ATOPE+: Supporting Personalized Exercise Interventions in Breast Cancer Care using Mobile Technologies and Machine Learning

PhD Thesis submitted by

Salvador Moreno Gutiérrez

To obtain the international PhD degree as part of the

Doctoral Programme in Information and Communication Technologies

Under the supervision of

Oresti Baños Legrán
Miguel Damas Hermoso

Department of Computer Architecture and Technology
University of Granada

16 March 2022

Editor: Universidad de Granada. Tesis Doctorales
Autor: Salvador Moreno Gutiérrez
ISBN: 978-84-1117-340-7
URI: http://hdl.handle.net/10481/74963

Salvador Moreno Gutiérrez
ATOPE+: Supporting Personalized Exercise Interventions in Breast Cancer Care using Mobile Technologies and Machine Learning
To obtain the international PhD degree as part of the Doctoral Programme in Information and Communication Technologies - 16 March 2022
Department of Computer Architecture and Technology, University of Granada
Supervisors: Oresti Baños Legrán and Miguel Damas Hermoso

Universidad de Granada
Cuesta del Hospicio s/n
18071 and Granada, Spain

Acknowledgments

Escribir esta tesis ha sido un trayecto cargado de emociones y retos. Tengo la suerte de poder mirar atrás y ver todo lo que he crecido profesional y, sobre todo, personalmente. No podía imaginar cuántos y cuántas habéis sido partícipes de todo esto. Me habéis ayudado, no sólo a superar los obstáculos del camino, sino también a apreciar las victorias del mismo. Os doy las gracias de corazón.

En primer lugar, gracias a **Miguel Damas**, **Oresti Baños** y **Héctor Pomares** por vuestra dirección y dedicación este tiempo. Fuisteis vosotros quienes me abristeis las puertas a la investigación cuando apenas estaba terminando el grado y quienes me habéis estado aconsejando en todo momento. Gracias por todo vuestro apoyo y el trabajo que hemos compartido.

Gracias a **Víctor Amezcua** y **Pilar Gutiérrez**. Jamás me abría atrevido a meterme en el "berenjenal" de una investigación interdisciplinar sin vuestros ánimos. Tener que dejar nuestro trabajo a medias fue una pena, sin embargo, esa chispa fue imprescindible para traerme hasta aquí.

Gracias a **Irene Cantarero** y **Manuel Arroyo**. He tenido la inmensa suerte de poder colaborar con los dos mano a mano. Vuestra calidez y entusiasmo a la hora de trabajar me caló desde el primer momento y, además, ha conseguido que esta tesis tome una forma que jamás pude soñar. Me habéis enseñado a vivir una investigación muy cercana, implicada y con ánimo de aportar en el día a día de la gente. Estoy seguro de que el futuro nos aguarda más trabajo y momentos juntos que poder disfrutar.

Thanks to all those who made my research stay in the UT in Enschede memorable. **Bert-Jan van Beijnum** and **Hermie Hermens**, thanks to your welcoming reception I could made myself a place in the group from the very beginning of the stay. **Kostas Konsolakis** and **Miriam Cabrita**, thank you, not only for the work we shared —I learned a lot with our discussions—, but also for your amazing human quality. Lastly, **Jade** and **Anna**, thanks to you I could spend most of the stay with the greatest joy possible and "“a pint” of Mediterranean lifestyle.

Por supuesto, dar las gracias también a todos los que me han acompañado en paralelo en esta travesía por el doctorado. Me siento halagado de haberos tenido a

todos a mi alrededor en todo momento, me llevo mucho de cada uno de vosotros. **Carlos**, gracias por todo lo que hemos compartido desde hace tanto tiempo. Pierdo la cuenta de la cantidad de experiencias que hemos vivido juntos, y no se me ocurre a nadie mejor con el que haber peleado una tesis al mismo tiempo. Has sido mi guía en muchísimos momentos, y ha sido un gustazo ver lo bien que te ha ido en todo este proceso. Sé que nos queda mucho por delante y, conociéndonos, seguro que seguiremos compartiendo muchas "batallitas". Un abrazo enorme. **Paula**, ha sido increíble compartir todo este trabajo juntos durante la tesis, tus ganas y predisposición son contagiosas. Compartir proyecto ha hecho que crezcamos prácticamente a la par durante la tesis, pero también que los ánimos no faltaran, sobre todo en la recta final. **Ofelia**, compartir despacho ha sido lo de menos —tampoco es que hayamos coincidido mucho en él x)—, ya desde que te conocí en el máster sabía que estabas hecha de otra pasta. Han sido muchas las tardes en las que nos hemos juntado a hacer cualquier otra cosa que no fuese "la tesis", y que nos han traído risas y alegría. Que sepas que nos quedan muchas más de esas aunque hayamos finalizado esta etapa. **Manolo**, tú ya eres de otro mundo. Traes entusiasmo, curiosidad y generosidad a raudales allá donde te pongas. Hemos compartido nuestros mejores y peores momentos durante todo este tiempo, y siempre he sentido tu apoyo y ayuda muy cercanos y auténticos. Es una suerte tenerte como amigo. Un abrazo enorme.

Además, también tengo que dar las gracias los que me llevan acompañando desde hace mucho, también han aportado su granito de arena. Gracias a **Juangui**, **Arturo**, **David** y **Javi**. Vuestra amistad me lleva acompañando por mucho tiempo, hemos crecido juntos y, aunque cada uno va tomando su camino, hacéis posible que me sienta como en casa cada vez que nos juntamos. Gracias a **Salva**, **Zuñi**, **Jorge**, **Juanma** , **Josema** y **Lucas**. Sois increíbles chicos, movéis una energía tremenda, y sólo puedo sentirme afortunado de poder ir a vuestro lado. Han sido muchos los momentos que hemos compartido y guardo mucho de vosotros. Gracias también a la música por haberme juntado con gente tan increíble como **Saúl**, **Luis**, **Roberto**, **Edu** y **Fran**. Habéis conseguido que cumpla el sueño de montar no uno, ¡sino dos grupos! Además de componer, grabar, y hasta dar conciertos. He aprendido y disfrutado con vosotros lo más grande. ¡Nos queda mucho ruido por hacer¡.

Gracias, por supuesto, a toda mi familia, tengo la suerte de que es enorme y está llena de referentes. A **Mamá**, **Papá**, mi hermana **Maria Victoria** y mi hermano **Marcos** por ser los mejores ejemplos para mí, habérmelo dado todo, y por vuestro apoyo incondicional. A mis abuelos, por ser los mejores que nadie puede soñar. A **Mama-Paita** y **Papa-Pepe**, que están cargados de ternura y sabiduría al mismo tiempo; y a **la Mama** y **el Papa**, por todo vuestro cariño y apoyo. Siempre me

acuerdo de vuestras anecdotillas viéndome crecer. Gracias también a mis tías y tíos, que tanto me han influido en todo este tiempo, **Julia**, **Migue**, **Pilar**, **Pedro**, **María Victoria**, **Gonzalo**, **María del Mar**, **José Luis**, **Jesús** y **Ana**. Por todos esos días en Alfacar, también con los primos. Son muchos los recuerdos que os guardo con muchísimo cariño.

Por último, gracias a ti, **Patri**, por haber sido mi luz en todo este camino. Todo lo bueno pasa por ti, la energía que desprendes es contagiosa, y gracias a ella he podido llegar hasta aquí. En ningún momento me ha faltado tu apoyo, hemos aprendido el uno del otro a la par que crecíamos, y no puedo ser más afortunado habiendo estado a tu lado en todo este tiempo. Te quiero muchísimo.

Antes de irme (y de hacer esta sección demasiado larga), no puedo evitar al menos hacer mención a mucha otra gente que también ha tenido su papel en esto: **Maria José, Adrián, Vero, Marcial, Eugenia, Tao, Ramón, Ana, Gregor, Fran, Javi, Andrés, Nono, Zoraida, Francisco, Coco, Mantequilla** y **Truman**. Gracias a todos.

Salva
(Dollo)

Grants and Funding

The work of Salvador Moreno Gutiérrez was supported by the *Spanish Ministry of Education, Culture, and Sport* with a Predoctoral Fellowship as part of the *University Faculty Training (FPU) Programme* with grant number FPU16/04201.

In addition, the contributions presented in this thesis were partially supported by the *Health Research Funds of the Carlos III Health Institute* under Project PI18/01840M; and the *Spanish Ministry of Science, Innovation, and Universities (MICINN)* under Projects PGC2018-098813-B-C31 and RTI2018-101674-B-I00.

Abstract

Alleviating the burden of breast cancer has become in one of the biggest challenges of our times. The advances in surgery, radiotherapy, and systemic therapy have improved the survival rates of patients with breast cancer, but have also produced a higher number of patients suffering short- and long-term side effects, with high the risk of recurrence, developing comorbidities, and death. Therapeutic exercise poses a means to address this issues; however, exercise interventions in patients with cancer are often adhered to the same therapeutic exercise guidelines. This results in one-size-fits-all exercise prescriptions for all adults, regardless their individual exercise capabilities and needs, which may lead to inadequate training adaptation.

The mobile health (mHealth) paradigm has enabled the remote and individual monitoring of health through wearable sensors and smartphones. Personalizing training adaptation with an mHealth approach has already been successfully conducted in sports settings, and the literature suggests that similar strategies may translated to patients with chronic conditions such as breast cancer. However, recent works do not target the adjustment of training doses to the individual needs of the patients.

This thesis presents three contributions to support the personalization of therapeutic exercise intervention in patients with breast cancer. First, ATOPE+, an mHealth system to support the remote monitoring of patients' training load through heart rate variability (HRV), self-reported wellness, and Fitbit physical activity and sleep data. ATOPE+ also integrates a decision-support system with expert rules that automatically trigger daily exercise recommendations for patients. Second, the ATOPE+Breast dataset, an open dataset describing the continuous evolution of training load during therapeutic exercise intervention for 23 patients with breast cancer. Third, a clustering approach to assess training needs in patients with breast cancer. Data science and artificial intelligence (AI) are leveraged in this approach to better understand the different states of the patient throughout an exercise intervention, and eventually serve as a tool to make more informed decisions when prescribing an exercise dose.

The potential of these contributions may lead to new research directions in the personalization of therapeutic exercise interventions in real-life scenarios, specially regarding the application of mHealth and AI to improve chronic conditions.

Resumen

Aliviar las secuelas del cáncer de mama se ha convertido en uno de los mayores retos de nuestros tiempos. Los avances en cirugía, radioterapia y terapia sistémica han mejorado las tasas de supervivencia en pacientes con cáncer de mama, pero también han traído consigo un elevado número de pacientes que sufren de efectos secundarios a corto y largo plazo, con elevado riesgo de recurrencia, de desarrollo comorbilidades, y de mortalidad. El ejercicio terapéutico se plantea como una solución para estos problemas; sin embargo, las intervenciones de ejercicio físico van normalmente dirigidas bajo las mismas guías de ejercicio terapéutico para pacientes con cáncer. Esto conlleva la entrega de la misma prescripción de ejercicio para todos los adultos, independientemente de sus capacidades y necesidades de entrenamiento específicas, lo que puede llevar a un mala adaptación durante el entrenamiento.

El paradigma de salud móvil (mSalud) ha permitido la monitorización remota e individualizada de la salud a través de sensores vestibles y teléfonos inteligentes. La personalización de la adaptación al entrenamiento con tecnologías mSalud ya se ha llevado a cabo con éxito en entornos deportivos, y la literatura sugiere que estrategias similares se pueden trasladar a pacientes con condiciones crónicas como el cáncer de mama. Sin embargo, trabajos recientes se olvidan del ajuste individual de las dosis de entrenamiento a las necesidades de cada paciente.

Esta tesis plantea tres contribuciones para la personalización de intervenciones de ejercicio físico terapéutico en pacientes con cáncer de mama. Primero, ATOPE+, un sistema mSalud para la monitorización remota de la carga de entrenamiento en pacientes con cáncer mediante la medición de variabilidad de la frecuencia cardíaca (VFC), bienestar autoreportado, y actividad física y sueño Fitbit. ATOPE+ también incluye un sistema de ayuda a la toma de decisiones que, mediante reglas expertas, entrega recomendaciones de ejercicio diarias para las pacientes. Segundo, ATOPE+Breast, un conjunto de datos abierto que describe la evolución continua de la carga del entrenamiento a lo largo de una intervención de ejercicio terapéutico para 23 pacientes con cáncer de mama. Tercero, un análisis, basado en algoritmos de agrupamiento, orientado a la evaluación de necesidades de entrenamiento en pacientes con cáncer de mama. La ciencia de datos y la inteligencia artificial (IA) permiten desde este análisis mejorar el entendimiento de los diferentes estados de la paciente a lo largo de una intervención en ejercicio físico, así como, en última instancia, servir como herramienta para tomar decisiones más informadas al preescribir dosis de ejercicio terapéutico.

El potencial de estas contribuciones permite la apertura de nuevas líneas de investigación dirigidas a la personalización de intervenciones de ejercicio terapéutico en escenarios de la vida real, especialmente en la aplicación del mHealth y la IA en la mejora de condiciones crónicas.

Contents

1 Introduction

> *Without deviation from the norm, progress is not possible.*
>
> — **Frank Zappa**

1.1 Thesis Goal

The goal of this thesis is to investigate how to support personalized therapeutic exercise interventions in patients with breast cancer using mobile technologies, data science, and machine learning. In particular, this thesis focuses on the continuous monitoring of training load in patients with breast cancer during a therapeutic exercise intervention. This work aims to contribute with the development of systems to support improved decision-making for experts when prescribing personalized exercise doses for patients with breast cancer.

1.2 Context

1.2.1 Breast Cancer in the Last Years

Cancer is the plague of the 21st century; it is a chronic disease that affects all populations regardless of wealth or social status. In 2018, cancer diagnoses raised to 18.1 million people globally, and 9.6 million died from the disease (Bray et al., 2018). Moreover, in 2020, cancer was responsible for one in six deaths globally, and one in five people had already faced cancer diagnosis during their lifetime (World Health Organization, 2020).

Population growth and aging may be the principal causes of the increase in cancer incidence in the last years, but other factors related to social and economic growth also play an essential role. These factors are mainly related to lifestyle habits in diet and physical activity (World Health Organization, 2020; Wild, Weiderpass, &

Stewart, 2020). For instance, a daily intake of red or processed meat may increase the risk of colorectal cancer (World Health Organization, 2015); and lower levels of physical activity may increase the risk of several types of cancer, such as bladder, breast, colon, endometrial, esophageal, kidney, and stomach cancer (McTiernan et al., 2019; Patel et al., 2019; Rezende et al., 2018). Such habits are typical of richer countries (Popkin, Adair, & Ng, 2012), and Europe is a fair representation of this case. With only 9% of the world population, the old continent held 23.4% of cancer diagnoses and 20.3% of cancer-related deaths in 2018 (Bray et al., 2018). This high magnitude of cancer in Europe made the European Union include it as one of the Missions of the Horizon Europe program for 2030 (European Commission, 2021).

Breast cancer (BC) is one of the most prevailing types of cancer affecting the population. Accounting for both sexes, BC was the most commonly diagnosed cancer (11.6% of the total cases) in 2018, tied with lung cancer (11.6%), and followed by colorectal (10.2%) and prostate (7.1%) cancer. Among females only, BC was the leading cause of cancer incidence with 2.1 million new cases (24.2% of the total cases) and mortality with 630 000 deaths (15.0% of the total deaths) (Bray et al., 2018).

BC is present all around the globe. It is the most frequently diagnosed cancer in most countries (154 of 185) and the leading cause of cancer mortality in more than 100 countries. BC incidence rates are the highest in the more developed countries, such as Australia and New Zealand, Northern Europe (e.g., Finland, Sweden, Denmark), Western Europe (e.g., Belgium, The Netherlands), Southern Europe (e.g., Spain, Italy, Portugal), and Northern America (Bray et al., 2018). Although hereditary factors account for 5% to 10% of BC cases, the higher BC incidence found in more developed countries is often associated with non-hereditary factors (Wild et al., 2020; Ziegler et al., 1993). These risk factors are associated with the socioeconomic context and lifestyle habits, such as reproduction (nulliparity, late age at first birth, and fewer children), menstruation (early age at menarche, later age at menopause), exogenous hormone intake (oral contraceptive use and hormone replacement therapy), nutrition (alcohol consumption), and anthropometry (greater weight, weight gain during adulthood, and body fat distribution) (Wild et al., 2020; Ziegler et al., 1993). Conversely, physical activity and breastfeeding are known as protective factors (Wild et al., 2020; Brinton, Gaudet, & Gierach, 2017; Ballard-Barbash et al., 2012; Schmid & Leitzmann, 2014).

Despite the incidence increase in BC, its outcomes are improving. The advances in BC detection and its treatment have resulted in decreased mortality within the last two decades (Bray et al., 2018). Nevertheless, this has also led to an increase of patients and survivors that suffer significant short- and long-term side effects (Patsou, Alexias, Anagnostopoulos, & Karamouzis, 2018), raising the risk of disease recurrence, developing comorbidities, or death (Wild et al., 2020). In fact, by the end of 2020, there were 7.8 million women alive who had been diagnosed with BC in the previous five years, making it the most prevalent cancer in the world (Wild et al., 2020). Hence, preventing or reducing the side effects of BC treatment is of great importance (Peterson & Ligibel, 2018).

Therapeutic exercise (TE) poses a means to address the short and long-term side effects of cancer and its treatment (Ballard-Barbash et al., 2012; Schmid & Leitzmann, 2014). TE is a subset of physical activity (PA) consisting of structured and repetitive planned movements and activities with a therapeutic aim. On the other hand, the definition of PA is broader; PA consists of any movement produced by skeletal muscles involving energy expenditure. TE and PA have consistently reported benefits to patients with cancer (Garcia & Thomson, 2014), and both are often recommended for prevention and treatment purposes (American Cancer Society, 2016; World Health Organization, 2021; Patel et al., 2019). Together with medical and surgical treatments, TE improves survival and reduces recurrence and mortality risks (Pollán et al., 2020) due to its positive impact on factors related to the quality of life (Lahart, Metsios, Nevill, & Carmichael, 2015; Cormie, Zopf, Zhang, & Schmitz, 2017; Peterson & Ligibel, 2018).

The benefits of TE interventions made the research community seek new means to deliver TE interventions in remote environments leveraging mobile technologies (Muller et al., 2018). To date, mobile health (mHealth) PA interventions are a feasible, cost-effective way to improve overall activity levels, body composition, quality of life, and self-reported symptoms in patients with cancer (Schaffer et al., 2019) and survivors (Roberts, Fisher, Smith, Heinrich, & Potts, 2017).

1.2.2 Mobile Technologies in Breast Cancer Care

Mobile technologies have quickly spread around the globe in the last ten years. The number of smartphone users worldwide raised from 751 million in 2011 to 4704 million in 2021, and it is expected to grow to 5575 million in 2025 (Statista - The Statistics Portal, 2021b). More specifically, Europe raised from 125 million users in 2011 and 634 million in 2021, expecting to grow to 709 million in 2025 (Statista - The Statistics Portal, 2021a). These absolute numbers translate into 60.4% of the world population and 84.8% of the European population having a smartphone in 2021.

A similar trend can be drawn for the number of wearable devices connected, such as wearable wrist-worn activity trackers, smartwatches, or even smartrings. In 2021, there were 928.8 million units connected in the world (11.9% of its population) and 215.3 million in Europe (28.8% of its population) (Statista - The Statistics Portal, 2021c). The wide adoption of mobile technologies and their sensing capabilities offers new opportunities for continuous and personal monitoring.

Smartphones, smartwatches, and wrist-worn activity trackers —among other *smart* devices— provide users with enormous capabilities for communication, access to information, and personal sensing. These devices are typically equipped with sensors like accelerometers and gyroscopes for activity recognition; barometric sensors to measure altitude changes; GPS (Global Positioning System) for tracking services; ambient light sensor for automatic screen brightness adjustment; front cameras and fingerprint scanners for biometric authentication; or photoplethysmography for heart rate monitoring. These embedded sensors enable the unobtrusive monitoring of the users' daily activities and their surrounding context. Besides, smartphones provide an interface for communication, the connection of wearable sensors (e.g., via Bluetooth), and the self-report of people's wellbeing. Smartphones and wearable activity trackers have enabled advances for a more objective and personalized approach in several health applications such as promoting physical activity (Brickwood, Watson, O'Brien, & Williams, 2019), monitoring mental health (Garcia-Ceja et al., 2018), or even fighting the effects of the COVID-19 pandemic (Khan et al., 2021).

The promise of mobile technologies and wearable activity trackers in oncology has been present in recent literature. Clinical experts' defended that including continuous, objective, and quantified measures of the physical activity and surrounding context of patients could provide new advances into personalizing cancer treatment (Kelly & Shahrokni, 2016; Beg, Gupta, Stewart, & Rethorst, 2017). Moreover, the ergonomics and affordability of commercial wearable activity trackers could enable longer monitoring times, instead the research-grade activity trackers typically used (Peddle-Mcintyre et al., 2018). This potential led the research community to leverage mobile and wearable technologies in numerous cancer-related studies successfully (Martin et al., 2021; Faro et al., 2021; Beauchamp, Pappot, & Holländer-Mieritz, 2020; Dorri, Asadi, Olfatbakhsh, & Kazemi, 2020; Schaffer et al., 2019).

To date, patients with breast cancer and survivors are the main targets of mHealth and eHealth studies related to cancer (Martin et al., 2021; Dorri et al., 2020; Chung et al., 2020). Even in some of the systematic reviews targeting all cancer types and the use of mobile technologies, breast cancer gets most of the attention (Schaffer et al., 2019; Gresham et al., 2018). The gathered evidence presents breast cancer patients as an ideal target population for implementing these novel health intervention paradigms.

Finally, the new coronavirus diseases (COVID-19) strike highlighted the need for tools to deliver remote interventions. The recurrent saturation of hospital resources and the risk of getting infected required new treatment strategies for immunosuppressed patients, like patients with cancer, who had a twofold increased risk of getting infected with the COVID-19 compared to the general population (Yu, Ouyang, Chua, & Xie, 2020). The inclusion of telemedicine strategies was strongly recommended to minimize the exposure of the most vulnerable patients and prioritize individual assistance (Al-Shamsi et al., 2020).

1.3 Motivation & Objectives

Therapeutic exercise poses a means to address the short- and long-term side effects of cancer and its treatment; however, the personalization of exercise interventions still presents a challenge. Personalizing a therapeutic exercise intervention consists of tailoring it to each patient's needs, characteristics, or possibilities with an adapted and evidence-based prescription following frequency, intensity, time, and type (Campbell et al., 2019). Typically, the tailoring of an exercise intervention relies on patient's self-management to regulate the intensity of the sessions. Intervention programs are already scheduled to meet the weekly exercise recommendations, while any adjustment is only based on demographic variables (e.g., age, height, weight) and surgery or systemic treatment (e.g., chemotherapy, radiotherapy) dates.

Mobile technologies may improve the personalization of exercise interventions for patients with breast cancer through remote and real-time assessment of their status. Smartphones and wearable sensors may enable the daily monitoring of biomarkers to evaluate individual training needs. Furthermore, data science techniques and machine learning algorithms may provide a means to understand such biomarkers' role during a therapeutic exercise intervention. There are already available solutions looking to optimize training and performance that leverage such tools to a certain extent for amateur and professional athletes —even commercially, such as HRV4Training or EliteHRV. However, there is little research focused on personalizing therapeutic exercise interventions in patients with cancer.

Hence, as its main objective, this thesis aims to investigate how to support personalized therapeutic exercise interventions in patients with breast cancer using mobile technologies, data science, and machine learning. To pursue its goal, this thesis states the following supportive objectives.

Objective 1: Develop an mHealth expert system to support personalized therapeutic exercise interventions in patients with breast cancer.

An mHealth intervention involves the use of mobile technologies from patients and/or clinical experts. In order to make an mHealth intervention successful, it is of utmost importance to meet both patients' and experts' needs from the beginning (Marcolino et al., 2018). Moreover, the requirement analysis cannot be exclusively addressed solely from a clinical (Granja, Janssen, & Johansen, 2018) or technological (Banos, Villalonga, et al., 2015) perspective, but include both.

The clinical experts delivering a remote therapeutic exercise intervention (i.e., physiotherapists) may have multiple needs, from communicating with a patient to displaying patients' information. Nevertheless, the most critical requirement is assessing the training needs of the patient, and just providing a wearable activity tracker may not be enough to monitor patients' health (Schaffer et al., 2019). Patients' assessments must be gathered with objective and quantifiable health status measures with multiple biomarkers. Furthermore, the different monitoring methods used (e.g., questionnaires, wearable sensors, activity trackers) must provide reliable information to support medical decisions.

The patient's health status must be presented to the experts in a clean and structured manner to support their decisions regarding the prescription of exercise doses. Knowledge-based systems provide an excellent framework to support the relation of medical knowledge (e.g., a base of rules, an ontology) to match the data gathered and, eventually, trigger recommendations. This expert-driven approach may work in parallel with a machine-driven approach, in which data science techniques and machine learning algorithms could provide different insights and perspectives for a more refined analysis and decision-making process.

The patients with breast cancer receiving a remote therapeutic exercise intervention have their own needs too. Patients must be able to record their health status autonomously to inform the clinical experts on their status and, eventually, receive their exercise prescription. However, this should be achieved in the most straightforward manner. Any medical tool or system developed must be easy to use, reliable, and never imply any risks for the patient. Simplicity-from-design may provide a friendly mHealth environment for patients with low technological skills —such as the often found among the elderly, a part of the population with the highest cancer incidence risks and rates.

The surgery, radiotherapy, and systemic treatment (e.g., hormonal therapy, targeted therapy, or chemotherapy) of breast cancer care may translate into functional limitations like impairments in the arm and shoulder (Hidding, Beurskens, Wees, Laarhoven, & Sanden, 2014). Therefore, the sensors and protocols followed need to consider such limitations to provide a safe mHealth intervention for the patients.

There are also shared requirements among patients and clinical experts. One of them is acknowledging that technology cannot replace the contact with a clinical expert (Granja et al., 2018). The supervision of an expert during exercise intervention is necessary to ensure adherence to the protocol and its correct execution. The different exercises need to be carefully instructed to be executed correctly, thus avoiding any harm. Besides, the physiotherapist is responsible for supervising that the patient is adjusting exercise intensity to the recommended levels. The developed mHealth system must consider these needs to make the intervention successful. Another shared requirement is that the patient sees the mHealth system as an assistant for the clinical expert, not a substitute. Physiotherapists may use the information to support their decisions and establish solid communication regarding the patient's status despite the remote environment. Such information and communication channels may enable quick, personalized adjustments of the individual exercise prescription.

With the spread of the COVID-19 pandemic, the availability of adequate mHealth interventions has become essential. Due to the immunosuppression related to cancer treatment, the general recommendation is to minimize patients' exposure and prioritize their individualized assistance to avoid any risk of getting infected with COVID-19 (Al-Shamsi et al., 2020). Despite the current availability of vaccines, the mHealth environment for the exercise intervention needs to be prepared for new outbursts or variants of COVID-19. Thus, the set-up mHealth environment needs to be prepared for in-situ assistance from the physical therapist, but also consider a fully-remote intervention scenario.

Through this objective, this thesis aims to design and develop an mHealth expert system capable of supporting personalized therapeutic exercise interventions in patients with breast cancer. For it to be possible, the system must contain experts' and patients' needs, besides collecting the necessary biomarkers to assess the patients' health state throughout a therapeutic exercise intervention.

Objective 2: Conduct a monitorization experiment of patients with breast cancer through therapeutic exercise intervention, and generate a longitudinal dataset with training load measures.

Exercise interventions in patients with breast cancer typically rely on self-management to adjust the intensity of the intervention. However, intervention response may differ from one patient to another at a physiological level, especially regarding whether the patient was already used to exercising or not, resulting in different individual adaptations to the training intervention. Besides, exercise programs are predefined and scheduled to meet a minimum of training amount a week, setting aside factors the patient's status at the moment of training. All these limitations may result in the undertraining or overtraining of the patient throughout intervention (Jones, Eves, & Scott, 2018; Carter et al., 2021).

The literature collects different attempts in personalizing mHealth exercise and physical activity interventions with the help of smartphones and wearable activity trackers (Beauchamp et al., 2020; Dorri et al., 2020; Schaffer et al., 2019). Some of the studies leveraged the monitoring capabilities of digital activity trackers, presenting the information to the expert to make more informed decisions and/or to the patients to increase their self-awareness on daily physical activity levels. Other studies even used activity tracker data to adjust the baseline for the physical activity levels recommended during an intervention. These approaches came with several benefits, such as increasing overall physical activity levels (Schaffer et al., 2019; Dorri et al., 2020) and increasing adherence due to a sense of accountability (Gell, Tursi, Grover, & Dittus, 2020). Nevertheless, the methods used did not mean an improvement compared to the traditional scheduled exercise interventions (Uhm et al., 2017), suggesting that self-monitoring with a wearable activity tracker or leveraging steps-count for personalizing recommendations may not be enough to personalize exercise interventions.

Exercise load monitoring poses a solution to assess individual training needs of patients with breast cancer during an exercise intervention. Although it has been well established for professional sports, exercise load monitoring principles may be translated into clinical practice with measures of heart rate variability, wellness, and physical activity levels to evaluate the readiness level to train (Carter et al., 2021).

Collecting quality and reliable longitudinal training load data in free-living environments is one of the main challenges researchers face when assessing training needs in patients. Clinical interventions often rely on sophisticated equipment that limits measurements to lab settings. This limitation results in a lack of longitudinal descriptions of training load in interventions with chronic conditions such as breast cancer, only reporting values such as pre-training, post-training, de-training, or control (Y.-H. Lee, Lai, Lee, Tsai Lai, & Chang, 2018; Guo et al., 2015; Niederer et al., 2012; Caro-Moran et al., 2016; Dias Reis et al., 2017; De Couck & Gidron, 2013; Freitag et al., 2018).

Through this objective, this thesis aims to design and conduct the longitudinal data collection of training load in patients with cancer enrolled in a therapeutic exercise intervention. This work plans to build and curate an open dataset available to the scientific community to enable and share novel methodological approaches to the longitudinal analysis of training needs in patients with breast cancer.

Objective 3: Identify the factors reflecting the individual readiness state of patients with breast cancer during therapeutic exercise intervention using a data science and machine learning approach.

Exercise load monitoring poses a solution to assess individual training needs of patients with breast cancer during an exercise intervention. Exercise load monitoring is well established in professional sports to assess readiness levels to train and perform; however, little is known about translating exercise load monitoring principles into clinical practice (Carter et al., 2021). Professional athletes rely on biomarkers like heart rate analysis, wellness questionnaires, and ratings of their perceived recovery to evaluate how much impact the last training had on their bodies (Miguel, Oliveira, Loureiro, García-Rubio, & Ibáñez, 2021). The role and significance of such biomarkers when assessing training needs in patients with breast cancer —or any other type— is still unknown.

The massive inflow of data that smartphones and wearable activity trackers provide may be integrated into a digital phenotype for patients with cancer (Fonseka & Woo, 2021; Carissa A. Low, 2020). Digital phenotyping is the moment-by-moment quantification of the individual-level human phenotype *in situ* using data from smartphones and other personal digital devices (Torous, Onnela, & Keshavan, 2017). Eventually, the construct of digital phenotyping is nurtured by the several biomarkers addressing the different health layers monitored in a patient. Digital phenotyping is a prevalent field in mental health that focuses on assessing (Torous et al., 2017; Insel, 2017), but it is less studied in oncology (Fonseka & Woo, 2021). Although some studies show the possibilities of mobile technologies and machine learning in cancer, these studies are mostly preliminary work. These works address different dimensions of cancer, such as the assessment of symptom severity (Carissa A. Low et al., 2017; Carissa A. Low et al., 2021) or fatigue (Sada et al., 2021). The analysis of biomarkers monitoring the exercise needs of patients with breast cancer may contribute to the state of the art on digital phenotyping applied to cancer.

Through this objective, this thesis aims to identify factors reflecting the individual readiness state of patients with breast cancer. Novel methodologies of analysis based on data science and machine learning techniques may provide new insights on how to translate exercise load monitoring into cancer care. Unsupervised learning may enable the search of unknown patterns among the different variables presented (e.g., clustering algorithms). Supervised learning may enable the search for factors reflecting expert knowledge (e.g., logistic regression, random forests, support vector machines) along with feature importance analysis (e.g., gini/entropy criteria).

1.4 Outline

This thesis is structured in six chapters:

Chapter 1 introduces the context for this thesis by presenting the state of breast cancer in the last years and the state of the use of mobile technologies in breast cancer care. Next, it lays out the motivation and objectives of this work, bringing up the current challenges and opportunities in designing mHealth systems, collecting longitudinal data during therapeutic exercise intervention, and identifying factors reflecting the individual readiness state of patients during therapeutic exercise intervention.

Chapter 2 provides an overview of the state of the art for digital health systems, the role of physical activity and exercise in patients with breast cancer, and the monitoring of exercise load in breast cancer. The section for digital health systems describes the principal tools to design a digital health system, decision support systems, the current status of mobile health systems and frameworks, especially those applied to cancer, and commercial activity trackers. The section for physical activity and exercise in patients with breast cancer lays out the benefits of these approaches and the limitations found due to the lack of personalization strategies. The section for exercise load monitoring in breast cancer describes the principal technologies for assessing training load.

Chapter 3 presents the design of an mHealth system to support the personalization of therapeutic exercise intervention in patients with breast cancer, describes the development of the different elements composing the system, and provides a usability evaluation with physiotherapists experienced in therapeutic exercise and patients with breast cancer and survivors.

Chapter 4 describes the longitudinal collection of training load measures for patients with breast cancer during a therapeutic exercise intervention. This chapter also opens the dataset collected to the research community.

Chapter 5 investigates the patterns for training load data across a therapeutic exercise intervention. In particular, it describes a novel methodology to clean, process, and select the most relevant features identifying the individual readiness level of breast cancer patients to exercise.

Finally, Chapter 6 concludes stating the main achievements and contributions of this thesis. It also presents future opportunities on the research directions opened by this work.

2 State of the Art

2.1 Digital Health Systems

In the last decades, information and communication technologies (ICT) have provided numerous solutions to challenging problems related to healthcare, such as the administration of electronic health records (Dinh-Le, Chuang, Chokshi, & Mann, 2019), the use of machine learning for cancer detection in images (Ravì et al., 2017), or the Internet of Things (IoT) paradigm for the remote monitoring of chronic conditions using wearable sensors (Siow, Tiropanis, & Hall, 2018). All these systems are known as *digital health systems*.

A digital health system (DHS) supports health interventions and management through electronic and mobile technologies. A DHS rely on different approaches to assist patients and healthcare providers in their needs. Some of these DHSs are based on electronic health (eHealth) methods, such as storing and delivering information or enabling doctor-patient communication through mail or websites. In contrast, other DHSs are based on mobile health (mHealth) approaches, such as SMS messaging, in-app communication, or sensor monitoring.

The mHealth approach may leverage smartphone and wearable sensors for activity recognition (Nweke, Teh, Mujtaba, & Al-garadi, 2019) and context monitoring in different applications, like mental health, diabetes, or cancer (Carissa A. Low, 2020). These mobile interventions provide the possibility of obtaining large datasets generated by ecologically valid measures of behavior, thinking, emotion, and physiology (i.e., in real-time and everyday contexts).

This section reviews the state of the art of DHSs applied to cancer, specifically the mHealth approaches. First, this section covers how to design a digital health system. Second, the concept of decision support systems is introduced. Third, a portfolio of mobile health frameworks, systems, and applications is presented, focusing on cancer. Finally, this section also provides a view on the role of commercial activity trackers for health applications.

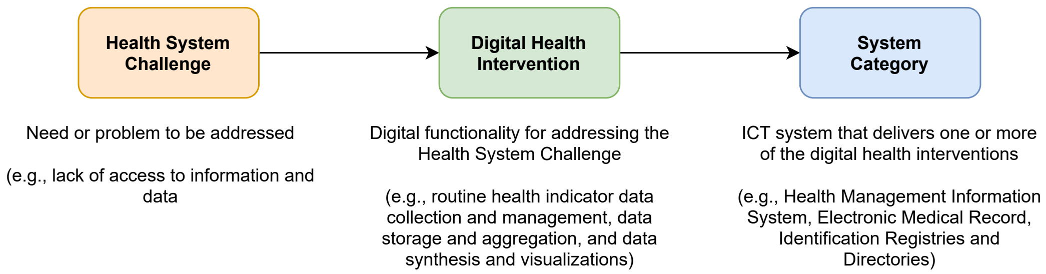
2.1.1 Designing a Digital Health System

Crafting a DHS is far from trivial; it requires a refined analysis of a health challenge to be addressed while considering the needs of the patients and healthcare providers. There are several standards detailing the development processes of systems in general, like the ISO/TC 215 for Health Informatics, the IEEE 1233-1998 for Developing System Requirements Specifications, or the IEEE 1012-2016 for System, Software, and Hardware Verification and Validation. However, these guidelines are closer to the technical side than healthcare managers and providers. The WHO published a taxonomy of digital health interventions (World Health Organization, 2018) to ensure good communication among the different professionals involved in developing a DHS (i.e., funding stakeholders, engineers, clinicians, researchers) while contemplating already defined standards. This taxonomy provides a framework with bridging language for health professionals with ICT professionals, which is of utmost importance to develop valuable systems targeting healthcare needs.

First, a *health system challenge* needs to be defined to design a DHS (i.e., the problem to be addressed). Second, a *digital health intervention* is designed to address that health system challenge. Third, a *system category* or *type* is assigned to the necessary DHS delivering such intervention. Figure 2.1 shows the linkage for the three elements. Finally, a detailed classification can be found for each element in Appendix A.

Taxonomies like the WHO/RHR/18.06 advocate for the inclusion of users in the design and development of mHealth systems. This co-design strategy aims to avoid the pitfalls resulting from a lack of communication and integration among stakeholders.

Fig. 2.1.: Linkages across Health System Challenges, Digital Health Interventions, and System Categories (adapted from WHO/RHR/18.06 (World Health Organization, 2018)) .



2.1.2 Decision Support Systems

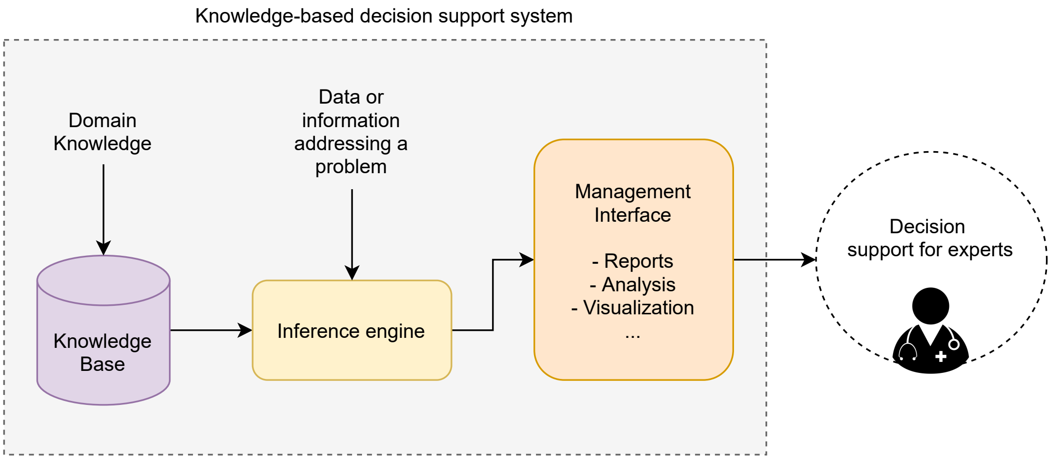
A decision support system (DSS) is a type of recommender system that facilitate the decision-making process in an organization by presenting helpful information from a combination of raw or processed data (Aggarwal, 2016). DSSs inform a user or an expert to make decisions, even suggesting a recommendation according to the available data. For instance, a DSS may help a physician decide which drug is most appropriate for a patient according to the patient's history and a drug trial database. A DSS may also be considered an implementation form of a knowledge management system (Wiig, 1997) that facilitates organizational processes within a company or workgroup. For instance, a DSS may help a paper company refill stock according to information from the sales department and warehouse information.

DSSs are typically structured as knowledge-based systems relying on a knowledge base and an inference engine (Figure 2.2). The knowledge base represents facts about the world, which may be organized as a rule base or as an ontology. The inference engine matches the available data with the knowledge base to trigger the appropriate rules. This paradigm applied to health may support many different applications, even allowing individual health recommendations for patients (Ertuğrul & Elçi, 2019). In research, knowledge-based DSSs are also useful, allowing to leverage domain knowledge while refining it (e.g., applying general rules to a different cohort of patients, adding new observation variables that may relate to the already known rules) (Karpatne et al., 2017). Moreover, a data-driven analysis approach may work parallel to a knowledge-driven approach in these systems, enabled by unsupervised or prediction methods (e.g., unsupervised machine learning algorithms, time series prediction).

In order to leverage domain knowledge, most DSSs need sophisticated data processing to transform the raw data into useful information. Recent works in the literature address this issue, processing data through layered architectures until the generation of recommendations —or support of decisions.

Mining Minds is an example of this combination of rich data processing to leverage domain knowledge. Banos et al. (2016) developed the Mining Minds digital health and wellness framework to provide personalized support leveraging the concepts of context-awareness and knowledge-based reasoning with smartphones, wearables, and the Internet of Things. Mining Minds presents a hierarchical multi-layer architecture (Figure 2.3) to transform raw data from multimodal sources (e.g., smartphone sensors, wearable ECG, external API). First, a *data curation layer* is responsible for acquiring, curating, and persisting the data obtained in order to be processed for a

Fig. 2.2.: Architecture of a knowledge-based decision support system.



higher level of understanding. Second, the *information curation layer* represents the core for the inference and modeling of the user context, translating the data from the first layer into categories like physical activities, emotional states, locations, and social patterns. Third, the *knowledge curation layer* enables the creation of health and wellness knowledge using domain expert knowledge or engineered knowledge through expert-driven or data-driven approaches. Finally, a service curation layer provides service communication between layers, analytics, and rule-authoring tools for the experts supervising an intervention conducted through Mining Minds.

Despite the flexibility and adaptability of Mining Minds to conduct health interventions, the availability of this framework is closed to its developers in its research facilities. Afzal et al. (2018) used Mining Minds (Banos et al., 2016) to generate context-aware personalized recommendations in 40 contextually different scenarios (e.g., home, sleeping, windy weather, and neutral emotion; restaurant, sleeping, windy weather, sadness) tested in 50 participants. Ali et al. (2018) used Mining Minds (Banos et al., 2016) with a knowledge-based reasoning and recommendation framework to generate personalized recommendations to promote active lifestyles and reduce sedentary behavior.

Mining Minds architecture participates from the mobile health paradigm to the integration of multimodal mobile sources of data such as wearable and smartphone sensors. There are several mHealth frameworks and systems that leverage the rich analysis of data with domain knowledge but also with machine learning algorithms.

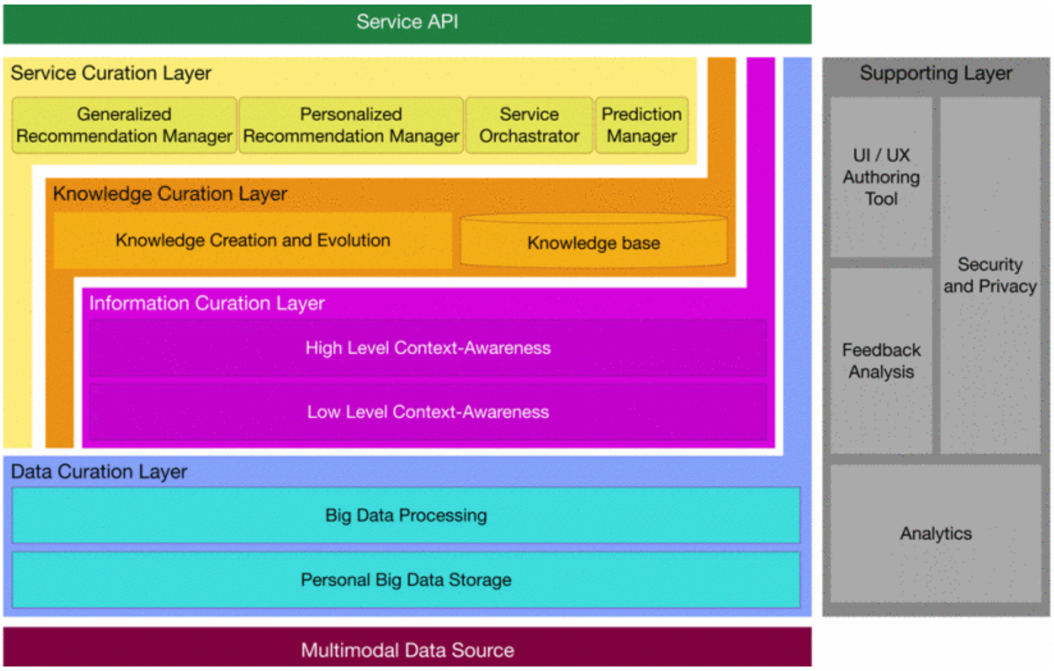


Fig. 2.3.: Mining Minds architecture (Banos, Jaehun Bang, et al., 2015).

Once introduced the concept of DSS, in order to provide a better perspective of its possibilities in mobile health, the following section presents a review of recent mobile health frameworks, systems, and their health applications.

2.1.3 Mobile Health

Mobile health (mHealth) is the practice of medicine and public health supported by mobile devices. mHealth is usually referred to as using mobile communication devices (e.g., mobile phones, smartphones, tablets) and/or wearable devices (e.g., smartwatches, activity trackers) to assist health services or interventions. mHealth applications include the use of mobile devices in collecting community and clinical health data; the delivery or sharing of information with practitioners, patients, and researchers; the real-time monitoring of patient's health; the remote assistance and provision of care; or the collaboration with health practitioners.

Mobile health has allowed a more objective and personalized approach to health, and different definitions have been raised addressing this matter. For example, the *Quantified Self* approach, by Swan (2009), aims to capture individuals' contexts and perceptions of their behaviors, health, and environments, practically in real-time.

Another definition by Estrin (2014) is the *small-data* paradigm, which also focuses on the individual, but, opposed to a so-called *big-data* approach, small-data proposes to focus on individualizing treatments based on the specifics of a single subject. This definition states that, although randomized control trials are our primary source of truth (the so-called *big-data* approach), they also deny individuality in favor of general domain knowledge. Therefore, shifting the mentality when applying those rules to individuals is necessary. A combination of personal monitoring with automatic data processing may provide improved individualization processes in healthcare (Hekler et al., 2019).

Several mHealth frameworks and systems are addressing different multiple health issues. Some of them even leverage —to some extent— these personalization capabilities with intelligent adaptation of rules and mechanisms while providing health support. These frameworks, systems, and applications are described in the following subsections.

Mobile Health Frameworks

The broad applicability of an mHealth approach to different health conditions encouraged researchers to develop and publish generic sensing mHealth frameworks. These mHealth frameworks differ from the ad-hoc mHealth sensing applications in their rich configuration capabilities, which enable the support of different conditions or diseases —instead of a single one. A systematic review by Kumar, Jeuris, Bardram, and Dragoni (2021) identified the existing mHealth frameworks up to 2018 and which health studies, application areas, and stakeholders do they target.

This review found 37 frameworks in total (28 frameworks published in scientific peer-reviewed literature, and 9 unpublished). Nevertheless, only 9 of them were classified as *end-to-end*, i.e., capable of providing support for all aspects of running an mHealth study: data collection and storage, data processing, visualization, participant recruitment, and monitoring study progress. Supplying all these functionalities enables the integration of all stakeholders in an mHealth intervention. The stakeholders are *researchers*, *developers*, and *end-users*.

The *researcher* —or study investigator— designs the mHealth study, deciding which data to collect to answer a specific research question. They are typically domain experts (e.g., physiotherapists) and inexperienced in software development. This group requires support for setting up new studies, personalized interventions, fine-tuning data sampling methods and frequencies, triggering surveys or questionnaires, requiring user consent, recruiting participants, and monitoring the progress of ongoing studies.

The *developer* implements the mHealth application over a specific framework for data collection and analysis. The use of an existing framework avoids implementing everything from scratch. Developers expect the framework to provide secure, modular, and extensible application programming interfaces (APIs) for mobile phone and server-side development.

End-users are the individuals (or patients) to whom the mHealth intervention is targeted. End-users may expect the application to work seamlessly and not cause troubles while using the mobile phone with other applications. Since they typically use their personal phones, battery drain should be avoided.

Out of the 9 end-to-end selected frameworks, only 7 covered non-functional features by ensuring extensibility, scalability, security, privacy, open, and documentation. Finally, only 4 of them were selected for analysis since (AWARE (Ferreira, Kostakos, & Dey, 2015), Beiwe (John Torous, Kiang, Lorme, & Onnela, 2016), Bridge (Sage Bionetworks, 2019), and mCerebrum (Hossain et al., 2017)) were still maintained and receiving updates.

Ferreira et al. (2015) developed AWARE as an Android-based open-source effort to develop an extensible and reusable platform for capturing, inferring, and generating context on mobile devices. AWARE provides a client-server framework that enables the collection of unobtrusive passive sensor data from smartphones. AWARE follows a modular approach with a client and a server side (Figure 2.4). First, the AWARE client app (installed in the smartphone) enables sensor data acquisition and communication with the server. Second, plugins can be added to the client to manage different sensor acquisition processes. On the server side, data are stored while providing an interface to manage the connected devices and the conduction of research studies (e.g., enrolling participants, supervising data acquisition). AWARE has been used in numerous studies, like monitoring symptom severity in cancer patients during chemotherapy (Carissa A. Low et al., 2017) or monitoring the fluctuation of affective states through experience sampling methods (Bailon et al., 2019).

John Torous et al. (2016) developed Beiwe, a research framework for transparent, customizable, and reproducible biomedical research. Beiwe features a study portal, smartphone app, database, and data modeling and analysis tools. Beiwe is supported over an Android-based app, which provides a clean interface for passing validated surveys such as the Patient Health Questionnaire 8 (PHQ-8), the General Anxiety Disorder Questionnaire 7 (GAD-7), or the Pittsburg Sleep Quality Index (PSQI).

Beiwe also supports experience sampling methods and the unobtrusive monitoring of patient context trough smartphone sensors. Beiwe also provides a rich dashboard for data visualization and analysis, and it was used to monitor schizophrenia spectrum illness (John Torous et al., 2016).

Sage Bionetworks (Seattle – Washington, USA) developed Bridge (Sage Bionetworks, 2019), a research platform to support Android or iOS smartphone-based mHealth interventions. Bridge's architecture revolves around six key components: Bridge services, iOS and Research Kit software development kits (SDKs), Android SDKs, Java REST Client, Bridge Study Manager, and Synapse. Bridge services are a set of REST-based web services that allow mobile apps to receive study configuration like surveys or task schedules, besides managing participant registration and consent while securely receiving participant data. The SDKs for iOS and Android provide open source libraries for building mHealth apps for Android- and iOS-based smartphones. The Java REST Client allows for integration with the Bridge server. The Bridge study manager presents a web interface for managing and monitoring the mHealth study. Finally, Synapse allows data scientists to carry out, track and communicate their research in real time.

Hossain et al. (2017) developed mCerebrum, an open-source mobile sensing software platform for developing and validating digital biomarkers in interventions. mCerebrum supports high-rate data collections from Android-based smartphone sensors with real-time data quality assessment. mCerebrum features a scalable storage architecture to ensure quick response despite fast-growing data volumes, a

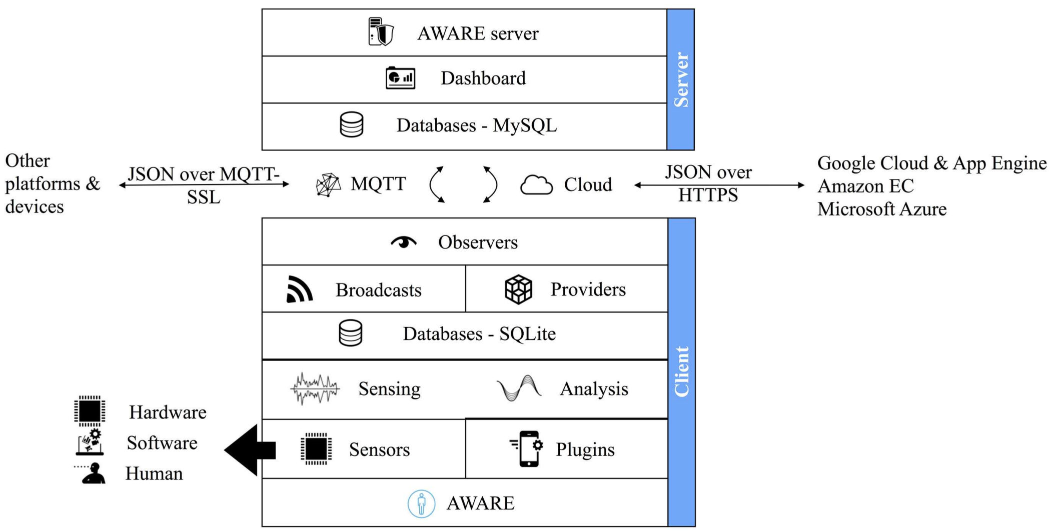


Fig. 2.4.: AWARE architecture (Ferreira, Kostakos, & Dey, 2015) .

micro-batching efficient sharing of data to enable real-time computation of multiple biomarkers without saturating CPU or memory, and a reconfigurable scheduler to manage CPU-and-memory-load- and context-aware prompts to participants. The evaluation of mCerebrum against other platforms shows that the design of mCerebrum supports higher data rates, storage throughput, and lower CPU usage than other platforms like AWARE (Ferreira et al., 2015).

Outside the systematic review by Kumar et al. (2021), other mHealth frameworks were published later than when the review was conducted, like CAMS (Bardram, 2020). Bardram (2020) developed the CARP mobile sensing framework (CAMS) as a cross-platform, reactive programming framework for digital phenotyping. CAMS aims to provide unobtrusive monitoring of sensor data from smartphones in mHealth interventions. CAMS ensures extensibility, maintainability, and adaptability, besides supporting external wearable sensors such as Bluetooth ECG monitors. The design of CAMS revolves around the concepts of reactive application programming interfaces (APIs) in order to allow non-blocking sensing and data processing through stream-based programming. CAMS was compared against platforms like AWARE (Ferreira et al., 2015) and mCerebrum (Hossain et al., 2017), outperforming them in battery consumption. Since its mobile app client is based on Flutter (Google LLC, Ireland), CARP is the only available (and active) cross-platform enabling mHealth interventions simultaneously in Android and iOS devices. CAMS has been used to develop MUBS [1] to support behavioral activation as part of treatment for depression and to develop mCardia [2] for monitoring cardiovascular diseases.

Mobile Health Systems and Applications

An mHealth intervention involves the use of mobile technologies from the patient and/or the clinical experts. In order to make an mHealth intervention successful, it is essential to meet patients' and experts' needs, provide technical support, and engage the users in the development and implementation of the tools from the beginning (Marcolino et al., 2018). Several mHealth systems in the literature address specific health issues and conditions with ad-hoc implementations, presented in the following.

Burns et al. (2011) developed the mobile app *Mobilyze!* for a successful intervention with patients with major depressive disorders. Mobilyze! enabled ecological

[1] Available in Google Play: https://play.google.com/store/apps/details?id=com.cachet.mubs01
[2] Available in Google Play: https://play.google.com/store/apps/details?id=com.cachet.reafelapp

momentary intervention and context monitoring of patients' mood, emotions, cognitive/motivational states, activities, environmental context, and social context based on at least 38 concurrent phone sensor values (e.g., global positioning system, ambient light, recent calls). A multilevel architecture enabled processing context data to feed a machine learning model enabling a behavioral activation intervention.

Cinaz, Arnrich, La Marca, and Tröster (2013) used an ECG recorder to monitor increased workload via analysis of HRV. Machine learning models (linear discriminant analysis, k-nearest neighbors, and support vector machines) classified the workload LEVEL during office work for seven subjects. Sympathetic nervous system activity was associated with increased workloads through HRV parameters like RMSSD, HF, and LF/HF ratio (HRV parameters will be detailed in subsection 2.3.1).

Banos, Moral-Munoz, et al. (2015) developed the mobile app *mDurance* to support trunk endurance assessment. mDurance used wearable inertial sensors to track patient posture and electromyography to measure the electrical activity produced by trunk muscles. This information facilitated the expert's assessment routine, reducing the impact of human errors.

Alharthi, Alharthi, Guthier, and El Saddik (2019) developed a mobile-based context-aware acute stress prediction system (CASP) to predict a user's stress status based on their current contextual data. CASP leveraged ECG signals, smartphone sensors, and machine learning models to identify the stress status

Mehrotra, Tsapeli, Hendley, and Musolesi (2017) developed *MyTraces* to investigate causality between users' emotional states and mobile phone interaction. MyTraces collected information related to phone interaction such as notifications, phone usage, application usage, and communication.

Castro, Favela, Quintana, and Perez (2015) developed *InCense* to assess functional status in older adults using mobile phones. InCense leveraged smartphone sensors (accelerometer, Wi-Fi, GPS) to gather behavioral data. They detected activity types (sedentary, light, moderate, and high) and time in each type to monitor habits and symptoms related to frailty and decreased functional status.

Freigoun et al. (2017) developed the mobile app *Just Walk* to promote physical activity in sedentary, overweight adults (Freigoun et al., 2017; Phatak et al., 2018). Just Walk was designed on the basis of system identification and control engineering principles, featuring the use of multisine signals as pseudo-random inputs for providing daily step goals and reward targets. The maximum-step goal was selected as a factor of the initial baseline level of physical activity.

Specifically for cancer, several mHealth approaches successfully targeted the promotion of healthy habits, physical activity, and exercise, present in different reviews (Dorri et al., 2020; Schaffer et al., 2019; Gomersall et al., 2019). However, most of these applications limit their mHealth approach to personalized remote communication (e.g., in-app messages, SMS) and rarely leverage sensors or activity trackers beyond improving patients' self-management. This limitation is tied to the tiny amount of research addressing how digital biomarkers (or digital phenotyping) may work for cancer care (Carissa A. Low, 2020; Fonseka & Woo, 2021). Nevertheless, we can find some examples in the literature.

Carissa A. Low et al. (2017) used AWARE (Ferreira et al., 2015) to develop a smartphone app to monitor severity symptoms during chemotherapy through context data in gastrointestinal cancer patients (Figure 2.5). They leveraged smartphone sensors and a Fitbit activity tracker to extract features reflecting mobility, activity, sleep, phone usage (e.g., duration of interaction with phone and apps), and communication (e.g., number of incoming and outgoing calls and messages). They successfully predicted symptom severity using machine learning models, namely random forests, with different combinations of the features extracted. Machine learning models allowed to build population (88.1% accuracy) and individual models (range 78.1% to 100% accuracy) and measure the importance of each biomarker monitored at predicting symptom severity.

Chung et al. (2020) used the smartphone app *WalkON* to promote exercise and reduce stress in breast cancer survivors. They monitored physical activity with daily steps measured with the smartphone and stress with the self-reported distress thermometer questionnaire. The WalkON app allowed for a community-based approach to boost patients' engagement in physical activity. In this community-based approach, all participants shared their step count with other patients.

Carissa A. Low et al. (2021) leveraged Fitbit and smartphone sensor data to estimate daily symptom burden before and after pancreatic surgery in cancer patients requiring it. They developed their app using AWARE (Ferreira et al., 2015) to collect sensor data along with self-reported ratings of daily symptoms. Day-level behavioral features reflecting mobility and activity patterns, sleep, screen time, heart rate, and communication were used to classify severity symptom levels for the next day. The machine learning models used (light gradient boosting machine) predicted high severity with 73.5% accuracy. The most relevant digital biomarkers enabling this prediction were related to physical activity bouts, sleep, heart rate, and location.

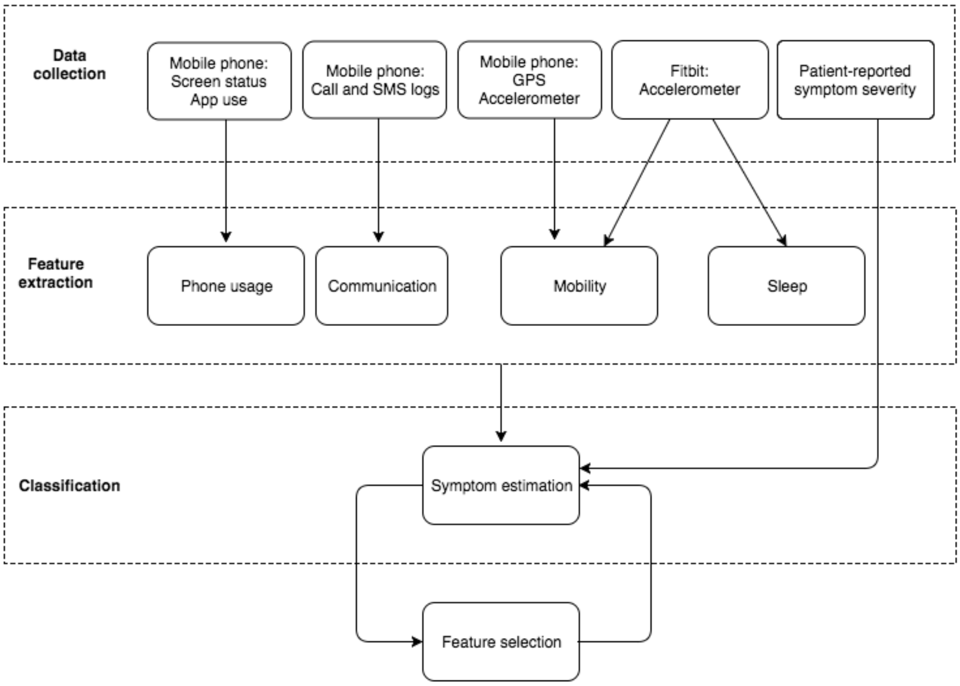


Fig. 2.5.: Data collection and analysis for the estimation of symptom severity in patients with cancer (Carissa A. Low et al., 2017).

2.1.4 Commercial Activity Trackers

An activity tracker is an electronic wearable device for monitoring fitness-related metrics like step count, distance walked or run, calorie consumption, and, in some cases, heart rate. Commercial activity trackers (CATs) are an affordable, off-the-shelf solution that combines rich monitoring capabilities with ergonomics, ease of use, and enhanced aesthetics. As opposed to research-grade activity trackers, commercial activity trackers do not ensure the validity of the measures gathered, and their use for medical purposes is not supported by the manufacturer. Figure 2.6 shows two examples of commercial and research-grade activity trackers.

Activity trackers have been used in clinical activity trials for decades. Research-grade activity trackers like ActiGraph GT3X (ActiGraph, LLC), OMRON HJ-72OITC (OMRON Corporation), or ActivPAL (PAL Technologies Ltd.), include 3-axis accelerometers, gyroscopes, and magnetometers, to accurately monitor physical activity levels, gait, and even postures. These hip-worn devices are the gold standard for activity monitoring; however, their high cost and unattractiveness for patients often limited its use to clinical trials driven by resourceful groups. Moreover, their traditional lack of real-time syncing capabilities hampered leveraging the data for continuous

(a) Fitbit Charge 5.

(b) Actigraph GT9X Link.

Fig. 2.6.: Examples of commercial (a) and research-grade (b) activity trackers.

adaptive interventions. Although nowadays some research-grade activity monitors include real-time syncing, like the wearable wrist-worn Empatica (Empatica Inc., Boston, USA), their high cost still limits its use in research.

The availability of CATs and its growing presence in the population gathered interest from the research community to improve digital health interventions (Wright, Collier, Brown, & Sandberg, 2017) since CATs may allow for continuous, scalable, unobtrusive, and ecologically valid data collection. CATs enable real-time data syncing through smartphone connections, allowing real-time monitoring and decision-making. The willingness of users to share CAT data with providers, family, and friends (Rising, Gaysynsky, Blake, Jensen, & Oh, 2021) ease its inclusion in digital health interventions and research. Moreover, using CATs as either the primary component of an intervention or as part of a broader physical activity intervention has the potential to increase physical activity participation (Brickwood et al., 2019). However, most of these commercial devices reached the market without health certificates or clinical validations, and the algorithms filtering and processing the data are private and often hidden to the public. These constraints pose reliability issues, requiring validation studies before using CATs in clinical settings (Fuller et al., 2020).

Several brands reached the market in the last decade —although not all survived the competition. Apple Inc, Fitbit, Garmin, Jawbone, Mio, Misfit, Polar, Samsung, UnderArmour, Withings, and Xiaomi were the most used brands in clinical trials since 2013, according to a recent systematic review (Fuller et al., 2020). Validity varied by study type (controlled or free-living), measurement type (step count, energy expenditure, heart rate), brand, and device. Fitbit was by far the most studied brand, offering a cost-effective solution for valid step count (Fuller et al., 2020). Due to its popularity among users and researchers, the following subsections will review the validity of Fitbit devices by measure, besides describing its use in cancer care.

Fitbit Validity by Measure

Fitbit step count. Step count is a simple yet reliable measure of overall physical activity levels (Migueles et al., 2021). The aforementioned systematic review found Fitbit devices to accurately measure step count in both laboratory and free-living settings, with a slight general underestimation close to -3% (Fuller et al., 2020). This good overall accuracy in measuring step count matches previous research (Xie et al., 2018; Straiton et al., 2018).

Another work shows how the accuracy of CATs to measure step count may be tied to the overall mobility capacities of the patient (Wong, Mentis, & Kuber, 2018). This limitation should be considered when monitoring a population with conditions that may compromise overall mobility (Ummels, Beekman, Theunissen, Braun, & Beurskens, 2018).

Energy expenditure. Although there are reasonable mean and median accuracy values for energy expenditure, Fitbit devices provide inaccurate measures with the gold standard (Fuller et al., 2020). Fitbit underestimated energy expenditure 48.4% of the time and overestimated 39.5% of the time (Fuller et al., 2020). Machine learning may bring a solution to correct the Fitbit misclassification by building classification models with energy expenditure labels from a gold standard like the Actigraph GT3X (Winfree & Dominick, 2018). However, such a solution would require the prior building of a classification model with data from patients with the same characteristics as patients in which the final digital health interventions would take place.

The aggregation of moderate and vigorous intensity levels as moderate-to-vigorous physical activity (MVPA) time, often used in physical activity guidelines, poses as a solution to the inaccuracies of Fitbit when classifying the intensity of physical activity. There is agreement on using MVPA measures as sufficiently reliable indicators (Reid et al., 2017; Redenius, Kim, & Byun, 2019) to, for instance, promote PA goals in breast cancer patients (Hartman, Nelson, Myers, et al., 2018).

Heart rate. Fitbit showed an acceptable measurement error for heart rate within the $\pm 3\%$ range in most of the studies included in the systematic review (Fuller et al., 2020). This validity has even been tested in children undergoing surgery (Pelizzo et al., 2018). Nevertheless, Fitbit may also underestimate heart rate depending on activity intensity, presumably due to movement and presence of artifacts (Thiebaud et al., 2018; Fuller et al., 2020). Due to the HR measuring changes among Fitbit devices, it is essential to report a sample description including wearable details such as device, firmware, and sensor type (Nelson et al., 2020).

Sleep. Fitbit uses HR and movement to detect sleep and classify its stages. Fitbit sleep measurements are reasonably satisfactory for general purposes, especially to measure the onset and offset of sleep, total sleep time, sleep quality, and sleep efficiency (Haghayegh, Khoshnevis, Smolensky, Diller, & Castriotta, 2019; Liang & Martell, 2018). However, measuring sleep structure (light, deep, and REM sleep) still presents moderate results and requires further research to assess its validity (Haghayegh et al., 2019).

Fitbit Use in Cancer Care

Although the validity of Fitbit measures is limited, various researchers have tried to leverage Fitbit devices differently. The most common objectives were to monitor physical activity levels, promote physical activity, and reduce sedentary behavior. Moreover, using Fitbit activity trackers may increase adherence and commitment in trial participation, hence beneficial for monitoring purposes; however, it may not be enough to promote physical activity levels.

Nyrop et al. (2018) used Fitbit Zip to monitor 127 patients with breast cancer who were instructed to walk at least 150 min/week during chemotherapy for 6 to 12 weeks. They collected analyzable data for 79% of their patients; however, only 24% attained the objective.

Hartman, Nelson, and Weiner (2018) used Fitbit One along with ActiGraph GT3X+ accelerometer in 42 patients with breast cancer and, although greater adherence to Fitbit was associated with greater increases in ActiGraph-measured MVPA, there was no general increase in physical activity for the patients adhered to wearing the Fitbit. The mean adherence was 88.13% of valid days over 12 weeks of monitoring.

Dreher et al. (2019) used Fitbit devices to promote physical activity levels in patients with breast cancer during the entire chemotherapy (6 to 9 months). Adherence was very low, with a mean number of valid days across the 9-month study period of 44.5%; besides, sustained adherence was only present in patients already used to wearable activity trackers. The long monitoring periods may have also damaged adherence.

Within the context of an exercise rehabilitation program, the role of Fitbit devices may be different. Gell et al. (2020) completed a 12-week exercise intervention in 19 female cancer survivors involving the use of tailored messages, Fitbit, and scheduled health coach interventions to support independent physical activity maintenance. Patients reported high acceptance and satisfaction with the remote monitoring of their physical activity, and communication, hence attributing physical activity maintenance to the accountability enabled by technology despite being anonymous and remote.

Few works leveraged Fitbit data as digital biomarkers (Carissa A. Low, 2020). Carissa A. Low et al. (2017) used Fitbit data (along with smartphone sensor data) to predict symptom burden in 14 patients undergoing chemotherapy for gastrointestinal cancer during 4 weeks. Several day-level features were extracted from Fitbit data reflecting behavior, physical activity, and sleep. A machine learning model (random forests) based only on Fitbit features obtained 77.6% accuracy when predicting symptom severity. Carissa A. Low et al. (2021) later used a similar approach with 44 cancer patients requiring pancreatic surgery. A machine learning model (LightGBM) predicted high severity symptom rates with 73.5% accuracy. They found that features reflecting activity patterns (e.g., number, total duration, and maximum duration of active bouts) were very relevant to that model and that they could be extracted from Fitbit data. Carissa A Low et al. (2018) also used Fitbit to predict the risk of 30-day ad 60-day readmission after cancer surgery. The possibilities posed by Fitbit data in providing relevant behavioral biomarkers have recently made it a primary source in data analysis libraries (Vega et al., 2021).

2.2 Physical Activity and Exercise in Breast Cancer Care

The overall benefits of physical activity and exercise are widely known; however, there are subtle differences between them. Physical activity (PA) consists of any movement produced by skeletal muscles involving energy expenditure, whereas exercise is a subset of PA that consists of structured and repetitive planned movements of activities. The controlled, systematic nature of exercise prescriptions allows for a therapeutic aim when targeting populations with chronic conditions or specific needs.

Psychiatric, neurological, metabolic, cardiovascular, and pulmonary diseases, along with musculoskeletal disorders and cancers, benefit from the prescription of exercise in their treatment (Pedersen & Saltin, 2015). In cancer care, physical activity and exercise pose a means to address the short- and long-term side effects related to cancer and its treatment (Ballard-Barbash et al., 2012; Cormie et al., 2017; Fong et al., 2012; Garcia & Thomson, 2014; Bluethmann, Vernon, Gabriel, Murphy, & Bartholomew, 2015; Pollán et al., 2020), and are generally recommended for prevention and treatment purposes (American Cancer Society, 2016; World Health Organization, 2021; Patel et al., 2019; McTiernan et al., 2019). Moreover, the benefits of exercise are not exclusive of the moment of cancer treatment. Exercising is beneficial before, during, and after cancer treatment, across all cancer types, and for a variety of cancer-related impairments (Stout, Baima, Swisher, Winters-Stone, & Welsh, 2017). Particularly for breast cancer care, exercise plans typically involve cardio and muscle strengthening. The prescribed exercises need to consider the impact of mastectomy (i.e., the surgical removal of one or both breasts), hence the design of specific exercises like Codman's pendulum to prevent postoperative edema of the upper limb.

Breast cancer treatment often combines surgery with radiotherapy and/or systemic therapy such as hormonal therapy, targeted therapy, or chemotherapy (Moo, Sanford, Dang, & Morrow, 2018). Unfortunately, these treatments involve high levels of toxicity and numerous short and long-term side effects for the patient, such as fatigue, heart problems, infertility, or cancer recurrence (Brinton et al., 2017). The underlying mechanisms are not yet fully understood, but accumulated evidence includes cardiotoxicity, chronic inflammation, autonomic imbalance, HPA-axis dysfunction, and/or mitochondrial damage (LaVoy, Fagundes, & Dantzer, 2016). Fortunately, exercise programs may influence such mechanisms towards reducing the burden of cancer treatment (Ginzac et al., 2019; Scott, Nilsen, Gupta, & Jones, 2018; Dias Reis et al., 2017; LaVoy et al., 2016). For breast cancer care, the evidence of physical activity and exercise interventions reducing the risk of death and recurrence is clear (Schmid & Leitzmann, 2014; Lahart et al., 2015), besides being beneficial for fatigue, depression, and sleep disturbance (Tomlinson, Diorio, Beyene, & Sung, 2014).

Exercise interventions before, during, and after breast cancer therapy are feasible and effective (K. Lee, 2021); however, more exercise is not always better. First of all, overtraining may provide drug-like effects that cause significant perturbation to the subject homeostasis (i.e., an alteration of the relatively stable state of equilibrium of the organism, maintained by self-regulating processes) (Hawley, Hargreaves, Joyner, & Zierath, 2014). Hence, the exercise needs to be adjusted (dosed) to the target patient population, so it is safe and tolerable (Jones, 2015), besides considering its correct adoption and maintenance (Garcia & Thomson, 2014). Moreover, the relationship between exercise and exercise-related benefits is not directly proportional; there is often a saturation point in which the benefits do not increase with more exercise. For instance, a review with meta-analysis (Pedisic et al., 2020) found that running participation was associated with a 23% reduced risk of cancer mortality. This significant reduction in mortality can be expected for any dose of running, even just once a week or 50 minutes a week, but there was no evidence for more reduced risk with higher amounts of running.

The possibilities of the mobile health (mHealth) and the Internet of Things (IoT) paradigms applied to cancer care motivated researchers to seek for more personalized intervention approaches. In fact, to date, mHealth exercise interventions with digital activity trackers are feasible and effective in patients with breast cancer (Schaffer et al., 2019; Dorri et al., 2020). Nevertheless, the proposed personalization mechanisms found in the literature still present several limitations.

Gell, Grover, Humble, Sexton, and Dittus (2017) aimed to promote physical activity with personalized text messages and Fitbit self-monitoring of physical activity levels in breast cancer survivors. Text messages were tailored with information from the Fitbit activity tracker, written in an informative voice, with inspirational content, or as a voice of authority, depending on the patient's preferences. The integration of Fitbit information within the messages was performed manually by the clinical experts, who checked that the physical activity records matched the exercise recommendations for the population. This intervention successfully promoted physical activity to the recommended levels; however, the personalization did not really target the physical activity levels adequate for that person —they were still just following general recommendations. The only personalized method was text-based communication, which ultimately activated the behavior-change mechanisms of the patients. Hartman, Nelson, and Weiner (2018) delivered a similar intervention and obtained similar results, but replacing the text messages with emails and scheduled telephone calls.

The importance of tailored communication with the patient is highlighted when the results of the studies above (Gell et al., 2017; Hartman, Nelson, & Weiner, 2018) are compared to others that exclusively relied on self-monitoring with wearable activity trackers. In these other studies (Dreher et al., 2019; Nyrop et al., 2018; Mendoza et al., 2017), just providing the patient with a wearable activity tracker was not associated with increased physical activity levels during cancer treatment.

Moreover, Gomersall et al. (2019) conducted an exercise rehabilitation program with 40 cancer patients and survivors. Although patients did not wear a commercial activity tracker for self-monitoring, the results showed that, compared to standard procedure, delivering personalized text messages during the trial was associated with higher physical activity levels and less sitting time. Other works prefer to rely on more sophisticated behavior change methods to boost physical activity levels, such as social cognitive theory (Duan et al., 2021; Chung et al., 2020; Maxwell-Smith et al., 2018), even combining it with systems engineering (Phatak et al., 2018).

The works mentioned above suggest a small to irrelevant effect of just giving a wearable activity tracker to a patient for physical activity or exercise promotion. The work of Uhm et al. (2017) aligns with this statement. They compared a self-monitored exercise intervention (using wearable activity trackers) against a standard scheduled intervention in patients with breast cancer (no activity trackers). The self-monitoring arm used a smartphone and a pedometer to provide information and monitor the prescribed exercises, aiming to increase participation; however, there were no significant differences in physical activity levels, physical function, and quality of life between both arms of the study.

The mechanisms for personalizing exercise and physical activity interventions in patients with breast cancer are dominated by behavior change techniques in which tailored communication is essential. Although such methods may introduce objective monitoring of physical activity levels, its sole presence when promoting exercise or physical activity levels may not be significant. Moreover, despite the heterogeneity in their designs, most of the studies mentioned adhered to the national exercise guidelines for patients with cancer (Campbell et al., 2019). The application of the same guidelines in all studies results in the delivery of the same exercise prescription for all adults, ignoring the adequate exercise needs of each patient, hence increasing the risk for wrong training adaptation due to overtraining or undertraining. The following section will introduce how a sports science approach may enable adaptive dosing during exercise intervention in patients with breast cancer.

2.3 Exercise Load Monitoring for Breast Cancer Care

Exercise guidelines for patients with breast cancer and survivors (Campbell et al., 2019) prescribe aerobic exercise alone or combined with resistance training at moderate intensity (i.e., 50% to 70% of a pre-established physiological parameter, such as age-predicted maximum heart rate). The exercise prescription must be distributed in two to five sessions per week, with 10 to 60 minutes per session, aiming at achieving 150 minutes of moderate-intensity exercise or 75 minutes of vigorous-intensity exercise a week.

Such exercise prescription ends up as a one-size-fits-all solution, making all patients receive a similar frequency, intensity, time, and type, of exercise intervention, besides setting aside factors like age, histology, or oncogenic somatic genotype (Jones et al., 2018). This application of general exercise guidelines to the individual characteristics of patients may result in their undertraining or overtraining during the intervention (Jones et al., 2018; Carter et al., 2021). To improve the exercise dosing, it is of utmost importance to consider the individual physiologic status during intervention —and not only in the beginning to gather baseline measures.

Exercise load monitoring poses a solution to tailor exercise programs for patients with cancer (Jones et al., 2018). Exercise load (or training load) monitoring is the description of the amount of exercise done by a patient. A baseline assessment of a patient's physiologic status may enable a first adaptation to the exercise programs. However, continuous follow-up, with a daily assessment of the training needs or capabilities of the patient, may enable undulating (i.e., adaptive) exercise programs (Carter et al., 2021).

Exercise load measures are typically categorized as internal or external (Bourdon et al., 2017), depending on whether such measures refer to health aspects occurring internally or externally. External training load (ETL) gathers objective, comparable measures of the exercise done such as time and distance run, number of sprints, or number of jumps (Bourdon et al., 2017). ETL measures are often recorded leveraging GPS, accelerometers, and gyroscopes embedded in activity trackers. Such measures are helpful to compare the amount of exercise performed by different trainees[3]. However, ETL does not consider the internal processes of training assimilation (Wallace, Slattery, & Coutts, 2009). Internal training load (ITL) is the individual and relative physiological and psychological stress felt by the trainees due to training and the rest of the demands in their daily lives (Bourdon et al., 2017; Halson, 2014). ITL enables monitoring training adaptation by looking into physiological biomarkers (e.g., average heart rate, heart rate variability, or saliva tests) and/or perceived self-reported status (e.g., rating of perceived exertion after training; or rating of perceived fatigue, muscle soreness, stress and sleep satisfaction before training) (Miguel et al., 2021).

[3] The literature addressing exercise load monitoring often refers to athletes or healthy people, and rarely to patients under chronic conditions. In order to adapt the nomenclature to *patients* without misleading the reader, the text will refer to *trainees* when discussing about methods tested in healthy people that may be applied to patients with cancer and survivors.

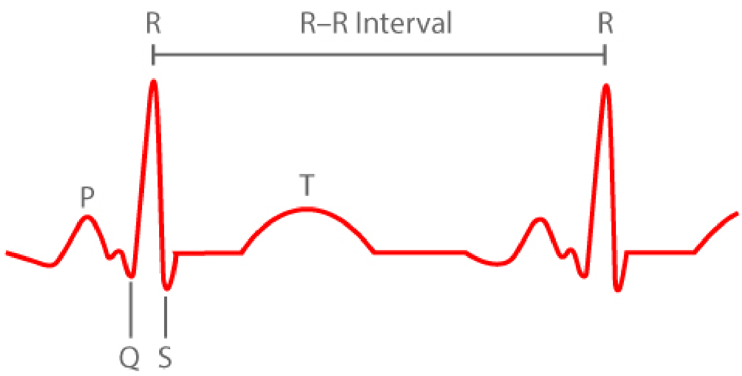


Fig. 2.7.: ECG trace for HRV extraction. HRV is composed successive R-R intervals, where R is the peak of the QRS complex found in the ECG wave (source: http://www.markwk.com/hrv-for-beginners.html).

The study of training principles and variations of ITL and ETL is mainly based on research with professional athletes (Bourdon et al., 2017; Wallace et al., 2009; Halson, 2014; Miguel et al., 2021). Little attention has been given to patients with cancer in this matter, but there is evidence on how exercise has shown improvements in HRV parameters linked to cardiac autonomic balance (Niederer et al., 2012; Dias Reis et al., 2017). Moreover, the lack of exercise is linked to worsened HRV parameters for patients with breast cancer (Caro-Moran et al., 2016), and changes in HRV have been used to determine whether the body is responding to physical exercise (i.e., as a measure in ITL) (Borresen & Lambert, 2008; Kaikkonen, 2015).

Short-term (5 min) measures of resting HRV aimed at capturing cardiac parasympathetic activity (Buchheit, 2014) is one of the most useful monitoring variables for training adaptation. However, measures of HR and HRV may not report on all aspects of wellness, fatigue, and performance, so their use is recommended with training logs and wellness questionnaires to monitor training status fully (Buchheit, 2014). The following subsections will extend the information on HRV and self-reported wellness for ITL monitoring.

2.3.1 Heart Rate Variability

Heart rate variability (HRV) is the measure of variations in time intervals between consecutive heartbeats. These intervals are measured as R-R intervals, i.e., the time between R peaks in ECG waves (Figure 2.7). HRV is controlled by the autonomic nervous system (ANS) (Shaffer & Ginsberg, 2017). The ANS is the main homeostatic regulator of our body, controlling physiologic tasks like heart rate, blood pressure, and breathing. The ANS is divided into two different components: the sympathetic and parasympathetic nervous systems (Simó, Navarro, Yuste, & Bruna, 2018).

The sympathetic nervous system prepares the body to fight or run from danger (Simó et al., 2018). When sympathetic activity is dominant, blood is shunted away from the internal organs and sent into the muscles and limbs of the body (e.g., arms, legs) to enable quick action. Sympathetic activity involves increased use of nutrients and hormones, as well as greater tissue destruction, thus resulting in a catabolic effect on the body (i.e., a breakdown), which requires subsequent recovery. Sympathetic activity is dominant while exercising or working on something that requires an increased delivery of blood to muscles, including stress.

The parasympathetic nervous system controls the digestion and elimination processes (Simó et al., 2018). An increased parasympathetic activity provokes an anabolic effect, i.e., the rebuilding and repair of the body. Parasympathetic activity also stimulates immune function while sleeping at night.

Maintaining the balance between sympathetic and parasympathetic activity is essential for the body's health, especially for patients with cancer (Simó et al., 2018). Since HRV reflects the status of both branches of the ANS, HRV is useful to assess the balance of sympathetic and parasympathetic activity (Shaffer & Ginsberg, 2017).

Overview of HRV metrics

HRV measures are classified according to their duration as long-term (24 h), short-term (5 min), or ultra-short-term (less than 5 min). There are several ways to analyze HRV. Time-domain, frequency-domain, and non-linear features can be extracted from HRV recordings; however, the availability of such features and their interpretation vary according to the duration and measurement conditions (Shaffer & Ginsberg, 2017). Since lying-down resting 5-min monitoring of HRV is one of the most reliable measures of training adaptation (Buchheit, 2014), this overview of HRV metrics focuses on features extracted in such conditions.

Time-domain features. Time-domain features of HRV measure the amount of variability in RR intervals (see Table 2.1). These values may be expressed in original units or as the natural logarithm (Ln) of original units to obtain a normal-like distribution.

The standard deviation of the time interval between successive R-R intervals (SDNN) is measured in ms. Both sympathetic activity and parasympathetic activity contribute to SDNN. However, the primary source of variation in short-term resting conditions is

Tab. 2.1.: HRV time-domain features.

Parameter	Unit	Description
SDNN	ms	Standard deviation of the time interval between successive R-R intervals.
SDSD	ms	Standard deviation of differences between succesive R-R intervals.
RMSSD	ms	Root mean square of the succesive differences of R-R intervals.
NN50	au	Number of successive R-R intervals differences greater than 50ms.
pNN50	%	Percentage of successive R-R intervals differences greater than 50ms.
NN20	au	Number of successive R-R intervals differences greater than 20ms.
pNN20	%	Percentage of successive R-R intervals differences greater than 20ms.
Mean NN	ms	Mean of R-R intervals.
Median NN	ms	Median of R-R intervals.
Range NN	ms	Difference between maximum and minimum R-R intervals.
CVSD	%	Coefficient of variation of succesive differences (RMSSD divided by Mean NN)
CVNN	%	Coefficient of variation of R-R intervals (SDNN divided by Mean NN)
Mean HR	bpm	Mean HR.
Max HR	bpm	Maximum HR.
Min HR	bpm	Minimum HR.
SD HR	bpm	Standard deviation of HR.

parasympathetic activity, especially with slow-paced breathing (Shaffer & Ginsberg, 2017).

The standard deviation of differences between successive RR intervals (SDSD) is measured in ms. Variations of SDSD are primarily mediated by parasympathetic activity and present strong correlations with HF and RMSSD (Shaffer & Ginsberg, 2017).

The root mean square of successive differences of RR intervals (RMSSD) is measured in ms. RMSSD is the primary time-domain measure to estimate parasympathetic activity, preferred over the rest of the features due to its stability and low sensitivity to breathing patterns across different measuring conditions (Shaffer, McCraty, & Zerr, 2014). Moreover, the logarithm of the RMSSD (LnRMSSD) is the gold standard to monitor training conditioning in professional athletes (Buchheit, 2014; Plews et al., 2017).

The number of successive RR-interval differences greater than 50 ms (NN50) and its percentage (pNN50) are mediated by parasympathetic activity. pNN50 may be a more reliable index than SDNN. The number of successive RR-interval differences greater than 20 ms (NN20) and its percentage (pNN20) are also mediated by parasympathetic activity (Shaffer & Ginsberg, 2017).

The mean (Mean NN), median (Median NN), and range (Range NN) of RR intervals describe where are the variations of other HRV parameters placed, and they are useful to compute other statistical parameters such as coefficient of variation and measure the presence of parasympathetic plateau (Kiviniemi et al., 2004). The coefficient of variation of the successive differences (CVSD) is the RMSSD divided by Mean NN is measured as a percentage. The CVSD is valid to assess adequate adaptations win training (Plews, Laursen, Stanley, Kilding, & Buchheit, 2013). The coefficient of variation of RR intervals (CVNN) is the SDNN divided by Mean NN. The CVNN is also helpful to measure training adaptations; however, the CVSD is preferred (Shaffer & Ginsberg, 2017).

The mean, maximum, minimum, and standard deviation of HR are measured in beats per minute. In resting short-term HR measurements, these variables represent the overall conditioning of the body and may be treated as basal conditions in the short term (Shaffer & Ginsberg, 2017). Changes in such parameters during training interventions are slow, hence making them baseline variables rather than indicators of ANS activity (Plews et al., 2013; Buchheit, 2014).

Frequency-domain features. Frequency-domain features of HRV estimate the distribution of absolute or relative power into four frequency bands (see Table 2.2). Fast Fourier Transformation (FFT) or autoregressive (AR) modeling can be used to separate HRV into the ultra-low frequency (ULF, ≤ 0.003 Hz), very-low frequency (VLF, 0.003 – 0.04 Hz), low frequency (LF, 0.04 – 0.15 Hz), and high frequency (HF, 0.15 – 0.40 Hz) bands. Higher power in higher frequencies is normally associated with parasympathetic activity, whereas higher power in lower frequencies with sympathetic activity —although measurement conditions may alter this interpretation.

ULF (≤ 0.003 Hz) requires at least 24 h recordings to be analyzed, and there is disagreement about the degree of contribution of parasympathetic and sympathetic activity to this band (Shaffer & Ginsberg, 2017).

VLF (0.003 – 0.04 Hz) requires at least 5 min of recording. VLF power may be fundamental to health. Low power on this band is associated with adverse outcomes such as all-cause mortality and high inflammation. This band appears to be regulated

Tab. 2.2.: HRV frequency-domain features.

Parameter	Unit	Description
TP	ms^2	Total power of the spectral density.
ULF	ms^2	Absolute power of the ultra low frequency band (≤ 0.003 Hz).
VLF	ms^2	Absolute power of the very low frequency band (0.003 – 0.04 Hz).
LF	ms^2	Absolute power of the low frequency band (0.04 – 0.15 Hz).
HF	ms^2	Absolute power of the high frequency band (0.15 – 0.40 Hz).
LF/HF	%	Ratio of LF and HF power.
LF nu	nu	Normalized LF power (HF divided by the sum of LF and HF).
HF nu	nu	Normalized HF power (HF divided by the sum of LF and HF).

by the heart's intrinsic nervous system and sympathetic activity (Shaffer & Ginsberg, 2017).

LF (0.04 – 0.15 Hz) requires at least 2 min of recording. LF power may be influenced by parasympathetic and sympathetic activity. However, in resting conditions, LF power mainly reflects baroreflex activity (Shaffer & Ginsberg, 2017). Its normalized version (LF nu) divides LF by the sum of HF and LF.

HF (0.15 – 0.40 Hz) requires at least 1 min of recording. HF mainly reflects parasympathetic activity. Lower HF power is correlated with stress and anxiety. Breathing rates may influence HF, causing a misrepresentation of vagal tone; however, under controlled conditions at stable breathing rates, the natural logarithm of HF (LnHF) can be used to estimate vagal tone (Shaffer & Ginsberg, 2017). Its normalized version (HF nu) divides LF by the sum of HF and LF.

The LF/HF ratio may estimate the ratio between sympathetic and parasympathetic activity under resting lying-down conditions. A high LF/HF ratio indicates sympathetic dominance (Shaffer & Ginsberg, 2017).

Total power (TP) is the sum of the energy in the four bands for 24h recordings and VLC, LF, and HF for short-term recordings (Shaffer & Ginsberg, 2017).

Non-linear features. Non-linear features of HRV describe the unpredictability of a time series (see Table 2.3). For RR analysis, a Poincaré plot is a graph of RR(n) on the x-axis versus RR(n + 1) on the y-axis. In other words, the RR intervals plot against the immediate following RR intervals. The purpose is to measure non-linear estimations. All the dots are fitted into an ellipse, for which two parameters are required: SD1 and SD2. SD1 is the standard deviation of the projection of the Poincaré plot on the line perpendicular to the line of identity, SD2 is the standard

Tab. 2.3.: HRV non-linear features.

Parameter	Unit	Description
SD1	ms	Standard deviation of the projection of the Poincaré plot on the line perpendicular to the line of identity.
SD2	ms	Standard deviation of the projection of the Poincaré plot on the line of identity.
SD2/SD1	%	Ratio between SD2 and SD1

deviation of the projection of the Poincaré plot on the line of identity. SD1 is identical to RMSSD, and SD2 correlates with LF. The ratio SD1/SD2 requires a minimum of 5 min recordings, measures unpredictability of the series, and may be used to assess balance between sympathetic and parasympathetic activity (Shaffer & Ginsberg, 2017).

Monitoring Training Adaptation with HRV

Measures of HRV have enabled the monitoring of training adaptation in professional athletes (Buchheit, 2014; Plews et al., 2013). First, autonomic balance (equilibrium between sympathetic and parasympathetic nervous activity) is tightly bound to the workload/recovery processes of training (Kluess, Wood, & Welsch, 2000); and second, HRV is capable of reflecting changes in autonomic balance (Borresen & Lambert, 2008; Buchheit, 2014; Kaikkonen, 2015).

Daily monitoring of resting (supine) short-term HRV is one of the most reliable measures of parasympathetic activity, and it can track acute and chronic responses to training adaptation (Buchheit, 2014; Plews et al., 2013). The analysis of LnRMSSD [4] variability is the gold standard for monitoring training adaptation, and several reasons support it (Buchheit, 2014). First, LnRMSSD may be captured over minimal periods of time —typically 5 min, but it can be reduced to 1 min measurements under certain conditions (Forner-Llacer et al., 2020)— and remain comparable to longer recordings. Second, the sensitivity of LnRMSSD to breathing patterns is very low. Third, the day-to-day variations of time-domain variables (e.g., LnRMSSD, SDNN) are likely lower than frequency-domain variables (e.g., LF/HF). Fourth, the probability of finding outliers and ectopic beats is scarce in resting conditions, yet RMSSD supports degree zero (mean), linear, polynomial, and cubic spline interpolation without being affected (Giles & Draper, 2018).

[4] The analysis of LnRMSSD over RMSSD is preferred due to its similarity with a normal distribution.

In order to adjust training loads, it is crucial to determine the smallest worthwhile change (SWC) of a health or performance measure of the trainee (Will G Hopkins, 2004). The idea behind SWC is that we can measure a minimal change in relevant variables to the athlete status or performance, like LnRMSSD, to assess the impact of training in athletes (Buchheit, 2014). It is known that, during training adaptation, LnRMSSD may decrease the day after high-intensity training. Thus, it is essential to assess how significant changes are and the minimal change enabling the adjustment of training intensities. However, since many factors may interfere with HRV, the interpretation of such changes is not trivial (Buchheit, 2014). There are different approaches in the literature describing different SWC calculations, all of them aiming for within-trainee individual variations of the LnRMSSD. These approaches are described in the following.

Determining SWC with the coefficient of variation of the LnRMSSD. Some researchers rely on the analysis of the coefficient of variation (CV) of the LnRMSSD (Plews et al., 2013; Plews, Laursen, Kilding, & Buchheit, 2012; Buchheit, 2014). The coefficient of variation of a variable results from dividing the standard deviation of this variable during a time window by the mean computed for the same time window. Equation 2.1 describes CV for LnRMSSD.

$$CV_{LnRMSSD} = \frac{LnRMSSD_{std}}{LnRMSSD_{mean}} \tag{2.1}$$

Plews et al. (2013) and Plews et al. (2012) used a 7-day rolling window to compute the CV of LnRMSSD and track if there was a change greater than 0.5 on it (Will G Hopkins, 2004). In general, a lower CV of LnRMSSD indicates a reduced perturbation in homeostasis, and therefore trainees showing increased CV may benefit from a reduced training load.

Plews et al. (2014) studied how changes in CV of the LnRMSSD correlate with performance changes for professional and recreational runners. They found that, for recreational runners, a minimum of 5 measures in a week are required to establish solid LnRMSSD baselines against performance, and, for professional triathletes, it is a minimum of 3 measures.

Determining SWC with the standard deviation of the LnRMSSD. Other researchers (Vesterinen et al., 2016; Javaloyes, Sarabia, Lamberts, & Moya-Ramon, 2019) and commercial applications (HRV4Training) defined the SWC of the LnRMSSD for a determined rolling time window (e.g., 30 days) as a factor (f) of its standard

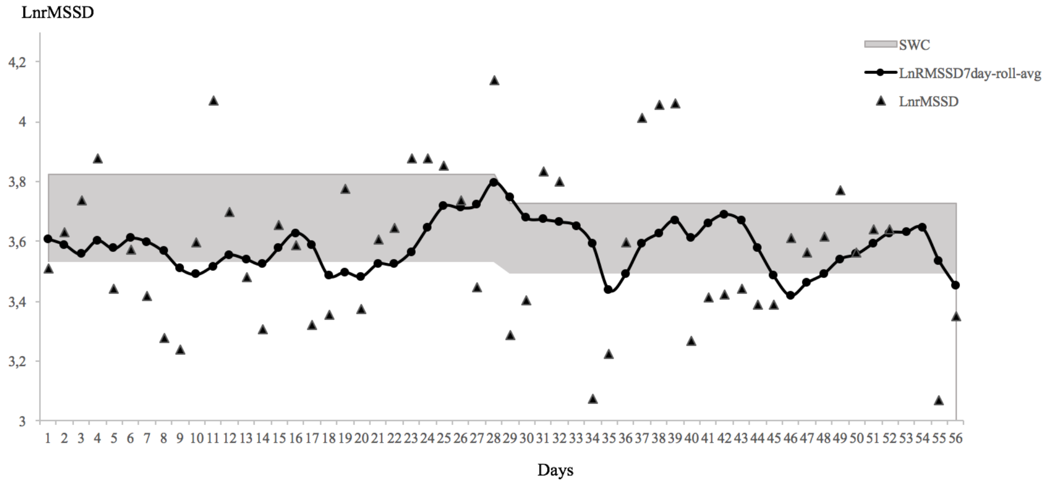


Fig. 2.8.: Example of individual response of HRV in a cyclist during training conditioning. SWC range thresholds are calculated out of 4-weeks baseline monitoring. If $LnRMSSD_{7day-roll-avg}$ fell outside SWC, training levels were readjusted (Javaloyes, Sarabia, Lamberts, & Moya-Ramon, 2019).

deviations ($LnRMSSD_{rolling\_std}$) around its mean ($LnRMSSD_{rolling\_mean}$). SWC thresholds are defined in Equation 2.2.

$$SWC = LnRMSSD_{rolling\_mean} \pm f \cdot LnRMSSD_{rolling\_std} \tag{2.2}$$

Then, a smaller window (e.g., 7 days) is used to compute the mean measure of LnRMSSD. This averaged measure ($LnRMSSD_{7days\_rolling\_mean}$) is checked to be inside the SWC to ensure adequate training.

The idea behind this SWC method is to monitor changes of the LnRMSSD outside the normality of its distribution in the previous days. Thus, if changes outside that normality curve are found, it may imply a perturbation in homeostasis, and therefore trainees would benefit from a reduced training load.

Vesterinen et al. (2016) used a 4-weeks rolling window of the RMSSD with a factor $f = 0.5$ to compute SWC. Javaloyes et al. (2019) used the averaged 30-days rolling window of the LnRMSSD with a factor $f = 0.5$ to compute SWC (Figure 2.8). The scientifically validated smartphone app HRV4Training[5] uses a 30-days rolling window and a factor f that may be adjusted to 0.5 or 1 depending on user preferences.

[5]HRV4Training: https://www.hrv4training.com/blog/the-big-picture

Role of other HRV parameters in training conditioning. The LF/HF was long considered a representation of sympathetic and parasympathetic activity balance; however, this view has been highly criticized. Among the most critical aspects is the loose relationship between LF power and sympathetic nerve activation, besides the non-linear and non-reciprocal relationship between sympathetic and parasympathetic activity (Billman, 2013). These limitations of LF and LF/HF encouraged researchers to shift their evaluations of parasympathetic activity to other parameters like the RMSSD and HF alone (Laborde, Mosley, & Thayer, 2017).

Other parameters like the SDNN, VLF, LF, HF, LF/HF ratio, SD1, and SD2, although they may be affected by respiration rate (Shaffer & Ginsberg, 2017), can be measured with reliability under resting (lying) conditions with the same outlier and ectopic beat treatment as RMSSD (Giles & Draper, 2018).

2.3.2 Self-Reported Wellness

HRV is useful to assess training needs; however, HRV measurements may be influenced by other factors such as physiological and genetic conditions, diseases, lifestyle habits, and even external factors (Shaffer et al., 2014; Sammito & Böckelmann, 2016). The impact of stress (Kim, Cheon, Bai, Lee, & Koo, 2018), sleep (Sajjadieh et al., 2020), and fatigue (Tran, Wijesuriya, Tarvainen, Karjalainen, & Craig, 2009) are the modulating factors of HRV getting more attention from the research community (Ltd., 2014; Plews et al., 2012).

High stress and elevated anxiety are frequently associated with low parasympathetic activity. Decreases in HF and RMSSD or increases in LF may characterize this association (Sajjadieh et al., 2020). High fatigue in healthy individuals may be associated with increased sympathetic arousal (Tran et al., 2009), hence associating it with higher LF/HF ratios. Similar results can be found in patients with breast cancer, focusing on parameters like HF and SDNN (Y.-H. Lee et al., 2018). Poor sleep quality is also associated with autonomic imbalance and worsened HRV parameters, like reduced SDNN and RMSSD (Sajjadieh et al., 2020).

Evaluating these stress-inducing factors is essential to assess recovery. Recovery periods are regularly needed to replenish the body physiologically and psychologically, but they only occur when physiological arousal diminishes, and parasympathetic activity dominates over sympathetic activity (Ltd., 2014).

Self-reported perceived wellness questionnaires like the Hooper Index (Hooper & Mackinnon, 1995) were defined to measure ITL by taking into account several factors like stress, fatigue, sleep quality, and muscle soreness. The Hooper Index has been associated with HRV; however, it has not been found significant enough to enable training intensity adjustment (Rabbani, Clemente, Kargarfard, & Chamari, 2019). Self-perceived wellness may influence performance; however, due to its subjective nature, it cannot provide enough reliability in assessing training needs. Therefore, the use of perceived wellness may be used for complementing the interpretation of physiological markers like the LnRMSSD (Buchheit, 2014).

The variables forming the Hooper Index have been studied separately in athletes. Stress may be associated with overtraining and underperformance (Gleeson, 2002; Aguilar Cordero et al., 2014). Sufficient sleep time and high sleep quality are associated with adequate recovery from training sessions (Carney et al., 2012; Beck, Schwartz, Towsley, Dudley, & Barsevick, 2004). High fatigue —which may be present in patients with cancer during and/or after systemic treatment (Cantarero-Villanueva et al., 2014)— is related to overtraining in athletes (Roldán-Jiménez, Bennett, & Cuesta-Vargas, 2015).

Plews et al. (2012) monitored Hooper Index and HRV in elite triathletes during training preparation for competition. They found that sleep quality presented consistent decline during non-functional over-reaching (i.e., being worse than the previous week), an already known relation with overtraining (Hooper & Mackinnon, 1995). Nevertheless, the rest of the factors monitored (stress, fatigue, muscle soreness) could not identify the manifestation of non-functional over-reaching or poor performance. Therefore, such subjective measures should be subject to the experts' interpretation and according to the physiological status reflected by HRV.

For patients with cancer, there are already successful experiences measuring the factors composing the Hooper Index (fatigue, muscle soreness, stress, and sleep quality) in remote environments with patients with cancer (Cantarero-Villanueva et al., 2014; Lozano-Lozano et al., 2018; Børøsund et al., 2020; Børøsund et al., 2018; Min et al., 2014), although they have been mainly related to factors like quality of life.

3

ATOPE+: An mHealth System to Support Personalized Therapeutic Exercise Interventions in Patients with Cancer

Publication Details

Reference: Moreno-Gutierrez, S., Postigo-Martin, P., Damas, M., Pomares, H., Banos, O., Arroyo-Morales, M., & Cantarero-Villanueva, I. (2021). ATOPE+: An mHealth System to Support Personalized Therapeutic Exercise Interventions in Patients With Cancer. IEEE Access, 9, 16878-16898. https://doi.org/10.1109/ACCESS.2021.3049398

Journal: IEEE Access

JCR Impact Factor 2020: 3.367 (2.454 w/o self-citations);

Rank & Category: 65/161 (Q2) in Computer Science Information Systems; 94/273 (Q2) in Engineering, Electrical and Electronic.

Preceding Note

This chapter presents an extension of the work published in the work described above. Specifically, the following sections include a more detailed analysis of the requirements (subsection 3.3.1), an improved description the ATOPE+ app implementation (section 3.3.3), an extension of usability results with patients with breast cancer (section 3.4.1), and a discussion addressing these new results (section 3.5).

3.1 Introduction

Therapeutic exercise (TE) poses a means to address the short- and long-term side effects of cancer and its treatment (Ballard-Barbash et al., 2012; Schmid & Leitzmann, 2014). TE and PA had consistently reported benefits to patients with cancer (Garcia & Thomson, 2014) and they are generally recommended for both prevention and treatment purposes (American Cancer Society, 2016; World Health Organization, 2021; Patel et al., 2019; McTiernan et al., 2019). Combined with medical and surgical treatments, TE improves quality of life (Lahart et al., 2015; Cormie et al., 2017; Peterson & Ligibel, 2018), reduces mortality risks, and reduces recurrence (Pollán et al., 2020). This has driven the research community to seek after new means to deliver TE interventions in remote environments by leveraging mobile technologies (Schaffer et al., 2019). In fact, to date, mHealth PA interventions are considered a feasible, cost-effective approach to improve overall activity levels, body composition, quality of life and self-reported symptoms in patients with cancer (Schaffer et al., 2019) and survivors (Roberts et al., 2017).

The personalization of TE interventions still presents a challenge. Personalizing (or tailoring) a TE intervention consists in fitting it to the needs, characteristics or possibilities of each patient with an adapted and evidence-based prescription following for frequency, intensity, time, and type (Campbell et al., 2019). Personalization strategies in remote TE interventions are often overlooked. Most of recent exercise interventions in patients with cancer adhered to the national exercise guidelines (Campbell et al., 2019), resulting in the the delivery of the same amount of exercise prescription for every patient. The common method to personalized a TE intervention is relying on patients' self-management (Campbell et al., 2019). This gap opens up opportunities to introduce mobile technologies to monitor objective, comparable, and quantifiable data about each patient's health and performance during the intervention process (Kelly & Shahrokni, 2016; Beg et al., 2017).

Internal training load (ITL) is the individual and relative physiological and psychological stress felt by the trainees due to training and the rest of the demands in their daily lives (Bourdon et al., 2017; Halson, 2014). Measures of ITL are widely used and established in sports to assess training conditioning (Miguel et al., 2021). Therefore, ITL measure may enable the monitoring of patient's training adaptation during TE. This approach may provide a way to optimize exercise prescription in an undulating manner (i.e., adaptive, flexible) to avoid the negative effects of undertraining or overtraining (Jones et al., 2018).

Heart rate variability (HRV) may provide reliable measures of ITL due to its relationship with reflecting autonomic balance (Shaffer & Ginsberg, 2017). Autonomic balance (i.e., the balance of sympathetic and parasympathetic nervous activity) plays a key role in the workload-recovery ratios of training (Kluess et al., 2000). This has made HRV a reliable tool to estimate ITL and monitor training adaptation (Buchheit, 2014, 2014; Kaikkonen, 2015). Nevertheless, HRV measurements typically require lab equipment like a Holter monitor to measure ECG (electrocardiography), hence any personalization process often becomes tedious and expensive. Fortunately, the remote monitoring of HRV has been consistently validated with different wearable devices and types of participants (Caminal et al., 2018; Hernando, Roca, Sancho, Alesanco, & Bailón, 2018; Perrotta, Jeklin, Hives, Meanwell, & Warburton, 2017), thus enabling its use in TE interventions. There are studies using HRV in patients with cancer (Dias Reis et al., 2017; Ha, Malhotra, Ries, O'Neal, & Fuster, 2019); however, and to the best of our knowledge, these works used HRV as an impact measure of a TE intervention (baseline and follow-up measures), and never as an ITL measure to adjust training prescription.

Other factors modulating HRV need to be taken into account. The impact of stress (Kim et al., 2018), sleep (Sajjadieh et al., 2020), and fatigue (Tran et al., 2009) are the modulating factors of HRV getting more attention from the research community (Ltd., 2014; Plews et al., 2012). There are already successful alternative previous experiences when measuring self-perceived wellness in patients with cancer in remote environments (Lozano-Lozano et al., 2018; Min et al., 2014; Børøsund et al., 2018; Børøsund et al., 2020). Nonetheless, there are no systems integrating HRV and self-reported measures adapted to patients with breast cancer.

In light of these opportunities, this chapter presents ATOPE+, an mHealth system to support personalized therapeutic exercise interventions in patients with cancer. ATOPE+ represents the technological drive of the ATOPE trial (ClinicalTrials.gov, NCT03787966; Postigo-Martin et al., 2021) by enabling the remote assessment of training load in patients with cancer and recommending optimal exercise dosage by means of a knowledge-based system. With the automatic generation of personalized training prescriptions, ATOPE+ is capable of providing undulating nonlinear exercise prescription, minimizing the risk of undertraining and overtraining throughout the TE intervention. To our knowledge, ATOPE+ is the first mHealth system combining measures of exercise load (HRV), modulating factors of HRV (recovery, sleep, distress, fatigue), and daily and training-specific physical activity levels (Fitbit activity tracker) to personalize TE interventions in patients with cancer. The contributions of this chapter are the following:

1. A novel approach to the personalization of TE interventions in patients with cancer using physiological variables related to training load in a remote context.

2. A novel mHealth architecture, and a description of its implementation, supporting the requirements of a TE intervention in patients with breast cancer, consisting of:

   - Heterogeneous physiological data collection: Bluetooth HRV for exercise load, in-app questionnaires for the modulating factors of HRV, and the Fitbit cloud for daily and in-training physical activity levels.

   - A multilevel architecture to transform physiological data into useful knowledge: data, information, and knowledge management layers.

   - An intelligent knowledge-based system to support the automatic generation of personalized training prescriptions.

3. A usability evaluation of ATOPE+ with experts (physical therapists with TE experience) and breast cancer patients using the Systems Usability Scale (SUS) and a semi-structured interview with the experts.

Overall, ATOPE+ allows clinical experts to simplify knowledge management and the decision-making process within the context of a TE intervention with the automatic pairing of data and diagnosing rules assessing the individual exercise needs of patients.

The rest of the chapter is structured as follows. Section 3.2 gathers related work to mHealth systems in general and applied to cancer. Section 3.3 describes ATOPE+ in its entirety: requirements, architecture, implementation, and use. Usability results with experts, patients with cancer, and survivors are presented in Section 3.4. Section 3.5 discusses the findings system and usability results. Section 3.6 closes the chapter with final conclusions and remarks.

3.2 Materials and Methods

3.2.1 Considerations for Design

The requirements of digital health systems are well-discussed in the literature, ranging from pure technical aspects (Banos, Villalonga, et al., 2015; Banos et al., 2016) to security concerns (O'Connor, Rowan, Lynch, & Heavin, 2017), or addressing what is necessary to deliver a successful intervention (Granja et al., 2018). The requirements of ATOPE+ thrive on them, but more specifically on the need to deliver personalized TE intervention in patients with breast cancer. The definition of these requirements was conducted through several meetings among the computer scientists, engineers and physical therapy professionals co-authoring the work referred in the beginning of the chapter. The WHO (World Health Organization, 2018) taxonomy of digital health interventions (described in Section 2.1.1) helped to ensure good communication among the different professionals involved in the development of a ATOPE+ by stating the health types of system challenges, system categories, and digital health interventions supported by ATOPE+. The European General Data Protection Regulation (GDPR) is also included in the requirements from design.

The mHealth systems and frameworks described in Section 2.1.3 inspired the design choices and requirements for ATOPE+. Specifically, the multilevel architecture described in works like Mining Minds (Banos et al., 2016) was considered to leverage data transformations into useful knowledge. Moreover, the combination of such architecture with an inference engine (e.g., a base of rules) was set as the basis for supporting personalized recommendations.

The different frameworks found were analyzed in depth (Kumar et al., 2021); however, none of the frameworks available supported our most relevant requirements fully. The first of these prior requirements was simplicity-from-design. Frameworks like AWARE (Ferreira et al., 2015) integrate context-aware measures by leveraging smartphone sensors. This feature would have introduced a significant overhead for the main monitoring target of the system, ITL. ITL can be measured integrating HRV and self-reported wellness (as described in Section 2.3), and these measures only depend on wearable Bluetooth ECG device and in-app questionnaires. The second prior requirement was enabling cross-platform deployment of the app in Android and iOS. Technical limitations had to be close to zero when targeting a restricted population such as patients with breast cancer. CAMS (Bardram, 2020) would have been an ideal choice, since it provides flexibility on designing apps and enables cross-platform; however, CAMS was not available at the time of developing ATOPE+. For the previous reasons, ATOPE+ was developed from scratch using Flutter.

Flutter (Google LLC, Ireland) is a framework that enables the fast development of natively compiled, cross-platform applications. Flutter ensures full native performance on both iOS and Android operating systems since it compiles Flutter code into native ARM machine code using Dart's compilers. Flutter has an extensive community supporting different features through a plenty of packages[1].

3.2.2 Usability Evaluation

Usability was evaluated using the Systems Usability Scale (SUS) (Brooke, 1996) and conducting a semi-structured interview (Adams, 2015). The purpose of using these two methods was to provide a comprehensive vision of the usability of ATOPE+ with an objective and comparable result (SUS) and a less constrained and more descriptive evaluation (semi-structured interview). The usability of the app and the web dashboard were addressed separately by the experts. In addition, patients with breast cancer and survivors evaluated the usability of the app alone.

The SUS scale is a ten-item Likert scale that gives a global view of subjective assessments of usability. Each item of the scale is scored from 1, *strongly disagree*, to 5, *strongly agree*, and the total SUS score is computed out of them, ultimately ranging from 0 to 100. The SUS is easy to administer, performs reliably on small sample sizes, and can effectively differentiate between usable and unusable systems. The SUS allows for usability comparison among systems in research and industry (ISO 9241-11). Sixty-eight points represent the minimum score to ensure good usability (Sauro, 2011).

SUS Questions

The SUS questions are as follows.

Q1. I think that I would like to use this system frequently.

Q2. I found the system unnecessarily complex.

Q3. I thought the system was easy to use.

Q4. I think that I would need the support of a technical person to be able to use this system.

Q5. I found the various functions in this system were well integrated.

Q6. I thought there was too much inconsistency in this system.

Q7. I would imagine that most people would learn to use this system very quickly.

[1] Dart packages for Flutter are available at https://pub.dev/

Q8. I found the system very cumbersome to use.

Q9. I felt very confident using the system.

Q10. I needed to learn a lot of things before I could get going with this system.

Semi-Structured Interview

The items conforming the semi-structured interview for experts addressed the smartphone app and dashboard of ATOPE+, and they are presented in the following.

I1. What were your general sensations using the ATOPE+ smartphone app?

I2. How do you consider ATOPE+ may help compared to traditional treatment?

I3. Do you think the measurement protocol is complex?

I4. Do you think the tutorials are clear enough?

I5. Do you think that patients may be enrolled in the daily use of the app?

I6. Is there anything from the ATOPE+ app that draw your attention? For good or bad.

I7. What is your opinion about patients following the app autonomously, using an app and wearable sensors in remote?

I8. Do you think this methodology could be applied to other populations, for instance, patients with other conditions)?

I9. Would you add extra functionalities to improve user experience?

I10. Do you think ATOPE+ could be used during restricted situations like the COVID-19 pandemic?

I11. What were your general sensations using the ATOPE+ dashboard?

I12. Is there anything from the ATOPE+ dashboard that draw your attention? For good or bad.

I13. Do you think the ATOPE+ dashboard may be helpful for the follow-up of patients and quick decision-making?

I14. What other features would you like to see in the dashboard?

3.3 ATOPE+

3.3.1 Requirements

Patients must follow the data gathering protocol under similar conditions every day. Patients should record their HRV in the morning right after waking up and emptying their bladder. HRV must be recorded in a lying position (Javaloyes et al., 2019), and a minimum of 5 minutes are required for analysis (Niederer et al., 2012). To establish a reliable baseline for HRV comparison, a minimum of 5 measures are required in the previous 7 days (Plews et al., 2014). Next, patients must record their perceived recovery status, distress, sleep quality, and fatigue using questionnaires. Patients must also use a wearable activity tracker to collect their overall and training-specific physical activity levels. These data will be used in post-intervention analysis to differentiate patients depending on the fulfillment of general physical activity guidelines (Dias Reis et al., 2017), and the level of agreement between the training intensity performed and the one indicated in the personalized exercise prescriptions.

First, to automatically generate the personalized exercise prescriptions according to expert knowledge, a knowledge-based system is required. The base of rules will determine the frequency, intensity, time and type of the exercise prescription depending on every day patient's data. Nevertheless, some rules may not always apply, so ATOPE+ must provide expert tools to be able to check and change the exercise prescription according to the expert's criteria on how patient's health is evolving. Two interfaces must be available to address the needs of patients and clinical personnel separately and interact with the knowledge-based system.

Patients must have access to an smartphone app to gather their data, interact with the experts, and receive the personalized exercise prescriptions. The smartphone app must allow connection to external devices, such as Bluetooth ECG, and to ask for recovery, distress, sleep quality and fatigue perceptions through in-app questionnaires. Besides, the app must collect and process the minimum amount of data required for the intervention trial, thus meeting the GDPR minimization principle. Patients should be able to follow the data gathering protocol every day, thus the smartphone app must provide a very clear and intuitive flow through it. The number of wearable devices used must be as reduced as possible, as well as the number of questions asked, so that the protocol complexity and amount of time needed to follow it are minimized.

Clinical experts must be able to check patients' data and modify their exercise recommendations. A web interface must provide meaningful visual display of data related to the workload-recovery ratio of every patient along with the exercise prescriptions. The same web interface must provide means for checking and modifying the personalized exercise prescriptions.

ATOPE+ must be able to manage the heterogeneous data sources noted before: Bluetooth ECG, in-app questionnaires, and commercial activity tracker. Moreover, ATOPE+ should be able to transform the raw data into useful information, that is, the personalized exercise prescriptions. This collection of data must be as unobtrusive as possible for the patients to facilitate their engagement in the intervention.

Since ATOPE+ is to be used in a context of a randomized trial with multiple patients at the same time, it must be able to deal with high data volumes and the structured, semi-structured, or unstructured nature of the collected data. Consequently, data must be stored and processed efficiently to provide agile and efficient responsiveness.

Data persistence must be carefully managed to avoid data loss in likely deviations from the ideal scenario, like no internet connection or Bluetooth ECG disconnection. Therefore, data must be stored locally in the patients' smartphone before being sent to the cloud or server.

Data reliability must be ensured. Some scenarios might be prone to error, specially those regarding HRV measurement, such as ECG misplacement or ECG recording disruption by external events (e.g., a loud noise, a flash light or a phone call). ECG misplacement may be avoided with training and displaying in-app reminders on how to use the ECG device. Disruption risk may be minimized by lowering notification volume levels during the recording. Last, to ensure HRV reliably, HRV signals must be filtered by detecting, removing and interpolating outliers and ectopic beats (Peltola, 2012; Giles & Draper, 2018). Patients should be given the choice to record their HRV again voluntarily if they considered the recording conditions were not to be ideal, or if the automatic HRV processing rejects the validity of the measures.

The vast amount of data generated may help to assess the validity and pertinence of the training plans assigned to each patient. Thus, this data can be used to refine the existing expert-based rules or even create new ones. On the one hand, unsupervised learning algorithms may reveal these unforeseen relationships among the participants and their recovery process using clustering, anomaly detection or rule generation algorithms. On the other hand, supervised learning, specifically classification models combined with feature selection, may help to highlight the most relevant features for the recovery of patients. Building prediction models to assess the recovery of the patients may also help experts in deciding the best exercise prescriptions for patients when comparing best-case vs worst-case scenarios. Consequently, ATOPE+ must be able to implement intelligent automatic data-driven analysis and provide means to introduce new rules commanding the recommendations.

Finally, it is of utmost importance to ensure the security and privacy of the data. All online communications must be secured and encrypted with available standards. Access to the ATOPE+ centralized server must be protected through firewall. All the data within the system pseudoanonymized and encrypted. The risk for malicious data usage is increasing as sensitive data-driven systems like ATOPE+ emerge. Fortunately, regulatory and legal policies are already taking this into account such as the European GDPR, which is of mandatory application for our system.

According to this requirement analysis and the taxonomy for digital health systems published in (World Health Organization, 2018), the following types of *health system challenges*, *system categories*, and *digital health interventions* were identified. This classification helped during the reassessment of requirements in the different iterations of the process.

Health system challenges. (1) Information: lack of quality/reliable data, lack of access to information or data, and insufficient utilization of data and information; (3) Quality: poor patient experience, and inadequate supportive supervision; and (6) Efficiency: Inadequate workflow management. Challenges related to (1) Information and (3) Quality address the need for data-driven therapeutic exercise interventions that really assess the individual needs of patients. Challenge related to (6) Efficiency address the need for automatic integration of data with expert knowledge, besides the provision of appropriate tools for patients (smartphone app) and experts (web interface).

System categories. (Q) Knowledge management system, and (Y) Telemedicine. ATOPE+ is belongs to (Q) Knowledge management systems due to the incorporation of expert's knowledge to automatically diagnose exercise needs according to the data collected (knowledge-based system). ATOPE+ also belongs to a (Y) Telemedicine system (more specifically to an mHealth approach) due to the remote assessment of training needs with a smartphone app and wearable sensors.

Digital health intervention. (1.0) Clients: (1.1) Targeted client communication (transmit diagnostics results, or availability of result), and (1.4) Personal health tracking (self monitoring of health or diagnostic data by client). (2.0) Healthcare Providers: (2.3) Healthcare provider decision support (scree clients by risk or other health status), and (2.4) Telemedicine (remote monitoring of client health or diagnostic data by healthcare provider, transmission of medical data to healthcare provider). The type of digital health intervention according to the (1.0) Clients or patients feature the individual capabilities of self-monitoring exercise needs with ATOPE+; according to (2.0) Healthcare Providers; ATOPE+ represent a means for the remote monitoring of patients exercise needs and the support of decision-making during an exercise intervention.

3.3.2 System Architecture

The architecture of ATOPE+ is shown in Figure 3.1. The first and fundamental element of the architecture is the *smartphone app* (hereon, just *app*). The app is the main communication channel with the patient, for both gathering data and receiving exercise prescriptions. The app collects data from three sources: wearable Bluetooth ECG, in-app questionnaires, and a Fitbit device. ECG and questionnaire data are collected directly through the app, and stored in a local database to ensure persistence of data. If Internet connection is available, data are sent to the server to generate an exercise prescription. The generated prescription is then communicated to the patient's app almost immediately. Last, Fitbit data collected in the *Fitbit Cloud* through the Fitbit app. A description of all the data available is found in Table 3.1. The *ECG* variables collected are the *time domain*, *frequency domain* and *Poincaré plot* features, all of them useful for short-HRV measurement (5-min) and to estimate workload-recovery ratio (Kaikkonen, 2015; Buchheit, 2014). The modulating factors of HRV are gathered through the *in-app questionnaire* features (Kim et al., 2018; Sajjadieh et al., 2020; Tran et al., 2009; Ltd., 2014; Plews et al., 2012), already successfully measured in patients with cancer in remote environments (Lozano-Lozano et al., 2018; Min et al., 2014; Børøsund et al., 2018; Børøsund et al., 2020). Fitbit's physical activity and sleep data are collected in its entirety as an objective and comparable measure of the exercise load performed by the patient during the day and within training sessions.

The second element of the system is the *centralized secure server*, and it embodies the *knowledge-based system* and the *clinical web interface* stated in the requirements. Several modules comprise the centralized secure server. ATOPE+ downloads Fitbit data with the *Fitbit querier* and incorporates it into the database with the *Fitbit data* *adapter*. The *Fitbit querier* interacts directly with the Fitbit web API (Fitbit Inc, n.d.) to download the fine-grained activity data of every participant, while the *Fitbit data adapter* adapts and inserts the JSON files returned by the Fitbit API into the *relational database*. A secured and authenticated *RESTful API* enables communications with the smartphone app to capture patient's ECG and questionnaire data. The API also serves as a means to deliver the personalized exercise prescriptions, which are stored once generated. Before building the exercise prescriptions, the raw heterogeneous data needs to be processed in order to extract meaningful information out of it. Raw data enters the *data manager* to be preprocessed and time-synced. Besides, this module cleans the ECG data by automatically detecting, removing, and interpolating outliers (Giles & Draper, 2018) and ectopic beats (Nabil & Bereksi Reguig, 2015), required to ensure correct short-HRV analysis (Giles & Draper, 2018). The *information manager* transforms the processed data into useful information related to different health domains of the patient: *sleep analyzer*, *active and sedentary behavior analyzer*, *training load analyzer*, *fatigue analyzer* and *distress analyzer* (modules not shown in figure for the sake of simplicity). All the information generated gets stored and serves as input to the *knowledge manager* to generate the individual exercise prescriptions. This information comes through the *information adapter* to feed simultaneously the *feature selector* and *recommendation builder*. The *recommendation builder* is the inference engine that generates personalized recommendations[2] according to the expert knowledge in the base of rules. The cascading *feature selector* and *machine learning prediction model* represent an active part of the data-driven knowledge by providing recovery predictions for each patient individually. This tool may even assist the expert in evaluating the fitness of rules to patients individually.

The remaining modules of the system revolve around the expert, in our case, a physiotherapist. The *Web interface* allows the expert to: visualize the patient's data gathered; generate new recommendations or modify existing ones for the patients through the *recommendation manager*; and introduce, modify or remove rules in the system through the *rule editor*.

3.3.3 System Implementation

This section details the implementation of ATOPE+ for the smartphone app and the centralized secured server, and provides insight on the use of ATOPE+ from the both patients' and experts' perspectives.

[2]In the case of ATOPE+, the term *recommendation* matches the *personalized exercise prescription*. For the sake of simplicity, the text will only refer to *recommendations* in the description of the ATOPE+ system.

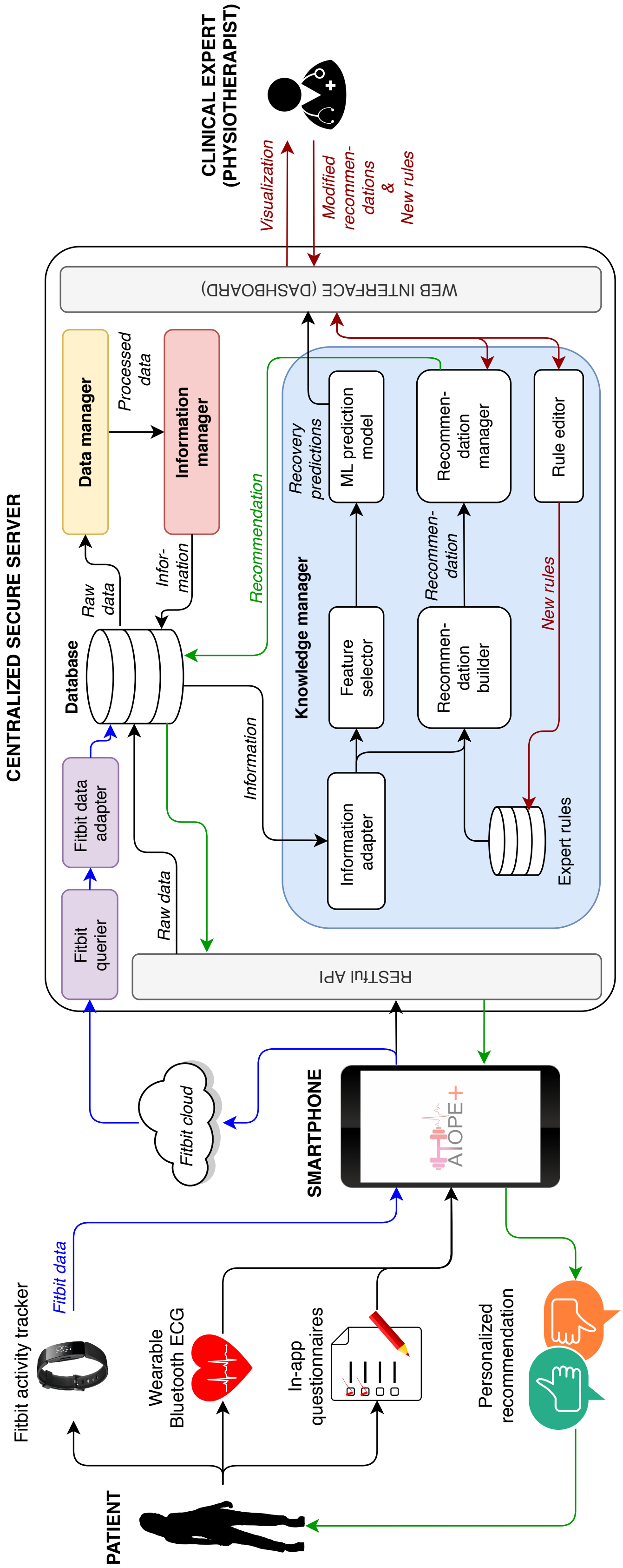


Fig. 3.1.: ATOPE+ architecture.

Tab. 3.1.: Data collected by ATOPE+. All variables are timestamped. ECG data are extracted from raw R-R signal with the Aura-healthcare *hrvanalysis* package (Champseix & contributors, 2020). Fitbit data are retrieved from its Web API (Fitbit Inc, n.d.) using the *python-fitbit* package (Orcas & contributors, 2019).

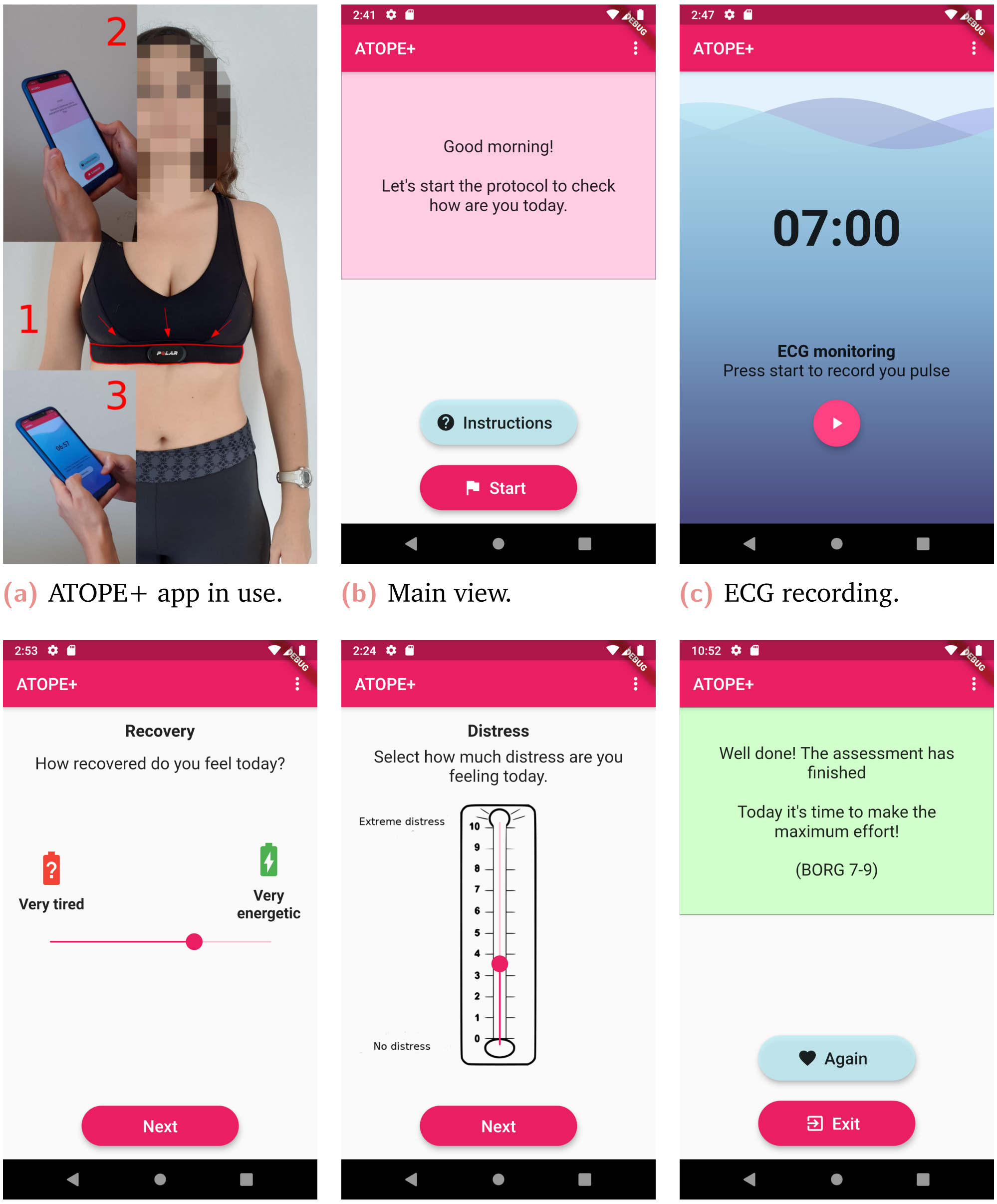
Source	Data	Type	Description
ECG	hr	int	Heart rate (beats per minute).
	rr	int	R-R interval in milliseconds.
	cvnni	float	Coefficient of variation equal to the ratio of sdnn divided by mean_nni.
	cvsd	float	Coefficient of variation of successive differences (rmssd divided by mean_nni.)
	cv_lnrmssd	float	Coefficient of variation of LnRMSSD 7-day rolling average.
	hf	float	Variance in R-R intervals in the high frequency (0.15 to 0.40 Hz).
	hfnu	float	Normalized hf power.
	lf	float	Variance in R-R intervals in the low frequency (0.04 to 0.15 Hz).
	lf_hf_ratio	float	lf/hf ratio as a quantitative mirror of the sympatho/vagal balance
	lfnu	float	normalized lf power.
	lnrmssd	float	Natural log of the root mean square of the successive differences.
	max_hr	float	Maximum heart rate.
	mean_hr	float	Mean heart rate.
	median_nni	float	Mean of R-R intervals.
	min_hr	float	Minimum heart rate.
	nni_20	int	Number of differences in successive R-R intervals greater than 20 ms.
	nni_50	int	Number of differences in successive R-R intervals greater than 50 ms.
	pnni_20	float	Proportion of NN20 divided by the total number of NN (R-R) intervals.
	pnni_50	float	Proportion of NN50 divided by the total number of NN (R-R) intervals.
	range_nni	float	Difference between the maximum and the minimum nn_interval
	ratio_sd2_sd1	float	Ratio between sd2 and sd1.
	sd1	float	Standard deviation of Poincare plot projection on the perpendicular to the line of identity.
	sd2	float	Standard deviation of Poincare plot projection on the line of identity.
	sdnn	float	Standard deviation of the NN (R-R) intervals
	sdsd	float	Standard deviation of differences between adjacent R-R intervals.
	std_hr	float	Standard deviation of heart rate.
	swc_lnrmssd	float	Smallest worthwhile change of LnRMSSD 7-day rolling average.
	total_power	float	Total power density spectral.
	vlf	float	Variance in R-R intervals in the very low frequency (0.003 to 0.04 Hz).
Wellness questionnaires	sleep_satisfaction	float	Sleep satisfaction in continuous Likert scale (0.0 – 10.0).
	sleep_time	int	Reported sleep time (minutes)
	distress	float	Distress in continuous Likert scale (0.0 – 10.0).
	recovery	float	Recovery in continuous Likert scale (0.0 – 10.0).
	fatigue	float	Fatigue in continuous Likert scale (0.0 – 10.0).
Fitbit's activity	steps	int	Steps count.
	intensity	int	PA level (0, sedentary; 1, lightly active; 2, fairly active; 3 very active)
	mets	int	METs (metabolic equivalents of task) expended
	calories	float	Calories expended.
Fitbit's sleep	sleep_level	string	Sleep stage ('deep', 'light', 'rem' and 'wake').
	nap	int	Number of sleep nap that day (0 is main sleep
	seconds	int	Duration in sleep stage (seconds).
Fitbit's training sessions	name	string	Name of activity.
	logtype	string	Type of activity ('auto_detected', 'manual', 'fitstar', 'mobile_run', 'tracker').
	active_duration	int	Duration of physical activity during session
	duration	int	Duration of session.
	calories	int	Calories expended in session.
	sed_time	int	Sedentary time in session.
	light_time	int	Light intense activity time in session.
	fair_time	int	Fair intense activity time in session.
	very_time	int	Very intense activity time in session.
	max_hr_normal	int	Max HR in normal level.
	max_hr_cardio	int	Max HR in cardio level.
	max_hr_fatburn	int	Max HR in fatburn level.
	max_hr_peak	int	Max HR in peak level (and in session).
	mean_hr	int	Mean hr in session.
	min_hr_cardio	int	Minimum HR in cardio level.
	min_hr_fatburn	int	Minimum HR in fatburn level.
	min_hr_normal	int	Minimum HR in normal level.
	min_hr_peak	int	Minimum HR in peak level.
	time_hr_cardio	int	Time in cardio zone.
	time_hr_fatburn	int	Time in fatburn zone.
	time_hr_normal	int	Time in normal zone.
	time_hr_peak	int	Time in peak zone.

Smartphone app

Taking into account the importance of cross-platform app development (essentially, Android and iOS), the ATOPE+ app was implemented using Flutter (Google LLC, Ireland) and it is shown in Figure 3.2. An exemplary use of the app is pictured in Figure 3.2a, it shows the ECG Polar H10 (Polar USA) position (1) and the start of the HRV recording protocol (2 and 3). Opening the app, the main view (Figure 3.2b) welcomes the patient with a message, instructions, and the option to start the protocol. Once the protocol has started, the app scans for available Bluetooth ECG devices to select one. Once the ECG is connected, the view lets the patient to start recording their HRV by pressing a *Play* button Figure 3.2c. The HRV recording is framed in a 7 minute countdown, out of which only the central 5 minutes are analyzed. Right after the countdown, the app notifies the patient with sound and vibration and the HRV data are sent to the server to be processed. The protocol is followed by the questions for sleep quality, recovery, fatigue and distress perception. Questions for sleep quality, fatigue and recovery perception (Hooper & Mackinnon, 1995) follow the design pictured in Figure 3.2d, a continuous Likert scale ranging 0 to 10 with labels in its extreme values. The distress view (Figure 3.2e) adapts the clinically validated NCCN Distress Thermometer (Cutillo et al., 2017) with a continuous slider too. Once the questions are finished, the responses are sent to the server to join the HRV data already processed and receive an automatic personalized exercise prescription, as shown in Figure 3.2f. This last view also provides the patient with the option to record their HRV again voluntarily, for example, if they think the HRV recording conditions were not ideal.

The ATOPE+ app includes some mechanisms to ensure data transfer to the server. All data are stored locally before being sent. If connection fails at the time of sending HRV or question data, the app will ask the patient to check the Internet connection and try again. Once the Internet connection is back, the data previously stored is sent to the server. This will only happen at the end of the HRV recording or after the last question is answered, thus avoiding disruptions of the protocol in the case of connection loss. Besides, if the patient were to exit the app in the middle of the protocol, a warning dialog would pop up to alert the patient they are about to exit the app, and inviting them to continue the protocol.

To ensure data reliability, different strategies are used to handle the HRV and the questions. Regarding HRV, first of all, the ECG Polar H10 (Polar USA) was selected due to being validated for HRV monitoring in at rest and during exercising (Gilgen-Ammann, Schweizer, & Wyss, 2019). If the server detects a problem while processing the HRV signal (e.g. less than 5 minutes recorded or an excessive amount of outliers



(a) ATOPE+ app in use. (b) Main view. (c) ECG recording.

(d) Recovery perception. (e) Distress. (f) Exercise prescription.

Fig. 3.2.: ATOPE+ smartphone app. The figures show an exemplary use of the app (a), the most representative views seen throughout the protocol (b-e), and the display of an exercise prescription once the protocol is finished (f).

and/or ectopic beats), its response will trigger in the app an error message, asking the patient again to record their HRV. Phone notifications can be very disruptive, thus the smartphone is automatically set up to silent mode while recording HRV; however in order to make the app not too obtrusive, silent mode is just applied to notifications and messages, while the volume of phone calls' ring remains unmodified. Volumes are brought back to the previous state once finished. Regarding the questions, the

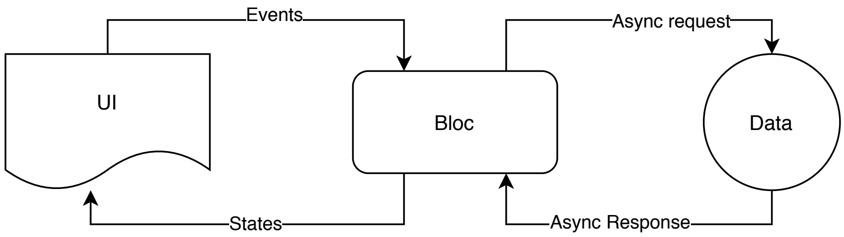


Fig. 3.3.: BLoC Design Pattern

patient is forced to answer them before advancing to the next question. This is done by disabling the *Next* buttons at the moment questions are presented, only enabling the *Next* button once the patient has actively selected a score on the slider. Moreover, the initial position of the slider is randomized for every question, which has proven to be an effective mechanism to mitigate anchoring in the responses (Gehlbach & Barge, 2012).

The ATOPE+ smartphone app is built following the BLoC Design Pattern (Business Logic Component) (Felix Angelov, n.d.)). BLoC is a design pattern to address state management in Flutter applications. BLoC serves to separate the states, and therefore the views, from the business logic of the app. Every communication between BLoC elements is performed via asynchronous data streams, making the logic responsive to events in the user interface (reactive programming).

Figure 3.3 depicts the BLoC design pattern. The business logic in the *bloc* is triggered by *events* captured in the *user interface (UI)* (e.g. tap), then, the *bloc* handles the *data* with the logic programmed in it. After this task is finished, the *bloc* sends a new *state* to rebuild the UI with a new view. BLoC was implemented in the app using the *bloc* and *flutter_bloc* Dart packages (Felix Angelov, n.d.).

Figure 3.4 describes the implementation of the BLoC design pattern in the the ATOPE+ app from the *bloc–state–UI* perspective. The app structure revolves around 4 main blocs: *Authentication*, *Login*, *Recommendation* and *Protocol*.

The *Authentication* bloc manages the verification, use and request of OAuth 2.0 client credentials to be able to send data and retrieve recommendations safely and unambiguously. When valid credentials are found, the *Authentication* bloc yields the *Authenticated* state to allow the *Recommendation* bloc to manage. If no credentials are found or there are connection or authentication problems, the bloc yields the *Unauthenticated* state thus releasing the *Login* bloc to command.

The *Login* bloc handles the required user interaction to the OAuth 2.0 authentication, in our case Authorization Code Grant. *LoginPage* and *ContinuePage* are the main

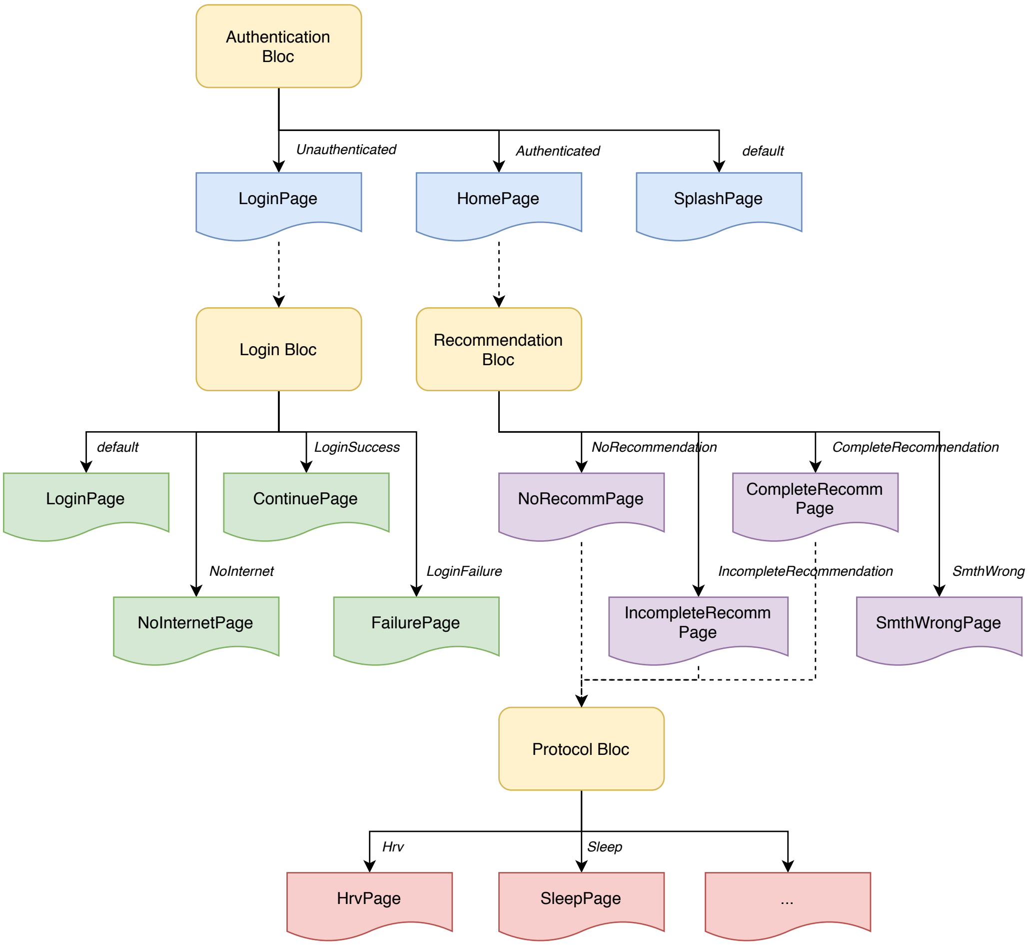


Fig. 3.4.: Bloc–state–UI structure of the ATOPE+ app. Blocs connect with UI (pages or views) through states.

views handling this interaction using a *WebView* to interact with the server webapp. Conversely, *NoInternetPage* and *FailurePage* deal with connection problems and exceptions that might occur when trying to use already saved OAuth 2.0 credentials, such as no internet connection, the expiration of refresh tokens, or any malformation of client credentials. Depending on the issue, *Authentication* bloc may need to renew OAuth 2.0 credentials (e.g. refresh token is revoked or outdated) or not (e.g. no internet connection).

The *Recommendation* bloc manages the app once OAuth 2.0 credentials are verified. This bloc handles communication with the server API when fetching patient's recommendations, and tells the *ProtocolBloc* (if available) the data to ask the patient in the protocol. This allows a more responsive experience for both experts and patients, so that questions are asked depending on the rules applied to each patient. If there is no recommendation available, *Recommendation* bloc yields the *NoRecommendation* state, thus allowing *Protocol* bloc to start the data gathering process and go through the different views that define it. If the recommendation is missing because there is missing or corrupted data, the *IncompleteRecommendation* state yields a view to communicate the Protocol bloc which data should ask again for. If all data is retrieved successfully to generate a recommendation in the server, *Recommendation* bloc yields a *CompleteRecommendation* state where the corresponing message is shown.

The *CompleteRecommendation* state also allows the patient to voluntarily record a new HRV session in the case of an involuntary disruption that may void HRV recording, for example, a phone call.

Finally, if OAuth 2.0 authentication works and yet there is a problem on the server side, the *Recommendation* bloc yields a *SmthWrong* state that displays a comforting message to the patient, and sends the issue directly to the server administrator.

Another main concern while implementing the app was usability. The number of interactions was minimized by including the least amount of elements in the screens (see snapshots in Figure 3.2). Patients are only required to login the first time they use the app to start using it. The protocol follows a straightforward path in the scheme of *one view, one question*, with icons to ease question identifying. There are no preferences to configure, all are controlled from the server side. Font and element sizes are high and controlled to avoid disruption of accessibility options that the smartphone might have enabled.

ATOPE+ was implemented using Flutter (Google LLC, Ireland). The app uses SQLite for data storage and AES encryption to secure it. The communications with the server are unambiguously authorized with OAuth 2.0 authorization protocol. OAuth 2.0 credentials are first obtained using Deep Linking (Nielsen, 2002) in Android and Universal Links (Apple, n.d.) in iOS. The ATOPE+ smartphone app was tested and built for Android versions over 4.4 (API 19) and iOS 8.0. HRV recordings were tested with a Polar H10 (Polar USA) device over BLE (bluetooth low energy) protocol.

Centralized secure server

The centralized secure server of ATOPE+ is responsible for storing and processing the data along with providing communication means for both patients and experts. As stated in the system architecture, the different layers conforming the system transform the data into useful information to, eventually, trigger the expert rules and provide the patients with personalized exercise prescriptions.

Data processing is different for HRV and the Fitbit data. For HRV processing, the *data manager* checks if its length is a minimum of 5 minutes. If so, the data manager looks for outliers in the HRV signal to be removed and linearly interpolated (Giles & Draper, 2018); ectopic beats are also detected and linearly interpolated (Nabil & Bereksi Reguig, 2015). Next, the *information manager* extracts time domain, frequency domain and Poincare features out of the clean HRV signal. Relevant features for estimating the workload-recovery ratio like the smallest worthwhile change (SWC) (Will G Hopkins, 2004; Javaloyes et al., 2019; Vesterinen et al., 2016) of the natural log of the root mean square of the successive differences (LnRMSSD) and the coefficient of variation (CV) (Buchheit, 2014; Plews et al., 2013) of the LnRMSSD are also extracted for a 7-day time window. The minute-by-minute Fitbit data are aggregated to match daily time windows and the training periods to extract features referred to both time windows.

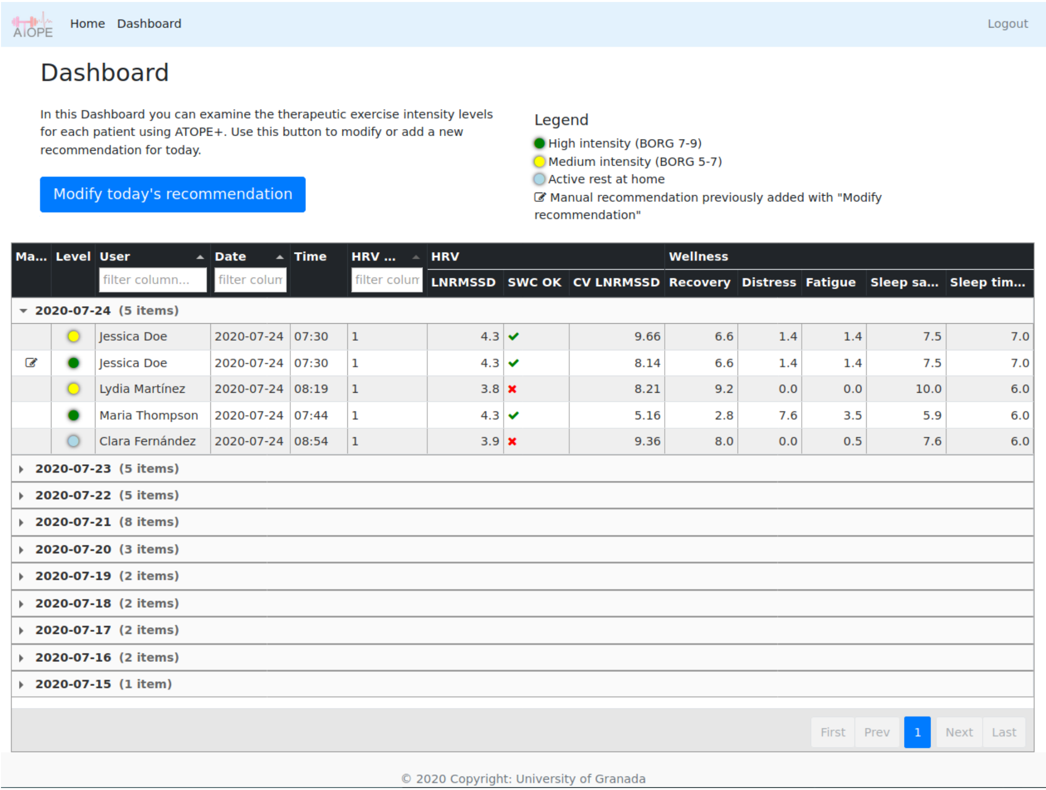
The base of rules permits defining rules depending on thresholds referred to question responses, HRV features and Fitbit features. For instance, an expert rule may define a *high intensity exercise prescription* if *SWC* is negative and *sleep satisfaction* value is greater or equal to 7.

The server implements a Dashboard as the expert web interface (Figure 3.5). The dashboard displays patient data and the exercise prescriptions given. The main view is shown in Figure 3.5a. Data are shown in a paginated table that groups the exercise prescriptions day by day. Data can be filtered by patient's name, date and the attempt to record HRV signal. The first column indicates if the exercise prescription shown has been manually added through the modification dialog. The second column shows the exercise description levels (to showcase, three different levels are defined). The table follows with patient's name, date and time of the exercise prescription. LnRMSSD, SWC and CV variables follow are the HRV features presented. Last, all the responses to the questions are presented under the wellness heading.

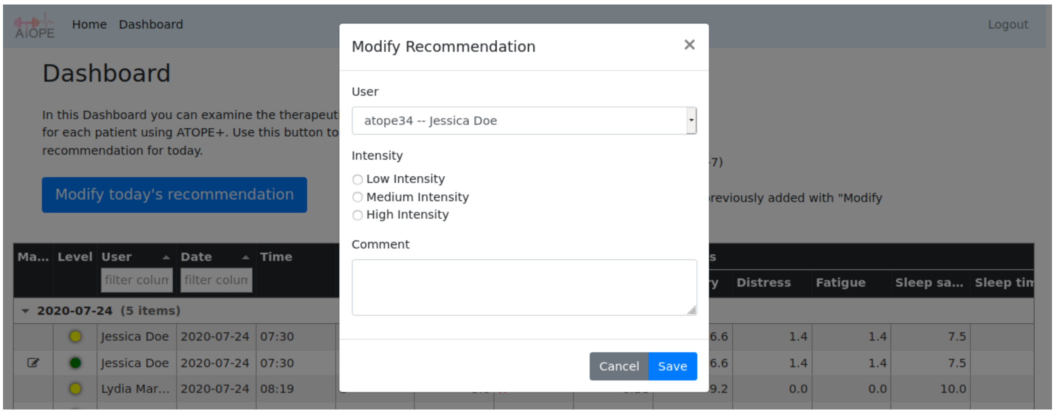
A dialog to create or modify exercise prescriptions is shown in Figure 3.5b. The dialog allows the expert to create or modify the exercise prescription for the day checked, by selecting the user and the intensity level of it. The expert can also provide a free comment on why the modification was necessary.

In order to ensure speed, stability, modularity and scalability, the different services composing the ATOPE+ server are implemented using Docker (Merkel, 2014). Docker enables the execution of programs in isolated environments by directly leveraging the host operating system resources. The implementation is divided in three services: *relational database*, *web application* and *reverse proxy*. Each service is a Docker container. All the containers are interconnected through a *Docker network*. The *relational database* runs on a MySQL 5.7 container. All its ports are closed to the outside, and its communications with the web application service are done via a *Docker network*. To ensure high speed performance in queries, data tables are partitioned to the number of participants to be enrolled in the ATOPE trial. The *web application* service is built over Flask 1.0.2 in a Python 3.7 container. This service features role-based authorization for users, an OAuth 2.0 authenticated RESTful API to connect with the ATOPE+ app, and the ATOPE+ dashboard. Last, the *reverse proxy* service exposes the web application securely to the internet over HTTPS through an uWSGI interface. The host machine runs Ubuntu 18.04 as operating system.

Regarding data security and privacy, Patients' data are gathered and stored meeting the European General Data Protection Regulation. The server is located within the facilities of the University of Granada (Granada, Spain) and its physical access is limited to the researchers participating in the ATOPE project and system administrators. All the data stored is pseudoanonymized (random UUID generation) and encrypted (LUKS1 with aes-xts-plain64 encryption). All online communications of the ATOPE+ system (ATOPE+ application and server) are secured via HTTPS connections with SSL/TLS encryption. Moreover, all the communications between the ATOPE+ smartphone app and the server are tokenized via OAuth 2.0 authorization to provide a secure delegated access for every patient. All communications with the database are made locally through a secured (HTTPS) web application. A firewall in the server limits the number of available ports for connections, only enabling ports 22 (SSH) and 443 (HTTPS).



(a) Main view.



(b) Dialog for exercise prescription modification.

Fig. 3.5.: ATOPE+ Dashboard.

3.4 Results

3.4.1 Usability Evaluation

Experts' Evaluation

Eight experts (6 female, 2 male; age 34.00 ± 7.03 years old), physiotherapists with TE experience in patients with cancer and survivors, used the ATOPE+ app and dashboard for seven days to test the whole system.

All the experts filled the SUS individually (Figure 3.6). The scores were computed and averaged for the app and the dashboard of ATOPE+. Both scored *A, excellent*, that is, over 80.3 points, 90th percentile. The app scored 91.6 ± 7.8 points (mean $\pm$ standard deviation) and the web dashboard 85.6 ± 20.9.

The answers to the app SUS are shown in Figure 3.7a. All the experts found the app likely to be used frequently (Q1), did not find it unnecessarily complex (Q2) and thought of it easy to use (Q3). Six of the experts did not consider the support of a technical person necessary to use the app (Q4). Every expert considered the functions of the app were well integrated (Q5) and that there was no inconsistency (Q6). Seven out of the eight experts imagined most people could learn to use the system very quickly (Q7). None of the experts found the app cumbersome to use (Q8), all of them were confident using it (Q9) and did not need to learn many things before using the system (Q10).

The results to the dashboard SUS evaluation are shown in Figure 3.7b. Seven of the experts found the dashboard likely to be used frequently (Q1), did not find it unnecessarily complex (Q2) and thought of it easy to use (Q3). Six of the experts did not consider necessary the support of a technical person to use the dashboard (Q4). Seven out of the eight experts considered the functions of the dashboard were well integrated (Q5), that there was no inconsistency (Q6), and that most people could learn to use the system very quickly (Q7). One of the experts found the dashboard cumbersome to use (Q8). Seven experts were confident when using the dashboard (Q9) and five did not need to learn many things before using the system (Q10).

The semi-structured interview was conducted to showcase the impressions of the experts from a more qualitative perspective. Regarding the app, all experts reported from "*good*" to "*very good sensation using the app.*" For example, expert #2 said, "*The overall sensation was very good, very intuitive to follow and with a very good connection to the Polar (ECG device)*", and expert #8 said, "*Quite a good sensation.*

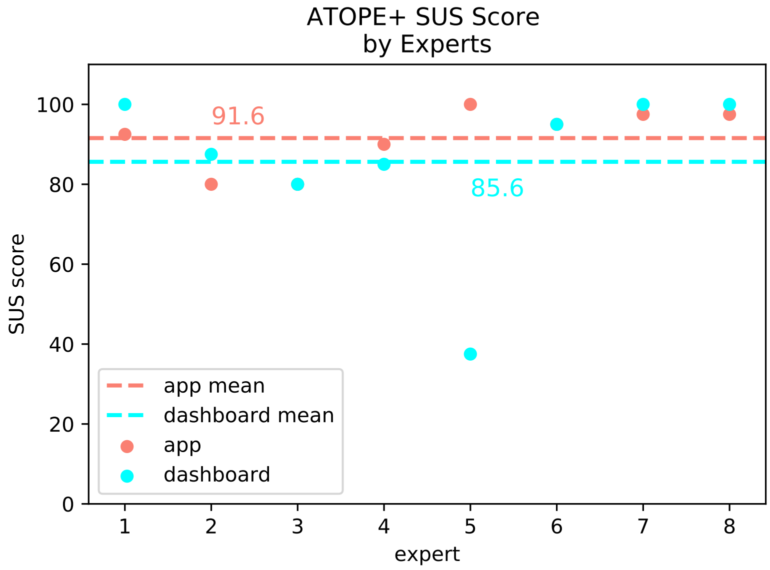
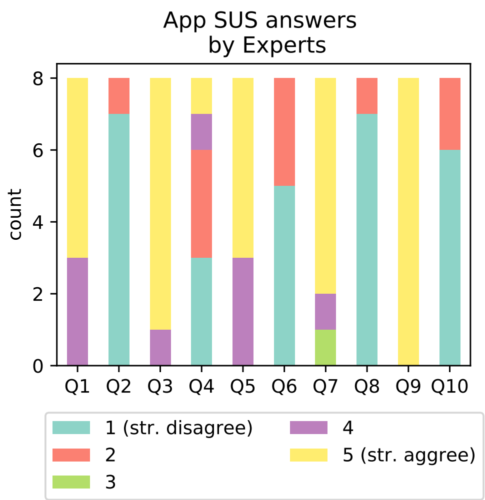
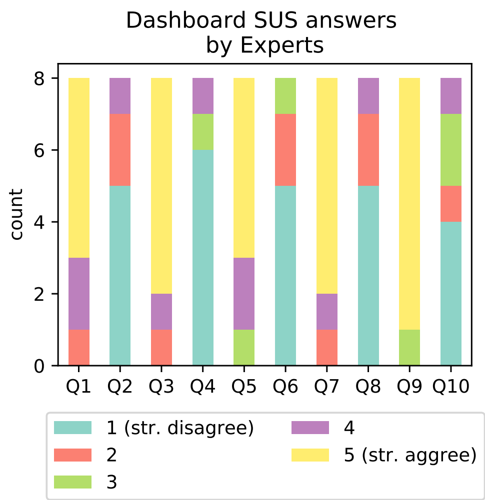


Fig. 3.6.: SUS score for the ATOPE+ app and dashboard by expert.



(a) Experts' app evaluation.



(b) Experts' dashboard evaluation.

Fig. 3.7.: Experts' SUS evaluation of ATOPE+ detailed by question. Each bar shows the score count for each of the ten SUS questions. Each color represents the type of answer.

Very simple to use, clean, with no (unnecessary) ornaments and very intuitive." All the experts highlighted the straightforwardness in the use of the app during the interviews. Experts also contemplated the need for training on how to use the app for some of the less skilled patients. As expert #2 reported, "*It is plenty accessible. It will always depend a little on the technological skills of the patient, but they can always receive training during the first and second week of the intervention.*"

To further detail the impressions on the use of the app, the experts were asked about protocol complexity, the clarity of the instructions given and the perspective of the patients using the app during the entire TE intervention. They all agreed on the simplicity of the protocol, the clarity of the in-app instructions and the ease for patients to use it daily.

Some of the experts underlined the importance of delivering and adequate feedback. Quoting expert #4, "*The app may foster patient's autonomy and adherence thanks to the personalized feedback, thus improving her results at the end of intervention.*" Expert #6 reported, "*Patients can learn to use this app easily and engage well, specially if the feedback presented to them is realistic and useful, and they actually see it translated into the (TE) intervention.*"

Taking into account the use of monitoring devices such as the wearable ECG (Polar H10), expert #3 said, "*It is not complicated (to attach the Polar H10), it may be even preferable to sleeping with the wristband (the Fitbit). For patients with breast cancer before surgery this would not be a problem. For those after surgery, they may need some extra attention and be carefully trained on how to use it.*" Conversely, expert #4 addressed, "*ATOPE+ needs to be careful in the number of elements participating within the measures, since each one represents a higher grade of complexity, thus rising the probability of errors,*" right after highlighting the potential of using portable monitoring devices.

Along with the positive feedback, there was place to express concerns, constructive critiques and suggestions. Expert #1 was "*worried if patients could maintain the daily use of the app.*" "*They can just forget, specially once you are in the middle of the treatment and stressed,*" she continued. Next, the expert suggested, "*A daily notification in the morning could help to remind the patients to start the protocol right after waking-up.*" Expert #4 was concerned about the validity of the HRV measures, since measuring conditions can be critical. He proposed, aside from the technical issues of filtering and processing the HRV signal, "*You can ask the participant, right at the start of the protocol, if the environment conditions are actually ideal, in the form of a checklist: 1) did you empty your bladder? 2) did you drink coffee/tea? 3) are you in a calm settled environment?.*" Most of the experts agreed that a chat/video-chat with the participant would be also very useful to establish a more solid communication, and the feedback messages could be improved just by mentioning the patient's name. Expert #6 even contemplated the idea of including "*gamification elements to foster patient's engagement, with very visual feedback.*"

Regarding the web dashboard, all experts reported a good experience while using it and that it was useful to check patients' assessment and make quick decisions on their TE intervention. They all found the option to modify the exercise prescription very intuitive. Expert #3 said, "*The dashboard is pretty intuitive, you can easily take a quick look at the evaluations of each patient.*" Some of them requested some features such as the display of data in graphics with trends and visual alerts of anomalies or values out of range.

All the experts agreed on the potential of ATOPE+ to improve the TE intervention process compared to the traditional treatment. In words of expert #7, "*ATOPE+ may provide a further objective and personalized assessment.*" Expert #4 said, "*ATOPE+ addresses the personalization and monitoring process in a new and unprecedented way.*" Expert #1 added that using ATOPE+ would mean an optimization of resources for both patients and medical personnel:

> *The dashboard is very convenient, it saves a lot of time. The remote personalized assessment alleviates a lot of evaluation tasks from the experts and saves unnecessary journeys from the patients at the medical center to be assessed, thus saving time and resources for us all. Patients could be sent home again due to not being in optimal conditions to perform TE that day. (Expert #1)*

All the experts acknowledged the possibility of using ATOPE+ in a 100% remote environment such as the COVID-19 pandemic context. They also agreed on the need to make some minor adjustments. Expert #7 commented, "*Since ATOPE+ is focused mainly now in (workload-recovery ratio) assessment, it would be necessary to provide more material to complete the TE program.*" Expert #6 added, "*The engagement with the program would need to be carefully studied. It is not trivial, maybe via technical means such as gamification and/or available communication channels, and also individual supervision by medical personnel.*"

The experts also foresaw the possibility of extrapolating the use of ATOPE+ to other kinds of patients. Quoting expert #5, "*This methodology could be used with patients with other types of cancer (different to breast cancer), cardiopathy or neurological conditions.*"

Patients' Evaluation

Twenty-five breast cancer patients and survivors (8 patients, 17 survivors; age 49.00 ± 7.21 years old) filled the SUS individually after using it for a minimum of 7 days (Figure 3.8). The scores were computed and averaged for the ATOPE+ app. The app

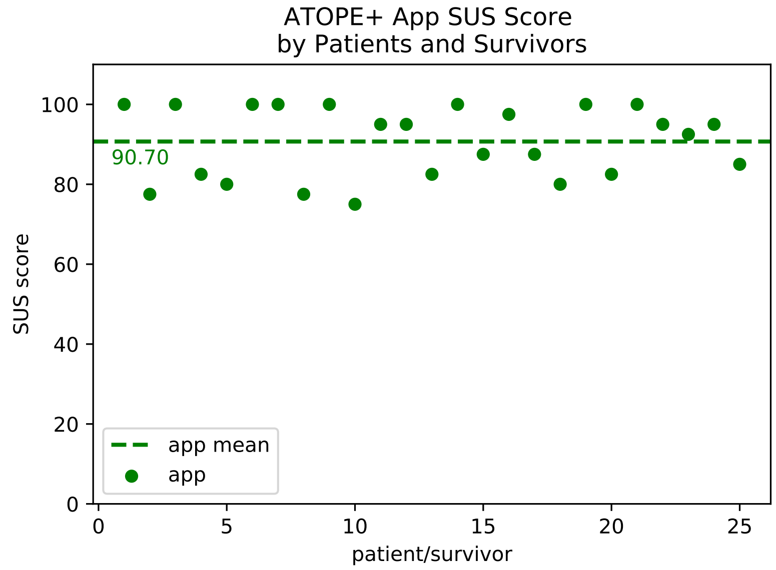


Fig. 3.8.: SUS score for the ATOPE+ app by patients with breast cancer and survivors.

scored *A, excellent*, i.e., over 80.3 points, 90th percentile. The app scored 90.7 ± 8.9 points (average $\pm$ standard deviation).

The answers to the app SUS are shown in Figure 3.9. Twenty-four (96%) of the patients and survivors found the app likely to be used frequently (Q1), and did not find it unnecessarily complex (Q2). All of them thought the app was easy to use (Q3). Twenty-one (84%) of the patients and survivors did not consider the support of a technical person necessary to use the app (Q4). Twenty-three (92%) of them considered the functions of the app were well integrated (Q5) and that there was no inconsistency (Q6). Twenty-two (88%) patients and survivors imagined most people could learn to use the system very quickly (Q7). None of the patients and survivors found the app cumbersome to use (Q8). Finally, twenty-three (92%) of them were confident using the app (Q9) and did not need to learn many things before using the system (Q10).

At the end of the SUS questionnaire, patients and survivors were also asked for voluntary comments in a free-text cell. Nine of them (36%) wrote something on it. Seven patients and survivors highlighted the straight-forwardness of the app, that it was easy to use, quick, and reliable; and five of these seven also appreciated the initial training on how to use the app. One patient pointed out that it would be interesting to access the app without the need for an ECG band, and another that the band may produce itching in the skin in some days.

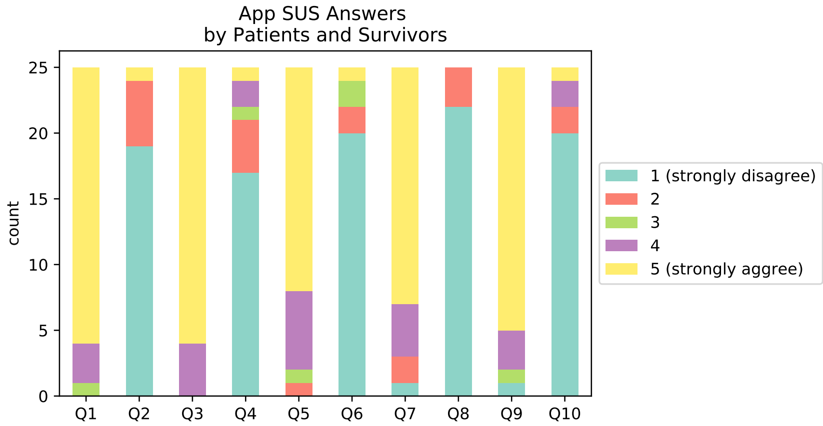


Fig. 3.9.: Patients' and survivors' app evaluation.

3.5 Discussion

3.5.1 Principal Findings

This chapter presented ATOPE+, an mHealth system to support personalized therapeutic exercise interventions in patients with cancer. The system architecture and implementation were thoroughly described. A usability evaluation was conducted by clinical experts, patients with breast cancer and survivors to show the potential of the system and the usability of its elements. The system, the results obtained, implications and recommendations for future studies are further discussed in this section.

ATOPE+ Development

A system like ATOPE+ can only emerge from the interdisciplinary cooperation among the physical therapy, medical, engineering and computer science fields. On the clinical side, the relevance of ATOPE+ is rooted in enabling individual remote monitoring of key variables to workload-recovery in patients with cancer. On the technological side, the relevance of ATOPE+ is drawn from the integration of commercial wearable monitoring devices, a data processing pipeline and clinical expert knowledge into a knowledge-based system to automatically provide personalized prescriptions of exercise dosage. Overall, ATOPE+ allows clinical experts to simplify knowledge management and decision-making within the context of a TE intervention by integrating in one tool the process of diagnosing and providing patients' with their individual exercise needs.

ATOPE+ is heavily inspired by the systems presented in the State of the Art (Chapter 2). The multilevel architecture present in most of the mHealth systems referenced (Banos, Amin, et al., 2015; Ferreira et al., 2015; Mehrotra et al., 2017; Castro et al., 2015; Alharthi et al., 2019) demonstrated its added value handling knowledge management, specially after being tested in different health applications such as promoting physical activity, general wellbeing and mental health. The small presence of similar approaches with patients with cancer in TE interventions served as a major impulse for this work, specially noting the lack of sophisticated personalization strategies. Most of the personalization strategies found were based on self-management and/or self-monitoring of physical activity with wearable devices (Schaffer et al., 2019). ATOPE+ takes a different approach by rooting its personalization strategy in the physiological foundations of workload-recovery ratio assessment by means of HRV (and its most relevant modulating factors: sleep satisfaction, distres, recovery perception and fatige.)

Despite the availability of general and open frameworks for the development of health applications such as AWARE (Ferreira et al., 2015), Beiwe (John Torous et al., 2016) or Bridge (Sage Bionetworks, 2019), an ad-hoc implementation of ATOPE+ provided a better support for our requirements. First, these frameworks present some limitations on handling knowledge with a base of rules and delivering recommendations. Although they provide dashboards for the monitoring of variables, their architecture does not include a knowledge based nor an inference engine. Second, none of the frameworks available at the moment were cross-platform (Android and iOS), they were all Android-based, hence limiting the patient recruitment capabilities of the application. Only Bridge (Sage Bionetworks, 2019) provided the tools to develop two apps for each platform (Android and iOS); however Flutter provided a simpler solution by supporting cross-platform with the same SDK (software development kit). Finally, these frameworks provide plenty of options for the users in their preferences. Despite some of them could be hidden to the end-use, the aim was to provide the most simple experience for patients from design.

Usability

Usability results were consistent and promising for ATOPE+. For the experts, the overall good scores in the SUS scale matched the answers to the semi-structured interview, for both the app and the dashboard. All the experts agreed on the potential of ATOPE+ to improve the personalization process in a TE intervention with patients with cancer. Moreover, while there was some critique pointing out possible improvements for ATOPE+, none of the commentaries or suggestions were deemed as major issues.

All experts agreed on the ease of use of the ATOPE+ app and the straightforwardness of the data collection protocol. The experts also reported that there was no need for the support of a technical person, that patients would only require a training period to use the app. This is an important result for ATOPE+, since providing an intuitive app experience is imperative to make the system accessible to all the patients, specially those with less technology skills such as the elder generations.

These comments match the usability evaluation of patients and survivors, who also reported very good app usability. Patients highlighted the simplicity and straightforwardness of the protocol, and there was no evaluation below 68 points, the minimum usability score. These results shows that the inclusion of from-design simplicity into ATOPE+ was successful for the two end-users of the system, and ATOPE+ can be presented as ready to be used in the context of a clinical trial.

The experts agreed on the possibility of using ATOPE+ in a fully remote environment, only requiring some adjustments on the intervention protocol. Experts' expectations on this topic were confirmed, since patients successfully evaluated the usability of ATOPE+ in a completely remote scenario.This is particularly relevant now in a COVID-19 pandemic context. The immunosuppression often related to cancer treatment may put patients at the very high risk of getting infected with COVID-19. Patients with cancer appear to have an estimated twofold increased risk of contracting SARS-CoV-2 (severe acute respiratory syndrome coronavirus-2) than the general population (Yu et al., 2020). Recent literature already recommended reducing this risk by minimizing exposure and prioritizing individualized assistance, suggesting the inclusion of telemedicine strategies as a means to do so (Al-Shamsi et al., 2020). Hence, a tool like ATOPE+ may become of interest in the uncertainty of the following times until the COVID-19 disease is set under control.

The experts shared their ideas on the need for providing objective feedback to patients on their performance. On the one hand, some of the experts were reluctant to include more information than the daily exercise prescription. These experts were concerned about patients becoming too self-aware on their performance and even trying to figure out the inner logic of ATOPE+. On the other hand, other experts were supportive of providing the maximum amount of feedback by wrapping it in a game-based context. These experts considered that a gamification strategy might provide them with tools to promote the fulfillment of the exercise prescriptions. Nevertheless, they also acknowledged that gamification strategies might need to be as tailored as the exercise prescription to become effective (Zhao, Arya, Orji, &

Chan, 2020). Since ATOPE+ is not a system that focuses on patient's self-regulation, future use of ATOPE+ will limit its feedback to the one provided by the commercial wearable used and the daily exercise prescriptions in the near term.

The SUS dashboard evaluation by expert #5 stands out as an outlier in Figure 3.6. This expert was particularly interested in the visual display of data, its trends, the presence of outliers and alerts. Future efforts will focus on providing these tools effectively. The rest of the experts considered that these extra tools would be valuable, but also that the information displayed was sufficient to evaluate the adequacy of the exercise prescriptions.

3.5.2 Limitations and Future Work

Our usability results present some limitations. Usability results were gathered between March and September of 2021, a period in which COVID-19 still menaced the health of immunosuppressed people such as patients with cancer despite its lower incidence at the time. Moreover, the clinical saturation caused by the rush of new cases, and the fear of patients of getting infected, hindered the recruitment possibilities. This situation forced the usability evaluation to be conducted combining the experience of breast cancer patients and survivors in a fully remote scenario. Although exercise interventions are conducted similarly for both patients and survivors (Campbell et al., 2019; Pollán et al., 2020), their psychological states may differ due to the stage of their recovery.

Future work may not be limited to patients with breast cancer. All the experts foresaw extending the use of ATOPE+ to other types of cancer and diseases, thus opening other lines of work such as lung or colorectum cancer —both hold the highest incidence in Europe and in the United States. Our long-term research will also aim to describe the most relevant variables related to the workload-recovery ratio that influence the decision making when prescribing exercise dosage in patients with cancer.

3.6 Conclusion

This chapter described ATOPE+, an mHealth system to support personalized therapeutic exercise interventions in patients with cancer. ATOPE+ enables the remote assessment of workload-recovery ratio to provide optimal exercise dosage by means of a knowledge-based system. Thus, ATOPE+ allows for undulating adaptive prescription of exercise, minimizing the risk of overtraining and undertraining throughout the TE intervention. To our knowledge, ATOPE+ is the first mHealth system combining measures of exercise load (HRV), modulating factors of HRV (recovery, sleep, distress, fatigue), and daily and training-specific physical activity levels (Fitbit activity tracker) to personalize therapeutic exercise interventions in patients with cancer. Overall, ATOPE+ allows clinical experts to simplify knowledge management and decision-making within the context of a TE intervention.

ATOPE+ presents a novel concept to personalization in TE interventions in patients with cancer, by using physiological variables related to training load in a remote context. The architecture of ATOPE+ is designed to collect physiological data from heterogeneous sources (wearable ECG, in-app questionnaires, Fitbit cloud), transform the data into useful information, and provide individual exercise prescriptions by means of an knowledge-based system.

Multiple tests with patients with breast cancer in TE intervention were conducted successfully. A usability evaluation was conducted to determine how medical personnel, and patients and survivors would receive ATOPE+. Results showed good satisfaction with the tool as simple, straightforward and easy to use. The experts perceived ATOPE+ as a promising tool to improve therapeutic exercise evaluations.

4 ATOPE+Breast: Continuous Monitoring of Training Load in Patients with Breast Cancer during Therapeutic Exercise Intervention

Publication Details

Reference: Moreno-Gutiérrez, S., Postigo-Martín, P., Damas, M., Banos, O., Pomares, H., Arroyo-Morales, M., & Cantarero-Villanueva, I. (2022). ATOPE+Breast, Continuous Monitoring of Training Load in Patients with Breast Cancer during Therapeutic Exercise Intervention. Zenodo, Dataset. https://doi.org/10.5281/zenodo.6322773

Preceding Note

The aim of this chapter is to contribute with an open dataset describing the continuous monitoring of training load in patients with breast cancer. The dataset was published in Zenodo in the reference above.

Nevertheless, for this dataset to be helpful, the reliability of ATOPE+ had to be assessed first (besides the usability presented in the previous chapter). The reliability of ATOPE+ was done in a separate work (Postigo-Martin et al., 2022) that is not part of this thesis. In such work, ATOPE+ was found as a valid and reliable tool to assess autonomic balance (LnRMSSD), sleep satisfaction, emotional distress, and potentially fatigue in breast cancer survivors. This reliability assessment enabled the publication of these data.

4.1 Introduction

Multiple studies address therapeutic exercise interventions in patients with cancer (Y.-H. Lee et al., 2018; Niederer et al., 2012; Caro-Moran et al., 2016; Dias Reis et al., 2017; Schaffer et al., 2019); however, these works typically measure the impact of the intervention by comparing health assessments (e.g., HRV) before and after the intervention. There are no available datasets describing the status of breast cancer patients during an exercise intervention. Hence, the objective of this chapter is to introduce a dataset of such nature: the *ATOPE+Breast* dataset.

ATOPE+Breast (ATOPE+ for patients with breast cancer) describes the daily status of patients with breast cancer during therapeutic exercise intervention with daily measures of HRV, self-reported wellness, physical activity, and sleep. Besides, the ATOPE+Breast dataset contains information about training sessions, such as intensity recorded, demographic data, treatment details, initial evaluations of quality of life, physical activity levels, previous medical conditions, and risk factors.

This chapter is structured as follows. Section 4.2 presents the materials and methods used to collect the dataset, describing the different variables and how they are coded. Section 4.3 provides notes on data availability (source), and a description of the data collected. Finally, section 4.4 provides an overview of the limitations presented by the dataset.

4.2 Materials and Methods

4.2.1 Study Design

The ATOPE trial (registration number NCT03787966, Clinicaltrials.gov; protocol published by Postigo-Martin et al., 2021) aims to compare the beneficial effects of a therapeutic exercise intervention in patients with breast cancer during treatment versus a therapeutic exercise intervention in patients before treatment.

The ATOPE program is a 12-to-18-sessions program lasting 6 to 8 weeks, depending on the treatment scheduled for each patient, supervised (1-on-1) program of therapeutic exercise that consists of multimodal therapeutic exercise (aerobic, strength, motor control exercises, myofascial techniques, and breathing exercises) implemented by a physical therapist expert in therapeutic exercise. Recovery strategies are followed at the end of each session. The whole duration of the sessions is approximately 1 hour.

ATOPE+ supports the ATOPE program enabling a personalized therapeutic exercise intervention. ATOPE+ provides daily tailored recommendations of the exercise needs for every patient depending on training load data (HRV and self-reported wellness) according to expert rules already protected by intellectual property (registration number 2010285737407; SafeCreative ATOPE+, 2020). The physical therapists used the recommendations and the data collected to support their decision-making process when prescribing exercise doses. The first two weeks of the intervention were scheduled with 2-3 training sessions to obtain solid HRV baselines and for patients to learn the exercises.

4.2.2 Participants

The participants were 23 patients with breast cancer (48.8 $\pm$ 12.2 years old) from the ATOPE trial. Through questionnaires, patients reported their demographic data, quality of life, and initial physical activity levels. Previous conditions and risk factors were noted by a physician after an initial screening. Patients were provided with a Polar H10 ECG monitor (Polar Electro Ltd.) and the ATOPE+ app installed on their smartphones to record their HRV and self-reported wellness every day. In addition, patients were provided with a Fitbit Inspire HR to monitor their daily physical activity levels and sleep patterns. A description of the collection of the ATOPE+Breast dataset is shown in Figure 4.1.

4.2.3 Eligibility

The inclusion criteria contained women with newly diagnosed, histologically confirmed, unresected stage I–IIIa BC. In addition, all had to be: (1) >18 years old; (2) scheduled for surgery, chemotherapy, and/or radiotherapy; and (3) predisposed to developing cardiotoxicity, as described by the American Society of Clinical Oncology Guidelines (Armenian et al., 2017).

The exclusion criteria were: (1) a previous history of malignancy; (2) having undergone previous cancer treatment; (3) pregnancy; (4) having a psychiatric or cognitive disorder that prevents patients from following exercises correctly, and/or acute or chronic condition that prevents exercise; and (5) any absolute contraindication for high-intensity exercise.

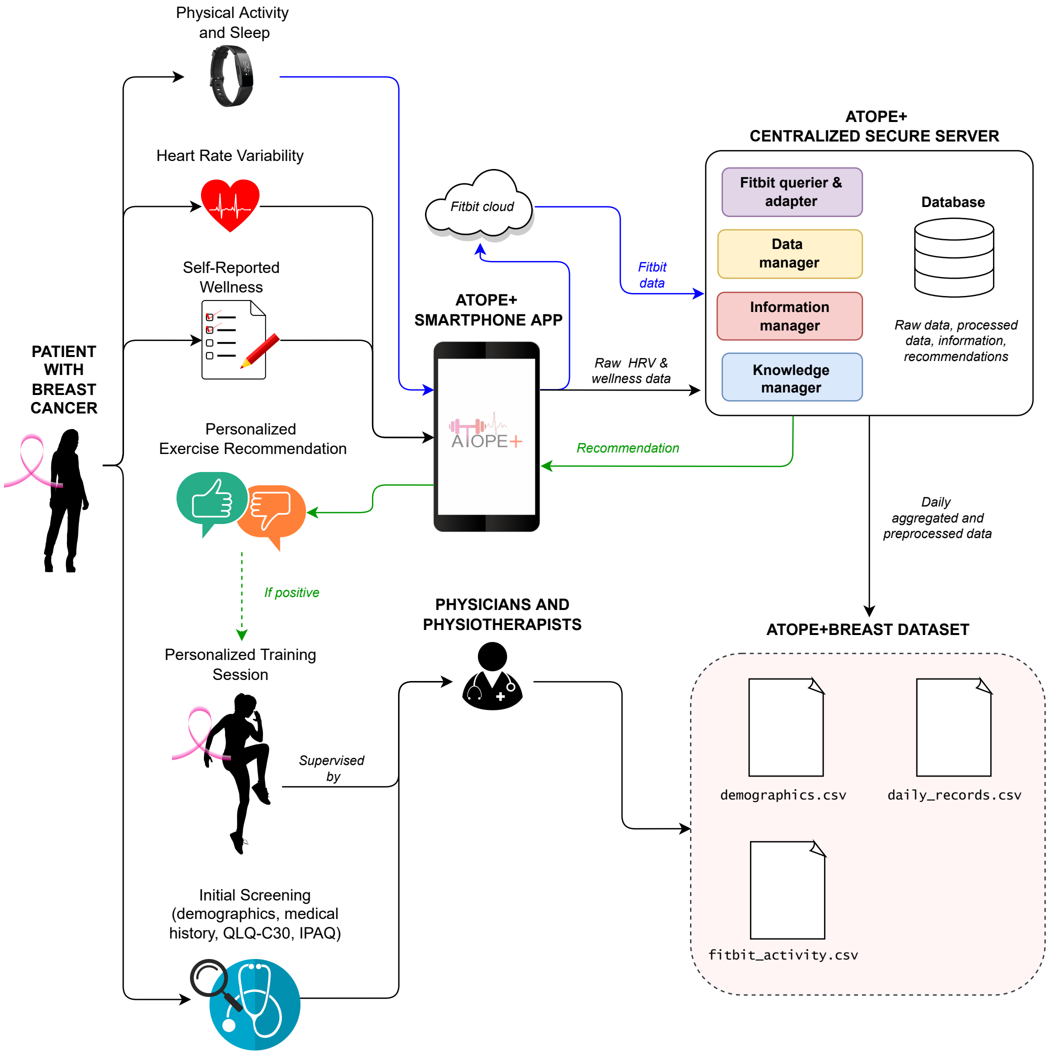


Fig. 4.1.: ATOPE+Breast dataset collection.

4.2.4 Data Collection

ATOPE+ was used with a Bluetooth ECG Polar H10 and a Fitbit Inspire HR to record HRV, self-reported wellness, physical activity, and sleep data. Physicians and physical therapists —experts in therapeutic exercise— collected demographic and medical history data, along with training session details. A total of 306 training sessions were conducted —and noted— with the 23 patients; also, 681 measurements of HRV and self-reported wellness were recorded daily during the monitoring period. In addition, 845 measures of Fitbit daily physical activity, and 687 measures of Fitbit sleep were collected. A summary of the monitoring and training data collected is illustrated in Figure 4.2.

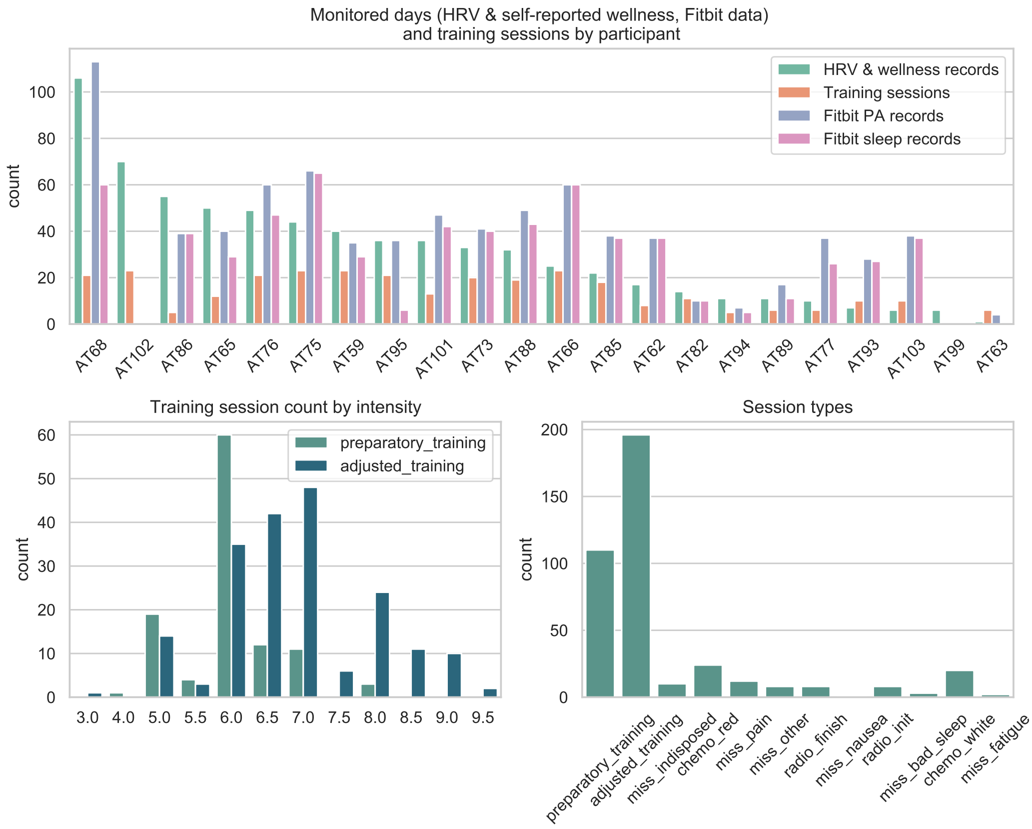


Fig. 4.2.: Description of monitoring sample by participants.

Heart Rate Variability

Patients were instructed to measure their HRV in a lying position and answer the self-reported wellness questionnaires with the ATOPE+ app every morning at home, right after waking up and emptying their bladder, and before starting the day (e.g., drinking coffee, having breakfast, checking emails), to avoid influences of circadian rhythm (Plews et al., 2012; Buchheit, 2014).

HRV measurements consisted of 10 minutes of R-R intervals. Outliers and ectopic beats of HRV signal were replaced with linear interpolation (Peltola, 2012; Giles & Draper, 2018). ATOPE+ removed the first and last 2.5 minutes of the recording, using the 5 minutes in the middle of the R-R signal for short-term HRV analysis (Shaffer & Ginsberg, 2017). ATOPE+ extracted the daily logarithm of the RMSSD (LnRMSSD) for every recording since it is the most accepted and reliable resting HRV-related parameter to measure the adaptation of in response to training (Plews et al., 2013; Buchheit, 2014). ATOPE+ also extracted the remaining time-domain, frequency-domain, and Poincaré plot HRV features previously described in Table 3.1.

HRV signals were checked before its computation. If Bluetooth disconnections were detected, or there was insufficient HRV data for short-HRV analysis, the patient was automatically asked for second monitoring of her HRV signals.

SWC of the LnRMSSD. ATOPE+ calculated the smallest worthwhile change (SWC) (Will G Hopkins, 2004) of the LnRMSSD out of a fixed number of standard deviations around the mean LnRMSSD (Vesterinen et al., 2016; Javaloyes et al., 2019). Nevertheless, the criteria defined for SWC slightly differed from those used for professional athletes. Because of the fragile psychological state of being diagnosed with a potentially terminal illness like breast cancer, the frail physiological state induced by systemic treatment, and the possible lack of exercise habits, training adaptation in patients with breast cancer may be subject to higher variability than the one found in professional athletes. Moreover, the objective of professional athletes is to improve performance, while for patients is to achieve an adequate training load to reduce toxicity levels —and avoid overtraining, which may cause more damage to the patient.

Hence, the observing window for SWC was reduced from comparing the evolution of a 7-day-rolling-mean window of the LnRMSSD against the SWC computed for a 4-weeks-rolling window of LnRMSSD measures (Figure 2.8, Javaloyes et al., 2019) to comparing the daily values of LnRMSSD against the SWC computed for a 5-to-7-days-rolling window. SWC thresholds were defined with a factor f of 0.5 as in (Vesterinen et al., 2016; Javaloyes et al., 2019):

$$SWC_{thresholds} = LnRMSSD_{rolling\_mean} \pm 0.5 \cdot LnRMSSD_{rolling\_std} \tag{4.1}$$

where $LnRMSSD_{rolling\_mean}$ and $LnRMSSD_{rolling\_std}$ are the mean and standard deviation computed for the available LnRMSSD of the previous 7 days, with at least 5 of them available (Plews et al., 2014).

The categorical variable SWC_{ok} was True if the LnRMSSD did not fell outside the SWC thresholds; however, to provide a continuous variable of the SWC approach proposed, the following normalized SWC variable was defined:

$$SWC = \frac{LnRMSSD - LnRMSSD_{rolling\_mean}}{LnRMSSD_{rolling\_std}} \tag{4.2}$$

CV of the LnRMSSD. ATOPE+ calculated the CV of the LnRMSSD (W. G. Hopkins, 2000) using a 5-to-7-days-rolling window to establish sufficiently solid baselines for the comparison of daily measures of LnRMSSD (Plews et al., 2014), as previously defined in Equation 2.1.

Self-Reported Wellness

As described in subsection 2.3.2, stress (Kim et al., 2018), sleep (Sajjadieh et al., 2020), and fatigue (Tran et al., 2009) are the modulating factors of HRV getting more attention from the research community (Ltd., 2014; Plews et al., 2012). Such factors were successfully measured in remote conditions in patients with cancer (Cantarero-Villanueva et al., 2014; Lozano-Lozano et al., 2018; Børøsund et al., 2020; Børøsund et al., 2018; Min et al., 2014).

ATOPE+ recorded perceived recovery, sleep time, sleep satisfaction, emotional distress, and peripheral fatigue using self-reported in-app questionnaires. Features were previously described in Table 3.1. Participants were required to actively answer every item in the questionnaire before advancing to the next question. This was done by enabling the Next button only once the participant had actively selected a score on the slider. Moreover, the initial position of the Likert-like sliders was randomized for every question to mitigate anchoring or learning effects in the responses (Gehlbach & Barge, 2012).

Recovery. ATOPE+ used the gold standard Perceived Recovery Scale (Laurent et al., 2011) to assess the level of recovery perceived by the patients every morning. ATOPE+ used a self-reported continuous Likert-type scale, with scores from 0 (*Very tired*) to 10 (*Very energetic*) labeled on the extremes, to measure recovery.

Sleep. ATOPE+ used the gold standard consensus diary to measure sleep patterns and sleep quality (Carney et al., 2012). ATOPE+ used a self-reported continuous Likert-type scale, with scores from 0 to 10, and labeled on the extremes, to measure recovery; besides, ATOPE+ used a digital-clock-like counter to enable the self-report of sleep time. In addition, ATOPE+ incorporated sleep data (time and stages) from a Fitbit Inspire HR when used during the night.

Emotional Distress. ATOPE+ used the gold standard NCCN Emotional Distress Thermometer (Cutillo et al., 2017) to measure emotional distress. ATOPE+ used a self-reported continuous Likert-type scale, with scores from 0 (*No distress*) to 10 (*Extreme distress*) labeled on the extremes and embedded in the picture of the original NCCN Distress Thermometer.

Peripheral Fatigue. ATOPE+ used the gold standard Borg-CR 10 scale (Soriano-Maldonado et al., 2015) to evaluate the level of perceived fatigue after a physical exertion test. Participants performed 10 repetitions at 40 beats per minute (marked by a metronome sound included in ATOPE+) of the *Sit to Stand Test*, frequently used to induce fatigue in lower extremities (Hatton, Menant, Lord, Lo, & Sturnieks, 2013). ATOPE+ used a self-reported continuous Likert-type scale, with scores from 0 (*No fatigue*) to 10 (*Extreme fatigue*) labeled on the extremes.

Demographics and Initial Screening

Patients were asked about their age, sex, ethnic origin, studies, marital status, employment situation, number of members in the family unit, income level, smoking habits, alcohol intake habits, menopause stage, dominant side, operation side, and cancer history in the family.

Ongoing medical conditions were assessed and checked with medical history: hypertension, dyslipidemia, diabetes, and cardiovascular diseases. Weight (kg), body fat mass (FM; kg), percentage of body fat (PBF, %), visceral fat area (VFA, cm^2), and body mass index (BMI; kg/m^2) were estimated using an InBody 720 impedanciometer, which provides reliable results (McLester, Nickerson, Kliszczewicz, & McLester, 2020). Hypertension was assessed using the clinically validated OMRON M3 (HEM-7200-E2 (V)). Treatment details during the intervention were also noted if they were receiving chemotherapy, radiotherapy, or still waiting for surgery (treatment-naive).

Quality of Life

Quality of life was assessed with the gold standard European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire (QLQ-C30) (Giesinger et al., 2016). The QLQ-C30 computes several items addressing different health facets: physical, tasks, emotional, cognitive, social, fatigue, nausea, pain, dyspnoea, insomnia, appetite, constipation, diarrhea, and economics. A combination of all scores may compute a total value for global health.

Physical Activity

Daily physical activity levels were monitored with a Fitbit Inspire HR, counting steps, calories, METs (metabolic equivalent of task), and the intensity of the activity recorded, as well as the timing for each variable.

Moreover, patients evaluated their baseline physical activity levels in a week by filling the gold standard International Physical Activity Questionnaire (IPAQ) (Wanner et al., 2016). The IPAQ assigns different MET values for the minutes dedicated to an activity a predefined number of days. These activities are reduced to three levels of intensity: walking, moderate, and vigorous. These levels ultimately may compute a total MET score.

$$\begin{aligned} METs/week &= WalkMETs + ModMETs + VigMETS = \\ &= 3.3 \cdot walkmin/week \ + \ 4.0 \cdot modmin/week + \\ &\quad + \ 8.0 \cdot vigmin/week \end{aligned} \tag{4.3}$$

These levels enable physical activity classification into low, moderate, and vigorous, as described elsewhere (Wanner et al., 2016).

Training Sessions

Training sessions aimed for determined intensities depending on the status of the patients. Physical therapists directed training sessions, double-checking the intensity given with the patient's sensations during and at the end of every session. Post-exercise BORG intensities (Soriano-Maldonado et al., 2015) were recorded for every session.

4.3 Data Records

4.3.1 Data Availability

Study data were stored and made openly available at Zenodo (https://zenodo.org/record/6322773). Anonymity was ensured by changing original IDs and removing dates from medical appointments (e.g., training, chemotherapy, and radiotherapy sessions). All data records were stored in three comma-separated values (CSV) files, described below.

4.3.2 Data Description

Aggregated

Demographics, initial screening, QLQ-C30, and IPAQ data were stored in the file `demographics.csv`. Table 4.1 describes the coding of demographics and initial screening, and Table 4.2 describes the coding of QLQ-C30 and IPAQ data. A visual summary of demographic and initial screening data is illustrated in Figures 4.3, and 4.4. Figure 4.5 describes QLQ-C30 and IPAQ data.

ATOPE+ data (HRV, self-reported wellness, Fitbit's sleep and physical activity, and training session) were stored in `records.csv`. HRV and self-reported data were previously described in Table 3.1. Training data, SWC and CV variables, and daily aggregated Fitbit variables of sleep and steps are described in Figure 4.6, and self-reported wellness with Fitbit steps and Fitbit sleep in Figure 4.7.

Detailed physical activity Fitbit data were stored in `fitbit_activity.csv`, including the steps, mets, calories and time in each intensity level of physical activity detected. Table 4.4 describes the coding of Fitbit physical activity variables. Figure 4.8 describes the sample.

Tab. 4.1.: Demographic and initial screening features in `demographics.csv`.

Variable	Unit	Type	Description
age	years	int	Age.
sex	category	string	Sex.
ethnic_origin	category	string	Ethnic origin.
studies	category	string	Level of studies.
marital_status	category	string	Marital or civil status.
employment_situation	category	string	Employment situation.
n_family_unit	members	int	Number of members in the family unit.
income	category	string	Level of income.
smoking	category	string	Smoking habits.
alcohol	category	string	Alcohol consumption habits.
menopause	category	string	Menopausal status.
dominant_side	category	string	Left- or right-handed.
operation_side	category	string	Side for breast surgery.
cancer_history_family	category	string	History of cancer in family.
breast	category	string	Breast cancer in family.
colon	category	string	Colon cancer in family.
prostate	category	string	Prostate cancer in family.
lung	category	string	Lung cancer in family.
bladder	category	string	Bladder cancer in family.
ovary	category	string	Ovary cancer in family.
stomach	category	string	Stomach cancer in family.
intestine	category	string	Intestine cancer in family.
throat	category	string	Throat cancer in family.
thyroid	category	string	Thyroid cancer in family.
leukemia	category	string	Leukemia in family.
blood	category	string	Blood cancer in family.
hypertension	category	string	Hypertension in medical conditions.
dyslipidemia	category	string	Dyslipidemia in medical conditions.
diabetes	category	string	Diabetes in medical conditions.
cardiovascular_disease	category	string	Cardiovascular disease in medical conditions.
height	cm	int	Height.
weight	kg	float	Weight.
bmi	kg/m^2	float	Body mass index (BMI).
body_fat_mass	kg	float	Fat mass.
body_fat_percentage	%	float	Percentage of body fat.
visceral_fat_area	cm^2	float	Visceral fat area.
treatment	category	string	Treatment during intervention.

Tab. 4.2.: QLQ-C30 and IPAQ features in `demographics.csv`.

Variable	Unit	Type	Description
qlqc30_physical	%	float	Physical activity score.
qlqc30_tasks	%	float	Tasks score.
qlqc30_emotional	%	float	Emotional score.
qlqc30_cognitive	%	float	Cognitive score.
qlqc30_social	%	float	Social score.
qlqc30_fatigue	%	float	Fatigue score.
qlqc30_nausea	%	float	Nausea score.
qlqc30_pain	%	float	Pain score.
qlqc30_dyspnoea	%	float	Dyspnoea score.
qlqc30_insomnia	%	float	Insomnia score.
qlqc30_appetite	%	float	Appetite score.
qlqc30_constipation	%	float	Constipation score.
qlqc30_diarrhoea	%	float	Diarrhoea score.
qlqc30_economic	%	float	Economic score.
qlqc30_global_health	%	float	Global health score.
ipaq_walking	METs/week	int	METs a week for walking-like activities.
ipaq_moderate	METs/week	int	METs a week for moderate activities.
ipaq_vigorous	METs/week	int	METs a week for vigorous activities.
ipaq_sitting	METs/week	int	METs a week for sitting-like activities.
ipaq_total_met	METs/week	int	Total METS a week.
physical_activity_level	category	string	IPAQ level of physical activity.

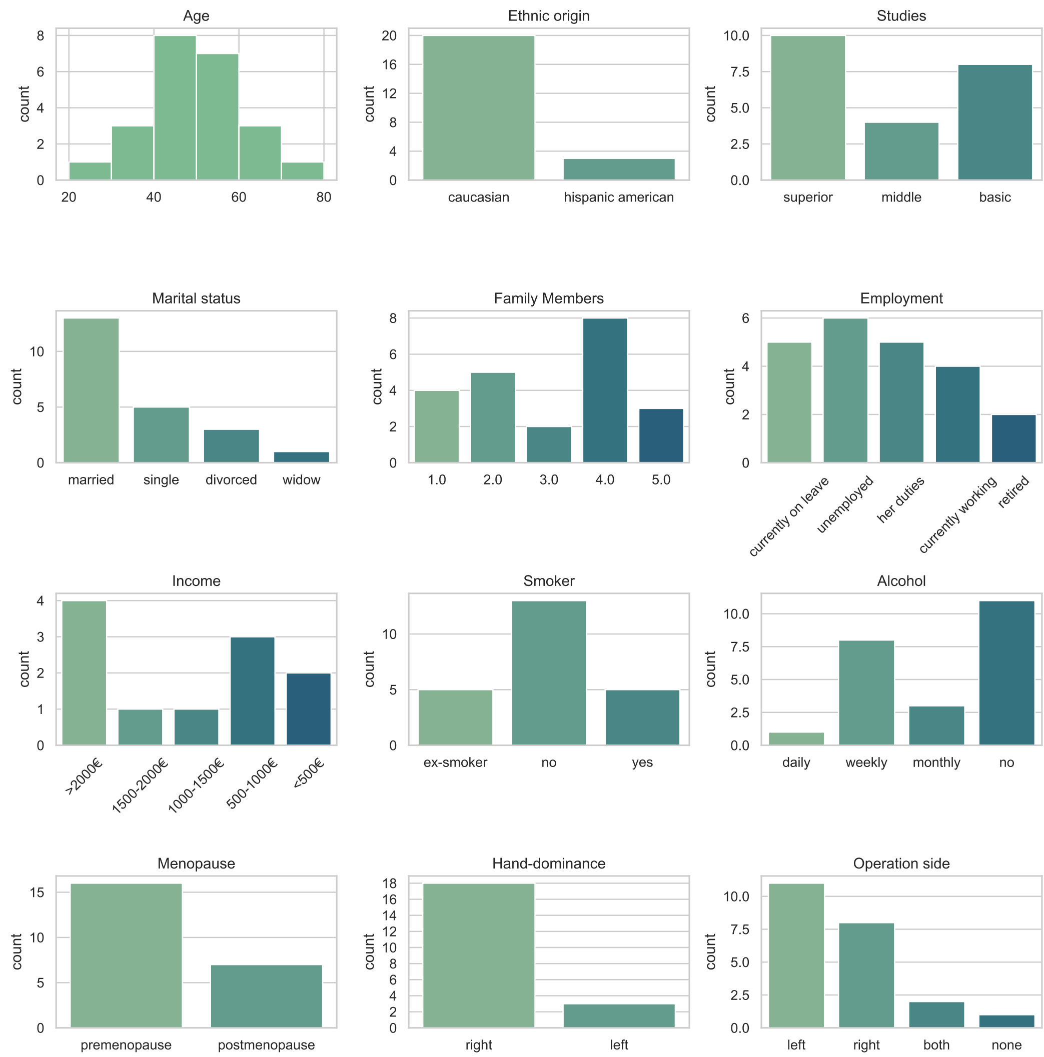


Fig. 4.3.: Description of participants (1/2).

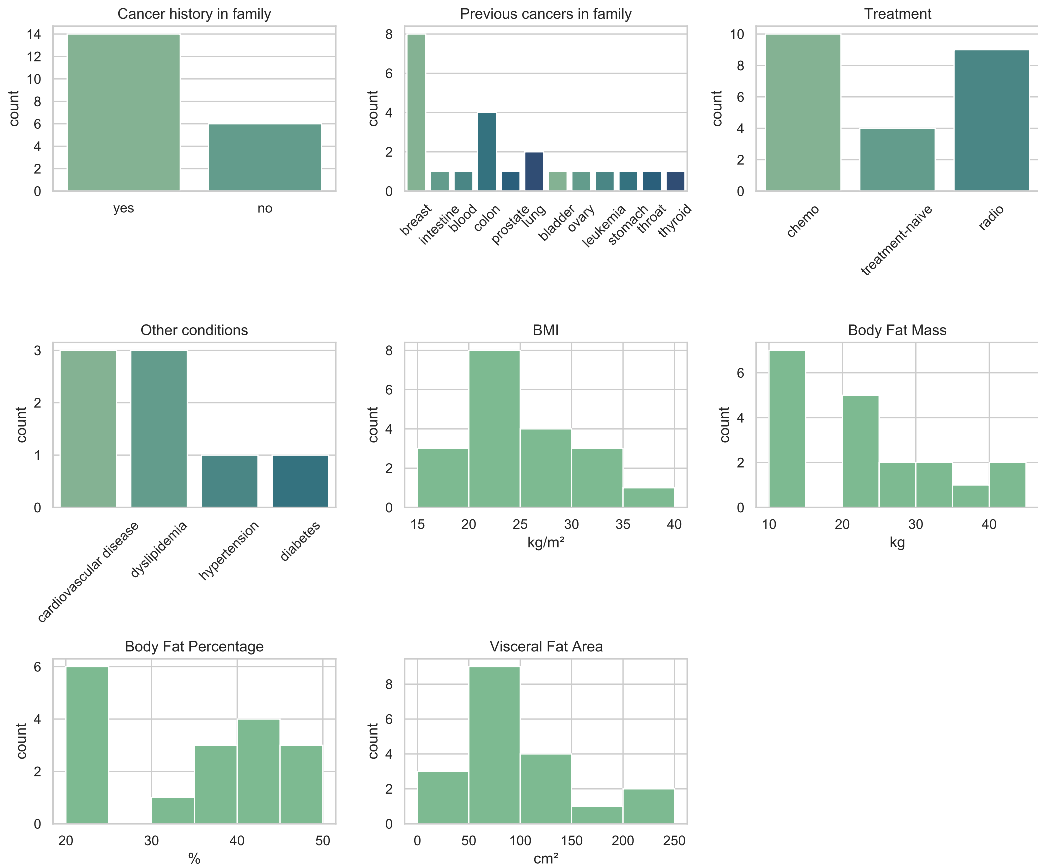


Fig. 4.4.: Description of participants (2/2).

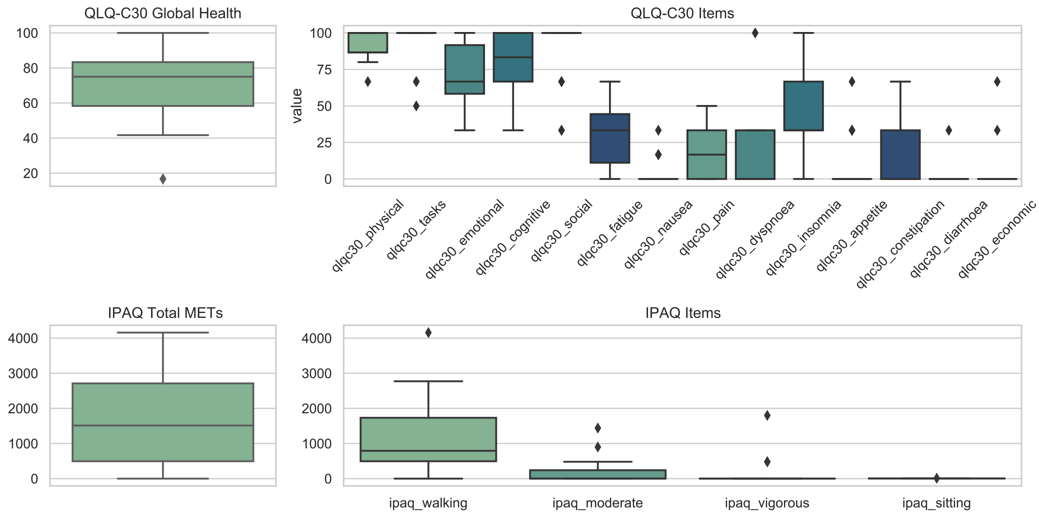


Fig. 4.5.: Initial physical activity (IPAQ) and quality of life (QLQ-C30) data for participants.

Tab. 4.3.: Training data, Fitbit steps, and Fitbit sleep stored in `records.csv`. The label n.u. is for *normalized units*. Fitbit sleep data may be categorized in *classic* or *stages*. *Classic* sleep data are retrieved from Fitbit API during the first days of monitoring; once Fitbit has enough sleep, it infers sleep *stages*, replacing the *classic* categories.

Variable	Unit	Type	Description
study_day	days	int	Study day.
session_type	category	string	Type of training session or relevant event.
session_number	session	int	Training session number.
training_borg	points	float	Post-training BORG intensity.
swc_lnrmssd_ok		bool	True if LnRMSSD inside SWC thresholds.
cv_lnrmssd_ok		bool	True if CV below defined threshold.
swc_lnrmssd	n.u.	float	Normalized SWC of the LnRMSSD.
cv_lnrmssd	%	float	Coefficient of variation of the LnRMSSD.
sleep_asleep	seconds	int	Asleep sleep time (classic).
sleep_awake	seconds	int	Awake sleep time (classic).
sleep_restless	seconds	int	Restless sleep time (classic).
sleep_deep	seconds	int	Deep sleep time (stages).
sleep_light	seconds	int	Light sleep time (stages).
sleep_rem	seconds	int	REM sleep time (stages).
sleep_unknown	seconds	int	Unknown sleep time (stages).
sleep_wake	seconds	int	Wake sleep time (stages).
sleep_total	seconds	int	Total sleep time.
steps_sedentary	steps	int	Sedentary steps.
steps_light	steps	int	Light-intensity steps.
steps_moderate	steps	int	Moderate-intensity steps.
steps_vigorous	steps	int	Vigorous-intensity steps.
steps_total	steps	int	Total steps.

Tab. 4.4.: Fitbit physical activity data by intensity level (*light*, *moderate*, *sedentary*, *vigorous*) and in total in `fitbit_activity.csv`.

Variable	Unit	Type	Description
intensity	category	string	Level of intensity detected.
steps	steps	int	Number of steps detected.
mets	METs	float	Number of METs inferred.
calories	cal	float	Number of calories inferred.
activity_time	seconds	int	Time in each intensity level.

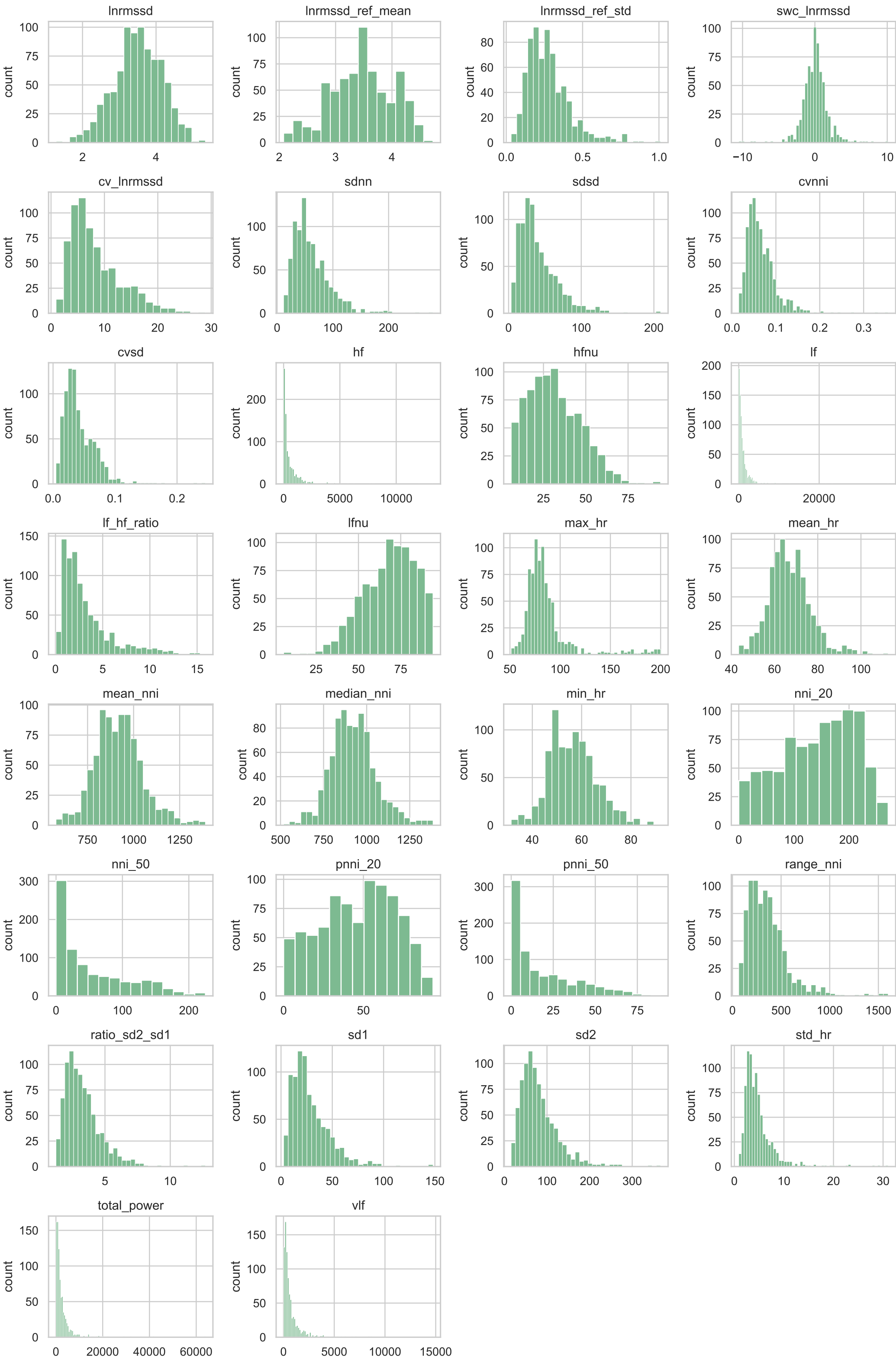


Fig. 4.6.: Description of HRV data.

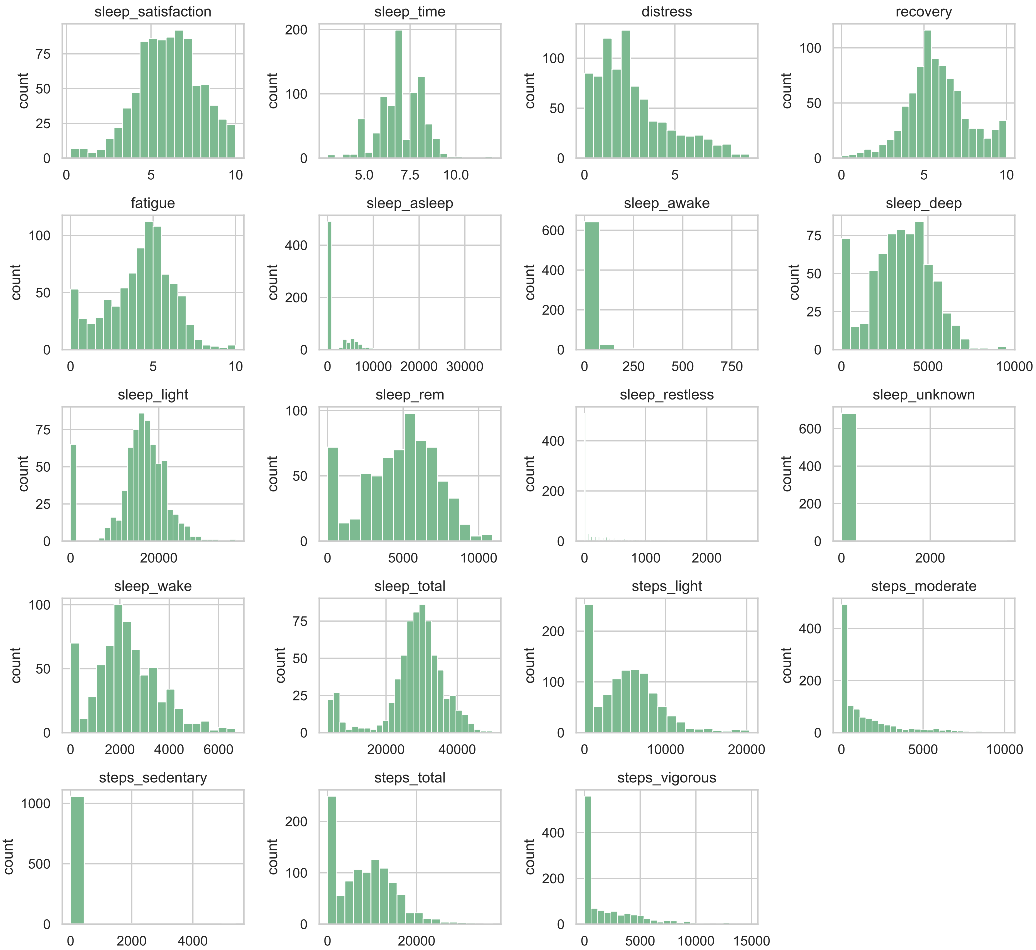


Fig. 4.7.: Description of self-reported wellness and Fitbit data.

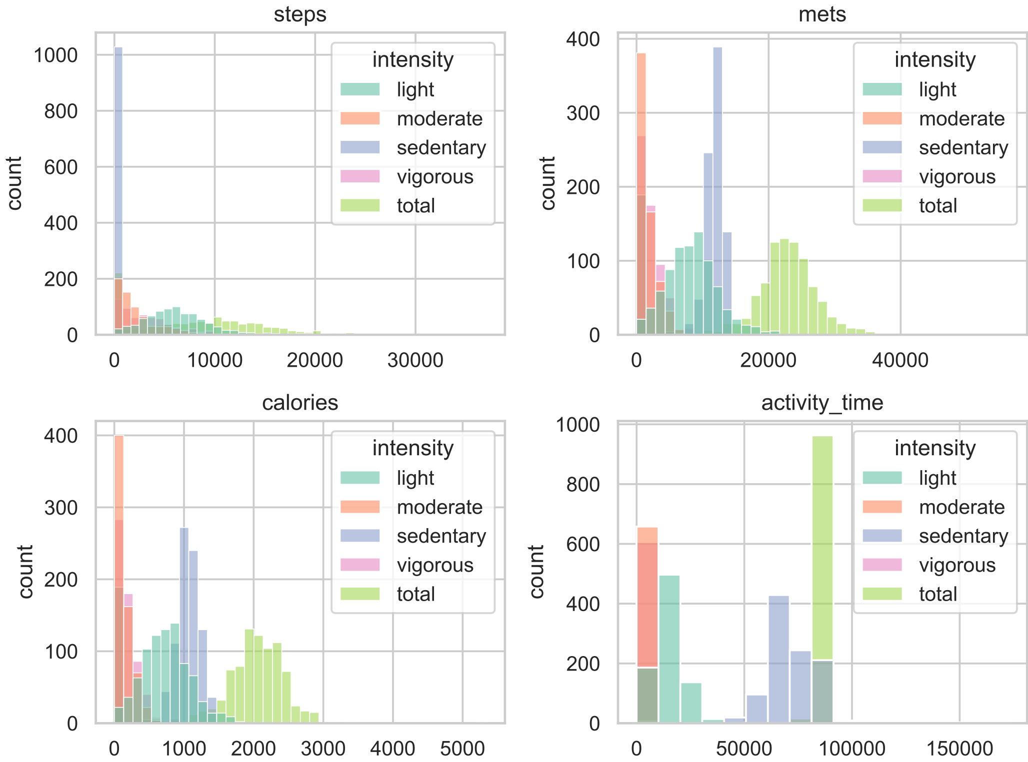


Fig. 4.8.: Description of Fitbit activity data by intensity level.

4.3.3 Longitudinal Analysis

The dataset provides means for individual longitudinal visualizations. Figure 4.9 draws an example for patient AT86. The remaining figures are attached as supplementary material in Appendix B.

4.4 Limitations and Future Work

The first limitation of the ATOPE+Breast dataset is the restricted sample size, out of which only preliminary results may be drawn. The COVID-19 restrictions impeded a faster recruitment of patients in the past two years. Future versions of the ATOPE+Dataset will extend the sample with data from more patients as the ATOPE trial continues. Second, the interpretation of the SWC used to decide the intensity of exercise recommendations is slightly different from the SWC methods found in sport. Future work will assess the validity of such SWC interpretation when the ATOPE trial is finished and post-intervention results are available.

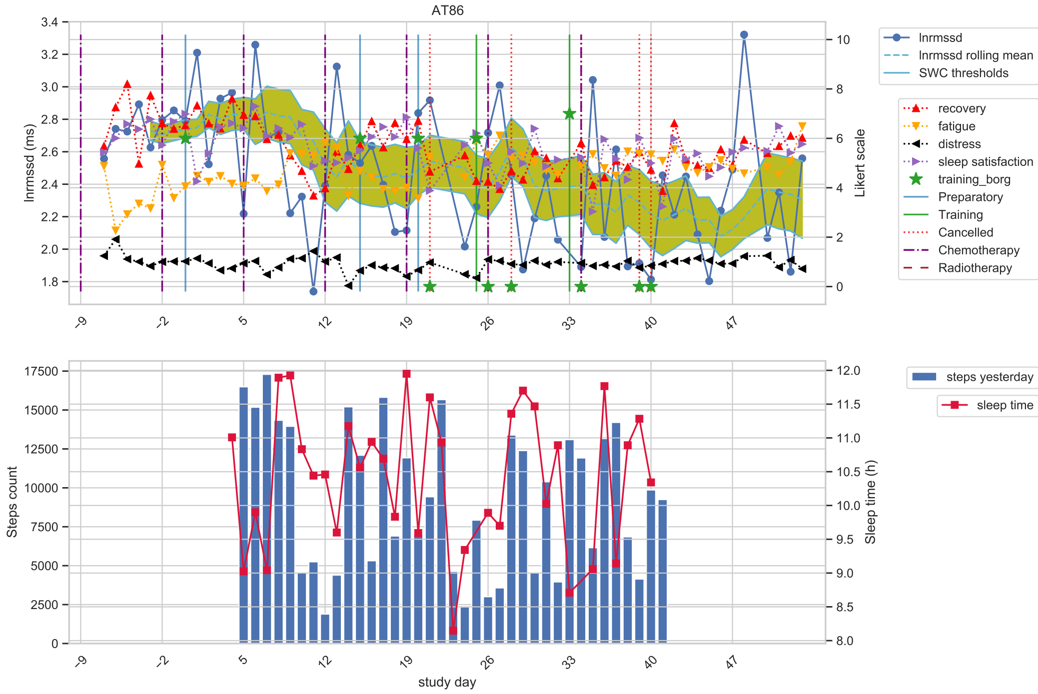


Fig. 4.9.: Longitudinal visualization of selected data for AT86.

4.5 Ethics Declarations

This study has the approval of the Biomedical Research Ethics Committee (Granada, Spain) (0507-N-18, July 27, 2018). Prior to the start of the study, all participants received written and verbal information and provided their voluntary written informed consent in conformity with the Declaration of Helsinki.

5 A Clustering Approach to Assess Training Needs in Patients with Breast Cancer

Preceding Note

This chapter aims to contribute with a novel approach to assess training needs in patients with breast cancer by leveraging data science and AI algorithms. Despite the possibilities of these tools, and in order to ensure their reliability and high quality, the nature of unsupervised learning algorithms (e.g., clustering) requires results to be contrasted by experts with their finest interpretation and discussion. Therefore, the crafting and interpretation of the results presented in this chapter were conducted along with Irene Cantarero-Villanueva (PhD), Paula Postigo-Martín (MSc), and Manuel Arroyo-Morales (PhD, MD), all of them experts in therapeutic exercise in patients with cancer, and with affiliation to the Department of Physical Therapy and the Unit of Excellence on Exercise and Health (UCEES) in the University of Granada.

All the results can be run with the code available in GitHub (https://github.com/salvador-moreno/atope-breast-clustering-analysis).

5.1 Introduction

Machine learning (ML) algorithms have provided numerous advances in precision medicine. ML is a subset of artificial intelligence that enables the building of models based on sample data (or training data) in order to make predictions or assist in decisions. For instance, ML has provided means to predict severity symptoms in patients with cancer out of context-monitoring data (Carissa A. Low et al., 2017; Carissa A. Low et al., 2021); to enhance tumor diagnostic capabilities through automatic image analysis (D'Amore, Smolinski-Zhao, Daye, & Uppot, 2021); or to remotely monitor and diagnose chronic conditions (Castelyn, Laranjo, Schreier, & Gallego, 2021).

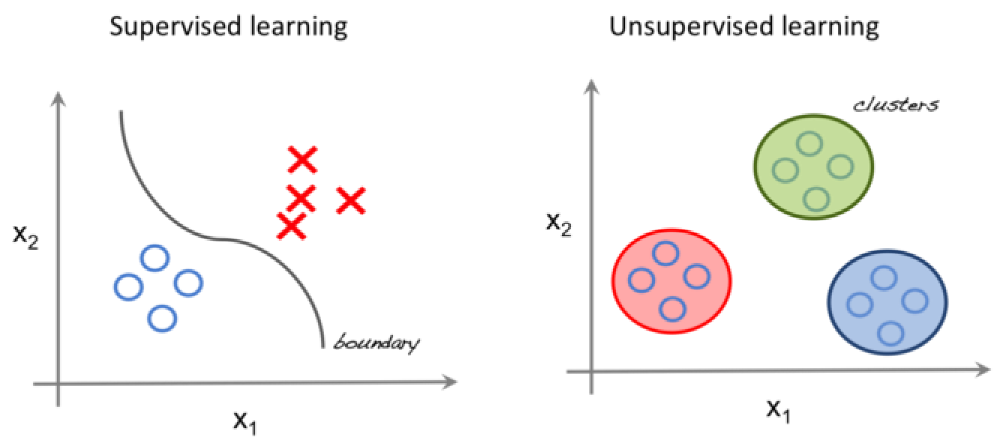


Fig. 5.1.: Supervised vs. unsupervised learning. In this example, the objective of the supervised learning algorithm is to build a model able to discern between the data points manually labeled as *circles* and *crosses*. For that, the model computes a *boundary* in the space domain of the input variables. Conversely, the unsupervised learning algorithm seeks for non-described patterns in the data, namely, sufficiently separated groups of data to be considerate separate clusters. (source: https://tinyurl.com/2w5f24ak).

Supervised and unsupervised learning are the most common approaches to leverage ML algorithms. Supervised learning trains a model with a training set of data that assigns a specific set of inputs to the desired output to predict or classify. For instance, an ML algorithm may be trained to find if a magnetic resonance image (MRI) contains a tumor when trained with a sufficient number of MRI labeled with tumors and their locations. In other words, in a supervised learning approach, an ML model is explicitly told what to learn out of a specific set of data. This output may be a continuous variable (a regression problem) or a categorical variable with a fixed number of labels (a classification problem). Logistic regression, decision-tree-based, support vector machine (SVM), and deep learning algorithms are ML algorithms typically used for supervised learning.

Conversely, unsupervised learning algorithms provide a more exploratory approach. Unsupervised learning seeks for patterns, structures, or clusters among the input data, without any explicit target or indication besides the configuration of the algorithm used and the codification of input data. Association rules, clustering techniques, and dimensionality reduction (e.g., principal component analysis) are typical examples of unsupervised learning algorithms. Figure 5.1 provides a visual example of supervised vs. unsupervised learning.

The evolution of training load parameters (like the LnRMSSD) during training adaptation has been widely studied in professional athletes (Plews et al., 2012;

Plews et al., 2013; Buchheit, 2014; Javaloyes et al., 2019). Nevertheless, little is known about the training adaptation process in patients with breast cancer while undergoing a therapeutic exercise intervention (Carter et al., 2021). The objective of this chapter was to analyze the data presented in the previous chapter with an unsupervised learning approach in order to seek for training adaptation patterns during an exercise intervention in patients with cancer. Specifically, this chapter uses a clustering approach to find different profiles training adaptation out of the most features found in the dataset.

This chapter is structured as follows. Section 5.2 describes the methods used for analyzing the ATOPE+Breast dataset. Next, section 5.3 reports the results delivered by data cleaning, feature selection, and clustering analysis processes. Finally, section 5.4 discusses the principal findings, practical implications, and outlook.

5.2 Materials and Methods

5.2.1 ATOPE+Breast Dataset

The data analyzed was the ATOPE+Breast dataset, described in Chapter 4. Specifically, the analysis focused on 681 instances containing complete HRV measurements and self-reported wellness from 23 patients. In addition, part of the analysis focused on 488 measures containing Fitbit steps for the previous day and on 328 measures also including both Fitbit steps and sleep. The analysis was preceded by data cleaning and preprocessing methods described in subsection 5.2.3.

5.2.2 Clustering Algorithms

Data clustering is the unsupervised classification of samples into groups. In other words, clustering algorithms provide a means to group similar samples into one group called *cluster*. Each cluster has a maximum within-cluster similarity and a minimum between cluster similarity based on certain indexes that depend on the algorithms used (Saxena et al., 2017).

In order to allow the clustering algorithms to identify relevant clusters correctly, two basic principles must be taken into account. First, the clustering algorithm must be selected according to the nature of the data used (e.g., integer/float numbers, categorical variables, images), the relationship among clusters desired (i.e., partitional or hierarchical), and how the data may be distributed among the clusters (e.g., into small/big clusters, different geometries, dense/sparse clusters, or a combination of such conditions). Second, finding clusters in high-dimensional spaces (i.e., with several features) is computationally expensive and may degrade learning performance. Therefore, a feature selection process should precede the training of clustering models. Moreover, although data-driven methods may select the features feeding the clustering algorithm (e.g., maximum correlation), it is highly advisable to include expert knowledge in the process —especially in health applications— to enable meaningful discussions and interpretations of the results achieved (Alelyani, Tang, & Liu, 2018). Both approaches were combined into the feature selection process of the data used.

For clustering analysis, the general-purpose K-means clustering algorithm (Saxena et al., 2017) was used. The centroid initialization algorithms *k-means++* was used to ensure the best centroid initialization for clustering (Arthur & Vassilvitskii, 2007). Clustering algorithms were run using scikit-sklearn (0.24.2) over Python 3.6.9.

K-Means Clustering

The K-Means algorithm is the most known and benchmarked clustering algorithm available (Saxena et al., 2017) to explore groups of data in a given dataset automatically. The K-Means algorithm clusters the data by separating samples in k groups of equal variance by minimizing a criterion known as the inertia or within-cluster sum-of-squares (described below). Although K-Means requires the number of clusters k to be manually specified, it scales well to a large number of samples and has been used in a wide range of application areas in several fields.

Let x_i be any observation in $X \subset \mathbf{R}^m, m \in \mathbf{N}$, with $i \in [1, n], n \in \mathbf{N}$; and μ_j the centroids of clusters $C \subset \mathbf{R}^m$, with $j \in [1, k], k \leq n$. The K-Means algorithm divides the set of n observations of X into k disjoint clusters. Each one of these clusters is described by the mean of the samples (μ_j) conforming the cluster. These means are called *centroids*. Usually, these centroids are not present in X, although they share the same space. A flow diagram of the K-Means algorithm is pictured in Figure 5.2. The K-means algorithm aims to choose centroids that minimize the inertia, or within-cluster sum-of-squares criterion

$$\sum_{i=1}^{n} \min_{\mu_j \in C}(||x_i - \mu_j||^2) \tag{5.1}$$

Inertia is commonly interpreted as a measure of how internally coherent clusters are. However, it suffers from some drawbacks. First, inertia assumes that clusters are convex and isotropic, therefore this measure responds poorly to elongated clusters or manifolds with irregular shapes. The methods described in the next section aim to lower the effect of this limitation. Second, inertia is not a normalized metric. Lower inertia values are better, with zero being optimal, but inertia does not allow comparison when using different datasets or even different feature sets.

5.2.3 Data Cleaning and Preprocessing

According to the *No Free Lunch* concept (Wolpert & Macready, 1997), no algorithm can be good under all circumstances. Each algorithm has its merit under some specific data natures but fails on others. K-means uses distance-based measurements to determine the similarity between data points; therefore, in order to avoid introducing noise and skew into the clustering process, the data input needs to be as clean as possible.

K-means is very sensitive to outliers and noisy data; hence the first step for cleaning was to remove outliers. This step is crucial since extreme outliers may impact centroid calculation and cluster shapes. The detection of outliers was done using interquartile range (IQR) for univariate distributions. Let $x \in X \subset \mathbf{R}$, where X is one of the continuous variables of the dataset:

$$outlier(x) = \begin{cases} True, & \text{if } x > Q3(X) + 1.5 \cdot IQR(X) \\ True, & \text{if } x < Q1(X) - 1.5 \cdot IQR(X) \\ False, & \text{otherwise} \end{cases} \tag{5.2}$$

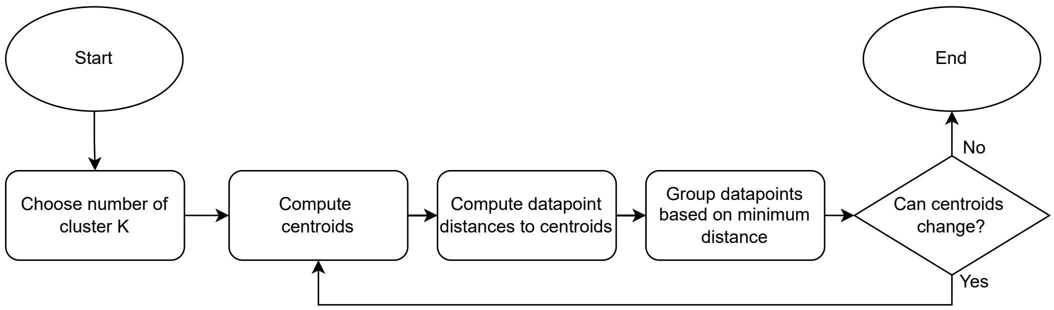


Fig. 5.2.: Flow diagram of the K-Means algorithm.

where $Q1$ and $Q3$ are the first and third quartile functions, and IQR is the interquartile range function, $IQR(X) = Q3(X) - Q1(X)$.

Outliers were replaced with a k-nearest-neighbors (KNN) model (Aggarwal, 2015). KNN algorithm looks for similar instances within the dataset to infer a new value to replace the outliers.

To further help find isometric clusters, data were transformed into normal-like distributions using logarithmic and power operations for left-skewed and right-skewed distributions, respectively. Skewness thresholds greater than 1 and lower than -1 were set to determine if transformations were needed (Gravetter & Wallnau, 2014). These transformations resulted in more isometric distributions, with more symmetry around the means. Moreover, since IQR outlier detection assumes normality of the distribution, these transformations were done before outlier detection and replacement.

In order to assign the same importance to all variables during the computing of distances, we standardized the data by removing the mean and scaling to unit variance (also known as z-score normalization). Let x be an observation in $X \subset \mathbf{R}$, where X is one of the continuous variables of the dataset, and μ, σ the mean and variance functions, respectively:

$$z(x) = \frac{x - \mu(X)}{\sigma(X)} \tag{5.3}$$

Finally, variance and correlation analysis were used for feature selection. A maximum variance threshold was defined in order to ensure a variable could provide enough information to the algorithm. Let x be an observation in $X \subset \mathbf{R}$, where X is one of the continuous variables of the dataset, and $p \in [0, 1]$:

$$Var(X) = p(1 - p) \tag{5.4}$$

Variance threshold was set for p=0.8, which, broadly, means that at least 20% of the observations in X should differ enough from the remaining 80%. Those who did not meet the criteria were removed. Highly correlated variables ($r > 0.8$) were also discarded in order to feed the clustering algorithm with the least amount of redundant information. Gold-standard variables for training conditioning like LnRMSSD were selected over the rest when removing highly correlated variables.

Data cleaning involved interpreting missing data and zeroes, further explained in the results section for each variable assessed. Data cleaning and preprocessing used pandas (1.1.5) and numpy (1.19.5) over Python 3.6.9. Visualizations were made with matplotlib (3.3.4), pandas (1.1.5), and seaborn (0.11.2).

5.2.4 Clustering Validation

The formation of clusters is important, but it is also necessary to assess their quality and validity. Several evaluation criteria have been developed (Saxena et al., 2017; Aggarwal, 2015), all of them falling into internal and external validation criteria, depending on the reference taken.

Internal validation criteria examine the clustering structure directly from the original data, not taking into account any prior knowledge. These methods are very dependent on the clustering algorithms used. Nevertheless, most of them are based on the concept of testing how similar are the objects conforming a cluster and how different and distanced those clusters are. Conversely, external validation criteria are based on some pre-defined structure, knowledge, or ground truth about the data, hence requiring interpretation to validate them.

For this analysis, the internal validation criteria used was the silhouette score (Rousseeuw, 1987) to measure the coherence of the clusters found. The silhouette coefficient is an internal clustering quality method that measures how similar an object is to their own cluster compared to other clusters. Silhouette coefficient ranges from -1 to $+1$. The higher (positive) the value, the better the objects are matched to its own cluster; lower (negative) values indicate wrong clustering. Nonetheless, despite finding a high value of silhouette score for a determined number of clusters, post-hoc analysis is required to determine its accuracy when representing the reality of the problem addressed. Silhouette coefficient may be calculated with any distance metric like Euclidean or Manhattan distance. The results presented used Euclidean distance.

External validation relied primarily on analyzing the feature values displayed in the clusters against the state of the art in the Discussion (section 5.4). Nevertheless, the different clustering results for HRV and Wellness were compared with confusion matrices to assess their coherence. The comparison among clusters with a confusion matrix (Aggarwal, 2015) involved comparing clustering applied to HRV, baseline wellness, and z-scored wellness separately, along with the interpretation of the feature values displayed for every cluster.

5.2.5 Feature Importance Analysis for Clustering

Clustering models assign objects to a cluster based on its distance to the centroids of the clusters. All the features participate in this process; however, it is often difficult to assess which features are the most relevant to assign an object to a cluster. The most common approach is to use supervised learning with the results obtained after clustering, as shown in Figure 5.3. After training a clustering model with the data, the resulting dataset contains all entries labeled with the cluster assigned to each object. These labels may be used as a target to learn for a supervised learning algorithm (e.g., linear regression, random forests, support vector machines). This approach enables the analysis of feature importance for the classification model trained.

For this analysis, and for the sake of interpretability, feature importance analysis used random forests with bootstrapping as classifiers to train the models. Classifiers were trained with 5-Fold Cross-Validation. Finally, feature importance was assessed by accounting which features provided the highest reduction of *Gini impurity* (Aggarwal, 2015).

Training Supervised Classification Models

Random forests are a combination of decision tree classifiers in which randomness has explicitly been inserted into the building process of each decision tree. Random forests are done by selecting different variables and bootstrapping the dataset at the moment of building each one of the trees composing the model. These mechanisms ensure a low correlation between the different decision trees conforming the model, hence making random forests robust to errors, outliers, and overfitting (Aggarwal,

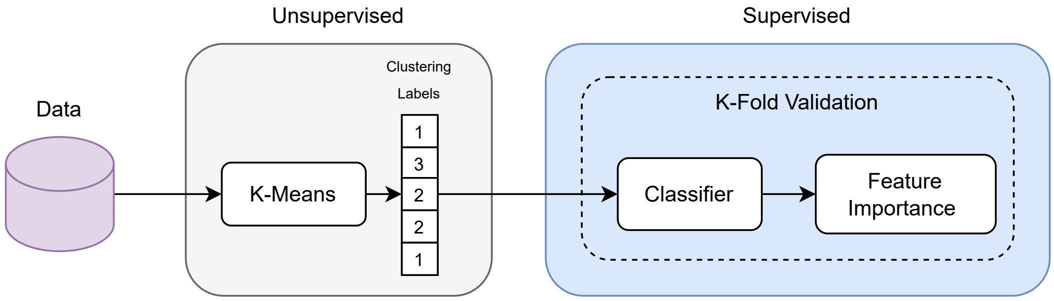


Fig. 5.3.: Feature importance analysis of clustering results with a supervised approach.

2015). Once the random forest model is built, predictions are computed as the mode of the outputs delivered by every decision tree.

Decision-tress predict the value of a target variable by learning linear rules inferred from the data features. The combination of multiple decision trees into a random forest enables non-linear combinations of the variables used to predict the target.

To ensure the robustness of feature importance analysis, 5-Fold Cross-Validation (CV) was used to train the supervised classifier. CV divides the dataset into N-1 parts for training the model and 1 for testing, using all data points for testing at least once. The CV process used in this analysis used N=5, thus dividing the dataset into 5 folds. Weighted F1-score was used to optimize and assess the quality of classification results in each iteration. F1-score is the harmonic mean of precision ($\frac{TP}{TP+FP}$, where TP is for true positive, and FP is for false positive) and recall ($\frac{TP}{TP+FN}$, where FN is for false negative):

$$F1\ score = 2\left(\frac{1}{precission} + \frac{1}{recall}\right)^{-1} = 2\frac{precission \cdot recall}{precission + recall} \tag{5.5}$$

Weighted F1-score computed a total score by weighting the F1-scores of every class with its number of instances relative to the total amount of data, which is recommended when classes are imbalanced.

Assessing Feature Importance

In each split of the trees conforming the Random Forests, the chosen features are the ones that maximize the reduction of pre-defined error criteria, such as Gini Impurity or Entropy (Aggarwal, 2015). This selection process ranks the features according to the reduction of error achieved in each split for every decision tree while building the model. Finally, in order to find the importance of each feature, this metric is averaged across all decision trees in the model. The analysis presented in this chapter used Gini Impurity.

Linear support vector machines (SVM) were also used to contrast unexpected feature analysis results (Aggarwal, 2015). Linear SVM can be used as classifiers dividing the search space with a linear kernel. The coefficients resulting from the training process indicate how much relevance each feature has to determine the mathematical space in which the cluster classes are contained.

5.3 Results

All the results presented can be drawn with the code published in GitHub (https://github.com/salvador-moreno/atope-breast-clustering-analysis).

5.3.1 Data Cleaning and Preprocessing

Skewness was calculated for all ATOPE+ variables, shown in Table 5.1, and visually inspected in order to decide transformations. A summary of the outlier detection and replacement process is described in Table 5.2. The following subsections describe in detail the cleaning process followed for HRV, wellness, and Fitbit data.

Tab. 5.1.: Skewness for HRV, wellness, and Fitbit variables. Values with absolute value higher than 1 are marked in bold.

	Skewness		Skewness
lnrmssd	–0.26	pnni_50	**1.09**
lnrmssd_ref_mean	–0.19	range_nni	**2.14**
lnrmssd_ref_std	**1.42**	ratio_sd2_sd1	**1.62**
swc_lnrmssd	–0.23	sd1	**1.79**
cv_lnrmssd	**1.14**	sd2	**1.93**
sdnn	**1.88**	std_hr	**3.76**
sdsd	**1.79**	total_power	**7.69**
cvnni	**2.74**	vlf	**5.75**
cvsd	**2.41**	sleep_satisfaction	-0.22
hf	**6.95**	sleep_time	-0.10
hfnu	0.60	distress	**1.03**
lf	**8.58**	recovery	0.24
lf_hf_ratio	**1.73**	fatigue	-0.44
lfnu	–0.60	steps_light_yesterday	0.26
max_hr	**2.66**	steps_moderate_yesterday	**1.37**
mean_hr	0.64	steps_total_yesterday	0.28
mean_nni	0.39	steps_vigorous_yesterday	**1.20**
median_nni	0.37	sleep_deep	0.00
min_hr	0.43	sleep_light	0.13
nni_20	–0.35	sleep_rem	-0.06
nni_50	0.95	sleep_wake	0.49
pnni_20	–0.21	sleep_total	0.19

Tab. 5.2.: Summary of outliers detected and replaced (absolute and percentage) for HRV, wellness, and Fitbit variables.

	Outliers	%		Outliers	%
lnrmssd	3	0.44	pnni_50	9	1.32
lnrmssd_ref_mean	0	0.00	range_nni	9	1.32
lnrmssd_ref_std	26	3.82	ratio_sd2_sd1	3	0.44
swc_lnrmssd	33	4.85	sd1	3	0.44
cv_lnrmssd	0	0.00	sd2	6	0.88
sdnn	3	0.44	std_hr	19	2.79
sdsd	3	0.44	total_power	6	0.88
cvnni	8	1.17	vlf	14	2.06
cvsd	5	0.73	sleep_satisfaction	6	0.88
hf	1	0.15	sleep_time	4	0.59
hfnu	4	0.59	distress	33	4.85
lf	9	1.32	recovery	15	2.20
lf_hf_ratio	3	0.44	fatigue	5	0.73
lfnu	0	0.00	sleep_deep	3	0.44
max_hr	50	7.34	sleep_light	7	1.03
mean_hr	17	2.50	sleep_rem	0	0.00
mean_nni	10	1.47	sleep_wake	12	1.76
median_nni	12	1.76	sleep_total	16	2.35
min_hr	8	1.17	steps_light_yesterday	29	4.26
nni_20	0	0.00	steps_moderate_yesterday	4	0.59
nni_50	4	0.59	steps_total_yesterday	14	2.06
pnni_20	0	0.00	steps_vigorous_yesterday	21	3.08

HRV

Skewness was computed for HRV variables (Table 5.1) and visually inspected with Figure 4.6. The variables `cv_lnrmssd`, `sdnn`, `sdsd`, `cvnni`, `cvsd`, `hf`, `lf`, `lf_hf_ratio`, `range_nni`, `ratio_sd2_sd1`, `sd1`, `sd2`, `std_hr`, `total_power`, `vlf`, and `lfnu` were selected for its transformation using logarithmic and square transformations (square only used for `lfnu`). Transformed HRV variables are shown in Figure 5.4. All variables except (`cv_lnrmssd` and `swc_lnrmssd`) were then inspected for outliers with IQR, and replaced using KNN (example shown in Figure 5.5, results in Table 5.2).

Since `cv_lnrmssd` and `swc_lnrmssd` were extracted from multiple `lnrmssd` values (time-dependent), KNN outlier replacement did not provide good results. Hence, `cv_lnrmssd` was only transformed, and its outliers were not removed nor replaced due to its dependency with `lnrmssd`. For `swc_lnrmssd`, it was not transformed due to its low skew, and its outliers were detected using IQR. Instead of using KNN

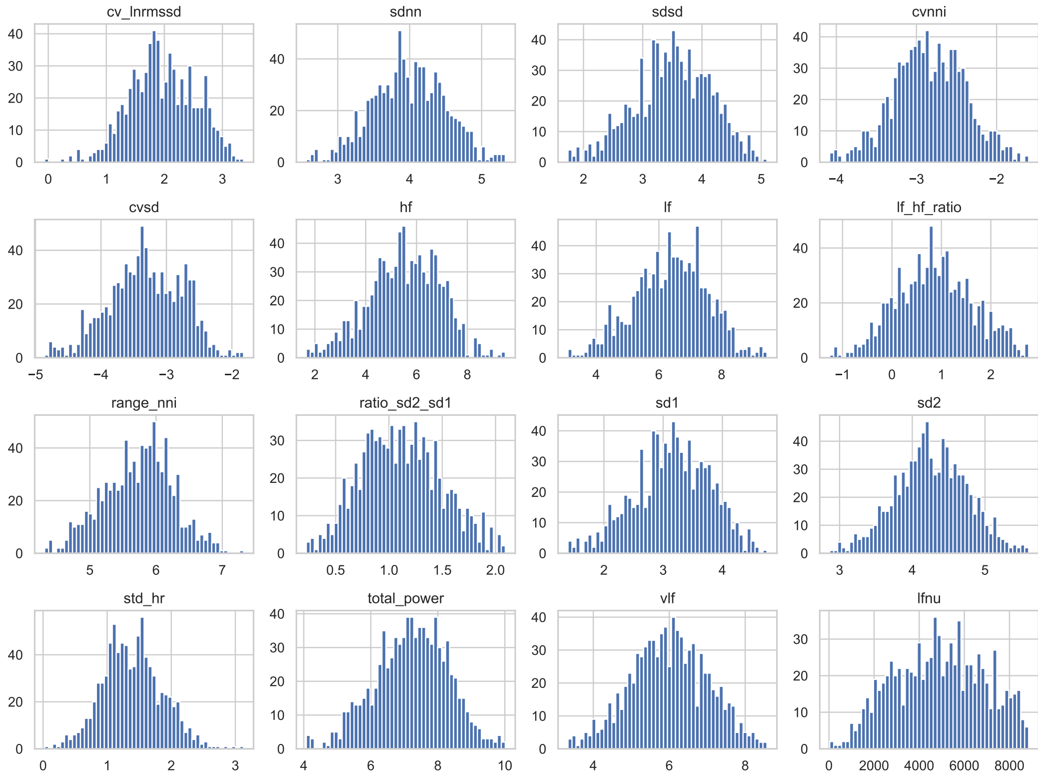


Fig. 5.4.: Transformed HRV variables.

for replacement, outliers were replaced with uniformly-distributed random values between 1x and 1.5x times the IQR distance below Q1, for lower values, and over Q3, for higher values (shown in Figure 5.6, results in Table 5.2).

Wellness

Skewness was calculated for wellness (Table 5.1) and visually inspected in Figure 4.7. Only `distress` presented high skewness and needed transformation. An anchoring effect was found for value 0 in `distress` and `fatigue`. In order to reduce the amount of bias in the clustering algorithm, zero values were uniformly replaced with uniformly-distributed random values within $(0, 0.5]$. The transformation and outlier replacement of `distress` is shown in Figure 5.7. The results for outlier detection and replacement are in Table 5.2.

Due to the subjective nature of perceived wellness, the distributions of self-reported wellness were expected to vary among individuals. These differences may hinder inter-individual comparisons and the extraction of generalized results. Thus, in

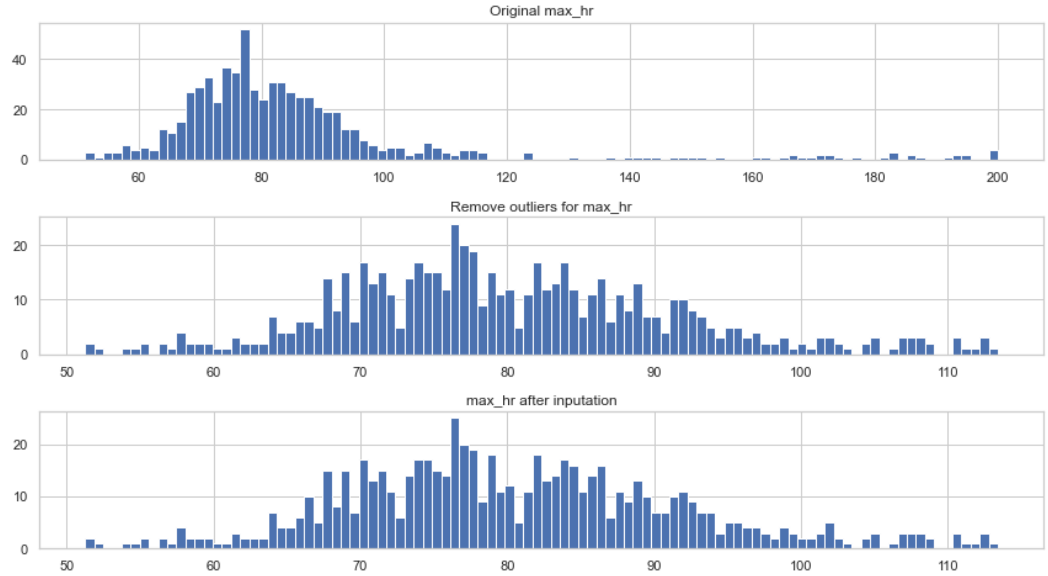


Fig. 5.5.: Outlier detection, removal, and imputation for `max_hr`.

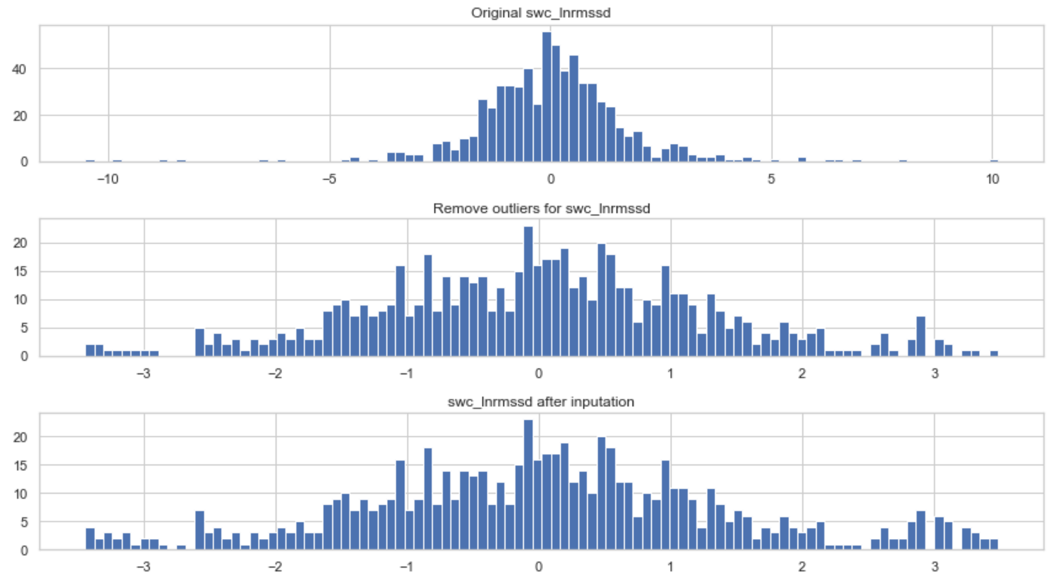


Fig. 5.6.: Outlier detection, removal, and imputation for `swc_lnrmssd`.

order to allow the search of comparative experiences among participants, we added individually z-scored wellness variables to our dataset (Figure 5.8), named adding "`_zscored`" at the end (e.g., `recovery_zscored`).

Fitbit Steps and Sleep

Fitbit steps variables were first shifted in date to align with the steps done the previous day (hence variables were renamed adding "`_yesterday`" in the end; e.g., `steps_vigorous_yesterday`). This change was made since the steps data available at the moment of HRV and wellness recordings were gathered the previous day. Fitbit sleep *classic* variables were removed from cleaning and analysis due to its high number of missing data. The reason for this is that the Fitbit API only provides such sleep classification during the first nights of sleeping (e.g., 4 nights); then, once their algorithm has sufficient data, it provides the more detailed *stages* labels.

For sleep, a high number of instances had value 0 for `sleep_deep` (70), `sleep_light` (65), `sleep_rem` (68), and `sleep_wake` (67). Entries with `sleep_total` or `sleep_light` equal to zero were discarded for analysis with Fitbit data. Zero values for `sleep_deep`, `sleep_rem`, and `sleep_wake` were replaced by NA to avoid skewing the clustering algorithm.

For steps, a high number of instances also had zero values for `steps_light_yesterday` (229), `steps_moderate_yesterday` (409), `steps_total_yesterday` (204), and `steps_vigorous_yesterday` (461). Entries with `steps_total_yesterday` equal to zero were discarded from analysis with Fitbit data; for the rest, zero values were replaced with NA.

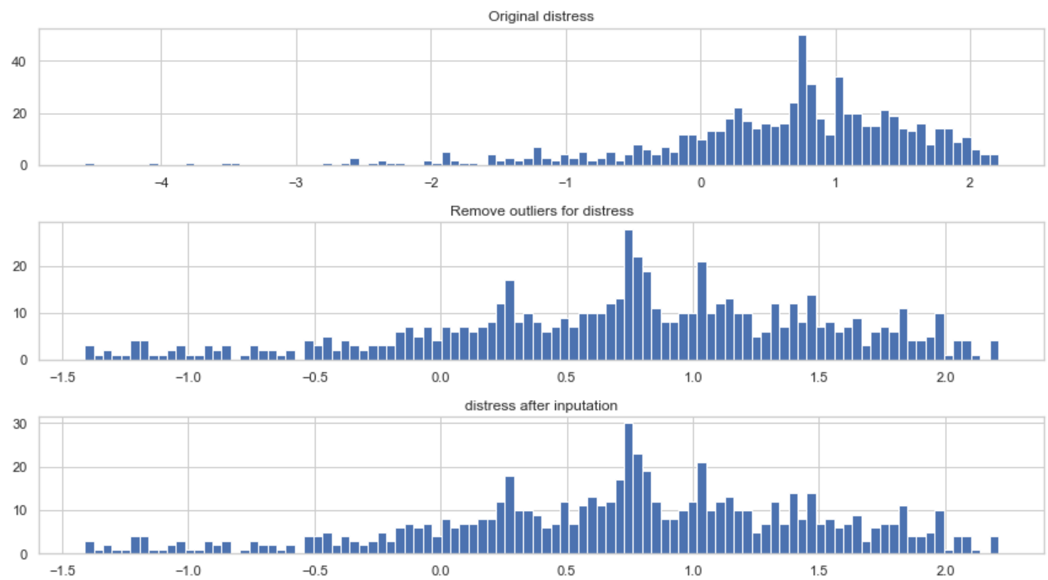


Fig. 5.7.: Outlier detection, removal, and imputation for `distress` after logarithmic transformation.

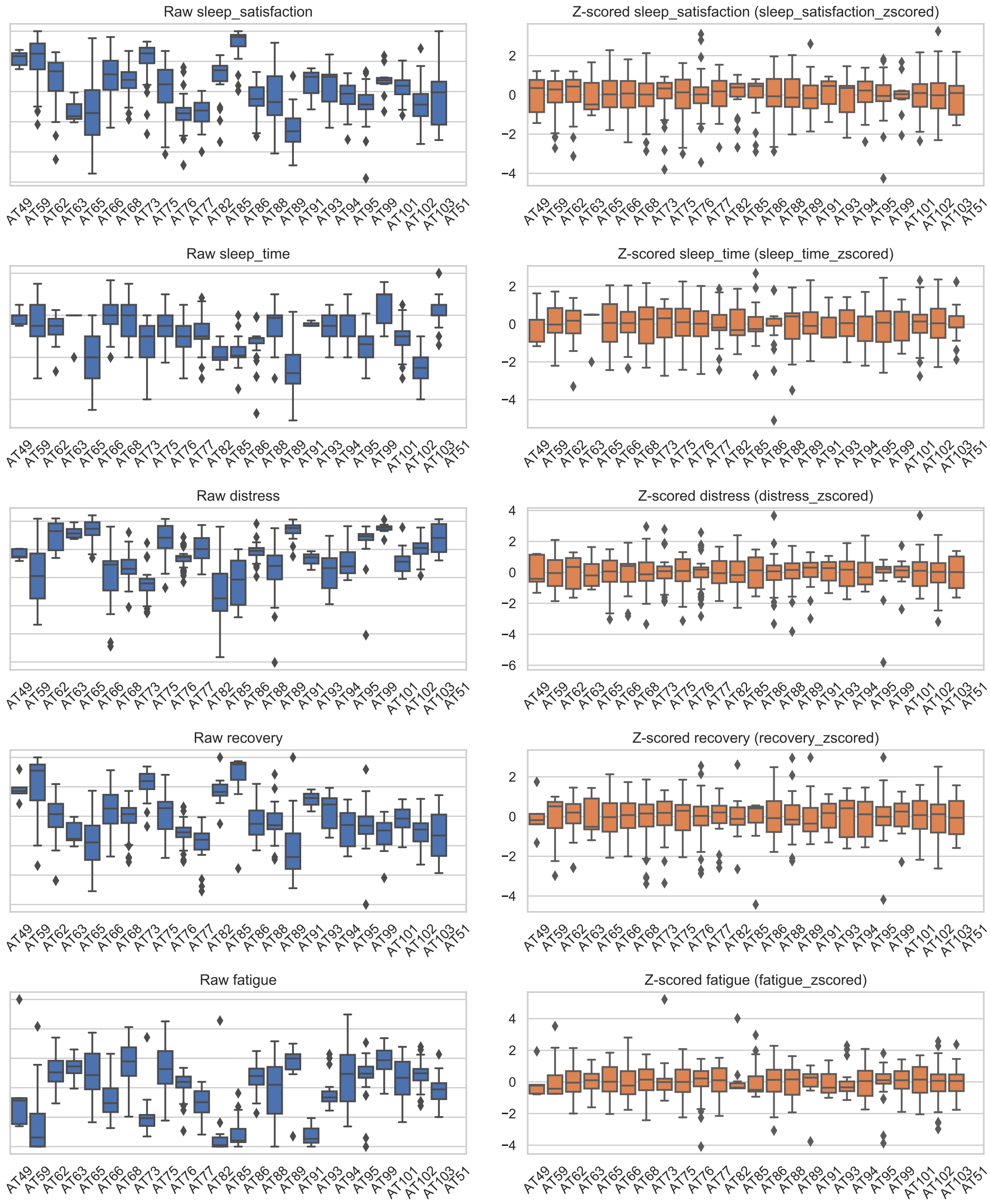


Fig. 5.8.: Raw and z-score normalization of wellness variables. Both versions of are preserved for the analysis.

Skewness was calculated for Fitbit steps and sleep variables (Table 5.1) and visually inspected against Figure 4.7. The variables `steps_vigorous_yesterday` and `steps_moderate_yesterday` were selected to be transformed with a logarithm. Outliers were detected with IQR and replaced with KNN for all Fitbit variables, a summary of it is described in Table 5.2.

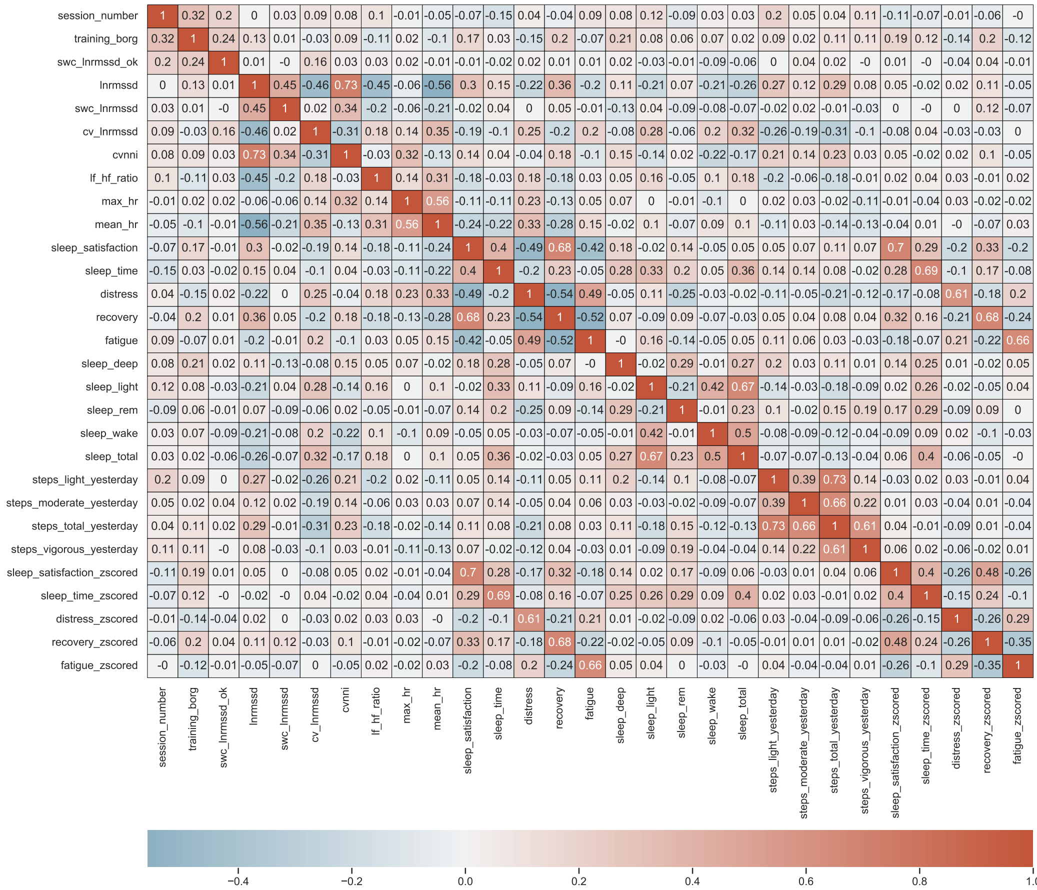


Fig. 5.9.: Correlation matrix after feature selection.

5.3.2 Feature Selection

Variance and correlation analysis provided the first step for feature selection. For correlation analysis, `lnrmssd` was used as reference for the rest of features since it is the gold standard to measure training adaptation.

Variance analysis discarded `cv_lnrmssd_ok`, `lnrmssd_ref_std`, `ratio_sd2_sd1`. Correlation analysis discarded several variables. Fifteen features were discarded due to its high correlation (> 0.8) with `lnrmssd`: `sdnn`, `sdsd`, `cvsd`, `hf`, `lf`, `min_hr`, `nni_20`, `nni_50`, `pnni_20`, `pnni_50`, `range_nni`, `sd1`, `sd2`, `total_power` and `lnrmssd_ref_mean`. Two features were discarded due to its high correlation with `cvnni`: `vlf` and `std_hr`. Two features were discarded due to its high correlation with `lf_hf_ratio`: `hfnu` and `lfnu`. Two features were discarded due to its high correlation with `mean_hr`: `mean_nni` and `median_nni`. The resulting features and its correlation matrix are shown in Figure 5.9.

5.3.3 Classic Clustering

This subsection describes the different clustering experiments run, the hypothesis on which each experiment is based, the interpretations of the results, and its limitations. The objective is to find groups of measures that may define the intervention process across time. In other words, the target is to find groups of training load measures indicating how are patients coping with training adaptation at the moment of the measure, and how prepared are they for a new training session on that day. This may be done by taking into account how variables describe patients' status. Some variables refer to the state of the patient at the moment of measurement (e.g., `lnrmssd`, `lf_hf_ratio`, `cvnni`, `recovery`, `distress`), whereas some others reflect an evolution of these parameters in time (`swc_lnrmssd`, `cv_lnrmssd`), and to the overall status of the patient `mean_hr`, `max_hr`.

To avoid skewing the algorithm with the accumulated effect of the intervention in the patients, the experiments avoided to include any temporal reference to the data (`study_day`) and any reference to the treatment received by the patients (`treatment`, `session_type`), as well as any other condition gathered in `demographics.csv`.

Classic clustering experiments were labeled as *CXX-K*, being *XX* the number of the clustering experiment, and *K* the number of clusters set for training. Besides, all the data fed to the algorithms was normalized with z-score normalization.

C01. All features: HRV, Wellness, and Fitbit

Experiment C01 tests if all features may be valuable to clustering when used at the same time. The results for 3 clusters (C01-3) are presented in the following (Figure 5.10). However, since 2 to 5 clusters were tested, the results are detailed in Appendix C, section C.1.

Three clusters (C01-3) were representative of the limitations of this first clustering experiment (Figure 5.10). At first sight, features like `sleep_satisfaction`, `distress`, `recovery`, `fatigue`, `cv_lnrmssd`, `lnrmssd`, `cvnni`, `steps_total_yesterday`, and `sleep_total` are of clear importance to separate the measures in 3 clusters. Moreover, checking how clustering labels are heterogeneously distributed by patient, clusters may not be clearly assessing changes during the intervention, but rather profiling types of patients. For instance, cluster 1 is mainly associated with measures of AT59, AT73, and AT85.

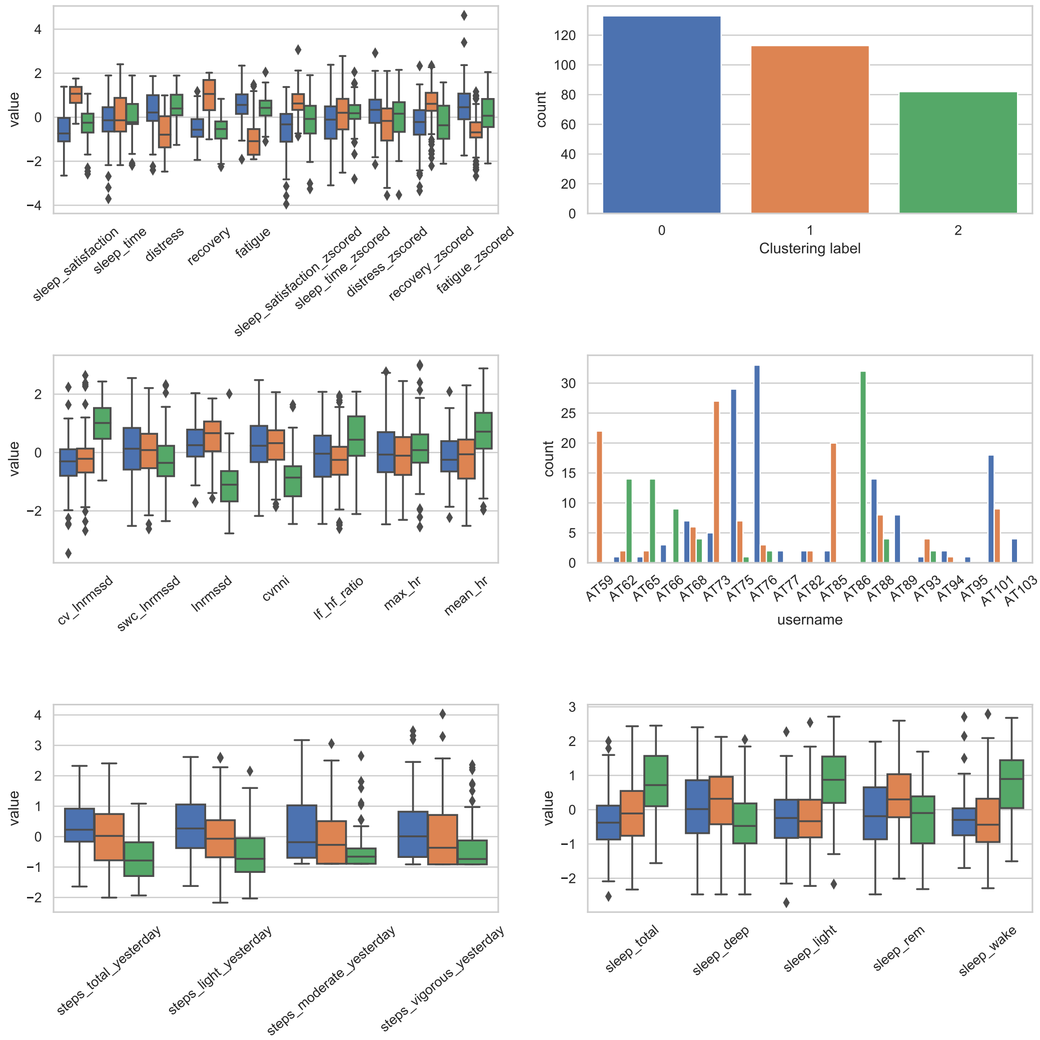


Fig. 5.10.: K-Means clustering experiment C01-3.

This result is replicated for clusters 0 and 2, and it is unlikely that most of patients were so adhered to just one profile during the entire intervention (e.g. the lnrmssd may start high, but it may also decrease with exercise, then go up again in the end). For instance, patients AT75, AT76, and AT101 were mainly associated with cluster 0; and patients AT62, AT65, and AT86 with cluster 2.

Another limitation is that the inclusion of all variables previously selected (HRV, wellness, and Fitbit steps and sleep) restricted the data sample to 328 measures. The intersection of HRV and self-reported wellness with Fitbit data limits this approach. Since HRV measures are the ones better able to reflect training adaptations, the next approach discards the use of Fitbit data.

C02. HRV and Wellness

Experiment C02 aimed to overcome the limitations posed by the restricted data sample in C01. This was done by removing Fitbit features, hence raising the data sample to 681 measures (almost twice than for C01). The focus was shifted to the analysis of HRV and Wellness variables, both typically used to monitor training adaptation. Two to five clusters were tested. The results are pictured in detail in Appendix C, section C.2.

As well as for C01, the clustering results for 3 clusters (Figure 5.11) may be representative of the C02 approach. However, in this case, the relevance of each variable is not as clear to the eye. Therefore, a feature importance analysis is shown in Table 5.3.

Although all variable presented more separation for every cluster compared to C01-3, the distribution of labels was still very skewed by participant. For instance, label 1, which may be associated to bad training adaptation (i.e., low `lnrmssd`, high `cv_lnrmssd`, low `recovery`, low `sleep_satisfaction`), was almost exclusive of participants AT65, AT86 and AT102.

Feature importance was analyzed with random forests classifiers in 5-Fold CV with a weighted F1-core of 86.96%. The top six most relevant features were `mean_hr`, `recovery`, `fatigue`, `lnrmssd`, `sleep_satisfaction`, and `distress`. The self-reported wellness in this top gathered 42.42% of importance, whereas all the z-scored wellness features only 13.22%.

Checking how raw wellness features are distributed across participants (Figure 5.8), it is clear that every participant perceives their wellness in way different forms, with different ranges and different mean values. This may compromise the clustering results, since the algorithm is getting its main skew from the patients' perception of training adaptation, instead of their physiological training adaptation. This is why a patient like AT59 was assigned the cluster label with better wellness (e.g., high `recovery`) and HRV features (e.g., low `cv_lnrmssd`, high `lnrmssd`) in C01 and C02.

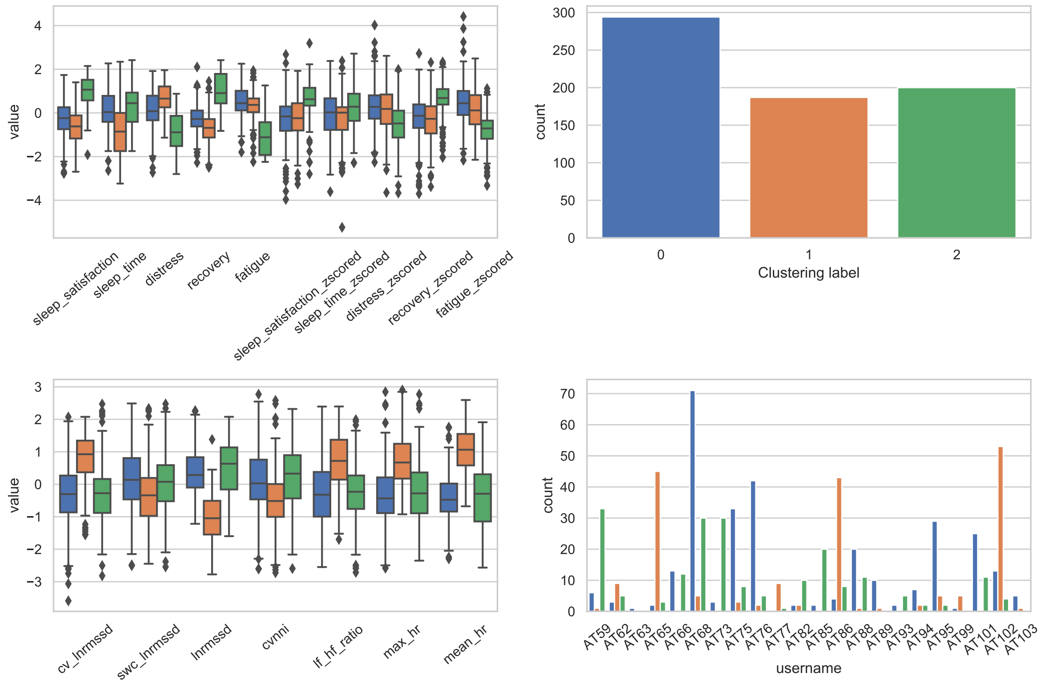


Fig. 5.11.: K-Means clustering experiment C02-3.

Tab. 5.3.: Feature importance for K-Means clustering experiment C02-3.

	feature importance (%)					
	CV1	CV2	CV3	CV4	CV5	Mean
cv_lnrmssd	4.72	6.98	3.31	5.58	6.25	5.37
swc_lnrmssd	0.84	1.05	0.90	0.91	0.79	0.90
lnrmssd	13.03	12.76	11.13	13.95	14.34	13.04
cvnni	1.39	1.03	0.81	1.21	1.45	1.18
lf_hf_ratio	2.41	1.60	2.28	2.24	1.92	2.09
max_hr	4.73	4.59	5.72	2.98	4.40	4.48
mean_hr	15.48	13.66	13.83	16.58	13.87	14.68
sleep_satisfaction	10.59	7.31	7.49	9.61	9.82	8.96
sleep_time	2.56	2.32	3.20	2.54	2.45	2.61
distress	7.93	8.18	8.32	6.50	8.10	7.81
recovery	12.96	14.72	11.58	12.81	12.40	12.89
fatigue	10.39	15.13	13.97	11.31	12.99	12.76
sleep_satisfaction_zscored	2.56	2.77	3.01	2.36	2.73	2.68
sleep_time_zscored	0.62	0.75	0.39	0.79	0.48	0.60
distress_zscored	1.11	0.96	1.54	1.46	1.16	1.25
recovery_zscored	1.90	2.34	2.13	2.94	1.24	2.11
fatigue_zscored	6.77	3.86	10.39	6.24	5.64	6.58

5.3.4 Layered Clustering

The previous clustering experiments revealed skew problems due to the nature of self-reports. To overcome this limitation, this section describes a layered clustering approach in which each layer matches a dimension of patients' health. The layers are *physiological status (L01)* for HRV features, *baseline perceived-wellness status (L02)* for raw wellness features, and *relative perceived-wellness status (L03)* for normalized wellness. Fitbit steps and sleep were discarded from these analysis due to the reduced sample found that joins HRV and well measures (328 measures with Fitbit data, 681 measures without Fitbit data).

Clustering experiments were labeled as *LXXY-K*, with *XX* referring to the layer number or health dimension (e.g., L01 for physiological status), *Y* being an A–Z letter indicating the iteration in feature selection, and *K* pointing out the number of clusters found.

The selection of the final *K* for each experiment was done checking silhouette coefficients and making interpretations of the clusters found. Although a higher silhouette score represents a better structured clustering, it does not necessarily mean a better representation of the reality. That is why all the selected results are complemented with its interpretation.

L01. Physiological Status

Assessing the physiological status is essential to understand the patient's adaptation to training during exercise intervention. For this purpose, HRV features pose the best option. Different combinations of HRV features are described in the following feature selection process combined with the interpretation of the results.

A trade-off analysis comparing silhouette scores (Table C.3, Table C.5, and Table C.7 in Appendix C) and interpretations against the state of the art with different values of k (2 to 5) for the following experiments resulted in choosing 4 clusters for L01.

Graphic descriptions of the three clustering approaches for L01 with $k = 4$ clusters are displayed in Figures 5.12, 5.13, and 5.14. Their respective feature importance analysis are detailed in Tables 5.4, 5.5, and 5.6. Their numerical description are detailed in Tables C.4, C.6, and C.8.

The legend for interpretations is $\uparrow$ for high values, $\downarrow$ for lower values, $\sim$ for intermediate values, and $\updownarrow$ to indicate wide distributions (prone overlapping with other clusters).

L01A. All HRV features. The clustering results for L01A with 2 to 5 clusters are detailed in Appendix C, section C.3. The results with 4 clusters are representative of the possibilities of this approach (Figure 5.12). Cluster labels are balanced, and its distribution across participants is not as skewed as in previous experiments. An interpretation of the clusters is presented in the following:

Cluster 0 (R↓). Regular-to-bad overall status (↓ `lnrmssd`, ↓ `cvnni`, ↓ `max_hr`, ↓ `mean_hr`) with recent (↓↕ cv_lnrmssd) physiological stress (↓ `swc_lnrmssd`, ~↕ `lf_hf_ratio`)

Cluster 1 (G). Good overall status (↑↑ `lnrmssd`, ↑ `cvnni`, ↓ `max_hr`, ↓ `mean_hr`) recently (↓↕ cv_lnrmssd) recovering from physiological stress (↑ `swc_lnrmssd`, ↓ `lf_hf_ratio`)

Cluster 2 (B). Bad overall status (↓↓ `lnrmssd`, ↓ `cvnni`, ↑ `max_hr`, ↑ `mean_hr`) with accumulated (↑ cv_lnrmssd) and recent physiological stress (↓ `swc_lnrmssd`, ↑ `lf_hf_ratio`)

Cluster 3 (R↑). Regular-to-good overall status (↑↑ `lnrmssd`, ↑↑ `cvnni`, ↑ `max_hr`, ↑ `mean_hr`) without accumulated (↓ cv_lnrmssd) physiological stress and uncertain recent stress (~↕ `swc_lnrmssd`, ~↕ `lf_hf_ratio`)

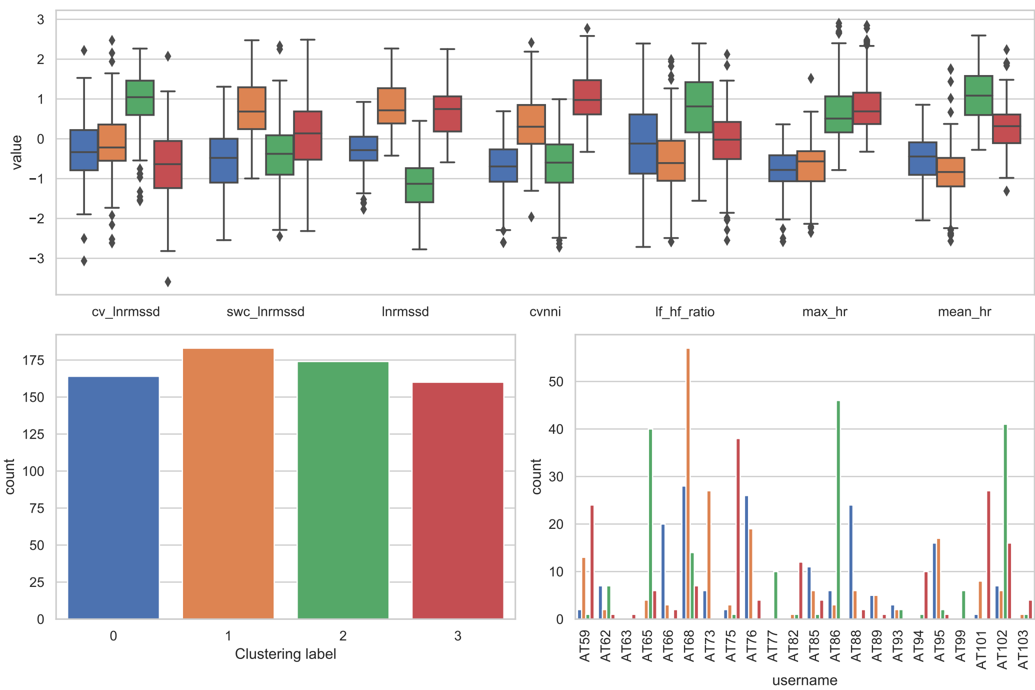


Fig. 5.12.: K-Means clustering experiment L01A-4. A numerical description is in Table C.4.

Tab. 5.4.: Feature importance for K-Means clustering experiment L01A-4.

	Feature importance (%)					
	CV1	CV2	CV3	CV4	CV5	Mean
cv_lnrmssd	9.43	10.41	7.64	6.78	11.59	9.17
swc_lnrmssd	9.36	6.45	9.96	9.42	6.92	8.42
lnrmssd	24.72	24.40	21.38	24.85	24.22	23.91
cvnni	15.21	16.17	16.44	15.62	13.88	15.46
lf_hf_ratio	3.08	3.76	3.58	4.26	3.91	3.72
max_hr	22.53	23.69	24.90	24.72	25.54	24.28
mean_hr	15.67	15.12	16.10	14.34	13.95	15.04

Feature importance analysis (random forests mean weighted-F1-score 87.63%) in Table 5.4 revealed that `max_hr` and `mean_hr` accumulated a 39.32% importance. Such level of reliance may jeopardize the quality of the clusters, since both features, measured in resting conditions, cannot assess acute training adaptation in the short term (Buchheit, 2014; Plews et al., 2013; Shaffer & Ginsberg, 2017).

L01B. All HRV features except max_hr and mean_hr. The clustering results for L01B with 2 to 5 clusters are detailed in Appendix C, section C.4. The results with 4 clusters are representative of the possibilities of this approach, illustrated in Figure 5.13, and numerically described in Appendix C, Table C.6. Clustering labels are balanced and its distribution across participants seems less skewed compared to previous experiments. Moreover, attending to the distribution of clusters, all features seem to be contributing to the clustering. An interpretation of the clusters is presented in the following:

Cluster 0 (G). Good overall status (↑↑ `lnrmssd`, ↑↑ `cvnni`), no accumulated physiological stress (↓ `cv_lnrmssd`), and recent recovery from physiological stress (↑↑ `swc_lnrmssd`, ∼↕ `lf_hf_ratio`)

Cluster 1 (B). Bad overall status (↓↓ `lnrmssd`, ↓ `cvnni`), accumulated (↑ `cv_lnrmssd`) and recent physiological stress (↓ `swc_lnrmssd`, ↑↑ `lf_hf_ratio`)

Cluster 2 (R↑). Regular-to-good overall status (↑ `lnrmssd`, ↑ `cvnni`), no accumulated (↓ `cv_lnrmssd`) but recently increased physiological stress (↓ `swc_lnrmssd`, ∼↕ `lf_hf_ratio`)

Cluster 3 (R↓). Regular-to-bad overall status (↓ `lnrmssd`, ↓ `cvnni`), accumulated (↑ `cv_lnrmssd`) but low recent physiological stress (∼ `swc_lnrmssd`, ↓ `lf_hf_ratio`)

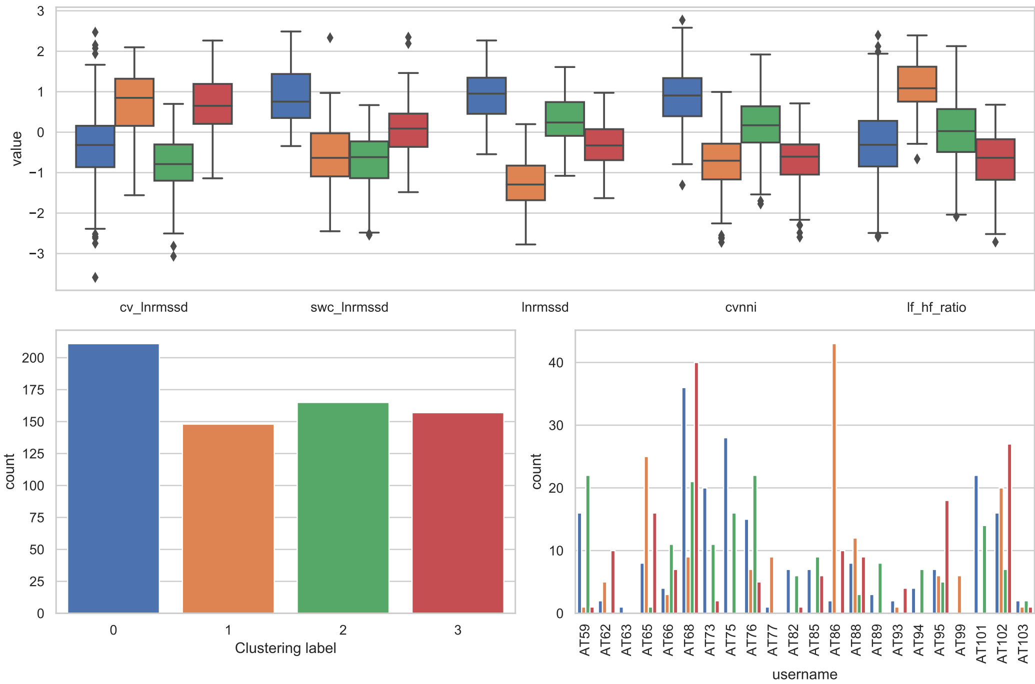


Fig. 5.13.: K-Means clustering experiment L01B-4. A numerical description is in Table C.6.

Tab. 5.5.: Feature importance for K-Means clustering experiment L01B-4.

	feature importance (%)					
	CV1	CV2	CV3	CV4	CV5	Mean
cv_lnrmssd	14.51	16.32	15.12	14.17	13.43	14.71
swc_lnrmssd	22.62	21.43	24.11	20.77	22.19	22.22
lnrmssd	24.01	26.85	23.68	26.07	25.15	25.15
cvnni	19.48	15.91	17.75	19.99	20.52	18.73
lf_hf_ratio	19.38	19.49	19.35	19.00	18.71	19.19

Feature importance analysis (random forests mean weighted-F1-score 88.92%) detailed in Table 5.5 revealed that all features are making a balanced contribution to the clustering, being `lnrmssd` the most relevant with 25.15% mean importance.

L01C. All HRV features except max_hr, mean_hr, and lf_hf_ratio. Due to controversies on the use of `lf_hf_ratio` (Billman, 2013), this feature was removed for the last version of L01. The clustering results for L01C with 2 to 5 clusters are detailed in Appendix C, section C.5. The results with 4 clusters are representative of the possibilities of this approach, illustrated in Figure 5.14, and numerically described in Appendix C, Table C.8. Labels are balanced across the clustering, and its distribution

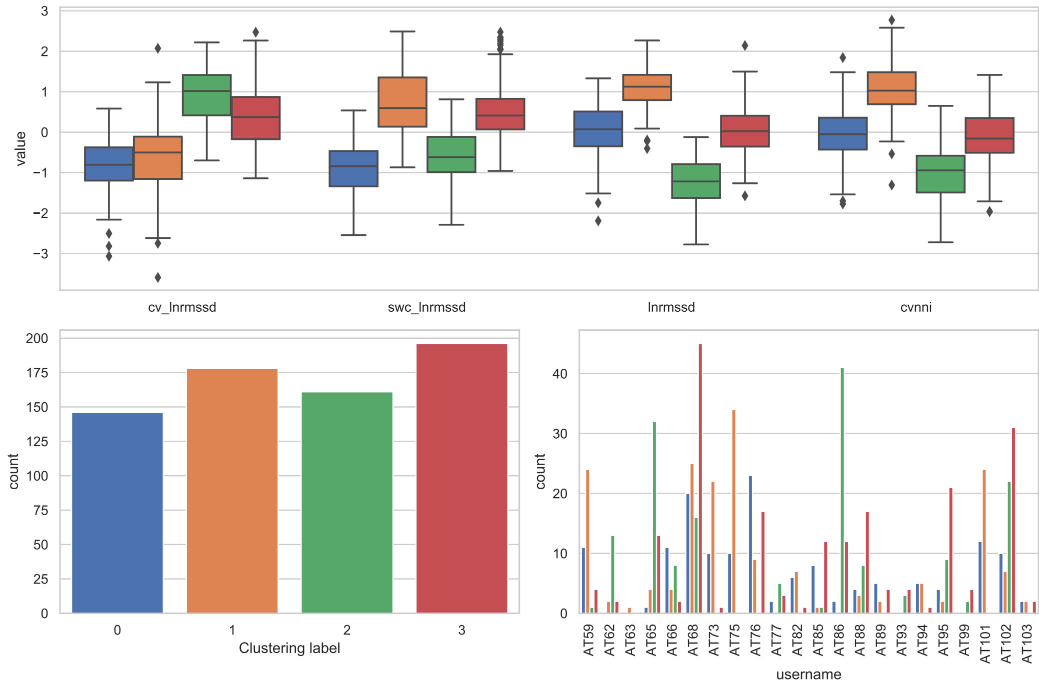


Fig. 5.14.: K-Means clustering experiment L01C-4. A numerical description is in Table C.8.

across participants is not as skewed as in previous experiments. Moreover, attending to the distribution of clusters, all variables seem to be contributing to the clustering. An interpretation of the clusters is presented in the following:

Cluster 0 (R↓). Regular overall status (∼ `lnrmssd`, ∼ `cvnni`), no accumulated (↓ `cv_lnrmssd`) but recent physiological stress (↓ `swc_lnrmssd`)

Cluster 1 (G). Good overall status (↑ `lnrmssd`, ↑ `cvnni`), no accumulated physiological stress (↓ `cv_lnrmssd`), and recovery from recent physiological stress (↑ `swc_lnrmssd`)

Cluster 2 (B). Bad overall status (↓ `lnrmssd`, ↓ `cvnni`) with accumulated (↑ `cv_lnrmssd`) and recent physiological stress (↓ `swc_lnrmssd`)

Cluster 3 (R↑). Regular overall status (∼ `lnrmssd`, ∼ `cvnni`) with accumulated (↑ `cv_lnrmssd`) physiological stress but recent recovery (↑ `swc_lnrmssd`)

Feature importance analysis (random forests mean weighted-F1-score 87.48%) detailed in Table 5.6 revealed that all features made a balanced contribution to the clustering, being `lnrmssd` the most relevant with 25.15% mean importance.

Tab. 5.6.: Feature importance for K-Means clustering experiment L01C-4.

	feature importance (%) CV1	CV2	CV3	CV4	CV5	Mean
cv_lnrmssd	19.39	23.93	22.16	21.23	19.14	21.17
swc_lnrmssd	26.98	26.36	27.47	25.97	25.64	26.49
lnrmssd	38.54	33.69	36.22	36.85	38.46	36.75
cvnni	15.08	16.02	14.16	15.94	16.76	15.59

L02. Baseline Wellness

The psychological status of patients may play an important role in their adherence to a TE intervention. Raw wellness features pose a solution to identify the baseline wellness status of patients. Different combinations of features are described in the following feature selection process combined with their interpretation.

After interpreting the distribution of clusters for $k = 2, 3, 4, 5$ against the state of the art, and comparing the silhouette scores obtained (Table C.9 and Table C.11 in Appendix C), clustering with $k = 3$ clusters was considered the best option for L02. Graphic descriptions of the two clustering approaches for L02 are displayed in Figures 5.15 and 5.16. Feature importance analysis are summarized in 5.7 and 5.9. Numerical descriptions of the main clusters are detailed in Tables C.10 and C.12.

L02A. Raw wellness features. The clustering results for L02A with 2 to 5 clusters are detailed in Appendix C, section C.6. The results with 3 clusters are representative of the possibilities of this approach, illustrated in Figure 5.15, and numerically described in Appendix C, Table C.10. Labels are balanced across the clustering labels, but, as expected, its distribution among participants is skewed to their perception levels. All variables seem to be contributing to the clustering, with the exception of `sleep_time`, which may not be helping to differentiate label 1 from 2, and `fatigue`, which may not be helping to differentiate label 0 from 2. An interpretation of the clusters is presented in the following:

Cluster 0 (B). Bad wellness status (↓ `sleep_satisfaction`, ↓ `sleep_time`, ↑ `distress`, ↓ `recovery`, ↑ `fatigue`)

Cluster 1 (G). Good wellness status (↑ `sleep_satisfaction`, ∼↕ `sleep_time`, ↓ `distress`, ↑ `recovery`, ↓ `fatigue`)

Cluster 2 (R). Regular wellness status (∼ `sleep_satisfaction`, ↑ `sleep_time`, ∼↕ `distress`, ∼ `recovery`, ↑ `fatigue`)

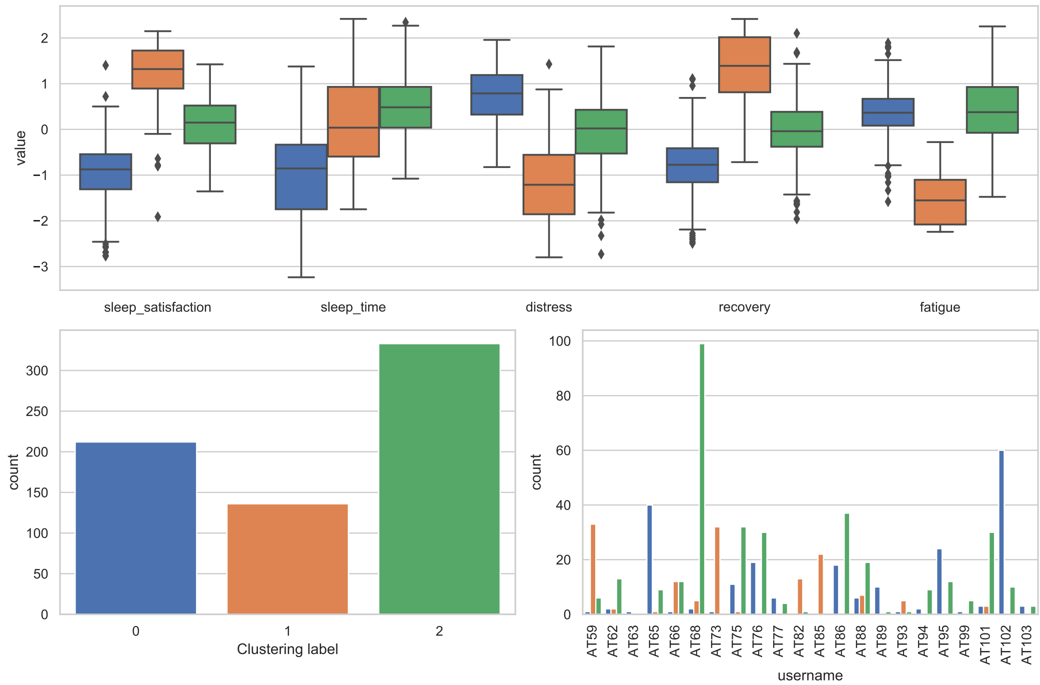


Fig. 5.15.: K-Means clustering experiment L02A-3. A numerical description is in Table C.10.

Tab. 5.7.: Feature importance for K-Means clustering experiment L02A-3.

	Feature importance (%)					
	CV1	CV2	CV3	CV4	CV5	Mean
sleep_satisfaction	24.79	18.78	25.13	26.60	22.23	23.51
sleep_time	20.17	23.69	21.38	17.78	18.52	20.31
distress	8.00	8.01	7.06	7.29	9.54	7.98
recovery	19.74	19.90	17.97	19.36	21.65	19.72
fatigue	27.30	29.62	28.46	28.98	28.06	28.48

Feature importance analysis (random forests mean weighted-F1-score 89.88%) in Table 5.7 revealed that all features made a balanced contribution to the clustering except for `distress`.

Despite the high relevance reported for `sleep_time` in the feature analysis for L02A-3, after a visual inspection of its distribution in clusters a second feature importance analysis was performed. A secondary feature importance analysis based on linear SVM (SVM mean weighted-F1-score 96.80%) in Table 5.8 indicated that the relevance of `distress` may be higher. In addition, the overlapping of `sleep_time` for cluster labels 1 and 2 suggest that it may not be relevant to distinguish good from regular baseline wellness.

Tab. 5.8.: Feature importance (Linear SVM) for K-Means clustering experiment L02A-3.

	Feature importance (%)					
	CV1	CV2	CV3	CV4	CV5	Mean
sleep_satisfaction	23.95	21.58	23.55	25.11	23.44	23.53
sleep_time	17.22	16.20	16.63	16.32	17.35	16.74
distress	15.21	16.31	17.14	14.93	16.72	16.06
recovery	23.39	25.04	24.03	23.96	24.69	24.23
fatigue	20.23	20.87	18.64	19.68	17.80	19.44

L02B. Raw wellness features except sleep_time. The clustering results for L02B with 2 to 5 clusters are detailed in Appendix C, section C.7. The results with 3 clusters are representative of the possibilities of this approach, illustrated in Figure 5.16, and numerically described in Appendix C, Table C.12. Labels are balanced across the clustering labels, but, as expected, its distribution among participants is skewed to their perception levels. All variables seem to be contributing to the clustering in a balanced way. An interpretation of the clusters is presented in the following:

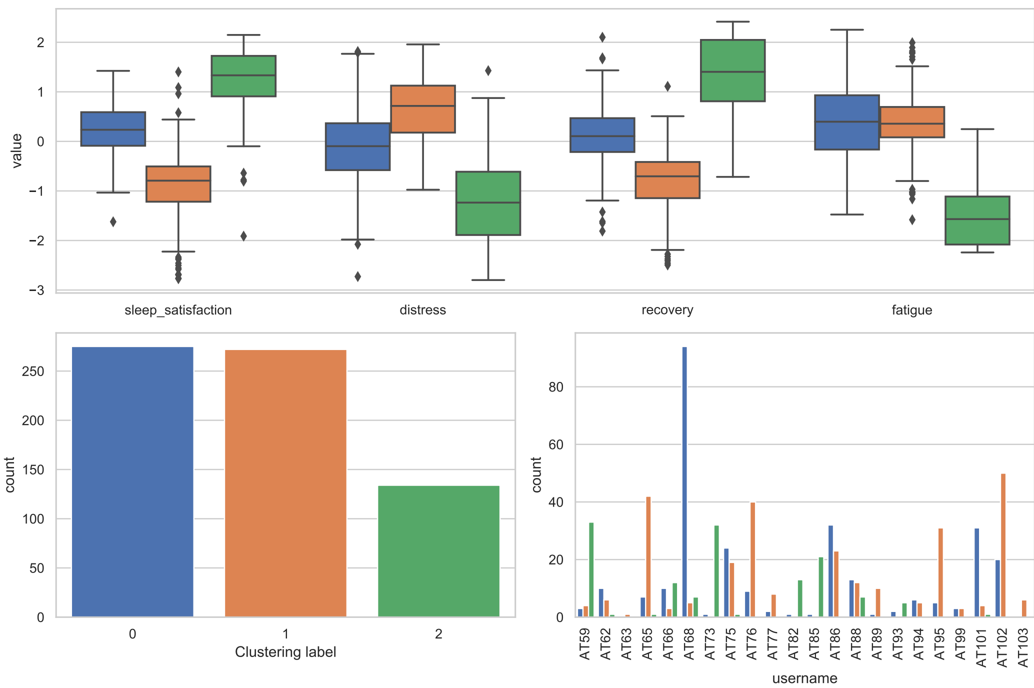


Fig. 5.16.: K-Means clustering experiment L02B-3. A numerical description is in Table C.12.

Tab. 5.9.: Feature importance for K-Means clustering experiment L02B-3.

	Feature importance (%)					
	CV1	CV2	CV3	CV4	CV5	Mean
sleep_satisfaction	34.78	29.38	34.79	38.09	29.34	33.28
distress	10.11	11.52	15.35	9.23	11.59	11.56
recovery	24.86	28.77	20.09	19.68	26.72	24.02
fatigue	30.25	30.34	29.77	33.00	32.35	31.14

Cluster 0 (R). Regular wellness status (∼ `sleep_satisfaction`, ∼↕ `distress`, ∼ `recovery`, ↑ `fatigue`)

Cluster 1 (B). Bad wellness status (↓ `sleep_satisfaction`, ↑ `distress`, ↓ `recovery`, ↑ `fatigue`)

Cluster 2 (G). Good wellness status (↑ `sleep_satisfaction`, ↓ `distress`, ↑ `recovery`, ↓ `fatigue`)

Feature importance analysis (random forests mean weighted-F1-score 92.03%) in Table 5.9 revealed that all features made a balanced contribution for the classifier.

L03. Relative Wellness

Normalized wellness features pose a solution to identify the wellness levels relative to their baseline. The features selected for these layers are the same selected as for L02 to allow coherent comparisons. Using $k = 3$ clusters allowed to find lower, normal, and higher values than compared to baseline.

L03A. Normalized wellness features except sleep_time. The clustering results for L03A with 2 to 5 clusters are detailed in Appendix C, section C.8. The results with 3 clusters are representative of the possibilities of this approach, illustrated in Figure 5.17, and numerically described in Appendix C, Table C.14. Labels are balanced across the clustering labels, but, as expected, its distribution among participants is skewed to their perception levels. All variables seem to be contributing to the clustering in a balanced way. An interpretation of the clusters depicted in Figure 5.17 is presented in the following:

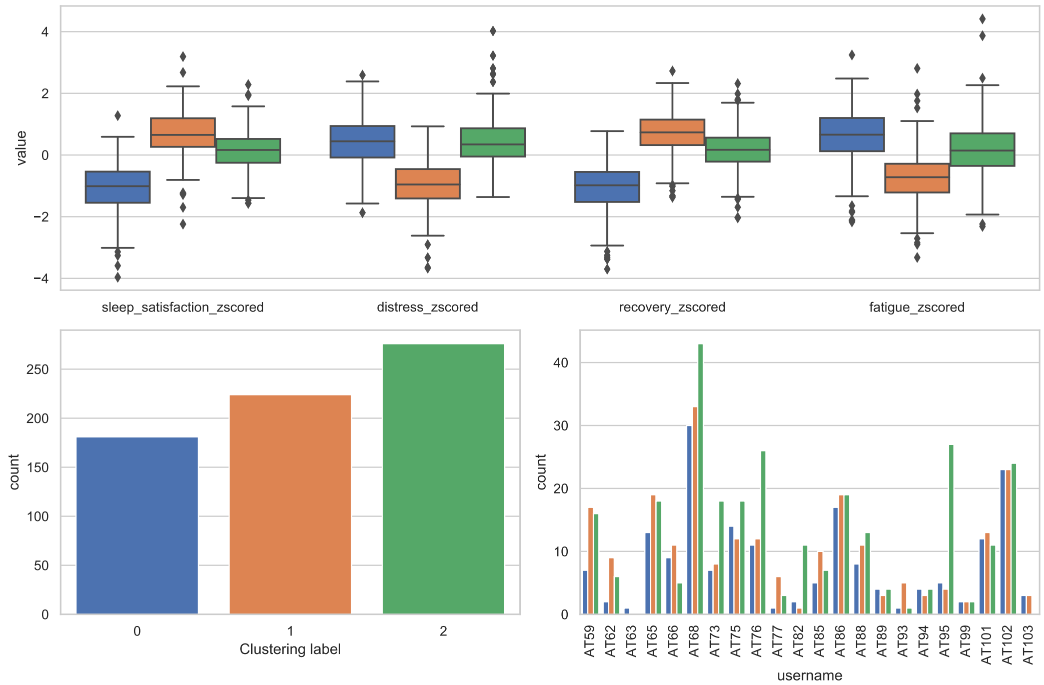


Fig. 5.17.: K-Means clustering experiment L03A-3. A numerical description is in Table C.14.

Cluster 0 (W). Worsened relative wellness status (↓ `sleep_satisfaction_zscored`, ↑`distress_zscored`, ↓ `recovery_zscored`, ↑ `fatigue_zscored`)

Cluster 1 (I). Improved relative wellness status (↑ `sleep_satisfaction_zscored`, ↓`distress_zscored`, ↑ `recovery_zscored`, ↓ `fatigue_zscored`)

Cluster 2 (M). Maintained relative wellness status (∼ `sleep_satisfaction_zscored`, ↑`distress_zscored`, ∼ `recovery_zscored`, ∼ `fatigue_zscored`)

Feature importance analysis (random forest mean weighted-F1-score 89.98%) in Table 5.10 revealed that all features made a balanced contribution for the classifier.

Tab. 5.10.: Feature importance for K-Means clustering experiment L03A-3.

	Feature importance (%)					
	CV1	CV2	CV3	CV4	CV5	Mean
sleep_satisfaction_zscored	29.30	30.57	28.82	26.00	27.50	28.44
distress_zscored	27.13	31.61	27.67	29.85	31.69	29.59
recovery_zscored	26.90	26.10	27.38	29.92	27.26	27.51
fatigue_zscored	16.68	11.71	16.13	14.22	13.55	14.46

Clustering Coherence Among Layers

Looking for an internal validation of the clustering results, HRV, baseline wellness, and relative wellness were paired into confusion matrices for its interpretation. The confusion matrix for HRV and baseline wellness is in Table 5.11; for HRV and normalized wellness in Table 5.12. The clustering results matched were L01C-4 for HRV, L02B-3 for baseline wellness, and L03A-3 for relative wellness.

First, Table 5.11 showed that, HRV label 0 (regular status, recent physiological stress) had higher coincidence (46.82%) with baseline wellness label 1 (bad wellness status). Second, HRV label 1 (good status, positive adaptation to training) also shared a high amount of instances (48.40%) with baseline wellness label 1 (bad wellness status). Conversely, HRV label 2 (bad status, sustained physiological stress) had the least amount of coincidence (8.49%) with baseline-wellness label 1 (bad wellness status), having more presence with baseline-wellness label 0 (42.54%, regular wellness status) and label 2 (48.98%, good wellness status). Finally, HRV label 3 (regular status, positive adaptation to training) had the higher coincidence with baseline-wellness label 2 (41.26%, good wellness status), followed by label 0 (34.16%, regular wellness status), and label 1 (24.58%, bad wellness status).

These results are not aligned with the typical expectations that would match good (HRV label 1) or regular (HRV label 3) HRV profiles with the best wellness cluster label (baseline wellness label 2).

The confusion matrix for HRV and relative wellness did not match that expectation either (Table 5.11). In fact, the incidence of the three relative wellness clusters was almost uniformly distributed ($\sim 33\%$) across the four HRV labels.

This is a sign of how much skewness wellness variables can introduce in the analysis. For the sake of exemplifying this result, the incidence of HRV label 2 (bad overall status, accumulated physiological stress) is analyzed in the following. Despite HRV

Tab. 5.11.: Confusion matrix for HRV (L01C-4) and baseline wellness (L02B-3) clusters. All values are expressed as percentage (%) of the total amount of instances for each HRV label (i.e., the values in a row must sum 100%).

	Label baseline wellness		
Label HRV	0 (R)	1 (B)	2 (G)
0 (R↑)	26.44	46.82	26.74
1 (G)	31.59	48.40	20.01
2 (B)	42.54	8.49	48.98
3 (R↓)	34.16	24.58	41.26

Tab. 5.12.: Confusion matrix for HRV (L01C-4) and normalized wellness (L03A-3) clusters. All values are expressed as percentage (%) of the total amount of instances for each HRV label (i.e., the values in a row must sum 100%).

	Label normalized wellness		
Label HRV	0 (W)	1 (I)	2 (M)
0 (R↑)	31.95	33.50	34.55
1 (G)	32.39	35.28	32.33
2 (B)	37.37	35.12	27.51
3 (R↓)	31.80	29.89	38.31

label 2 was distributed across all participants (Figure 5.14), the patients AT86, AT65, and AT102 recorded its highest incidence, in this order. These patients were very different among them.

AT86 was a patient undergoing chemotherapy, with vigorous physical activity levels, QLQ-C30 global health score of 75%, and low adherence to the protocol (5 training sessions, BORG intensity 6.20 ± 0.44).

AT65 was a patient undergoing chemotherapy, with low physical activity levels, QLQ-C30 global health score of 16.67%, and medium adherence to the protocol (12 sessions, BORG intensity 6.42 ± 0.80).

AT102 was a patient undergoing chemotherapy, with moderate levels of physical activity, a QLQC30 global health of 58.33% and high adherence to the protocol (23 training sessions, BORG intensity 6.52 ± 1.18).

The common factor across these three patients was undergoing chemotherapy treatment, hence a deteriorated overall status due to the secondary effects. Figure 5.18 displays the evolution of HRV labels found for the three patients across the intervention; and Figure 5.19 displays the evolution of baseline wellness. The three patients mostly reported good (cluster 1) and regular (cluster 0) wellness status; however, the HRV profiles were typically stuck at HRV cluster 2 (bad overall status, accumulated physiological stress), with some oscillations to cluster 3 (regular overall status, recent recovery from physiological stress) for the whole intervention. The only exception was for AT102, who continued her improvement until the end of the intervention.

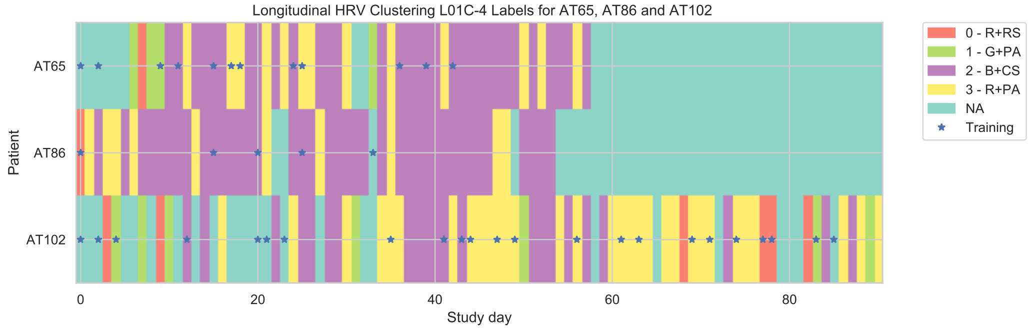


Fig. 5.18.: Longitudinal clustering exploration for HRV (L01C-4) for patients AT65, AT86, and AT102.

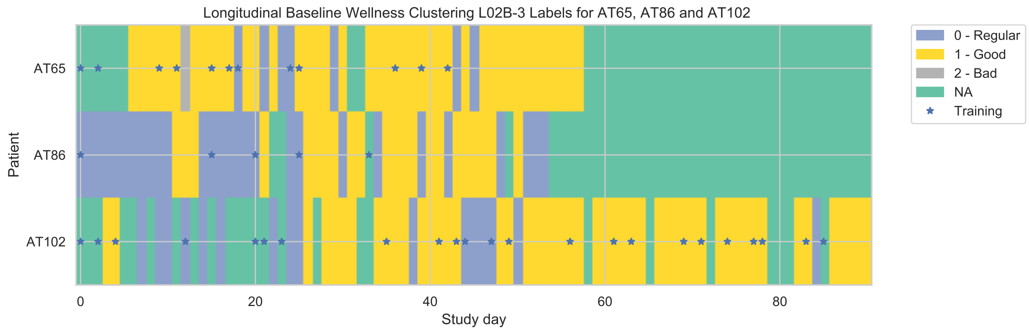


Fig. 5.19.: Longitudinal clustering exploration for baseline wellness (L02B-3) for patients AT65, AT86, and AT102.

5.4 Discussion

5.4.1 Principal Findings

Feature Selection & Clustering

The feature selection process showed high levels of redundant information across HRV variables. Therefore, it was of utmost importance to take into account expert knowledge when selecting the most relevant variables. Since the literature on HRV analysis already stated that resting short-HRV measures of the LnRMSSD were the most solid assessment training load (Buchheit, 2014; Plews et al., 2013; Shaffer & Ginsberg, 2017), it was used to cascade the feature selection process.

The clustering process revealed how the different variables may play multiple roles in the assessment. Moreover, an approach like the layered clustering was mandatory to fully leverage the data without introducing skew in the analysis. Domain knowledge about the variables was needed during the clustering process in order to remove features like `max_hr` or `mean_hr`, which may skew the clusters due to its slow changes in time (Shaffer & Ginsberg, 2017), or `lf_hf_ratio`, due to its discussed inability to reflect balance between sympathetic and parasympathetic activity (Billman, 2013). This process enabled finding clusters of training load representative of the training adaptation processes of patients with breast cancer during TE intervention.

HRV Clusters

Clustering results for HRV (L01) are coherent with the HRV norms found in the state of the art, specially for L01B-4 (section C.4, Table C.6) and L01C-4 (section C.5, Table C.8). First of all, both address the physiological status of the patient from different temporal perspectives at the same time. The features `cv_lnrmssd` and `swc_lnrmssd` incorporate the dimension of time by leveraging measures of the previous 7 days, whilst `lnrmssd`, `cvnni` (and `lf_hf_ratio` for L01B-4) are representative of the acute state of the patient at the moment of the recording. According to discrepancies with the use of LF/HF to represent autonomic balance (Billman, 2013), the use of an approach like L01C-4 is discouraged in favor of L01B-4. Nevertheless, the results obtained for both approaches are discussed in the following.

CV of LnRMSSD. Clusters 0 (G) and 2 (R↑) for L01B-4 had low values of `cv_lnrmssd` ($6.66 \pm 3.78\%$ and $4.89 \pm 1.69\%$, respectively), which are close to the ones found for healthy people like elite triathletes at baseline (Plews et al., 2012; `cv_lnrmssd`$< 4\%$. Plews et al., 2014; `cv_lnrmssd`$= 6.7 \pm 2.9\%$). The `cv_lnrmssd` values in L01C-4 for clusters 0 (R↓) and 1 (G) ($4.75 \pm 1.56\%$ and $5.59 \pm 2.66\%$, respectively) are also comparable in the same manner.

Conversely, clusters 1 (B) and 3 (R↓) for L01B-4 had higher values for `cv_lnrmssd` ($11.42 \pm 4.65\%$ and $11.38 \pm 4.61\%$, respectively), comparable to the ones found for athletes during non-functional overreaching (Plews et al., 2012; mean `cv_lnrmssd` $\approx 8\%$), and to recreational runners during the first week of training (Plews et al., 2014; `cv_lnrmssd`$= 10.1 \pm 3.4\%$). Clusters 2 (B) and 3 (R↑) for L01C-4 ($12.95 \pm 4.54\%$ and $9.76 \pm 4.19\%$, respectively) are also comparable in the same manner.

LnRMSSD. Clusters 0 (G) and 2 (R↑) for L01B-4 had high values of `lnrmssd` ($4.04 \pm 0.38\ ms$ and $3.65 \pm 0.33\ ms$) that are comparable to patients with increased survival (Guo et al., 2015; $4.02 \pm 3.85\ ms$), cancer patients after an exercise intervention (Freitag et al., 2018; $3.66\ ms$), and healthy women (Caro-Moran et al., 2016; $4.07 \pm 3.10\ ms$. De Couck and Gidron, 2013; $3.74 \pm 2.71\ ms$). The `lnrmssd` values in L01C-4 for clusters 0 (R↓), 1 (G) and 3 (R↑) ($3.51 \pm 0.40\ ms$ and $4.15 \pm 0.32\ ms$, $3.48 \pm 0.32\ ms$ respectively) are also comparable in the same manner.

Conversely, clusters 1 (B) and 3 (R↓) for L01B-4 had low values of `lnrmssd` ($2.66 \pm 0.40\ ms$ and $3.29 \pm 0.34\ ms$) that are comparable to patients with reduced survival (Guo et al., 2015; $3.26 \pm 2.56\ ms$), cancer patients before an exercise intervention (Freitag et al., 2018; $2.77\ ms$), cancer patients without exercise intervention (Caro-Moran et al., 2016; $3.36 \pm 3.16\ ms$. De Couck and Gidron, 2013; $3.16 \pm 3.21\ ms$). The `lnrmssd` values in L01C-4 for cluster 2 (B) ($2.69 \pm 0.38\ ms$) are also comparable in the same manner.

SWC of LnRMSSD. Clusters 0 (G), 1 (B) and 2 (R↑) for L01B-4 had high absolute values for the normalized `swc_lnrmssd` (1.31 ± 1.02, -0.8 ± 1.13, and -0.98 ± 0.96) that are comparable to one of the factors used by HRV4Training to compute SWC ($f = 1$). This higher value than the typically used ($f = 0.5$) may come as a result of shrinking the window for LnRMSSD analysis to 1 week, instead of the typical 4 weeks for professional athletes (Buchheit, 2014; Plews et al., 2013; Javaloyes et al., 2019). Cluster 3 sits on the fence with a value of 0.17 ± 0.92. For L01C-4, clusters are also comparable and are better divided. L01C-4 presented clusters 0 (R↓) and

2 (B) with negative values (-1.27 ± 0.97 and -0.81 ± 0.89), and 1 (G) and 3 (R↑) with positive values (1.06 ± 1.14 and 0.77 ± 0.90) for `swc_lnrmssd`.

Positive values of `swc_lnrmssd` may be found in clusters 0 (G) and 3 (R↓) for L01B-4, and in clusters 1 (G) and 3 (R↑) for L01C-4, may be associated with good training adaption if they are not too high (Buchheit, 2014). Otherwise, negative values of `swc_lnrmssd` (clusters 1 (B), 2 (R↑) and 3 (R↓) for L01B-4, and in clusters 0 (R↓) and 2 (B) for L01C-4) may be associated with accumulated fatigue; however, such interpretations may be done assessing its relation to HR changes (Buchheit, 2014). Therefore a measure like `cvnni` is adequate to accompany this measure.

CVNNI. Cluster 0 (G) for L01B-4 had the highest values of `cvnni` (0.0906 ± 0.0293), which were in company of the highest `lnrmssd` values. Cluster 2 (R↑) for L01B-4 had middle values for `cvnni` (0.0648 ± 0.0192), also in company of high `lnrmssd` values. Clusters 1 (B) and 3 (R↓) L01B-4 had the lowest `cvnni` values (0.0437 ± 0.0140 and 0.0441 ± 0.0121). A different pattern was found for L01C-4. Cluster 1 (G) had the highest values of `cvnni` (0.0965 ± 0.0285), which were in company of the highest `lnrmssd` values. Clusters 0 (R↓) and 3 (R↑) had middle values (0.0589 ± 0.0171 and 0.0574 ± 0.0163), also with middle values for `lnrmssd`. Finally, cluster 2 (B) had the lowest values for `cvnni` (0.0384 ± 0.0114).

`cvnni` is the SDNN divided by the mean duration of RR intervals. There is a point of saturation in which measures like SDNN and the RMSSD do not raise linearly with the duration of RR intervals Plews et al., 2013. Therefore, `cvnni` may provide robustness and finer interpretation to RMSSD values. High `lnrmssd` values should be in company of high `cvnni` in order to be outside saturation point; conversely, if high `lnrmssd` values are in company of lower `cvnni` values, that may indicate accumulated fatigue Plews et al., 2013.

LF/HF. L01B-4 used `lf_hf_ratio` during clustering. Cluster 1 (G) reported the highest values for `lf_hf_ratio` (6.23 ± 1.92), which are values for sympathetic activity (Shaffer & Ginsberg, 2017). Clusters 0 (R↓) and 2 (B) displayed intermediate values (2.42 ± 2.02 and 2.85 ± 1.90), also implying certain level of sympathetic activity. Finally, cluster 3 (R↑) reported the lowest values for `lf_hf_ratio` (1.55 ± 0.75), which may reflect higher parasympathetic activity.

All the values reported for `lf_hf_ratio` are inside the typically reported values in the literature (Shaffer and Ginsberg, 2017; 2.8 ± 2.6 range $1.1 - 11.6$), even for breast cancer survivors undergoing an exercise intervention (Dias Reis et al., 2017; control group and detraining means ranging $1.3 - 2.3$, exercise group scoring 1.04 ± 0.54).

Wellness Clusters

Coherence analysis among clustering layers revealed that wellness data may be too skewed by the personal perception or attitudes of patients towards cancer and its treatment. Therefore, wellness data may not be useful to monitor training adaptation and adjust training prescriptions.

The baseline wellness clustering found different profiles in which each patient was typically reporting good, regular, or bad overall wellness status. This presence of different baseline profiles is aligned with other works in the literature for patients with cancer. For instance, Li et al. (2017) identified 3 classes of of copers using latent profile analysis. Adaptive coper had the best psychological adjustment, negative coper had the worst, and inconsistent coper had relatively high levels of psychological stress. Guimond, Ivers, and Savard (2020) identified 2 to 3 clusters of patients using latent profile analysis, finding that higher levels of avoidance were related to more severe symptoms.

This result contrast with the use of wellness in sports. Self-reported wellness may provide a complementary measure to HRV when adjusting training needs in professional athletes (Hooper & Mackinnon, 1995; Buchheit, 2014). This may be tied to the strong differences of training for a competition and fighting against cancer. The psychological states of patients with cancer may be more fragile and volatile compared to athletes', hence self-reported wellness measures cannot be interpreted in the same manner.

Finally, these baseline wellness clusters may be combined with the relative wellness of patients across the intervention in order to develop adherence strategies that consider the psychological status besides the physiological one. It is of utmost importance to break the asthenia cycles in which patients fall during systemic treatment. For this to happen, adherence to the therapeutic needs to be as higher as possible, and wellness measures may provide extra information to HRV on how to deal with patients at the moment of facing another exercise session.

5.4.2 Practical Implications

The methodology followed for data cleaning, preprocessing, and feature selection, joined with the layered clustering approach, may be put to use in other treatment scenarios. Such approach is not limited to patients with breast cancer, it could be put to use with patients with other conditions and be leveraged for interventions.

The clusters presented may be used directly in ATOPE+ as the *ML prediction model* module in the *knowledge manager* (see ATOPE+ architecture in Figure 3.1). This classification may enable a better assessment of patients' training needs. Moreover, different feature groups may be used for the clustering depending on the needs of the expert driving the intervention, since they may provide different interpretations of exercise load (for instance, using or not using `lf_hf_ratio`). The separation of HRV and wellness clusters may enable a better informed decision-making for experts with the individualization of training load and adherence strategies.

HRV clusters may enable an improved decision support for experts when assessing training exercise needs for patients. Figures 5.18 and 5.19 are representative of the information display that may be given while delivering an exercise intervention. Moreover, the values found for each feature in the clusters set narrower preliminary norms than the ones typically found in the literature for such profiles.

Wellness profiles may play a role into psychological intervention with the patients, or even into the designing of specific enrollment strategies to improve adherence to therapeutic exercise programs. Relative levels of wellness —which did not match any pattern with HRV clusters— can also be useful to this purpose. For instance, for a patient typically found in the "bad wellness" profile, it may be helpful to take advantage of those moments in which she is better than usual (despite being worse than the rest of patients), to foster her participation in the therapeutic exercise program.

5.4.3 Limitations and Future Work

The results presented should be considered preliminary. They are based on a sample of 23 patients with breast cancer with different adherence profiles during an exercise intervention. Future work should extend the results with the data from more patients. The methodology presented may be used with patients with other chronic conditions, such as other types of cancer, cardiovascular, or neurological conditions.

Future work may study the impact of training sessions in the recovery of patients, i.e., the evolution of patient across the different HRV profiles found. Moreover, future work should asses the validity and degree of improvement of different personalized interventions based on the assistance given by these tools and analysis. In particular, the efficacy of the rules registered for the ATOPE trial (registration number 2010285737407; SafeCreative ATOPE+, 2020) will be assessed.

Factors facilitating adherence to the program may be studied by taking into account demographics and the wellness profiles found (which should be complemented with measures related to the coping strategies of the patients), as well as its interaction with HRV profiles.

The relevance of Fitbit measures into this context may be studied too. The use of recent novel tools (Vega et al., 2021) may enable the collection of sophisticated context-aware measures describing the behavior or patients beyond physical activity and sleep levels.

6 Conclusion

6.1 Achievements

Therapeutic exercise poses a means to address the short- and long-term side effects of cancer and its treatment. Nevertheless, tailoring the exercise intervention to patients' training needs and capabilities has not been sufficiently addressed. Training load monitoring already provided means to personalize training intensities in professional athletes; however, this approach did not reach other potential targets such as patients with breast cancer in therapeutic exercise interventions. In order to maximize the tailoring of therapeutic exercise interventions, it is mandatory to improve the understanding of the training/recovery processes of patients. Therefore, longitudinal studies that monitor the patients' training/recovery processes are necessary to pursue personalization improvements.

These opportunities bring us back to the primary goal of this thesis: to investigate how to support personalized therapeutic exercise interventions in patients with breast cancer using mobile technologies, data science, and machine learning. This goal could only be achieved with the support of a robust interdisciplinary collaboration involving oncologists, physical therapists, engineers, and computer scientists capable of developing the necessary tools to conduct longitudinal studies in real-life scenarios. Conforming this collaboration is, without any doubt, the most relevant achievement of this thesis. None of the stated objectives would have been fulfilled without a strong partnership that enabled conducting longitudinal monitoring studies with patients with breast cancer during therapeutic exercise intervention.

We live in unprecedented times where data are more present and available than ever. Data science techniques and machine learning algorithms are still evolving and improving exponentially, delivering us the best tools to acquire new knowledge. Nonetheless, such techniques on their own may not be helpful without adequate context and interpretation. This thesis represents how, in the synergy of interdisciplinary collaborations, science can trace a better way towards more individualized, information-driven, and human-centered research.

This thesis intended to contribute to the state of the art relying on such collaboration principles. More precisely, this work aimed to provide valuable information on the design and development of data collection tools, the conduction of longitudinal experiments, and novel analysis approaches to the continuous assessment of training needs —all applied to patients with breast cancer in a therapeutic exercise intervention. The following describes the achievements of the objectives established for this thesis.

Objective 1: Develop an mHealth expert system to support personalized therapeutic exercise interventions in patients with breast cancer.

The traditional health assessment tools employed in therapeutic exercise interventions were designed to be used in a controlled environment with lab equipment, primarily in pre- and post-intervention situations. Nowadays, the availability of mobile devices and wearable sensing have improved the possibilities for the remote and reliable monitoring of training load. Nevertheless, these tools primarily target athletes or healthy populations aiming to improve their sports performance. In order to apply such methodology to therapeutic exercise intervention, patients' and experts' needs have to be considered from design.

In this work, the ATOPE+ mHealth expert system has been developed. ATOPE+ supports personalized therapeutic exercise interventions in patients with breast cancer, providing an end-to-end solution for the remote monitoring of training load (HRV and self-reported questionnaires), physical activity, and sleep using a smartphone and wearable sensors. ATOPE+ enables the automatic generation of personalized recommendations through data processing and pairing with expert rules. These recommendations allow experts to provide undulating nonlinear exercise prescriptions, hence adapting the intervention to the exercise needs of the patient. In addition, ATOPE+ provides the functionalities of an expert system with a dashboard, presenting information regarding the daily training/recovery status of patients and training recommendations triggered by expert rules.

ATOPE+ has been evaluated with the conduction of a pilot study, assessing its validity and usability. This thesis presented a usability evaluation conducted with physical therapists experienced in delivering therapeutic exercise intervention, patients with breast cancer, and survivors. Usability assessment presented excellent acceptability for both patients and experts, highlighting the simplicity and straightforwardness of using ATOPE+. Nevertheless, some difficulties were faced during the usability evaluation due to the strike of the COVID-19 pandemic. The rush of new COVID-19 cases and the fear of getting infected —especially for immunosuppressed populations like patients with cancer— hindered recruitment possibilities. This limitation forced the usability evaluation to be conducted with a mixture of breast cancer patients and survivors. On the other hand, assessing usability in a fully remote environment may be interpreted as a strength to the usability evaluation. In sum, ATOPE+ has been proved to simplify the knowledge management and decision-making process within the context of a therapeutic exercise intervention by enabling the remote and reliable monitoring of training load in patients with breast cancer.

Through these achievements, the first objective of this thesis has been fulfilled since a functional mHealth system has been designed, developed, and tested in a real-life scenario with patients with breast cancers, survivors, and physical therapists experienced in therapeutic exercise.

Objective 2: Conduct a monitorization experiment of patients with breast cancer through therapeutic exercise intervention, and generate a longitudinal dataset with training load measures.

Continuous training load monitoring has often been ignored in therapeutic exercise interventions. Performing individual daily assessments of training load depended on time-consuming tasks and sophisticated lab equipment. To date, most of the research has focused on ensuring that the minimum levels of recommended physical activity are met. In order to improve the tailoring approaches, there is a need to describe and understand the training/recovery processes of patients with breast cancer through therapeutic exercise interventions; however, there are no available examples of longitudinal monitoring of training load.

In this work, this lack of examples has been addressed by conducting a longitudinal experiment involving a population of patients with breast cancer enrolled in a therapeutic exercise intervention. Heart rate variability and self-reported wellness assessed training load, besides including daily physical activity levels and sleep patterns from wearable activity trackers. Patients were enrolled in an intervention lasting 6 to 8 weeks in which they monitored their daily status with ATOPE+. Moreover, the monitoring was maintained for some patients even after the intervention. This experiment has resulted in one of the first openly available longitudinal datasets with daily measures of training load (HRV and self-reported wellness), physical activity, and sleep, including information about training sessions, demographic data, quality of life, and treatment details too. The presented data were curated to enable different types of analysis and explorations.

Prior to the conduction of this study, the reliability of ATOPE+ was successfully assessed in separate work. ATOPE+ was found as a valid and reliable tool to assess autonomic balance (LnRMSSD), sleep satisfaction, emotional distress, and, potentially, fatigue in breast cancer survivors. This reliability provides the dataset with high applicability despite the reduced sample size. Moreover, this work also provides details on the conduction of the experiment, such as study design, the participants, eligibility criteria, and data collection, in order to facilitate the conduction of similar experiments in the future to the research community.

Through these achievements, the second objective of this thesis has been fulfilled with the publication of an openly available dataset describing the longitudinal evolution of training load, physical activity, and sleep of patients with breast cancer through therapeutic exercise intervention.

Objective 3: Identify the factors reflecting the individual recovery state of patients with breast cancer during therapeutic exercise intervention using a data science and machine learning approach.

Training adaptation has been widely studied in professional athletes to adjust to individual training needs. Nevertheless, little is known about the training adaptation process in patients with breast cancer during therapeutic exercise intervention. Although the same principles may be applied to patients with cancer and athletes, there is no previous knowledge on the evolution and the daily adaptation to training for patients. The frailty and immunosuppression induced by cancer and its treatment may play a role in the training/recovery process that still remains uncovered.

In this work, a novel methodology for the analysis of the longitudinal training load data is presented. This methodology consists of a clustering approach in which each dimension of the patient's recovery is grouped in different layers to support tailored information-driven decision-making support for the prescription of exercise. Three layers were drawn out of the analysis. The first layer was conformed by features representing the physiological status of the patient (i.e., HRV); the second layer by wellness features representing the baseline self-reported wellness status of the patient; and the third layer by z-scored (individually normalized) wellness features, which represented variations of wellness around the results of the second layer. Preliminary results for each layer represent the potential of this methodology and how its application may enable the personalized prescription of training intensities and the development of tailored adherence strategies. Moreover, the results also suggest how self-reported wellness may be prone to individual skewness in perception, highlighting the need to rely on physiological markers to assess training needs.

In addition, this methodology describes the data cleaning and preprocessing used to introduce the least amount of possible skew. A data-driven feature selection process is explained and later combined with an expert-driven interpretation of clustering results. This combination of data- and expert-driven methods helped narrow down the most relevant features for assessing training/recovery cycles in patients. Moreover, this process also showed how a traditional clustering approach, in which all features were part of the analysis, could derive skewed results that were not representative of the patients' conditions.

Through this achievement, the third objective of this thesis has been fulfilled by presenting a novel approach to study training adaptation in patients with breast cancer through therapeutic exercise intervention. Moreover, this work presented different approaches to assessing training needs, along with the interpretations and norms for the variables used in each clustering approach. This methodology may be used in other target populations such as patients with other types of cancer or even other chronic conditions such as cardiovascular or neurological.

6.2 Contributions

Section 6.1 described the fulfillment of the objectives of this thesis. Now, the main contributions of this thesis are listed:

- Identification of the requirements and challenges posed by the systems addressing the personalization of therapeutic exercise interventions in patients with cancer.
- Definition and development of an mHealth system (ATOPE+) to support the requirements of a therapeutic exercise intervention in patients with breast cancer.
- Evaluation of usability of an ATOPE+ in a real scenario with physiotherapists experts in therapeutic exercise and with patients with breast cancer and survivors.
- Collection and curation of a dataset describing the continuous monitoring of training load in patients with breast cancer enrolled in a therapeutic exercise

intervention. This dataset includes training load data, as well as demographics, treatment, and intervention details from the participants, enabling the investigation of the effects of therapeutic exercise intervention individually in patients. The dataset is publicly available to the research community in Zenodo (https://doi.org/10.5281/zenodo.6322773).

- Identification of readiness profiles in patients with breast cancer during therapeutic exercise intervention using data science and machine learning. The profiles obtained may serve to establish preliminary norms on the follow-up of future interventions with patients with breast cancer. Besides, the methodology followed may be used in other and broader cohorts of patients. The code for the analysis is publicly available to the research community in GitHub (https://github.com/salvador-moreno/atope-breast-clustering-analysis).

6.3 Outlook

The contributions of this thesis open up multiple research directions to continue and extend the work presented. This section describes some of the possible outlooks.

The design and development of mHealth systems for personalized intervention in patients with cancer pose multiple improvements. Defining adherence strategies along with personalization may lead to better interventions in those patients less eager to exercise. Gamification and behavior-change techniques may serve this purpose.

The personalization of therapeutic exercise should not be limited to patients with breast cancer. There are plenty of chronic conditions that may benefit from regular exercising, and mHealth systems addressing it may contemplate, from design, the inclusion of different cohorts of patients. This perspective would enable the investigation of the personalization of therapeutic exercise in patients with other types of cancer, cardiovascular diseases, or neurologic impairments.

The possibilities of AI in assisting data analysis and exploration are endless. Analyzing the impact of individual training sessions in patients' training adaptation employing supervised ML algorithms is a natural continuation of the work presented. The inclusion of prediction models and risk assessment would assist experts in decision-making when prescribing exercise doses. Finally, a reinforcement learning strategy might further individualize the prescription of exercise needs with an extensive definition of the metrics to optimize during therapeutic exercise intervention.

Conclusión 7

7.1 Logros

El ejercicio terapéutico se plantea como una solución para paliar los efectos secundarios del cáncer y su tratamiento a corto y largo plazo. Sin embargo, las necesidades de personalización durante una intervención de ejercicio terapéutico no se han trabajado lo suficiente. La monitorización de la carga de ejercicio es crucial para la personalización del entrenamiento en atletas profesionales; pero, a pesar de su utilidad, estos métodos aún se han intentado trasladar a otros escenarios, como pacientes con cáncer de mama en intervención de ejercicio terapéutico. Para poder mejorar la adaptación del ejercicio terapéutico a las necesidades de las pacientes, es necesario aumentar el nivel de entendimiento de los procesos de entrenamiento y recuperación en el que se ven envueltas las pacientes. Por ello, es necesario realizar estudios longitudinales que monitoricen estos procesos de forma continua para, en última instancia, buscar mejoras en la personalización de la intervención.

Estas oportunidades nos devuelven al principal objetivo de esta tesis: investigar cómo asistir en la personalización de las intervenciones de ejercicio terapéutico en pacientes con cáncer de mama utilizando las tecnologías móviles, la ciencia de datos y la inteligencia artificial. Este objetivo sólo podría alcanzarse bajo el paraguas de una colaboración interdisciplinar sólida que integrara profesionales de la oncología, fisioterapia, ingeniería e informática, una colaboración capaz de desarrollar las herramientas para llevar a cabo estudios longitudinales con pacientes en un entorno real y fuera del laboratorio. Sin duda alguna, consolidar esta colaboración ha sido el logro más importante de esta tesis, y ninguno de los objetivos dispuestos habrían sido posibles sin la fuerte relación entre todo el equipo.

Vivimos en una época sin precedentes en la que los datos están más presentes y disponibles que nunca. La ciencia de datos y la inteligencia artificial siguen evolucionando de manera exponencial, proporcionando las mejores herramientas para adquirir nuevos conocimientos. Aun así, estas técnicas pueden no ser tan útiles si su aplicación se da sin un contexto e interpretación adecuados. Esta tesis representa cómo, en la sinergia de las colaboraciones interdisciplinares, la ciencia puede trazar un mejor camino hacia una investigación más centrada en las personas y la asistencia en salud de manera informada para hacerla aún más personalizada.

La intención de esta tesis ha sido contribuir al estado de la materia desde los principios de colaboración dispuestos. Concretamente, este trabajo ha perseguido proveer a la comunidad científica con información valiosa referida al diseño y desarrollo de herramientas para la recogida de datos, a la conducción de experimentos longitudinales, y al desarrollo de nuevos métodos para el análisis continuo de las necesidades de entrenamiento —todo esto aplicado a pacientes con cáncer de mama en intervención de ejercicio terapéutico. A continuación se describen los logros referidos a cada uno de los objetivos de esta tesis.

Objetivo 1: Desarrollar un sistema experto de mSalud para asistir intervenciones de ejercicio terapéutico personalizado en pacientes con cáncer de mama.

Tradicionalmente, los métodos para evaluar a las pacientes durante una intervención de ejercicio terapéutico han dependido de equipamiento de laboratorio y mediciones en entornos controlados exclusivamente en situaciones de pre- y post-intervención. A día de hoy, los dispositivos móviles y sensores vestibles han mejorado las posibilidades para la monitorización remota y fiable de la carga del ejercicio. Sin embargo, estas herramientas han sido desarrolladas principalmente para atletas profesionales y gente sana que busca mejorar su rendimiento deportivo. Por tanto, de cara a aplicar dichos métodos en un contexto de intervención de ejercicio terapéutico, las necesidades de pacientes y expertos tienen que considerarse desde el principio.

En este trabajo se ha desarrollado el sistema experto de mSalud (salud móvil) ATOPE+. ATOPE+ permite intervenciones de ejercicio terapéutico personalizadas en pacientes con cáncer de mama, otorga una solución para la monitorización remota de la carga del ejercicio (variabilidad de la frecuencia cardíaca y bienestar auto-reportado), actividad física y sueño utilizando teléfonos móviles inteligentes y sensores vestibles. ATOPE+ permite generar recomendaciones de ejercicio terapéutico de manera automática gracias al procesamiento de datos y su emparejamiento con reglas expertas. Estas recomendaciones permiten a los expertos dar prescripciones de ejercicio no lineales y ondulatorias, es decir, que se adaptan a las necesidades de ejercicio de las pacientes de forma continua. Además, ATOPE+ provee las funcionalidades de un sistema experto con un tablero de mandos para el profesional desempeñando la intervención. Este tablero de mandos presenta información relacionada con los ciclos de entrenamiento y recuperación, y con las recomendaciones de ejercicio diarias para cada una de las pacientes y cada uno de los factores que la provocan.

ATOPE+ se ha evaluado con un estudio piloto, midiendo su validez y usabilidad. Esta tesis recoge la evaluación de usabilidad realizada con fisioterapeutas expertos en ejercicio terapéutico, pacientes con cáncer de mama y supervivientes. Se encontró una usabilidad excelente para pacientes y expertos, destacando la simplicidad y facilidad a la hora de utilizar ATOPE+, tanto su aplicación móvil como su tablero de mandos web. La pandemia del COVID-19 conllevó ciertas dificultades en su evaluación. El rápido aumento de casos unido al miedo a contagiarse —especialmente relevante para población inmunodeprimida como pacientes con cáncer— mermaron la velocidad de reclutamiento. Esta limitación forzó que la evaluación de usabilidad se realizase con una mezcla de pacientes con cáncer de mama y supervivientes. Por otro lado, también implicó evaluar la usabilidad en un entorno completamente remoto, lo cual es una fortaleza de cara a los buenos resultados. En suma, ATOPE+ se ha demostrado útil al simplificar la gestión del conocimiento y la toma de decisiones en el contexto de una intervención de ejercicio terapéutico con pacientes con cáncer de mama a través de la monitorización remota y fiable de la carga del ejercicio.

Con estos logros, el primer objetivo de la tesis se ha cumplido ya que un sistema experto mSalud ha sido exitosamente diseñado, desarrollado, y testado en un entorno real con pacientes con cáncer de mama, supervivientes, y fisioterapeutas expertos en ejercicio terapéutico.

Objetivo 2. Llevar a cabo un experimento de monitorización con pacientes con cáncer de mama durante una intervención de ejercicio terapéutico, y generar un conjunto de datos longitudinal con mediciones de la carga del ejercicio.

La monitorización continua de la carga del ejercicio ha sido normalmente ignorada en intervenciones de ejercicio terapéutico. Realizar mediciones diarias dependía de tareas prolongadas en el tiempo y equipamiento de laboratorio muy específico. Hasta la fecha, la mayor parte de las intervenciones se han centrado en asegurar el cumplimiento de los niveles de ejercicio físico mínimos recomendados. De cara a mejorar la personalización de la prescripción de ejercicio, hay que mejorar la descripción y el entendimiento de los procesos de entrenamiento y recuperación de las pacientes con cáncer de mama a lo largo de las intervenciones de ejercicio terapéutico adaptado.

En esta tesis se ha trabajado para suplir la falta de ejemplos que describan la evolución de las pacientes a lo largo de una intervención de ejercicio terapéutico. La variabilidad del ritmo cardíaco y el bienestar auto-reportado midieron la carga del ejercicio; además se incluyó la monitorización de niveles de actividad física y patrones de sueño diario desde monitores de actividad vestibles. Las pacientes participaron en una intervención de entre 6 y 8 semanas de duración en la que monitorizaron su estado diario con ATOPE+. Además, la monitorización fue mantenida para algunas de las pacientes después de la intervención. Este experimento ha resultado en el primer dataset longitudinal con mediciones diarias de la carga interna, actividad física y sueño, además de incluir información sobre las sesiones de entrenamiento, datos demográficos, calidad de vida, y detalles del tratamiento. Los datos ATOPE+Breast se presentan curados y listos para permitir distintos tipos de análisis y exploraciones.

Antes de la realización de este estudio, la fiabilidad de ATOPE+ había sido correctamente evaluada en un trabajo previo. ATOPE+ se presenta así como una herramienta válida y fiable para la medición de equilibrio en el sistema nervioso autónomo (LnRMSSD), satisfacción del sueño, estrés emocional y, potencialmente, fatiga en supervivientes de cáncer de mama. Esta fiabilidad provee al dataset ATOPE+Breast de una alta aplicabilidad a pesar de las restricciones de la muestra. Además, este trabajo provee detalles en el diseño del estudio, participantes, criterio de elegibilidad y recogida de datos de cara a facilitar la realización estudios futuros por parte de la comunidad científica.

A través de estos logros, se ha cumplido el segundo objetivo de esta tesis. Se ha publicado el dataset ATOPE+Breast describiendo la evolución de la carga del ejercicio en pacientes con cáncer de mama a lo largo de una intervención de ejercicio físico, además de sus niveles de actividad física y patrones de sueño.

Objetivo 3: Identificar los factores reflejando el estado de recuperación individual de las pacientes con cáncer de mama durante una intervención de ejercicio terapéutico utilizando técnicas de ciencia de datos y aprendizaje automático.

La adaptación al entrenamiento ha sido ampliamente estudiado en atletas profesionales para ajustar las necesidades de entrenamiento individuales. Aun así, poco se conoce sobre los procesos de adaptación al entrenamiento en pacientes con cáncer de mama durante una intervención de ejercicio terapéutico. A pesar de que se pueden aplicar los mismos principios fisiológicos entre atletas y pacientes con cáncer, no hay investigaciones que estudien la evolución y adaptación diaria de los pacientes durante el entrenamiento. La fragilidad e inmunosupresión inducida por el cáncer y su tratamiento pueden jugar un rol aún desconocido en los procesos de entrenamiento y recuperación.

En este trabajo se presenta una metodología novedosa para el análisis de la carga interna de las pacientes con cáncer de mama. Esta metodología consiste en aprovechar algoritmos de agrupamiento de forma laminar, es decir, agrupando cada dimensión de la recuperación de las pacientes en diferentes capas para mejorar la personalización de las prescripciones de ejercicio asistiendo en la toma decisiones con una información mejor y más completa. Tres capas conforman este análisis. La primera capa la componen características que reflejan el estado fisiológico de la paciente (variabilidad del ritmo cardíaco); la segunda el bienestar auto-reportado por las pacientes; y la tercera el bienestar auto-reportado normalizado individualmente para todas las pacientes (z-score). Esta última capa representa variaciones del bienestar en comparación con la segunda capa. Los resultados preliminares para cada capa representan el potencial de esta metodología, y cómo su aplicación puede permitir la personalización del ejercicio terapéutico y el desarrollo de estrategias de adherencia más sofisticadas. Además, los resultados sugieren que el bienestar auto-reportado puede estar muy sesgado por la percepción de las pacientes durante su recuperación, y que no se corresponden con los patrones encontrados en atletas, resaltando así la importancia que los marcadores fisiológicos pueden tener al medir las necesidades de entrenamiento.

Este trabajo también describe el proceso de limpieza y preprocesamiento de los datos para reducir los sesgos lo máximo posible. Se explica un proceso de selección de características que además se combina con la interpretación de expertos de cada uno de los resultados de clustering. Esta combinación de métodos basados en datos y en conocimiento experto permite discernir las variables más relevantes para medir los ciclos de entrenamiento y recuperación en pacientes. Este proceso también muestra cómo un abordaje clásico de agrupamiento, en el que todas las variables seleccionadas son parte del análisis, puede resultar en resultados sesgados que no sean verdaderamente representativos para las condiciones de las pacientes.

A través de estos logros, se ha cumplido el tercer objetivo de esta tesis. Se ha presentado una metodología novedosa para el estudio de la adaptación al entrenamiento en pacientes con cáncer de mama durante una intervención de ejercicio terapéutico. Además, este trabajo presenta distintas propuestas para la evaluación de las necesidades de entrenamiento, así como sus interpretaciones y los valores encontrados para cada una de las variables utilizadas en los resultados de agrupamiento. Esta metodología podría utilizarse en otras poblaciones, como pacientes con otros tipos de cáncer u otras condiciones de tipo crónico, como cardiovasculares o neurológicas.

7.2 Contribuciones

La sección 7.1 describió el cumplimiento de los objetivos de esta tesis. A continuación, se listan las principales contribuciones de esta tesis:

- Identificación de los requerimientos y retos mostrados por los sistemas utilizados hasta la fecha para la personalización del ejercicio terapéutico en pacientes con cáncer.

- Definición y desarrollo de un sistema mSalud (ATOPE+) para asistir en intervenciones de ejercicio terapéutico en pacientes con cáncer de mama.

- Evaluación de la usabilidad de ATOPE+ en un escenario real de intervención de ejercicio terapáutico con pacientes con cáncer de mama, supervivientes y fisioterapeutas expertos en ejercicio terapéutico.

- Coleción y curación del dataset ATOPE+Breast describiendo la monitorización de la carga del ejercicio en pacientes con cáncer de mama durante una intervención de ejercicio terapéutico. Este dataset incluye datos de la carga del ejercicio, así como detalles demográficos, de tratamiento, de la intervención y de las pacientes, permitiendo el estudio de los efectos de la intervención de manera particular en cada paciente. Este dataset está públicamente disponible para la comunidad científica en Zenodo (https://doi.org/10.5281/zenodo.6322773).

- Identificación de perfiles de recuperación en pacientes con cáncer de mama durante el ejercicio terapéutico utilizando ciencia de datos y la inteligencia artificial. Los perfiles obtenidos pueden servir para establecer valores preliminares para el seguimiento de pacientes con cáncer de mama en intervenciones futuras. Además, la metodología empleada podría utilizarse en otras cohortes de pacientes. El código para el análisis está disponible para la comunidad científica en GitHub (https://github.com/salvador-moreno/atope-breast-clustering-analysis).

7.3 Trabajo Futuro

Las contribuciones de esta tesis abren múltiples líneas de investigación para continuar y extender el trabajo presentado. Esta sección describe algunas de estas posibles líneas de trabajo.

El diseño y desarrollo de sistemas mSalud para intervenciones personalizadas en pacientes con cáncer aún presenta un amplio grado de mejora. Incluir el desarrollo de estrategias de adherencia en la personalización puede conllevar mejores intervenciones en aquellas pacientes menos dispuestas a hacer ejercicio. Integrar herramientas desde la gamificación y el cambio conductual pueden ayudar a este propósito.

La personalización del ejercicio terapéutico no debe limitarse a pacientes con cáncer de mama. Existen multitud de condiciones crónicas que pueden beneficiarse del ejercicio, y los sistemas mSalud pueden incluir en su diseño distintas cohortes de pacientes desde el principio para facilitar su uso. Esta perspectiva facilitaría la investigación de estos métodos de ejercicio físico en otros tipos de cáncer, enfermedades cardiovasculares o neurológicas, por ejemplo.

Las posibilidades de la inteligencia artificial en el análisis y exploración de los datos recogidos son inmensas. La continuación natural de este trabajo sería analizar el impacto de las sesiones de ejercicio en la adaptación al entrenamiento de las pacientes utilizando algoritmos de aprendizaje automático supervisado. La inclusión de modelos de predicción y evaluación del riesgo asistirían mejor a los expertos a la hora de prescribir ejercicio. Finalmente, una estrategia de aprendizaje por refuerzo podría terminar de ayudar a la individualización del ejercicio tras una definición exhaustiva de las métricas a optimizar durante una intervención de ejercicio terapéutico.

Bibliography

Adams, W. C. (2015). Conducting Semi-Structured Interviews. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (pp. 492–505). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Afzal, M., Ali, S. I., Ali, R., Hussain, M., Ali, T., Khan, W. A., … Lee, S. (2018). Personalization of wellness recommendations using contextual interpretation. *Expert Systems with Applications*, *96*, 506–521.

Aggarwal, C. C. (2015). *Data Mining*. Cham: Springer International Publishing.

Aggarwal, C. C. (2016). *Recommender Systems*. Cham: Springer International Publishing.

Aguilar Cordero, M. J., Sánchez López, A. M., Mur Villar, N., García García, I., Rodríguez López, M. A., Ortegón Piñero, A., & Cortés Castell, E. (2014). Salivary cortisol as an indicator of physological stress in children and adults; a systematic review. *Nutricion Hospitalaria*, *29*(5), 960–968.

Alelyani, S., Tang, J., & Liu, H. (2018). Feature Selection for Clustering: A Review. In C. C. Aggarwal & C. K. Reddy (Eds.), *Data Clustering* (1st ed., pp. 29–60). Chapman and Hall/CRC.

Alharthi, R., Alharthi, R., Guthier, B., & El Saddik, A. (2019). CASP: Context-aware stress prediction system. *Multimedia Tools and Applications*, *78*(7), 9011–9031.

Ali, R., Afzal, M., Sadiq, M., Hussain, M., Ali, T., Lee, S., & Khattak, A. M. (2018). Knowledge-based reasoning and recommendation framework for intelligent decision making. *Expert Systems*, *35*(2), e12242.

American Cancer Society. (2016). Summary of the ACS Guidelines on Nutrition and Physical Activity. Accessed 2020-03-16.

Apple. (n.d.). Universal Links for Developers. Accessed 26 June 2020.

Armenian, S. H., Lacchetti, C., Barac, A., Carver, J., Constine, L. S., Denduluri, N., … Lenihan, D. (2017). Prevention and Monitoring of Cardiac Dysfunction in Survivors of Adult Cancers: American Society of Clinical Oncology Clinical Practice Guideline. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *35*(8), 893–911.

Arthur, D. & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (pp. 1027–1035). SODA '07. USA: Society for Industrial and Applied Mathematics.

Bailon, C., Damas, M., Pomares, H., Sanabria, D., Perakakis, P., Goicoechea, C., & Banos, O. (2019). Smartphone-Based Platform for Affect Monitoring through Flexibly Managed Experience Sampling Methods. *Sensors*, *19*(15), 3430.

Ballard-Barbash, R., Friedenreich, C. M., Courneya, K. S., Siddiqi, S. M., McTiernan, A., & Alfano, C. M. (2012). Physical Activity, Biomarkers, and Disease Outcomes in Cancer Survivors: A Systematic Review. *JNCI Journal of the National Cancer Institute*, *104*(11), 815–840.

Banos, O., Amin, M., Khan, W., Ali, T., Afzal, M., Kang, B., & Lee, S. (2015). Mining Minds: An innovative framework for personalized health and wellness support. ICST.

Banos, O., Bilal Amin, M., Ali Khan, W., Afzal, M., Hussain, M., Kang, B. H., & Lee, S. (2016). The Mining Minds digital health and wellness framework. *BioMedical Engineering OnLine*, *15*(S1), 76.

Banos, O., Jaehun Bang, Taeho Hur, Hameed Siddiqi, M., Huynh-The Thien, Le-Ba Vui, ... Sungyoung Lee. (2015). Mining human behavior for health promotion. (pp. 5062–5065). IEEE.

Banos, O., Moral-Munoz, J. A., Diaz-Reyes, I., Arroyo-Morales, M., Damas, M., Herrera-Viedma, E., ... Villalonga, C. (2015). mDurance: A Novel Mobile Health System to Support Trunk Endurance Assessment. *Sensors*, *15*(6), 13159–13183. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

Banos, O., Villalonga, C., Garcia, R., Saez, A., Damas, M., Holgado-Terriza, J. A., ... Rojas, I. (2015). Design, implementation and validation of a novel open framework for agile development of mobile health applications. *BioMedical Engineering OnLine*, *14*(Suppl 2), S6.

Bardram, J. E. (2020). The CARP Mobile Sensing Framework – a Cross-platform, Reactive, Programming Framework and Runtime Environment for Digital Phenotyping. *arXiv:2006.11904 [cs]*. arXiv: 2006.11904.

Beauchamp, U. L., Pappot, H., & Holländer-Mieritz, C. (2020). The Use of Wearables in Clinical Trials During Cancer Treatment: Systematic Review. *JMIR mHealth and uHealth*, *8*(11), e22006.

Beck, S. L., Schwartz, A. L., Towsley, G., Dudley, W., & Barsevick, A. (2004). Psychometric evaluation of the Pittsburgh sleep quality index in cancer patients. *Journal of Pain and Symptom Management*, *27*(2), 140–148.

Beg, M. S., Gupta, A., Stewart, T., & Rethorst, C. D. (2017). Promise of Wearable Physical Activity Monitors in Oncology Practice. *Journal of Oncology Practice*, *13*(2), 82–89.

Billman, G. (2013). The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Frontiers in Physiology*, *4*, 26.

Bluethmann, S. M., Vernon, S. W., Gabriel, K. P., Murphy, C. C., & Bartholomew, L. K. (2015). Taking the next step: A systematic review and meta-analysis of physical activity and behavior change interventions in recent post-treatment breast cancer survivors. *Breast Cancer Research and Treatment*, *149*(2), 331–342.

Børøsund, E., Mirkovic, J., Clark, M. M., Ehlers, S. L., Andrykowski, M. A., Bergland, A., ... Solberg Nes, L. (2018). A Stress Management App Intervention for Cancer Survivors: Design, Development, and Usability Testing. *JMIR formative research*, *2*(2), e19.

Børøsund, E., Varsi, C., Clark, M. M., Ehlers, S. L., Andrykowski, M. A., Sleveland, H. R. S., ... Nes, L. S. (2020). Pilot testing an app-based stress management intervention for cancer survivors. *Translational Behavioral Medicine*, *10*(3), 770–780.

Borresen, J. & Lambert, M. I. (2008). Autonomic control of heart rate during and after exercise : Measurements and implications for monitoring training status. *Sports Medicine (Auckland, N.Z.) 38*(8), 633–646.

Bourdon, P. C., Cardinale, M., Murray, A., Gastin, P., Kellmann, M., Varley, M. C., ... Cable, N. T. (2017). Monitoring Athlete Training Loads: Consensus Statement. *International Journal of Sports Physiology and Performance*, *12*(Suppl 2), S2161–S2170.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424.

Brickwood, K.-J., Watson, G., O'Brien, J., & Williams, A. D. (2019). Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. *JMIR mHealth and uHealth*, *7*(4), e11819.

Brinton, L. A., Gaudet, M. M., & Gierach, G. L. (2017). Breast Cancer. In M. Thun, M. S. Linet, J. R. Cerhan, C. A. Haiman, & D. Schottenfeld (Eds.), *Cancer Epidemiology and Prevention* (4th ed., pp. 861–888). New York: Oxford University Press.

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. Jordan, B. Thomas, B. Weerdmeester, & I. McClelland (Eds.), *Usability Evaluation in Industry*. London: Taylor and Francis.

Buchheit, M. (2014). Monitoring training status with HR measures: Do all roads lead to Rome. *Frontiers in physiology*, *5*, 73.

Burns, M. N., Begale, M., Duffecy, J., Gergle, D., Karr, C. J., Giangrande, E., & Mohr, D. C. (2011). Harnessing Context Sensing to Develop a Mobile Intervention for Depression. *Journal of Medical Internet Research*, *13*(3).

Caminal, P., Sola, F., Gomis, P., Guasch, E., Perera, A., Soriano, N., & Mont, L. (2018). Validity of the Polar V800 monitor for measuring heart rate variability in mountain running route conditions. *European Journal of Applied Physiology*, *118*(3), 669–677.

Campbell, K. L., Winters-Stone, K. M., Wiskemann, J., May, A. M., Schwartz, A. L., Courneya, K. S., ... Schmitz, K. H. (2019). Exercise Guidelines for Cancer Survivors: Consensus Statement from International Multidisciplinary Roundtable. *Medicine and Science in Sports and Exercise*, *51*(11), 2375–2390.

Cantarero-Villanueva, I. (2020). *Attenuating Cancer Treatment-related Toxicity in Oncology Patients With a Tailored Physical Exercise Program: Study Protocol of the ATOPE Trial in Women Recently Diagnosed With Breast Cancer* (Clinical trial registration No. NCT03787966). clinicaltrials.gov. submitted: December 14, 2018. Bethesda (MD): National Library of Medicine (US).

Cantarero-Villanueva, I., Fernández-Lao, C., Díaz-Rodríguez, L., Cuesta-Vargas, A. I., Fernández-de-las-Peñas, C., Piper, B. F., & Arroyo-Morales, M. (2014). The Piper Fatigue Scale-Revised: Translation and psychometric evaluation in Spanish-speaking breast cancer survivors. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *23*(1), 271–276.

Carney, C. E., Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Krystal, A. D., Lichstein, K. L., & Morin, C. M. (2012). The Consensus Sleep Diary: Standardizing Prospective Sleep Self-Monitoring. *Sleep*, *35*(2), 287–302.

Caro-Moran, E., Fernandez-Lao, C., Galiano-Castillo, N., Cantarero-Villanueva, I., Arroyo-Morales, M., & Diaz-Rodriguez, L. (2016). Heart Rate Variability in Breast Cancer Survivors After the First Year of Treatments: A Case-Controlled Study. *Biological Research for Nursing*, *18*(1), 43–49.

Carter, S. J., Baranauskas, M. N., Ballinger, T. J., Rogers, L. Q., Miller, K. D., & Nabhan, D. C. (2021). Exercise load monitoring: Integrated approaches to advance the individualisation of exercise oncology. *BMJ Open Sport — Exercise Medicine*, *7*(3), e001134.

Castelyn, G., Laranjo, L., Schreier, G., & Gallego, B. (2021). Predictive performance and impact of algorithms in remote monitoring of chronic conditions: A systematic review and meta-analysis. *International Journal of Medical Informatics*, 104620.

Castro, L. A., Favela, J., Quintana, E., & Perez, M. (2015). Behavioral data gathering for assessing functional status and health in older adults using mobile phones. *Personal and Ubiquitous Computing*, *19*(2), 379–391.

Champseix, R. & contributors. (2020). Aura-healthcare/hrvanalysis. original-date: 2018-09-13T09:27:19Z. Association AURA.

Chung, I. Y., Jung, M., Park, Y. R., Cho, D., Chung, H., Min, Y. H., … Lee, J. W. (2020). Exercise Promotion and Distress Reduction Using a Mobile App-Based Community in Breast Cancer Survivors. *Frontiers in Oncology*, *9*.

Cinaz, B., Arnrich, B., La Marca, R., & Tröster, G. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. *Personal and Ubiquitous Computing*, *17*(2), 229–239.

Cormie, P., Zopf, E. M., Zhang, X., & Schmitz, K. H. (2017). The Impact of Exercise on Cancer Mortality, Recurrence, and Treatment-Related Adverse Effects. *Epidemiologic Reviews*, *39*(1), 71–92.

Cutillo, A., O'Hea, E., Person, S., Lessard, D., Harralson, T., & Boudreaux, E. (2017). NCCN Distress Thermometer: Cut off Points and Clinical Utility. *Oncology nursing forum*, *44*(3), 329–336.

D'Amore, B., Smolinski-Zhao, S., Daye, D., & Uppot, R. N. (2021). Role of Machine Learning and Artificial Intelligence in Interventional Oncology. *Current Oncology Reports*, *23*(6), 70.

De Couck, M. & Gidron, Y. (2013). Norms of vagal nerve activity, indexed by Heart Rate Variability, in cancer patients. *Cancer Epidemiology*, *37*(5), 737–741.

Dias Reis, A., Silva Garcia, J. B., Rodrigues Diniz, R., Silva-Filho, A. C., Dias, C. J., Leite, R. D., & Mostarda, C. (2017). Effect of exercise training and detraining in autonomic modulation and cardiorespiratory fitness in breast cancer survivors. *The Journal of Sports Medicine and Physical Fitness*, *57*(7-8), 1062–1068.

Dinh-Le, C., Chuang, R., Chokshi, S., & Mann, D. (2019). Wearable Health Technology and Electronic Health Record Integration: Scoping Review and Future Directions. *JMIR mHealth and uHealth*, *7*(9), e12861.

Dorri, S., Asadi, F., Olfatbakhsh, A., & Kazemi, A. (2020). A Systematic Review of Electronic Health (eHealth) interventions to improve physical activity in patients with breast cancer. *Breast Cancer (Tokyo, Japan)*, *27*(1), 25–46.

Dreher, N., Hadeler, E. K., Hartman, S. J., Wong, E., Acerbi, I., Rugo, H. S., … Melisko, M. (2019). Fitbit usage in breast cancer patients undergoing chemotherapy. *Clinical Breast Cancer*.

Duan, Y., Shang, B., Liang, W., Du, G., Yang, M., & Rhodes, R. E. (2021). Effects of eHealth-Based Multiple Health Behavior Change Interventions on Physical Activity, Healthy Diet, and Weight in People With Noncommunicable Diseases: Systematic Review and Meta-analysis. *Journal of Medical Internet Research*, *23*(2), e23786.

Ertuğrul, D. Ç. & Elçi, A. (2019). A survey on semanticized and personalized health recommender systems. *Expert Systems*, *e12519*, 1–23.

Estrin, D. (2014). Small data, where n = me. *Communications of the ACM, 57*(4), 32–34.

European Commission. (2021). European Union Mission Cancer Implementation Plan.

Faro, J. M., Mattocks, K. M., Nagawa, C. S., Lemon, S. C., Wang, B., Cutrona, S. L., & Sadasivam, R. S. (2021). Physical Activity, Mental Health, and Technology Preferences to Support Cancer Survivors During the COVID-19 Pandemic: Cross-sectional Study. *JMIR Cancer*, *7*(1), e25317.

Felix Angelov. (n.d.). Bloc: A predictable state management library for Dart. Available online: https://bloclibrary.dev/ (accessed on 15 June 2020).

Ferreira, D., Kostakos, V., & Dey, A. K. (2015). AWARE: Mobile Context Instrumentation Framework. *Frontiers in ICT*, *2*.

Fitbit Inc. (n.d.). Fitbit Development: Web API. Available online: https://dev.fitbit.com/build/reference/web-api/ (accessed on 12 July 2019).

Fong, D. Y. T., Ho, J. W. C., Hui, B. P. H., Lee, A. M., Macfarlane, D. J., Leung, S. S. K., … Cheng, K.-K. (2012). Physical activity for cancer survivors: Meta-analysis of randomised controlled trials. *Bmj-British Medical Journal*, *344*, e70.

Fonseka, L. N. & Woo, B. K. (2021). Consumer Wearables and the Integration of New Objective Measures in Oncology: Patient and Provider Perspectives. *JMIR mHealth and uHealth*, *9*(7), e28664.

Forner-Llacer, F., Aranda-Malaves, R., Aranda Malavés, R., Calleja-Gonzalez, J., Perez-Turpin, J. A., & Gonzalez-Rodenas, J. (2020). Minimal Stabilization Time for Ultra-short Heart Rate Variability Measurements in Professional Soccer. *International Journal of Sports Medicine*, *41*(14), 1032–1038.

Freigoun, M. T., Martín, C. A., Magann, A. B., Rivera, D. E., Phatak, S. S., Korinek, E. V., & Hekler, E. B. (2017). System identification of Just Walk: A behavioral mHealth intervention for promoting physical activity. In *2017 American Control Conference (ACC)* (pp. 116–121). ISSN: 2378-5861.

Freitag, N., Weber, P. D., Sanders, T. C., Schulz, H., Bloch, W., & Schumann, M. (2018). High-intensity interval training and hyperoxia during chemotherapy: A case report about the feasibility, safety and physical functioning in a colorectal cancer patient. *Medicine*, *97*(24), e11068.

Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M. A., Simango, B., … Taylor, N. G. A. (2020). Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. *JMIR mHealth and uHealth*, *8*(9), e18694.

Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard, K. J., & Tørresen, J. (2018). Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing*, *51*, 1–26.

Garcia, D. O. & Thomson, C. A. (2014). Physical Activity and Cancer Survivorship. *Nutrition in Clinical Practice*, *29*(6), 768–779.

Gehlbach, H. & Barge, S. (2012). Anchoring and Adjusting in Questionnaire Responses. *Basic and Applied Social Psychology*.

Gell, N. M., Grover, K. W., Humble, M., Sexton, M., & Dittus, K. (2017). Efficacy, feasibility, and acceptability of a novel technology-based intervention to support physical activity in cancer survivors. *Supportive Care in Cancer*, *25*(4), 1291–1300.

Gell, N. M., Tursi, A., Grover, K. W., & Dittus, K. (2020). Female cancer survivor perspectives on remote intervention components to support physical activity maintenance. *Supportive Care in Cancer*, *28*(5), 2185–2194.

Giesinger, J. M., Kieffer, J. M., Fayers, P. M., Groenvold, M., Petersen, M. A., Scott, N. W., … EORTC Quality of Life Group. (2016). Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. *Journal of Clinical Epidemiology*, *69*, 79–88.

Giles, D. A. & Draper, N. (2018). Heart Rate Variability During Exercise: A Comparison of Artefact Correction Methods. *Journal of Strength and Conditioning Research*, *32*(3), 726–735.

Gilgen-Ammann, R., Schweizer, T., & Wyss, T. (2019). RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. *European Journal of Applied Physiology*, *119*(7), 1525–1532.

Ginzac, A., Passildas, J., Gadéa, E., Abrial, C., Molnar, I., Trésorier, R., … Durando, X. (2019). Treatment-Induced Cardiotoxicity in Breast Cancer: A Review of the Interest of Practicing a Physical Activity. *Oncology*, *96*(5), 223–234.

Gleeson, M. (2002). Biochemical and immunological markers of over-training. *Journal of Sports Science & Medicine*, *1*(2), 31–41.

Gomersall, S. R., Skinner, T. L., Winkler, E., Healy, G. N., Eakin, E., & Fjeldsoe, B. (2019). Feasibility, acceptability and efficacy of a text message-enhanced clinical exercise rehabilitation intervention for increasing 'whole-of-day' activity in people living with and beyond cancer. *BMC Public Health*, *19*(Suppl 2).

Granja, C., Janssen, W., & Johansen, M. A. (2018). Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. *Journal of Medical Internet Research*, *20*(5), e10235.

Gravetter, F. J. & Wallnau, L. B. (2014). *Essentials of Statistics for the Behavioral Sciences* (8th). Wadsworth, Belmont, CA: Cengage Learning.

Gresham, G., Schrack, J., Gresham, L. M., Shinde, A. M., Hendifar, A. E., Tuli, R., … Piantadosi, S. (2018). Wearable activity monitors in oncology trials: Current use of an emerging technology. *Contemporary Clinical Trials*, *64*, 13–21.

Guimond, A.-J., Ivers, H., & Savard, J. (2020). Clusters of Psychological Symptoms in Breast Cancer: Is There a Common Psychological Mechanism? *Cancer Nursing*, *43*(5), 343–353.

Guo, Y., Koshy, S., Hui, D., Palmer, J. L., Shin, K., Bozkurt, M., & Yusuf, S. W. (2015). Prognostic Value of Heart Rate Variability in Patients with Cancer. *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*, *32*(6), 516–520.

Ha, D., Malhotra, A., Ries, A. L., O'Neal, W. T., & Fuster, M. M. (2019). Heart rate variability and heart rate recovery in lung cancer survivors eligible for long-term cure. *Respiratory Physiology & Neurobiology*, *269*, 103264.

Haghayegh, S., Khoshnevis, S., Smolensky, M. H., Diller, K. R., & Castriotta, R. J. (2019). Accuracy of Wristband Fitbit Models in Assessing Sleep: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, *21*(11), e16273.

Halson, S. L. (2014). Monitoring training load to understand fatigue in athletes. *Sports Medicine*, *44 Suppl 2*, S139–S147.

Hartman, S. J., Nelson, S. H., Myers, E., Natarajan, L., Sears, D. D., Palmer, B. W., … Patterson, R. E. (2018). Randomized controlled trial of increasing physical activity on objectively measured and self-reported cognitive functioning among breast cancer survivors: The memory & motion study. *Cancer*, *124*(1), 192–202.

Hartman, S. J., Nelson, S. H., & Weiner, L. S. (2018). Patterns of Fitbit Use and Activity Levels Throughout a Physical Activity Intervention: Exploratory Analysis from a Randomized Controlled Trial. *JMIR mHealth and uHealth*, *6*(2), e29.

Hatton, A. L., Menant, J. C., Lord, S. R., Lo, J. C. M., & Sturnieks, D. L. (2013). The effect of lower limb muscle fatigue on obstacle negotiation during walking in older adults. *Gait & Posture*, *37*(4), 506–510.

Hawley, J. A., Hargreaves, M., Joyner, M. J., & Zierath, J. R. (2014). Integrative biology of exercise. *Cell*, *159*(4), 738–749.

Hekler, E. B., Klasnja, P., Chevance, G., Golaszewski, N. M., Lewis, D., & Sim, I. (2019). Why we need a small data paradigm. *BMC Medicine*, *17*(1), 133.

Hernando, D., Roca, S., Sancho, J., Alesanco, Á., & Bailón, R. (2018). Validation of the Apple Watch for Heart Rate Variability Measurements during Relax and Mental Stress in Healthy Subjects. *Sensors*, *18*(8), 2619.

Hidding, J. T., Beurskens, C. H. G., Wees, P. J. v. d., Laarhoven, H. W. M. v., & Sanden, M. W. G. N.-v. d. (2014). Treatment Related Impairments in Arm and Shoulder in Patients with Breast Cancer: A Systematic Review. *PLOS ONE*, *9*(5), e96748.

Hooper, S. L. & Mackinnon, L. T. (1995). Monitoring overtraining in athletes. Recommendations. *Sports Medicine*, *20*(5), 321–327.

Hopkins, W. G. [W. G.]. (2000). Measures of reliability in sports medicine and science. *Sports Medicine (Auckland, N.Z.) 30*(1), 1–15.

Hopkins, W. G. [Will G]. (2004). How to Interpret Changes in an Athletic Performance Test. *Sportscience*, *8*, 1–7.

Hossain, S. M., Hnat, T., Saleheen, N., Nasrin, N. J., Noor, J., Ho, B.-J., . . . Kumar, S. (2017). mCerebrum: A Mobile Sensing Software Platform for Development and Validation of Digital Biomarkers and Interventions. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems* (pp. 1–14). SenSys '17. New York, NY, USA: Association for Computing Machinery.

Insel, T. R. (2017). Digital Phenotyping: Technology for a New Science of Behavior. *JAMA*, *318*(13), 1215–1216.

International Standards Organization (ISO). (2014). Health informatics – Capacity-based eHealth architecture roadmap – Part 2: Architectural components and maturity model ISO/TR 14639-2:2014. Geneva: International Standards Organization.

Javaloyes, A., Sarabia, J. M., Lamberts, R. P., & Moya-Ramon, M. (2019). Training Prescription Guided by Heart-Rate Variability in Cycling. *International Journal of Sports Physiology and Performance*, *14*(1), 23–32.

Jones, L. W. (2015). Precision Oncology Framework for Investigation of Exercise As Treatment for Cancer. *Journal of Clinical Oncology*, *33*(35), 4134–4137.

Jones, L. W., Eves, N. D., & Scott, J. M. (2018). Bench-to-Bedside Approaches for Personalized Exercise Therapy in Cancer. *American Society of Clinical Oncology Educational Book*.

Kaikkonen, P. (2015). *Post-exercise heart rate variability: A new approach to evaluation of exercise-induced physiological training load* (Doctoral dissertation, Jyväskylä University, Jyväskylä). ISBN 978-951-39-6218-0 ISSN 0356-1070.

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., . . . Kumar, V. (2017). Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Transactions on Knowledge and Data Engineering*, *29*(10), 2318–2331. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

Kelly, C. M. & Shahrokni, A. (2016). Moving beyond Karnofsky and ECOG Performance Status Assessments with New Technologies. *Journal of Oncology*, *2016*, 1–13.

Khan, M., Mehran, M. T., Haq, Z. U., Ullah, Z., Naqvi, S. R., Ihsan, M., & Abbass, H. (2021). Applications of artificial intelligence in COVID-19 pandemic: A comprehensive review. *Expert Systems with Applications*, *185*, 115695.

Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., & Koo, B.-H. (2018). Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investigation*, *15*(3), 235–245.

Kiviniemi, A. M., Hautala, A. J., Seppänen, T., Mäkikallio, T. H., Huikuri, H. V., & Tulppo, M. P. (2004). Saturation of high-frequency oscillations of r-r intervals in healthy subjects and patients after acute myocardial infarction during ambulatory conditions. *American Journal of Physiology-Heart and Circulatory Physiology*, *287*(5), H1921–H1927.

Kluess, H. A., Wood, R. H., & Welsch, M. A. (2000). Vagal modulation of the heart and central hemodynamics during handgrip exercise. *American Journal of Physiology. Heart and Circulatory Physiology*, *278*(5), H1648–1652.

Kumar, D., Jeuris, S., Bardram, J. E., & Dragoni, N. (2021). Mobile and Wearable Sensing Frameworks for mHealth Studies and Applications: A Systematic Review. *ACM Transactions on Computing for Healthcare*, *2*(1), 8:1–8:28.

Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research – Recommendations for Experiment Planning, Data Analysis, and Data Reporting. *Frontiers in Psychology*, *8*, 213.

Lahart, I. M., Metsios, G. S., Nevill, A. M., & Carmichael, A. R. (2015). Physical activity, risk of death and recurrence in breast cancer survivors: A systematic review and meta-analysis of epidemiological studies. *Acta Oncologica*, *54*(5), 635–654.

Laurent, C. M., Green, J. M., Bishop, P. A., Sjökvist, J., Schumacker, R. E., Richardson, M. T., & Curtner-Smith, M. (2011). A Practical Approach to Monitoring Recovery: Development of a Perceived Recovery Status Scale. *The Journal of Strength & Conditioning Research*, *25*(3), 620–628.

LaVoy, E. C. P., Fagundes, C. P., & Dantzer, R. (2016). Exercise, inflammation, and fatigue in cancer survivors. *Exercise Immunology Review*, *22*, 82–93.

Lee, Y.-H., Lai, G.-M., Lee, D.-C., Tsai Lai, L.-J., & Chang, Y.-P. (2018). Promoting Physical and Psychological Rehabilitation Activities and Evaluating Potential Links Among Cancer-Related Fatigue, Fear of Recurrence, Quality of Life, and Physiological Indicators in Cancer Survivors. *Integrative Cancer Therapies*, *17*(4), 1183–1194.

Lee, K. (2021). Clinical Exercise Prescription for Cardiovascular Health in Breast Cancer Survivors. *The Korean Journal of Sports Medicine*, *39*(3), 77–90. Publisher: The Korean Society of Sports Medicine.

Li, L., Li, S., Wang, Y., Yi, J., Yang, Y., He, J., & Zhu, X. (2017). Coping Profiles Differentiate Psychological Adjustment in Chinese Women Newly Diagnosed With Breast Cancer. *Integrative Cancer Therapies*, *16*(2), 196–204.

Liang, Z. & Martell, M. A. C. (2018). Validity of Consumer Activity Wristbands and Wearable EEG for Measuring Overall Sleep Parameters and Sleep Structure in Free-Living Conditions. *Journal of Healthcare Informatics Research*, 1–27.

Low, C. A. [Carissa A.]. (2020). Harnessing consumer smartphone and wearable sensors for clinical cancer research. *npj Digital Medicine*, *3*(1), 1–7.

Low, C. A. [Carissa A.], Dey, A. K., Ferreira, D., Kamarck, T., Sun, W., Bae, S., & Doryab, A. (2017). Estimation of Symptom Severity During Chemotherapy From Passively Sensed Data: Exploratory Study. *Journal of Medical Internet Research*, *19*(12), e420.

Low, C. A. [Carissa A.], Li, M., Vega, J., Durica, K. C., Ferreira, D., Tam, V., . . . Dey, A. K. (2021). Digital Biomarkers of Symptom Burden Self-Reported by Perioperative Patients Undergoing Pancreatic Surgery: Prospective Longitudinal Study. *JMIR Cancer*, *7*(2), e27975.

Low, C. A. [Carissa A], Bovbjerg, D. H., Ahrendt, S., Choudry, M. H., Holtzman, M., Jones, H. L., . . . Bartlett, D. L. (2018). Fitbit step counts during inpatient recovery from cancer surgery as a predictor of readmission. *Annals of Behavioral Medicine*, *52*(1), 88–92.

Lozano-Lozano, M., Galiano-Castillo, N., Martín-Martín, L., Pace-Bedetti, N., Fernández-Lao, C., Arroyo-Morales, M., & Cantarero-Villanueva, I. (2018). Monitoring Energy Balance in Breast Cancer Survivors Using a Mobile App: Reliability Study. *JMIR mHealth and uHealth*, *6*(3), e67.

Ltd., F. T. (2014). *Stress and Recovery Analysis Method Based on 24-hour Heart Rate Variability*. Firstbeat Technologies Ltd. Available in: https://assets.firstbeat.com/firstbeat/uploads/2015/10/Stress-and-recovery_white-paper_20145.pdf (accessed on 20 September 2020).

Marcolino, M. S., Queiroz Oliveira, J. A., D'Agostino, M., Ribeiro, A. L., Moreira Alkmim, M. B., & Novillo-Ortiz, D. (2018). The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. *Jmir Mhealth and Uhealth*, *6*(1), e23.

Martin, E., Meglio, A. D., Charles, C., Ferreira, A., Gbenou, A., Blond, M., . . . Luis, I. V. (2021). Use of mHealth to Increase Physical Activity Among Breast Cancer Survivors With Fatigue: Qualitative Exploration. *JMIR Cancer*, *7*(1), e23927.

Maxwell-Smith, C., Cohen, P. A., Platell, C., Tan, P., Levitt, M., Salama, P., . . . Hardcastle, S. J. (2018). Wearable Activity Technology And Action-Planning (WATAAP) to promote physical activity in cancer survivors: Randomised controlled trial protocol. *International Journal of Clinical and Health Psychology*, *18*(2), 124–132.

McLester, C. N., Nickerson, B. S., Kliszczewicz, B. M., & McLester, J. R. (2020). Reliability and Agreement of Various InBody Body Composition Analyzers as Compared to Dual-Energy x-Ray Absorptiometry in Healthy Men and Women. *Journal of Clinical Densitometry*, *23*(3), 443–450.

McTiernan, A., Friedenreich, C. M., Katzmarzyk, P. T., Powell, K. E., Macko, R., Buchner, D., . . . 2018 PHYSICAL ACTIVITY GUIDELINES ADVISORY COMMITTEE*. (2019). Physical Activity in Cancer Prevention and Survival: A Systematic Review. *Medicine and Science in Sports and Exercise*, *51*(6), 1252–1261.

Mehrotra, A., Tsapeli, F., Hendley, R., & Musolesi, M. (2017). MyTraces: Investigating Correlation and Causation between Users' Emotional States and Mobile Phone Interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 1–21.

Mendoza, J. A., Baker, K. S., Moreno, M. A., Whitlock, K., Abbey-Lambertz, M., Waite, A., . . . Chow, E. J. (2017). A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: A pilot study. *Pediatric Blood & Cancer*, *64*(12), 1–9.

Merkel, D. (2014). Docker: Lightweight Linux containers for consistent development and deployment. *Linux Journal*, *2014*(239), 2:2.

Miguel, M., Oliveira, R., Loureiro, N., García-Rubio, J., & Ibáñez, S. J. (2021). Load Measures in Training/Match Monitoring in Soccer: A Systematic Review. *International Journal of Environmental Research and Public Health*, *18*(5), 2721.

Migueles, J. H., Aadland, E., Andersen, L. B., Brønd, J. C., Chastin, S. F., Hansen, B. H., . . . Ortega, F. B. (2021). GRANADA consensus on analytical approaches to assess associations with accelerometer-determined physical behaviours (physical activity, sedentary behaviour and sleep) in epidemiological studies. *British Journal of Sports Medicine*.

Min, Y. H., Lee, J. W., Shin, Y.-W., Jo, M.-W., Sohn, G., Lee, J.-H., . . . Ahn, S. H. (2014). Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: A feasibility study. *Journal of Medical Internet Research*, *16*(5), e135.

Moo, T.-A., Sanford, R., Dang, C., & Morrow, M. (2018). Overview of Breast Cancer Therapy. *PET clinics*, *13*(3), 339–354.

Moreno Gutiérrez, S., Cantarero Villanueva, I., Postigo Martín, E. P., Arroyo Morales, M., Martín Martín, L., López Garzón, M. d. l. C., . . . Baños Legrán, O. (2020). *ATOPE+*. 2010285737407. España.

Muller, A. M., Maher, C. A., Vandelanotte, C., Hingle, M., Middelweerd, A., Lopez, M. L., . . . Wark, P. A. (2018). Physical Activity, Sedentary Behavior, and Diet-Related eHealth and mHealth Research: Bibliometric Analysis. *Journal of Medical Internet Research*, *20*(4), e122.

Nabil, D. & Bereksi Reguig, F. (2015). Ectopic beats detection and correction methods: A review. *Biomedical Signal Processing and Control*, *18*, 228–244.

Nelson, B. W., Low, C. A., Jacobson, N., Areán, P., Torous, J., & Allen, N. B. (2020). Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. *npj Digital Medicine*, *3*(1), 1–9.

Niederer, D., Vogt, L., Thiel, C., Schmidt, K., Bernhörster, M., Lungwitz, A., ... Banzer, W. (2012). Exercise Effects on HRV in Cancer Patients. *International Journal of Sports Medicine*, *34*(01), 68–73.

Nielsen, J. (2002). Deep Linking is Good Linking. Available online: https://www.nngroup.com/articles/deep-linking-is-good-linking/ (accessed on 26 July 2020).

Nweke, H. F., Teh, Y. W., Mujtaba, G., & Al-garadi, M. A. (2019). Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Information Fusion*, *46*, 147–170.

Nyrop, K. A., Deal, A. M., Choi, S. K., Wagoner, C. W., Lee, J. T., Wood, A., ... Muss, H. B. (2018). Measuring and understanding adherence in a home-based exercise intervention during chemotherapy for early breast cancer. *Breast Cancer Research and Treatment*, *168*(1), 43–55.

O'Connor, Y., Rowan, W., Lynch, L., & Heavin, C. (2017). Privacy by Design: Informed Consent and Internet of Things for Smart Health. *Procedia Computer Science*. The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017) / The 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2017) / Affiliated Workshops, *113*, 653–658.

Orcas & contributors. (2019). Python-Fitbit: Fitbit API Python Client Implementation (version 0.3.1). Available online: https://github.com/orcasgit/python-fitbit (accessed on 12 July 2019). ORCAS.

Patel, A. V., Friedenreich, C. M., Moore, S. C., Hayes, S. C., Silver, J. K., Campbell, K. L., ... Matthews, C. E. (2019). American College of Sports Medicine Roundtable Report on Physical Activity, Sedentary Behavior, and Cancer Prevention and Control. *Medicine and Science in Sports and Exercise*, *51*(11), 2391–2402.

Patsou, E. D., Alexias, G. T., Anagnostopoulos, F. G., & Karamouzis, M. V. (2018). Physical activity and sociodemographic variables related to global health, quality of life, and psychological factors in breast cancer survivors. *Psychology Research and Behavior Management*, *11*, 371–381.

Peddle-Mcintyre, C. J., Cavalheri, V., Boyle, T., Mcveigh, J. A., Jeffery, E., Lynch, B. M., & Vallance, J. K. (2018). A Review of Accelerometer-based Activity Monitoring in Cancer Survivorship Research. *Medicine and Science in Sports and Exercise*, *50*(9), 1790–1801.

Pedersen, B. K. & Saltin, B. (2015). Exercise as medicine - evidence for prescribing exercise as therapy in. *Scandinavian Journal of Medicine & Science in Sports*, *25*, 1–72.

Pedisic, Z., Shrestha, N., Kovalchik, S., Stamatakis, E., Liangruenrom, N., Grgic, J., ... Oja, P. (2020). Is running associated with a lower risk of all-cause, cardiovascular and cancer mortality, and is the more the better? a systematic review and meta-analysis. *British Journal of Sports Medicine*, *54*(15), 898–905.

Pelizzo, G., Guddo, A., Puglisi, A., De Silvestri, A., Comparato, C., Valenza, M., ... Calcaterra, V. (2018). Accuracy of a Wrist-Worn Heart Rate Sensing Device during Elective Pediatric Surgical Procedures. *Children*, *5*(3), 38.

Peltola, M. A. (2012). Role of Editing of r–r Intervals in the Analysis of Heart Rate Variability. *Frontiers in Physiology*, *3*.

Perrotta, A. S., Jeklin, A. T., Hives, B. A., Meanwell, L. E., & Warburton, D. E. (2017). Validity of the Elite HRV Smartphone Application for Examining Heart Rate Variability in a Field-Based Setting: *Journal of Strength and Conditioning Research*, *31*(8), 2296–2302.

Peterson, L. L. & Ligibel, J. A. (2018). Physical Activity and Breast Cancer: An Opportunity to Improve Outcomes. *Current Oncology Reports*, *20*(7), 50.

Phatak, S. S., Freigoun, M. T., Martín, C. A., Rivera, D. E., Korinek, E. V., Adams, M. A., ... Hekler, E. B. (2018). Modeling individual differences: A case study of the application of system identification for personalizing a physical activity intervention. *Journal of Biomedical Informatics*, *79*, 82–97.

Plews, D. J., Laursen, P. B., Kilding, A. E., & Buchheit, M. (2012). Heart rate variability in elite triathletes, is variation in variability the key to effective training? a case comparison. *European Journal of Applied Physiology*, *112*(11), 3729–3741.

Plews, D. J., Laursen, P. B., Meur, Y. L., Hausswirth, C., Kilding, A. E., & Buchheit, M. (2014). Monitoring Training With Heart-Rate Variability: How Much Compliance Is Needed for Valid Assessment? *International Journal of Sports Physiology and Performance*, *9*(5), 783–790.

Plews, D. J., Laursen, P. B., Stanley, J., Kilding, A. E., & Buchheit, M. (2013). Training Adaptation and Heart Rate Variability in Elite Endurance Athletes: Opening the Door to Effective Monitoring. *Sports Medicine*, *43*(9), 773–781.

Plews, D. J., Scott, B., Altini, M., Wood, M., Kilding, A. E., & Laursen, P. B. (2017). Comparison of Heart-Rate-Variability Recording With Smartphone Photoplethysmography, Polar H7 Chest Strap, and Electrocardiography. *International Journal of Sports Physiology and Performance*, *12*(10), 1324–1328.

Pollán, M., Casla-Barrio, S., Alfaro, J., Esteban, C., Segui-Palmer, M. A., Lucia, A., & Martín, M. (2020). Exercise and cancer: A position statement from the Spanish Society of Medical Oncology. *Clinical & Translational Oncology: Official Publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico*, *22*(10), 1710–1729.

Popkin, B. M., Adair, L. S., & Ng, S. W. (2012). Global nutrition transition and the pandemic of obesity in developing countries. *Nutrition Reviews*, *70*(1), 3–21.

Postigo-Martin, P., Gil-Gutiérrez, R., Moreno-Gutiérrez, S., Lopez-Garzon, M., González-Santos, Á., Arroyo-Morales, M., & Cantarero-Villanueva, I. (2022). *mHealth system (ATOPE+) to support exercise prescription in breast cancer survivors: A validity and reliability, cross-sectional observational study (ATOPE study) (Preprint)*. Journal of Medical Systems.

Postigo-Martin, P., Peñafiel-Burkhardt, R., Gallart-Aragón, T., Alcaide-Lucena, M., Artacho-Cordón, F., Galiano-Castillo, N., ... Cantarero-Villanueva, I. (2021). Attenuating Treatment-Related Cardiotoxicity in Women Recently Diagnosed with Breast Cancer via a Tailored Therapeutic Exercise Program: Protocol of the ATOPE Trial. *Physical Therapy*, (pzab014).

Rabbani, A., Clemente, F. M., Kargarfard, M., & Chamari, K. (2019). Match Fatigue Time-Course Assessment Over Four Days: Usefulness of the Hooper Index and Heart Rate Variability in Professional Soccer Players. *Frontiers in Physiology*, *10*. Publisher: Frontiers.

Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G.-Z. (2017). Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics*, *21*(1), 4–21.

Redenius, N., Kim, Y., & Byun, W. (2019). Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. *BMC medical research methodology*, *19*(1), 29.

Reid, R. E. R., Insogna, J. A., Carver, T. E., Comptour, A. M., Bewski, N. A., Sciortino, C., & Andersen, R. E. (2017). Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. *Journal of Science and Medicine in Sport*, *20*(6), 578–582.

Rezende, L. F. M. d., Sá, T. H. d., Markozannes, G., Rey-López, J. P., Lee, I.-M., Tsilidis, K. K., . . . Eluf-Neto, J. (2018). Physical activity and cancer: An umbrella review of the literature including 22 major anatomical sites and 770 000 cancer cases. *British Journal of Sports Medicine*, *52*(13), 826–833.

Rising, C. J., Gaysynsky, A., Blake, K. D., Jensen, R. E., & Oh, A. (2021). Willingness to Share Data From Wearable Health and Activity Trackers: Analysis of the 2019 Health Information National Trends Survey Data. *JMIR mHealth and uHealth*, *9*(12), e29190.

Roberts, A. L., Fisher, A., Smith, L., Heinrich, M., & Potts, H. W. W. (2017). Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: A systematic review and meta-analysis. *Journal of Cancer Survivorship*, *11*(6), 704–719.

Roldán-Jiménez, C., Bennett, P., & Cuesta-Vargas, A. I. (2015). Muscular Activity and Fatigue in Lower-Limb and Trunk Muscles during Different Sit-To-Stand Tests. *PloS One*, *10*(10), e0141675.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Sada, Y. H., Poursina, O., Zhou, H., Workeneh, B. T., Maddali, S. V., & Najafi, B. (2021). Harnessing digital health to objectively assess cancer-related fatigue: The impact of fatigue on mobility performance. *PLOS ONE*, *16*(2), e0246101.

Sage Bionetworks. (2019). Bridge — Platform Overview. Available online: https://developer.sagebridge.org/articles/overview.html (accessed on 7 January 2022).

Sajjadieh, A., Shahsavari, A., Safaei, A., Penzel, T., Schoebel, C., Fietze, I., . . . Kelishadi, R. (2020). The Association of Sleep Duration and Quality with Heart Rate Variability and Blood Pressure. *Tanaffos*, *19*(2), 135–143.

Sammito, S. & Böckelmann, I. (2016). Factors influencing heart rate variability. *International Cardiovascular Forum Journal*, *6*(0), 18–22. Number: 0.

Sauro, J. (2011). *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices*. Denver, CO: CreateSpace Independent Publishing Platform.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., . . . Lin, C.-T. (2017). A review of clustering techniques and developments. *Neurocomputing*, *267*, 664–681.

Schaffer, K., Panneerselvam, N., Loh, K. P., Herrmann, R., Kleckner, I. R., Dunne, R. F., . . . Fung, C. (2019). Systematic Review of Randomized Controlled Trials of Exercise Interventions Using Digital Activity Trackers in Patients With Cancer. *Journal of the National Comprehensive Cancer Network*, *17*(1), 57–63.

Schmid, D. & Leitzmann, M. F. (2014). Association between physical activity and mortality among breast cancer and colorectal cancer survivors: A systematic review and meta-analysis. *Annals of Oncology*, *25*(7), 1293–1311.

Scott, J. M., Nilsen, T. S., Gupta, D., & Jones, L. W. (2018). Exercise Therapy and Cardiovascular Toxicity in Cancer. *Circulation*, *137*(11), 1176–1191.

Shaffer, F. & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*, 258.

Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: An integrative review of the heart's anatomy and heart rate variability. *Frontiers in Psychology*, *5*, 1040.

Al-Shamsi, H. O., Alhazzani, W., Alhuraiji, A., Coomes, E. A., Chemaly, R. F., Almuhanna, M., . . . Xie, C. (2020). A Practical Approach to the Management of Cancer Patients During the Novel Coronavirus Disease 2019 (COVID-19) Pandemic: An International Collaborative Group. *The Oncologist*, *25*(6).

Simó, M., Navarro, X., Yuste, V. J., & Bruna, J. (2018). Autonomic nervous system and cancer. *Clinical Autonomic Research*, *28*(3), 301–314.

Siow, E., Tiropanis, T., & Hall, W. (2018). Analytics for the Internet of Things: A Survey. *ACM Computing Surveys*, *51*(4), 1–36.

Soriano-Maldonado, A., Ruiz, J. R., Álvarez-Gallardo, I. C., Segura-Jiménez, V., Santalla, A., & Munguía-Izquierdo, D. (2015). Validity and reliability of rating perceived exertion in women with fibromyalgia: Exertion-pain discrimination. *Journal of Sports Sciences*, *33*(14), 1515–1522.

Statista - The Statistics Portal. (2021a). Forecast of the number of smartphone users in Europe from 2010 to 2025.

Statista - The Statistics Portal. (2021b). Forecast of the number of smartphone users in the World from 2010 to 2025.

Statista - The Statistics Portal. (2021c). Number of connected wearable devices worldwide by region from 2015 to 2022.

Stout, N. L., Baima, J., Swisher, A. K., Winters-Stone, K. M., & Welsh, J. (2017). A Systematic Review of Exercise Systematic Reviews in the Cancer Literature (2005-2017). *PM&R*. Contemporary Issues in Cancer Rehabilitation, *9*(9, Supplement 2), S347–S384.

Straiton, N., Alharbi, M., Bauman, A., Neubeck, L., Gullick, J., Bhindi, R., & Gallagher, R. (2018). The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: A systematic review. *Maturitas*, *112*, 85–93.

Swan, M. (2009). Emerging Patient-Driven Health Care Models: An Examination of Health Social Networks, Consumer Personalized Medicine and Quantified Self-Tracking. *International Journal of Environmental Research and Public Health*, *6*(2), 492–525.

Thiebaud, R. S., Funk, M. D., Patton, J. C., Massey, B. L., Shay, T. E., Schmidt, M. G., & Giovannitti, N. (2018). Validity of wrist-worn consumer products to measure heart rate and energy expenditure. *DIGITAL HEALTH*, *4*, 2055207618770322.

Tomlinson, D., Diorio, C., Beyene, J., & Sung, L. (2014). Effect of exercise on cancer-related fatigue: A meta-analysis. *American Journal of Physical Medicine & Rehabilitation*, *93*(8), 675–686.

Torous, J., Onnela, J.-P., & Keshavan, M. (2017). New dimensions and new tools to realize the potential of RDoC: Digital phenotyping via smartphones and connected devices. *Translational psychiatry*, *7*(3), e1053.

Torous, J. [John], Kiang, M. V., Lorme, J., & Onnela, J.-P. (2016). New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. *JMIR Mental Health*, *3*(2), e5165.

Tran, Y., Wijesuriya, N., Tarvainen, M., Karjalainen, P., & Craig, A. (2009). The Relationship Between Spectral Changes in Heart Rate Variability and Fatigue. *Journal of Psychophysiology*, *23*(3), 143–151. Publisher: Hogrefe Publishing.

Uhm, K. E., Yoo, J. S., Chung, S. H., Lee, J. D., Lee, I., Kim, J. I., . . . Hwang, J. H. (2017). Effects of exercise intervention in breast cancer patients: Is mobile health (mHealth) with pedometer more effective than conventional program using brochure? *Breast Cancer Research and Treatment*, *161*(3), 443–452.

Ummels, D., Beekman, E., Theunissen, K., Braun, S., & Beurskens, A. J. (2018). Counting Steps in Activities of Daily Living in People With a Chronic Disease Using Nine Commercially Available Fitness Trackers: Cross-Sectional Validity Study. *JMIR mHealth and uHealth*, *6*(4), e70.

Vega, J., Li, M., Aguillera, K., Goel, N., Joshi, E., Khandekar, K., . . . Low, C. A. (2021). Reproducible Analysis Pipeline for Data Streams: Open-Source Software to Process Data Collected With Mobile Devices. *Frontiers in Digital Health*, *3*, 168.

Vesterinen, V., Nummela, A., Heikura, I., Laine, T., Hynynen, E., Botella, J., & Häkkinen, K. (2016). Individual Endurance Training Prescription with Heart Rate Variability. *Medicine & Science in Sports & Exercise*, *48*(7), 1347–1354.

Wallace, L. K., Slattery, K. M., & Coutts, A. J. (2009). The ecological validity and application of the session-RPE method for quantifying training loads in swimming. *Journal of Strength and Conditioning Research*, *23*(1), 33–38.

Wanner, M., Probst-Hensch, N., Kriemler, S., Meier, F., Autenrieth, C., & Martin, B. W. (2016). Validation of the long international physical activity questionnaire: Influence of age and language region. *Preventive Medicine Reports*, *3*, 250–256.

Wiig, K. M. (1997). Knowledge Management: An Introduction and Perspective. *Journal of Knowledge Management*, *1*(1), 6–14.

Wild, C., Weiderpass, E., & Stewart, B. (Eds.). (2020). *World Cancer Report: Cancer Research for Cancer Prevention*. Lyon, France: International Agency for Research on Cancer.

Winfree, K. N. & Dominick, G. (2018). Modeling Clinically Validated Physical Activity Assessments Using Commodity Hardware. *IEEE Journal of Biomedical and Health Informatics*, *22*(2), 335–345.

Wolpert, D. & Macready, W. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82.

Wong, C. K., Mentis, H. M., & Kuber, R. (2018). The bit doesn't fit: Evaluation of a commercial activity-tracker at slower walking speeds. *Gait & Posture*, *59*, 177–181.

World Health Organization. (2021). Cancer. Available online: https://www.who.int/news-room/fact-sheets/detail/cancer (accessed on 19 October 2021).

World Health Organization. (2015). Cancer: Carcinogenicity of the consumption of red meat and processed meat. Available online: https://www.who.int/news-room/q-a-detail/cancer-carcinogenicity-of-the-consumption-of-red-meat-and-processed-meat (accessed on 23 September 2021).

World Health Organization. (2018). *Classification of digital health interventions*. Geneva: World Health Organization.

World Health Organization. (2020). *WHO report on cancer: Setting priorities, investing wisely and providing care for all*. Geneva: World Health Organization.

Wright, S. P., Collier, S. R., Brown, T. S., & Sandberg, K. (2017). An analysis of how consumer physical activity monitors are used in biomedical research. *The FASEB Journal*, *31*(1_supplement), 1020.24–1020.24.

Xie, J., Wen, D., Liang, L., Jia, Y., Gao, L., & Lei, J. (2018). Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. *JMIR mHealth and uHealth*, *6*(4), e94.

Yu, J., Ouyang, W., Chua, M. L. K., & Xie, C. (2020). SARS-CoV-2 Transmission in Patients With Cancer at a Tertiary Care Hospital in Wuhan, China. *JAMA oncology*, *6*(7), 1108–1110.

Zhao, Z., Arya, A., Orji, R., & Chan, G. (2020). Physical Activity Recommendation for Exergame Player Modeling using Machine Learning Approach. In *Proceedings of the IEEE 8th International Conference on Serious Games and Applications for Health* (p. 10). Vancouver, Canada.

Ziegler, R. G., Hoover, R. N., Pike, M. C., Hildesheim, A., Nomura, A. M. Y., West, D. W., . . . Hyer, M. B. (1993). Migration Patterns and Breast Cancer Risk in Asian-American Women. *JNCI: Journal of the National Cancer Institute*, *85*(22), 1819–1827.

List of Figures

List of Tables

Curriculum Vitae

Personal Information

Salvador Moreno Gutiérrez
ResearchGate | LinkedIn | salvadormoreno@ugr.es
Granada, Spain

Education

2017 - 2022	PhD studies at CASIP research group, Department of Computer Architecture and Technology, University of Granada, Spain
2016 - 2017	MSc in Data Science and Computer Engineering, Higher Technical School of Computer Science and Telecommunication Engineering, University of Granada, Spain.
2012 - 2016	BD in Industrial Electronic Engineering, Faculty of Sciences, University of Granada, Spain.

Research/Work Experience

2017 - Today	Research Assistant at CASIP research group, Department of Computer Architecture and Technology, University of Granada, Spain
2017 - Today	National Research Grant Holder, University Faculty Training (FPU), grant number FPU16/04201.
2019 - 2019	Research Stay at Biomedical Signals and Systems (BSS) research group, University of Twente, Enschede, The Netherlands.
2016 - 2017	National Collaboration Grant at CASIP research group, Department of Computer Architecture and Technology, University of Granada, Spain
2015 - 2016	University of Granada Starting Grant at CASIP research group, Department of Computer Architecture and Technology, University of Granada, Spain

Publications List

International Journals (SCI-indexed)

Postigo-Martín, P., Gil-Gutiérrez, R., Moreno-Gutiérrez, S., López-Garzón, M., González-Santos, A., Arroyo-Morales, M., & Cantarero-Villanueva, I. (2022). mHealth system (ATOPE+) to support exercise prescription in breast cancer survivors: A validity and reliability, cross-sectional observational study (ATOPE study) (Preprint). **Journal of Medical Systems** [Under review]

Moreno-Gutiérrez, S., Banos, O., Damas, M., Pomares, H., Postigo-Martín, P., Cantarero-Villanueva, I., & Arroyo-Morales, M. (2021). Monitoring Internal Load with a Mobile App in Professional Soccer Players after the COVID-19 Lockdown (Readiness Soccer): Observational Study (Preprint). **JMIR mHealth and uHealth** [Under review]

Moreno-Gutierrez, S., Bailon, C., Konsolakis, K., Damas, M., Banos, O., Pomares, H., Cabrita, M. (2021) Extracting Interpretable Knowledge from Mobile and Wearable Data for the Support of Healthy Aging, **Journal of Ambient Intelligence and Humanized Computing** [Under review]

Moreno-Gutierrez, S., Postigo-Martin, P., Damas, M., Pomares, H., Banos, O., Arroyo-Morales, M., & Cantarero-Villanueva, I. (2021). ATOPE+: An mHealth System to Support Personalized Therapeutic Exercise Interventions in Patients With Cancer. **IEEE Access**, 9, 16878-16898.

Postigo-Martin, P., Peñafiel-Burkhardt, R., Gallart-Aragón, T., Alcaide-Lucena, M., Artacho-Cordón, F., Galiano-Castillo, N., Fernández-Lao, C., Martín-Martín, L., Lozano-Lozano, M., Ruíz-Vozmediano, J., Moreno-Gutiérrez, S., Illescas-Montes, R., Arroyo-Morales, M., & Cantarero-Villanueva, I. (2021). Attenuating Treatment-Related Cardiotoxicity in Women Recently Diagnosed with Breast Cancer via a Tailored Therapeutic Exercise Program: Protocol of the ATOPE Trial. **Physical Therapy**, pzab014.

International Conferences

Morales, J.C., Moreno-Gutierrez, S., Bailon, C., Pomares, H., Rojas, I., Herrera, L.J. (2019) Long and Short Term Prediction of Power Consumption using LSTM Networks. In: Proceedings of the 6th International Conference on Time Series and Forecasting (**ITISE 2019**), Granada, Spain, September 25–27.

Moreno, S., Ortega, J., Damas, M., Díaz, A., González, J., & Pomares, H. (2018). Prediction of Energy Consumption in a NSGA-II-based Evolutionary Algorithm. Proceedings of the Genetic and Evolutionary Computation Conference Companion (**GECCO '18**), 239-240.

Moreno, S., Damas, M., Rojas, I., Amezcua, V., Gutierrez-Pastor, P., & Banos, O. (2018). Ubiquitous Assessment of the Recovery of Cancer Patients Using Consumer-Level Activity Trackers. Proceedings of the 12th International Conference on Ubiquitous Computing and Ambient Intelligence (**UCAmI 2018**), 2, 1204.

Moreno, S., Damas, M., Pomares, H., Moral-Munoz, J. A., & Banos, O. (2016). First Approach to Automatic Performance Status Evaluation and Physical Activity Recognition in Cancer Patients. 2016 15th International Conference on Ubiquitous Computing and Communications and 2016 8th International Symposium on Cyberspace and Security (**IUCC-CSS**), 116-123.

Moreno, S., Damas, M., Pomares, H., Moral-Muñoz, J. A., & Banos, O. (2016). Designing a Gamified System to Support Cancer Patients' Recovery. Proceedings -Extended Abstracts- of the 4th International Work-Conference on Bioinformatics and Biomedical Engineering (**IWBBIO**), 433-436.

Patents & Intellectual Property

Moreno Gutiérrez, S., Cantarero Villanueva, I., Postigo Martín, E. P., Arroyo Morales, M., Martín Martín, L., López Garzón, M. de la C., Lozano Lozano, M., González Santos, Á., Galiano Castillo, N., Fernández Lao, C., Damas Hermoso, M., Pomares Cintas, H., & Baños Legrán, O. (2020). ATOPE+ (**Safe Creative, 2010285737407**).

Datasets

Moreno-Gutiérrez, S., Postigo-Martín, P., Damas, M., Banos, O., Pomares, H., Arroyo-Morales, M., & Cantarero-Villanueva, I. (2022). ATOPE+Breast, Continuous Monitoring of Training Load in Patients with Breast Cancer during Therapeutic Exercise Intervention. **Zenodo, Dataset**.

A

Digital Health Systems Taxonomy

This appendix gathers the classification for health system challenges, digital health interventions and system categories defined for digital health interventions by the WHO/RHR/18.06 taxonomy (World Health Organization, 2018).

Health System Challenges

The *health system challenges* defined in the taxonomy (World Health Organization, 2018) are presented in the following.

1. **Information.** Lack of population denominator; delayed reporting of events; lack of quality/reliable data; communication roadblocks; lack of access to information or data; insufficient utilization of data and information; lack of unique identifier.

2. **Availability.** Insufficient supply of commodities; insufficient supply of services; insufficient supply of equipment; insufficient supply of qualified health workers.

3. **Quality.** Poor patient experience; Insufficient health worker competence; low quality health commodities; low health worker motivation; insufficient continuity of care; inadequate supportive supervision; poor adherence to guidelines.

4. **Acceptability.** Lack of alignment with local norms; programs which do not address individual beliefs and practices.

5. **Utilization.** Low demand for services; geographic inaccessibility; low adherence to treatments; loss to follow up.

6. **Efficiency.** Inadequate workflow management; lack of or inappropriate referrals; poor planning and coordination; delayed provision of care; inadequate access to transportation.

7. **Cost.** High cost of manual processes; lack of effective resource allocation; client-side expenses; lack of coordinated payer mechanism.

8. **Accountability**. Insufficient patient engagement; unaware of service entitlement; absence of community feedback mechanisms; lack of transparency in commodity transactions; poor accountability between levels of the health sectors; inadequate understanding or beneficiary populations.

System Categories

The *system categories* defined in the taxonomy (World Health Organization, 2018) –which adapted from the International Standards Organization (International Standards Organization (ISO), 2014)– are described in the following.

A. **Census, population, information and data warehouse.** Stores data regarding information of the population in a certain region.

B. **Civil registration and vital statistics.** Provides information to epidemiologists, statisticians, demographers, and others working in public health about vital statistics from the people in a region.

C. **Client applications.** Client program that targets an end-user (e.g., patient, healthcare provider) that consumes services provided by a server program.

D. **Client communication system.** Manages communication between healthcare providers and patients.

E. **Clinical terminology and classifications.** Utilized by consumers, healthcare providers, quality and utilization management personnel, researchers, and other administrative staff (accounting, billing, and coding personnel), to facilitate communication between healthcare providers and consumers in healthcare for data collection purposes.

F. **Community-based information system.** Involves data collection, management, and analysis of health services that exist within a community outside of health facilities.

G. **Data interchange interoperability and accessibility.** Enables automatic, inter-organizational computer-to-computer communication without the need for human interaction.

H. Electronic medical record. Allows to create, gather, manage, and consult the electronic health records of patients by authorized clinicians and staff within one health care organization.

I. Emergency response system. Connects an user (patient) with a response coordinator to assesses the situation and send urgent assistance if necessary.

J. Environmental monitoring system. Provides facility managers the ability to monitor their entire site to help prevent environmental disasters capable of causing costly damages.

K. Facility management information system. Supports workflow processes, providing facility managers with data to assist decision-making and to help measure management performance.

L. Geographic information system. Contains geographic data, combined with software tools for management, analysis and visualization.

M. Health finance and insurance information system. Ensure the raise of sufficient funds and providing financial risk protection.

N. Health management information system. Supports planning, management, and decision-making in health facilities and organizations, as well as tracking service quality.

O. Human resource information system. Enables the data entry, tracking, and information needs of the human resources department, payroll, management, and accounting functions of a business.

P. Identification registries and directories. Provide information about healthcare members and registrants, including data like registration and membership status, identification, city, state, province and country.

Q. Knowledge management system. Enables the process of knowledge management by storing and retrieving knowledge to improve understanding, collaboration, and process alignment within an organization or team.

R. Laboratory and diagnostics information system. Receives and stores requests for tests, as well as results from laboratory technicians or laboratory instruments.

S. Learning and training system. Enables the creation and management of educational courses, lessons, and training materials for employees to improve and develop in their career.

T. Logistics management information system. Stores logistics information and provides reports that aggregate, analyze, validate and display data to support logistic decision-making and management in the supply chain.

U. Pharmacy information system. Stores and retrieves information in a pharmacy, enabling services like prescription management, inventory management, or drug interaction monitoring.

V. Public health and disease surveillance system. Provides ongoing, systematic collection, analysis, and interpretation of health-related data critical to the planning, implementation, and evaluation of public health.

W. Research information system. Captures information about the current research activities of a unit (e.g., university, hospital), aggregating, curating, managing and utilizing information and metadata to summarize a view of research output and its impact.

X. Shared health record and health information repositories. Enable the collection and storage of electronic health records of individual patients in a centralized server, capable of being shared across different healthcare environments.

Y. Telemedicine. Allows the distribution of health-related services and information using ICT, enabling long-distance patient-clinician contact, care, advice, reminders, education, intervention, monitoring, and remote assistance.

Digital Health Intervention Types

The *digital health intervention* types defined in the taxonomy (World Health Organization, 2018), depend on the stakeholders involved and functionalities needed: clients, healthcare providers, health system managers, and data services. Clients and healthcare providers are described in the following.

Clients:

1. **Targeted client communication.** Transmit health event alerts to specific population groups; transmit targeted health information to client(s) based on health status or demographics; transmit targeted alerts and reminders to client(s); transmit diagnostics result or availability of result to client(s)

2. **Untargeted client communication.** Transmit untargeted health information to an undefined population; transmit untargeted health event alerts to undefined group.

3. **Client to client communication.** Peer group for clients.

4. **Personal health tracking.** Access by client to own medical records; self monitoring of health or diagnostic data by client; active data capture/documentation by client.

5. **Citizen based reporting.** Reporting of health system feedback by clients; reporting of public health events by clients.

6. **On-demand information services to clients.** Client look-up of health information.

7. **Client financial transactions.** Transmit or manage out of pocket payments by clients; transmit or manage vouchers to client(s) for health services; transmit or manage incentives to client(s) for health services.

Healthcare Providers:

1. **Client identification and registration.** Verify client unique identity; enrol client for health services/clinical care plan.

2. **Client health records.** Longitudinal tracking of clients' health status and services; manage client's structured clinical records; manage client's unstructured clinical records; routine health indicator data collection and management.

3. **Healthcare provider decision support.** Provide prompts and alerts based according to protocol; provide checklist according to protocol; screen clients by risk or other health status.

4. **Telemedicine.** Consultations between remote client and healthcare provider; remote monitoring of client health or diagnostic data by healthcare provider; transmission of medical data to healthcare provider; consultations for case management between healthcare providers.

5. **Healthcare provider communication.** Communication from healthcare providers to supervisor; communication and performance feedback to healthcare providers; transmit routine news and workflow notifications to healthcare providers; transmit non-routine health event alerts to healthcare providers; peer group for healthcare providers.

6. **Referral coordination.** Coordinate emergency response and transport; manage referrals between points of service within health sector; manage referrals between health and other sectors.
7. **Health worker activity planning and scheduling.** Identify clients in need of services; schedule healthcare providers activities.
8. **Healthcare provider training.** Provide training content to healthcare providers; assess capacity of healthcare providers.
9. **Prescription and medication management.** Transmit or track prescription orders; track client's medication consumption; report adverse drug events.
10. **Laboratory and diagnostics imaging management.** Transmit diagnostic result to healthcare provider; transmit and track diagnostic orders; capture diagnostic results from digital devices; track biological specimens.

B Longitudinal Visualizations

Longitudinal visualizations for for clean HRV, Wellness and Fitbit data are shown in the following pictures for all participants. The features `lnrmssd`, the SWC of the lnrmssd as in Equation 4.1, `recovery`, `fatigue`, `distress`, `sleep_satisfaction`, `training_borg`, `session_type`, `steps_total_yesterday`, and `sleep_total` data of participants are shown in this section.

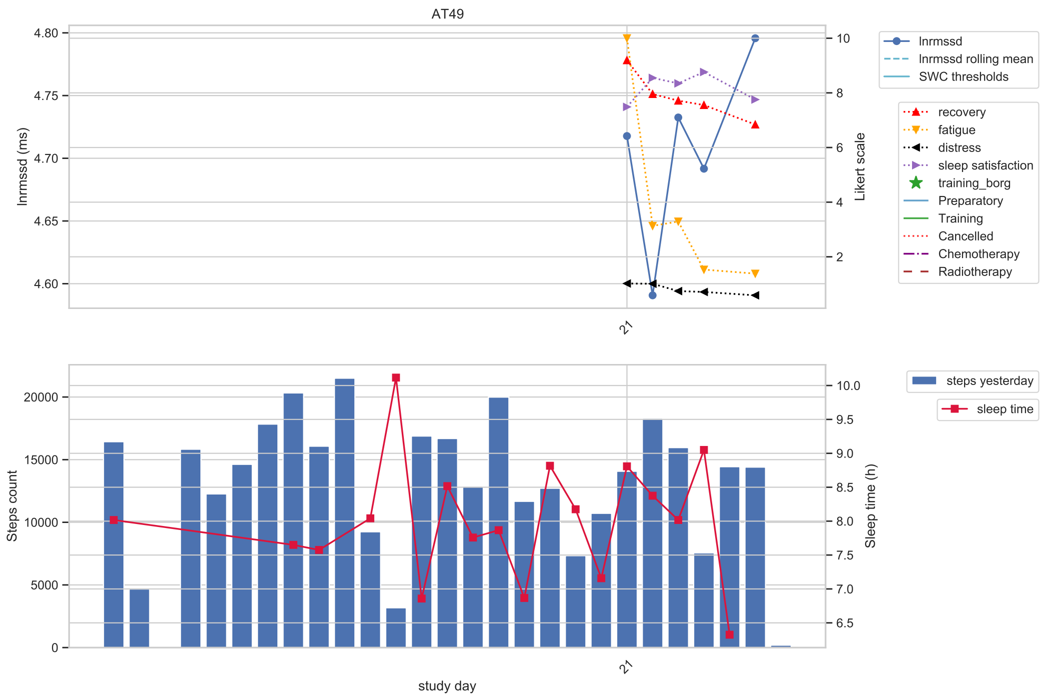


Fig. B.1.: AT49

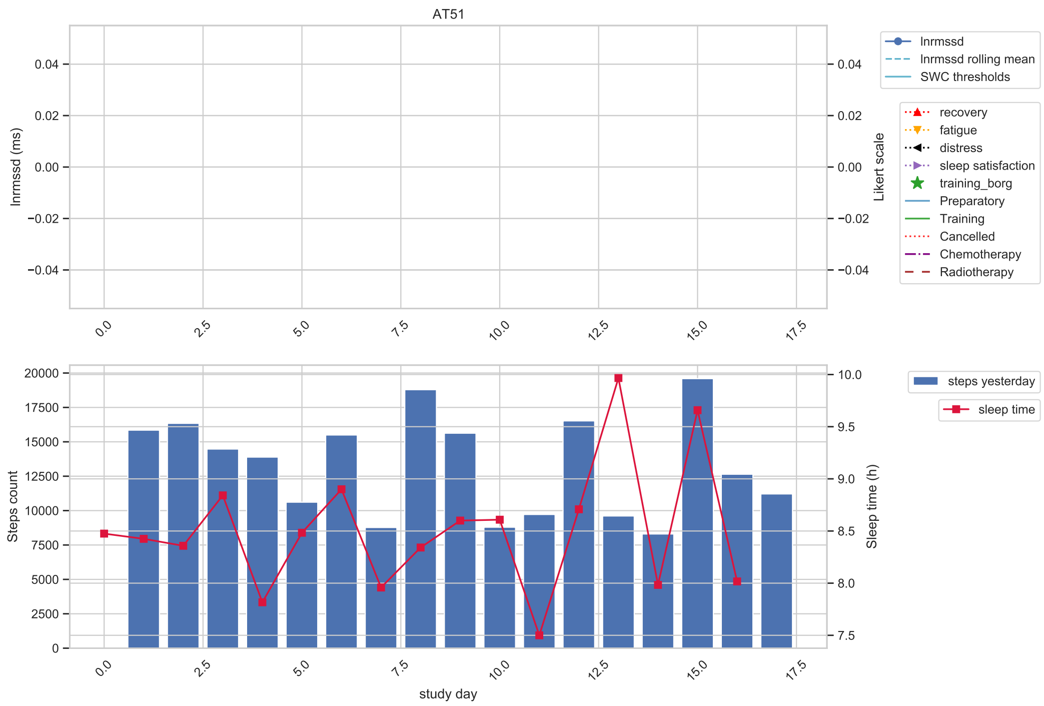


Fig. B.2.: AT51

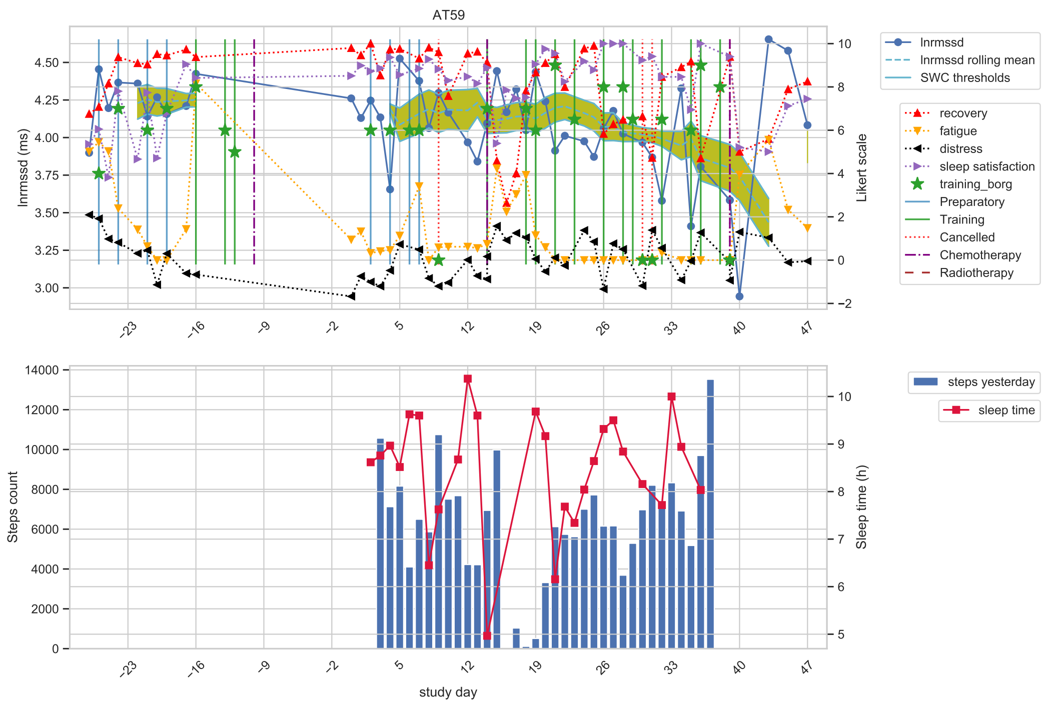


Fig. B.3.: AT59

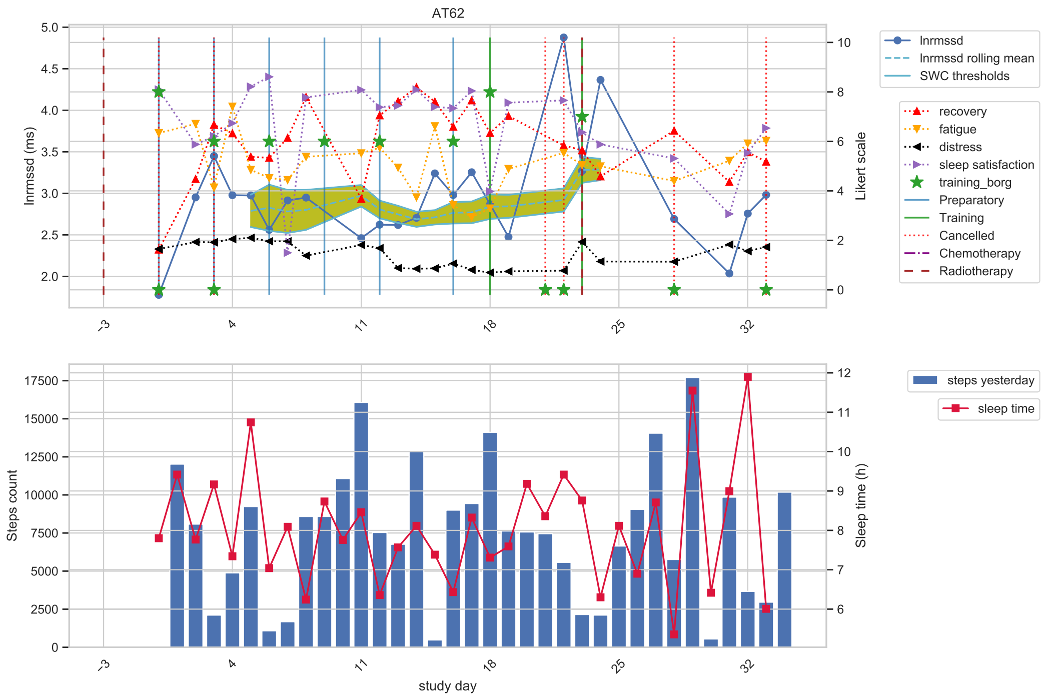


Fig. B.4.: AT62

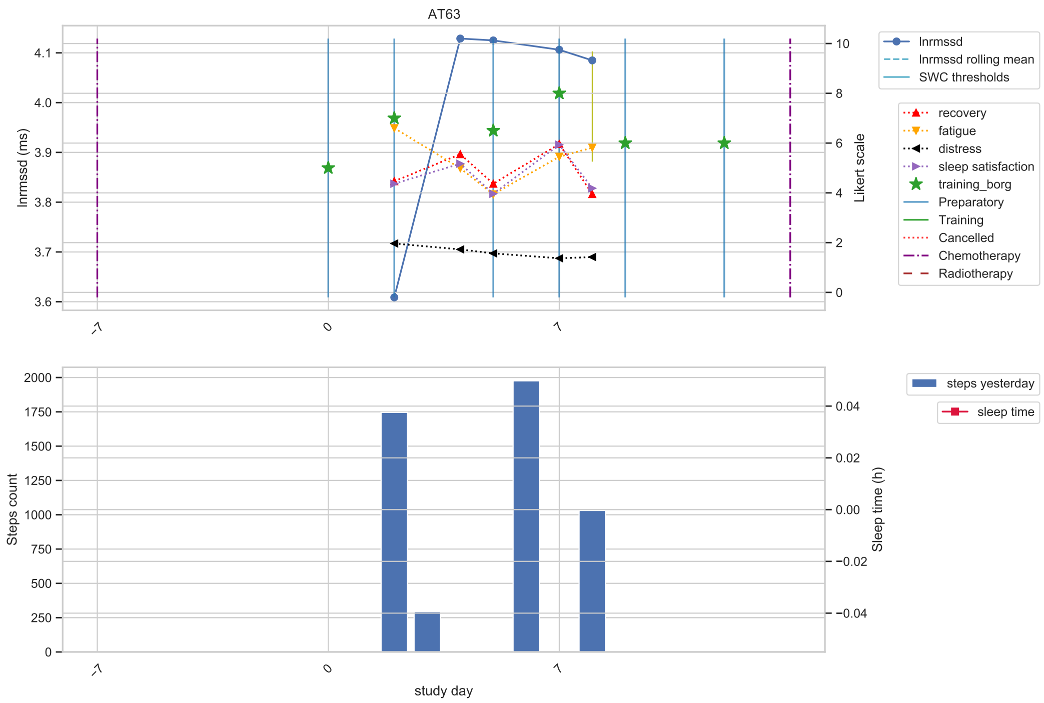


Fig. B.5.: AT63

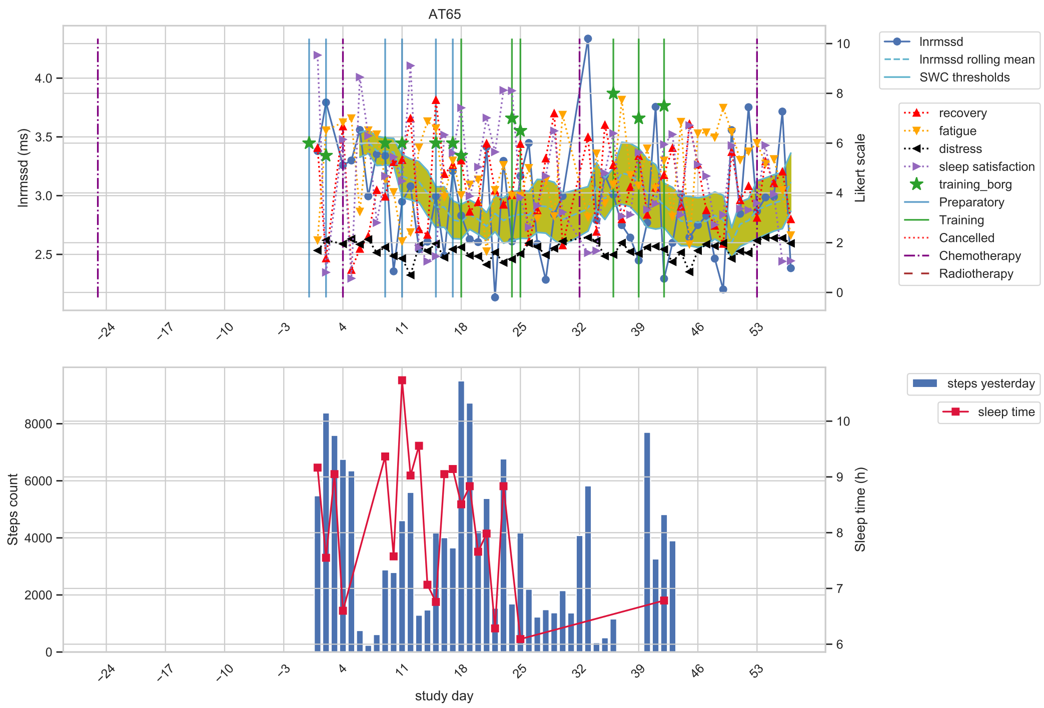


Fig. B.6.: AT65

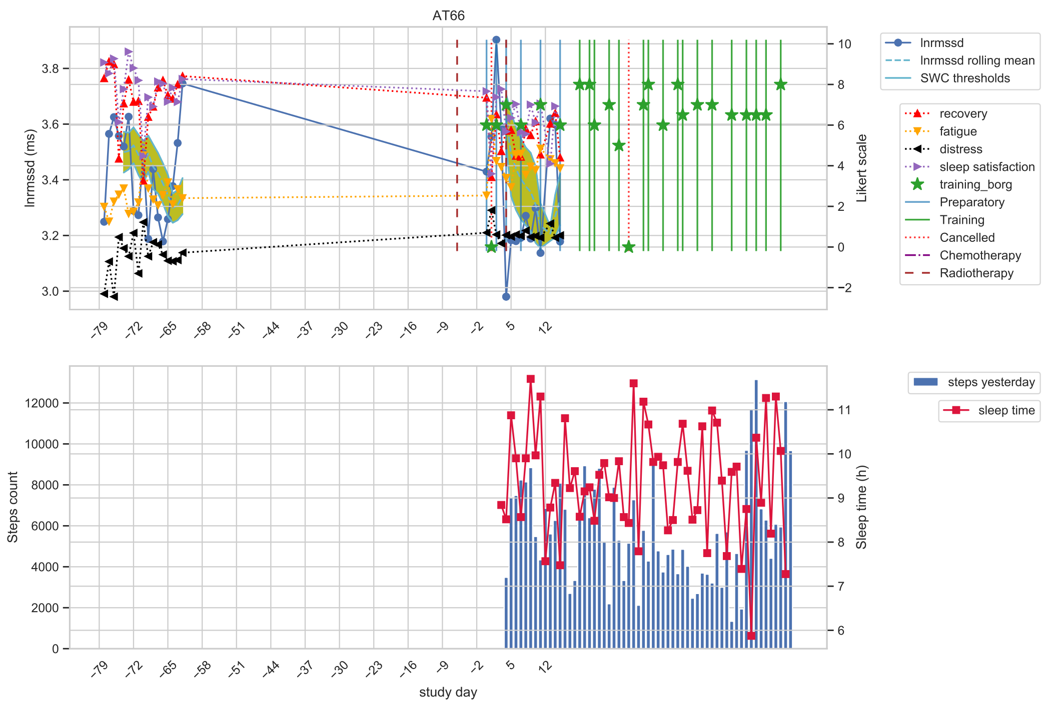


Fig. B.7.: AT66

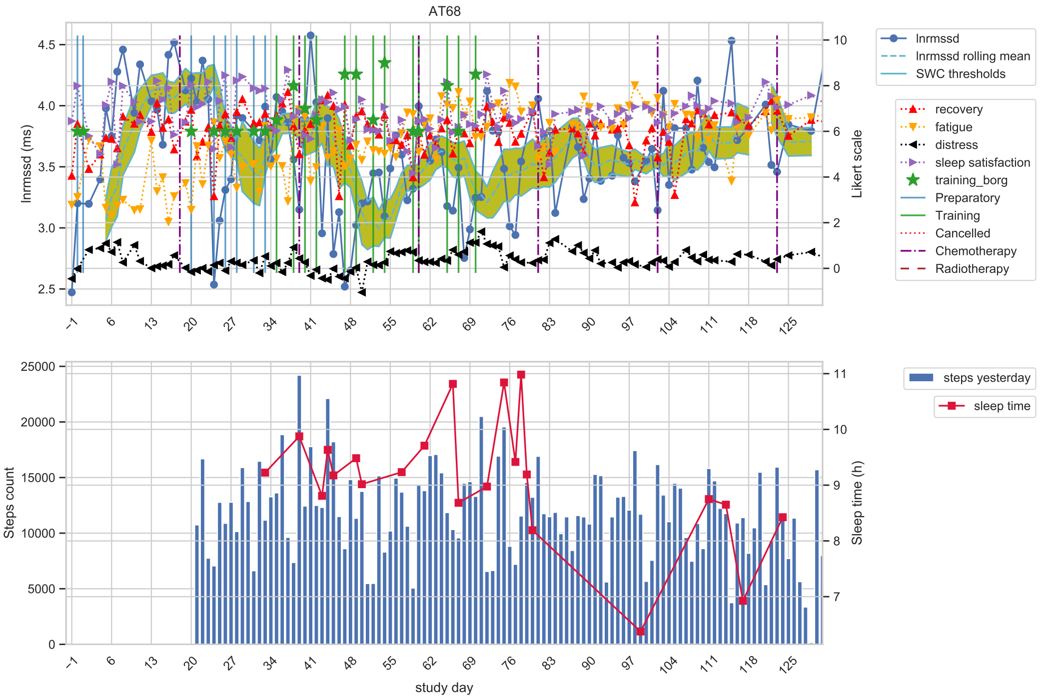


Fig. B.8.: AT68

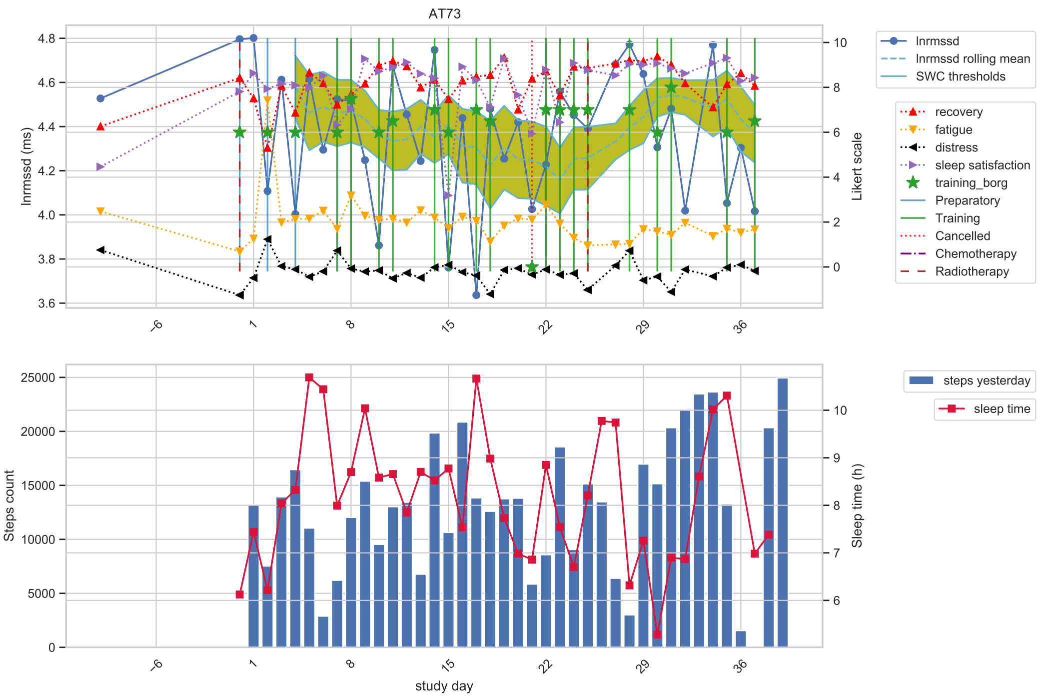


Fig. B.9.: AT73

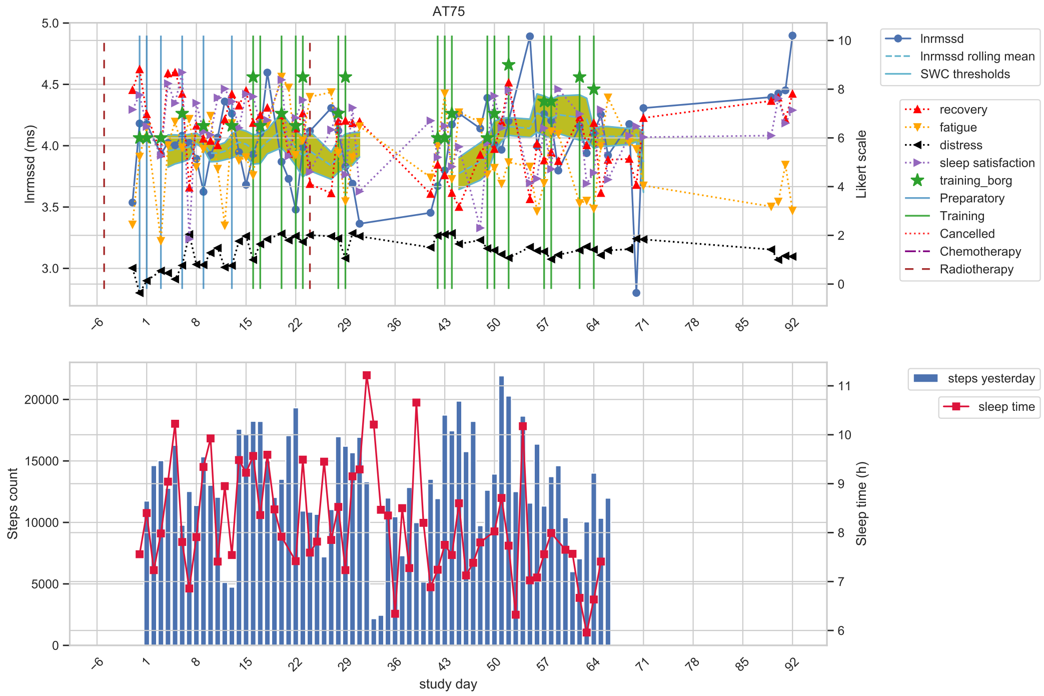


Fig. B.10.: AT75

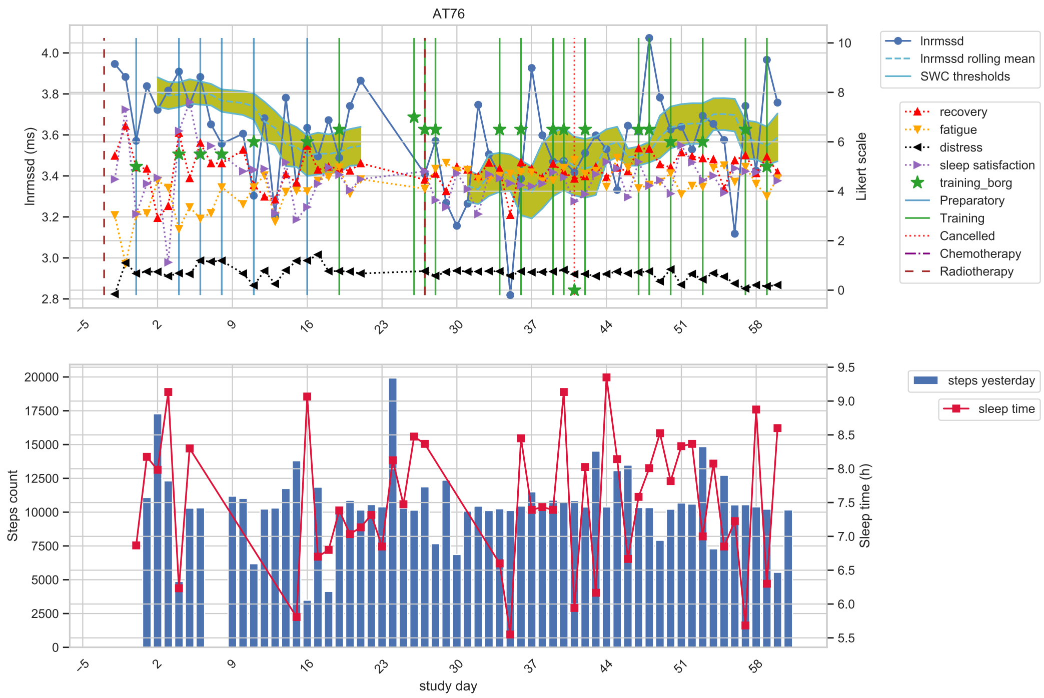


Fig. B.11.: AT76

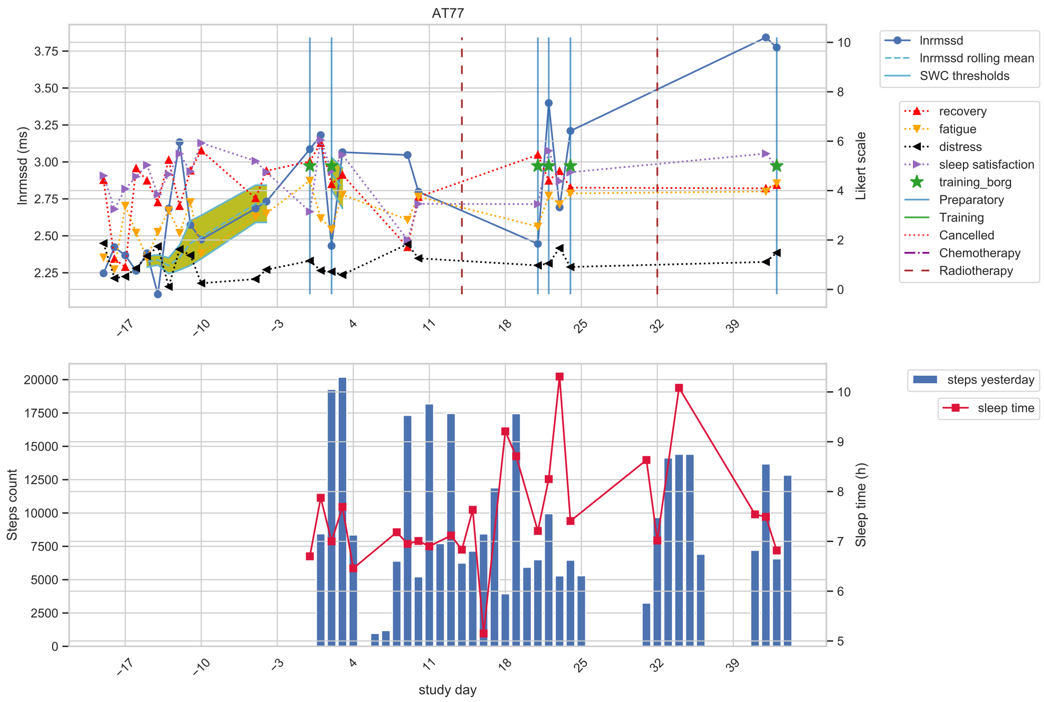


Fig. B.12.: AT77

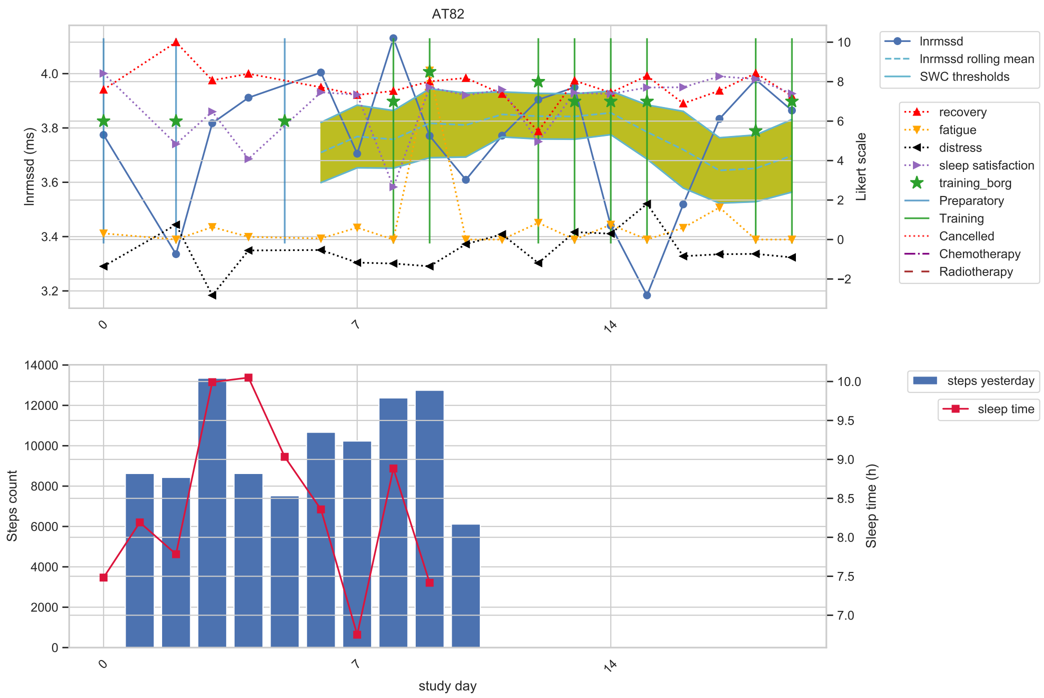


Fig. B.13.: AT82

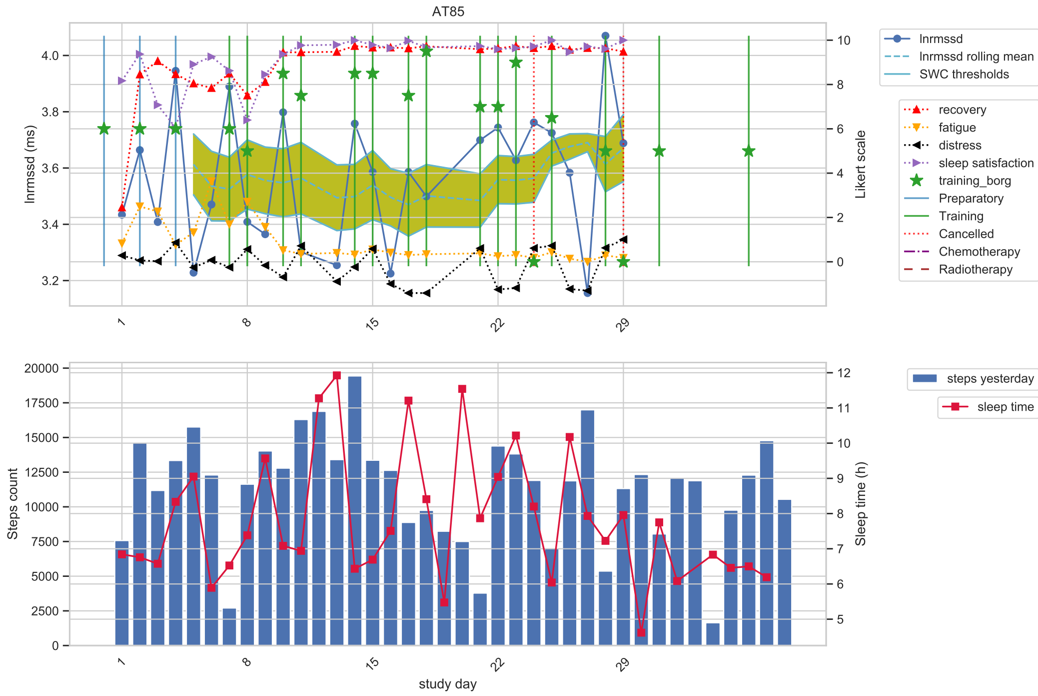


Fig. B.14.: AT85

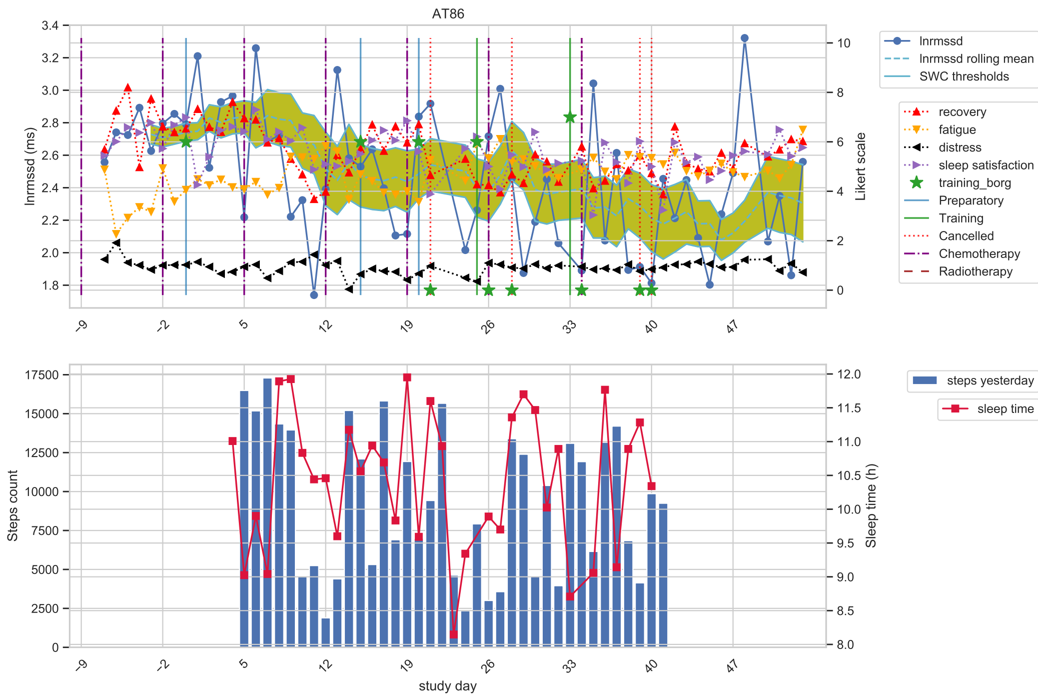


Fig. B.15.: AT86

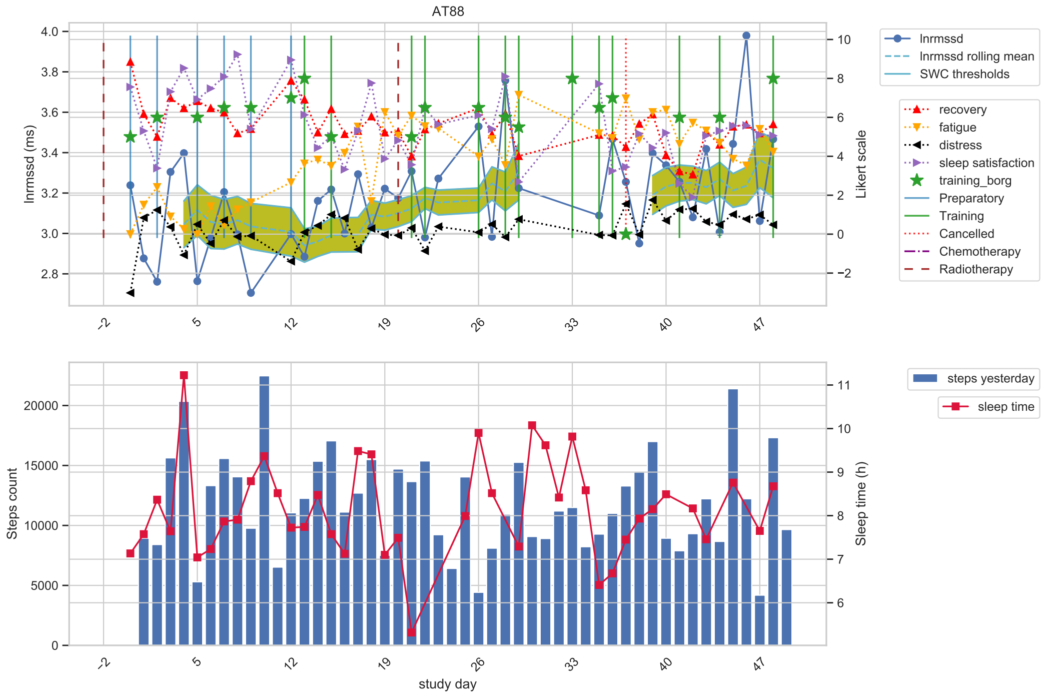


Fig. B.16.: AT88

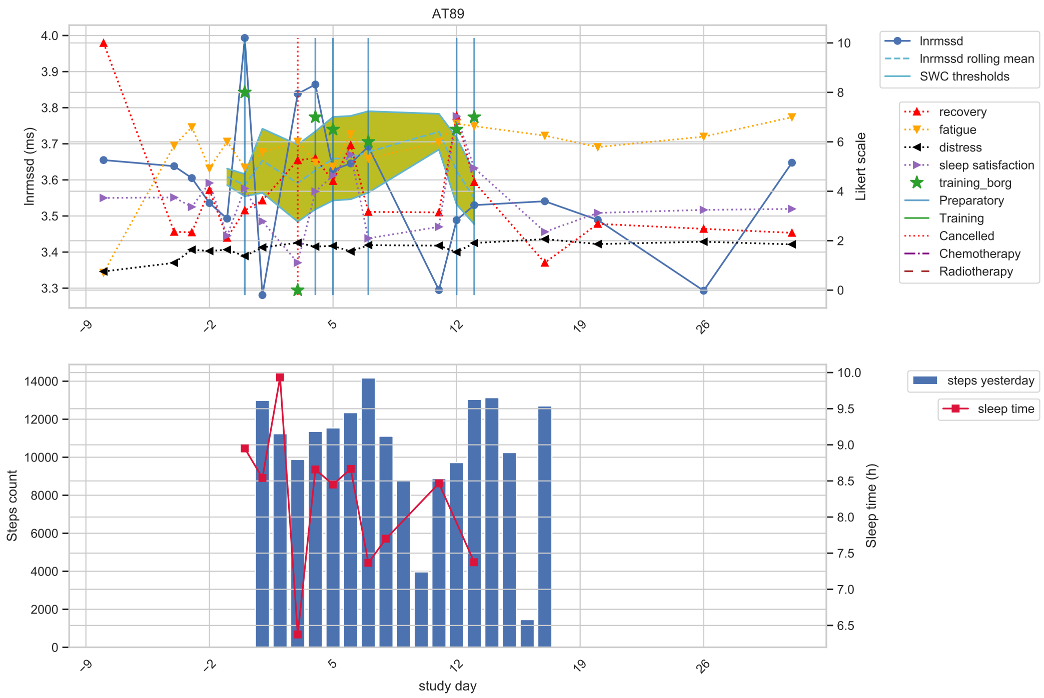


Fig. B.17.: AT89

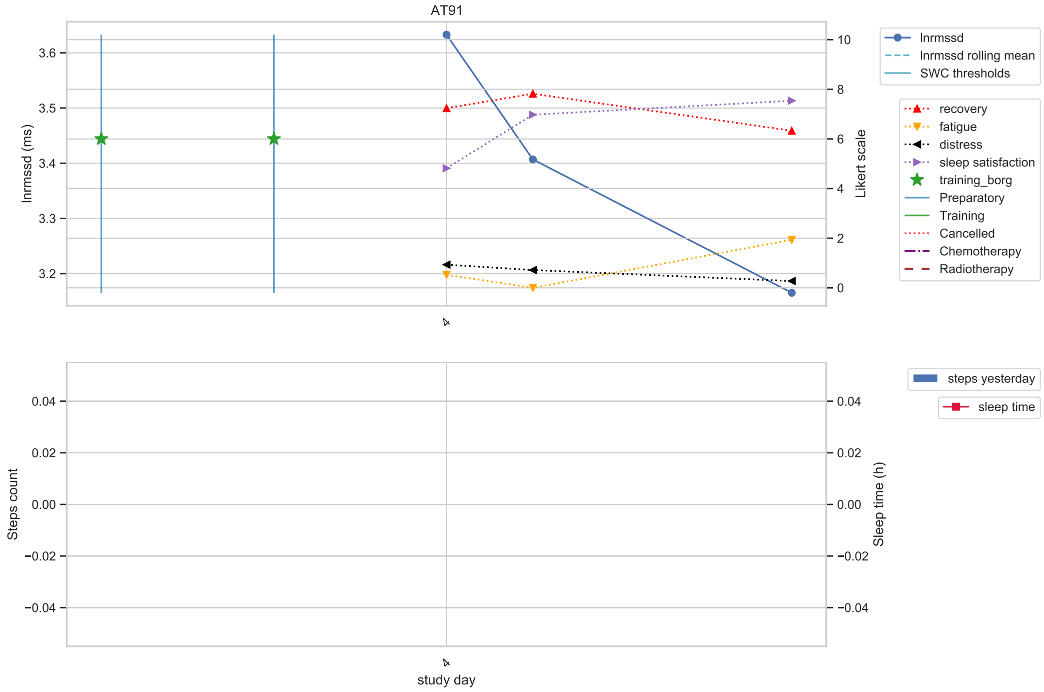


Fig. B.18.: AT91

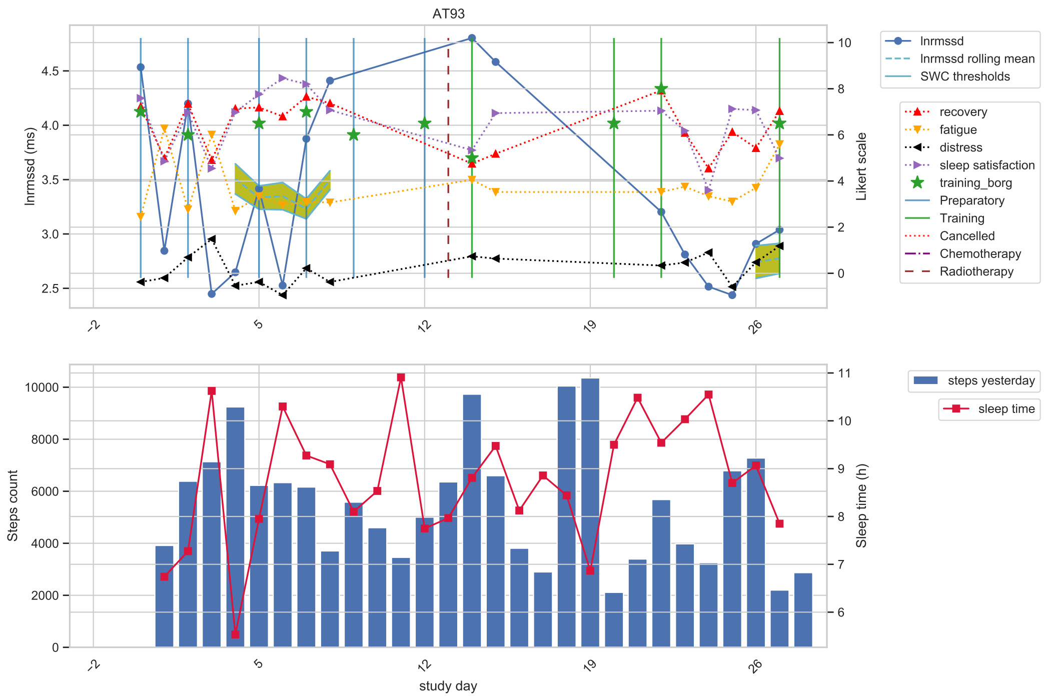


Fig. B.19.: AT93

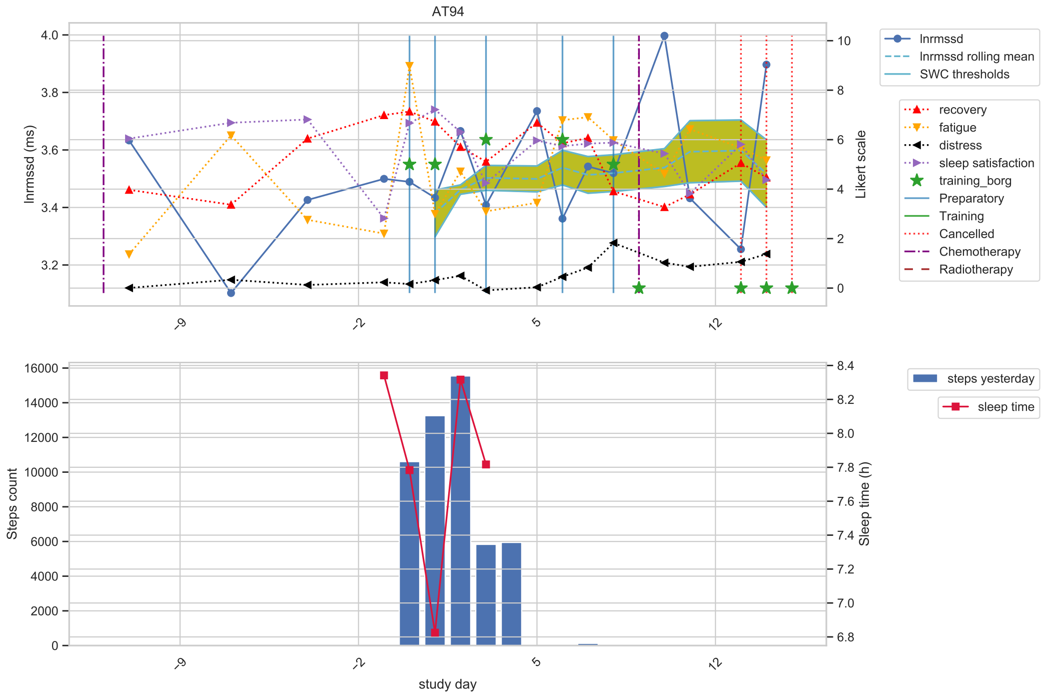


Fig. B.20.: AT94

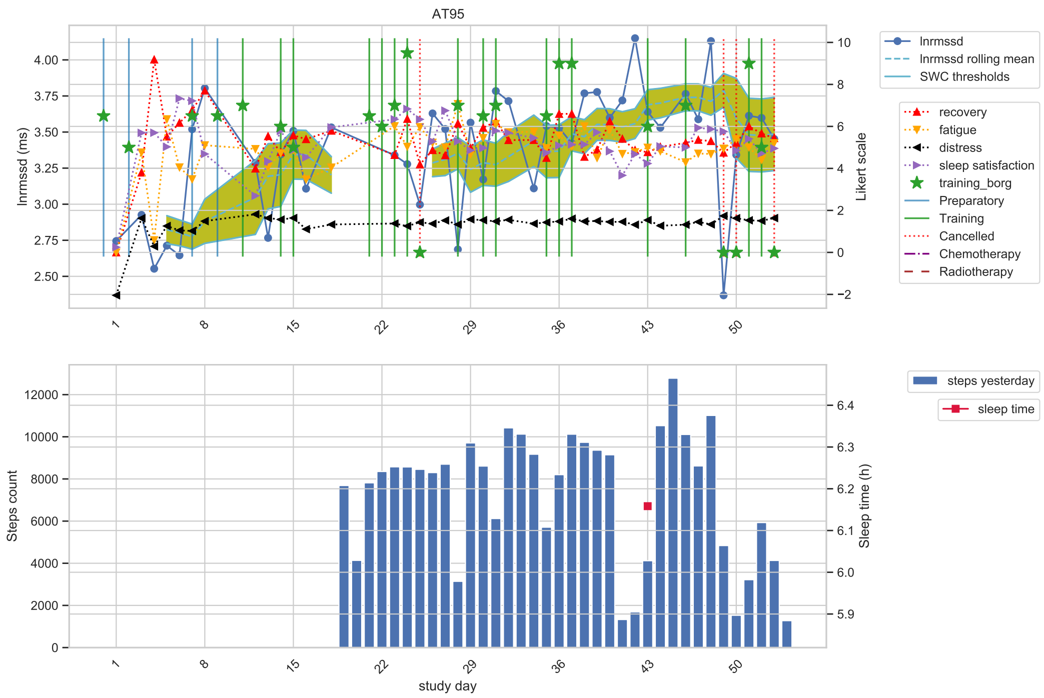


Fig. B.21.: AT95

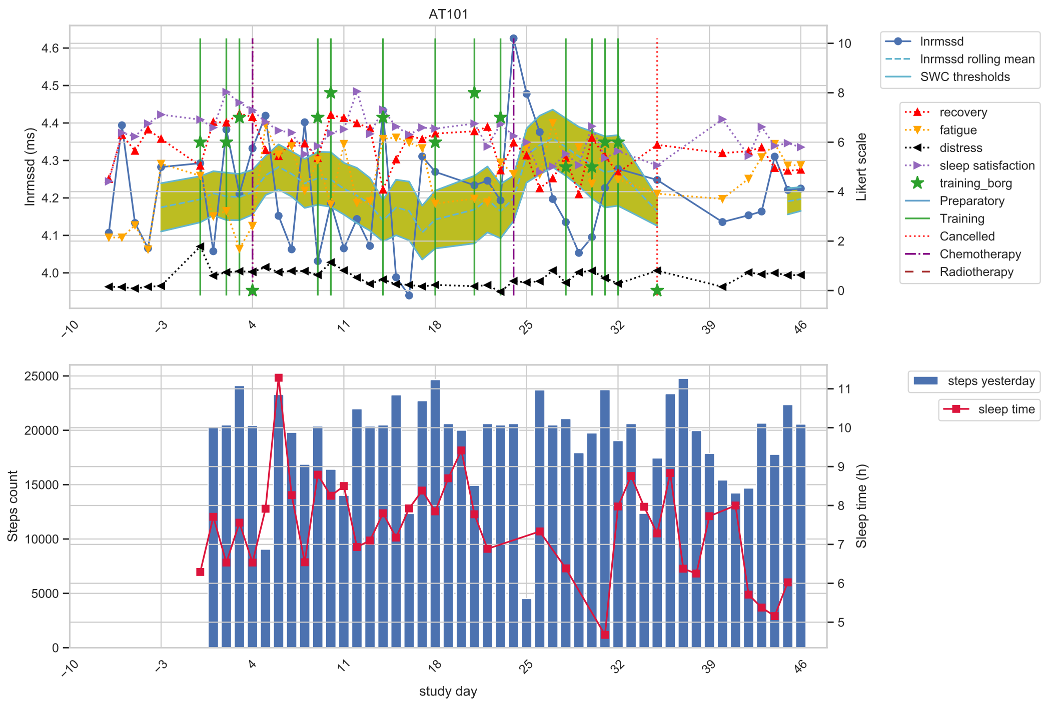


Fig. B.22.: AT101

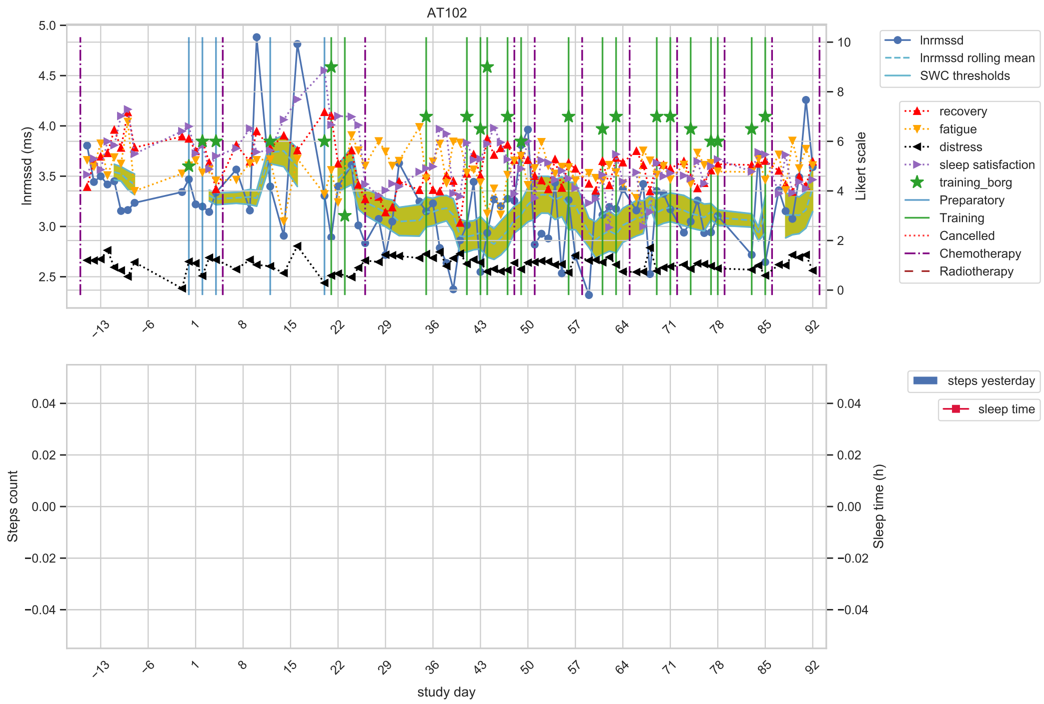


Fig. B.23.: AT102

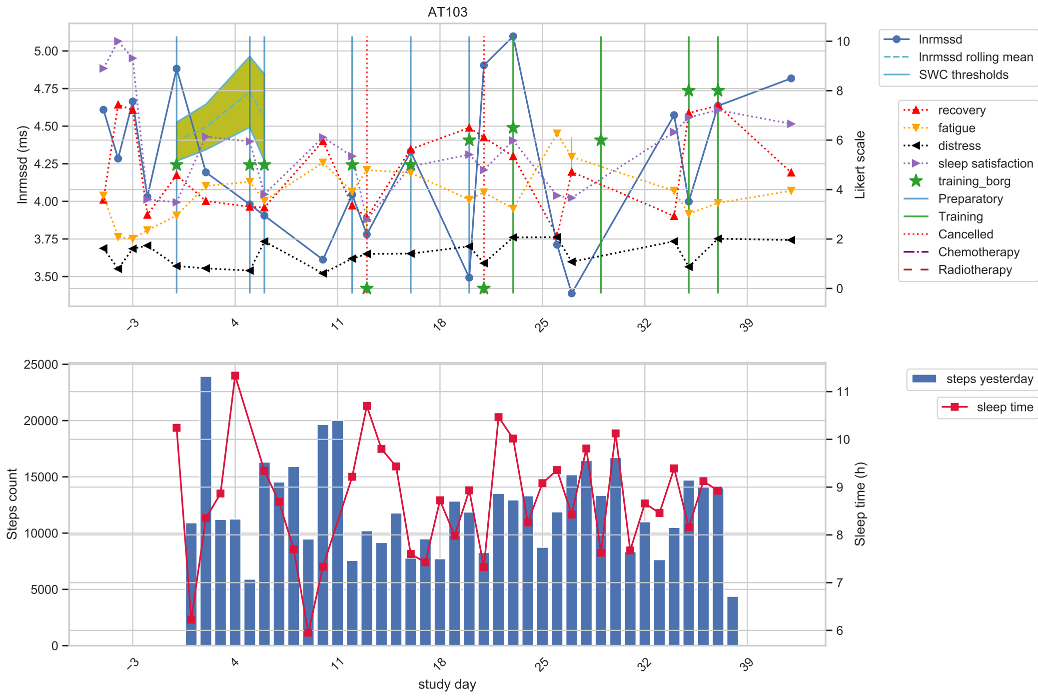


Fig. B.24.: AT103

C Clustering Experiments

Clustering experiments were labeled as *CXX-K*, being *XX* the number of the clustering experiment, and *K* the number of clusters set for training. K-Means run over scikit-sklearn (0.24.2) with the following configuration:

```
sklearn.cluster.KMeans(n_clusters=K, init='k-means++', n_init=10, max_iter=300,
tol=0.0001, verbose=0, random_state=42, copy_x=True, algorithm='auto')
```

Feature importance analysis were conducted analyzing *gini* importance out of Random Forests classifiers with its parameters optimized over a grid to reach the highest weighted f1-score:

```
sklearn.ensemble.RandomForestClassifier(n_estimators=100, criterion='gini',
max_depth=5, min_samples_leaf=5, min_samples_split=10, max_features='auto',
bootstrap=True, oob_score=True, random_state=42, class_weight='balanced')
```

C.1 C01. All Variables: HRV, Wellness, and Fitbit

Results for C01 with all features selected (HRV, Wellness, Fitbit) are displayed in Table C.1, Figure C.1, Figure C.2, Figure C.3 and Figure C.4.

Tab. C.1.: Silhouette scores for K-Means clustering experiment C01.

n_clusters	silhouette
2	0.113564
3	0.111387
4	0.105895
5	0.101964

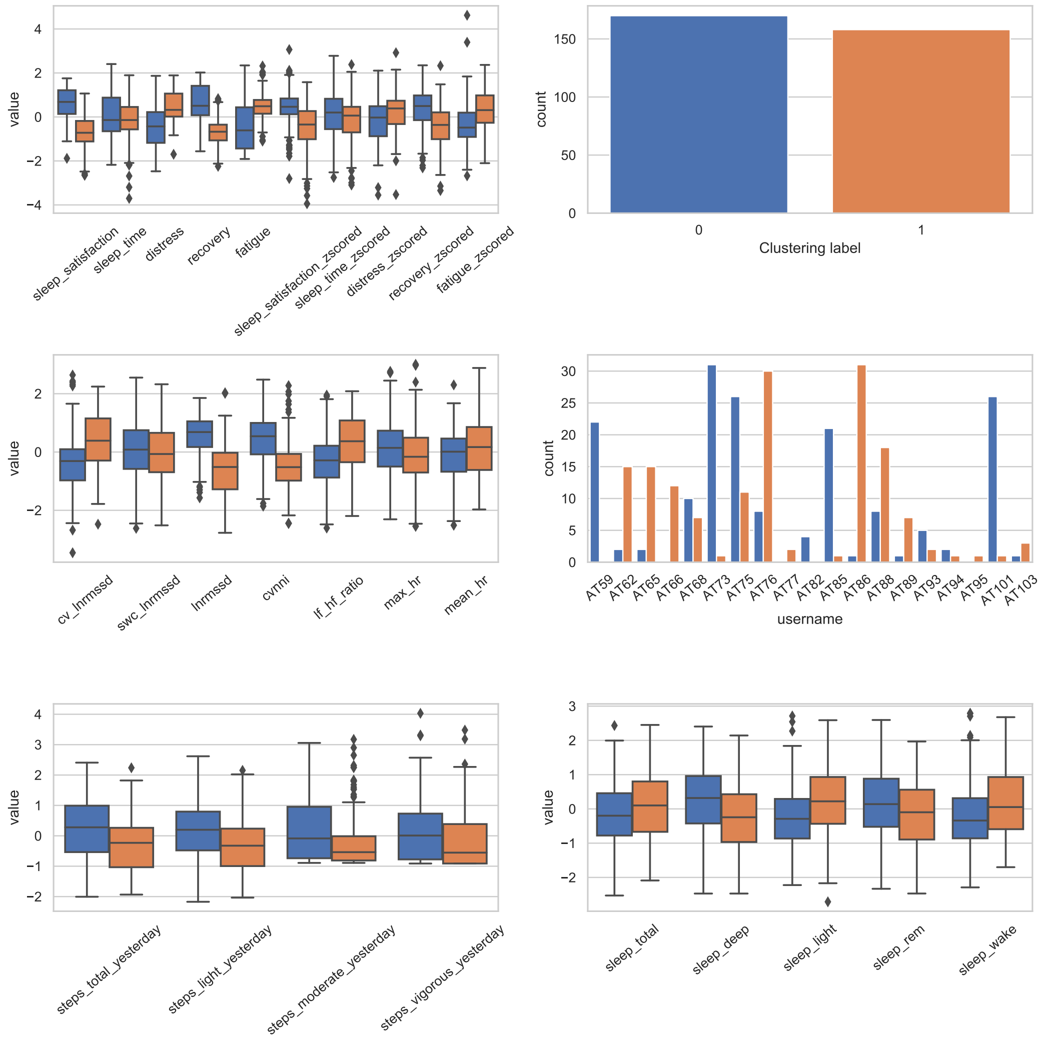


Fig. C.1.: K-Means clustering experiment C01-2.

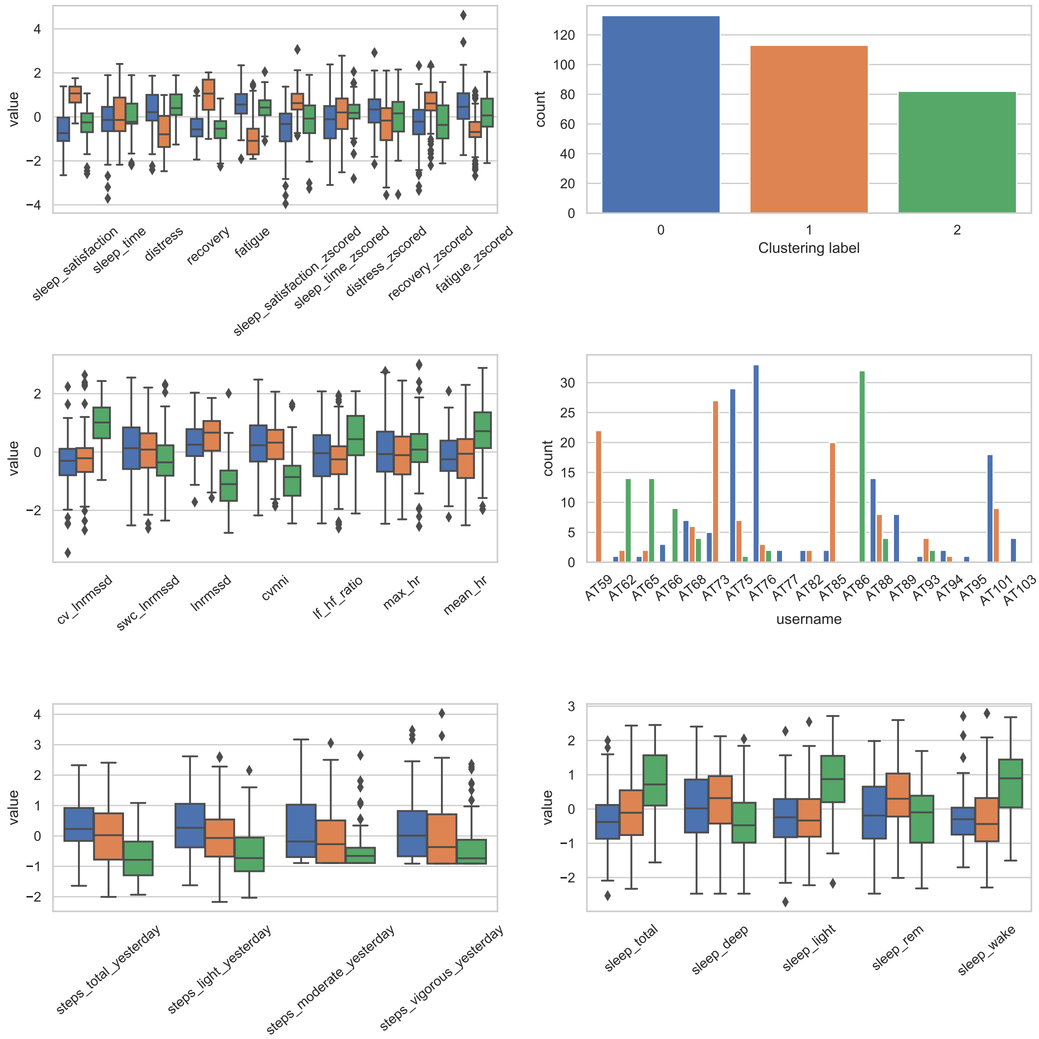


Fig. C.2.: K-Means clustering experiment C01-3.

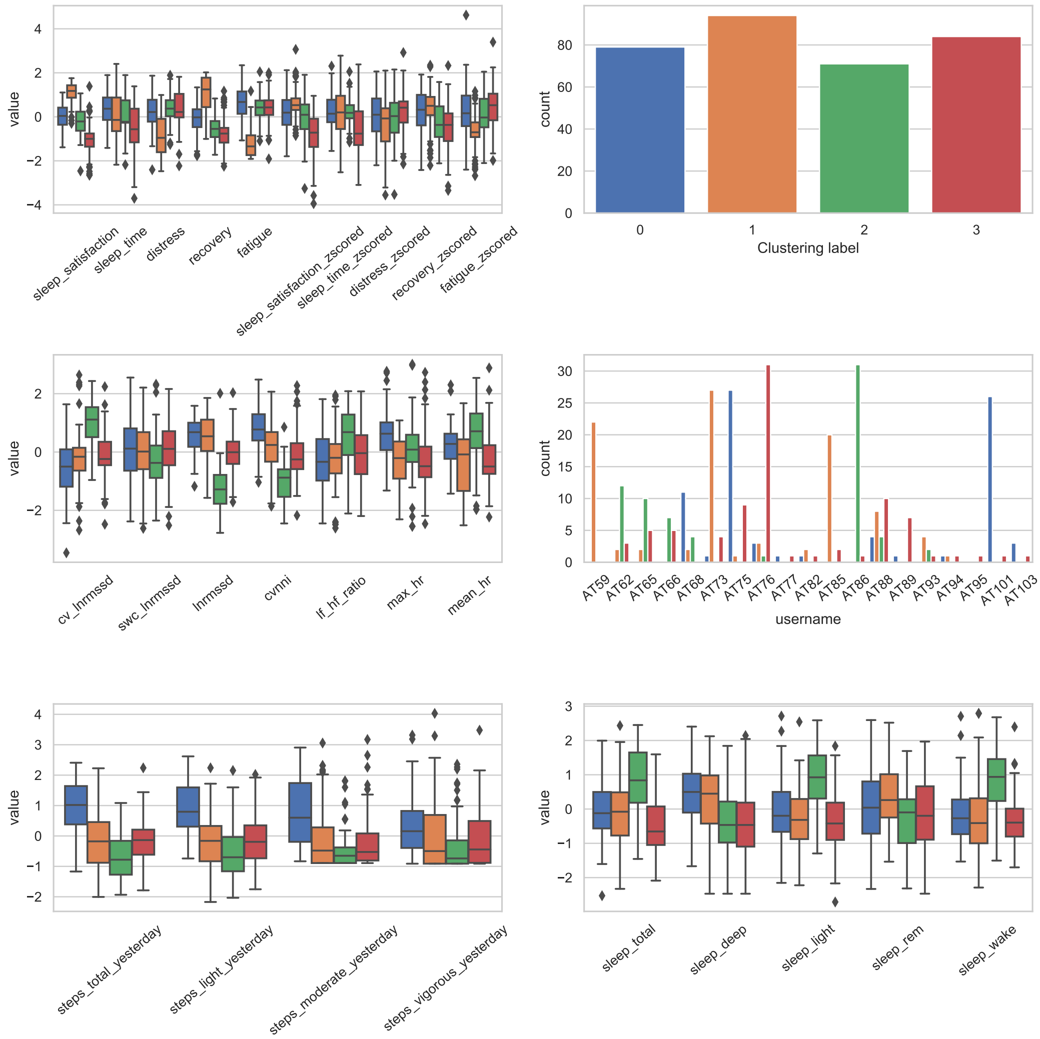


Fig. C.3.: K-Means clustering experiment C01-4.

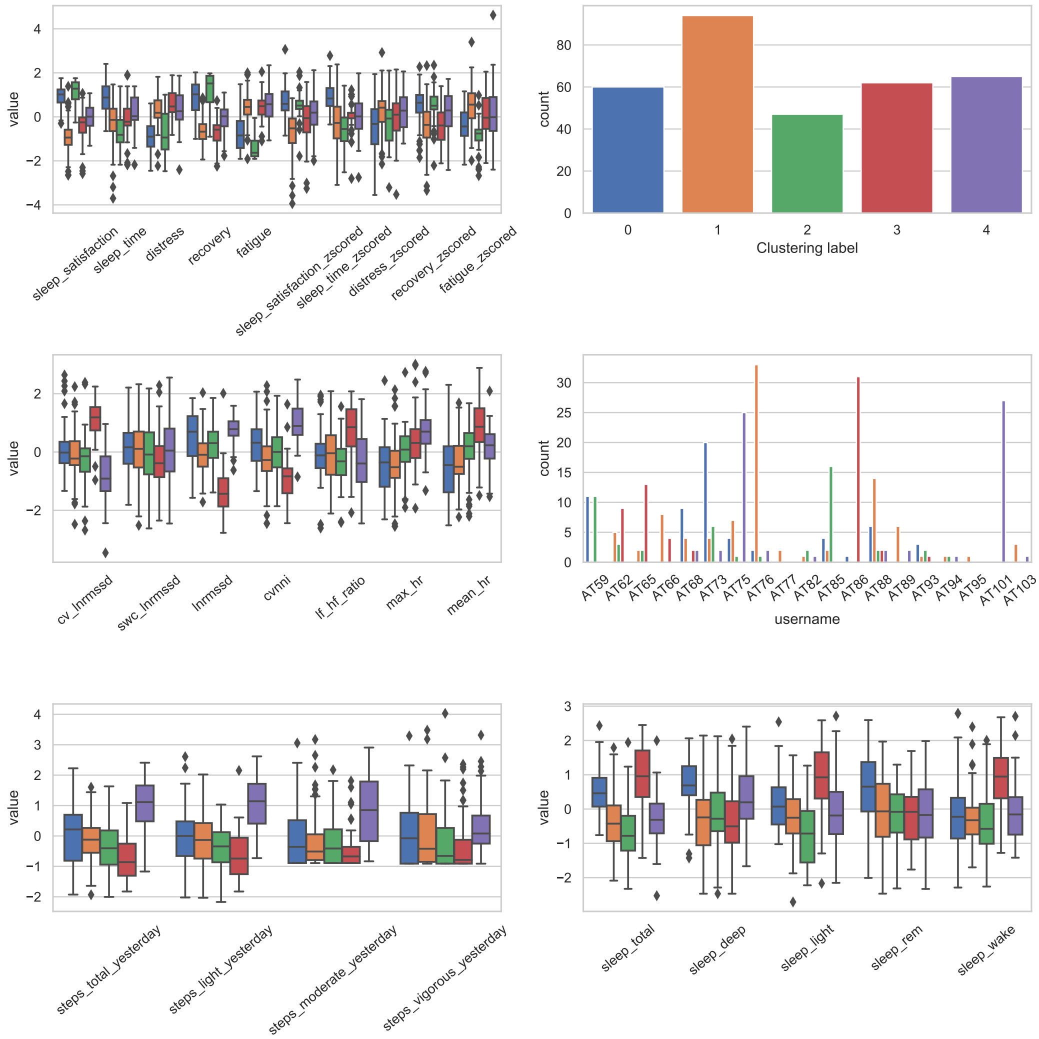


Fig. C.4.: K-Means clustering experiment C01-5.

C.2 C02. HRV and Wellness

Results for C02 with HRV and Wellness features are displayed in Table C.2, Figure C.5, Figure C.6, Figure C.7 and Figure C.8.

Tab. C.2.: Silhouette scores for K-Means clustering experiment C02.

n_clusters	silhouette
2	0.155735
3	0.136509
4	0.130735
5	0.114407

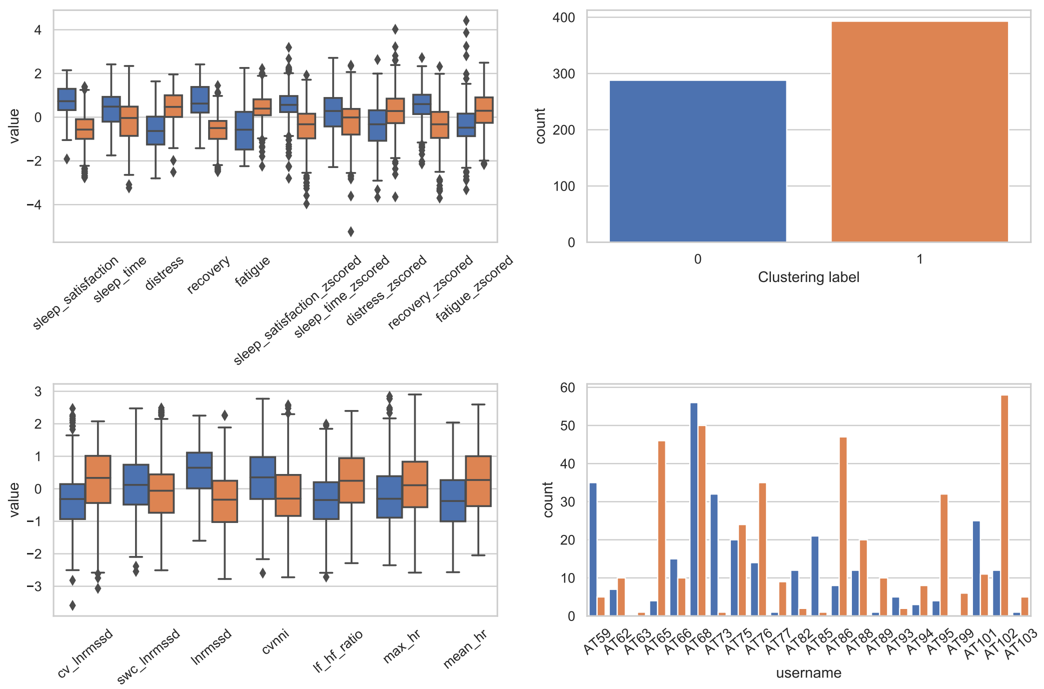


Fig. C.5.: K-Means clustering experiment C02-2.

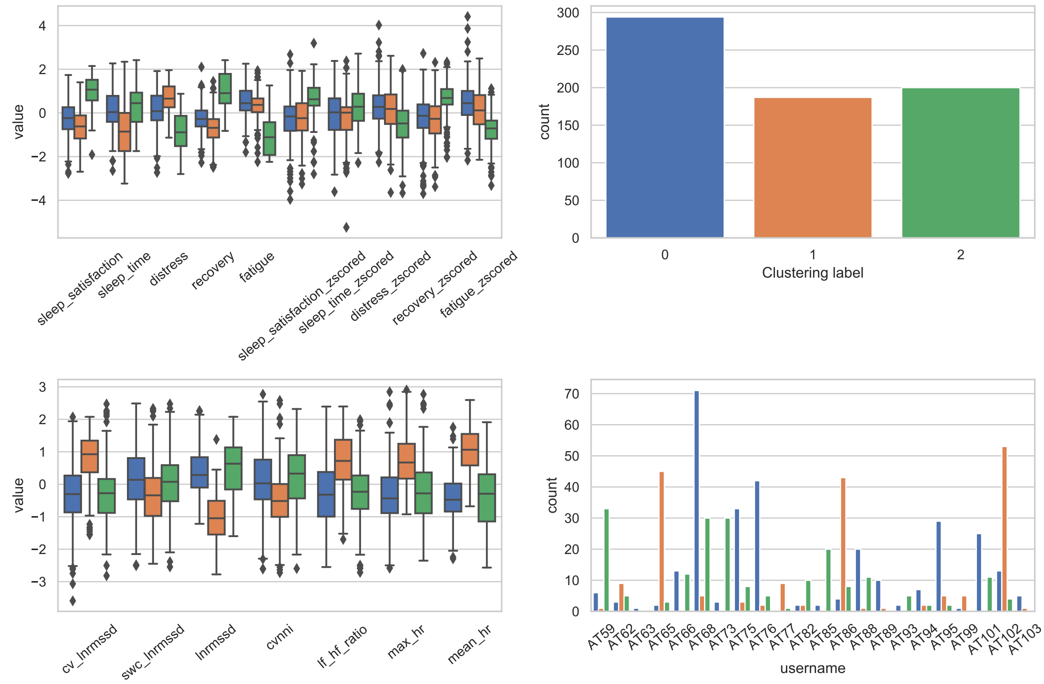


Fig. C.6.: K-Means clustering experiment C02-3.

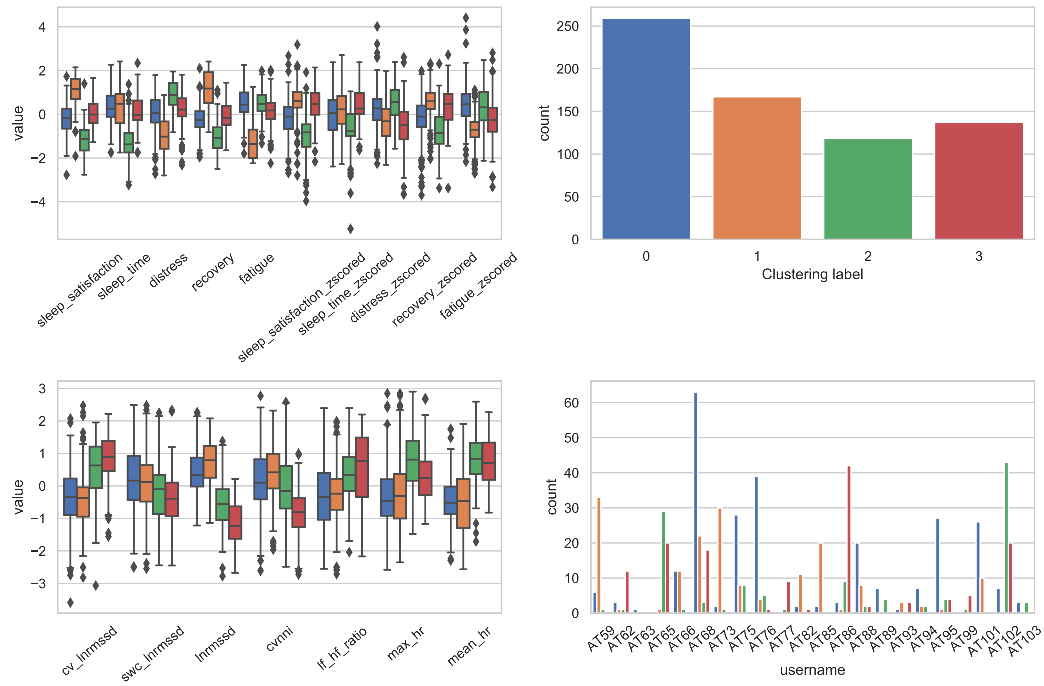


Fig. C.7.: K-Means clustering experiment C02-4.

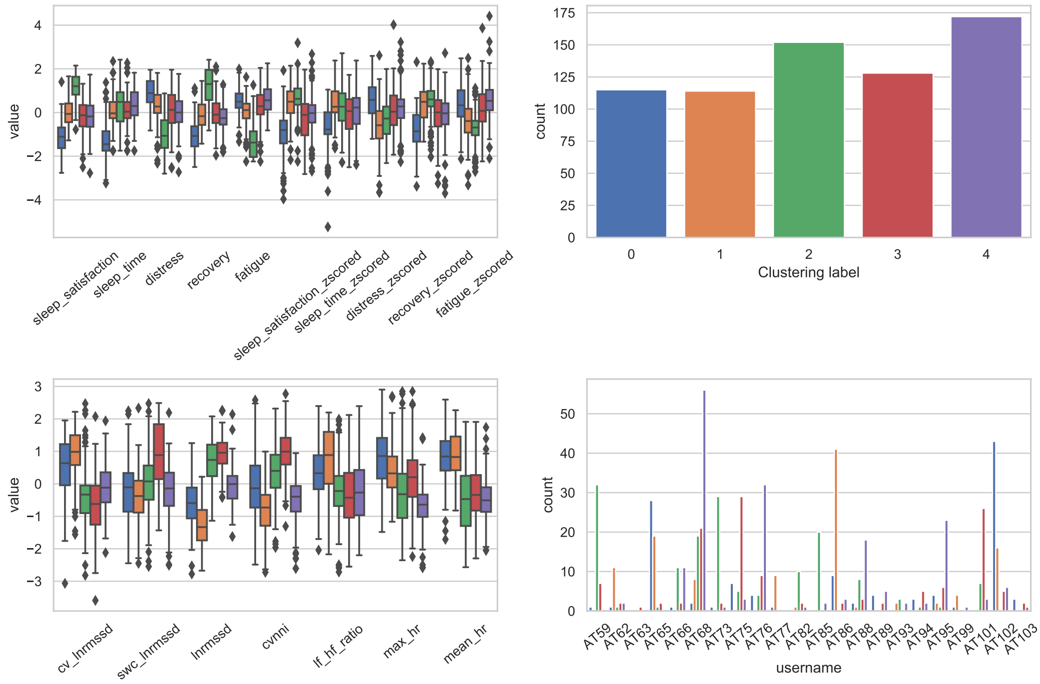


Fig. C.8.: K-Means clustering experiment C02-5.

C.3 L01A. All HRV features

Results for L01A with HRV and Wellness features are displayed in Table C.3, Figure C.9, Figure C.10, Figure C.11 and Figure C.12.

Tab. C.3.: Silhouette scores for K-Means clustering experiment L01A.

n_clusters	silhouette
2	0.252863
3	0.212174
4	0.187604
5	0.177786

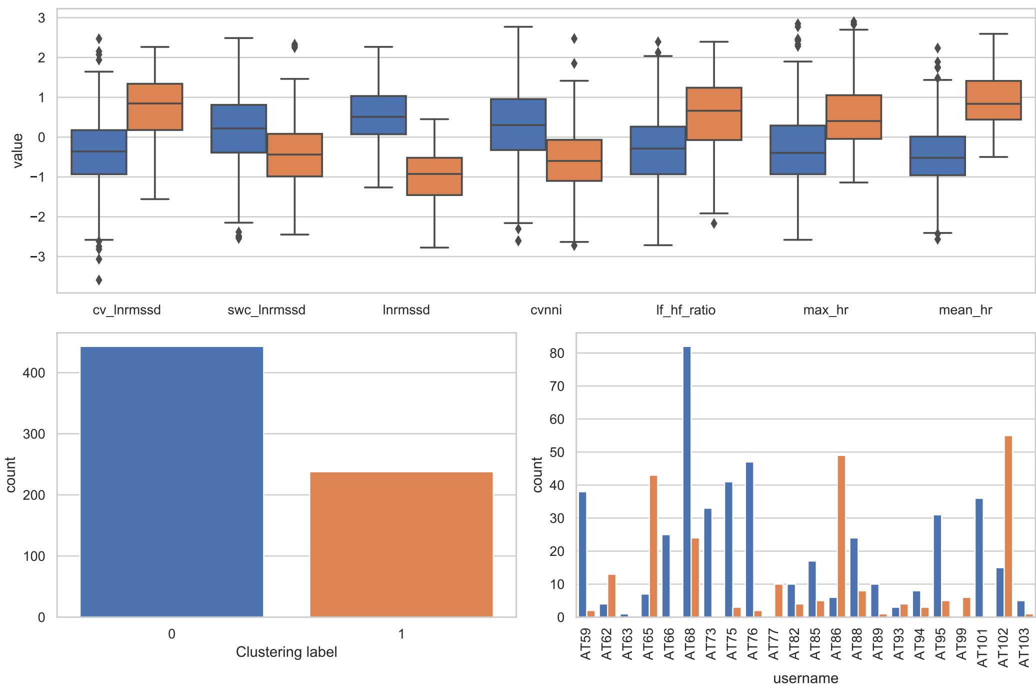


Fig. C.9.: K-Means clustering experiment L01A-2.

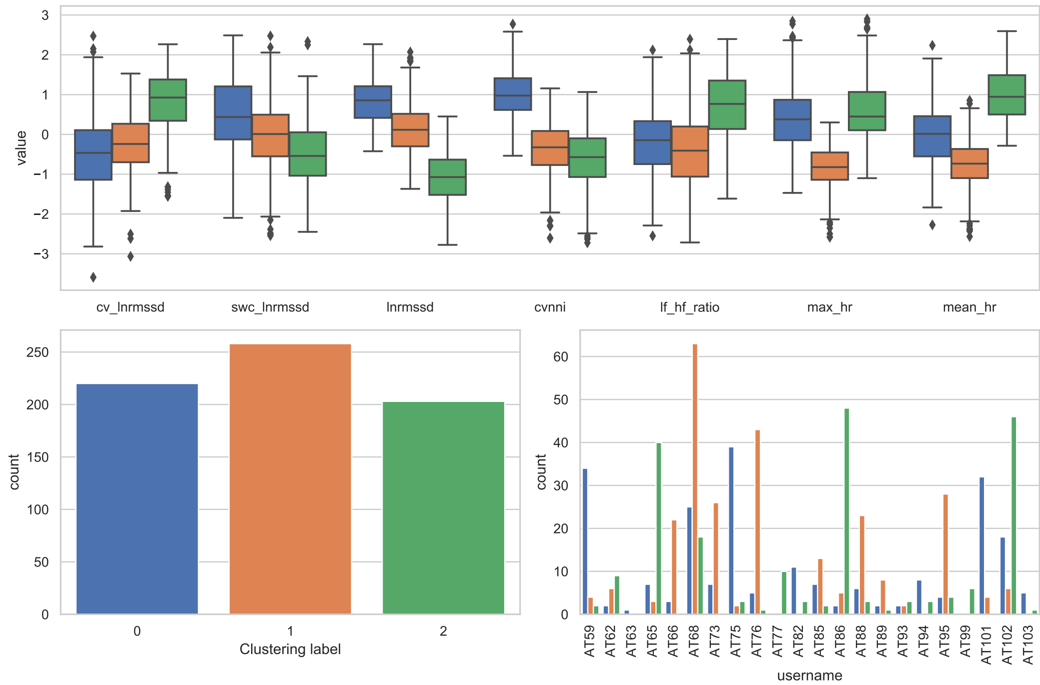


Fig. C.10.: K-Means clustering experiment L01A-3.

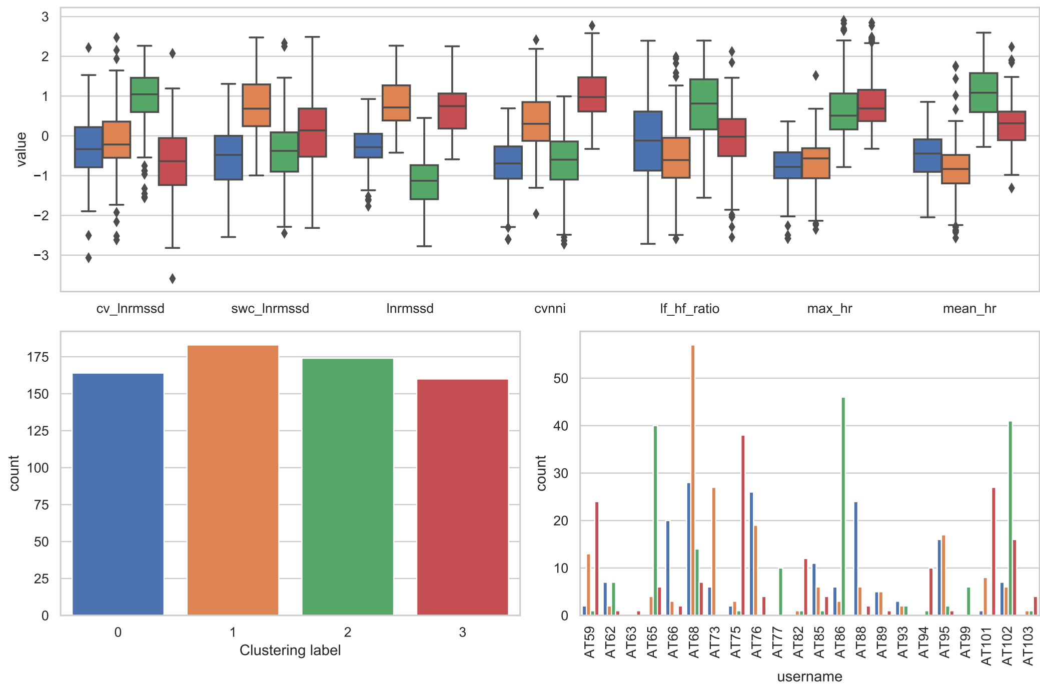


Fig. C.11.: K-Means clustering experiment L01A-4.

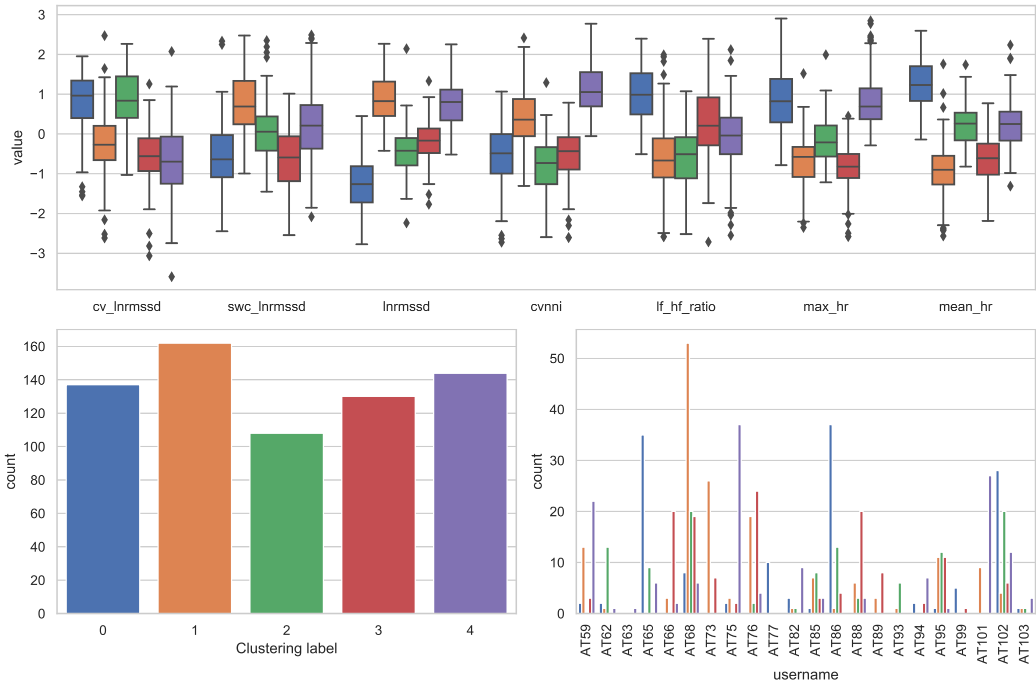


Fig. C.12.: K-Means clustering experiment L01A-5.

Tab. C.4.: Numerical description of features (de-normalized) for L01A-4.

	label	0	1	2	3
cv_lnrmssd	count	164.00	183.00	174.00	160.00
	mean	6.73	7.50	13.19	5.73
	std	3.27	3.91	4.60	3.02
	min	1.26	1.63	2.96	0.94
	25%	4.58	5.24	10.03	3.55
	50%	5.92	6.32	12.91	4.99
	75%	8.10	8.76	16.33	6.93
	max	25.07	28.92	25.74	23.09
swc_lnrmssd	count	164.00	183.00	174.00	160.00
	mean	-0.74	1.12	-0.54	0.21
	std	1.11	1.09	1.15	1.34
	min	-3.49	-1.35	-3.36	-3.18
	25%	-1.49	0.37	-1.21	-0.69
	50%	-0.63	0.98	-0.49	0.22
	75%	0.04	1.83	0.16	0.99
	max	1.85	3.47	3.27	3.48
lnrmssd	count	164.00	183.00	174.00	160.00
	mean	3.29	3.98	2.73	3.88
	std	0.32	0.39	0.40	0.35
	min	2.37	3.21	1.74	3.11
	25%	3.13	3.72	2.48	3.59
	50%	3.30	3.92	2.77	3.94
	75%	3.51	4.27	3.01	4.14
	max	4.05	4.89	3.76	4.88
cvnni (x100)	count	164.00	183.00	174.00	160.00
	mean	4.36	7.05	4.58	9.47
	std	1.22	2.27	1.42	2.83
	min	1.85	2.45	1.76	5.00
	25%	3.61	5.46	3.57	7.53
	50%	4.26	6.58	4.45	8.82
	75%	5.14	8.35	5.42	10.95
	max	7.80	16.51	8.90	19.29
lf_hf_ratio	count	164.00	183.00	174.00	160.00
	mean	2.97	1.96	5.03	2.66
	std	2.60	1.65	3.08	1.75
	min	0.28	0.31	0.70	0.32
	25%	1.19	1.03	2.66	1.58
	50%	2.13	1.46	4.41	2.30
	75%	3.77	2.25	7.07	3.25
	max	15.03	10.99	15.06	12.16
max_hr	count	164.00	183.00	174.00	160.00
	mean	71.63	72.61	88.32	90.12
	std	6.17	6.84	8.78	7.86
	min	51.24	53.81	71.60	76.82
	25%	68.42	68.42	82.33	84.71
	50%	71.64	74.07	86.27	88.31
	75%	75.81	76.97	92.59	93.64
	max	84.63	97.72	113.42	112.78
mean_hr	count	164.00	183.00	174.00	160.00
	mean	62.37	59.16	76.30	69.17
	std	5.77	6.30	5.69	5.39
	min	48.54	43.94	64.21	55.05
	25%	58.65	56.09	71.95	65.70
	50%	62.73	59.29	76.26	69.47
	75%	65.88	62.42	80.63	72.07
	max	74.23	82.23	89.63	86.47

C.4 L01B. All HRV features except max_hr and mean_hr

Results for L01B with HRV and Wellness features are displayed in Table C.5, Figure C.13, Figure C.14, Figure C.15 and Figure C.16.

Tab. C.5.: Silhouette scores for K-Means clustering experiment L01B.

n_clusters	silhouette
2	0.273946
3	0.204155
4	0.199767
5	0.196936

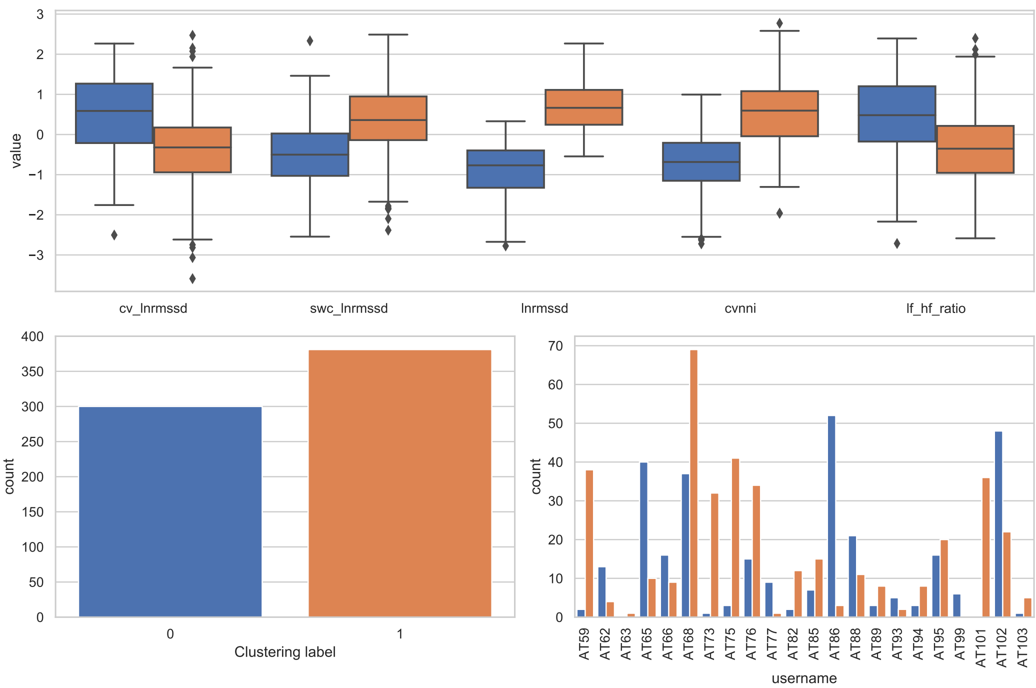


Fig. C.13.: K-Means clustering experiment L01B-2.

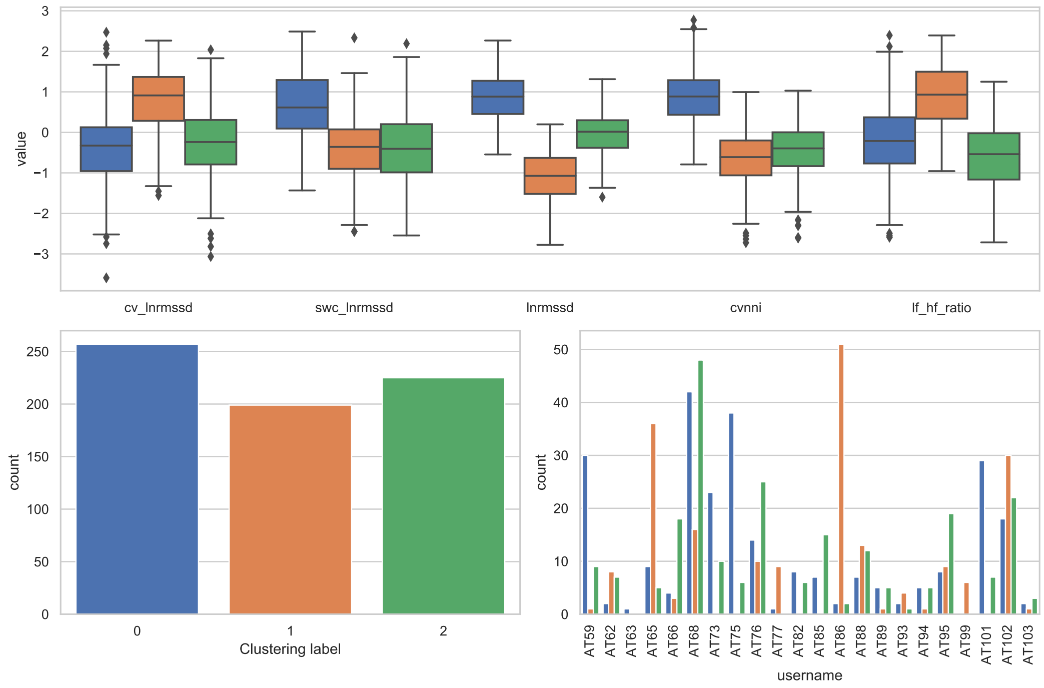


Fig. C.14.: K-Means clustering experiment L01B-3.

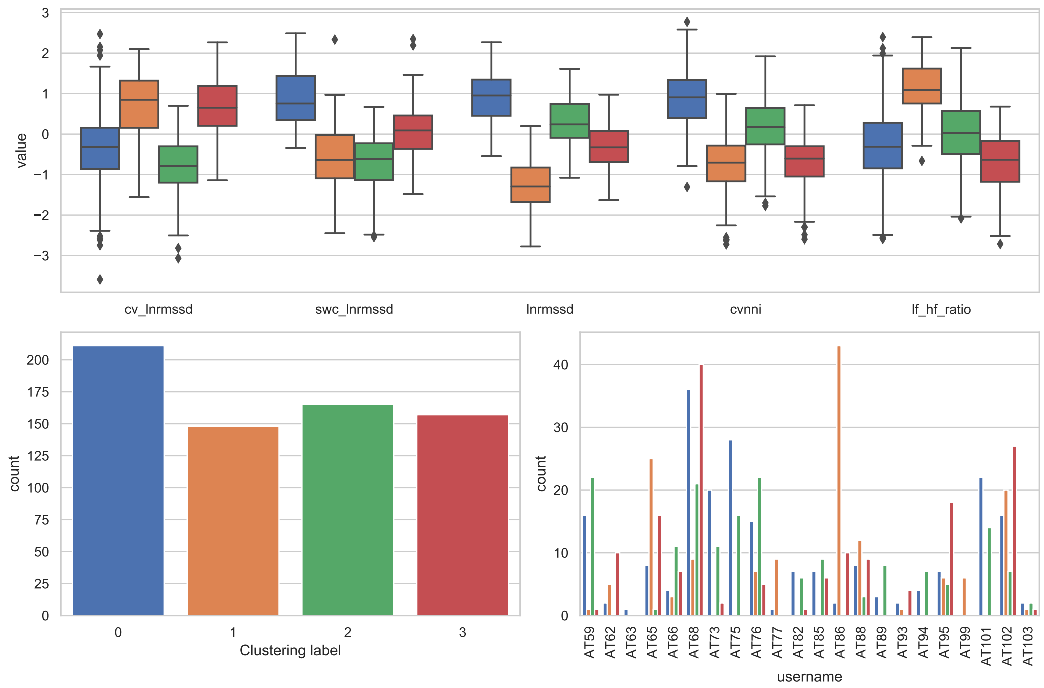


Fig. C.15.: K-Means clustering experiment L01B-4.

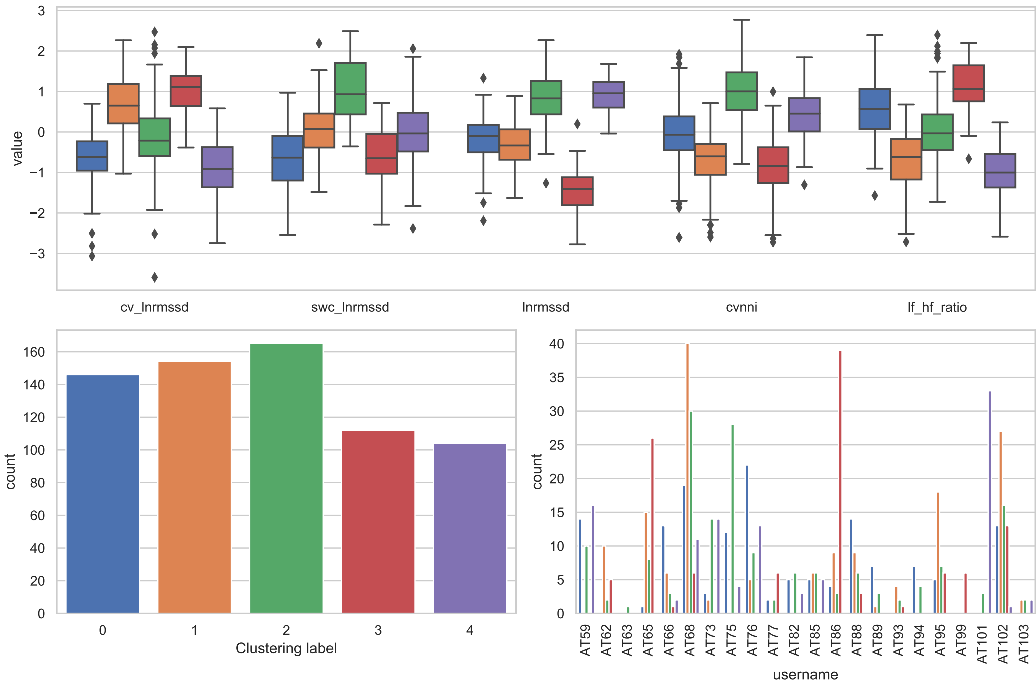


Fig. C.16.: K-Means clustering experiment L01B-5.

Tab. C.6.: Numerical description of features (de-normalized) for L01B-4.

	label	0	1	2	3
cv_lnrmssd	count	211.00	148.00	165.00	157.00
	mean	6.66	11.42	4.89	11.38
	std	3.78	4.65	1.69	4.61
	min	0.94	2.96	1.26	3.75
	25%	4.39	7.81	3.63	8.03
	50%	5.98	11.55	4.58	10.34
	75%	7.82	15.09	6.03	14.03
	max	28.92	23.42	10.62	25.74
swc_lnrmssd	count	211.00	148.00	165.00	157.00
	mean	1.31	-0.80	-0.98	0.17
	std	1.02	1.13	0.96	0.92
	min	-0.44	-3.36	-3.49	-2.02
	25%	0.52	-1.48	-1.54	-0.47
	50%	1.08	-0.85	-0.82	0.16
	75%	2.03	-0.00	-0.28	0.67
	max	3.48	3.27	0.96	3.29
lnrmssd	count	211.00	148.00	165.00	157.00
	mean	4.04	2.66	3.65	3.29
	std	0.38	0.40	0.33	0.34
	min	3.13	1.74	2.80	2.45
	25%	3.76	2.42	3.42	3.04
	50%	4.07	2.67	3.62	3.27
	75%	4.32	2.96	3.94	3.52
	max	4.89	3.60	4.48	4.08
cvnni (x100)	count	211.00	148.00	165.00	157.00
	mean	9.06	4.37	6.48	4.41
	std	2.93	1.40	1.92	1.21
	min	3.27	1.76	2.66	1.86
	25%	6.85	3.47	5.16	3.65
	50%	8.56	4.24	6.21	4.43
	75%	10.33	5.10	7.62	5.06
	max	19.29	8.90	13.31	7.87
lf_hf_ratio	count	211.00	148.00	165.00	157.00
	mean	2.42	6.23	2.85	1.55
	std	2.02	2.82	1.90	0.75
	min	0.31	1.40	0.46	0.28
	25%	1.21	4.21	1.60	0.94
	50%	1.84	5.44	2.39	1.43
	75%	2.91	8.23	3.65	2.04
	max	15.06	15.03	12.22	3.97

C.5 L01C. All HRV features except max_hr, mean_hr, and lf_hf_ratio

Results for L01C with HRV and Wellness features are displayed in Table C.7, Figure C.17, Figure C.18, Figure C.19 and Figure C.20.

Tab. C.7.: Silhouette scores for K-Means clustering experiment L01C.

n_clusters	silhouette
2	0.311806
3	0.257434
4	0.234708
5	0.238176

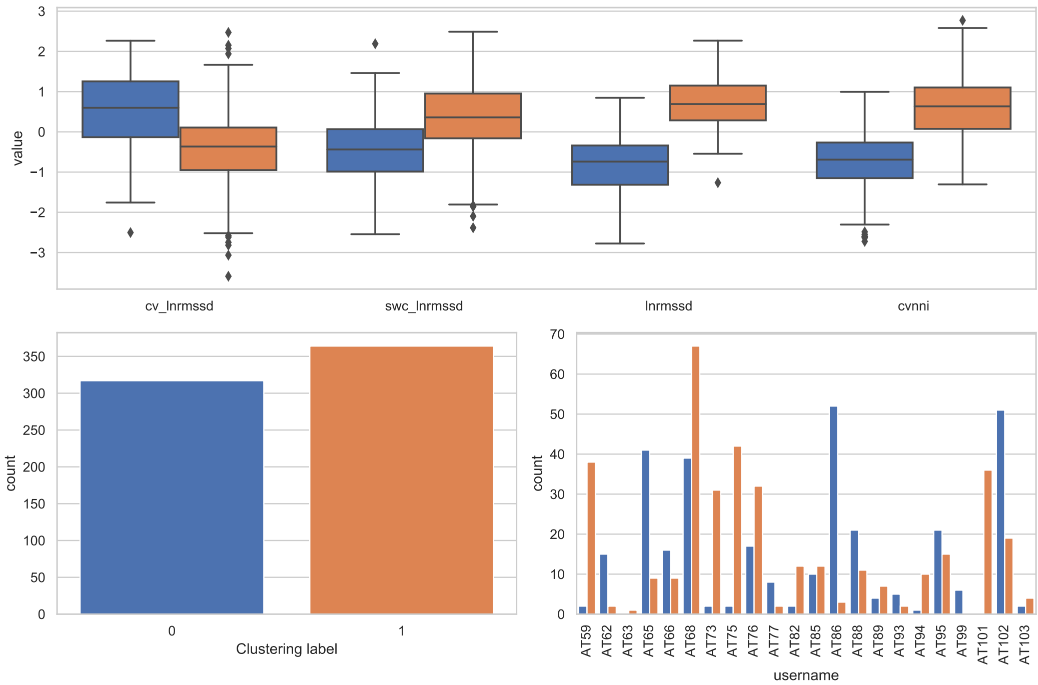


Fig. C.17.: K-Means clustering experiment L01C-2.

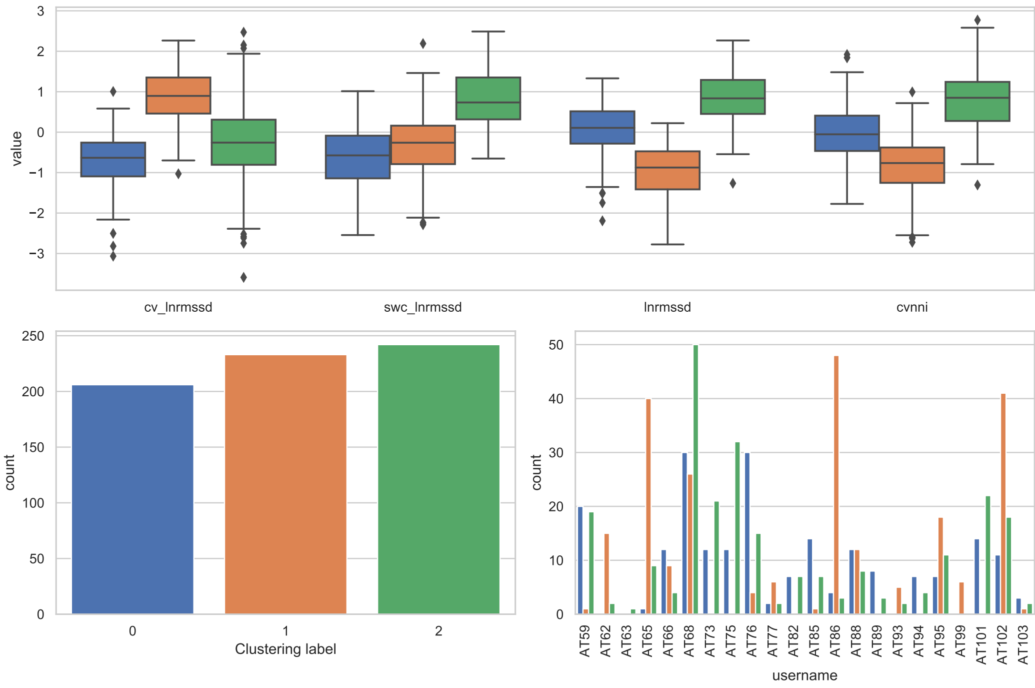


Fig. C.18.: K-Means clustering experiment L01C-3.

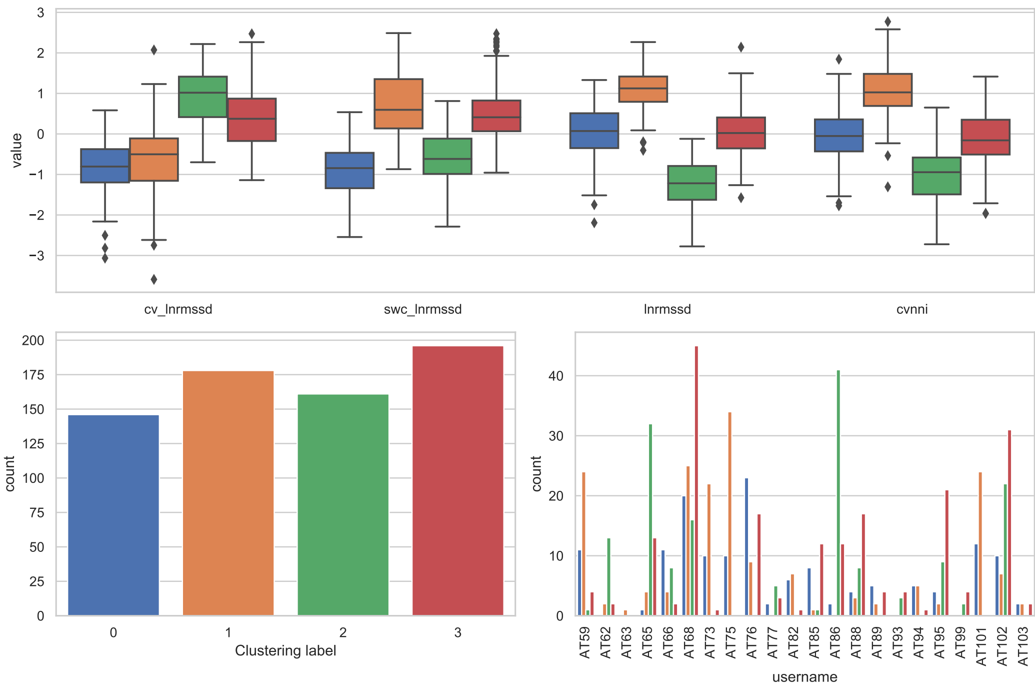


Fig. C.19.: K-Means clustering experiment L01C-4.

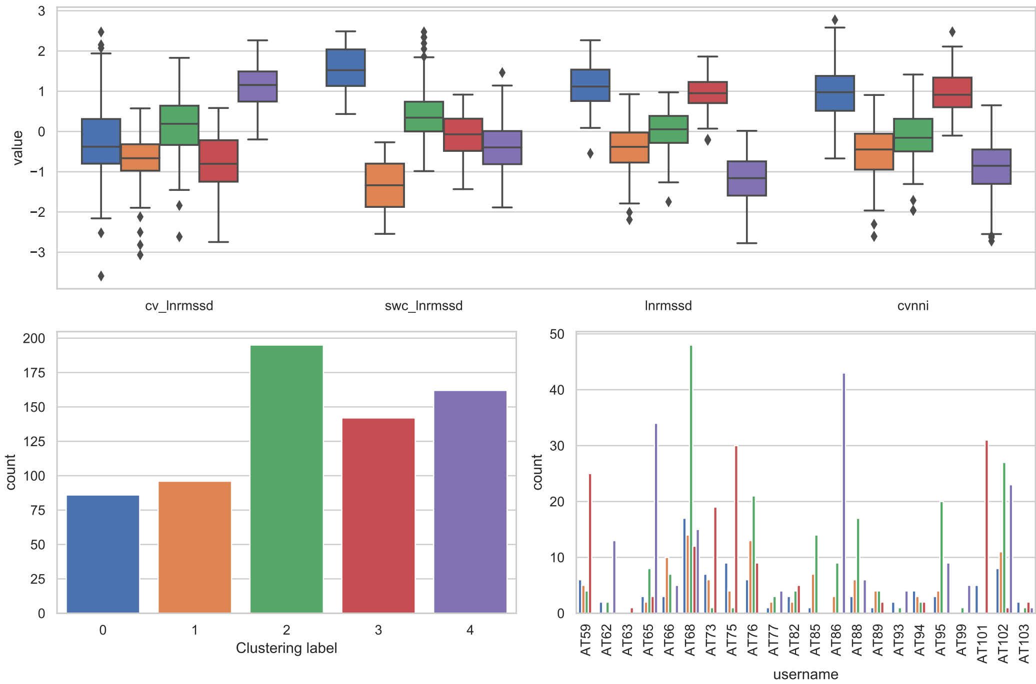


Fig. C.20.: K-Means clustering experiment L01C-5.

Tab. C.8.: Numerical description of features (de-normalized) for L01C-4.

	label	0	1	2	3
cv_lnrmssd	count	146.00	178.00	161.00	196.00
	mean	4.75	5.59	12.95	9.76
	std	1.56	2.66	4.54	4.19
	min	1.26	0.94	4.82	3.75
	25%	3.63	3.72	9.05	6.48
	50%	4.54	5.38	12.71	8.84
	75%	5.78	6.72	15.91	11.71
	max	9.94	23.09	25.07	28.92
swc_lnrmssd	count	146.00	178.00	161.00	196.00
	mean	-1.27	1.06	-0.81	0.77
	std	0.97	1.14	0.89	0.90
	min	-3.49	-1.17	-3.14	-1.29
	25%	-1.82	0.22	-1.33	0.13
	50%	-1.14	0.86	-0.82	0.60
	75%	-0.61	1.91	-0.13	1.18
	max	0.78	3.48	1.16	3.47
lnrmssd	count	146.00	178.00	161.00	196.00
	mean	3.51	4.15	2.69	3.48
	std	0.40	0.32	0.38	0.32
	min	2.10	3.22	1.74	2.49
	25%	3.26	3.97	2.46	3.25
	50%	3.52	4.18	2.71	3.49
	75%	3.79	4.36	2.98	3.73
	max	4.31	4.89	3.40	4.81
cvnni (x100)	count	146.00	178.00	161.00	196.00
	mean	5.89	9.65	3.84	5.74
	std	1.71	2.85	1.14	1.63
	min	2.66	3.27	1.76	2.45
	25%	4.78	7.79	3.01	4.62
	50%	5.64	9.02	3.82	5.38
	75%	6.74	11.00	4.48	6.72
	max	12.88	19.29	7.66	10.69

C.6 L02A. Wellness features

Results for L02A with HRV and Wellness features are displayed in Table C.9, Figure C.21, Figure C.22, Figure C.23 and Figure C.24.

Tab. C.9.: Silhouette scores for K-Means clustering experiment L02A.

n_clusters	silhouette
2	0.374903
3	0.271060
4	0.223430
5	0.211223

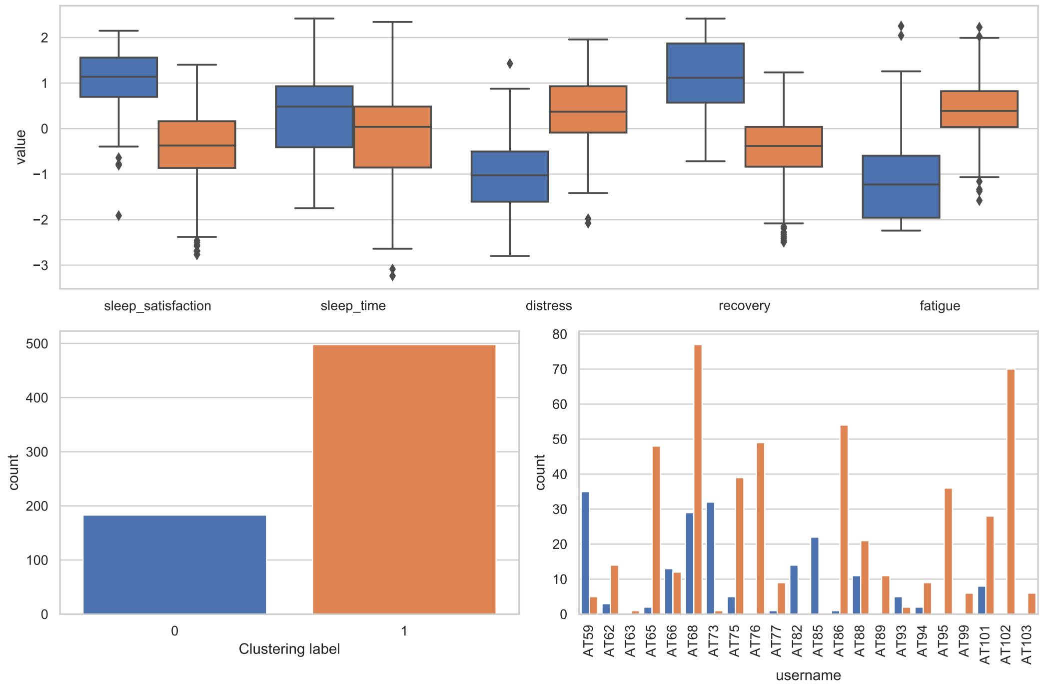


Fig. C.21.: K-Means clustering experiment L02A-2.

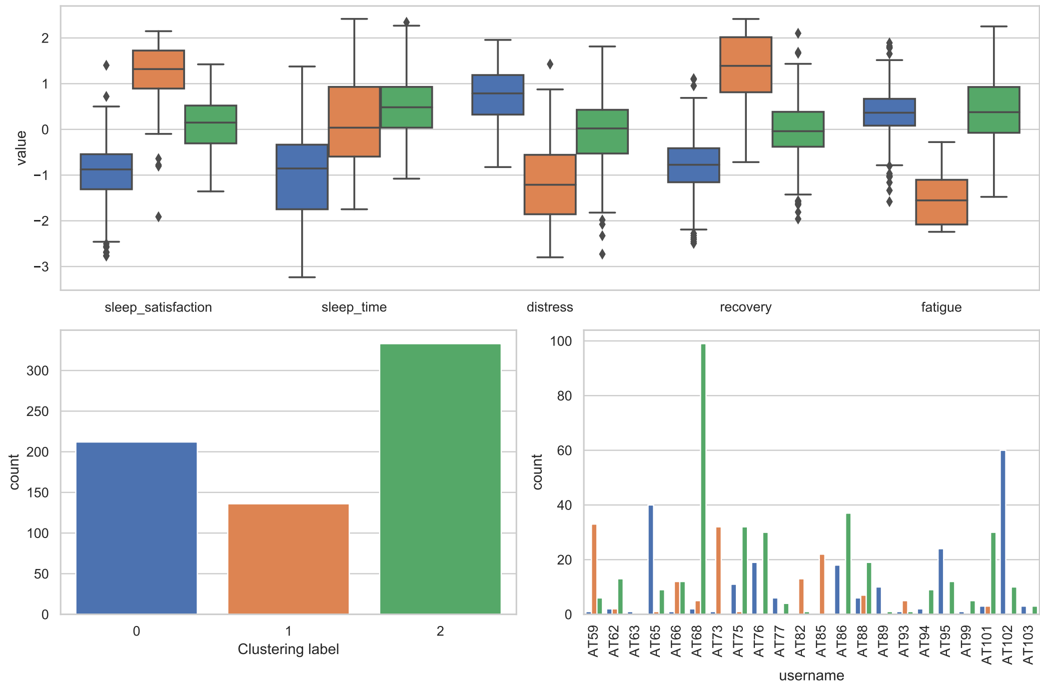


Fig. C.22.: K-Means clustering experiment L02A-3.

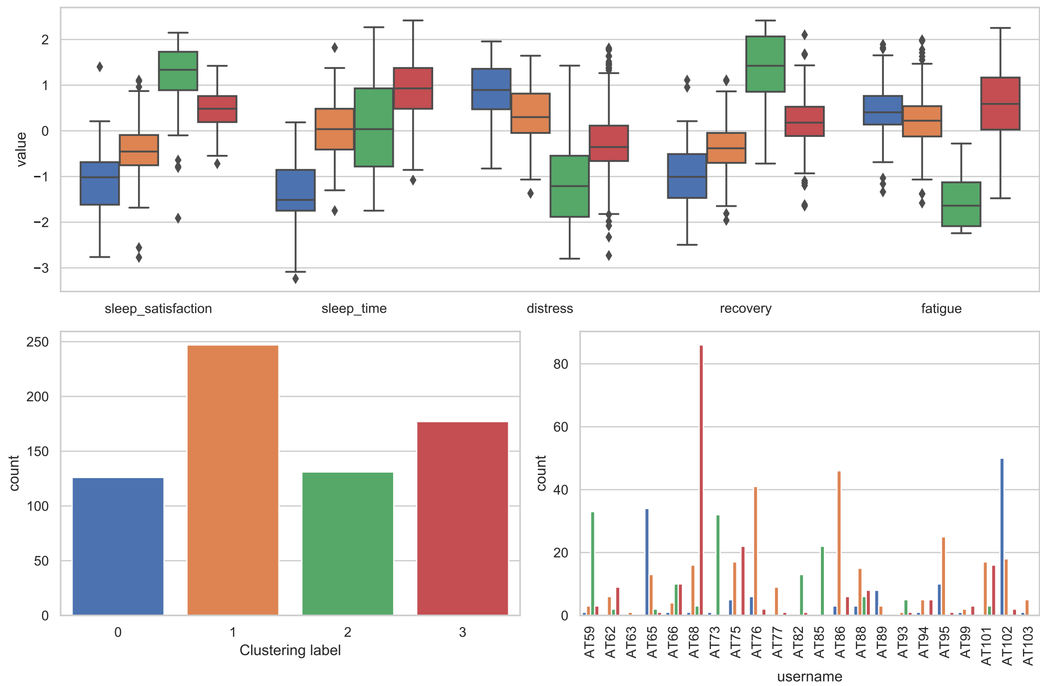


Fig. C.23.: K-Means clustering experiment L02A-4.

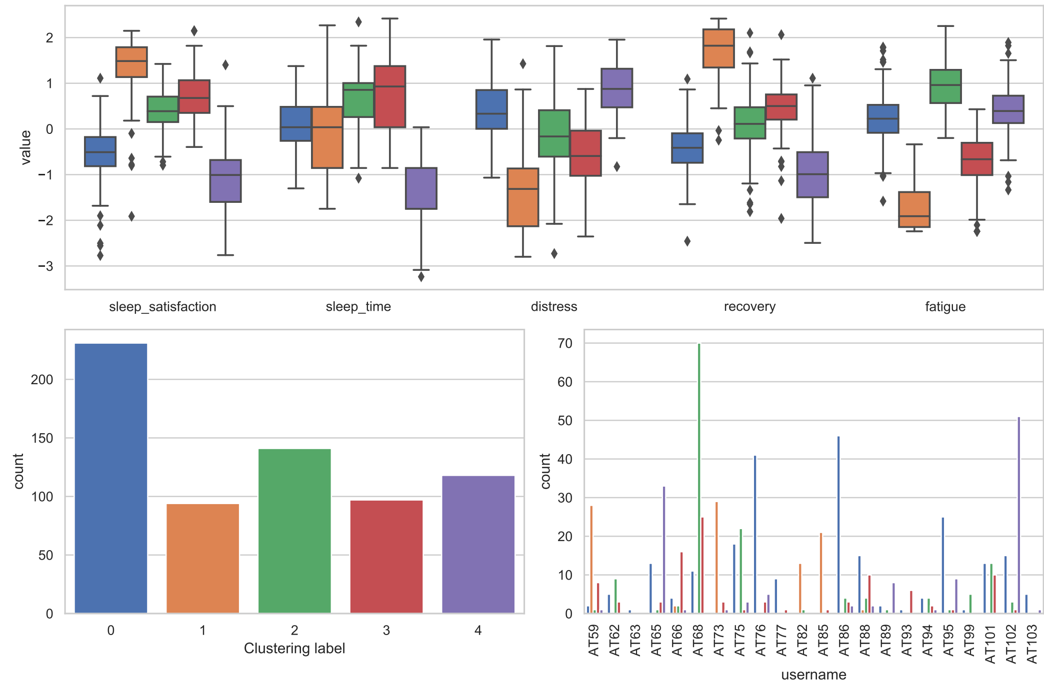


Fig. C.24.: K-Means clustering experiment L02A-5.

Tab. C.10.: Numerical description of features (de-normalized) for L02A-3.

	label	0	1	2
sleep_satisfaction	count	212.00	136.00	333.00
	mean	4.37	8.36	6.31
	std	1.24	1.19	1.04
	min	1.11	2.66	3.67
	25%	3.75	7.73	5.56
	50%	4.53	8.50	6.39
	75%	5.14	9.23	7.06
	max	8.65	10.00	8.69
sleep_time	count	212.00	136.00	333.00
	mean	5.93	7.17	7.53
	std	0.91	1.10	0.74
	min	3.33	5.00	5.75
	25%	5.00	6.29	7.00
	50%	6.00	7.00	7.50
	75%	6.58	8.00	8.00
	max	8.50	9.67	9.58
distress	count	212.00	136.00	333.00
	mean	4.02	1.07	2.43
	std	1.80	0.85	1.51
	min	1.10	0.24	0.26
	25%	2.63	0.50	1.38
	50%	3.75	0.82	2.09
	75%	5.08	1.35	2.85
	max	9.13	6.10	8.18
recovery	count	212.00	136.00	333.00
	mean	4.53	8.17	5.83
	std	1.07	1.21	1.01
	min	1.75	4.70	2.64
	25%	3.97	7.24	5.26
	50%	4.61	8.20	5.83
	75%	5.21	9.24	6.53
	max	7.73	9.90	9.38
fatigue	count	212.00	136.00	333.00
	mean	4.93	1.33	5.05
	std	1.11	1.10	1.32
	min	1.26	0.00	1.46
	25%	4.43	0.31	4.13
	50%	4.96	1.31	4.99
	75%	5.54	2.17	6.04
	max	7.88	3.74	8.56

C.7 L02B. Wellness features except sleep_time

Results for L02B with HRV and Wellness features are displayed in Table C.11, Figure C.25, Figure C.26, Figure C.27 and Figure C.28.

Tab. C.11.: Silhouette scores for K-Means clustering experiment L02B.

n_clusters	silhouette
2	0.442257
3	0.279222
4	0.248510
5	0.248656

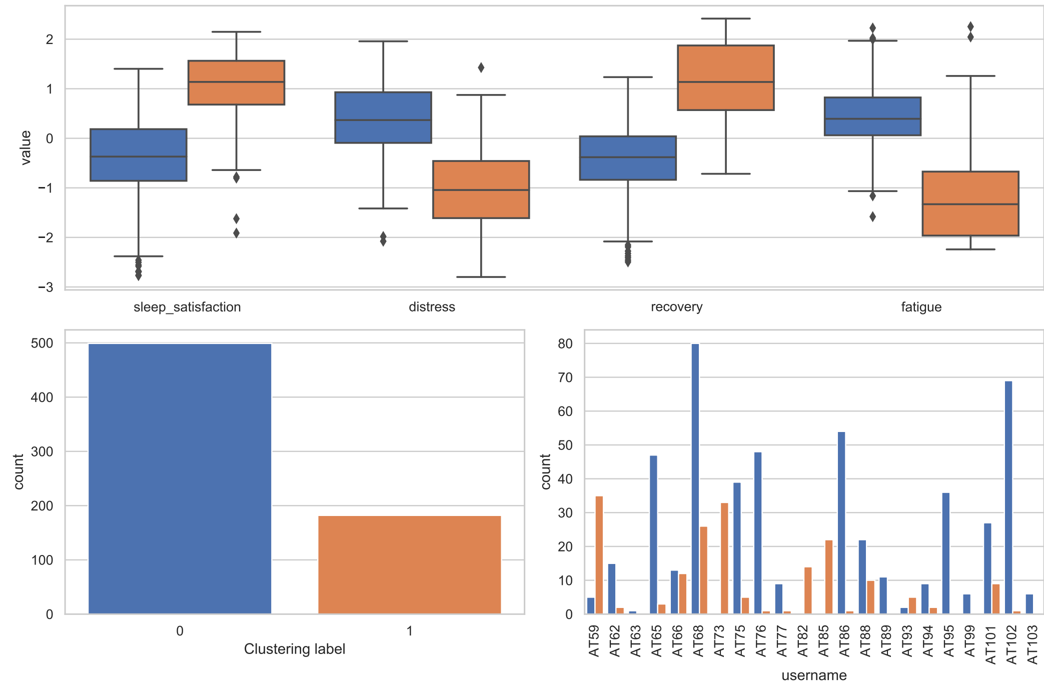


Fig. C.25.: K-Means clustering experiment L02B-2.

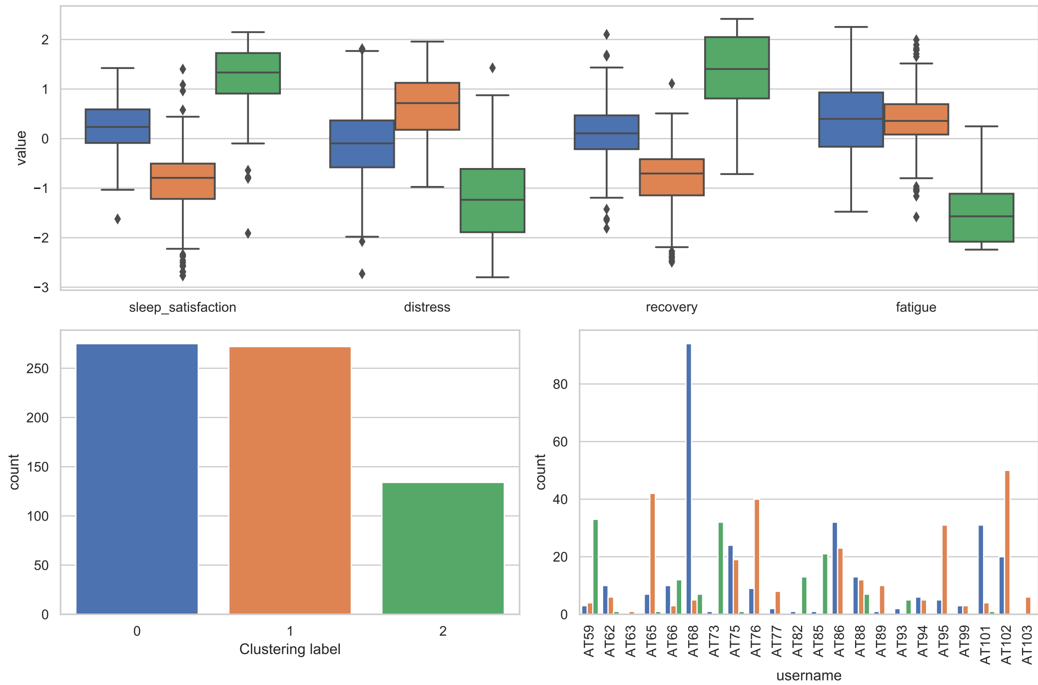


Fig. C.26.: K-Means clustering experiment L02B-3.

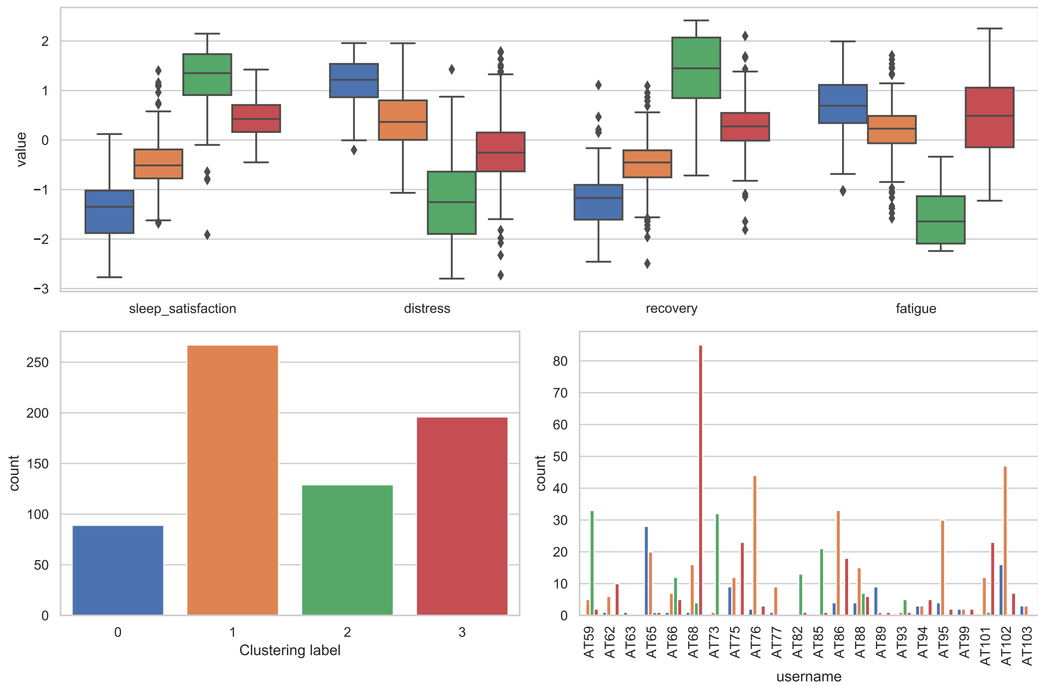


Fig. C.27.: K-Means clustering experiment L02B-4.

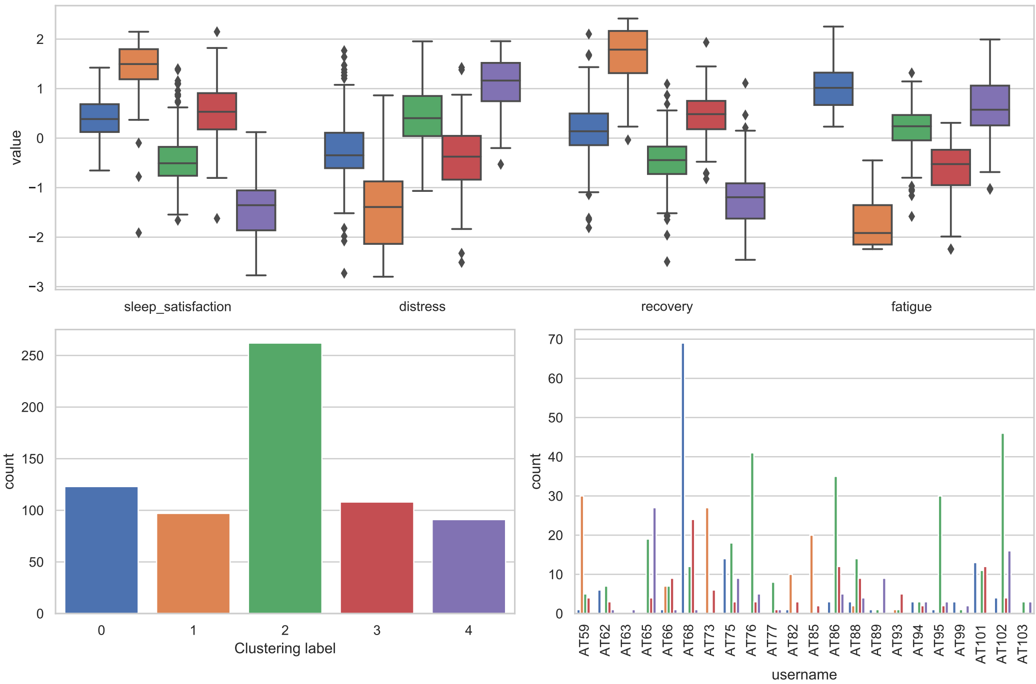


Fig. C.28.: K-Means clustering experiment L02B-5.

Tab. C.12.: Numerical description of features for L02B-3.

	label	0	1	2
sleep_satisfaction	count	275.00	272.00	134.00
	mean	6.57	4.54	8.39
	std	0.91	1.20	1.18
	min	3.19	1.11	2.66
	25%	5.96	3.92	7.76
	50%	6.54	4.68	8.53
	75%	7.18	5.20	9.24
	max	8.69	8.65	10.00
distress	count	275.00	272.00	134.00
	mean	2.25	3.86	1.04
	std	1.40	1.78	0.84
	min	0.26	0.98	0.24
	25%	1.32	2.35	0.49
	50%	1.91	3.55	0.80
	75%	2.72	4.84	1.29
	max	8.18	9.13	6.10
recovery	count	275.00	272.00	134.00
	mean	6.08	4.58	8.17
	std	0.93	0.98	1.22
	min	2.89	1.75	4.70
	25%	5.54	3.99	7.23
	50%	6.07	4.72	8.22
	75%	6.67	5.20	9.29
	max	9.38	7.73	9.90
fatigue	count	275.00	272.00	134.00
	mean	5.00	4.98	1.32
	std	1.39	1.09	1.12
	min	1.46	1.26	0.00
	25%	3.96	4.43	0.30
	50%	5.03	4.95	1.28
	75%	6.04	5.59	2.15
	max	8.56	8.07	4.74

C.8 L03A. Normalized wellness features except sleep_time

Results for L03A with HRV and Wellness features are displayed in Table C.13, Figure C.29, Figure C.30, Figure C.31 and Figure C.32.

Tab. C.13.: Silhouette scores for K-Means clustering experiment L03A.

n_clusters	silhouette
2	0.266132
3	0.210561
4	0.196307
5	0.201947

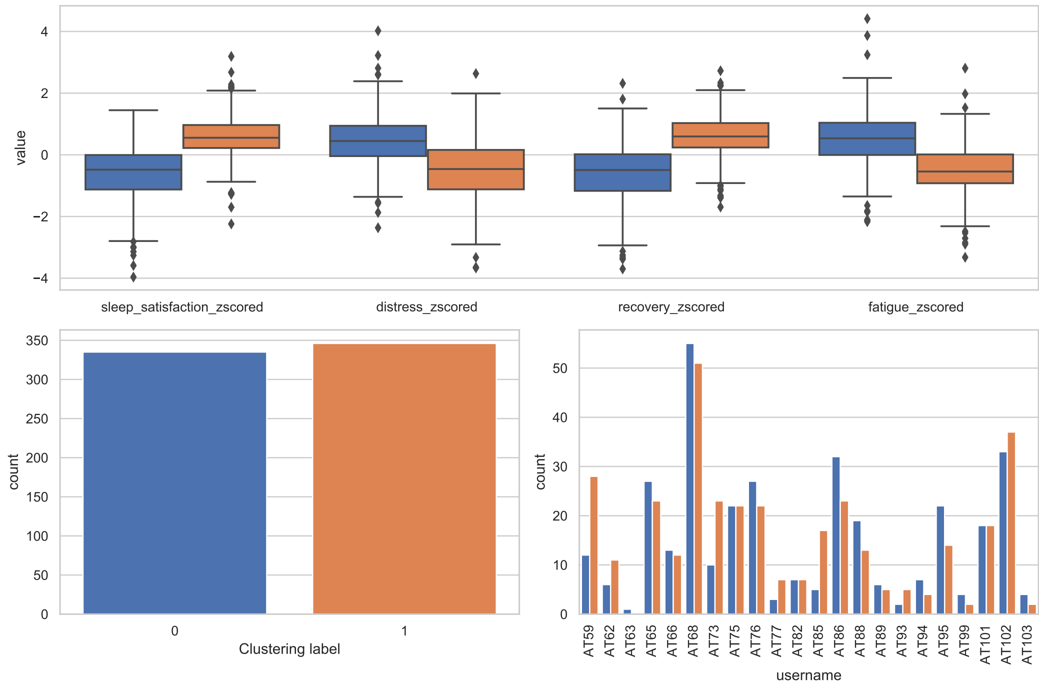


Fig. C.29.: K-Means clustering experiment L03A-2.

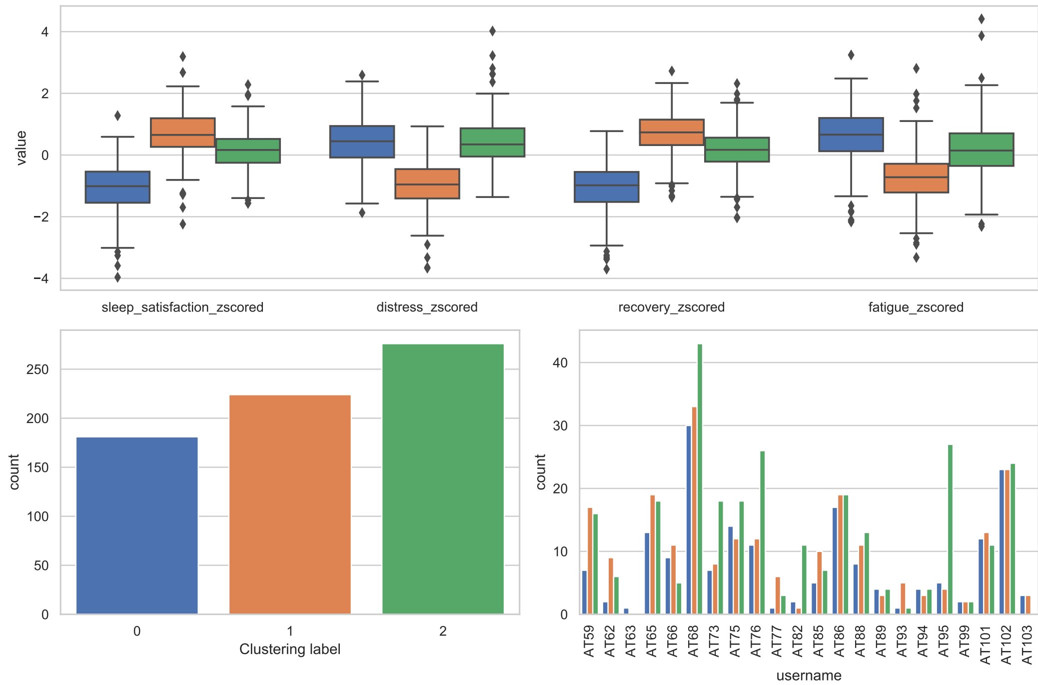


Fig. C.30.: K-Means clustering experiment L03A-3.

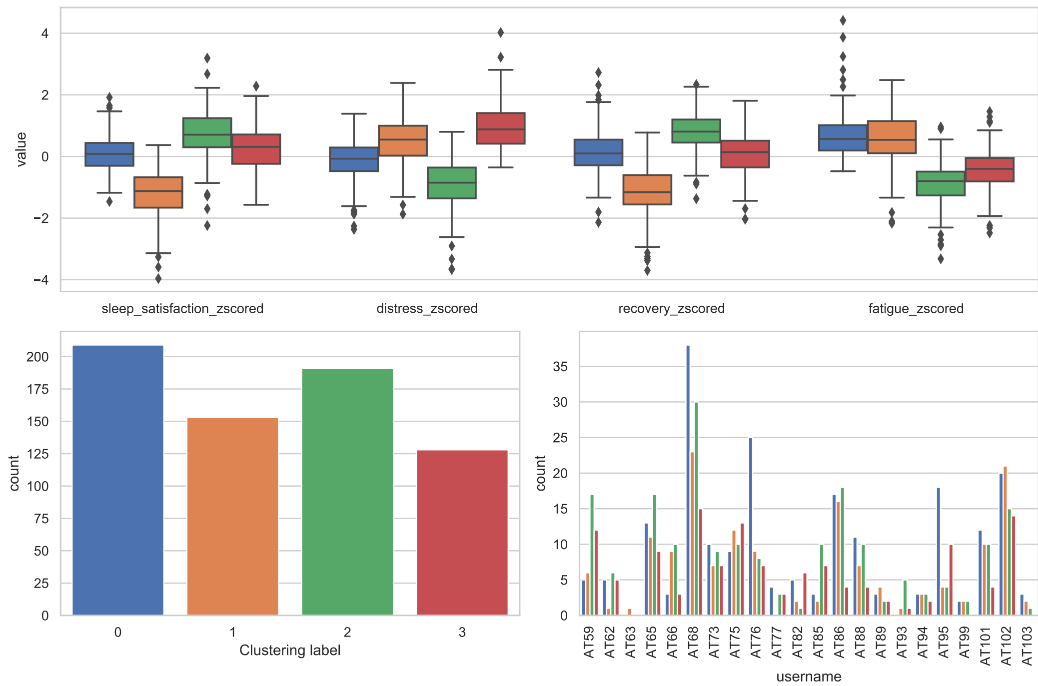


Fig. C.31.: K-Means clustering experiment L03A-4.

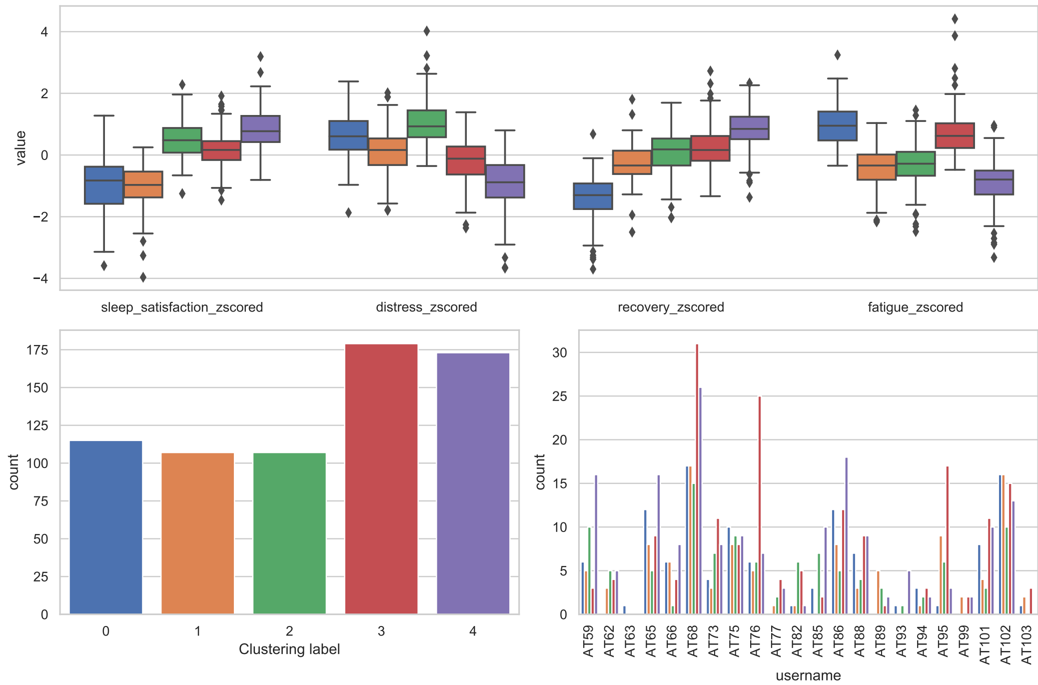


Fig. C.32.: K-Means clustering experiment L03A-5.

Tab. C.14.: Numerical description of features for L03A-3.

	label	0	1	2
sleep_satisfaction_zscored	count	181.00	224.00	276.00
	mean	-1.03	0.68	0.17
	std	0.80	0.74	0.59
	min	-3.80	-2.14	-1.49
	25%	-1.47	0.27	-0.22
	50%	-0.96	0.65	0.18
	75%	-0.50	1.17	0.52
	max	1.25	3.10	2.22
distress_zscored	count	181.00	224.00	276.00
	mean	0.43	-0.84	0.42
	std	0.69	0.70	0.69
	min	-1.70	-3.35	-1.24
	25%	-0.07	-1.28	-0.04
	50%	0.41	-0.87	0.32
	75%	0.87	-0.41	0.80
	max	2.38	0.86	3.69
recovery_zscored	count	181.00	224.00	276.00
	mean	-1.00	0.63	0.16
	std	0.76	0.66	0.60
	min	-3.39	-1.25	-1.86
	25%	-1.39	0.30	-0.19
	50%	-0.90	0.68	0.16
	75%	-0.50	1.06	0.52
	max	0.72	2.50	2.13
fatigue_zscored	count	181.00	224.00	276.00
	mean	0.58	-0.62	0.19
	std	0.79	0.80	0.73
	min	-1.93	-2.98	-2.07
	25%	0.14	-1.07	-0.29
	50%	0.62	-0.62	0.16
	75%	1.11	-0.23	0.66
	max	2.96	2.57	4.02