pattern visible that a specific category is mostly increasing or decreasing. However, the scatter plot illustrates why a general prediction of emergency visits does not work: the amount of changes very individual without generally valid patterns.

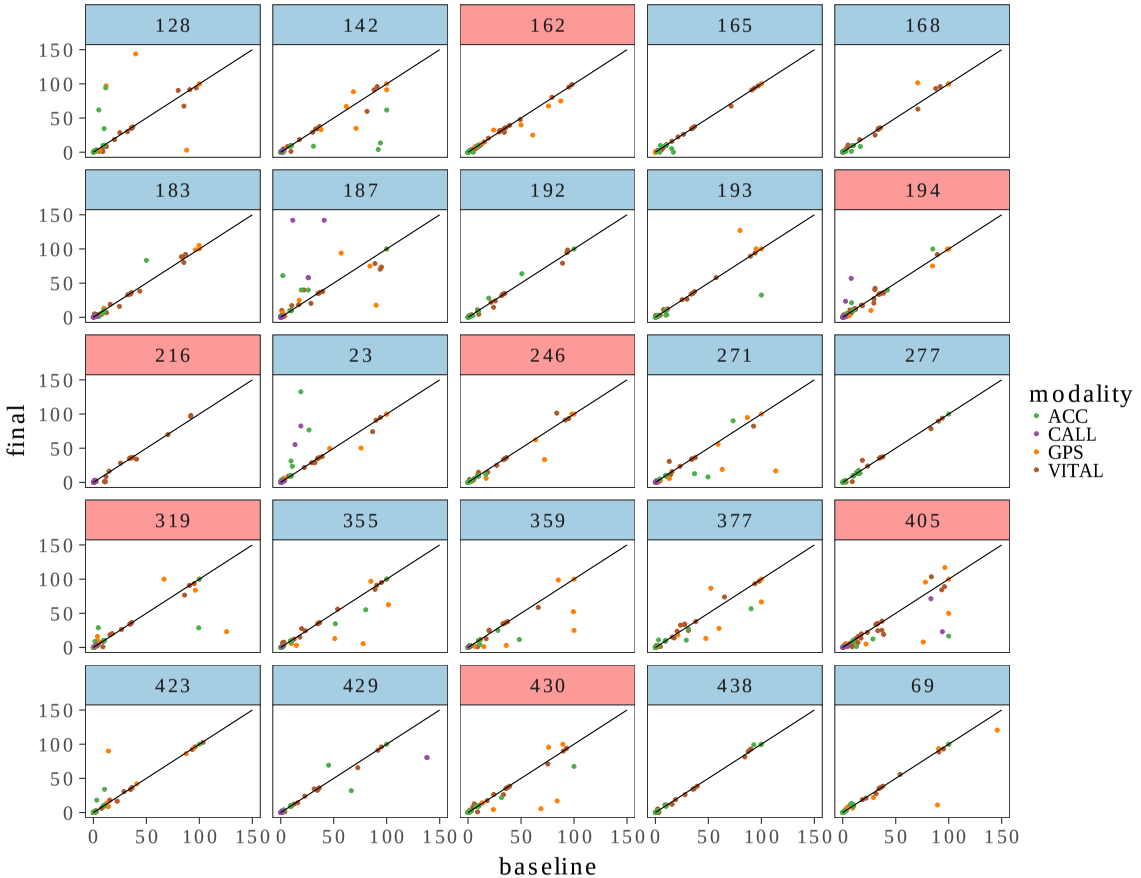


Figure A.13: Scatter plot for all patients. The black line indicates a EV/BL ratio of 1. Red text strips mark the patients with emergency visit.

A.8 Recommendations for future work

The following recommendations are based on best practices developed for the conducted palliative care study.

A.8.1 Recommendations concerning the contact with patients

Weekly contact

Most patients appreciated the personal weekly contact. Interviewers should take the time to listen to patients.

Showing personal interest

Also questionnaires with a technical focus should be formulated in a personal way to show compassion, e. g.,

impersonal How is the device handling?

personal How are you with handling the devices?

In the study, personal formulations were used which was appreciated by the patients.

A.8.2 Recommendations concerning the study design

Study design

A future study design should take into consideration a control group with a healthy population. Measurements of the control group can give estimations for normal ranges. In addition, differences in the behaviour between patient group and control group are probably more distinct than between patients or over the time in one patient.

Sample size

The power analysis has shown that larger sample sizes are needed. In order to test three hypothesis as resulting from section 5.6.1, at least 84 patients would be needed to avoid an underpowered study (ref. power analysis in section A.5).

Inclusion criteria

Depending on the features that should be examined, inclusion criteria should be more strict by adding exclusion criteria such as co-morbidities like heart diseases or sitting in a wheel chair.

Experience based sampling using digital questionnaires

In the palliative care study, the self-reports collected through the digital questionnaires on the smartphone showed little variation. Even days with emergency visits did not necessarily affect the reported pain and distress. For example, a patient might fill out the questionnaire once in the morning, still feeling good, and later in the afternoon, his situation changes leading to an ambulatory emergency visit. After already recovering from the situation during night and the next day, he already feels better so that the reported symptoms are again on the previous level. Hence, instead of reporting the symptoms once per day at a random point of time, future studies should request two reports: one in the morning and one in the evening.

Systematic log

A log file with categorised entries (medical, organisational, technical) was kept to store information about unusual events, important informations from the patient concerning the devices, their health condition, absences, etc. It is also useful to explain implausible looking data, e. g., Everion® step count is underestimated in patients with walking aids. Furthermore, emergency visits should be logged in detail to allow to consider the differing circumstances during data analysis.

In a future study, additional information about the patients should be recorded systematically. Helpful for data analysis are

- medication of patients
- hospitalisation reasons and relevant unplanned visits of physicians, including a diary over the course of events that led to the emergency visit gained through a qualitative interview
- circumstances of daily living such as family, work, etc.
- information on co-morbidities, e. g., limited vision, physical limitations like the need of walking aids

A.8.3 Organisational recommendations

Documents and information exchange between organisations

Especially hospitals follow very strict policies concerning IT infrastructure. Encrypted e-mails are a practicable communication solution preserving patients' privacy. However, for collaborative editing of documents, e. g., the aforementioned log file, a solution, e. g., shared drive, has to be established in collaboration with the IT departments of the involved organisations.

Network provider

Depending on the local provider landscape and study protocol (e. g., provided SIM cards), much time consuming effort can be reduced by using contracts instead of pre-paid cards. In addition, if patients keep their SIM cards and change the smartphone, time for adaption of SIM card size and support for data transfer to the new phone has to be planned in advance in order to guarantee on time study start.

Mobile hotspots

In case of using Wi-Fi for data transmission, mobile hotspots should be in stock. Some patients even did not know if they have Wi-Fi or not. This can be clarified during a home visit.

Stock of devices

A stock of devices for 3 patients (depending on the lead time for all material) should be always ready, since the period of announcement is varying and can be short-term (e. g., discharge next morning).

Human resources

The expenditure of time should not be underestimated. The preparation of the study including the study design, system implementation and application for ethical approval needed more than one year of one person working full time on the project. For conducting the study, three persons were in charge for screening and recruitment of patients, one person (50 %) was responsible for the technical organisation, briefing of patients and final interviews, one person (50 %) was responsible for the weekly interviews and data input in the database.

A.9 Tools for data processing

For data processsing, the following software packages were used:

- Python
  - Python 3.5.1
  - Pandas 0.21.0
  - Numpy 1.11.3
  - scikit-learn 0.19.1
  - feather [148]
- R
  - R 3.5.0 x64 [162]
  - tidyverse [163]
  - dbscan [136]
  - ggplot2 [148]
  - stargazer [164]
  - pwr [160]

Bibliography

[1] World Palliative Care Alliance, *Global atlas of palliative care at the end of life*. Worldwide Palliative Care Alliance, 2014, no. January.

[2] B. W. Stewart and C. P. Wild, *World Cancer Report 2014*. International Agency for Research on Cancer, 2014.

[3] S. Gibson and R. McConigley, "Unplanned oncology admissions within 14 days of non-surgical discharge: a retrospective study," *Support. Care Cancer*, vol. 24, no. 1, pp. 311–317, jan 2016.

[4] M. O. Delgado-Guay, Y. J. Kim, S. H. Shin, G. Chisholm, J. Williams, J. Allo, and E. Bruera, "Avoidable and unavoidable visits to the emergency department among patients with advanced cancer receiving outpatient palliative care," *J. Pain Symptom Manage.*, vol. 49, no. 3, pp. 497–504, 2015.

[5] E. Basch, A. M. Deal, M. G. Kris, H. I. Scher, C. A. Hudis, P. Sabbatini, L. Rogak, A. V. Bennett, A. C. Dueck, T. M. Atkinson, J. F. Chou, D. Dulko, L. Sit, A. Barz, P. Novotny, M. Fruscione, J. A. Sloan, and D. Schrag, "Symptom monitoring with patient-reported outcomes during routine cancer treatment: A randomized controlled trial," *J. Clin. Oncol.*, vol. 34, no. 6, pp. 557–565, 2016.

[6] E. Basch, A. M. Deal, A. C. Dueck, H. I. Scher, M. G. Kris, C. Hudis, and D. Schrag, "Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment," *Jama*, vol. 318, no. 2, p. 197, 2017.

[7] S. V. Pakhomov, S. J. Jacobsen, C. G. Chute, and V. L. Roger, "Agreement between patient-reported symptoms and their documentation in the medical record," *Am. J. Manag. Care*, vol. 14, no. 8, pp. 530–539, 2008.

[8] R. Stasi, L. Abriani, P. Beccaglia, E. Terzoli, and S. Amadori, "Cancer-Related Fatigue: Evolving Concepts in Evaluation and Treatment," *Cancer*, vol. 98, no. 9, pp. 1786–1801, 2003.

[9] J. E. Bower, "Cancer-related fatigue: Mechanisms, risk factors, and treatments," *Nat. Rev. Clin. Oncol.*, vol. 11, no. 10, pp. 597–609, oct 2014.

[10] G. A. Curt, "The Impact of Fatigue on Patients with Cancer: Overview of FATIGUE 1 and 2," *Oncologist*, vol. 5, no. 90002, pp. 9–12, 2000.

[11] A. M. Berger, L. H. Gerber, and D. K. Mayer, "Cancer-related fatigue: Implications for breast cancer survivors," *Cancer*, vol. 118, no. SUPPL.8, pp. 2261–2269, 2012.

[12] M. L. van der Lee and B. Garssen, "Mindfulness-based cognitive therapy reduces chronic cancer-related fatigue: a treatment study," *Psychooncology*, vol. 21, no. 3, pp. 264–272, 2012.

[13] J. Sadja and P. J. Mills, "Effects of Yoga Interventions on Fatigue in Cancer Patients and Survivors: A Systematic Review of Randomized Controlled Trials," *Explore (NY).*, vol. 9, no. 4, pp. 232–243, 2013.

[14] O. Minton, A. Berger, A. Barsevick, F. Cramp, M. Goedendorp, S. A. Mitchell, and P. C. Stone, "Cancer-related fatigue and its impact on functioning," *Cancer*, vol. 119, no. SUPPL11, pp. 2124–2130, 2013.

[15] G. Prue, J. Rankin, J. Allen, J. Gracey, and F. Cramp, "Cancer-related fatigue: A critical appraisal," *Eur. J. Cancer*, vol. 42, no. 7, pp. 846–863, 2006.

[16] G. A. Kelley and K. S. Kelley, "Exercise and cancer-related fatigue in adults: A systematic review of previous systematic reviews with meta-analyses," *BMC Cancer*, vol. 17, no. 1, pp. 1–17, 2017.

[17] V. G. Motti and K. Caine, "Human Factors Considerations in the Design of Wearable Devices," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 58, no. 1, pp. 1820–1824, sep 2014.

[18] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, 2007.

[19] A. Shcherbina, C. Mikael Mattsson, D. Waggott, H. Salisbury, J. W. Christle, T. Hastie, M. T. Wheeler, and E. A. Ashley, "Accuracy in wrist-worn, sensor-based measurements of heart rate

and energy expenditure in a diverse cohort," *J. Pers. Med.*, vol. 7, no. 2, 2017.

[20] C. Bauer, "On the (In-)Accuracy of GPS Measures of Smartphones," *Proc. Int. Conf. Adv. Mob. Comput. Multimed. - MoMM '13*, no. December 2013, pp. 335–341, 2013.

[21] S. Korpilo, T. Virtanen, and S. Lehvävirta, "Smartphone GPS tracking—Inexpensive and efficient data collection on recreational movement," *Landsc. Urban Plan.*, vol. 157, pp. 608–617, 2017.

[22] P. Mohebbi, E. Stroulia, and I. Nikolaidis, "Indoor Localization: A Cost-Effectiveness vs. Accuracy Study," *Proc. - IEEE Symp. Comput. Med. Syst.*, vol. 2017-June, pp. 552–557, 2017.

[23] W. T. Abraham, P. B. Adamson, R. C. Bourge, M. F. Aaron, M. R. Costanzo, L. W. Stevenson, W. Strickland, S. Neelagaru, N. Raval, S. Krueger, S. Weiner, D. Shavelle, B. Jeffries, and J. S. Yadav, "Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: A randomised controlled trial," *Lancet*, vol. 377, no. 9766, pp. 658–666, 2011.

[24] G. Hindricks, N. Varma, S. Kacet, T. Lewalter, P. Søgaard, L. Guédon-Moreau, J. Proff, T. A. Gerds, S. D. Anker, and C. Torp-Pedersen, "Daily remote monitoring of implantable cardioverter-defibrillators: Insights from the pooled patient-level data from three randomized controlled trials (IN-TIME, ECOST, TRUST)," *Eur. Heart J.*, vol. 38, no. 22, pp. 1749–1755, 2017.

[25] M. Landolina, G. B. Perego, M. Lunati, A. Curnis, G. Guenzati, A. Vicentini, G. Parati, G. Borghi, P. Zanaboni, S. Valsecchi, and M. Marzegalli, "Remote monitoring reduces healthcare use and improves quality of care in heart failure patients with implantable defibrillators: The evolution of management strategies of heart failure patients with implantable defibrillators (EVOLVO) study," *Circulation*, vol. 125, no. 24, pp. 2985–2992, 2012.

[26] G. Hindricks, M. Taborsky, M. Glikson, U. Heinrich, B. Schumacher, A. Katz, J. Brachmann, T. Lewalter, A. Goette, M. Block, J. Kautzner, S. Sack, D. Husser, C. Piorkowski, and P. Søgaard,

"Implant-based multiparameter telemonitoring of patients with heart failure (IN-TIME): A randomised controlled trial," *Lancet*, vol. 384, no. 9943, pp. 583–590, 2014.

[27] G. F. Gensini, C. Alderighi, R. Rasoini, M. Mazzanti, and G. Casolo, "Value of Telemonitoring and Telemedicine in Heart Failure Management," *Card. Fail. Rev.*, vol. 3, no. 2, pp. 116–121, 2017.

[28] A. Bhimaraj, "Remote monitoring of heart failure patients." *Methodist Debakey Cardiovasc J*, vol. 9, no. 1, pp. 26–31, 2013.

[29] N. Bashi, M. Karunanithi, F. Fatehi, H. Ding, and D. Walters, "Remote Monitoring of Patients With Heart Failure: An Overview of Systematic Reviews." *J. Med. Internet Res.*, vol. 19, no. 1, p. e18, jan 2017.

[30] G. Giunti, E. Guisado-Fernandez, and B. Caulfield, "Connected Health in Multiple Sclerosis: A Mobile Applications Review," *2017 IEEE 30th Int. Symp. Comput. Med. Syst.*, no. June, pp. 660–665, 2017.

[31] P. A. Arean, K. A. Hallgren, J. T. Jordan, A. Gazzaley, D. C. Atkins, P. J. Heagerty, and J. A. Anguera, "The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial." *J. Med. Internet Res.*, vol. 18, no. 12, p. e330, dec 2016.

[32] J. Firth, J. Torous, J. Nicholas, R. Carney, A. Pratap, S. Rosenbaum, and J. Sarris, "The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials," *World Psychiatry*, vol. 16, no. 3, pp. 287–298, 2017.

[33] M. Allsop and S. Taylor, "Information and communication technology for managing pain in palliative care: a review of the literature," *... Palliat. Care*, pp. 1–9, 2014.

[34] N. Andebe, P. Waiganjo, and J. Weru, "mHealth in palliative care for cancer patients & care givers," in *2017 IST-Africa Week Conf.* IEEE, may 2017, pp. 1–10.

[35] E. Rincon, F. Monteiro-Guerra, O. Rivera-Romero, E. Dorronzoro-Zubiete, C. L. Sanchez-Bocanegra, and E. Gabarron, "Mobile Phone Apps for Quality of Life and Well-Being

Assessment in Breast and Prostate Cancer Patients: Systematic Review." *JMIR mHealth uHealth*, vol. 5, no. 12, p. e187, dec 2017.

[36] F. Kawsar, M. M. Haque, R. Love, D. Roe, R. Dowla, T. Ferdousy, R. Selim, and S. Hossain, "e-ESAS : Improving Quality of Life for Breast Cancer Patients in Developing Countries," *Proc. 2nd ACM Int. Work. Pervasive Wirel. Healthc. - MobileHealth '12*, vol. 2, pp. 9–14, 2012.

[37] A. R. Kim and H.-a. A. Park, "Web-based Self-management Support Interventions for Cancer Survivors: A Systematic Review and Meta-analyses," *Stud Heal. Technol Inf.*, vol. 216, no. January 2000, pp. 142–147, 2015.

[38] A. Seiler, V. Klaas, G. Tröster, and C. P. Fagundes, "eHealth and mHealth interventions in the treatment of fatigued cancer survivors: A systematic review and meta-analysis," *Psychooncology.*, vol. 26, no. 9, pp. 1239–1253, sep 2017.

[39] C. Brunelli, S. Kaasa, A. K. Knudsen, M. J. Hjermstad, A. Pigni, and A. Caraceni, "Comparisons of Patient and Physician Assessment of Pain-Related Domains in Cancer Pain Classification: Results From a Large International Multicenter Study," *J. Pain*, vol. 15, no. 1, pp. 59–67, 2014.

[40] S. Lundström, "Is it reliable to use cellular phones for symptom assessment in palliative care? Report on a study in patients with advanced cancer." *J. Palliat. Med.*, vol. 12, no. 12, p. 1087, 2009.

[41] A. V. Bennett, R. E. Jensen, and E. Basch, "Electronic Patient-Reported Outcome Systems in Oncology Clinical Practice," *CA. Cancer J. Clin.*, vol. 62, pp. 336–347, 2012.

[42] S. Benedetto, C. Caldato, E. Bazzan, D. C. Greenwood, V. Pensabene, and P. Actis, "Assessment of the fitbit charge 2 for monitoring heart rate," *PLoS One*, vol. 13, no. 2, pp. 1–10, 2018.

[43] J. A. Bunn, J. W. Navalta, C. J. Fountaine, and J. D. Reece, "Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015-2017." *Int. J. Exerc. Sci.*, vol. 11, no. 7, pp. 503–515, 2018.

[44] S. Gillinov, M. Etiwy, R. Wang, G. Blackburn, D. Phelan, A. M. Gillinov, P. Houghtaling, H. Javadikasgari, and M. Y. Desai, "Variable accuracy of wearable heart rate monitors during aerobic exercise," *Med. Sci. Sports Exerc.*, vol. 49, no. 8, pp. 1697–1703, 2017.

[45] O. D. Incel, M. Kose, and C. Ersoy, "A Review and Taxonomy of Activity Recognition on Mobile Phones," *Bionanoscience*, vol. 3, no. 2, pp. 145–171, jun 2013.

[46] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 1, pp. 1–12, 2010.

[47] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *J. Neuroeng. Rehabil.*, vol. 9, no. 1, p. 21, 2012.

[48] G. M. Harari, N. D. Lane, R. Wang, B. S. Crosier, A. T. Campbell, and S. D. Gosling, "Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges." *Perspect. Psychol. Sci.*, vol. 11, no. 6, pp. 838–854, nov 2016.

[49] B. Logan, J. Healey, M. Philipose, E. M. Tapia, and S. Intille, "A Long-Term Evaluation of Sensing Modalities for Activity Recognition," *UbiComp 2007 Ubiquitous Comput.*, vol. 4717, pp. 483–500, 2007.

[50] J. Seiter, S. Feese, B. Arnrich, G. Tröster, O. Amft, L. Macrea, and K. Maurer, "Activity monitoring in daily life as an outcome measure for surgical pain relief intervention using smartphones," in *Proc. 17th Annu. Int. Symp. Int. Symp. wearable Comput. - ISWC '13*. New York, New York, USA: ACM Press, sep 2013, p. 127.

[51] C. Burton, B. McKinstry, A. Szentagotai Tatar, A. Serrano-Blanco, C. Pagliari, and M. Wolters, "Activity monitoring in patients with depression: A systematic review," *J. Affect. Disord.*, vol. 145, no. 1, pp. 21–28, 2013.

[52] M. Rabbi, M. H. Aung, M. Zhang, and T. Choudhury, "MyBehavior: automatic personalized health feedback from user be-

haviors and preferences using smartphones," *Ubicomp '15, Sept. 7–11, 2015, Osaka, Japan*, pp. 707–718, 2015.

[53] M. Lin, N. D. Lane, M. Mohammod, X. Yang, H. Lu, G. Cardone, S. Ali, A. Doryab, E. Berke, A. T. Campbell, and T. Choudhury, "BeWell+: Multi-dimensional Wellbeing Monitoring with Community-guided User Feedback and Energy Optimization," *Proc. Conf. Wirel. Heal. - WH '12*, pp. 1–8, 2012.

[54] N. Lane, M. Mohammod, M. Lin, X. Yang, H. Lu, S. Ali, A. Doryab, E. Berke, T. Choudhury, and A. Campbell, "BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing," in *Proc. 5th Int. ICST Conf. Pervasive Comput. Technol. Healthc.* IEEE, 2011, p. 8.

[55] D. Ben-Zeev, R. Wang, S. Abdullah, R. Brian, E. A. Scherer, L. A. Mistler, M. Hauser, J. M. Kane, A. Campbell, and T. Choudhury, "Mobile Behavioral Sensing for Outpatients and Inpatients With Schizophrenia," *Psychiatr. Serv.*, vol. 67, no. 5, pp. 558–561, 2016.

[56] R. Wang, M. S. H. Aung, S. Abdullah, R. Brian, A. T. Campbell, T. Choudhury, M. Hauser, J. Kane, M. Merrill, E. A. Scherer, V. W. S. Tseng, and D. Ben-zeev, "CrossCheck: Towards passive sensing and detection of mental health changes in people with schizophrenia," *Proc. Int. Conf. Ubiquitous Comput.*, pp. 1–12, 2016.

[57] V. Tseng, S. Abdullah, M. Merrill, M. Aung, F. Wittleder, and T. Choudhury, "Assessing mental health issues on college campuses: Preliminary findings from a pilot study," *UbiComp 2016 Adjun. - Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, 2016.

[58] S. Saeb, M. Zhang, C. J. Karr, S. M. Schueller, M. E. Corden, K. P. Kording, and D. C. Mohr, "Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study," *J. Med. Internet Res.*, vol. 17, no. 7, p. e175, jul 2015.

[59] A. Grunerbl, A. Muaremi, V. Osmani, G. Bahle, S. Ohler, G. Troester, O. Mayora, C. Haring, and P. Lukowicz, "Smartphone Based Recognition of States and State Changes in Bipolar

Disorder Patients." *IEEE J. Biomed. Heal. informatics*, vol. 19, no. 1, pp. 140–148, 2014.

[60] M. Faurholt-Jepsen, M. Vinberg, E. M. Christensen, M. Frost, J. Bardram, and L. V. Kessing, "Daily electronic self-monitoring of subjective and objective symptoms in bipolar disorder–the MONARCA trial protocol (MONitoring, treAtment and pRediCtion of bipolAr disorder episodes): a randomised controlled single-blind trial." *BMJ Open*, vol. 3, no. 7, pp. 1–10, 2013.

[61] O. Mayora, B. Arnrich, J. Bardram, C. Drager, A. Finke, M. Frost, S. Giordano, A. Grunerbl, C. Raring, R. Raux, P. Lukowicz, A. Muaremi, S. Ohler, A. Puiatti, N. Reichwaldt, C. Scharnweber, G. Troester, L. Vedel, and G. Wurzer, "Mobile Health Systems for Bipolar Disorder," *Proc. MindCare*, vol. 5, no. March, pp. 424–429, 2013.

[62] A. Muaremi, B. Arnrich, and G. Tröster, "Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep," *Bionanoscience*, vol. 3, no. 2, pp. 172–183, 2013.

[63] L. G. Morrison, C. Hargood, V. Pejovic, A. W. A. Geraghty, S. Lloyd, N. Goodman, D. T. Michaelides, A. Weston, M. Musolesi, M. J. Weal, and L. Yardley, "The Effect of Timing and Frequency of Push Notifications on Usage of a Smartphone-Based Stress Management Intervention: An Exploratory Trial," *PLoS One*, vol. 12, no. 1, 2017.

[64] D. Lopez Martinez and R. Picard, "Wearable Technologies for Multiple Sclerosis The future role of wearable stress measurement in improving quality of life," *Second Int. Conf. Smart Portable, Wearable, Implant. Disabil. Devices Syst.*, 2016.

[65] Y. Moon, R. S. McGinnis, K. Seagers, R. W. Motl, N. Sheth, J. A. Wright, R. Ghaffari, and J. J. Sosnoff, "Monitoring gait in multiple sclerosis with novel wearable motion sensors," *PLoS One*, vol. 12, no. 2, pp. 1–19, 2017.

[66] S. Mazilu, U. Blanke, A. Calatroni, E. Gazit, J. M. Hausdorff, and G. Tröster, "The role of wrist-mounted inertial sensors in detecting gait freeze episodes in Parkinson's disease," *Pervasive Mob. Comput.*, pp. –, 2016.

[67] Z. A, M. S, T. C, and E. al, "Using smartphones and machine learning to quantify parkinson disease severity: The mobile parkinson disease score," *JAMA Neurol.*, 2018.

[68] J. Timmerman, M. D.-v. Weering, T. Tönis, H. Hermens, and M. Vollenbroek-Hutten, "Relationship between patterns of daily physical activity and fatigue in cancer survivors," *Eur. J. Oncol. Nurs.*, vol. 19, no. 2, pp. 162–168, apr 2015.

[69] E. H. HON and S. T. LEE, "ELECTRONIC EVALUATION OF THE FETAL HEART RATE. VIII. PATTERNS PRECEDING FETAL DEATH, FURTHER OBSERVATIONS." *Am. J. Obstet. Gynecol.*, vol. 87, pp. 814–26, nov 1963.

[70] M. Malik, J. Bigger, A. Camm, and R. Kleiger, "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology." *Eur. Heart J.*, vol. 17, no. 3, pp. 354–381, mar 1996.

[71] J. M. Karemaker, "Autonomic integration: the physiological basis of cardiovascular variability," *J. Physiol.*, vol. 517, no. 2, pp. 316–316, jun 1999.

[72] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, J. S. Suri, U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: A review," *Med. Biol. Eng. Comput.*, vol. 44, no. 12, pp. 1031–1051, dec 2006.

[73] P. A. Lotufo, L. Valiengo, I. M. Benseñor, and A. R. Brunoni, "A systematic review and meta-analysis of heart rate variability in epilepsy and antiepileptic drugs," *Epilepsia*, vol. 53, no. 2, pp. 272–282, feb 2012.

[74] J. F. Thayer, S. S. Yamamoto, and J. F. Brosschot, "The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors," *Int. J. Cardiol.*, vol. 141, no. 2, pp. 122–131, may 2010.

[75] G. E. Billman, "The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance," *Front. Physiol.*, vol. 4, p. 26, feb 2013.

[76] W. von Rosenberg, T. Chanwimalueang, T. Adjei, U. Jaffer, V. Goverdovsky, and D. P. Mandic, "Resolving Ambiguities in the LF/HF Ratio: LF-HF Scatter Plots for the Categorization of Mental and Physical Stress from HRV," *Front. Physiol.*, vol. 8, jun 2017.

[77] Y. Wang, H. Wu, E. Huang, C. Wang, S. Hseu, and Y. Kou, "A Novel Prognostic Factor in Patients With Brain Metastasis: Heart Rate Variability," *Int. J. Radiat. Oncol.*, vol. 93, no. 3, pp. E472–E473, nov 2015.

[78] C. Arab, D. P. M. Dias, R. T. d. A. Barbosa, T. D. de Carvalho, V. E. Valenti, T. B. Crocetta, M. Ferreira, L. C. de Abreu, and C. Ferreira, "Heart rate variability measure in breast cancer patients and survivors: A systematic review," *Psychoneuroendocrinology*, vol. 68, pp. 57–68, 2016.

[79] E. K. Masel, P. Huber, T. Engler, and H. H. Watzke, "Heart rate variability during treatment of breakthrough pain in patients with advanced cancer: A pilot study," *J. Pain Res.*, vol. 9, pp. 1215–1220, dec 2016.

[80] S. A. Alsirafy, "Heart rate variability measurement: is there a role in predicting the admission outcome of palliative cancer patients?" *Ann. Palliat. Med.*, vol. 4, no. 1, 2015.

[81] K. K. Kim, J. S. Kim, Y. G. Lim, and K. S. Park, "The effect of missing RR-interval data on heart rate variability analysis in the frequency domain," *Physiol. Meas.*, vol. 30, no. 10, pp. 1039–1050, oct 2009.

[82] K. K. Kim, Y. G. Lim, J. S. Kim, and K. S. Park, "Effect of missing RR-interval data on heart rate variability analysis in the time domain," *Physiol. Meas.*, vol. 28, no. 12, pp. 1485–1494, dec 2007.

[83] D. Nabil and F. Bereksi Reguig, "Ectopic beats detection and correction methods: A review," *Biomed. Signal Process. Control*, vol. 18, pp. 228–244, 2015.

[84] M. Senožetnik, L. Bradeško, B. Kažič, D. Mladeni, and T. Šubic, "Spatio-temporal clustering methods," *http://optimumproject.eu/news/44/67/Spatio-temporal-Clustering-Methods.html*, pp. 1–4, 2016.

[85] H. F. Tork, "Spatio-Temporal Clustering Methods Classification," *Dr. Symp. Informatics Eng.*, no. December, pp. 1–12, 2012.

[86] Y. Yuan and M. Raubal, "A framework for spatio-temporal clustering from mobile phone data," in *Work. Complex Data Min. a GeoSpatial Context Proc. Agil. 2012*, 2012, pp. 22–26.

[87] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-based Algorithm for Discovering Clusters a Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proc. Second Int. Conf. Knowl. Discov. Data Min.*, ser. KDD'96. AAAI Press, 1996, pp. 226–231.

[88] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "Optics: Ordering points to identify the clustering structure," *ACM Sigmod Rec.*, pp. 49–60, 1999.

[89] D. Birant and A. Kut, "ST-DBSCAN: An algorithm for clustering spatial-temporal data," *Data Knowl. Eng.*, vol. 60, no. 1, pp. 208–221, 2007.

[90] L. O. Alvares, V. Bogorny, B. Kuijpers, J. A. F. de Macedo, B. Moelans, and A. Vaisman, "A model for enriching trajectories with semantic geographical information," *Proc. 15th Annu. ACM Int. Symp. Adv. Geogr. Inf. Syst. - GIS '07*, no. i, p. 1, 2007.

[91] A. Palma, V. Bogorny, and B. Kuijpers, "A clustering-based approach for discovering interesting places in trajectories," *SAC '08*, pp. 863–868, 2008.

[92] J. A. M. Rocha, V. C. Times, G. Oliveira, L. O. Alvares, and V. Bogorny, "DB-SMoT: A direction-based spatio-temporal clustering method," *2010 IEEE Int. Conf. Intell. Syst. IS 2010 - Proc.*, pp. 114–119, 2010.

[93] R. Melamed, "Indoor Localization: Challenges and Opportunities," *IEEE/ACM Int. Conf. Mob. Softw. Eng. Syst.*, pp. 1–2, 2016.

[94] S. Xia, Y. Liu, G. Yuan, M. Zhu, and Z. Wang, "Indoor Fingerprint Positioning Based on Wi-Fi: An Overview," *ISPRS Int. J. Geo-Information*, vol. 6, no. 5, p. 135, 2017.

[95] J. Luo and X. Zhan, "Characterization of smart phone received signal strength indication for WLAN indoor positioning accuracy improvement," *J. Networks*, vol. 9, no. 3, pp. 739–746, 2014.

[96] J. H. Kang, W. Welbourne, B. Stewart, and G. Borriello, "Extracting places from traces of locations," *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 9, no. 3, p. 58, 2005.

[97] S. L. Lau, Y. Xu, and K. David, "Novel Indoor Localisation using an Unsupervised Wi-Fi Signal Clustering Method," *Futur. Netw. &MobileSummit 2011 Conf. Proc.*, pp. 1–8, 2011.

[98] Y. Xu, S. L. Lau, R. Kusber, and K. David, "DCCLA: Automatic Indoor Localization Using Unsupervised Wi-Fi Fingerprinting," in *Model. Using Context*, 2013, pp. 73–86.

[99] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A Database of German Emotional Speech," *Interspeech*, pp. 1517—1520, 2005.

[100] Z. Xie and L. Guan, "Multimodal Information Fusion of Audiovisual Emotion Recognition Using Novel Information Theoretic Tools," *Int. J. Multimed. Data Eng. Manag.*, vol. 4, no. 4, pp. 1–14, oct 2013.

[101] Z. XIE and L. GUAN, "Multimodal Information Fusion of Audio Emotion Recognition Based on Kernel Entropy Component Analysis," *Int. J. Semant. Comput.*, vol. 07, no. 01, pp. 25–42, 2013.

[102] S. Steininger, S. Rabold, O. Dioubina, and F. Schiel, "Development of the User-State Conventions for the Multimodal Corpus in SmartKom," *Proc. 3 Rd Int. Conf. Lang. Resour. Eval. Work. Multimodal Resour. Multimodal Syst. Eval. Las*, no. 27, pp. 33–37, 2002.

[103] D. Ververidis, D. Ververidis, C. Kotropoulos, and C. Kotropoulos, "A Review of Emotional Speech Databases," *Proc. 9th Panhellenic Conf. Informatics*, pp. 560–574, 2003.

[104] B. Schuller, B. Vlasenko, F. Eyben, G. Rigoll, and A. Wendemuth, "Acoustic emotion recognition: A benchmark comparison of performances," *Proc. 2009 IEEE Work. Autom. Speech Recognit. Understanding, ASRU 2009*, pp. 552–557, 2009.

[105] T. Vogt, E. André, and N. Bee, "EmoVoice - A framework for online recognition of emotions from voice," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5078 LNCS, pp. 188–199, 2008.

[106] J. Wagner, F. Lingenfelser, and E. André, "The social signal interpretation framework (SSI) for real time signal processing and recognition," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, pp. 3245–3248, 2011.

[107] T. Giannakopoulos, "pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis," *PLoS One*, vol. 10, no. 12, p. e0144610, dec 2015.

[108] T. Vogt and E. André, "COMPARING FEATURE SETS FOR ACTED AND SPONTANEOUS SPEECH IN VIEW OF Augsburg University , Germany Multimedia concepts and applications Applied Computer Science," *Multimed. Expo, 2005. ICME 2005. IEEE Int. Conf.*, pp. 474–477, 2005.

[109] ——, "Improving Automatic Emotion Recognition from Speech via Gender Differentiation," *Proc. Lang. Resour. Eval. Conf.*, no. January, pp. 1123–1126, 2006.

[110] I. Guyon, I. Guyon, A. Elisseeff, and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.

[111] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical Human Activity Recognition Using Wearable Sensors," *Sensors*, vol. 15, no. 12, pp. 31 314–31 338, dec 2015.

[112] R. J. Klement, M. Allgäuer, S. Appold, K. Dieckmann, I. Ernst, U. Ganswindt, R. Holy, U. Nestle, M. Nevinny-Stickel, S. Semrau, F. Sterzing, A. Wittig, N. Andratschke, and M. Guckenberger, "Support Vector Machine-Based Prediction of Local Tumor Control After Stereotactic Body Radiation Therapy for Early-Stage Non-Small Cell Lung Cancer," *Int. J. Radiat. Oncol.*, vol. 88, no. 3, pp. 732–738, mar 2014.

[113] G. A. Susto, M. Terzi, and A. Beghi, "Anomaly Detection Approaches for Semiconductor Manufacturing," *Procedia Manuf.*, vol. 11, no. June 2017, pp. 2018–2024, 2017.

[114] R. J. Bolton, D. J. Hand, F. Provost, L. Breiman, R. J. Bolton, and D. J. Hand, "Statistical Fraud Detection: A Review," *Stat. Sci.*, vol. 17, no. 3, pp. 235–255, 2002.

[115] Ghosh and Reilly, "Credit card fraud detection with a neural-network," *1994 Proc. Twenty-Seventh Hawaii Int. Conf. Syst. Sci.*, vol. 3, pp. 621–630, 1994.

[116] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation forest," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 413–422, 2008.

[117] O. Mazhelis, "One-class classifiers: a review and analysis of suitability in the context of mobile-masquerader detection," *South African Comput. J.*, vol. 36, pp. 29–48, 2006.

[118] E. Gurarie, R. D. Andrews, and K. L. Laidre, "A novel method for identifying behavioural changes in animal movement data," *Ecol. Lett.*, vol. 12, no. 5, pp. 395–408, 2009.

[119] R. J. Gutierrez, B. C. Boehmke, K. W. Bauer, C. M. Saie, and T. J. Bihl, "anomalyDetection: Implementation of Augmented Network Log Anomaly Detection Procedures," *R J.*, vol. 9, no. 2, pp. 354–365, 2017. [Online]. Available: https://journal.r-project.org/archive/2017/RJ-2017-039/index.html

[120] T. De Waal, J. Pannekoek, and S. Scholtus, "The editing of statistical data: Methods and techniques for the efficient detection and correction of errors and missing values," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 4, no. 2, pp. 204–210, 2012.

[121] N. A. James, A. Kejariwal, and D. S. Matteson, "Leveraging cloud data to mitigate user experience from 'breaking bad'," in *Proc. - 2016 IEEE Int. Conf. Big Data, Big Data 2016*, nov 2016, pp. 3499–3508.

[122] D. S. Matteson and N. A. James, "A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data," *J. Am. Stat. Assoc.*, vol. 109, no. 505, pp. 334–345, jan 2014.

[123] A. Cnaan, N. A. N. M. Laird, and P. Slasor, "Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data," *Stat.Med.*, vol. 16, no. 20, pp. 2349–2380, 1997.

[124] V. V. C. Klaas, A. Calatroni, M. Hardegger, M. Guckenberger, G. Theile, and G. Tröster, "Monitoring Patients in Ambulatory Palliative Care: A Design for an Observational Study," *Wirel.*

Mob. Commun. Healthc. 6th Int. Conf. MobiHealth 2016, Milan, Italy, Novemb. 14-16, 2016, Proc., vol. 192, pp. 207–214, 2017.

[125] V. C. V. Klaas, G. Troster, N. Buel, H. Walt, and J. Jenewein, "Smart-phone based monitoring of cancer related fatigue," in *2017 IEEE 13th Int. Conf. Wirel. Mob. Comput. Netw. Commun.*, vol. 2017-Octob. Rome, Italy: IEEE, oct 2017, pp. 249–256.

[126] S. J. Fusco, K. Michael, M. Michael, and R. Abbas, "Using a social informatics framework to study the effects of location-based social networking on relationships between people: A review of literature," in *2010 Ninth Int. Conf. Mob. Bus. 2010 Ninth Glob. Mobil. Roundtable*. IEEE, 2010, pp. 230–237.

[127] D. Fiorillo and F. Sabatini, "Quality and quantity: The role of social interactions in self-reported individual health," *Soc. Sci. Med.*, vol. 73, no. 11, pp. 1644–1652, 2011.

[128] M. Ture, J. Barth, F. Angst, A. Aeschlimann, U. Schnyder, N. Zerkiebel, J. Perseus, C. Renner, P. Imesch, B. Fuchs, G. F. Huber, H. Walt, C. Martin-Soelch, J. Jenewein, and i. Press), "Use of inpatient rehabilitation for cancer patients in Switzerland : Who undergoes cancer rehabilitation ?" *Swiss Med. Wkly.*, no. December, pp. 1–11, 2015.

[129] M. Csikszentmihalyi and R. Larson, *Validity and Reliability of the Experience-Sampling Method*. Dordrecht: Springer Netherlands, 2014, pp. 35–54.

[130] M. Bossola, E. Di Stasio, S. Giungi, F. Rosa, and L. Tazza, "Fatigue is associated with serum interleukin-6 levels and symptoms of depression in patients on chronic hemodialysis," *J. Pain Symptom Manage.*, vol. 49, no. 3, pp. 578–585, 2015.

[131] B. DiCicco-Bloom and B. F. Crabtree, "The qualitative research interview," *Med. Educ.*, vol. 40, no. 4, pp. 314–321, 2006.

[132] G. Theile, V. Klaas, G. Tröster, and M. Guckenberger, "mHealth Technologies for Palliative Care Patients at the Interface of In-Patient to Outpatient Care: Protocol of Feasibility Study Aiming to Early Predict Deterioration of Patient's Health Status," *JMIR Res. Protoc.*, vol. 6, no. 8, p. e142, aug 2017.

[133] L. CHOI, Z. LIU, C. E. MATTHEWS, and M. S. BUCHOWSKI, "Validation of Accelerometer Wear and Nonwear Time Classification Algorithm," *Med. Sci. Sport. Exerc.*, vol. 43, no. 2, pp. 357–364, feb 2011.

[134] C. J. de Ruiter, P. W. L. Verdijk, W. Werker, M. J. Zuidema, and A. de Haan, "Stride frequency in relation to oxygen consumption in experienced and novice runners," *Eur. J. Sport Sci.*, vol. 14, no. 3, pp. 251–258, 2014.

[135] M. Duarte, "Detection of peaks in data." [Online]. Available: http://nbviewer.jupyter.org/github/demotu/BMC/blob/master/notebooks/DetectPeaks.ipynb

[136] M. Hahsler and M. Piekenbrock, *dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms*, 2017. [Online]. Available: https://cran.r-project.org/package=dbscan

[137] N. Chopde and M. Nichat, "Landmark Based Shortest Path Detection by Using A* and Haversine Formula," *GH Raisoni Coll. Eng. . . .*, vol. 1, no. 2, pp. 298–302, 2013.

[138] D. Steinley, "Properties of the Hubert-Arable Adjusted Rand Index." Steinley, Douglas: Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel Street, Champaign, IL, US, 61820, steinley@cyrus.psych.uiuc.edu, pp. 386–396, 2004.

[139] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz, "Guidelines Heart rate variability," *Eur. Heart J.*, vol. 17, pp. 354–381, 1996.

[140] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Front. Public Heal.*, vol. 5, p. 258, 2017.

[141] M. Tsipouras, D. Fotiadis, and D. Sideris, "An arrhythmia classification system based on the RR-interval signal," *Artif. Intell. Med.*, vol. 33, no. 3, pp. 237–250, mar 2005.

[142] F. Shaffer, R. McCraty, and C. L. Zerr, "A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability," *Front. Psychol.*, vol. 5, no. September, p. 1040, 2014.

[143] A. Field, J. Miles, and Z. Field, *Discovering Statistics Using R*. SAGE Publications, 2012. [Online]. Available: https://books.google.ch/books?id=wd2K2zC3swIC

[144] N. K. Aaronson, S. Ahmedzai, B. Bergman, M. Bullinger, A. Cull, N. J. Duez, A. Filiberti, H. Flechtner, S. B. Fleishman, J. C. J. M. de Haes, S. Kaasa, M. C. Klee, D. Osoba, D. Razavi, P. B. Rofe, S. Schraub, K. C. a. Sneeuw, M. Sullivan, and F. Takeda, "The European Organisation for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology," *J. Natl. Cancer Inst.*, vol. 85, pp. 365–376, 1993.

[145] B. Hayes, "Gauss's day of reckoning," *Am. Sci.*, vol. 94, no. 3, 2006.

[146] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in {P}ython," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[147] A. Inselberg, "Multidimensional detective," *Proc. VIZ '97 Vis. Conf. Inf. Vis. Symp. Parallel Render. Symp.*, no. November 1997, pp. 100–107, 1997.

[148] H. Wickham, "ggplot2 - Elegant Graphics for Data Analysis, Second Edition," *Media*, vol. 35, no. July, p. 211, 2016. [Online]. Available: http://had.co.nz/ggplot2/book

[149] W. M. Grove and N. C. Andreasen, "Simultaneous tests of many hypotheses in exploratory research." *J. Nerv. Ment. Dis.*, vol. 170, no. 1, pp. 3–8, jan 1982.

[150] J. Hart, "Normal resting pulse rate ranges," *J. Nurs. Educ. Pract.*, vol. 5, no. 8, pp. 95–98, 2015.

[151] D. Nunan, G. R. H. Sandercock, and D. A. Brodie, "A quantitative systematic review of normal values for short-term heart rate variability in healthy adults," *PACE - Pacing Clin. Electrophysiol.*, vol. 33, no. 11, pp. 1407–1417, nov 2010.

[152] M. De Couck and Y. Gidron, "Norms of vagal nerve activity, indexed by Heart Rate Variability, in cancer patients," *Cancer Epidemiol.*, vol. 37, no. 5, pp. 737–741, 2013.

[153] J. M. Dekker, E. G. Schouten, P. Klootwijk, J. Pool, C. A. Swenne, and D. Kromhout, "Heart Rate Variability from Short Electrocardiographic Recordings Predicts Mortality from All Causes in Middle-aged and Elderly Men: The Zutphen Study," *Am. J. Epidemiol.*, vol. 145, no. 10, pp. 899–908, may 1997.

[154] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in openSMILE, the munich open-source multimedia feature extractor," *Proc. 21st ACM Int. Conf. Multimed. - MM '13*, no. May, pp. 835–838, 2013.

[155] V. T. Chang, S. S. Hwang, and M. Feuerman, "Validation of the Edmonton Symptom Assessment Scale." *Cancer*, vol. 88, no. 9, pp. 2164–71, may 2000.

[156] N. K. Aaronson, S. Ahmedzai, B. Bergman, M. Bullinger, A. Cull, N. J. Duez, A. Filiberti, H. Flechtner, S. B. Fleishman, C. J. M. D. Haes, S. Kaasa, M. Klee, D. Osoba, D. Razavi, B. Rofe, S. Schraub, K. Sneeuw, M. Sullivan, and F. Takeda, "The European Organization for Research and Treatment for Use in International Clinical Trials in Oncology," *J. Natl. Cancer Inst.*, vol. 85, no. 5, pp. 365–376, 1993.

[157] NCCN, "NCCN Distress Thermometer and Problem List for Patients," 2018. [Online]. Available: https://www.nccn.org/patients/resources/life{_}with{_}cancer/pdf/nccn{_}distress{_}thermometer.pdf

[158] A. Mehnert, D. Müller, C. Lehmann, and U. Koch, "The German Version of the NCCN Distress Thermometer: Validation of a Screening Instrument for Assessment of Psychosocial Distress in Cancer Patients," *Zeitschrift fur Psychiatr. Psychol. und Psychother.*, vol. 54, no. 3, pp. 213–223, 2006.

[159] FACIT, "FACIT Fatigue Scale (Version 4)," 2007.

[160] S. Champely, "pwr: Basic Functions for Power Analysis," 2018.

[161] J. Cohen, *Statistical power analysis for the behavioral sciences (2nd ed.)*. Lawrence Erlbaum, 1988.

[162] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.r-project.org/

[163] H. Wickham, *feather: R Bindings to the Feather 'API'*, 2016. [Online]. Available: https://cran.r-project.org/package=feather

[164] M. Hlavac, *stargazer: Well-Formatted Regression and Summary Statistics Tables*, Central European Labour Studies Institute (CELSI), Bratislava, Slovakia, 2018. [Online]. Available: https://cran.r-project.org/package=stargazer

Glossary

Notation	Description	Page List
act	activity index	59
ADL	activities of daily living	34
AP	Wi-Fi accesss point	7, 14
API	application programming interface	16
bar	air pressure (measured by the barometer)	61
BLE	Bluetooth Low Energy	24, 27, 31, 134
BLPE	blood perfusion	59
BLPW	blood pulse wave	59
CB-SMoT	clustering-based algorithm to identify stops and moves of trajectories	14
CBT	cognitive behavioural therapy	8
CE	Communaute Europeenne; compliance with directive EC 765/2008	10
class	activity class	59
cnt	message counter	59, 61, 62
CRF	cancer related fatigue	3, 4, 18, 76, 83, 132, 133, 169
DB-SMoT	direction-based Stops and Moves of Trajectories	14
DBSCAN	Density-Based Spatial Clustering of Applications with Noise	13, 15
ECOG	Eastern Cooperative Oncology Group	35, 138
EDM	E-Divisive with Medians	18
EE	energy expenditure	6, 9, 61
EORTC	European Organisation for Research and Treatment of Cancer	36, 39, 99, 103

Notation	Description	Page List
ESAS	Edmonton Symptom Assessment Scale	36, 39, 102, 138
ESM	experience-based sampling method	81
ETH	Eidgenoessische Technische Hochschule	25
FACIT-F	Functional Assessment of Chronic Illness Therapy – Fatigue	33, 76, 78, 79, 81, 83
FDA	Food and Drug Administration	10, 31
FIR	finite impulse response filter	42, 50
GPS	Global Positioning System	6, 7, 43, 49, 53–55, 81, 131
GSR	galvanic skin response	61
HF	high frequency band	11, 12, 71, 72
HR	heart rate	6, 9, 59, 61
HRV	heart rate variability	6, 11, 12, 61, 69, 71, 126, 131
ICD	implantable cardioverter defibrillator	7
ICT	information and communication technology	7, 8
IF	Isolation Forest	17
IPI	inter pulse interval	62, 69, 72
IQR	inter quartile range	81, 102
k-means	k-means	13
LF	low frequency band	11, 12, 71, 72
LOF	angle based outlier detection	17
LOF	local outlier factor	17
M.I.N.I	mini international neuropsychiatric interview	33
MAC	Media-Access-Control address	47
MFCC	Mel Frequency Cepstral Coefficients	16
MS	Multiple Sclerosis	8

Notation	Description	Page List
NCCN	National Comprehensive Cancer Network®	27, 39, 144
ncz	non-zero count	47, 48
OPTICS	Ordering Points To Identify the Clustering Structure	13, 53
PC	palliative care	2, 18, 132
PCA	principal component analysis	17
PCHIP	Piecewise Cubic Hermite Interpolating Polynomial	44
pNN_{50}	percentage of normal intervals that differ more than 50 ms	71, 72
POI	Point of interest	13, 14
PPG	photoplethysmography (optical method to measure heart rate)	6
q-q plot	quantile-quantile plot	116
QLQ-C30	quality of life questionnaire	36, 39, 86, 99, 100, 103, 122
QoL	quality of life	3, 100, 132, 140
rbf	radial basis function	67
RF	Random Forest	17
RHR	resting heart rate	112, 113, 117–119, 155
RHRV	resting heart rate variability	112, 113, 117, 119–122, 155
RMSSD	root mean square of successive differences	71, 72, 126
RR	respiration rate	6, 61
RSSI	receive signal strength indicator	7, 14, 47
SAMW	Swiss Academy of Medical Sciences	35
SDK	software development kit	16, 31, 130, 171
SDNN	standard deviation of normal intervals	71, 72

Notation	Description	Page List
SFTP	Secure File Transfer Protocol	31
SMoT	Stops and Moves of Trajectories	14
SpO_2	peripheral capillary oxygen saturation	6, 59
st-DBSCAN	spatio-temporal DBSCAN	13
steps	step count	59
SVM	Support Vector Machine	17, 42, 64, 66–68, 161
temp	(skin) temperature	61
USA	United States of America	31
UTC	Coordinated Universal Time	46–48, 59, 61, 62
VAS	visual analogue scale	28, 38, 48, 76, 79–81, 84, 86, 169
Vsteps	daily mean of the speed of steps as measured by the smartphone	117, 121–123, 125, 155
WHO	World Health Organisation	2
Wi-Fi	commonly used name for the IEEE 802.11 standard for wireless networking	6, 7, 14, 15, 25, 30, 34, 36, 37, 43, 47, 49, 54, 56, 131, 132, 173

Curriculum Vitae

Personal information

Vanessa Christina Klaas

Born on April 21, 1979, in Augsburg, Germany

Education

2015 - 2018	PhD studies (Dr. sc. ETH) in Information Technology and Electrical Engineering ETH Zürich, Switzerland.
2000 - 2007	Dipl.-Inform in Computer Science, LMU München, Germany.
2004 - 2005	Computer Science, Uniwersystet Warszawski, Warsaw, Poland, Socrates/Erasmus scholarship.
1999 - 2000	Studies of biology, Universite de Nice, France.
1999	Abitur, Rudolf-Diesel-Gymnasium Augsburg, Germany.

Work experience

2015 - 2018	Research assistant, Electronics Laboratory (IfE) ETH Zürich, Switzerland.
2012 - 2015	Systems Engineer, Biotronik SE & Co. KG Berlin, Germany.
2007 - 2012	Software & Systems Engineer, Berner & Mattner Systemtechnik GmbH, München & Berlin, Germany.