UNIVERSIDAD
DE GRANADA

Doctoral Programme in Information and Communication Technologies

PhD Dissertation

Uncovering the relationship between mood and sport performance using context-aware mobile sensing

presented by

Carlos Bailon Romacho

under the supervision of

Prof. Dr. Oresti Baños
Prof. Dr. Miguel Damas

January 20, 2022

Editor: Universidad de Granada. Tesis Doctorales
Autor: Carlos Bailón Romacho
ISBN: 978-84-1117-244-8
URI: http://hdl.handle.net/10481/73157

Acknowledgments

They say that the actual value of an experience relies on those who walk with you, and I have been so lucky to go through this road with the best people ever.

First of all, I wish to thank Miguel, Oresti, and Héctor. You have been more than supervisors to me; you have uncovered the research career to me and have taught me how to walk through it. Thanks to you, I have grown as a researcher and as a person, and I could never be grateful enough to you for having a bet on me and supporting me all the time.

I would also like to thank all those people I have had the fortune to work with during these years. To the psychology team, Daniel, Pandelis, and Carmen, thank you for placing your trust in me and giving rise to this amazing collaboration, which I hope to continue leading to many results in the future. To the ATC department, thank you for the beautiful reception you gave me, and the incredible team spirit and excellent atmosphere always present. Of course, to Julio, one of the best persons I have ever met because without your help, I would not be in this place today. To my colleagues Manolo and Ofelia, for those recovery coffee breaks at 11. And to my colleagues and supervisors at the UT, Kostas, Miriam, Bert-Jan, and Hermie, for making my research stay at Enschede unforgettable.

Likewise, I would like to say a few words to two very significant people, Lorenzo and Pedro, my high school technology teachers. Thank you for instilling your love for technology and teaching. I have always known that the career I have made so far started with you.

And last but not least, I would like to give a special thanks to the most important people in my life. To my family, dad, mom, Sergio, Hugo, thank you for believing in me, being my strongest support, and always giving me your wise advice. To my friends (or rather, brothers), thank you for your trust, for always accompanying me, and for being the best family I could have ever dreamed with. To my training group, thank you for all those days running, traveling, and growing together. And finally, to Salva, my brother in arms. I can barely remember how long we have been walking together, and I know that I could not have had a better mate during this adventure.

I do not know which new paths the future will bring to me, but it is clear to me that this time will always be engraved in my mind. Thanks for everything.

Carlos

Agradecimientos

Dicen que el verdadero valor de una experiencia está en las personas que te acompañan, y yo he tenido la enorme suerte de recorrer este camino con la mejor compañía posible.

En primer lugar, quiero dar las gracias a Miguel, Oresti y Héctor. Habéis sido mucho más que mis directores, me habéis descubierto el mundo de la investigación y me habéis enseñado a caminar por él. Gracias a vosotros he crecido como investigador y como persona, y nunca os podré agradecer lo suficiente el haber apostado por mí y haberme apoyado en todo momento.

También me gustaría agradecer su trabajo a todas las personas con las que he tenido la suerte de colaborar durante estos años. Al equipo de psicología, Daniel, Pandelis y Carmen, gracias por confiar en mí y dar lugar a esta maravillosa sinergia que espero siga dando muchos frutos en el futuro. Al departamento ATC, gracias por la fantástica acogida que he recibido desde el principio, y por el increíble compañerismo y buen ambiente presente en todo momento. Por supuesto, a Julio, una de las mejores personas que he conocido, porque sin tu ayuda no estaría en el lugar en el que estoy hoy. A mis compañeros Manolo y Ofelia, por esos cafelitos a las 11 que daban la vida. Y a mis compañeros y supervisores de la UT, Kostas, Miriam, Bert-Jan y Hermie, por hacer de mi estancia en Enschede algo inolvidable.

Igualmente, me gustaría dedicar unas palabras a dos personas muy importantes, Lorenzo y Pedro, mis profesores de tecnología del instituto. Gracias por transmitirme vuestro amor por la tecnología y por la docencia. Siempre he tenido claro que con vosotros comenzó el camino que he recorrido hasta llegar aquí.

Y para terminar, quiero dar un agradecimiento muy especial a las personas más importantes que tengo en la vida. A mi familia, papá, mamá, Sergio, Hugo, gracias por creer en mí, por ser mi mayor apoyo y por darme siempre

vuestros sabios consejos. A mis amigos (o mejor dicho, hermanos), gracias por vuestra confianza, por acompañarme siempre, y por ser la mejor familia con la que podría soñar. A mi grupo de entrenamiento, gracias por todos los días corriendo, viajando y creciendo juntos. Y por último a Salva, mi compañero de batallas. Me cuesta recordar cuantos años hace que seguimos el mismo camino, y tengo claro que no podría haber tenido alguien mejor a mi lado durante esta aventura.

No sé que nuevos caminos me deparará el futuro, pero si algo tengo claro es que esta etapa quedará grabada para siempre en mi memoria. Gracias por tanto.

Carlos

Contents

Abstract

Understanding how the changes in our mood have an impact on the events and experiences that shape our daily lives is an essential task in affective science. This fact becomes even more critical in the area of sports since mood is widely recognized as a contributing factor to determining the performance of athletes. As mood fluctuations are strongly influenced by the context surrounding us, identifying the situations and behaviors that trigger these mood changes is crucial to optimizing the performance of athletes during practices and competitions.

Traditional research in this field has focused on studying the mood of athletes at specific points in time, primarily within controlled settings, and exploring its relationship with a specific sports outcome. However, little research has been oriented towards exploring the mood variations during the athletes' daily lives, which is even more critical due to the dynamic nature of mood. This lack of longitudinal studies is partly caused by the limitations of existing data collection technologies and data analysis methodologies. In recent years, the emergence of mobile devices has enormously facilitated the task of continuously monitoring mood and context in free-living environments. However, there are still numerous challenges to be overcome when representing and analyzing an individual's daily life and extracting valuable knowledge from that information.

In this thesis, the use of mobile sensing and data science to explore the in-context mood fluctuations and their relationship with the performance of elite athletes is investigated. To that end, novel data collection systems and longitudinal analysis methodologies are proposed to overcome the limitations of traditional research approaches.

In particular, a mobile-based monitoring platform is developed to continuously capture people's mood and context during their daily lives. The

platform, which is tested and validated, is employed to gather affective, behavioral, and contextual data in two longitudinal experiments. The primary considerations and issues related to the development of these experiments during extended periods in free-living environments are gathered and discussed, contributing to the future design of more effective longitudinal data collection studies. Moreover, an innovative methodology for analyzing longitudinal behavior-related data is proposed to provide a new approach for analyzing the between-person differences in affective behavior and the within-person fluctuations of mood based on the surrounding context. This approach proves to be effective in analyzing how the mood of athletes behaves during their daily lives and leads to preliminary conclusions about the effect of these mood fluctuations on the performance of the studied population of athletes.

Through these contributions, this thesis aims to provide a complete picture of the process followed from the data collection to the extraction of interpretable information and offers a new perspective on how mood, behavior, and context can be analyzed together. Furthermore, its potential uncovers future research directions such as integrating new personal sensing devices for the unobtrusive collection of behavioral and contextual data, or extending longitudinal data analysis approaches to account for the temporal dimension of mood fluctuations.

Resumen

Comprender cómo los cambios en nuestro estado de ánimo afectan a los eventos y experiencias de nuestra vida es una tarea indispensable en la investigación emocional. Este hecho se hace aún más evidente en el ámbito deportivo, ya que está demostrado que el estado de ánimo es un factor que contribuye enormemente al rendimiento de los deportistas. Las variaciones en el estado de ánimo están muy influenciadas por el entorno o contexto que nos rodea, por lo que identificar las situaciones y comportamientos que desencadenan dichas variaciones es crucial para optimizar el rendimiento de los deportistas en sus entrenamientos y competiciones.

La investigación en este campo se ha enfocado tradicionalmente en estudiar el estado de ánimo de los deportistas en momentos específicos, principalmente en entornos controlados, y explorar cómo se relaciona dicho estado de ánimo con un resultado deportivo específico. Sin embargo, existe poca investigación orientada a explorar las variaciones en el estado de ánimo de los deportistas durante su vida diaria, lo cual es aún más importante dada la naturaleza dinámica de las emociones. La falta de estudios longitudinales en este campo está en parte motivada por las limitaciones de las tecnologías de captura de datos y metodologías de análisis actuales. El uso extendido de dispositivos móviles durante los últimos años ha facilitado enormemente la monitorización continua del estado de ánimo y el contexto en entornos no controlados. Sin embargo, a pesar de estos avances, aún quedan numerosas dificultades que resolver a la hora de representar y analizar la vida diaria de las personas y, sobre todo, de extraer conocimiento interpretable de dicha información.

En esta tesis se investiga el uso de tecnologías móviles y ciencia de datos para explorar las variaciones del estado de ánimo dentro del contexto de cada persona, y su relación con el rendimiento de los deportistas de élite. Para ello, se proponen novedosos sistemas de captura de datos y metodologías

de análisis longitudinal que superan las limitaciones de las estrategias de investigación tradicionales.

En concreto, se desarrolla una plataforma de monitorización basada en smartphones para recoger de forma continua datos de estado de ánimo y contexto durante la vida diaria de las personas. La plataforma es utilizada para capturar datos emocionales, comportamentales y contextuales en dos experimentos longitudinales. Como resultado, se discuten las principales consideraciones metodológicas relacionadas con la realización de dichos experimentos durante periodos de tiempo prolongados, contribuyendo así al futuro diseño de estudios longitudinales más efectivos. Además, en esta tesis se propone una innovadora metodología de análisis de datos longitudinales relacionados con el comportamiento. Esta metodología supone un nuevo enfoque para analizar las diferencias entre el comportamiento emocional de diferentes sujetos, así como las fluctuaciones individuales del estado de ánimo en base al contexto. Se comprueba cómo la metodología es capaz de analizar cómo el estado de ánimo de los deportistas se modula durante su vida diaria, y permite obtener conclusiones preliminares acerca del efecto de dichas fluctuaciones en el rendimiento de los deportistas estudiados.

Mediante estas contribuciones, esta tesis pretende proporcionar una visión global del proceso seguido desde la recogida de datos hasta la extracción de información interpretable. Asimismo, ofrece un nuevo punto de vista sobre cómo el estado de ánimo, el comportamiento y el contexto pueden ser analizados de forma conjunta. El potencial de este trabajo da lugar a varias líneas de investigación, como son la integración de nuevos dispositivos de monitorización para la captura de datos de comportamiento y contexto, o la extensión de las metodologías de análisis longitudinal para considerar la dimensión temporal de las variaciones en el estado de ánimo.

1 Introduction

> *Sprinting isn't just a physical sport. It's a mental game, too.*
>
> — **Usain Bolt**

1.1 Thesis goal

The goal of this thesis is to investigate the use of mobile sensing, data science, and context-aware computing to explore how people's mood fluctuates over time and how it is related to the performance of elite athletes. In particular, this thesis focuses on the leading technological and practical challenges related to the unobtrusive and continuous collection of affective and contextual data in the real world, and the extraction of interpretable knowledge that preserves the dynamic nature of such longitudinal data. In this respect, this work aims to contribute to the development of methods to help researchers achieve a deeper understanding of the dynamics of affective behavior, and design more advanced and individualized intervention strategies to optimize the performance of elite athletes.

1.2 Mood, context and sport performance

Affective states are an inherent part of our daily life. They provide our emotional context, shaping the background of everything we experience and do (Davidson, 1994). Furthermore, they are not just the scenario or backdrop where our emotional history takes place but they play an active role in determining the cognition content—how we think or remember—and the cognitive processes—how we reason. The study of affective states presents a considerable challenge for researchers who, during the last decades, have made progress towards the definition of the multiple constructs or terms that constitute the affective domain. In this context, it is important to differentiate between two of these constructs: *emotion* and *mood*. Emotions are defined as psychological experiences or feelings elicited by something—a person, an event, or a thing—, which have a specific trigger or target. In contrast, moods are defined as diffuse affective states which typically last longer than emotions and whose origin may not be clear. While emotions follow their eliciting stimuli closely, the mood is usually temporally remote from its cause. For example, a person can wake up in a sad mood in the morning as a result of a confrontation the previous week (Ekkekakis, 2012).

A fundamental characteristic of mood is that it continuously changes over time. In fact, the time-dynamic nature of mood is thought to be one of the primary reasons why people have affective experiences (Kuppens et al., 2010). Our daily lives are characterized by affective ups and downs, changes, and fluctuations following the events and situations we experience (Rüegger et al., 2020; Servia-Rodríguez et al., 2017). These mood fluctuations are strongly influenced by the behavior of the individuals—e.g., sleep patterns, physical activity, and social contact—and the situations which they find themselves in—e.g., weather, ambient noise, and location—, which is also known as their *context* (Dey, 2001; Parkinson, 1996). Therefore, identifying the daily life events, situations, or behaviors that could lead to a change in the individual's mood can help us not only to model the mood fluctuations over time but also to understand the source of these changes (Martinent et al., 2018).

Within the context of sport, the mood of an athlete is directly involved in practices and competitions and it is widely recognized to be a contributing factor in determining their performance and how it varies over time (Doron and Gaudreau, 2014; Lazarus, 2000). Athletes commonly experience fluctuations in their performance which may not be due exclusively to physical causes but also emotional ones. Traditionally, positive moods have been related to increased performance outcomes, while negative ones have been associated with malfunctioning and decreased performance (Rathschlag and Memmert, 2013). However, the mechanisms that mediate the impact of emotions on performance have been proved to be more complex and to involve a range of affective characteristics. For example, recent studies have explored the relationship between mood and the physiological responses taking part in the athlete's performance during a competition (Davis and Stenling, 2020) and concluded that high levels of physiological arousal can have both positive and negative effects on the sport performance (Robazza, 2006). Moreover, they suggested that each athlete has an individual range of arousal level in which the optimal performance is achieved (Robazza et al., 2008), which reflects the importance of accounting for the individual differences in mood which lead to optimal performance. For that reason, research into the link between mood and sport performance is gaining momentum. A SportDiscus (EBSCO Industries Inc., 2021) search for the keywords *mood* OR *emotion* in the title from 2010 to 2021 produced more results—1875—that the previous two decades combined—1681 results from 1990 to 2009. This research has traditionally focused on analyzing affective states pre-, during-, and post-competition, however, little research has been oriented towards analyzing the daily life factors that could modulate the performance. Furthermore, most of the studies on this topic are based on experiments carried out within controlled environments rather than in the wild, where several behavioral and contextual factors could modulate the athlete's mood, thus not considering its dynamic nature (Kanning and Schlicht, 2010). That is why researchers are encouraged to perform longitudinal analyses among professional athletes, monitoring the mood fluctuations over time and their effect on the daily sport performance (Davis and Stenling, 2020; Martinent et al., 2018; Nicholls et al., 2009).

1.3 Use of mobile technologies to assess mood and context

In order to reach a better understanding of how the mood fluctuations of athletes during their daily life affect their performance, it is necessary to uncover the behavioral and contextual factors that lead to changes in their mood. Identifying the optimum mood for a good performance, as well as the behaviors and situations that trigger that affective state, is a crucial task to implement emotional management strategies (Terry, 2003). To that end, we need to gather information about the behavior, context, and mood of those individuals repeatedly over time while they are engaged in their daily routines.

The continuous monitoring of all factors which influence each athlete entails a vast challenge, almost unreachable so far. Consequently, most studies and results presented during the last years are restricted to very specific aspects of the emotional behavior and, therefore, they reflect a limited representation of the mechanisms that trigger and modulate mood. Currently, with the recent emergence of mobile technologies, a new breed of intelligent mechanisms have raised to ubiquitously monitor affective processes and daily life context (Gravenhorst et al., 2015; Vales-Alonso et al., 2010). In particular, smartphones are at the leading edge because of their richness in terms of sensors and computing resources and the widespread use among every segment of the population—the number of smartphone users worldwide has grown from 3.668 million in 2016 to 6.378 million in 2021 and is expected to continue growing (Statista, 2021). As we carry the smartphone with us almost everywhere, these devices offer the potential to measure our context and behavior continuously, with minimal effort for the user. Moreover, the use of smartphones prevents individuals from carrying around extra devices that could modify their behavior. Their growing popularity as behavioral data-collection tools is partly due to their capability to both proactively elicit *in-situ* self-reports about specific behaviors or experiences and ubiquitously infer many elements of an individual's environment (Harari et al., 2016).

1.4 Challenges in real-world mood and context monitoring

Despite the widespread use of smartphones as continuous data collection tools, there are significant limitations and challenges to be overcome when representing and analyzing an individual's daily life in free-living environments. To become real-world applicable, mobile sensing (also known as smartphone sensing) must satisfy operational requirements that allow for the appropriate representation of the dynamic nature of mood, behavior, and context. In the following, the main challenges of real-world mood and context monitoring are detailed.

Heterogeneous data-collection mechanisms

Behavioral, contextual, and affective information of an individual can be collected using several techniques. Firstly, due to the diffuse nature of mood, this affective construct is commonly assessed subjectively asking individuals to self-report their mood at different points during their daily life, thus gathering its fluctuations over time (Bolger et al., 2003; Y. S. Yang et al., 2019). On the other hand, although self-reports are also employed to enquire about specific events or behaviors, the monitoring of the context surrounding the individuals during their daily life, as well as their behaviors and routines, is commonly carried out using sensor-based data (Gravenhorst et al., 2015). Off-the-shelf smartphones are presumably the richest devices in terms of sensors within our reach. They routinely record behaviors as sociability and mobility as part of their daily functioning. Using a combination of some of the multiple sensors available on most smartphones—such as microphone, accelerometer, GPS, etc.—, we can capture and describe a wide range of behavioral and contextual aspects, thus representing the individual's lifestyle and its changes over time.

Although both self-reports and sensor data can be obtained from the same source—smartphones—, their characteristics present several differences in terms of sampling rate—self-reports are usually gathered every few hours, while sensor data has a sampling rate of seconds or minutes—, output data types—categorical vs. numerical—, or missing data—while sensor-based data collection is transparent to the user, self-reports need to be proactively sent by the user, which usually increases the amount of missing data depending on the user's adherence to the data collection strategy. This issue has sparked a discussion about whether or not both approaches are alternative or complementary (Dogan et al., 2017; Tams et al., 2014). Although it is widely recognized that combining self-reported and objectively measured data from smartphones may help to better monitor and represent the daily life of an individual, the analysis of these data requires a thorough work of curation, cleaning, and aggregation of the data to appropriately merge them into a single dataset.

Data collection in free-living environments

A significant number of experiments aimed at gathering affective information of athletes have collected data during sessions performed in controlled settings (Davis and Stenling, 2020; Martinent et al., 2018). When assessing mood and context in the wild, we face issues and difficulties related to the collection of data in free-living environments. Although people carry their smartphones most of the time, they are not always in the optimal position or place to obtain relevant sensor data. For example, if the device is placed facing down on a surface or carried in a bag or backpack, the collection of ambient light data will be interrupted. The same situation may lead to incorrect accelerometry or activity recognition values. Similarly, during their daily life, users face several situations where their device has no internet collection available, or it is deliberately turned off, so it may not be able to collect certain data or send it to a remote server. Self-reported data also poses challenges inherent to the need to proactively perform the self-report. Although the smartphone can throw alarms or notifications to remind the user to report the desired information, they may be triggered in situations where the user is unable to answer—work meetings, school, sport practices, etc. Likewise, longitudinal data collection studies involve users' participation during extended periods of time, where their motivation may decrease, leading to anchored self-reports and occasional withdrawals.

Although some strategies can be implemented to solve the issues mentioned above or minimize the distortion caused by them, collecting data in free-living environments implies some degree of uncertainty. This situation forces to assume a certain amount of missing or noisy data. Therefore, it is crucial to design an appropriate data collection protocol and perform a complete pre-processing stage of the data, setting subject's exclusion criteria or estimating missing data when possible.

Background sensing with smartphones

When ubiquitously collecting behavioral and contextual data, it is necessary to deploy a service or application within the user's smartphone, which enables access to the desired sensors, gathers the data, and manages the data storage—either locally or remotely. To avoid biases and reach a reliable representation of the user's daily life, this process needs to be performed without any interaction from the user beyond the app installation. Consequently, these processes must keep running in the background during the regular operation of the smartphone, which involves additional power consumption.

During the last years, the increased use of mobile devices has led to a continuous improvement of the smartphone capabilities in both hardware—sensors, cameras, etc.—and software—operating systems (OSs). Due to these features, the energy requirements of smartphones are constantly increasing, making the battery duration a priority for the device user. For that reason, most smartphone OSs include battery optimization systems that monitor the energy consumption of the running apps and services and block those running in the background which show a high energy demand. This is a significant barrier to overcome when finding a balance between preserving the smartphone's battery life and keepinging the background sensing running. In this regard, native battery optimization systems can usually be disabled by asking the user for permission during the app installation process. However, some smartphone vendors using Android-based OSs implement personalization layers deployed over the OS—e.g., EMUI for Huawei devices, MIUI for Xiaomi devices, or One UI for Samsung devices—and include additional battery optimization mechanisms which cannot be natively accessed. Therefore, one of the major challenges when implementing background sensing solutions with smartphones is to keep the sensing running and prevent the device OS from killing the deployed apps or services.

Usability

Smartphone sensing applications must be designed to be easy to use by ordinary people without technological knowledge. Understandable and straightforward systems are proved to have a better reception than complex and cumbersome products, which increases the subjects' adherence to longitudinal studies and their predisposition to deal with unexpected issues. Therefore, end-applications must minimize the need for specific learning.

In addition, when designing apps for background sensing, it is crucial not to disturb the routine usage of the smartphone. End-users should not notice any difference in the behavior of the smartphone when the app is installed so that they do not modify their normal behavior. That way, a reliable estimation of the behavior and context can be reached. This situation is especially critical in terms of storage, network, and battery usage. The user must not perceive a significant decrease in the device's battery life, slower functioning, or even some app crash. From a designing perspective, these practical issues must be carefully addressed to guarantee an accurate inference of the user behavior from the gathered smartphone sensor data.

Nested structure of data

Once the data have been collected, several challenges can also be faced during the data analysis stage. One of the most important is the proper understanding of the internal structure of longitudinal behavioral, affective, and contextual data, which constitutes a key point when designing an effective analysis methodology to extract valuable knowledge.

One central insight about affective science is that the way in which mood fluctuates across time can be very different from one person to the next. Indeed, there exists a wide variety of patterns of affective changes across individuals. Therefore, the structure of longitudinal data of behavior and mood is inherently nested (Bauer et al., 2020). As several individuals are observed on multiple occasions over time, the evolution of their behavior and mood can be explored from the perspective of a group—between-person level—or a single individual—within-person level. The information obtained by separately analyzing each level can yield complementary insights about different aspects of the behavior dynamics. At the between-person level, studying the differences in behavior and context among all the individuals within a group during a long period uncovers long-term aspects of their physical and affective behavior, such as different affective profiles or behavioral lifestyles (Harari et al., 2016). On a deeper level, the within-person one, exploring the variations in the behavior and context of a single individual leads to a more granular understanding of the day-to-day fluctuations of that particular person, unveiling short-term aspects like the affective reaction to each context or situation (Müller et al., 2020; Yu et al., 2021). For that reason, this multilevel structure must be considered when designing analysis methodologies for longitudinal data of human behavior and affect.

Interpretability of the results

Usually, the complex structure of longitudinal data poses significant challenges when it comes to achieving a thorough understanding of the behavior and mood dynamics. Most of the time, the objective of behavioral research goes beyond the accurate prediction of future outcomes. The knowledge extraction techniques used are not intended to be *black box* models but to provide understandable information. Therefore, the analysis of behavioral and affective data should be aimed to extract highly interpretable information which could be directly applied to social, behavioral, or health interventions (Harari et al., 2016; VanDam and De Palma, 2019). The choice of the analysis techniques must consider that black box modeling is often employed when assumptions of the nature of the underlying system are hard to make, or the complexity of the underlying relationship is extremely high. As a matter of fact, this work aims to shed light on the relationship between multiple real-world aspects, which can only be achieved if the results obtained from the analyses are highly interpretable. For that reason, it is crucial to integrate expert knowledge into the design of data analysis methodologies and to assess the usage of each technique and evaluation metric individually, according to the problem studied.

1.5 Motivation and objectives

In the light of the difficulties and challenges presented in the previous sections, there is an opportunity to apply advanced monitoring and analysis techniques to delve into the relationship between the affective state of athletes and their sport performance. Therefore, this thesis aims to investigate the use of context-aware mobile sensing and data science to explore the underlying factors that modulate the mood of an individual over time and the effect of these fluctuations on the performance of elite athletes. In this way, this work aims at contributing to the design of intelligent monitoring systems for real-world behavioral, contextual, and affective data and the development of analysis strategies that provide a better understanding of the individualized mood and behavior dynamics. The goal of this thesis is based on the following supporting objectives:

Objective 1: Design and develop a smartphone-based automatic monitoring platform to simultaneously collect affective, behavioral, and contextual information in free-living environments.

The wide availability and rich sensing capabilities of smartphones make them the ideal tool for continuous data collection about people's behaviors and lifestyles. These mobile devices open the gate to much more precise and complete monitoring strategies than cross-sectional surveys, enhancing the quality and reliability of the information obtained. However, despite the current use of smartphones in behavioral and affective research in sport contexts, most longitudinal studies only monitor a limited and fixed part of the individual's behavior and context, representing a limited amount of behavioral, affective, and contextual features primarily in controlled settings.

Given the challenges described in Section 1.4, this objective aims to design and develop an automatic monitoring platform based on smartphones, capable of gathering affective, behavioral, and contextual information of athletes during their daily life, using a combination of subjective and objective data collection techniques. To that end, the platform's main usability and performance requirements must be investigated. Moreover, the usability and validity of the system should be evaluated in a real-life scenario to ensure its correct operation.

Objective 2: Conduct a long-term monitorization experiment within a population of elite athletes, and generate a longitudinal dataset with affective, behavioral, and contextual data.

Collecting quality and reliable behavioral and contextual data in free-living environments is one of the main challenges researchers face in many areas of knowledge. Current longitudinal studies that can be found with athlete populations gather performance data during specific sessions within controlled environments, and there is little to no information about their affective behavior between these sessions. Moreover, the availability of open datasets including behavioral, affective, and contextual information gathered in longitudinal experiments is limited, hindering the design and test of novel analysis methodologies.

Through this objective, this thesis aims to design and conduct various longitudinal data collection experiments in a variety of contexts. As a result, this work plans to make several datasets openly available to the scientific community and share the methodological and technical lessons learned from the implementation of such experiments. These results are intended to help researchers expedite the design of longitudinal studies and the subsequent data analysis.

Objective 3: Develop an individualized analysis methodology to explore the differences in affective behavior among athletes, the contextual factors which modulate their mood, and the effect of their mood on their performance.

Although it is known that the affective processes of athletes contribute to determining their performance in both practices and competitions, little knowledge is available about the mechanisms that generate these effects. This may be caused by the lack of effective methodologies to explore longitudinal data at multiple levels—both between-person and within-person—, which also obtain individualized information about the dynamic nature of each athlete's behavior, mood, and performance.

Therefore, one of the main objectives of this work is to develop a novel analysis methodology that, using advanced data science and machine learning techniques, can help to uncover the mood-context-performance relationship. In particular, this objective focuses on the design and implementation of the methodology on the previously generated dataset but preserving a generalized approach capable of extracting interpretable knowledge from data of diverse populations and target behaviors.

1.6 Outline

This thesis is structured in six chapters:

Chapter 1 presents the goal of the thesis, lays out the main characteristics of human mood, behavior, and context, brings up the current evidence of their effect over the sport performance—which constitutes the primary motivation of this research—, describes the main challenges related to this topic and details the objectives to achieve this goal.

Chapter 2 provides an overview of the current research on longitudinal affective and contextual data collection, presenting the fundamental concepts about mood and context assessment. This chapter also describes the principal technologies for developing smartphone-based data collection tools.

Chapter 3 presents the design of an automatic mood and context monitoring platform based on mobile technologies, describes the development of the multiple elements which constitute the system, and provides a validation and usability evaluation based on a pilot study.

Chapter 4 investigates the longitudinal collection of affective, behavioral, and contextual data in free-living environments using the developed monitoring platform. It describes the development of two data collection experiments on different target populations and summarizes the main challenges faced and lessons learned from these experiences.

Chapter 5 researches the analysis of longitudinal behavioral data. In particular, it describes a novel methodology to analyze the between-person and within-person mood variations and explore their relationship with the context surrounding the subjects and the performance of a group of elite athletes.

Finally, Chapter 6 gathers the main achievements, contributions, and final conclusions of this thesis. It also presents the future opportunities on the research line addressed by this work.

2 Fundaments and state of the art

2.1 Mood assessment

The diffuse nature of mood makes it difficult to quantify or even characterize the mood of an individual. Since researchers started to study affective experiences, several techniques have been developed to assess mood and emotions, both objectively and subjectively. A small number of studies have employed psychophysiological measures—e.g., EEG, EMG, skin conductance—to objectively estimate affective responses based on the activation of the autonomic nervous system (Ekkekakis, 2012). However, although some systems use wearable devices, which have the potential to be employed for ambulatory assessment, they are mainly employed in laboratory settings. Designing studies to record psychophysiological parameters in daily life situations can be a challenging task due to the quantity and variety of stimuli experienced in free-living environments (Wac and Tsiourti, 2014).

Moreover, using wearable devices forces the subject to carry external elements such as wristbands or chest straps. The presence of these elements may make subjects aware that they are being monitored and unconsciously modify their behavior to achieve what they think are “better results" in the study, thus hindering the actual relationship between context, behavior, and mood. Therefore, as each person has a subjective perspective of their affective experiences according to the events and situations perceived during their daily lives, the mood is usually assessed through self-reports in longitudinal studies. In this section, the main measures and models used for a subjective evaluation of mood are presented, and their particularities are discussed. Additionally, the existing longitudinal mood data collection techniques based on self-reports are described.

2.1.1 Mood assessment models

In the last years, more than 20 different measures of affective constructs have been used in studies of sport and exercise (Ekkekakis, 2012). When choosing one of them, it is critical to consider both the research goal and the desired response format of the self-reports. In the following, the two main categories of measures are presented and discussed.

Distinct-states measurements

On one side, there are measurements that assess a specific, narrowly defined affective state such as *anger*, *fear* or *happiness*. These measurements only represent a limited portion of the affective domain, allowing for a deeper analysis of that specific facet of affect but not being able to extract generalized conclusions about mood as a whole. For example, if a researcher assesses the effect of an exercise intervention on a set of specific affective states and results in no significant changes, it would be erroneous to conclude that exercise has no effect on the broad domain of mood since more mood components might have been involved.

One of the most popular models to measure specific affective states is the Profile of Mood States—POMS (McNair et al., 1971). It consists of a questionnaire that taps six distinct mood states: tension, depression, anger, vigor, fatigue, and confusion. The item pool of POMS is not intended to reflect the global domain of mood but only to reflect certain states that are considered of interest. Another example is the Multiple Affect Adjective Checklist—MAACL (Zuckerman and Lubin, 1985, 1965). It contains a 132-item pool of seven affect scales: anxiety, depression, hostility, positive affect, and sensation seeking. Despite its frequent use in earlier studies of exercise psychology, its popularity has declined in recent years.

Regarding the response format, singular affects and emotions are usually easier to identify by the subject, so the responses tend to be more transparent and precise. However, as these models evaluate a high number of items, they take longer to administer, making them more convenient for studies where the mood is assessed on specific occasions—such as pre and post sport practices—, rather than longitudinal ones, where they can increase respondent fatigue and reactivity to testing.

Dimensional measurements

On the other side, some measurements assess which is known as *affect dimensions*. These constructs are based on the concept that the global domain of mood can be modeled by several underlying dimensions, which are believed to account for the majority of the variance among distinct mood states. Using these measurements, a global overview of mood can be represented, thus allowing for the extraction of generalized results about an individual's mood. However, they are unable to assess, for example, changes in specific states.

One of the most frequently used models to measure affect dimensions is the Positive and Negative Affect Schedule—PANAS (Watson et al., 1988). This model assesses which are known as the "the two primary dimensions of mood": *positive activation* (PA), and *negative activation* (NA). These dimensions are theorized to be bipolar and orthogonal to each other and account for the majority of the variance among distinct mood states. PA reflects "the extent to which a person feels enthusiastic, active and alert", and NA represents a "general dimension of subjective distress and unpleasurable engagement". It consists of a set of 20 items, 10 for each dimension. Despite its popularity in psychological research, recent analyses have debated the accuracy of the model since the items of PANAS appear to represent a mixture of emotions and moods.

Another commonly used model is the Circumplex Model of Affect (Russell, 1980). It is based on the idea that two orthogonal and bipolar dimensions define the affective space: *affective valence* (V), and *perceived arousal* (A). The valence dimension measures the "self-perceived pleasure of the mood", ranging from *unpleasant* to *pleasant*. The arousal dimension reflects "whether the individual is likely to take action under the mood state"—i.e., the self-perceived level of activation—, and ranges from *quiet* to *active*. The various affective states are defined as combinations of these two basic dimensions in different degrees, resulting in the affective states arranged along the perimeter of the circle defined by the two dimensions (Figure 2.1). States

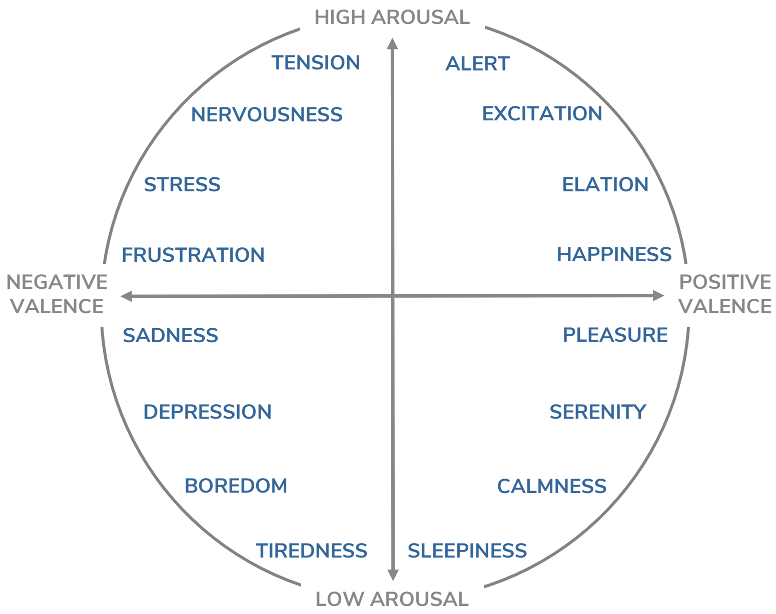


Figure 2.1 Rusell's circumplex model of affect (Russell, 1980). Affective states of the mood domain are built as combinations of valence and arousal dimensions.

which are close together represent similar mixtures of valence and arousal—e.g., happy and pleasant represent high valence and average activation. In contrast, states positioned diametrically from each other differ in terms of one dimension—e.g., tense and tired represent average valence and high and low arousal, respectively. The evaluation measure consists of only two items, one per affect dimension, which can be assessed independently or together using a 2-D response grid—Affect Grid (Russell et al., 1989). Due to its simplicity, this measurement has been widely used in sport psychology exercises.

One last measurement derived from Russell's model is the Self-Assessment Manikin—SAM (Bradley and Lang, 1994). It assesses the two affect dimensions mentioned before—valence and arousal—plus a third one called dominance, D. Although the SAM has been used in a few studies, it is intended to increase the user-friendliness and easiness of the responses. Its peculiarity is that it employs pictures of cartoon characters instead of numerical scales with verbal anchors. The valence scale depicts a character with facial expressions ranging from pleasure—smiling face—to displeasure—frowning face. The arousal scale facial expressions range from—eyes closed—to high activation—

shanking and heart pounding. Finally, the dominance scale depicts a figure ranging from submissiveness—small size—to dominance—large size.

As dimensional measures evaluate a substantially lower number of items than distinct-states ones, they only take a few seconds to administer, thus shortening interruptions of ongoing tasks and minimizing the respondent fatigue. This makes them highly convenient for studies where the mood is repeatedly measured over time to track its fluctuations. On the other hand, affect dimensions are more difficult to interpret by the subject, which could result in less reliable responses.

2.1.2 Longitudinal mood data collection

The measurement of mood and any other behavior through self-reports has been traditionally carried out using what is known as *diary studies*. In this popular research method, participants are asked to fill out self-report questionnaires regarding their mood, behavior, or experiences. Participants commonly answer the questionnaire at a predetermined timeslot or event or discuss the responses later with the researcher(s). Through this technique, diary studies allow for longitudinal data collection in free-living environments, not relying on direct observational methods that may skew the collected data (Bolger et al., 2003). However, this self-reporting strategy entails well-known biases inherent to the act of recalling due to humans' inability to reliability reconstruct past events and the limited reach of the data collection (Ilda et al., 2012). Therefore, the most common drawback in diary studies is that participants must remember and be sufficiently motivated to complete the questionnaires without fabricating the responses to forge the study compliance. Moreover, it can also suffer from *rosy retrospection*, a phenomenon in which participants evaluate past events more positively retrospectively than they did during the actual event (Mitchell et al., 1997).

These limitations raised the need for approaches to proactively elicit self-reports at randomized points in time, which gave rise to the Experience Sampling Method—ESM (Larson and Csikszentmihalyi, 1983). This technique collects data through *in situ* self-reports triggered at various points throughout the day. As participants record their answers in their natural environment, as opposed to in a laboratory or controlled environment, ESM self-reports provide a more accurate representation of the participant's behavior (van Berkel et al., 2018). Therefore, in contrast to punctual data collection methods—such as cross-sectional surveys—, ESM captures "the film rather than a snapshot of daily life reality of patients" (Myin-Germeys et al., 2018). The ESM relies heavily on the concept of the diary study but aims to mitigate some of the mentioned drawbacks by reducing reliance on the participants' ability to accurately reproduce past experiences. It also aims to preserve the *ecological validity* during studies, which is defined as "the occurrence and distribution of stimulus variables in the natural or customary habitat of an individual" (Hormuth, 1986).

Although ESM was initially used on electronic pagers, the increased availability of mobile devices during the last decades has given rise to new possibilities for smartphone-based ESM. Combining these widely available devices with the ESM technique enables the development of more powerful and insightful research probes (van Berkel et al., 2018). In addition, collecting data using smartphones allows researchers to track closely the data provided by the subjects as it is being collected, thus monitoring parameters as the compliance rate or the adherence to the study protocol. The ESM is usually deployed on smartphones through an application that manages the presentation of the self-reports, the data collection, and their delivery to a storage server. The questionnaire is presented on the phone upon receiving a notification, which could be triggered at multiple fixed or randomized time points during the day. The advantages of smartphone-based ESM studies are summarized as follows (van Berkel et al., 2018):

- *Improved data quality through validation.* By monitoring the response time and the response distribution, it is possible to detect a lack of attention when answering. That way, management strategies can be implemented to avoid issues related to paper-based approaches, such as data fabrication after the intended response time.
- *Context-aware responses.* Using the smartphone's sensors, many elements of the subject's behavior and environment can be inferred. These contextual factors are proved to have a substantial effect on self-reports, compromising their validity or even biasing the information (Papini et al., 2020; van Berkel, Goncalves, Koval, et al., 2019; Xie et al., 2019).

Moreover, collecting contextual data together with self-reports enriches the answers' quality by representing the context in which these answers were provided—see Section 2.2 for more information.

- *Real-time study status.* The study data can be received and even analyzed in real-time. That way, researchers can identify and solve possible issues during the course of the study.
- *Advanced question logic.* Digital questionnaires can follow complex logical question flows, where some questions can depend on previous inputs from the subject or the context detected during the presentation of the self-report. Moreover, real-time modifications can be made to the questionnaires to overcome adherence problems.

Among the recent applications of mobile ESM methodology to assess mood, we can find some examples which use questionnaires or short phone calls to assess the participants' mood during their daily lives (Courvoisier et al., 2010; Naragon-Gainey, 2019; Y. S. Yang et al., 2019). Mobile ESM systems that include the capture of context are also present in the literature. For example, MoodPrism (Rickard et al., 2016) uses mobile ESM questionnaires delivered at the smartphone to assess the emotional state and some specific behaviors, along with the collection of data from social networks to infer the mood through text analysis. MoodScope (LiKamWa et al., 2013) and MoodMiner (Ma et al., 2012) have also leveraged the potential of built-in smartphone sensors to objectively capture context in an attempt of predicting mood. Situational features such as location or noise are inferred from the smartphone sensors to contextualize the responses of the ESM questionnaires. Similar and also recent examples can be found in (Asselbergs et al., 2016; Burns et al., 2011).

2.2 Inference of behavior and context from sensor-based data

Similar to the assessment of mood, in traditional procedures for collecting data on behavior and context, researchers typically use retrospective self-reports which ask participants to estimate the frequency or duration of past behaviors or the context in which they have been placed. Questions such as "How many people did you talk to yesterday?"—social behavior—or "How much time did you spend indoors during last week?"—location—are associated with the typical biases and limitations of retrospective self-reporting methodologies (Paulhus and Vazire, 2007). Other methods have focused on recording behaviors in controlled laboratory studies where specific scenarios were presented. Several commentators have criticized the widespread reliance on self-reports and laboratory studies, which cannot fully capture the behavior and context of people's natural lives. For that reason, smartphones have been raised as powerful tools for behavior and context monitoring due to the great range of sensors equipped in off-the-shelf smartphones. They are posed to address the gap in research by allowing for the collection of data in naturalistic environments both objectively and unobtrusively (Harari et al., 2016). Many behavioral and contextual elements such as mobility, sleep, or social contact can be inferred by passively sensing data from the smartphone sensor like the accelerometer, microphone, and Global Positioning System (GPS). The utilization of ubiquitous sensor data to estimate behaviors and context has been described using various terms such as *reality mining* (Eagle and Pentland, 2006), *personal informatics* (Li et al., 2010), *digital phenotyping* (Jain et al., 2015; Torous et al., 2016) and *personal sensing* (Klasnja et al., 2009; Mohr et al., 2017).

The goal of mobile sensing is to convert raw sensor data into meaningful and interpretable information related to behavior and context. Although many approaches can be applied to achieve this goal, one of the most extended ones is the model developed by Mohr et al. (Mohr et al., 2017). In that work, the authors present a layered, hierarchical sensemaking model in which the raw sensor data gathered using smartphones is converted into features that contain high-level information. Figure 2.2 depicts the structure of the context inference model, which is described in the following paragraphs.

Raw sensor data At the lower layer of the model, we find the raw smartphone sensor data. The first data inputs in mobile sensing systems are the unprocessed, raw data collected by the sensors. For the most part, sensor data by itself does not contain enough information to infer context and be-

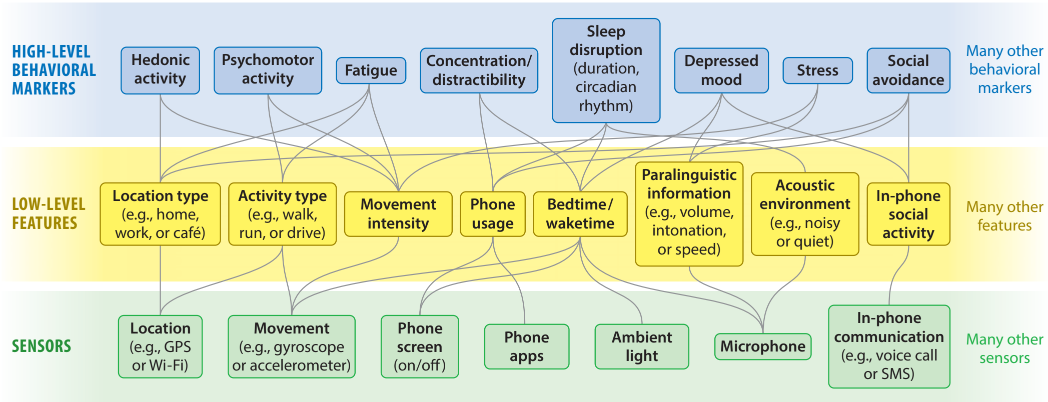


Figure 2.2 Layered, hierarchical context inference framework (Mohr et al., 2017). Green elements represent raw sensor inputs from the smartphone. Yellow elements represent features extracted from these inputs. Blue boxes represent high-level behavioral indicators.

havior, so it has to be further processed and transformed into higher-level data. Some of the most common sensors found in smartphones include the accelerometer, Bluetooth, GPS, light sensor, microphone, proximity sensor, and WiFi. Additionally to physical sensors, researchers also collect information related to smartphone usage, which is also considered a primary data input. These data include call logs, short message service (SMS) logs, app-use logs, notifications received, and battery status.

Low-level features Going up to the middle layer, raw sensor data are combined and transformed into more informative low-level features. This is, arguably, the most critical step in the context inference process (Bengio et al., 2013). These features are constructs that provide information about particular contextual or behavioral elements but are still too proximal to raw data and do not represent a human-readable behavior or context. The low-level features are usually extracted either based on expert knowledge or using algorithms such as slow feature analysis and stacked autoencoders, which automatically discover new feature representations (Vincent et al., 2010; Wiskott and Sejnowski, 2002). However, the first approach increases the interpretability of the features and ensures a closer representation of the subjects' reality. For example, during a data collection experiment, we could have captured raw data of ambient light measured in lux with a sample rate of one minute. Although these values are not informative enough by themselves, they can be used to extract features such as the periods when the individual has been outdoors—in combination with location data—, or the daily bedtime and waketime—based on the long darkness periods detected, along with noise, accelerometer, and phone screen data (Zhenyu Chen et al., 2013). Another example is the raw phone usage data. Although a list of calls and SMS messages does not provide direct behavioral information, relevant features might be extracted, such as the number and duration of outgoing calls and SMS messages, the ratio of missed calls, or the number of unique contacts called. These features will be later used as an input of some knowledge extraction techniques or to infer high-level social contact information.

High-level behavioral indicators Behavioral indicators, also known as behavioral markers, are higher-level features that reflect behaviors or situations and whose representation is close to human-understandable concepts. These features, placed in the upper layer of the model, are derived from the low-level features with similar procedures to the ones used in the previous stage. For example, sleep quality indicators can be extracted from lower-level features of bed/waketime, phone usage, or ambient noise. By combining the low-noise periods with the bed and waketime detected and the long periods with no phone screen activations, several behavioral indicators providing information about sleep can be inferred, such as the sleep duration, the circadian rhythms, or a categorical estimation of the sleep quality—low, average or good (Abdullah et al., 2014). Some studies have even used multimodal sensing schemes—accelerometer, microphone, light, phone usage, and battery state—to infer the sleep stages (Gu et al., 2016; J.-K. Min et al., 2014). In another example, phone usage features like duration, frequency, and regularity of calls, SMS, and social media notifications have also been used to classify a person's contacts into relationship domains—family, friend, or work colleague—to extract behavioral indicators about the amount and type of social contact (J.-K. Min et al., 2013). Similarly, location features such as the number of unique locations visited, together with phone usage features, can be used to infer and classify the individual's location—home, workplace, outdoors, gym, etc. These features have been combined with noise features—detection of silent periods—and physical activity data to infer study durations (Wang et al., 2015). Some researchers have even attempted to extract indicators about mood disorders using a combination of phone usage, noise, and physical activity features (Gravenhorst et al., 2015).

2.3 Mobile sensing frameworks

In order to implement both ESM and mobile sensing strategies, researchers may need to develop their own study-specific sensing, data management, and data analysis tools and methods from scratch. In addition to this, several technical features like resource optimization, sensor configuration, and integration to different operating systems and hardware model—specific smartphone vendors—need to be handled according to the study characteristics. Addressing these challenges requires substantial resources and technical skills, which is costly and time-consuming. To support the development and deployment of personal sensing studies, researchers have designed and built generic and reusable mobile sensing frameworks. They provide configurable modules and plugins which facilitate the development and implementation of the data acquisition, storage, management, and analysis procedures. These frameworks are intended to help researchers design and deploy personal sensing studies with a minimal need for developing and programming the sensing technology at low level. Therefore, the main part of them has typically been released as open-source so that other researchers can adapt them to their own needs and contribute to their development. In the following, the main functional features of mobile sensing frameworks are presented, and the most relevant frameworks used in behavior and context sensing are described.

2.3.1 Features of mobile sensing frameworks

Built-in smartphone sensors

To collect data from the built-in smartphone sensors, these frameworks typically use abstract interfaces with general purpose, often called *probes* (Kumar et al., 2021). The most commonly supported sensors are accelerometer, GPS, gyroscope, proximity, gravity, ambient light, audio, temperature, Bluetooth, and WiFi (Harari et al., 2016). Additionally, these frameworks also support the collection of phone usage data—e.g., calendar events, app usage, call, and SMS log—, which is known as *human-based probes* (Kumar et al., 2021). It is important to note that the support to access the smartphone sensors varies from OS to OS. In particular, Apple iOS implements a more stringent security and privacy policy, which restricts access to users' personal data and specific sensors.

Integration with external sensors

Some mobile sensing frameworks support the connection to external wearable sensors like Fitbit (Fitbit, 2021), which provide additional information about physiological parameters such as heart rate, electrocardiography (ECG), glucose level, or physical activity. This is typically implemented in two ways: through a wireless connection between the smartphone and the sensor—which provides direct data collection—or by retrieving the data from the device's data server through their web APIs (Kumar et al., 2021). The second option facilitates data collection and pre-processing—which can be done at the vendor's server—but does not allow real-time sensor readings access. For example, Fitbit has a 15-minutes delay from the moment the data is collected to the one it becomes available through the API.

ESM support

A high number of frameworks support the collection of self-reports through ESM. It is commonly carried out by triggering a notification whose activation opens the ESM questionnaire. The most usual scheduling approaches to trigger ESM are: (i) fixed points in time, (ii) random/pseudo-random times, (iii) based on contextual events, and (iv) remotely triggered by the researcher. The most common questionnaire-building schemas used to define the content and structure of the questionnaires are JSON (Ferreira et al., 2015) and XML (Hashemian et al., 2012).

Data storage

The data collected can be stored locally and/or on a remote server. The de facto local data storage used is SQLite (Hipp, 2021). Although some frameworks use local data storage, it usually serves as a temporary data buffer until it is sent to a remote server since the storage of high amounts of data may cause performance and storage issues and disturb the normal functioning of the subject's smartphone. Therefore, one essential feature of mobile sensing frameworks is the data offloading technique. The most common strategies are: (i) periodic data synchronization, (ii) event-based data synchronization, and (iii) on-demand data synchronization (Kumar et al., 2021). The most widely used strategy is synchronizing data with the server at regular intervals. When the server or device is unreachable, the data is queued at the local database file for later delivery. Event-based synchronization is popular among frameworks that only support low-frequency sensors or just ESM surveys. The data offloading could be triggered, for example, when the data buffer reaches a limit, when the device connects to a WiFi network, or when the device starts to charge. Finally, triggering data on demand is typically used in frameworks that sample data across devices—e.g., when devices are intended to issue commands to other devices for data upload, exchange, and synchronization.

Context-aware sampling

Some frameworks allow for the customization or adaptation of the data collection based on the subject's current activity (Bardram, 2020; Ferreira et al., 2015). This feature can optimize phone resources, achieve adaptability based on the application demands, and personalize data sampling according to the subject's preferences. For example, sensors with high sampling rates could be turned off based on the user context to reduce battery consumption—e.g., stop heart rate and location measurement if the user is not moving.

Back-end and remote study management

One interesting feature provided by some frameworks is a dashboard or web interface to manage the study configuration and deployment. This includes support for participant recruitment, creation and scheduling of ESM questionnaires, visualization of study progress, and remote configuration of sensors.

Data processing and analysis

Mobile sensing frameworks can also manage data processing. They can provide mechanisms to assess the data quality by warning when the missing data exceeds a threshold or when the sensors malfunction or need to be re-calibrated. On a subsequent step, some frameworks provide tools for data analysis. Behavioral indicators can be extracted in real-time at either the subject's smartphone or the remote server. They could also provide support for offline data processing on their back-end servers with custom plugins and libraries for data analysis.

2.3.2 Most relevant frameworks in research

According to a recent systematic review by Kumar et al. (Kumar et al., 2021), 28 mobile sensing frameworks have been published in scientific peer-reviewed literature. The majority—20—are built for Android OS since it allows for easier and less restrictive access to smartphone sensors. The rest—8—have been built for iOS and Nokia, six of them also supporting Android OS. Table 2.1 summarizes the most relevant frameworks existing in literature along with the features supported by each one. The table only contains those frameworks aimed to serve as application development tools, not those that only provide already implemented technologies for direct study deployment. Also, two of the most widely used ones are described in detail: AWARE (Ferreira et al., 2015) and CARP (Bardram, 2020).

AWARE

AWARE is an open-source mobile context instrumentation framework released in 2015 by Ferreira et al. (Ferreira et al., 2015), available for Android OS—although a version for iOS is also available, it follows a separate development and has a reduced number of features compared with the Android version. It was developed as an extensible and reusable platform for capturing, inferring, and generating context on smartphones. It follows a client-server architecture. The client is deployed as a regular Android mobile application and manages the access to the smartphone sensors, the data collection, and the data synchronization with the remote server. Researchers

Table 2.1 Mobile sensing frameworks for app development. Tick marks indicate whether each framework support the specified feature or not. Operating System—A = Android OS, I = iOS.

Framework	Ref.	OS	Last update	Built-in sensors	External sensors	ESM	Remote data storage	Context-aware sampling	Remote study management	Data analysis
Aware	Ferreira et al., 2015	A, I	2021	✓	✓	✓	✓	✓	✓	
CARP	Bardram, 2020	A, I	2021	✓	✓	✓	✓	✓	✓	✓
Healthopia	C. Min et al., 2011	A	2011	✓	✓			✓		
HealthOS	Lim et al., 2012	A	2012		✓		✓			✓
iEpi	Hashemian et al., 2012	A	2012	✓	✓	✓	✓	✓		✓
Lifestreams	Hsieh et al., 2013	A, I	2014							✓
mCerebrum	Hossain et al., 2017	A	2018	✓	✓	✓	✓	✓		✓
MobiCon	Lee et al., 2012	A	2012	✓	✓			✓		✓
MobiSens	Perera et al., 2014	A	2013	✓			✓	✓		✓
ODK Sensors	Brunette et al., 2012	A	2012	✓	✓					
Ohmage	Tangmunarunkit et al., 2015	A, I	2015	✓		✓	✓	✓	✓	✓
RADAR-base	Ranjan et al., 2018	A	2021	✓	✓	✓	✓		✓	✓
Sensus	Xiong et al., 2016	A, I	2019	✓	✓	✓	✓	✓	✓	✓
SensingKit	Katevas et al., 2014	A, I	2019	✓			✓			✓
TigerAware	Morrison et al., 2018	A, I	2021	✓	✓	✓	✓		✓	
UbiqLog	Rawassizadeh et al., 2013	A	2016	✓	✓		✓			
Zappa	Ruiz-Zafra et al., 2013	A	2015		✓		✓			✓

can use AWARE as a framework for developing their own ad-hoc app or use the generic client app provided by AWARE. At the server side, a back-end with a LAMP architecture—Linux, Apache, MySQL, PHP—can be deployed along with a study configuration dashboard to manage the definition of the study protocol and the data storage.

The client enables the collection of data from all the sensors available at each specific device, as well as human-based sensors. One of its main features is the so-called *plugins*. They are customizable code extensions that can be added to the main app to add extended functionalities such as modified data acquisition protocols, connection to external sensors, and real-time generation of context features from the raw data. For example, the client app can gather data from a Fitbit device by calling the API through a plugin. AWARE also supports an ESM questionnaire-building schema defined in JSON. Diverse ESM response formats can be defined and chained together to support step-by-step questionnaires. Regarding the data storage, the data is stored locally in SQLite databases and replicated to the remote server back-end as JSON objects via HTTP POSTs.

From the server side, a dashboard is available to set up the whole study protocol, add new plugins, and manage the participant's onboarding. In addition, the dashboard allows for the real-time monitoring of the ESM responses received by each participant. The configuration of the client app is also simple. Once the researcher has configured the study protocol at the dashboard, a QR code is generated, which can be scanned by the client app, automatically configuring all the sensors and sampling frequencies transparently to the user.

CARP

The CACHET Research Platform (CARP) is another open-source mobile sensing framework, released in 2020 by Bardram et al. (Bardram, 2020). The platform consists of several elements, which can be used together or independently, the CARP Mobile Sensing Framework (CAMS), the CARP Research Package, the CARP Web Services (CAWS), and the CARP Core. The last element, CARP Core, contains domain models which can be interacted with through application services. Dependencies on specific infrastructures are injected into the application services defined in CARP Core. It constitutes the basis on which the rest of the elements are deployed, defining the elements of the whole CARP system—protocols, studies, deployments, clients, etc. It is implemented in Kotlin, currently targeting Java (JVM) and JavaScript—with matching TypeScript declarations.

CAMS is the programming framework for developing client apps for mobile sensing. It is developed as a Flutter (Google, 2021) package, which constitutes one of the key features of this framework. Flutter is an open-source SDK for mobile app development created by Google. It allows for the development of user interfaces for Android OS, iOS, and web. Therefore, the mobile sensing apps developed with CAMS are cross-platform, thus broadening the target population of research studies. Client apps developed with CAMS can collect data from smartphone sensors and human-based sensors. However, the availability of the sensors may differ from one OS to another depending on the OS native restrictions. Sensor sampling is available through *Flutter packages*, which can be added independently to the client app. The ESM functionality is provided by the CARP Research Package. It is a flutter implementation of the Apple ResearchKit and Android ResearchStack, which provides functionalities related to the obtention of informed consents, and the creation and management of ESM questionnaires, which are defined in JSON. Regarding the data storage, the data is stored locally in a JSON file and supports several data synchronization methods: single data points, batches of data points, or files.

Finally, CAWS is a cloud-based runtime infrastructure for managing research studies. It implements an infrastructure according to the domain model of CARP Core and supports study user management, study protocol configuration, data collection, and the implementation of data analysis pipelines. It provides a range of sub-services to manage user authentication, data synchronization, and deployment of protocols to client apps. Additionally, the framework contains an open-source study management portal, which facilitates study management and data analysis.

3 Development of a smartphone-based platform for mood and context monitoring

3.1 Introduction

The continuous monitoring of mood and context in free-living environments is a considerable challenge that, nowadays, is easier to address thanks to mobile technologies. The development of smartphone-based tools aimed at unobtrusively collecting information about an individual's daily life is a relatively recent research field. However, it is a critical requirement to achieve a deep understanding of the affective, behavioral, and contextual factors that influence other elements such as sport performance. In fact, many inconsistencies found in sport-related studies of affective behavior are due to the low quality of the data collection technologies (Kanning et al., 2013).

Existing approaches for in-the-wild longitudinal data collection often use mobile ESM apps which do not capture the context surrounding the subjects, thus lacking crucial information about the mood fluctuations. Meanwhile, systems including context monitoring usually collect data from a small set of smartphone sensors and do not provide an integrated solution with both the smartphone app and back-end deployment. Moreover, existing works do not provide enough information about the system details so that it can be reused, adapted, or improved by the scientific community. In addition to this, although existing systems have achieved promising results and proved that these tools are technically feasible for sensing mood and context, the data collection tools employed do not allow for modifying the data acquisition methodology while the experiment is running. Since the unobtrusive data collection from the smartphone sensors is transparent to the user, the sensor sampling protocol can be designed based on previous tests. On the contrary, self-reported information requires direct interaction from the user, and it may be necessary to adapt specific settings, such as the content of ESM questionnaires, their scheduling, or the response format, to the particularities of each study. Existing approaches do not allow for modifying these parameters based on the issues encountered during the progress of the study, which could be helpful to improve the participants' compliance or the accuracy of the data acquired.

Given the present challenges of daily life longitudinal data collection, this chapter presents an integrated, multimodal platform for mood and context monitoring in free-living environments. The platform is aimed at providing a comprehensive, easy-to-use tool to perform longitudinal data collection experiments combining ESM and sensor-based data acquisition. The system uses smartphone sensors to unobtrusively capture context data and employs the ESM technique to gather self-reports about the subject's mood. As a novel feature, the platform integrates the real-time flexible management of the ESM questionnaires. In this chapter, the system and its key features are thoroughly described, and a usability analysis is performed based on a pilot study to test its validity and usability.

3.2 Monitoring platform

3.2.1 Platform architecture

To give an overview of the overall design and operation of the monitoring platform, its architecture is presented here. The platform relies on the open-source mobile context instrumentation framework AWARE (Ferreira et al., 2015). In Section 2.3, the most widely used frameworks for developing mobile sensing systems were presented and described. Considering the advantages and drawbacks and the features of each framework, AWARE was chosen to be used as the developing tool for the system. The main reason for this choice is the widespread use of the framework in research projects. Its extensive documentation and active community and development are essential considerations due to the number of issues arising during the development of technical systems. Moreover, according to Table 2.1, AWARE supports all the features of mobile sensing frameworks except the data analysis. However, since we are targeting a specific research objective, an ad-hoc offline data analysis pipeline can be implemented from scratch. On the other side, a key point that could be seen as a drawback is the target OS. AWARE is available for both Android OS and iOS, but their development is almost entirely independent, and the iOS version suffers from all the strong restrictions in terms of sensor access related to that OS. In most situations, Android OS permits third-party apps to access more sensors and phone logs than iOS. In addition, as of June 2021, Android claims 73% of the global market for smartphone devices, suggesting that a more representative sample could be accessed using Android devices. For that reason, the monitoring platform is developed, targeting only Android OS.

Figure 3.1 shows the architecture of the system. It follows a client-server model with four core elements: the (i) user smartphone on the client side, and the (ii) Dashboard, (iii) ESM Management Interface, and (iv) database on the server side. The *user smartphone* constitutes the data source. It uses a client application based on the AWARE framework, which enables and abstracts the data acquisition through the sensors, the management of ESM questionnaires, and the communication with the server. On the server side, the core entity is the *Dashboard*, which includes the server back-end, allowing for communication among all the platform elements. It manages the configuration and data synchronization of the client app, the data storage within the database, and the delivery of the ESM questionnaire files to the client app. In addition, the Dashboard includes a web platform where the researchers can set up the study protocol, the configuration of the smartphone sensors and ESM questionnaires, and manage the participants' onboarding. The second element of the server is the *ESM Management Interface* (EMI), a web-based platform that allows the researchers to configure the settings of the ESM questionnaires—e.g., number of questions, content, response format, scheduling. Even if the study is running, any modification of the ESM questionnaires is automatically reflected on the users' smartphones, thus allowing for real-time adaptation of the ESM protocol. Finally, the *database*

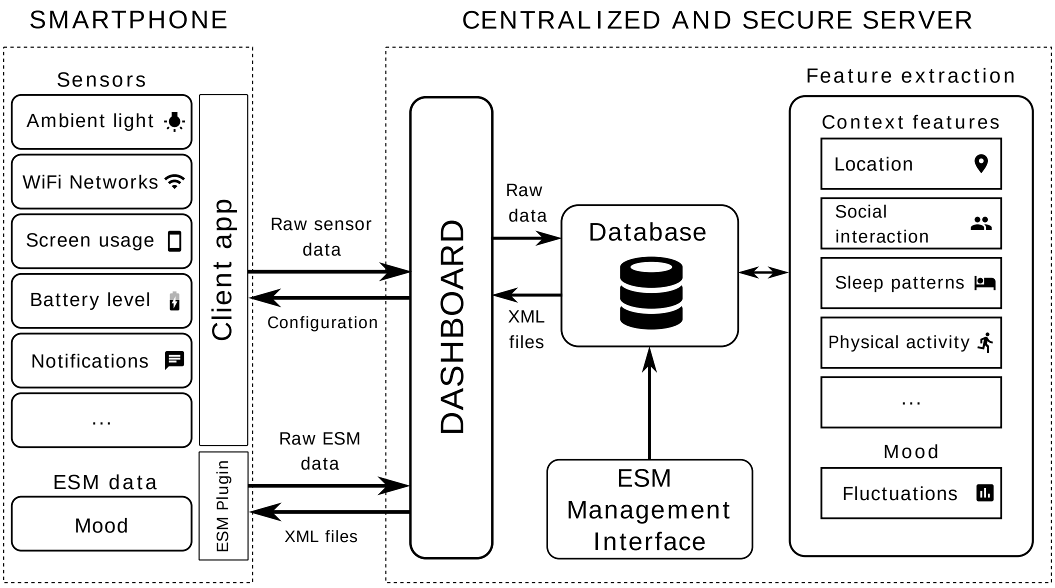


Figure 3.1 Architecture of the smartphone-based monitoring platform.

implements a centralized MySQL database where the data received by the client apps is stored.

The overall functioning of the monitoring platform is described in the following. First, the researcher designs the ESM questionnaires using the ESM Management Interface. The process includes the definition of the question content, the response format, and the trigger schedule. The EMI generates an XML file for each ESM configuration, which is stored in the database. Then, the researcher can configure the study protocol through the Dashboard, setting the active smartphone sensors, sampling frequencies, data synchronization rate, and ESM configuration—by choosing the desired XML file(s). Once the study is configured, the settings for the client app are stored in the database, and the Dashboard generates a QR code linking to that information.

On the client side, the users install the smartphone app on their devices and scan the QR code for the automatic configuration of the sensing protocol. From that moment, the data collection is enabled, and the app starts to store the data locally in a local database deployed on the smartphone storage unit. In the first installation of the app, the app synchronizes with the server a table with information about the hardware of the device and sensors along with a randomly generated Universally Unique IDentifier (UUID), which will be used to label de received data. Then, based on the data synchronization frequency established in the study configuration, the app periodically sends the data to the server and wipes the local database. The data is available in the database for offline data analysis and knowledge extraction.

3.2.2 Smartphone client app

The client app constitutes the front-end of the sensing system. According to the design of the AWARE framework, the client app follows a modular approach. Figure 3.2 shows the interaction among the main components of the application. AWARE supports the addition of customizable code extensions called *plugins* to the client app, whose primary goal is to collect and abstract sensor data to add extra features or process the raw data gathered. In this work, the AWARE core manages the sensor data acquisition, the ESM delivery, and the data storage and synchronization. Then, a plugin called "ESM Flexible Plugin" (Bailon, 2019) has been developed and added to the core app to manage the retrieval of the ESM configuration from the remote server and the building of the ESM questionnaires.

Foreground Service One of the most challenging tasks in mobile sensing is to keep the sensor gathering data in the background with no interaction from the user. There are multiple approaches to run background tasks in Android apps, being the most common ones: (i) WorkManager, (ii) ForegroundService, and (iii) AlarmManager. The WorkManager is the less strict and is recommended for those tasks which can be deferred. The AlarmManager is the most strict, being used when the task must be run at an exact time—such as an alarm clock. Finally, the Foreground Service is a balanced approach: it does not defer the tasks but does not set an exact execution time. Since the sensing task must adhere to the defined sampling frequency, but the starting time does not need to be strict, this app uses the Foreground Service to keep the sensing process running in the background. These services must show a permanent status bar notification so that users are actively aware that the app is performing a background task, which cannot be dismissed until the service is stopped. This app takes advantage of the permanent notification and uses

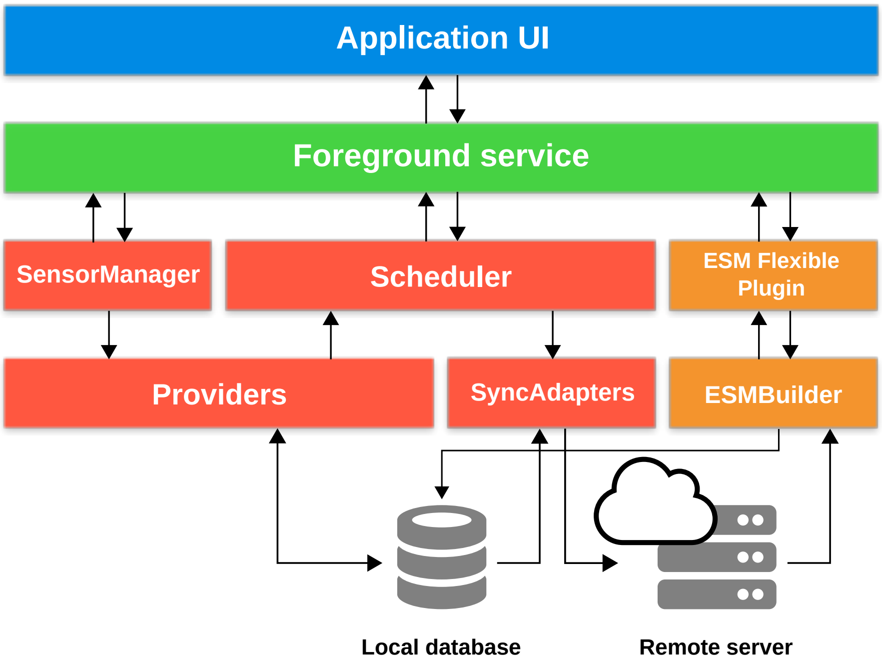


Figure 3.2 Architecture of the smartphone client app. Blue components represent the application UI. Green components represent the application services. Red components encompass the AWARE core features. Orange components constitute the ESM Flexible Plugin.

it to explicitly inform the users that the data collection is active, which is a recommended practice for data transparency (Raento et al., 2009).

SensorManager The Foreground Service manages the sensor data acquisition through the SensorManager. This module provides access to the device's sensors. When the app is configured, the SensorManager activates the desired sensors, stores the sampling frequencies in the database, and initializes the data collection. Then, it implements a listener that waits for new sensor data and sends it to the database according to the specified sampling frequency.

Scheduler All the processes that require to be triggered at a particular time are managed by the Scheduler. This service works dynamically with a database table—`schedules`—to periodically trigger the ESM delivery and the data synchronization and to ensure that the smartphone does not kill the app. At the first run, the app stores a list of those tasks that must be scheduled with the desired trigger time in the database. Then, each minute the Android

OS sends an Intent—Android messaging object to communicate among app components—with an *extra*—key-value pair that carry additional information within Intents—called `ACTION_TIME_TICK`. The Foreground Service filters that intent and starts the Scheduler service when received. This service checks if any task from those available in the database table must be executed and triggers them. Afterward, it reschedules the tasks, updating the database with the timestamp of the last and subsequent executions. As mentioned before, this technique is used to trigger the ESM questionnaires, trigger the data synchronization, and act as a watchdog, periodically restarting the Foreground Service in case the OS has stopped it. The schedules can be set to either a fixed timestamp or a pseudo-random one. Additionally, date constraints can be applied—weekday, month—, as well as contextual ones—for example, trigger when the phone is charging or when connected to a WiFi network. Pseudo-random schedules are provided with a time interval, and a random timestamp within the interval is generated when they are registered in the database. Each time the task is executed and the database updated, the random timestamps are generated again.

Providers An important design aspect that must be considered is data storage. Longitudinal studies with objective context monitoring entail weeks or months of data collection from several sensors. Even if the sampling rates are not particularly high, the large amount of data generated will not be able to be stored in the participant's device without altering its normal functioning. Therefore, the monitoring platform is designed to store data on a remote server. However, in order to reduce the number of data synchronizations and to dodge those situations in which the data could not be sent—unavailable or unstable internet connection, battery saving mode, etc.—, the system has an SQLite database (Hipp, 2021) deployed on the smartphone storage unit, acting as a data buffer. The data is stored in the database through the components called Providers. They are implemented as Android's ContentProviders. Their function is to both store and provide passive access to the data. Each provider has unique content URIs, one per table it stores. They also include the tables' columns and schema, which differ from one table to another. There are three common table columns among all providers: `_id`, an automatically generated database entry ID; `timestamp`, the time instance of the event; and `device_id`, the device UUID. There is one provider per smartphone sensor plus one that gathers information about the device—brand, model, manufacturer, OS version, and SDK level. Table 3.1 shows an example of the structure of the ESM provider.

Table 3.1 Structure of the app's ESM Provider.

Field	Type	Default	Description
`_id`	integer	NULL	Database entry ID, primary key
`timestamp`	real	0	Unix timestamp of the database entry
`device_id`	text		Device UUID
`esm_json`	text		The ESM questionnaire codified in JSON
`esm_status`	integer	0	Codified status of the ESM response
`esm_expiration_threshold`	integer	0	Expiration threshold of the ESM dialog
`esm_notification_timeout`	integer	0	Timeout of the ESM notification
`esm_user_answer_timestamp`	real	0	Unix timestamp of the ESM response
`esm_user_answer`	text		User's answer codified in JSON
`esm_trigger`	text		Element which triggered the ESM

SyncAdapters The synchronization of the data between the local database and the remote server is carried out using services called SyncAdapters. They encapsulate the code for the tasks that transfer data between the device and the server based on the scheduling provided by the Scheduler, which has previously been stored in the database. They operate asynchronously, and, like the Providers, there is one SyncAdapter per database table. As described in previous chapters—see Section 2.3.1—, there are several strategies to synchronize the data between the device and the storage server, with the following three being the most common: (i) periodic sync, (ii) event-based sync, and (iii) on-demand sync. Although event-based synchronization is attractive due to the possibility of synchronizing the data only when the user is, for example, at home with the phone charging, it often requires already processed contextual information. Running data processing algorithms in the own device can be risky since it can slow down the regular smartphone operation, disturbing the user experience. Therefore, the monitoring platform combines the two first data synchronization methods. The data is synchronized periodically with a predefined frequency, but some contextual constraints which do not require data processing are added: the data is only synchronized when the device is connected to a WiFi network, and the battery level is higher than 20%.

ESM Flexible Plugin In order to enable the flexible management of ESM questionnaires, a new AWARE plugin has been developed based on the structure proposed in (Wohlfahrt-Laymann et al., 2019). The ESM Flexible Plugin is initialized by the Foreground Service and enables the construction of the ESM questionnaires with customized response styles, using the configuration defined with the server's ESM Management Interface. This tool generates an XML file with the ESM configuration, which is retrieved using a component called ESMBuilder, which asynchronously gets the configuration files available and deserializes them using the Simple-XML serialization framework for Java ("Simple-XML Serialization framework for Java", 2012). Then, the ESM questionnaires are constructed in three steps. First, one or more `ESM` objects are created with all the parameters from the configuration. Second, one or more `scheduler` objects are created based on the configuration, with their associated ESM questionnaires. Finally, the schedules are stored in the database, ready to be triggered. The search for available configuration files is performed asynchronously each time there is an interaction with the app—the main Activity is executed, for example, when answering a questionnaire—to reduce the number of calls to the back-end. In each execution, the configuration files are deserialized, and the ESM and schedules are built, but the database is only updated if any change has been made to the ESM configuration. That way, the users can receive questionnaire modifications in real-time.

Data privacy Mobile sensing studies are particularly sensitive in terms of data security because smartphone data inherently reveals information about participants' daily lives. Researchers must attend to data security, especially during the collection, transmission, and storage stages of the data. On the client side, the system presented here implements various methods devoted to ensuring the privacy and security of the collected data. First, as the data is only synchronized when the participant is connected to a

WiFi network, it is transferred to the server using secure-sockets-layer (SSL) encryption. Regarding the data identification, when the app is installed on a smartphone, a random 128-bit UUID is generated. All the data collected by the smartphone sensors are only labeled by this UUID. This way, the system meets the requirements of the European General Data Protection Regulation (GDPR).

Example of use When the app is installed, the user has to grant all the permissions related to the smartphone sensors which will be used (Figure 3.3a), plus the permission to ignore Android's native battery optimization system—to keep the background sensing running. Then, the user can scan a QR code (Figure 3.3b) generated by the researcher with the server Dashboard, which includes a link to the configuration of the app—active sensors, active plugins, sampling frequencies, data synchronization frequency, etc. The user must also explicitly grant permission to join the study for the configuration to be applied. To that end, a screen is shown with specific information about the study, researcher contact information, and a text field to insert a participant code (Figure 3.3c). This code may be provided by the researcher in order to link the pseudo-random device UUID with it and relate the user data with other manually collected data—such as demographic or socioeconomic information. Finally, when the user clicks the "JOIN" button, the app is automatically configured, and the data collection starts with no need for further interaction. Figure 3.3d shows the main app screen, which displays the device UUID. This interface can be restricted so that it only displays a text with information about the study and the researcher.

3.2.3 Back-end

The server back-end is the system element that communicates with the client application and manages the data at the server side. It consists of the three major elements outlined in Section 3.2: the Dashboard, the ESM Management Interface, and the database. The back-end follows a modular approach based on Docker (Docker, Inc., 2021) containers. Docker is a platform that automatizes the deployment of applications into isolated environments called containers, providing an abstraction layer that separates the applications

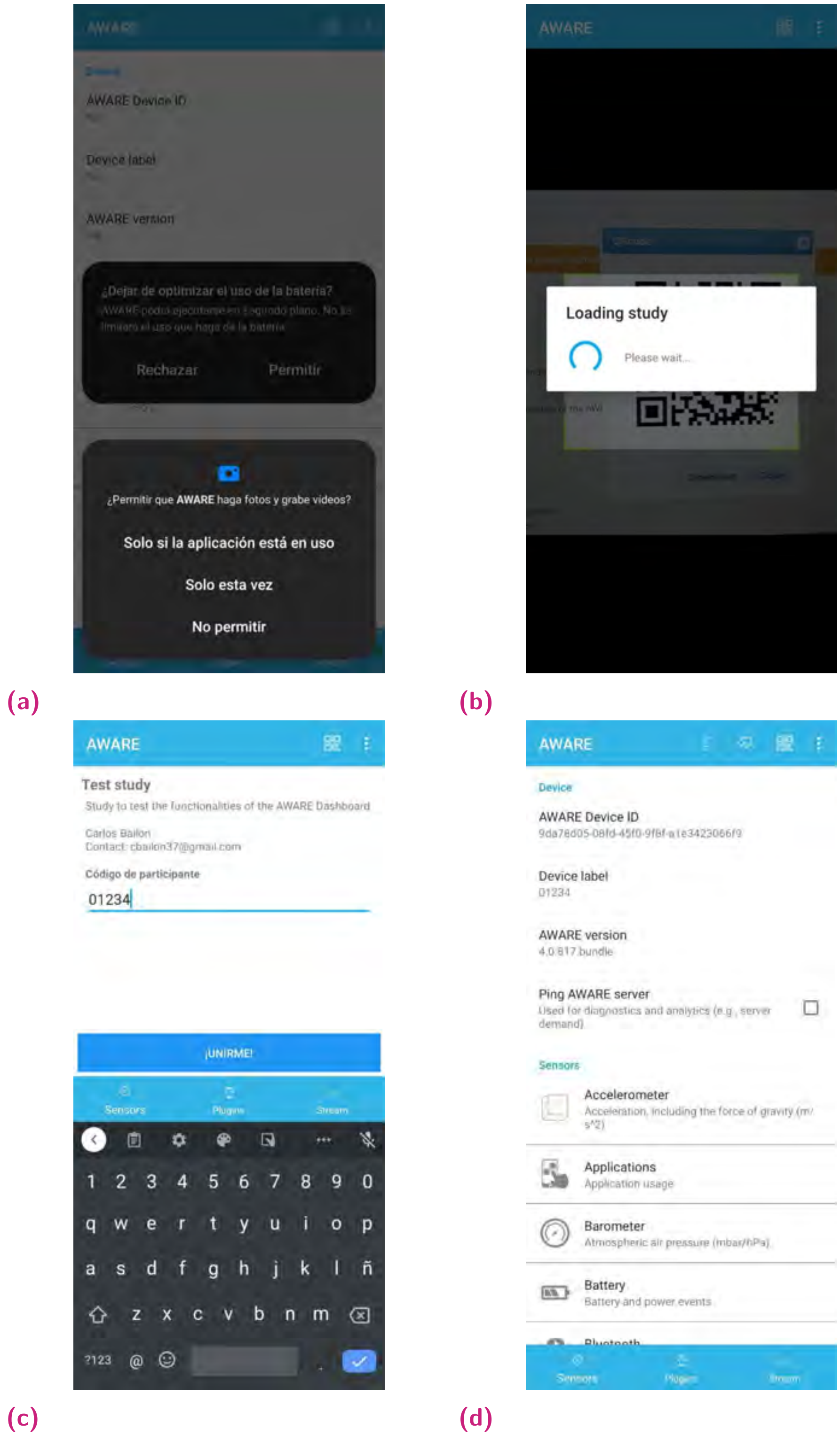


Figure 3.3 Screenshots of the client app: (a) permission request dialog, (b) loading study configuration after QR scan, (c) inserting participant code, (d) non-restricted interface with device ID.

from the host infrastructure. This isolation provides additional security to the deployed applications and facilitates the migration of applications to another host. As opposed to Virtual Machines (VMs), which have a full OS with its own memory management installed and emulate the resources for the guest OS and hypervisor, Docker containers share the host kernel and contain the minimal necessary resources to run the application, thus being more lightweight and faster than VMs.

The platform back-end is deployed following a multi-container architecture (Figure 3.4). The three elements of the back-end are deployed on three independent containers—called *services*—, which are connected among them through a docker network. Networks provide a communication channel between selected containers, which is isolated from the rest of the containers and the host. Using this network, containers can request data from each other through their internal ports—e.g., 3306 for the MySQL service. If desired, ports can be exposed and mapped to the host to access the container from external networks. Additionally, all the containers have SSL encryption to ensure security during the data and request transmission. In the following, the main aspects of the three back-end elements are described.

Dashboard and database The Dashboard runs on top of a LAMP—Linux, Apache, MySQL, PHP—infrastructure, deployed in two docker containers. The first one contains the Apache web server, the PHP dependencies, and the Dashboard web application. The second one contains the MySQL engine and the database.

The web application of the Dashboard is based on CodeIgniter (EllisLab, 2021), an open-source framework for web applications with PHP. It contains the user interface for study configuration and management and a REST API that manages the requests from the client app. Through the Dashboard, the researcher can set the sensor configuration (Figure 3.5b), select ESM questionnaires from the ESM Management Interface if necessary, and receive real-time feedback about the number of sensor data collected (Figure 3.5a), devices connected to the server, and daily ESM questionnaires answered by each device (Figure 3.5c). Multiple study configurations can be defined in the Dashboard, which will generate a QR code for each one. That code provides a link to the configuration and can be scanned by the client app to

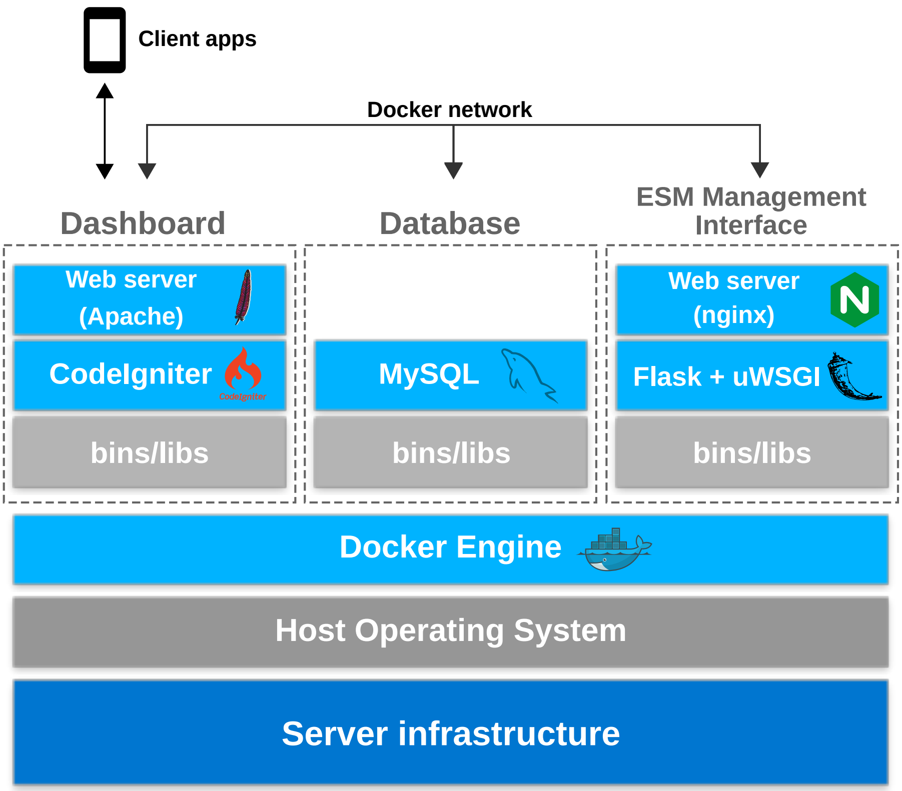


Figure 3.4 Architecture of the server back-end. Three docker containers are deployed, which contain the three core elements of the back-end: Dashboard, database and ESM Management Interface.

automatically configure the device. The access to the Dashboard is secured using OAuth 2.0 protocol and is only accessible to authorized researchers. Each researcher account is independent and allows only to manage its own studies. An additional administrator privilege level can be granted to specific accounts, which allows managing the studies created by all researchers.

The second container includes the MySQL engine and the system databases. There is one database called `aware_dashboard`, which contains information about the Dashboard: accounts, user privileges, available sensors, and studies configuration. The configuration is encoded as a JSON so that it can be downloaded by the client apps. These JSON objects, as well as the incoming data, are sent via HTTPS POSTs. Apart from the database mentioned earlier, a new database is created each time a study is configured to isolate the data of different studies. By default, the databases include two tables: `aware_logs`—log of data synchronization requests—and `aware_device`—information about each device that joined the study, including the device UUID and the label

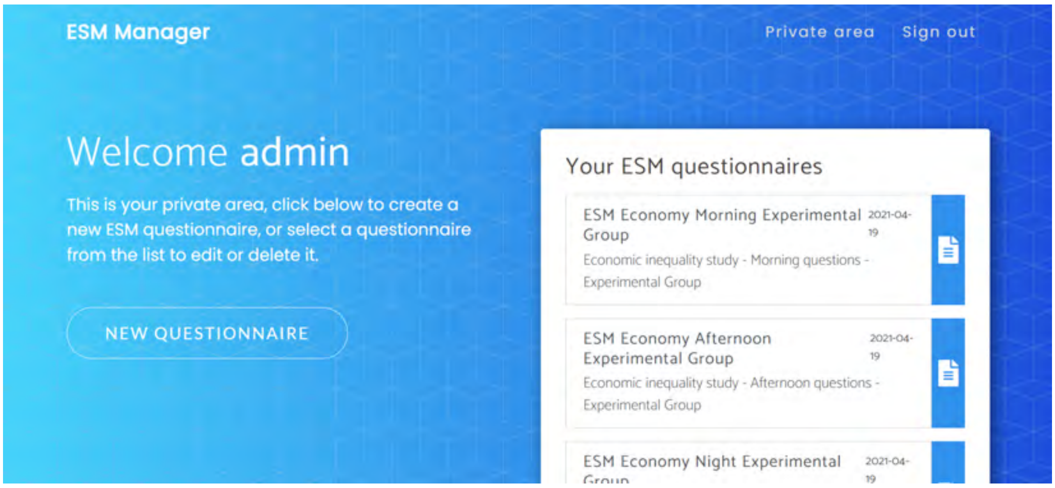
Figure 3.5 Screenshots of the platform Dashboard: (a) real-time count of sensor data received, (b) sensor configuration, (c) real-time daily visualization of ESM questionnaires received from each user.

inserted in the client app. Then, an extra table is added for each sensor activated in the study configuration, which stores the data from that sensor identified by the device UUIDs.

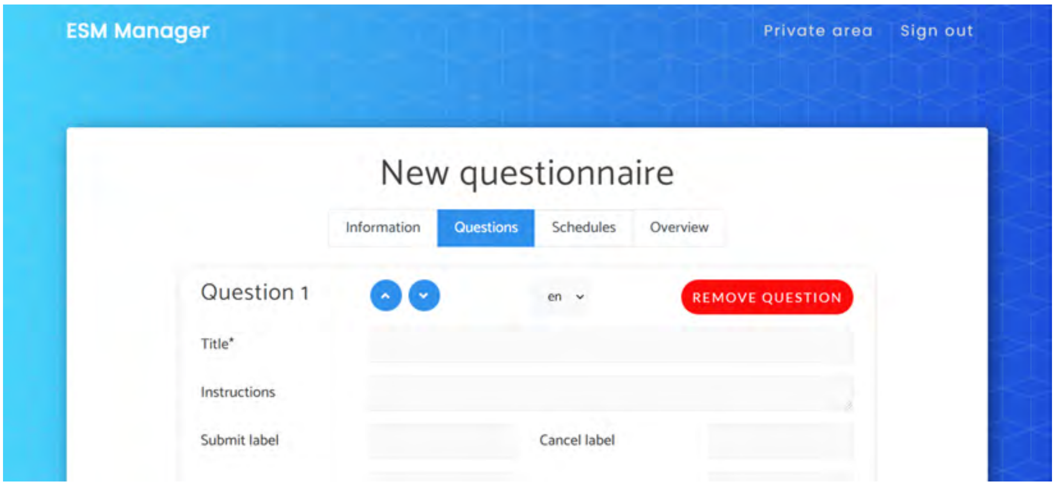
Regarding the sizing of the server resources, Ferreira et al. (Ferreira et al., 2015) evaluated the scalability and performance of AWARE with an increasing number of participants—1 thousand, 10 thousand, 100 thousand, 1 million, 10 million. They measured the server connection delay, using a Python script which mimics the synching mechanisms of the AWARE default client app, inserting 100 thousand data points per device in all sensors' tables. The evaluation machine had 4-core CPUs, 8GB RAM, and 300GB HDD storage. They found that the performance degrades upon 10 million participants. For a higher number of participants, using cloud-computing services—such as Amazon EC or Google Cloud—to run multiple instances of AWARE and do load-balance may be needed to keep the system's performance.

ESM Management Interface The last element of the system back-end is the ESM Management Interface (EMI). This web application is deployed on the third and last container of the Docker structure. It is based on Flask (Pallets Projects, 2021), a web microframework for developing web applications in Python. It is connected to the database and Dashboard through the Docker network and is served using a combination of uWSGI—interface to forward requests between web applications and web servers—and Nginx web server.

The EMI provides secure password-based access for researchers, which can access their private area (Figure 3.6a), where ESM questionnaires can be created, deleted, edited, or viewed. When creating a new ESM questionnaire, a dynamic form built with HTML and JavaScript is presented (Figure 3.6b). First, the form allows for the definition of the questionnaire information—name, description, etc. Second, one or more questions can be added, ordered, or removed. The questions support several locales—i.e., language configurations. If the device's default locale configuration is provided, it will be used when the question is loaded and triggered in the client app. If not, it will be displayed in the first language defined for the question. Additionally, several options can be configured for each question, depending on the response format selected. For example, for checkbox questions, the list of checkbox options must be provided, and for slider-based questions, the slider parameters—minimum, maximum, labels, step, initial value—must be configured. Two extra settings can be defined for each question: the notification



(a)



(b)

Figure 3.6 Screenshots of the ESM Management Interface: (a) real-time count of sensor data received, (b) sensor configuration, (c) real-time daily visualization of ESM questionnaires received from each user.

timeout—time that the questionnaire notification keeps visible before being dismissed by the smartphone—and the expiration threshold—available time to answer the questionnaire after being opened by the user. After configuring the questions, the third step allows for the configuration of the schedules. One or more schedules can be added for each questionnaire, and they can be either fixed—triggers at a fixed time, defined by minute, hour, weekday, and month—or pseudo-random—a specific number of triggers within one or more time intervals, and a minimum time separation between the intervals.

The questionnaire will be triggered at each of these intervals. Finally, an overview of the configuration is provided before storing it in the database.

Once the form has been submitted, the EMI creates an XML file with the questionnaire definition, which is stored in the database. When the study protocol is being configured in the Dashboard, one or more XML files—i.e., questionnaires—can be selected for each study, thus allowing for the trigger of multiple ESM questionnaires with their respective schedules. The questionnaires can be edited at any time within the EMI, which will open the HTML form and retrieve the configuration to pre-populate the form fields. The client app requests the existing questionnaires in each app execution and updates them if any change has been made. That way, the ESM methodology of the study can be modified in real-time, and the changes will be available for all the connected devices. The last option of the EMI is to visualize the XML code for the ESM questionnaire configurations. Listing 3.1 shows an example of an ESM questionnaire defined using the EMI.

```
<ESMDefinition
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <name>Test study configuration</name>
  <short_name>TSC</short_name>
  <description>
    ESM configuration for the test study
  </description>
  <Question>
    <ESM_Type>ESM_QuickAnswer</ESM_Type>
    <Title>Quick answer question</Title>
    <Instructions>Choose an answer</Instructions>
    <Locale>es</Locale>
    <Options>
      <Option>Yes</Option>
      <Option>No</Option>
      <Option>Maybe</Option>
    </Options>
  </Question>
  <Question>
    <ESM_Type>ESM_ScaleImage</ESM_Type>
    <Title>Slider question</Title>
    <Instructions>Move the slider</Instructions>
    <SubmitText>Send</SubmitText>
    <CancelText>Cancel</CancelText>
```

```
    <ScaleOptions>
      <ScaleStartRandom>true</ScaleStartRandom>
      <ScaleStartRandomValues>5</ScaleStartRandomValues>
      <ScaleStep>1</ScaleStep>
      <ScaleMin>0</ScaleMin>
      <ScaleMinLabel>Min</ScaleMinLabel>
      <ScaleMax>100</ScaleMax>
      <ScaleMaxLabel>Max</ScaleMaxLabel>
      <ScaleValueVisible>true</ScaleValueVisible>
    </ScaleOptions>
  </Question>
  <Schedule>
    <id>morning_weekend</id>
    <hour>9</hour>
    <minute>15</minute>
    <weekday>sabado</weekday>
    <weekday>domingo</weekday>
    <weekday>saturday</weekday>
    <weekday>sunday</weekday>
  </Schedule>
</ESMDefinition>
```

Listing 3.1 Example XML configuration file of ESM questionnaires generated by the ESM Management Interface.

3.3 Usability evaluation

To show the potential of the monitoring platform and assess its validity and usability, an evaluation has been made based on a pilot study. This evaluation aims to determine whether the platform is suitable for its use during people's daily lives and discuss whether the data gathered is representative of the participants' affective behavior. The perception of the platform by the end-users and its user-friendliness was also assessed. Furthermore, the results lead to recommendations for future studies conducted with this platform. In this section, the details and results of the pilot study and the usability evaluation are described and discussed.

3.3.1 Pilot study

As the platform is aimed to collect mood data through ESM self-reports and contextual and behavioral information using mobile sensing in free-living environments, the pilot study was designed to simulate these characteristics.

Participants A total of 22 participants—9 males, 13 females, 17-52 years old (M = 22.2, SD = 7.4)—were recruited for the study. All of them were required to have a smartphone with Android OS, which was the device used for data collection. Following the ethics approval from the Ethical Committee of the University of Granada, all the participants were informed about the aims of the study, and they read and signed an informed consent form prior to the beginning of the study. Participants were instructed about the installation of the client app and the procedure for answering ESM questionnaires. No training session for the app use was given since the data gathering and ESM delivery is automatic, and participants do not even need to be aware of the presence of the app so that they preserve their natural behavior. First, they were asked to follow the app installation instructions and scan the QR code for its automatic configuration. After that, they received instructions about disabling the brand-specific battery optimization system—different from the native Android one, which is disabled through a dialog during the app installation. Finally, they were asked to continue with their normal lives and answer the ESM questionnaires when received.

Procedure In the first place, three expert researchers in the psychology field designed the ESM questionnaire for mood assessment using the ESM Management Interface. The questionnaire included two questions to measure the affect dimensions following the Circumplex Model of Affect—see Section 2.1.1—, valence and arousal. Traditionally, these dimensions have been measured using discrete Likert scales. However, given the potential of mobile technologies, a 100-point scale selected with a slider bar was used for this study since it increases the answer's resolution. The valence question showed the text "*¿Cómo de feliz te sientes ahora mismo?*"—i.e., *How happy do you feel right now?*—and its slider ranged from -50 to 50. The arousal question displayed the text "*¿Cómo de activo te sientes ahora mismo?*"—i.e., *How excited* *do you feel right now?*—and its slider ranged from 0 to 100. The notification expiration property was set to 120 min, meaning that participants had 2 hours to open the notification before it was dismissed and disappeared from the notification bar—marking the question as *expired* in the database. This time is decided based on the time interval between successive questions—3 hours—since two answers too close in time may provide redundant information.

Regarding the question schedule, there is no agreement in the literature on the ideal number of notifications, usually ranging from 1 to 10 per day (Y. S. Yang et al., 2019). As it is suggested to gather the minimum required number of samples for obtaining valid data without cluttering participants (Stone et al., 1991; van Berkel et al., 2018), in this study, six questionnaires were delivered per day at pseudo-random times. The use of random or pseudo-random schedules is encouraged in literature (Bolger et al., 2003) to reduce the chance of biased reports. The following six evenly distributed intervals of one hour were selected: 07:00–08:00, 10:00–11:00, 13:00–14:00, 16:00–17:00, 19:00–20:00, 22:00–23:00. During these intervals, a random time for the questionnaire trigger is set in each execution of the client app so that the time is different for each device, thus minimizing the chance of concurrent back-end requests. When it triggers, a notification appears in the device indicating that a new questionnaire is available. Upon its click, the questionnaire window is displayed, and the questions are presented sequentially: once the valence level is reported, the arousal question appears on the screen. Additionally, on the penultimate study day, a reminder was sent as an extra ESM using the ESM Management Interface to notify the participants of the upcoming end of the study and provide instructions to stop and uninstall the application.

After that configuration, the researchers specified the study settings in the platform Dashboard, and the QR code was generated. Five smartphone sensors were activated, whose details are summarized in Table 3.2. The study duration was 14 days.

Validity indicators A number of context indicators regarding the ESM questionnaires were computed to assess the validity of the mood measurements. These indicators have been used in previous works as a measurement to assess the validity of the data acquired (Mcabe et al., 2012; Stone et al., 1991;

Table 3.2 Smartphone sensor data gathered during the pilot study.

Sensor	Sampling rate	Description
Screen state	-	Registers when the screen is locked or unlocked
Network	-	Type of network connected
Ambient light	1 minute	Intensity of ambient light in lux
Battery state	-	Registers when the device starts and ends charging

van Berkel, Goncalves, Koval, et al., 2019). They are measured concerning each questionnaire: the hour of the day and the relative day of study when the questionnaire was answered; the completion time for the questionnaire; the time elapsed between the questionnaire reception and its completion; and the response rate of the questionnaires.

Usability evaluation The goal of the last part of this assessment is to evaluate the platform's usability and check whether it is of interest to researchers in behavior and psychology. To that end, the System Usability Scale—SUS (Brooke, 1996)—was employed. This scale has become an industry-standard used to quantify the users' experience with a system. It consists of a 10-item questionnaire, where answers are scored using a 5-point scale ranging from *strongly disagree* to *strongly agree*. It provides a reliable indicator of the usability of a system and has been tested on a wide variety of works (Sauro, 2011). The SUS questions are the following:

(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.
(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.

At the end of the study, participants were asked to fill a survey with these questions regarding the use of the client app and provide feedback if desired. Similarly, the three experts were asked to provide their impressions about the use of the ESM Management Interface and the Dashboard and evaluate it with the SUS.

3.3.2 Results

ESM response rate The response rate of ESM questionnaires is an indicator of how well the data acquired is representative of the daily life aspects observed. It can be used to assess the validity of the monitoring platform towards assessing participants' mood. As a global measure, during the two-week data collection period, a total of 1848 questionnaires were triggered. Participants completed a total of 1369 questionnaires, leading to an overall response rate of 74.1%. They actively dismissed a total of 11 notifications by removing them from the notification bar—0.6%. The resting 468 questionnaires—25.3%—were automatically dismissed after 2 hours since they were not answered. This indicator needs to be measured individually to evaluate whether the valid data acquired for a particular subject is representative of the evolution of its mood. Figure 3.7 shows the rate of answered, dismissed, and expired questionnaires for each subject. The individual response rate ranges from 27.8%—subject P08—to 97.6%—subject P13—, with a mean value of 82.2% (±16.5%).

The client app triggered six questionnaires at different times during the day. The response rate varies depending on the time when the questionnaire was triggered. Figure 3.8 depicts the overall response rate versus the hour of the day when the notification was triggered. As the ESM schedule was pseudo-random, the hours are grouped in the figure within 3-hour intervals evenly spaced around the triggering limits. The response rate was considerably lower in the first half of the day—between 06:00 and 12:00—, increasing as the day progressed. The highest response rate was achieved between

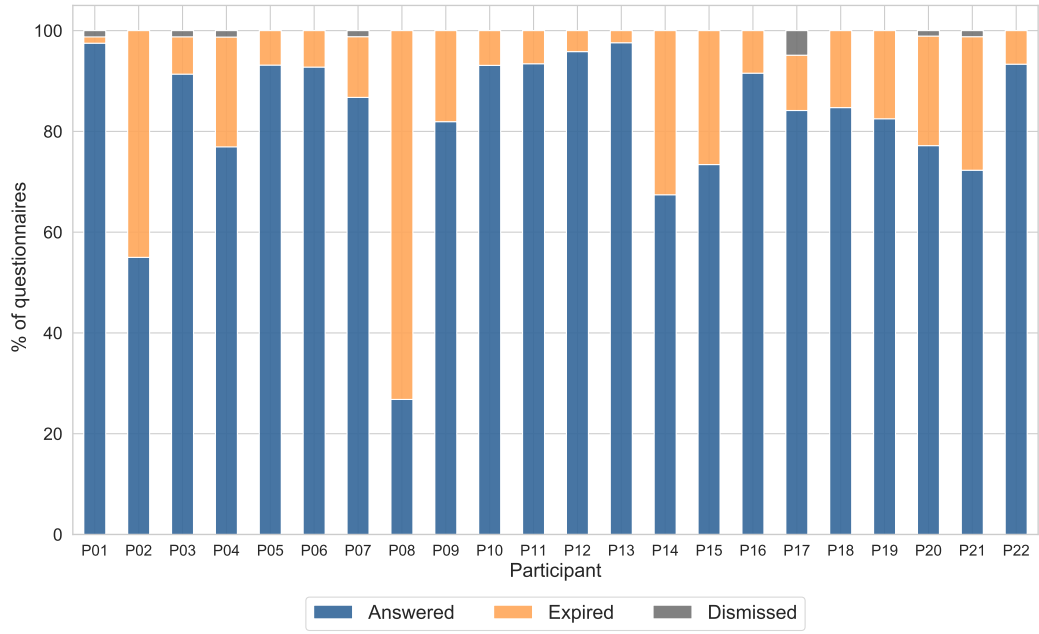


Figure 3.7 Rate of questionnaires answered—blue—, expired—orange—and actively dismissed—grey—for each subject during the whole study period.

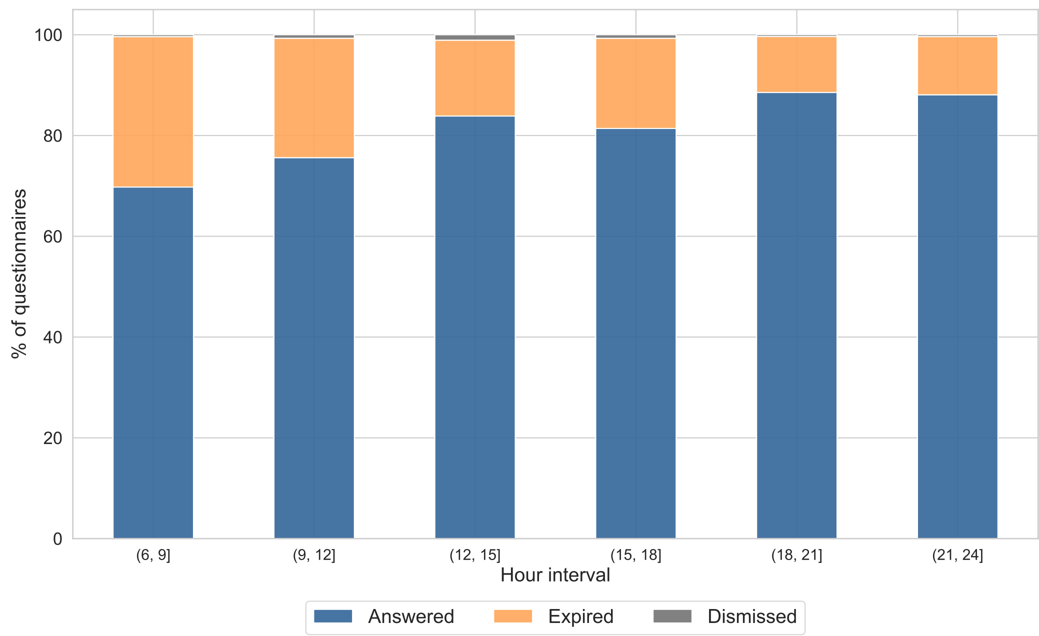


Figure 3.8 Overall rate of questionnaires answered—blue—, expired—orange—and actively dismissed—grey—per interval of daily hours.

18:00 and 00:00—88.1%—, showing an increase of 18.3% with respect to the beginning of the day.

The response rate also experiences daily fluctuations. Figure 3.9 shows the evolution of the overall response rate during the study. The central vertical line emphasizes the distinction between the two weeks of study duration. It can be seen that the response rate keeps significantly high during the first week, experiencing a progressive decrease during the second week. The mean values of response rate during each week are 86.9% and 77.9%, respectively. As mentioned before, on day 13, a reminder of the upcoming study end was sent, which could explain the slight increase of responses on that day, as the participants started again to keep questionnaires in mind.

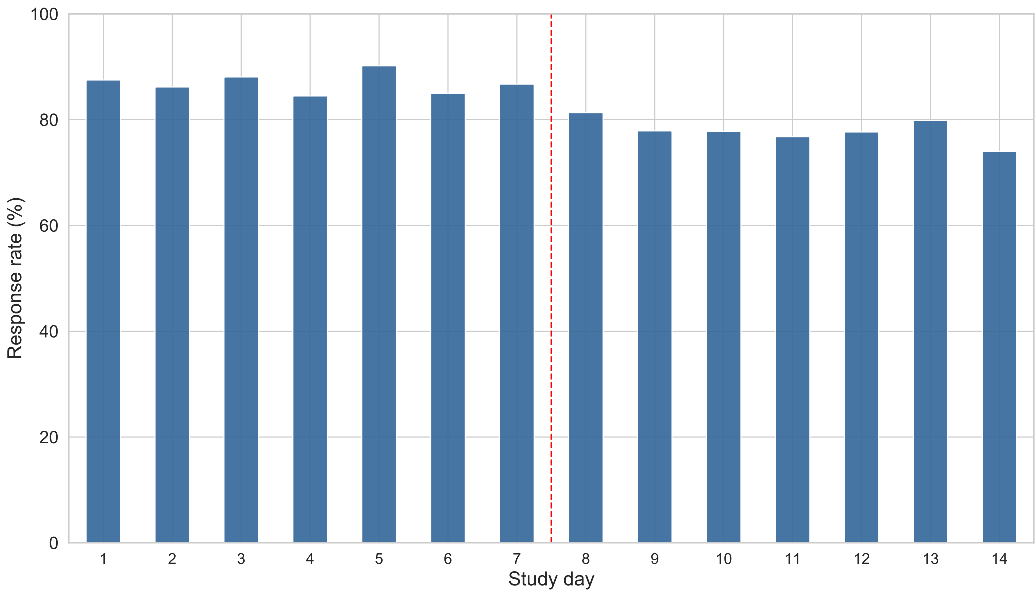


Figure 3.9 Overall response rate registered per day of study. The red dashed vertical line splits the graphic in the two weeks of the study.

ESM completion time The elapsed time between the notification click—questionnaire opening—and the answer submission is an indicator of the attention when answering the question. The average completion time for all questionnaires was 6.95 (±7.53) and 4.87 (±4.17) seconds for the valence—first—and arousal—second—questions, respectively. These values include the time spent by the app to load the questionnaire resources. Therefore, the

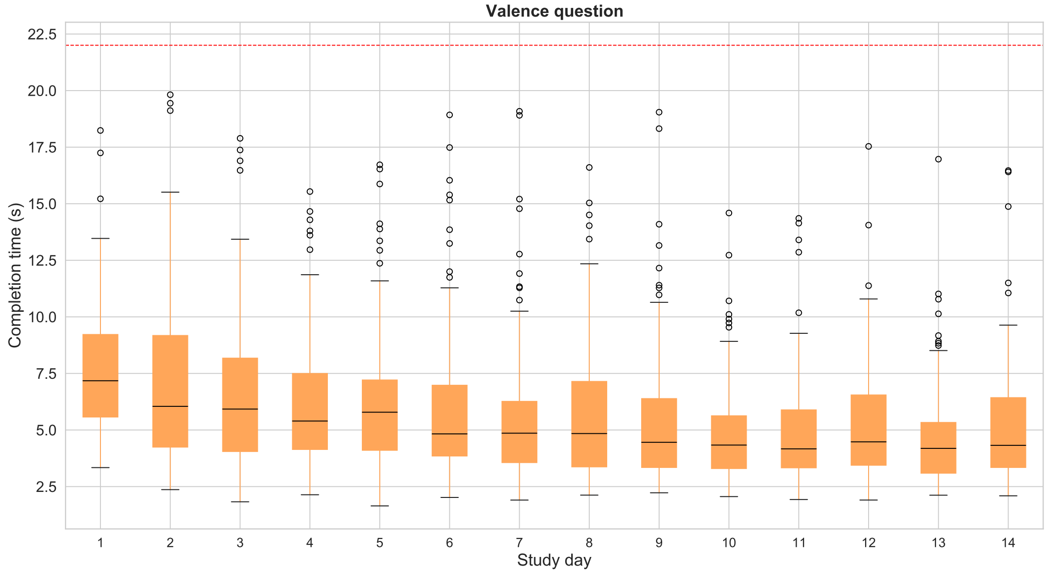


Figure 3.10 Daily completion times of the questionnaires for the valence question. The red dashed line marks the limit of non-valid response times—two standard deviations above the mean.

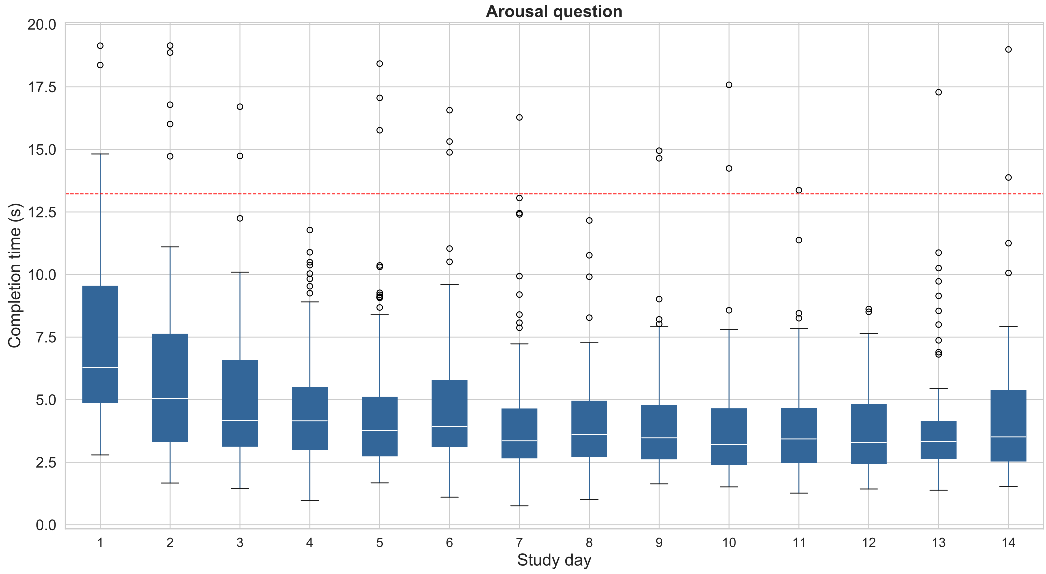


Figure 3.11 Daily completion times of the questionnaires for the arousal question. The red dashed line marks the limit of non-valid response times—two standard deviations above the mean.

absolute times were not examined, but the relative comparison among them. Following the recommendations of van Berkel, Goncalves, Koval, et al., 2019, answers with completion times two standard deviations above the mean are removed since they may be caused by problems experienced when loading the resources or by excessive inattention during the response. Figures 3.10 and 3.11 summarize the daily completion times for each question. This analysis shows two important results: first, it is worth mentioning that the average completion times of the valence question are substantially higher than the ones for the arousal question, presumably due to the order of presentation of the questions. Second, the completion time decreases considerably during the first week, remaining almost constant during the second one.

Time elapsed from ESM notification arrival to response The time elapsed between the reception of the ESM notification and the time when the participant submits the response is also assessed in this evaluation. The ESM notification persisted for 120 min after its arrival, so it could be responded out of the 1-hour triggering interval, thus modifying the time distribution of the measures. By computing this indicator, we can get an idea of the actual answer time of the questionnaires. Only answered and dismissed questions are included in this analysis since expired ones do not have a response time. Figure 3.12 depicts the elapsed times versus the hour of the day when the notification was triggered.

The figure shows that participants took more time to open the questionnaire in the early morning interval—06:00 to 09:00—, with an average elapsed time of 37.2 ($\pm$32.1) min. This time is considerably lower during the rest of the day, reaching its minimum at the end of the day—21:00 to 00:00—, with an average elapsed time of 18.3 ($\pm$25.9) min.

Usability A total of 20 of the 24 subjects completed the usability questionnaire about the client app at the end of the study, and the three experts did the same for the ESM Management Interface and Dashboard. In order to compute the global SUS score for each user, an individual score is given to each item, following the guidelines presented in Brooke, 1996:

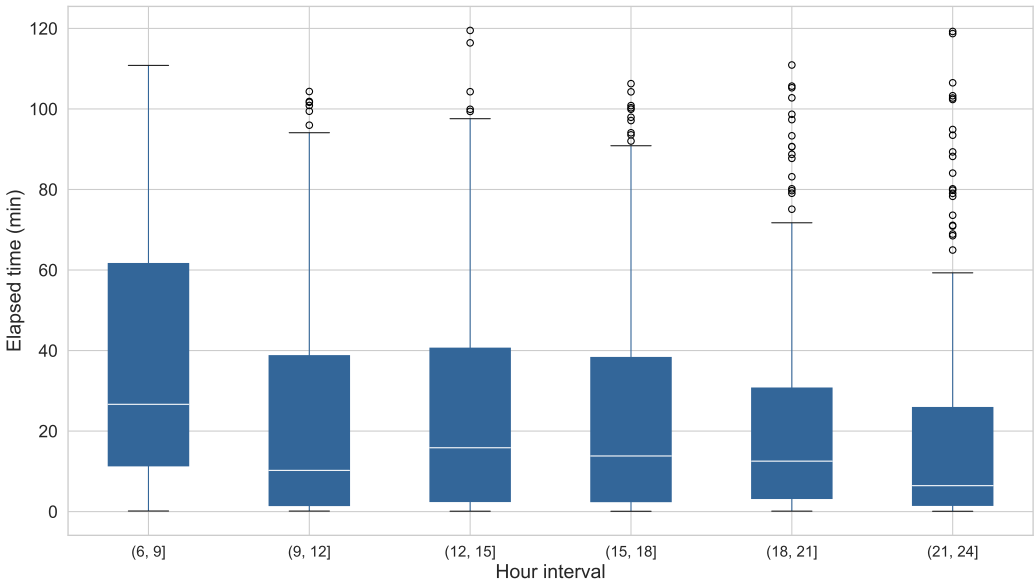


Figure 3.12 Time elapsed from the reception of the ESM notification to the subjects's response per interval of daily hours.

1. For odd-numbered items, the score is computed subtracting 1 to the user response.
2. For even-numbered items, the score is computed subtracting the user response to 5.
3. All the scores obtained—now ranging from 0 to 4, are added and multiplied by 2.5 to obtain the overall SUS score.

The average standard SUS score is 68 (Bangor et al., 2008; Sauro, 2011). A system with a score over 68 is considered to have a good usability level. Moreover, systems exceeding a score of 80.3 are considered to have an excellent usability level. Regarding the client app, the SUS scores obtained are depicted in Figure 3.13. The black dashed line indicates the average value of the 20 SUS scores, and the red one shows the aforementioned good usability threshold. Only two out of twenty ratings are under 68, and the mean SUS score of the app is 84.75, representing a high level of acceptability and ease of use and indicating that the system seems to be highly favorable for the end-users.

Participants were also asked to give voluntary feedback about the app's performance and to report possible issues experienced. First, most of them

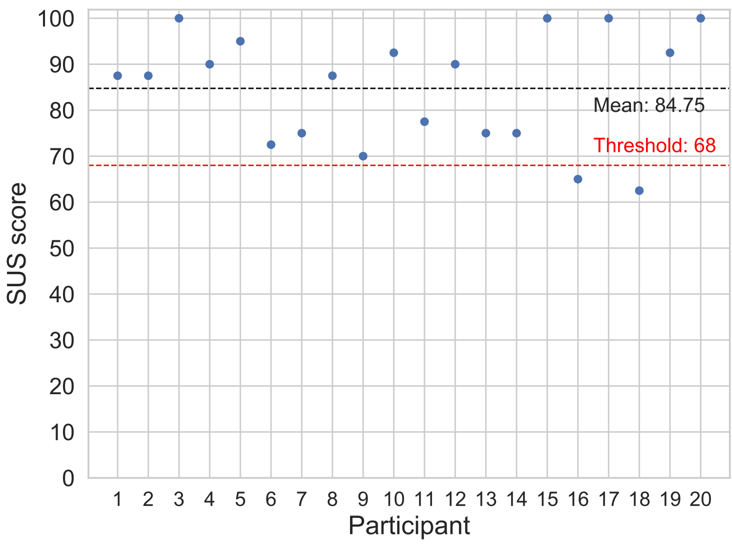


Figure 3.13 System Usability Scale—SUS—scores granted by the pilot study participants to the monitoring platform. The horizontal lines represent the mean SUS score of the system and the threshold value that indicates a good usability.

emphasized the user-friendliness of the ESM questionnaires. They remarked that the response was much less time-consuming than expected: "*The application is very intuitive, it automatically launches the questionnaires and you can answer very easily and quickly*"—P15. Some participants also gave indications about the procedure of responding to the questionnaires, pointing out that both valence and arousal were challenging to identify on such a broad scale, particularly the arousal question. They suggested assessing the mood using scales with more restrictive values—e.g., a Likert scale—as a more straightforward evaluation method. Participant P02 also suggested that "*It could be interesting to access to the questionnaire directly when unlocking the phone*". Although no relevant negative comments were reported, some participants noted a slight increase in battery drainage during the day: "*The only negative point could be that the battery drainage is noticeable even when the phone is not being used*"—P18. Despite this fact, almost all the participants pointed that, although the battery duration seemed to be a little lower, it still allowed for a typical daily operation of the smartphone.

Regarding the ESM Management Interface and Dashboard, the three experts gave a SUS score of 95.2, 87.5, and 95, respectively, leading to an average score of 91.67. This score shows excellent usability of the platform among the experts, which also provided their impressions about the platform usage. They appreciated its high level of customizable settings and the user-friendliness since the process of study and ESM questionnaire configuration is very straightforward: "*The tool offers several options and is very intuitive and easy to use since it guides the process so that the main issues are addressed*". Likewise, the experts were truly impressed with the easy way of making real-time changes to already designed questionnaires. The only negative issue reported is that the ESM Management Interface was not fully optimized for smartphones.

3.3.3 Discussion

Flexible management of ESM questionnaires The monitoring platform developed in this work provides a unified system that integrates all the functionalities of the tools employed in previous experiments on affective research. This platform is placed among the first systems which provide an end-to-end solution with both front-end elements for study participants and back-end tools for researchers. One of the most innovative features of the system is the high level of flexibility achieved with the study configuration. During the course of an experiment, it may be necessary to modify the content or schedule of ESM questionnaires or even send additional questions to maximize the completion rate, capture additional information or keep up the participants' engagement. For example, some researchers point that rewording the questions or modifying their order can increase the response rate and the answer quality, as novel or unfamiliar content has been proved to increase engagement (Attfield et al., 2011; Huang, 2003; Wenemark et al., 2011) and to mitigate anchoring (Gehlbach and Barge, 2012) during longitudinal studies. Some studies have also shown that reminders or motivational messages sent during the course of an experiment can prevent the engagement from dropping (Heron and Smyth, 2010; Naughton et al., 2016). These additional questions can also be sent after a period of missing data, asking participants to recall their mood during the previous time. Therefore, incorporating the ESM Management Interface to the platform supposes a key novelty point of this work. The possibility of remotely configuring and modifying the study protocol is a feature that only 6 of the 28 existing mobile sensing frameworks (Section 2.3.2) include, and any of them support real-time modification of the ESM questionnaires content and schedule (Kumar et al., 2021). Moreover, the opportunity to define multiple study configurations simultaneously provides an exciting feature: using the same app, researchers can test different configurations and modifications during an experiment without any impact on the subjects' apps. They can switch among the different configurations just by scanning new QR codes, thus being able to, for example, test a modification of ESM questionnaires and schedules before applying it to the study subjects.

Impact of the smartphones' battery optimization systems One of the main challenges encountered when developing the client app is the presence of battery and memory optimization systems in smartphones. These systems kill the running processes of those apps that are not frequently opened, a condition met by our app, which in fact is designed to not be opened by the user for other things rather than answering ESM questionnaires. This is a common issue found when implementing background sensing solutions, and since the leading OS vendors are working towards increasing the battery life of their devices, the solution is getting more and more complicated. Regarding the OS native battery optimization, it can be tackled, along with the use of a Foreground Service by explicitly asking the user permission to disable these systems, which is presented as an additional permission request after the permissions to access the sensors. However, the leading smartphone brands deploy personalization layers built over the native OS which extend the smartphone functionalities—e.g., EMUI for Huawei devices, MIUI for Xiaomi devices, or One UI for Samsung devices—, including brand-specific battery and memory optimization systems. These systems cannot be accessed through the application source code, and they must be manually disabled by the user, which entails additional app installation instructions, and the risk of losing data if some user does not disable them and the app gets killed by the OS. An additional aspect to keep in mind is that, despite the disabling of those systems, the app can suffer from periods where the OS disables the primary process, and it does not check for new questionnaires, with the risk of skipping one of them if this disability period coincides with the trigger time. To solve this problem, the client app includes a method to check for past schedules that were not triggered and trigger them despite the possible delay—usually in the order of a few minutes. This includes schedules for ESM questionnaire trigger or for data synchronization with the server.

Methodological recommendations based on the pilot study results The response rate of ESM questionnaires is an indicator of how well the data acquired is representative of the daily life aspects observed. A high response rate indicates that the self-reported information—in this study, the mood—has been captured in a wider variety of scenarios, so it is more likely to be contextually diverse. In contrast, a low response rate indicates a lack of samples, thus not representing faithfully the self-reported information. Although no gold standard has been agreed for acceptable response rate, some studies point to a compliance rate close to 80% to be representative of participants' daily lives (Stone et al., 1991; Y. S. Yang et al., 2019). The response rate of our study was above 77% for 17 out of 22 participants, thus considering the data valid for the majority of the sample. This marker also shows three participants—P02, P08, and P14—with compliance rates of 55%, 26.8%, and 67.4%, respectively. Although these results show that the monitoring platform is suitable for sampling mood through ESM, these three participants should be removed from subsequent analyses.

The response rate varied across the study and during each study day. It was considerably lower during the early hours of the morning—06:00 to 12:00—, experiencing an increase during the day. This result was expected, as people's schedules are usually busier in the morning, so participants can find themselves in situations where they are not aware of the mobile phone or able to use it—for example, at work or lectures. Although this issue is inherent to the use of smartphones as data collection tools, it could be interesting to include some recall questionnaires at particular times during the day using the ESM Management Interface. Despite the decrease of the response rate experienced for the second week of the study—9% with respect to the first week—, it remains close to 80%, so it can be considered that the data is still representative. In fact, a slight increase in the response rate can be noticed one day before the end of the study, which can be attributed to the reminder sent to the participants to inform them about the end of the study the next day. Based on this finding, both the good performance of the

ESM Management Interface and the effectiveness of the reminder approach to keep up the engagement level can be confirmed.

Regarding the completion time, when participants of a study have to answer the same questionnaire several times, they get used to it and gradually “automatize" the response. Literature links extremely low completion times to inattention when responding, and several studies recommend removing from the data those questionnaires with suspiciously fast completion times (Mcabe et al., 2012; van Berkel et al., 2018). In the pilot study, less than 5% of the responses were removed, so the remaining data is still enough to keep a high response rate. Surprisingly, on the first day, the completion time was substantially higher than the rest of the days, presumably due to the novelty of the questionnaires and the difficulty in identifying the valence and arousal levels. The daily average completion time decreases, especially during the first week, remaining constant during the second week due to the habituation effect. This automation process can be dangerous since paying less attention to the responses could decrease the data quality. In that case, reformulating the question text or changing the response format during the study through the ESM Management Interface could help maintain a high attention level. Another suggestion is to perform a learning phase before the beginning of the study, during which participants are taught to correctly answer the questions and, if necessary, to identify psychological aspects such as valence or arousal, thus avoiding these initial higher completion times. It is also worth mentioning the difference between valence and arousal questions. The second one was responded an average of three seconds faster. Further analyses may be required to determine if this finding is a result of the question order or if the arousal level is easier to identify than valence.

Finally, the analysis of the elapsed time from the notification reception to the response shows that, during the early hours of the day, this value is higher, reinforcing the previous affirmations about the response rate, which was lower during the first period of the day. For future studies, it may imply that, even if the participants respond to the questionnaires during the mornings, it may be necessary to design schedules with not evenly distributed intervals since data gathered through ESMs may have less quality.

Usability The results of the usability analysis show that the monitoring platform has good user-friendliness and acceptability among the study participants and experts. The average SUS scores prove that the use of the platform does not require technical skills, making it appropriate for its widespread use among the population and the scientific community. The feedback received has been highly valuable and has led to some suggestions for improvement. For example, it could be interesting to implement the questionnaire opening upon the smartphone unlock, making the response process simpler and faster. It may also help to increase the response rate, which has been explored and tested in previous research (van Berkel, Goncalves, Lovén, et al., 2019). Finally, the battery drainage should be optimized if more sensors are intended to be activated. Additionally, the main contribution to the battery drainage was found to be the data synchronizations (Ferreira et al., 2015). In this regard, using context-aware scheduling may help to reduce this effect. However, in this pilot study, our participants did not report a significant decrease in the battery life or performance of the devices.

Limitations and future work The pilot study was carried out with a sample of participants whose mean age was 22 years old. It may bias the type of situations the participants are involved in when receiving the questionnaires. For example, a significant part of them attended lectures or worked during the mornings, thus not being able to answer the questionnaires. Therefore, a more comprehensive range of ages would be recommended for future studies. It could also be interesting to conduct a more prolonged study for further confirmation of the results obtained. Additionally, although the usability results seem favorable, the platform should also be assessed by a higher number of experts to further confirm the usability levels reached in this study.

3.4 Conclusions

The continuous monitoring of the changes in mood and sport performance over time is a challenging but essential task in emotion research. Traditionally, sport-related studies of affective behavior use cross-sectional, laboratory-based data collection methods in pre- and post-exercise conditions. Existing approaches for longitudinal data collection in free-living environments present certain limitations, such as not considering the context surrounding the subjects or providing unstructured systems that hinder their usage by non-technical researchers. Moreover, the low quality of the employed data collection technologies often leads to inconsistencies and a lack of information.

Based on the present challenges of longitudinal data collection in the wild, and the limitations of the existing systems, in this chapter, an integrated, multimodal platform for mood and context monitoring is presented. The platform uses mobile technologies to collect data in free-living environments combining self-reports and sensor-based data acquisition. The presented platform constitutes an end-to-end solution that implements a client smartphone application for the data collection and a server back-end with various web applications which allow researchers to design, manage, and deploy longitudinal data collection experiments. Furthermore, the platform presents a novel feature that integrates the real-time flexible management of ESM questionnaires. Through this feature, the content and schedule of the ESM questionnaires can be modified or even extended on-the-fly, which helps to mitigate engagement issues or capture additional information.

The validity and usability of the system have been assessed through a pilot study. The results obtained prove the system's feasibility for sampling mood through ESM and collecting sensor-based data without a significant impact on the daily operation of the users' smartphones. The usability scores and the feedback obtained for both client application and back-end show a good level of acceptability and user-friendliness of the platform among researchers and end-users. Additionally, based on the pilot study results, a series of methodological recommendations have been made for the design of future data collection experiments. In particular, in this thesis, the platform is intended to be used to obtain mood and context information among sports-related populations, focusing on out-of-sport situations, a period that is not usually explored. A data collection experiment with such conditions is presented in Chapter 4.

4 Longitudinal collection of mood and context data in free-living environments

4.1 Introduction

The collection of longitudinal behavioral, affective, and contextual data in free-living environments is a considerable challenge, even with the appropriate data collection tools. Long-term experiments require deploying a significant number of resources, both in terms of equipment and time. Recruiting participants, developing and deploying the tools, monitoring the experiment for weeks or months, and curating the data to prepare it for knowledge extraction analyses are demanding tasks that are not always feasible. For that reason, researchers encourage the publication of open longitudinal datasets of mobile sensing data. Using already existing, validated, and curated data avoids duplication of effort, accelerates research, and opens the field to a broader range of knowledge areas beyond digital health (Huckvale et al., 2019). Mobile sensing studies usually generate very rich and heterogeneous datasets that may be analyzed for multiple purposes and whose spreading has a specific value in the standardization of data cleaning and validation pipelines and the replicability of the results. However, the current availability of open datasets, including longitudinal behavioral, affective, and contextual data is scarce, or even inexistent if we look for populations of elite athletes.

Additionally, the design of longitudinal study protocols involves setting multiple methodological parameters, which can vary extremely depending on the research objective, the target population, and the data collection tools. Not every study existing in the literature carries out a good design of the study parameters prior to the data collection, usually leading to inconsistent results which could be hard to replicate. Moreover, these parameters are not always shared along with the results. In light of the current need for longitudinal datasets and standardized methodological data collection designs, in this chapter, two longitudinal data collection experiments are presented based on two different topics. Additionally, the key methodological decisions of longitudinal studies are gathered and discussed.

4.2 Methodological considerations in longitudinal studies

Before running a longitudinal smartphone-based sensing study, there are a series of basic questions about the design that need to be addressed. In the following paragraphs, these key methodological decisions are summarized, and recommendations are made based on the experience of both small and large-scale smartphone-sensing studies found in the literature.

Study duration Longitudinal study designs may span from several hours to months. Their duration is typically dependent on various factors, with a particular focus on the research questions addressed—e.g., interested in daily fluctuations or monthly affective trends. To ensure that a variety of contexts and behaviors of participants' daily life are captured, a minimum duration of 1 week is recommended, since individuals use to keep a weekly routine that covers the majority of their activities (Hektner et al., 2007). No clear guidelines exist in the literature regarding the maximum duration. Early ESM studies find the quality of collected data to deteriorate after a period of 2 to 4 weeks (Stone et al., 1991) due to the potential effect of self-reflection among participants. In (van Berkel, Goncalves, Koval, et al., 2019), a considerable drop in the response quality was found in the third week of the study. In a review by Berkel et al. (van Berkel et al., 2018), the average study duration was found to be 32 days with a standard deviation of 57.1 days—due to several high outliers—and a median value of 14 days. In general, the reported duration of studies is less than one month, which could be a natural consequence of the ESM methodology, trying to reduce the participant burden.

Number of participants The number of participants reported by existing studies differs drastically from one to another, ranging from 1 to 1013 ($\pm$124.6) participants (van Berkel et al., 2018). The mean number of participants is 53, while the median is 19 participants, providing a representative insight into mobile sensing and ESM practices. This number of participants is in line with the local standards in the Human-Computer Interface (HCI) community—sample size with a mode of 12 and a median of 18 (Caine, 2016). Multiple methods are used to determine sample sizes, including power analysis, cost analysis, or usage guidelines—such as local standards. While power analysis is an objective approach to determine sample size, it is based on previous data on the topic, which do not abound in novel technologies like mobile sensing. For that reason, researchers consider it appropriate to guide sample sizes by existing local standards of the HCI community (van Berkel et al., 2018).

Sampling device One of the first critical considerations when designing a smartphone-based study is whether the participants should use their own smartphone or provide them with one. One significant benefit of participants using their own smartphone as the sampling device is that the recruitment can reach a more extensive and diverse sample if the experiment is opened to the general population. For example, the Emotion Sense application (Rachuri et al., 2010) registered thousand of users worldwide while available on Google Play. Even more important is that the fidelity and ecological validity of the data will be higher if collected from the participant's primary device since it will not affect the participant's daily activities (Raento et al., 2009; Wang et al., 2014). The main drawback of using participants' smartphones is the lack of standardization. The mixture of devices, brands, and OSs can introduce noise in the sensor data collected and differences in the availability of sensors (Harari et al., 2016). According to recent reviews, although the global amount of use of personal and provided devices in studies has been almost equal, the use of personal smartphones has been increasing during the last decades (van Berkel et al., 2018). This is considered a positive development, since “the less aware the subject is of the presence of the observing device, the less its presence should affect the study" (Raento et al., 2009).

ESM sampling schedule The goal of studies employing ESM is to collect rich data on the participants’ experiences. This makes it necessary to find a balance between the number of questionnaires sent to the participants and the minimization of their burden. Different types of questionnaire triggers include *signal contingent*—randomised alerts during the course of a given time span—, *interval contingent*—fixed alerts according to a predefined schedule—and *event contingent*—alerts according to predefined events—(Barrett and Barrett, 2001). Additionally, some studies encourage participant-initiated data submissions, which, although useful for events occurring irregularly, cannot be simply analyzed as one collection since the methodology differs from the ESM. The most common ESM scheduling in literature is the interval contingent trigger (van Berkel et al., 2018). For this type of schedule, researchers recommend avoiding sending ESM notifications during the night and presenting longer schedules during evening hours since smartphone users show higher levels of attentiveness during that period (Dingler and Pielot, 2015). Regarding the inquiry limit, several guidelines have been discussed in the literature. Some studies indicate sampling frequencies of five to eight ESM questionnaires per day as an optimal balance of recall and annoyance (Klasnja et al., 2008). Another review shows that the frequency of ESM questionnaires varies from 1 to 10 times per day (Y. S. Yang et al., 2019). The acceptability of a certain number of questionnaires depends highly on the time and effort required to complete the questionnaire. Therefore, shorter and easier questionnaires may allow a higher sampling rate.

ESM presentation and response format When designing the questionnaire layout, several parameters must be considered, including question readability and modality—i.e., response format. Questionnaires must be easily understandable by the participants and keep their interruption as brief as possible (van Berkel et al., 2018). In fact, they are recommended to take no more than 2 minutes to complete (Csikszentmihalyi and Larson, 2014). For that reason, the use of sliders or Likert-based scales is usual in studies for mood assessment (Schimmack, 2003). In particular, for those response formats which use bipolar scales, it is recommended to clearly communicate the bipolarity to the participants using, for example, opposing labels—e.g., pleasant–unpleasant—at the two slider ends.

4.3 Experiment 1: mood fluctuations among elite athletes

The first experiment designed with the monitoring platform aims to gather affective and contextual data of athletes during their daily lives. In the domain of sport, emotions are acknowledged by experts as an inherent part of the competitive and daily practice experience (Robazza, 2006). Several models have been proposed to explain the emotion-performance relationship and provide the theoretical foundations for research and applications. One of them is known as Individual Zones of Optimal Functioning—IZOF (Hanin, 2000)—and suggests that each athlete has an individual range of activation level in which the optimal performance is achieved. This model relates sport performance with anxiety to reflect the importance of the individualized study of the emotional factors that characterize elite athletes.

Current research on temporal dynamics of emotions in sport focuses mainly on the description of the affective states on pre-, mid-, and post-event (Robazza, 2006). However, researchers encourage the study of both short and long-term emotion dynamics across the athletes' daily life to obtain an appropriate overview of the emotional behavior of the athlete. As part of the motivation of this thesis, it aims to design and conduct a longitudinal experiment of data collection during the daily life of a group of elite athletes, which could provide affective, contextual, and behavioral data. These data could be used to explore relations among the long-term emotional profiles of athletes and their performance traits and study the short-term fluctuations of the mood based on the context surrounding the athlete. The dataset resulting from this experiment is one of the first longitudinal datasets of affective behavior and context within a sports population.

4.3.1 Materials and methods

Participants

Twenty-two adults—11 females; 11 males; mean age = 23.1 years; std age = 7.86 years—were recruited to participate in the experiment, sample size in line with the local standards of the HCI community. Ten of them were elite athletes—six or more training sessions per week—, and the other twelve were sedentary people—occasional leisure sport. Since the experiment aimed to gather data without altering participants' natural behavior, it is designed to operate with the subjects' own smartphones, which were required to have Android OS. All participants were informed in a joint session about the objective of the study, and they read and signed an informed consent with information about the study, risks and benefits, privacy protection, and participation rights. Afterward, they received a unique ID number which was used as the only means of identification during the study. The session ended with a detailed explanation about how to install the monitoring app, scan the QR code, and disable the brand-specific battery optimization systems. No training session for the app use was given. Finally, after the joint session, participants were asked to fill an intake questionnaire about anxiety during their daily life.

Ethical approval

The experiment protocol was reviewed and approved by the Human Research Ethics Committee of the University of Granada, ref.: 642/CEIH/2018.

Data collection

During the experiment, the mood and context of the participants were continuously monitored for two weeks using the monitoring platform described in Chapter 3. The mood was assessed using ESM. Participants received mood assessment questionnaires on their devices, which triggered a notification six times per day, at pseudo-random times within six evenly distributed one-hour intervals: 07:00–08:00, 10:00–11:00, 13:00–14:00, 16:00–17:00, 19:00–20:00, and 22:00–23:00. This hybrid approach between interval contingent and signal contingent schedules was used to avoid the fabrication of responses—which could happen if the questionnaire always triggers at a fixed time—and to cover a broader range of contextual situations. Participants received a notification from the app indicating that a new questionnaire was available, which could be opened by tapping the notification. That way, the mood was assessed *in-situ* during any daily life situation. The contextual information was inferred from the data gathered through the smartphone sensors. In the following paragraphs, the questionnaire content and the sensor-based sampling are described.

Intake questionnaire: DASS-21 The intake questionnaire was intended to measure the average anxiety levels of each participant in order to give the chance to examine the relationship between the mood dynamics and the IZOF model in future analyses. The anxiety was assessed using the Spanish version of the short form of the Depression, Anxiety and Stress scale—DASS-21 (Bados et al., 2005; Henry and Crawford, 2005). This questionnaire comprises 21 items of three subscales to assess depression, anxiety, and stress, with seven items each. Each item is evaluated with a Likert scale ranging from 0 to 3 based on how frequently the subject experiences the behavior described in the item, and the scores of each subscale are added together to obtain the final score. The labels of each score are: 0 = "*Did not apply to me at all*"; 1 = "*Applied to me to some degree, or some of the time*"; 2 = "*Applied to me to a considerable degree, or a good part of time*"; 3 = "*Applied to me very much or most of the time*". In this study, only the anxiety subscale was used since it is the affective state most related to the sport performance—based on the IZOF model. Detailed information about the questionnaire items is shown in Table 4.1.

Mood assessment questionnaire Participants' mood was assessed using the two affect dimensions proposed in the Circumplex Model of Affect—see Section 2.1.1—, valence and arousal. The quick evaluation offered by this model is appropriate for longitudinal studies since it helps to considerably reduce the participant burden. In pursuit of a simple, easy-to-use response format, the questions were answered using a slider-based Likert scale. The valence slider displayed the question "*How happy do you feel right now?*" and ranged from -50 to +50. The arousal slider displayed the question "*How*

Table 4.1 Spanish version of the DASS-21 questionnaire: anxiety subscale. Items are scored from 0 to 3—0 = “*Nada aplicable a mí*"; 1 = “*Aplicable a mí en algún grado, o una pequeña parte del tiempo*"; 2 = “*Aplicable a mí en grado considerable, o una buena parte del tiempo*"; 3 = “*Muy aplicable a mí la mayor parte del tiempo*".

	Item
1	Me doy cuenta de que tengo la boca seca
2	Se me hace difícil respirar
3	Siento que mis manos tiemblan
4	Estoy preocupado/a por situaciones en las cuales puedo tener pánico o en las que puedo hacer el ridículo
5	Siento que estoy a punto del pánico
6	Siento los latidos de mi corazón a pesar de no haber hecho ningún esfuerzo físico
7	Tengo miedo sin razón

excited do you feel right now?" and ranged from 0 to +100. Anchors were located at the two endpoints of both sliders, but instead of using text-based anchors—e.g. “*happy*"–“*sad*"—or the numerical value of the scale, face icons representing each extreme of the slider were used (Figure 4.1). For example, in the valence question, the minimum value is represented with a sad face and the maximum one, with a happy face; the arousal question employed a calm face and an excited face for the minimum and maximum activation, respectively. Emoticon-based scales have been proved to be more intuitive and user-friendly than numeric or text-based ones without introducing any type of bias (Meschtscherjakov et al., 2009). The initial position of the slider was randomly assigned in each questionnaire to minimize anchoring issues. Detailed information about the questions and response options is shown in Table 4.2.

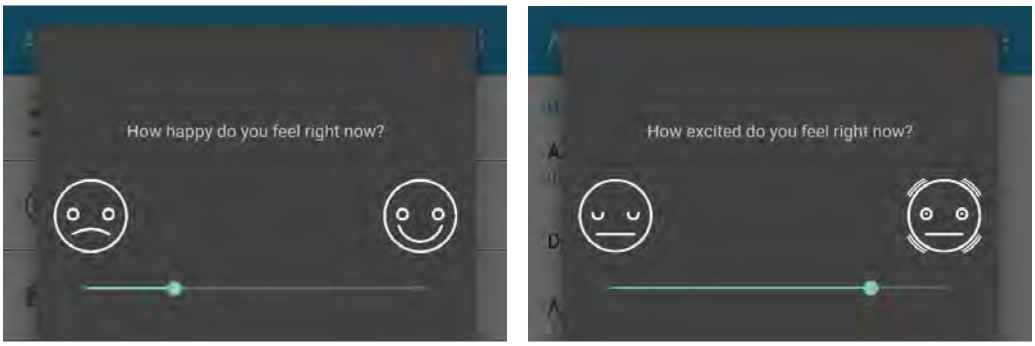


Figure 4.1 Screenshots of the ESM questions for assessing valence and arousal.

Smartphone sensors The smartphone sensors activated were selected based on the available literature (Harari et al., 2016; Mohr et al., 2017; Rüegger et al., 2020), to monitor a broad range of contextual elements. The sensor data was unobtrusively collected without any interaction required from the user. In subsequent steps of data processing, several behavioral markers will be extracted from the raw smartphone data using the hierarchical model described in Section 2.2. Table 4.3 summarizes the sensors used, their sampling frequencies, and the behavioral and contextual facets which are represented by the data of each sensor.

Table 4.2 Description of the mood assessment questionnaire.

Question	Response options
How do you feel right now?	Visual Analogue Scale ranging from 'Very bad' to 'Very good'
How physically active do you feel right now?	Visual Analogue Scale ranging from 'Not active' to 'Very active'

Table 4.3 Summary of the contextual sensor data collected.

Source sensor	Behavioral/contextual facets represented	Sampling frequency
Screen state	Sleep time (Ciman and Wac, 2019), phone usage (van Berkel, Goncalves, Koval, et al., 2019)	Each time the screen state changed
Notifications	Social interaction through phone (LiKamWa et al., 2013; van Berkel, Goncalves, Koval, et al., 2019)	Each time a notification was received
Foreground apps	Phone usage, social interaction through phone (LiKamWa et al., 2013)	Each time an app was in foreground
Battery level	Phone usage (Moshe et al., 2021)	Each time a charge/discharge started
Light sensor	Ambient light (Kim et al., 2019; Ma et al., 2012)	1 minute
WiFi network connections	Location (Servia-Rodríguez et al., 2017; van Berkel, Goncalves, Koval, et al., 2019)	5 minutes

4.3.2 Results: longitudinal mood and context dataset of athletes

As a result of this experiment, a longitudinal dataset containing emotional, contextual, and anxiety data of a population of athletes was obtained, curated, and stored at the Open Science Framework—OSF—repository (https://doi.org/10.17605/OSF.IO/2Z5B6). The data, associated metadata, and documentation are openly available for anyone interested, under the Creative Commons Attribution License—CC BY 4.0 (Creative Commons, 2021).

The responses to the DASS-21 intake questionnaire were used to compute the anxiety score of the participants, which is obtained by adding the scores of all the items. Values ranged from 0—lower average anxiety level during daily life—to 21—higher average anxiety level during daily life. The distribution of anxiety scores and the basic demographic information of the participants are illustrated in Figure 4.2.

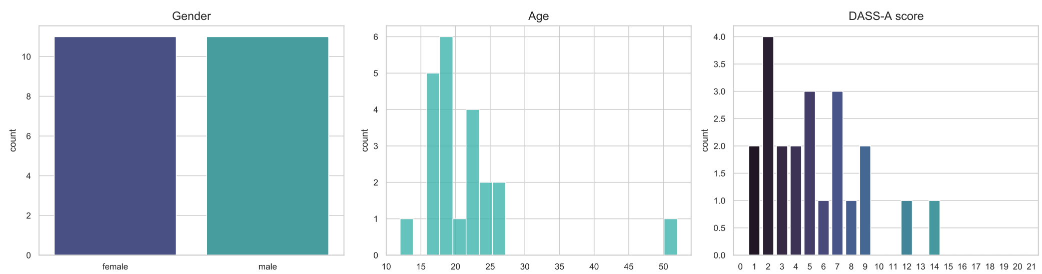


Figure 4.2 Description of the participant's sample and value distribution of the DASS-21 anxiety subscale.

The mood ratings registered are stored in the file `mood_raw.csv`. It contains 1,686 unique records. The variables, data types and value options are described in Table 4.4. The distribution of valence and arousal values registered is illustrated in Figure 4.3.

The contextual data gathered with the smartphone sensors is stored in a set of files, one per activated sensor. Table 4.5 summarizes the file names, their description, and the number of records in each one.

Table 4.4 Description of the file `mood_raw.csv`.

Variable name	Description	Type	Values
participant	Unique identifier of the participant	String	
answer_timestamp	Timestamp of the mood report	Datetime	YYYY-MM-DD HH:MM:SS
valence	Current valence rating	Integer	-50 to 50
valence_scale_ini	Initial value of the valence slider	Integer	-50 to 50
completion_time_valence	Time spent by the participant to complete the valence rating, in milliseconds	Numeric	
arousal	Current arousal rating	Integer	0 to 100
arousal_scale_ini	Initial value of the arousal slider	Integer	0 to 100
completion_time_arousal	Time spent by the participant to complete the arousal rating, in milliseconds	Numeric	
status	Status of the ESM questionnaire	Integer	One of {1 = dismissed, 2 = answered, 3 = expired}

Table 4.5 Sensor data files generated by the monitoring platform.

File name	Description	Records
screen.csv	Events of changes in the smartphone screen state	161,392
notifications.csv	Events of notifications received	73,275
applications.csv	Events of applications opened	221,000
battery_charges.csv	Events of battery charges	929
battery_discharges.csv	Events of battery discharges	928
light_raw.csv	Raw ambient light measurements	1,584,216
wifi.csv	Events of connections to WiFi networks	117,222

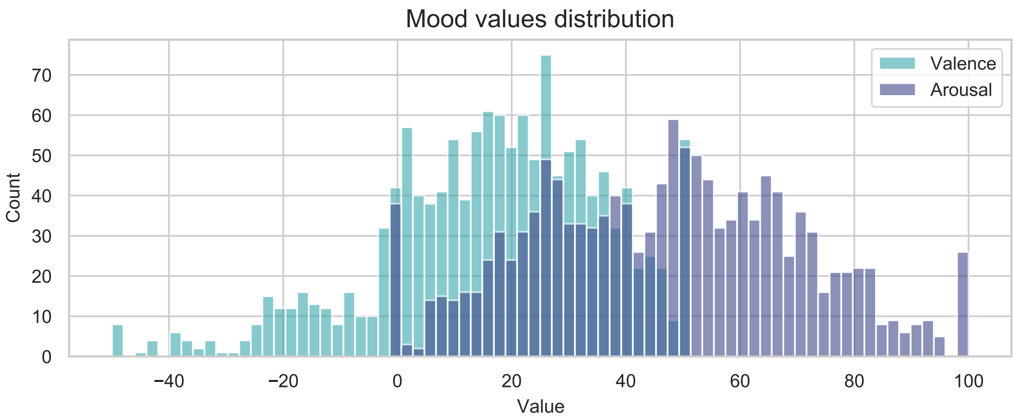


Figure 4.3 Distribution of valence and arousal values of all participants.

4.4 Experiment 2: mood fluctuations in lockdown conditions

The second experiment designed with the monitoring platform was motivated by the COVID-19 worldwide pandemic. On March 11th, 2020, the World Health Organisation—WHO—characterized this new coronavirus disease as a pandemic (Sohrabi et al., 2020; World Health Organization, 2020). Governments around the world adopted unprecedented confinement measures in an attempt to restrict the spread of the disease. These measures had a huge effect on people's daily activities and routines (Nussbaumer-Streit et al., 2020; Thompson, 2020). Spain quickly became one of the most affected countries worldwide, escalating from 9,785 to 239,429 diagnosed cases and from 136 to 27,117 deaths in two months—from early March to the end of May—(Roser et al., 2020). For that reason, on March 14th, the Spanish government imposed a widespread lockdown aimed at reducing social contact and avoiding the collapse of the national health system (Mitjà et al., 2020). The lockdown measures implemented rank among the most restrictive and prolonged and consisted in closing schools, universities, and non-essential industrial activity countrywide to drastically reduce the population's mobility (Tobías, 2020).

These confinement measures led to dramatic changes in people's behavior and lifestyle, with a recognized negative impact in mental health terms, whose levels of anxiety, stress, and depression were expected to rise (Brooks et al., 2020; Moccia et al., 2020; Odriozola-González et al., 2020). To best anticipate the needs for psychosocial support and evidence-based policy-making, it is essential to gather accurate information on the population's emotional response to the ongoing events. This critical issue was attempted to be addressed by several studies (Ahmad and Murad, 2020; Moccia et al., 2020; H. Yang and Ma, 2020), which employed cross-sectional surveys, thus capturing a static description of the population's evolving mood. However, research on mood dynamics has shown that the study of long-term mood variations may be critical for understanding and predicting psychological well-being (Houben et al., 2015).

Given the limitation mentioned above of existing datasets at that moment, a project named CovidAffect was initiated. It comprises the collection and curation of a database of individual changes in mood during the COVID-19 lockdown in Spain. Participants countrywide regularly reported their mood via the developed monitoring platform from March 28th to June 21th, 2020, when the nationwide state of alarm was lifted. As the lockdown forced people to stay at home, no contextual data was collected beyond basic socioeconomic and living information. The study provided longitudinal, openly available data of mood variations in the Spanish territory, which could be used to investigate aspects of the psychological impact of the COVID-19 crisis on the affected population. Notably, at the time of its release, this was the first dataset that offered the opportunity to study the behavior of mood dynamics in a lockdown situation.

4.4.1 Materials and methods

Participant onboarding

Since its official release, the project was broadcast in social media and Spanish national press and opened to all the population countrywide. Volunteer participants joined by accessing the project's website (https://covidaffect.info), which described the scope of the project, offered privacy information, and presented a summary of the data through an interactive map. No exclusion criteria were applied.

The onboarding process was developed as follows. Participants accessed the website and had to click the *participate* button. They were asked to fill an intake questionnaire on demographic, socioeconomic, and COVID-19 related data. To submit the questionnaire and further proceed with the registration, participants had to sign an informed consent with detailed information about the study, risks and benefits, privacy protection, and participation rights. Afterward, they received a unique ID number which was used as the only means of identification in the subsequent procedures, and they were asked to complete the first mood assessment. Finally, a video tutorial of the platform's client app installation procedure was provided. A layout of the participants' onboarding process is depicted in Figure 4.4.

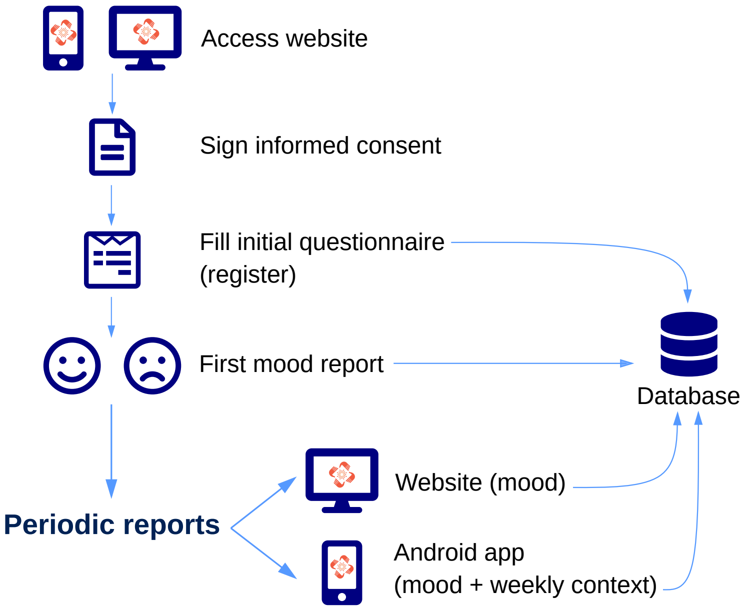


Figure 4.4 Participant onboarding process to CovidAffect study.

Ethical approval

The project was reviewed and approved by the Human Research Ethics Committee of the University of Granada, ref.: 1378/CEIH/2020.

Data collection

Since this experiment was intended to be available for all the population in a specific situation as a lockdown is, the participants' own smartphones were used as sampling devices, giving them instructions to install our client app. Since no contextual features were intended to be obtained, this experiment was limited to ESM sampling. Regarding the ESM trigger, interval contingent triggers are the most used ones in the literature. However, as participants get used to the trigger time, answers can be previously prepared and not truly reflect the current participant's mood. Therefore, a hybrid approach between interval and signal contingent scheduling was used. The client app triggered the mood assessment questionnaire at pseudo-random times during six, evenly distributed one-hour intervals: 07:00–08:00, 10:00–11:00, 13:00–14:00, 16:00–17:00, 19:00–20:00, and 22:00–23:00. Participants received a notification indicating that a new questionnaire was available, which displayed the mood rating screen. Given the low sampling rate of ESM, the data was synchronized with the server at each questionnaire submission. If not answered, the notifications expired after one hour to avoid questionnaire overlap and preserve the sample distribution over time. In addition, once a week, participants received a supplementary questionnaire with questions related to current health and lifestyle status to follow the fluctuations of these habits during the lockdown. This questionnaire was triggered at a fixed time during the evening—Fridays at 20:00—, following the methodological recommendations found in the literature. The content and periodicity of each questionnaire are summarized in Table 4.6, and their description is provided in the following paragraphs.

Table 4.6 Summary the characteristics of each CovidAffect questionnaire.

Questionnaire	Trigger frequency	Content
Intake	Once, when the participant registers at the website	15 questions including demographic, situational, socioeconomic and contextual information (Table 4.7).
Mood	Six times per day	2 questions rating subjective feeling and physical arousal (Table 4.8).
Context	Weekly	8 questions about contextual changes during the week (Table 4.9).

Intake questionnaire The intake questionnaire was a mandatory step during the registration of the participants at the website. It was designed to collect data on demographics, residence characteristics, employment, COVID-19 symptoms, and average physical and emotional status previous to the lockdown. Detailed information about the questions and response options is shown in Table 4.7.

Mood assessment questionnaire Mood fluctuations were monitored using the Circumplex Model of Affect—see Section 2.1.1. This model allows for a quick assessment of mood, evaluating only two affect dimensions, valence, and arousal. This fast evaluation reduces enormously the participant burden and is appropriate for longitudinal experiments. The questionnaire presented the two assessment questions on consecutive screens to maximize the participant's attention to each question. Sliders with continuous range were used as the response format—i.e., Visual Analogue Scale (VAS). Valence slider displayed the question "*How do you feel right now?*" and ranged from -50 to +50, with anchors located at -50—"*Very bad*"—and +50—"*Very good*". Arousal slider showed the question "*How physically active do you feel right now?*" ranged from 0 to +100 with anchors located at 0—"*Not active*"—to 100—"*Very active*". The initial value of the slider was randomly assigned in each questionnaire to minimize anchoring. Detailed information about the questions and response options is shown in Table 4.8, and a screenshot of the app questionnaire is shown in Figure 4.5.

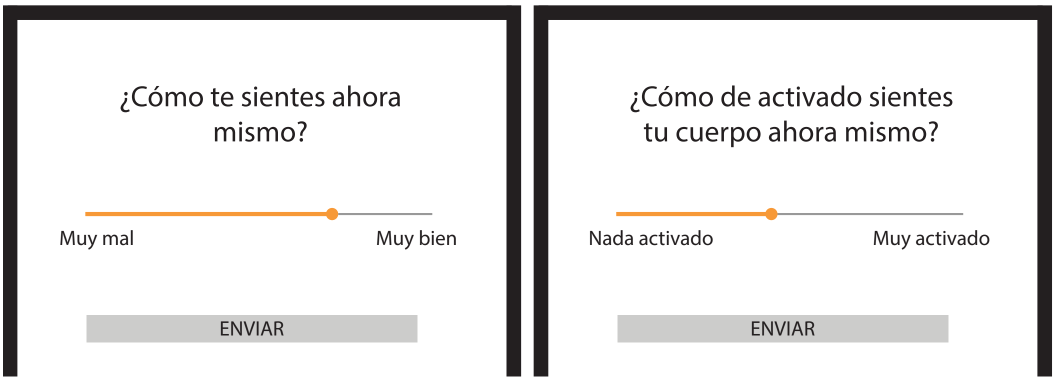


Figure 4.5 Valence and arousal rating screens triggered by the CoVidAffect smartphone app. English: '*How do you feel right now?*'—'*Very bad*' to '*Very good*'—, '*How physically active do you feel right now?*'—'*Not active*' to '*Very active*'.

Weekly context questionnaire A weekly questionnaire was delivered to track possible changes in the contextual or socioeconomic status during

Table 4.7 Description of the intake questionnaire.

Question	Response options
Gender	One of {'Male', 'Female', 'Other'}
Age	Integer (>= 16)
Postcode	Integer
How many people live with you?	One of {'I live alone', '1 person', '2 people', '3 people', '4 people', '5 people', '6 or more people'}
Age and relationship of people living with you[1]	Age: integer Relationship: one of {'Parent', 'Spouse/Couple', 'Child', 'Sibling', 'Grandparent', 'Grandchild', 'Other', 'No family relationship'}
Type of residence	One of {'Studio', 'Apartment', 'House', 'Rest home', 'Chalet', 'Other'}
How many rooms does your residence have?[2]	One of {'1', '2', '3', '+3'}
Do you have access to any of the following open spaces?	Multiple choice from of {'Balcony or terrace', 'Garden', 'Courtyard', 'Other', 'Any of these'}
Employment status before the start of the crisis	One of {'Employee', 'Self-employed', 'Unemployed', 'Student', 'Retired', 'Other'}
Current employment status	One of {'Not changed', 'Teleworking', 'Reduced workday', 'Increased workday', 'Temporary Employment Regulation, ERTE', 'Fired', 'New employment', 'Other'}
Net monthly income	One of {'Less than 500€', '500 to 999€', '1000 to 1499€', '1500 to 1999€', '2000 to 2499€', '2500 to 2999€' '3000 to 4999€', '5000 to 6999€', '7000 to 8999€', 'More than 9000€'}
Do you consider that the crisis has negatively affected your economic situation?	One of {'Yes', 'No'}
Regarding COVID-19	One of {'I don't have symptoms', 'I have symptoms but have not been diagnosed', 'I am diagnosed'}
From the people who live with you	One of {'No one have symptoms', 'One or more have symptoms but have not been diagnosed', 'One or more have been diagnosed'}
Before the lockdown, how many hours of physical activity did yo get weekly?	One of {'Less than 2h', '2 to 4h', '4 to 6h', '6 to 8h', 'More than 8h'}
Generally, how do you consider your mood before the crisis?	Visual Analogue Scale ranging from 'Very negative' to 'Very positive'

[1]This question is displayed only if the answer to the previous question is one or more. It adds two fields for each family member, one for age and one for relationships.

[2]This question is displayed only if the answer to the previous question is 'Apartment' or 'House'.

the study. It gathered data regarding changes in COVID-19 diagnosis, health status, habits, and employment, occurring during the past week. Detailed information about the questions and response options is shown in Table 4.9.

Table 4.8 Description of the mood assessment questionnaire.

Question	Response options
How do you feel right now?	Visual Analogue Scale ranging from 'Very bad' to 'Very good'
How physically active do you feel right now?	Visual Analogue Scale ranging from 'Not active' to 'Very active'

Table 4.9 Description of the weekly context questionnaire.

Question	Response options
Have you been diagnosed with COVID-19 during the past week?	One of {'Yes', 'No'}
Has someone living with you been diagnosed with COVID-19 during the past week?	One of {'Yes', 'No'}
Has some relative—not living with you—been diagnosed with COVID-19 during the past week?	One of {'Yes', 'No'}
How do you consider your overall health status?	Visual Analogue Scale ranging from 'Very bad' to 'Very good'
Has your employment status changed during the past week?	One of {'Not changed', 'Teleworking', 'Reduced workday', 'Increased workday', 'Temporary Employment Regulation, ERTE', 'Fired', 'New employment', 'Other'}
How many hours of physical activity did you get during the past week?	One of {'Less than 2h', '2 to 4h', '4 to 6h', '6 to 8h', 'More than 8h'}
How frequently have you been in contact with relatives, friends and neighbours during the past week?	One of {'Same as before the lockdown', 'Less than before the lockdown', 'More than before the lockdown'}
On average, how many hours do you sleep every night?	One of {'Rather less than usual', 'Slightly less than usual', 'Same as usual', 'Slightly more than usual', 'Rather more than usual'}

4.4.2 Results: CovidAffect dataset

As a result of this experiment, the CovidAffect dataset was obtained, curated, and stored at the Zenodo (https://doi.org/10.5281/zenodo.3774526) and OSF (https://doi.org/10.17605/osf.io/5CQZK) project repositories, in line with FAIR principles (GO FAIR, 2020) and RDA guidelines (RDA COVID-19 Working Group, 2020) for COVID-19-related datasets. The data, questionnaire description, associated metadata, and documentation are openly available for anyone interested, under the Creative Commons Attribution License—CC BY 4.0 (Creative Commons, 2021).

A total of 999 participants successfully registered and reported at least one ESM questionnaire. As the participation was voluntary, there was no fixed study length, and participants could withdraw at any moment, resulting in variable sample sizes over time. The average enrollment length was 23 days—in line with the duration of mobile sensing and ESM studies—, and participants reported an average of 68 mood assessments during the study. Figure 4.6 illustrates the sample size as a function of the enrollment duration. Researchers can select any study duration based on their research interest and obtain the corresponding sample size. For example, for a time window of one week, data from 154 participants are available. On the other end of the spectrum, 44 participants provided data for 60 consecutive days. The responses to each questionnaire were stored in three comma-separated values (CSV) files that are described hereafter.

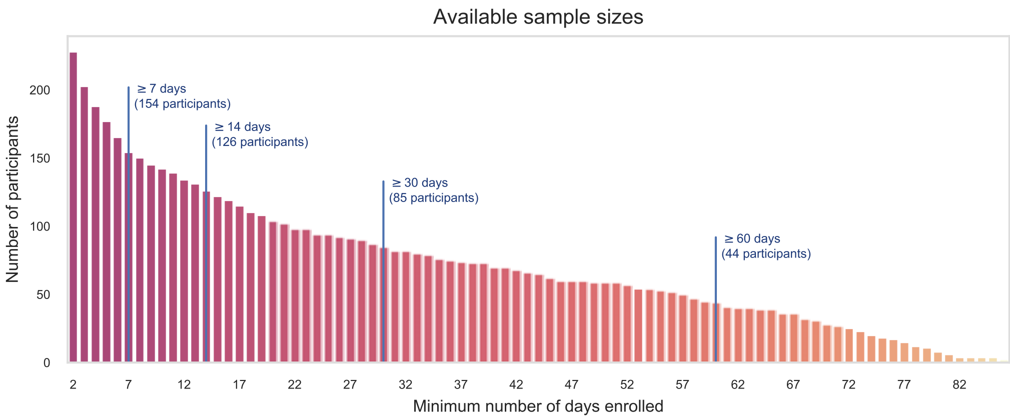


Figure 4.6 Available sample sizes for longitudinal analysis. For visualization purposes, participants with only one day of enrollment are not included.

The responses to the intake questionnaire are in the file `participants.csv`. It contains 999 unique records—one per study participant. The variables, data types and value options are described in Table 4.10. Demographic and baseline characteristics of the study participants based on this questionnaire are shown in Figure 4.7.

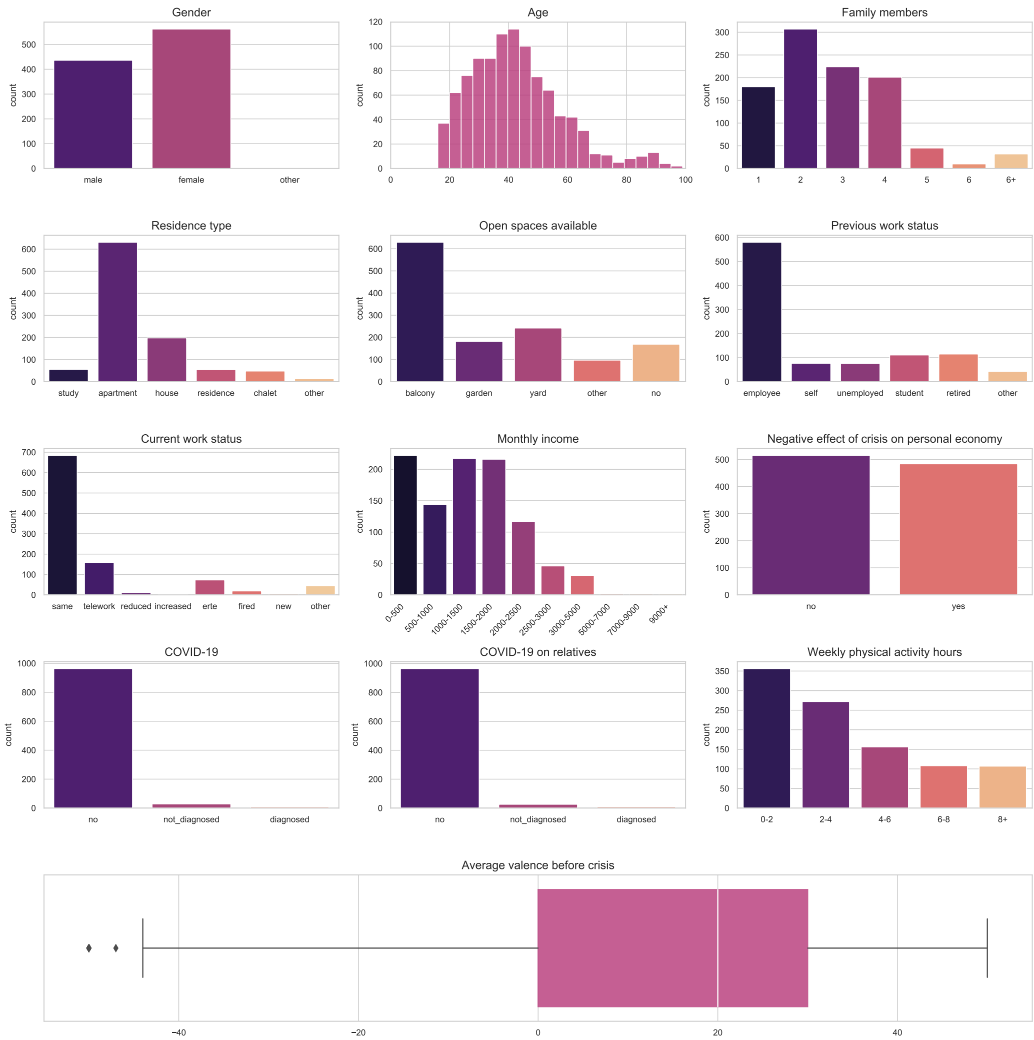


Figure 4.7 Description of the participant's sample.

The mood ratings registered are stored in the file `mood.csv`. It contains 17,452 unique records. The variables, data types and value options are described in Table 4.11. The number of mood questionnaires submitted each study day is illustrated in Figure 4.8. The geographical distribution of the participants within the Spanish national territory—based on the provided

Table 4.10 Description of the file `participants.csv`.

Variable name	Description	Type	Values
id	Unique identifier of the participant	Integer	
registered_date	Timestamp of the participant's registration at the website	Datetime	YYYY-MM-DD HH:MM:SS
sex	Gender of the participant	String	One of {'fem', 'masc', 'other'}
age	Age of the participant	Integer	16 to 100
postcode	Postcode of the participant	Integer	
family_members	Number of people in the residence	String	One of {'1', '2', '3', '4', '5', '6+'}
family_ages	Age of each co-resident specified in the variable family[a]	String	Ordered, comma-separated values for each co-resident from 0 to 100
family_relation	Relationship between the participant and each co-resident specified in the variable family[a]	String	Ordered, comma-separated values for each co-resident from of {'parent', 'partner', 'child', 'sibling', 'grandparent', 'grandchild', 'other', 'any'}
residence_type	Type of participant's residence	String	One of {'study', 'apartment', 'house', 'residence', 'chalet', 'other'}
residence_rooms	Number of rooms in the participant's residence[b]	String	One of {'1', '2', '3', '3+'}
open_spaces	Open spaces available at the participant's residence	String	Comma-separated values from of {'balcony', 'garden', 'yard', 'other', 'no'}
work_previous	Employment status of the participant before the crisis	String	One of {'employee', 'self', 'unemployed', 'student', 'retired', 'other'}
work_current	Current employment status of the participant	String	One of {'same', 'telework', 'reduced', 'increased', 'erte', 'fired', 'new', 'other'}
income	Net monthly income of the participant coded in 10 possible ranges, in euros	String	One of {'0-500', '500-1000', '1000-1500', '1500-2000', '2000-2500', '2500-3000', '3000-5000', '5000-7000', '7000-9000', '9000+'}
negative_economy	Binary flag to indicate the participant's perception on whether the crisis has negatively affected to his/her economic situation	Integer	One of {0, 1}
covid	Presence of COVID-19 symptoms in the participant	String	One of {'no', 'not_diagnosed', 'diagnosed'}
covid_family	Presence of COVID-19 symptoms in other co-residents	String	One of {'no', 'not_diagnosed', 'diagnosed'}
physical_activity	Number of hours dedicated to physical activity by the participant before the crisis	String	One of {'0-2', '2-4', '4-6', '6-8', '8+'}
valence	Valence rating of the participant before the crisis	Integer	-50 to 50

[1] This field is empty if *family_members* value is '1' or '6+'.

[2] This field is empty unless *type_living* value is 'apartment' or 'house'.

Table 4.11 Description of the file `mood.csv`.

Variable name	Description	Type	Values
participant	Unique identifier of the participant—*id* variable of *participants.csv*	Integer	
timestamp	Timestamp on which the questionnaire notification	Datetime	YYYY-MM-DD HH:MM:SS
answer_timestamp	Timestamp of the mood report	Datetime	YYYY-MM-DD HH:MM:SS
valence	Current valence rating	Integer	-50 to 50
arousal	Current arousal rating	Integer	0 to 100
valence_scale_ini	Starting position of *valence* input slider	Integer	-50 to 50
arousal_scale_ini	Starting position of *arousal* input slider	Integer	0 to 100

postcodes—is shown un Figure 4.9. The three provinces with more number of participants were Granada, Madrid and Cádiz.

The responses to the weekly questionnaire are stored in the file `context.csv`. Since not every participant completed the weekly questionnaires, it contains 395 unique records. The variables, data types and value options are described in Table 4.12.

4.5 Discussion

Application of the monitoring platform on real-life experiments During the last years, several diverse technologies and methodologies have been employed by researchers when addressing different problems involving the longitudinal collection of data in free-living environments. A multitude of behaviors can be better understood if they are longitudinally studied. People's mobility behaviors, workplace-related behaviors, affective fluctuations, or even clinical conditions such as depression or bipolarity are some of the targets whose long-term observation can yield insights unknown so far.

However, the monitoring tools employed usually follow *ad-hoc* designs for the specific problem studied, being only capable of collecting data within a

Table 4.12 Description of the file `context.csv`.

Variable name	Description	Type	Values
date	Date on which the questionnaire was answered	Date	DD/MM/YYYY
participant	Unique identifier of the participant—corresponding to the *id* variable of *participants.csv*	Integer	
covid_diagnosed	Binary flag to indicate whether the participant has been diagnosed with COVID-19 during the past week	String	One of {'Yes', 'No'}
covid_residence	Binary flag to indicate whether any co-resident was diagnosed with COVID-19 during the past week	String	One of {'Yes', 'No'}
covid_family	Binary flag to indicate whether any family member or close person was diagnosed with COVID-19 during the past week	String	One of {'Yes', 'No'}
perceived_health	Participant's perception on his/her general health status during the past week	Integer	1 to 5
work_changed	Changes in work status during the past week	String	One of {'no', 'telework', 'reduce', 'increase', 'erte', 'fired', 'new', 'other'}
physical_activity	Number of hours dedicated to physical activity during the past week	String	One of {'0-2', '2-4', '4-6', '6-8', '8+'}
social_contact	Social contact frequency during the past week compared to social contact before the crisis	String	One of {'same', 'less', 'more'}
sleep	Sleep quantity during the past week, compared to the average before the crisis	String	One of {'lot_less', 'less', 'same', 'more', 'lot_more'}

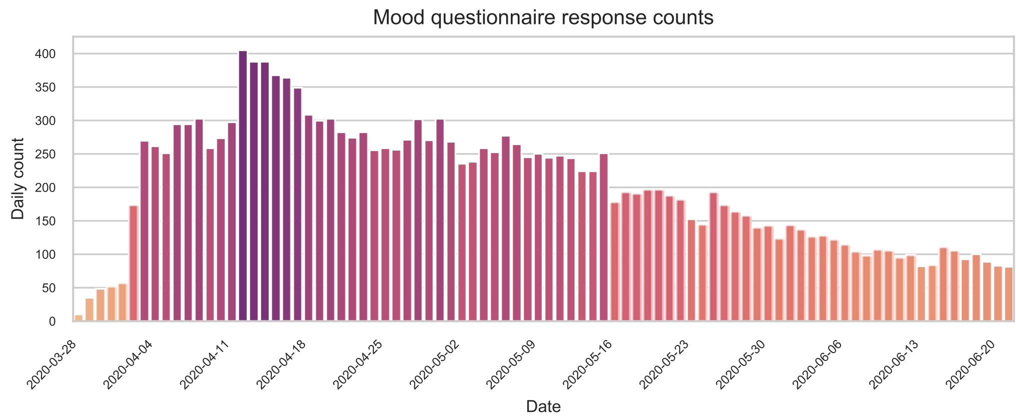


Figure 4.8 Daily amount of responses to mood questionnaire during study period.

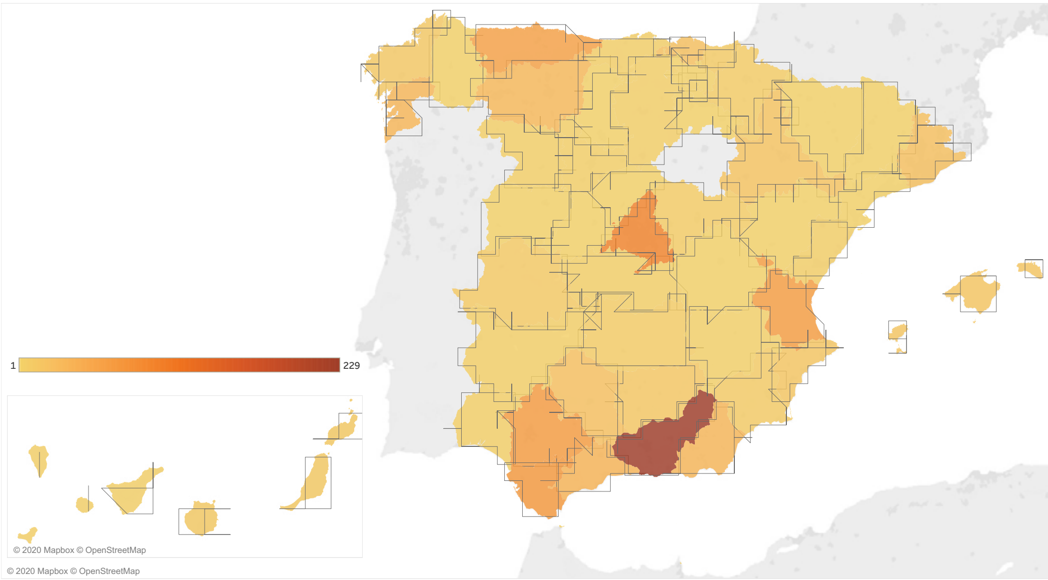


Figure 4.9 Geographical distribution of the study participants.

controlled population. Conducting a study open to all the population entails additional challenges since the data collection cannot be fully supervised, and mobile sensing tools are deployed in a wide range of devices, targeting people with diverse technical knowledge. This work shows how the monitoring platform developed can accommodate two different data collection experiments within both a controlled population and a public participation sample. The latter methodology is commonly known as *citizen science*, which can be described as the public participation in scientific research, in which experiments are opened to all the population, which can voluntarily join the data collection. Although both types of research differ in terms of methodological considerations, they can be deployed using the monitoring platform developed in this thesis, which has been proven to successfully gather data in both situations.

Participants in longitudinal experiments: controlled vs. open samples The progress of science depends on the interactions and collaborative efforts put by the scientific community, which provides a cumulative knowledge base that allows science to build on itself. In this sense, several difficulties and challenges have been faced during the design and deployment of the two real-life longitudinal experiments on this work. For that reason, one additional contribution of this work is to share these issues with the scientific community with the aim of setting the stage for future researchers interested in conducting longitudinal experiments in free-living environments. In the following, the main lessons learned from this experience are discussed—more information about the experience with the CovidAffect project can be found in a scientific dissemination article in The Conversation (Bailon et al., 2020).

The main methodological difference among the two experiments conducted is the studied population. Although the data collection was carried out in free-living environments in both of them, the first one involved a previously selected population that was specifically recruited and made particularly aware of the scientific and applied usefulness of the collected data. However, the second one involved the voluntary participation of anyone interested, with a nationwide open call for participation promoted through social media, press, and the project's website. This voluntary implication raised several insights that should be considered when conducting such experiments. First of all, one of the most critical issues is that a minimum sample size is not assured. Since participants can join and withdraw from the experiments without contacting the researchers, it is unknown to them if the participants whose data collection stopped were experiencing technical issues or just left the study. Figures 4.6 and 4.8 show that, although the initial reception of the project was excellent and many people joined, the participation starts being reduced after the second week of the study, following a constant decrease until the end of the experiment. This effect is also appreciated in the first experiment and aligns with the methodological recommendations available in the literature for longitudinal experiments involving ESM—two to four weeks of data collection.

Regarding the CovidAffect dataset, the decreasing sample is not necessarily a negative issue, since researchers can select the desired analysis period and obtain the corresponding sample size, provided that the sample characteristics are not relevant for the sake of the study. For example, if we want to analyze short-term mood fluctuations during one week, there is a sample of 154 participants; if the research objective is the long-term mood patterns during two months of lockdown, we will have 44 participants. It is important to emphasize that although this sample size is considerably shorter than the total amount of participants enrolled in the study—999—, it is larger than the average number of participants in existing longitudinal studies with ESM sampling—see Section 4.2. Moreover, the study duration surpasses by far the average study duration of this type of studies.

An important lesson learned from the experience mentioned above is that, in public sample projects, participants' withdrawal will normally be present. People's overloaded and busy lifestyles make it unavoidable. Moreover, nowadays—and especially in COVID-19 times—, the population is constantly bombarded with information, news, and surveys, and smartphones have reached a saturation point in which they are a source of stress and anxiety. Therefore, since longitudinal experiments by definition involve weeks of data collection, the targetted population should be comprised of as many participants as possible, so that afterward a reasonable withdrawal rate can be assumed. As a recommendation, if the study needs a minimum sample size, a mixed recruitment methodology can be used. First, a controlled minimum sample can be recruited following the traditional recruitment procedures, offering a reward for completing the study. Once the data collection starts with this sample, the experiment can also be opened to the public, allowing additional participants to join it voluntarily. This way, the minimum needed sample size can be assured irrespective of the number of voluntaries and their behavior during the experiment. However, the study protocol must be the same for both groups, providing them the same information and guidelines prior to the data collection, and using the same onboarding process for both of them, so that the behavior of the controlled sample is not modified.

Rewards can generally increase adherence to longitudinal studies. However, in publicly open samples, the large number of participants could make it challenging to provide enough rewards due to limited fundings. In that case, during the CovidAffect study, we found it effective to provide feedback that the participants can use to confirm that their data is being collected and they are actively contributing to the study. However, it is necessary to ensure that the feedback provided does not include information about the behavioral or contextual facets observed since this could cause behavior modifications if the participant thinks that the results are not being “good". For example, in the CovidAffect study, the project’s website included an interactive map (Figure 4.9), in which participants could select their city, see the number of data collected in real-time, and even change parameters such as the date limit, the age limit, or the gender to see how the participation has been evolved. Additionally, if they provided their participant ID, they could see their own amount of data collected during the specified period.

User interaction with longitudinal data collection tools The second set of lessons learned is related to the usage of the monitoring platform. We live in a world full of technological improvements, where we are getting used to interacting with multiple devices. Smartphones, smart TVs, virtual assistants, and other intelligent devices are part of our daily life and, because of this, the interaction with them must be as natural as possible. For that reason, users are getting used to smooth, easy-to-use human-computer interfaces. This also applies to any data collection tool aimed at capturing people’s natural behavior. App crashes, long server connection delays, and complex sequences to display and answer questionnaires have a negative effect on participation and could increase the withdrawal rate since people could find the interaction with the system too annoying. Therefore, polishing and easing the user experience of the client tools employed should be a priority during the design and development stages. Significant emphasis should be put in the interaction with ESM questionnaires, displaying them without the need to put the app in the foreground, and actively acknowledging the participant about the reception of the answer after the questionnaire submission.

Another critical issue faced is the presence of battery optimization systems in the participants’ devices. These systems were presented in Section 1.4 as a challenge in mobile sensing studies, and their role in the usability evaluation of the monitoring platform was discussed in Section 3.3.3. These systems are difficult to avoid in themselves but they can be managed with a controlled participant sample. During the preparation for the first experiment, a series of tests were run where the main part of the devices killed the client app—thus stopping the background sensing and ESM notification triggering—if the brand-specific battery optimization systems were not disabled. This issue could be solved in the first experiment by manually disabling them, participant by participant, during the joint session, which took place before the experiment started. However, this is impossible to accomplish when researchers are not in contact with the participants, which is the case of public participation studies such as CovidAffect. For that reason, it is highly advisable to include instructions about how to disable them for the most widely used smartphone brands, with plenty of screenshots or even a video tutorial about the procedure. Nevertheless, due to the large variety of devices and OS versions, a certain amount of errors and missing data should be assumed.

The final recommendation is related to the app installation procedure. Mobile sensing apps typically involve several installation steps: the installation itself, study configuration, possible intake questionnaires within the app, granting sensor access permissions, etc. Even when an installation guide is provided, not every participant reads it, or at least does it paying enough attention, which could lead to installation errors and participant frustration. In this regard, we recommend including the installation steps as built-in dialogs within the own app. These dialogs should include instructions about every step taken by the participant, such as how to accept the permissions, how to scan the QR code, or how to activate the accessibility settings. Including a progress bar of these steps and informing about the estimated installation time is also recommended.

Limitations of the resulting datasets The methodologies of the experiments are not exempt from certain limitations that should be noted before using the resulting datasets. Regarding the longitudinal dataset of athletes, it should be underlined that, although the study protocol includes data from smartphone sensors, the set of sensors selected may not cover every contextual aspect of the participant's daily life. More smartphone sensors could be included in future works. Regarding the CovidAffect dataset, a great part of the participants reported only the first mood rating, which was assessed through the project's website during the registration in the study. It could be due to issues with the app, which not allowed participants to further use it, or because they withdrew before starting the smartphone data collection. Although it is still useful as a wide one-shot mood rating sample, and it can provide an overview of the populations' mood at specific points in time, a preliminary analysis of the data showed significant differences in valence and arousal between participants with only one mood report and those with longitudinal data. The average valence of participants with only one mood record was 2.0 ($\pm$24.5), while the average valence of participants who adhered to the study protocol for two or more days was 9.8 ($\pm$17.6). Differences in arousal were less noticeable, with the average arousal of "one-shot" participants being 44.2 ($\pm$25.6) versus 48.4 ($\pm$18.9) for those with repeated mood recordings. The reason for these statistically significant differences ($p_{valence}$<0.001; $p_{arousal}$=0.009) may be partly the existence of a positive trend in participants with longitudinal recordings—average slope of 0.092 for valence and 0.135 for arousal. However, it could also indicate differences in other sample aspects not assessed in the study.

4.6 Conclusions

Collecting longitudinal behavioral, affective, and contextual data in real-world settings is a challenging task that requires a lot of equipment and time. Due to the difficulties present in data collection experiments, their deployment is not always feasible. Moreover, the design of these experiments in free-living environments involves several methodological decisions, which are not always adequately addressed and communicated in subsequent scientific publications. The use of already existing and validated datasets for the design and implementation of knowledge extraction analyses is highly encouraged by researchers. However, the current availability of open longitudinal datasets with such data is scarce.

In this chapter, two longitudinal data collection experiments are conducted using the monitoring platform presented in Chapter 3, and their methodology and resulting datasets are described. In the first experiment, affective and contextual data is collected from a group of athletes during their daily life. A controlled sample of 22 participants composed of elite athletes and sedentary people was monitored for two weeks, gathering mood information through ESM questionnaires and contextual data using the smartphone sensors. In addition, a survey regarding average anxiety levels is collected in the first stage of the study. The second experiment was motivated by the lockdown imposed in many countries worldwide due to the COVID-19 pandemic. In this citizen science experiment, the study was opened to all the Spanish population, who voluntarily joined the study during a maximum period of two months, self-reporting their mood six times per day.

These experiments resulted in two longitudinal datasets openly available for the scientific community and any other individual interested in using them. The first dataset is one of the first ones that longitudinally monitor elite athletes during their out-of-sport lives. The second one, the ground-breaking CovidAffect dataset, is one of the most extensive datasets—both in terms of sample size and study duration—, which studied the mood fluctuations of the population in lockdown conditions, and one of the first datasets doing so during the COVID-19 pandemic. Through these datasets, this work contributes to provide reliable, validated longitudinal data about affective fluctuations in two different scenarios, which is highly demanded and appreciated by the scientific community. Moreover, in this chapter, the main methodological considerations and lessons learned from the experience of these studies are discussed, thus contributing to the safer and more accessible design of future longitudinal experiments by researchers. Subsequently, several knowledge extraction methodologies can be applied to these datasets, one of them being designed, implemented, and tested in Chapter 5.

5 Methodology for the context-aware analysis of longitudinal affective data

5.1 Introduction

Mobile sensing is a flourishing technique in the field of computer science research. The collection of longitudinal data provides a more complete overview of the dynamic nature of every type of behavior, including affective ones. There are many opportunities for researchers to examine longitudinal behavioral data and extract valuable knowledge that was out of reach years ago. However, the methodological toolkit for analysis and knowledge extraction has only recently begun to enter behavioral research (Harari et al., 2016), making the analysis of longitudinal data a task ready to be explored.

In order to fully appreciate the analysis possibilities, it is necessary to understand the structure of human-based longitudinal data. Although each person behaves independently, the social nature of humans makes us prone to live in groups with similar lifestyles and routines. For that reason, the structure of longitudinal behavioral data is inherently nested (Bauer et al., 2020). As it was explained in Section 1.4 if individuals are observed on multiple occasions over time, the evolution of their behavior can be explored from two perspectives: studying the differences in behavior among all the individuals within a group can provide insight into the long-term aspects of their behavior and the similarities and differences among them—which is known as *between-person* level analysis. If we go one level deeper, studying the behavioral evolution of each individual separately, a more granular understanding of the short-term fluctuations of that particular individual can be reached, such as the behavioral reaction of that individual to a specific contextual scenario (Müller et al., 2020; Yu et al., 2021).

This detailed level of analysis leads to knowledge that is out of reach when analyzing sporadic data—one-shot questionnaires, cross-sectional surveys, etc.—(Müller et al., 2020; van Berkel et al., 2018). However, its complex structure poses significant challenges when attempting to understand the affect and behavior dynamics thoroughly. The ability to generalize meaningful patterns is often more challenging than the data collection itself and requires advanced techniques of analysis that capture the evolutionary nature of the longitudinal data. In addition, the context in which each behavior takes place gains huge informativeness when capturing affective fluctuations during long periods. Characterizing the situation surrounding the individual from smartphone-sensed data allows for a better understanding of the relation of their behavior with the context (Papini et al., 2020; Servia-Rodríguez et al., 2017). In this sense, researchers emphasize the need for novel methodologies to extract interpretable knowledge which could be directly applied to social, behavioral, or health interventions (Harari et al., 2016; VanDam and De Palma, 2019).

Among the main applications of longitudinal affective analysis, the description of mood patterns over time stands out as one of the most valuables. Uncovering the basic behavioral or affective lifestyles of people during their daily lives would provide a much-needed empirical basis on which more complex and sophisticated questions can be built (Harari et al., 2016). For example, a set of behavioral indicators can be used to group individuals in *between-person* classes with specific affective trends. At *within-person* level, examining short-term behavioral lifestyles such as daily activity or social patterns can provide complementary insights about the reason for the affective fluctuations. The current methodological approaches aiming at unveiling behavioral patterns usually focus on studying the relationships between a set of contextual or behavioral features and a target behavior—e.g., mood. For example, Yu et al. (Yu et al., 2021) use multilevel modeling to study the effect of certain factors on the behavior of a set of individuals in the hospitality management context. They highlight the additional information extracted using longitudinal rather than cross-sectional data. In another example, Müller et al. (Müller et al., 2020) apply factor analysis to explore the relationships between the mobility behavior of a group of individuals and their subjective well-being. They aggregate the data across the whole study period—between-person level—and each study day—within-person level—, obtaining complementary information about these relationships.

The second most widely employed analysis of longitudinal data is the prediction of life outcomes. Longitudinal data about behavior and context can be used to determine the key predictors of essential life aspects such as physical health, mental health, subjective well-being, mood, or performance. For example, Servia-Rodriguez et al. (Servia-Rodríguez et al., 2017) used neural networks to explore the predictability of mood using passively-sensed contextual data. They reported a limited accuracy due to the noise introduced by the scale of the uncontrolled study. Another study carried out by Eagle and Pentland (Eagle and Pentland, 2006) used machine learning models to identify and predict behavioral routines based on smartphone-sensed data.

Although these approaches represent a significant advance in the analysis of longitudinal data, methodologies that combine context characterization and longitudinal exploration of behavior at both between and within-person levels are scarce, and the extraction of interpretable knowledge is still a challenge. For that reason, in this work, a methodological approach for the analysis of longitudinal data has been designed, which employs data-science techniques to provide a novel point of view on how the dynamic nature of mood can be assessed. The proposed methodology is aimed at helping researchers to address such a complex problem, leading to a more granular exploration of the in-context between and within-person affective fluctuations. In particular, researchers in sport science can use this methodology to uncover affective profiles among elite athletes and examine their individual affective fluctuations, thus relating them to their average performance and being able to promote affective and behavioral interventions that optimize the subsequent performance.

5.2 Proposed multilevel analysis methodology

The analysis methodology proposed in this chapter has been designed and tested based on the dataset resulting from the first experiment conducted in this thesis (Section 4.3). In that experiment, twenty-two participants—both athletes and sedentary people—were monitored for two weeks using the monitoring platform designed in this thesis. Information about their mood and context were continuously gathered during their daily life using ESM self-reports and smartphone-sensed data, respectively. In addition, an initial survey with information about average anxiety levels—which is related to the sport performance following the IZOF model—was also fulfilled by all participants. The methodology is based on the concept that aggregating the data at different time frames can provide data that preserves the dynamic nature of the target behavior—in this case, mood—and allows for the extraction of knowledge at different levels.

The analysis methodology starts with a data pre-processing stage. Raw ESM and sensor data are cleaned and transformed into more interpretable features—low-level behavioral indicators. Then, two pipelines of data analysis are proposed. Following the first one, the mood data of each subject is aggregated along the whole study period, and subjects are clustered into mood profiles—i.e., which individuals have similar affective behavior in the long-term. This analysis reveals information about each individual's mood tendencies over time and the between-person differences in affective behavior. Finally, the mood profiles are related to the average anxiety levels of each individual to uncover whether the mood profile of an athlete is related to the average anxiety level, which leads to the optimal performance—in an attempt to look for the IZOF zone. In the second analysis pipeline, the data of each subject is aggregated within a specified time frame—i.e., hours, days. The context information of each time frame is used to cluster them, obtaining a finite number of types with similar contextual characteristics—e.g., several types of days with specific contextual traits. The behavior of each individual in each time frame is explored based on the characteristics of that type of time frame. This analysis helps to uncover the mood fluctuations of each individual and how they behave under each specific context—which changes over time—beyond their identified affective profile. In particular, it will be used to explore which type of day causes significant mood fluctuations in each athlete, which could lead to changes in performance.

It is worth emphasizing that, although the methodology has been built and tested on a specific topic, it is intended to be adapted to accommodate different problems by simply changing some parameters such as the behavioral indicators extracted or the aggregation time frames, according to the expert's knowledge. In Section 5.5, the most relevant aspects of the methodology are discussed, along with its other potential applications.

5.3 Data pre-processing

5.3.1 Data cleaning

The first step of the methodology, before the analysis itself, is to prepare the data. To that end, duplicate ESM entries and erroneous sensor values are removed. Together with the sensor data, the monitoring platform gathers information about the devices' sensor characteristics, including maximum and minimum values. Those values which exceed the maximum sensor value are considered erroneous. A total amount of 158 data were removed. After that, a participants' exclusion criteria is set based on the compliance rate of ESM questionnaires, which requires a high response rate to accurately represent the daily life of individuals in a variety of scenarios. Although there is no gold standard, according to a recent review, most studies point to a minimum response rate close to 80% to be representative (Y. S. Yang et al., 2019). Therefore, participants with less than 80% of answered questionnaires are excluded from the analysis, as well as those who, due to technical errors, did not receive a significant amount of ESM notifications in the client app. From the studied sample, two participants—P09 and P26—were excluded based on these criteria, leading to a sample of 20 subjects.

5.3.2 Extraction of mood, behavioral and contextual indicators

Once the raw data is cleaned, it is transformed into behavioral indicators which better represent the measured behavior and context and increase the interpretability of the results obtained during the data analysis stage. The behavioral indicators are extracted based on the aggregation time frames selected for each type of analysis. In this particular case, for the group-level analysis—between-person—, the data of each subject is aggregated during the whole study period—two weeks—, and behavioral indicators are extracted based on the aggregated data. In the individual-level analysis—within-person—, a time frame of one day was selected, so the indicators are extracted from the data of each subject during each day—daily aggregated data. Determining the aggregation time frame is a crucial step of the methodology, usually selected based on expert knowledge. It will depend on the data nature and the desired granularity of the analysis. For this sample, and our research topic, a time frame of one day was selected due to the frequency of the athletes' practices. As they perform one training session per day, their mood will likely fluctuate following that pattern. As a result of the data aggregation, two different datasets are obtained: the first one—group-level—contains one row per participant with the aggregated data of each one for the whole study period; the second one—individual-level—contains one row per participant and study day—e.g., for 20 participants and 15 study days, it will have 300 rows. In the following, the procedure followed to extract the behavioral indicators of mood and context is described.

Affective behavior indicators

The indicators extracted to characterize the mood data—valence and arousal self-reports—must represent the dynamic nature of these data. For this data, six frequently studied measured of affect dynamics were extracted (Dejonckheere et al., 2019). Although that work proves that complex affect dynamics beyond the mean and standard deviation add limited information to the prediction of psychological well-being, the authors also argue that these dynamic measures do have a practical value when describing emotional trajectories. They indicate that these measurements provide meaningful

Table 5.1 List of behavioral indicators extracted from the mood self-reported data.

Behavioral indicator (name)	What does it represent?	Mathematical description
Mean (M)	Average levels of emotion	Sum of all measures divided by the total number of measures
Standard deviation (SD)	Emotional variability	Sum of all squared differences between each measure and the mean level of that particular dimension, divided by the total number of measures minus one
Relative standard deviation (RELSTD)	Relative emotional variability	SD of an affect dimension divided by the maximum possible SD of that dimension, given its mean level
Mean of successive squared differences (RMSSD)	Emotional instability	Square root of the mean of all squared differences between two successive measures
Correlation valence/arousal (CORR)	Valence/arousal interaction	Within-person correlation between valence and arousal
Auto-regressive slope (AR)	Emotional inertia	Person-specific auto-regressive slope in a multilevel AR(1) model, in which one measure at time t-1 predicts the one at time t

information about mood fluctuations when acknowledging the temporal dimension, which could be overlooked by the mean levels. Table 5.1 summarizes the mood indicators extracted from the raw mood data and their description. Each measure—except the valence/arousal correlation—is computed twice, using valence and arousal values. Therefore, in subsequent steps, a suffix will be added to the indicator names to differentiate between valence (_V) and arousal (_A).

Behavioral and contextual indicators

The indicators about the behavior of the individuals and the context surrounding them were extracted from the raw sensor data. To that end, the layered context inference model described in Section 2.2 has been used. It must be taken into account that low-level features were extracted as behavioral indicators in most cases. Since these indicators aim not to provide a direct representation of the data but to serve as the input of the subsequent knowledge extraction techniques, there is no need for a more complex, higher-level aggregation. In the following paragraphs, each contextual aspect represented by the data is described, along with the procedure followed to extract the corresponding behavioral indicators. For a complete list of the indicators extracted, see Table 5.3.

Ambient light Exposure to ambient light during the day has an essential effect on mood (Blume et al., 2019). It can affect mood by modulating the availability of neurotransmitters involved in mood regulation, such as serotonin, and by modulating the circadian rhythms, which has an effect on mood fluctuations. The mood is sensed using the ambient light sensor available in smartphones, which collects the ambient light intensity—measured in lux—in every sample. Based on the literature, the main statistics of the raw light data—mean, maximum, minimum, and standard deviation—during the corresponding time frames are extracted as ambient light exposure indicators.

Sleep The daily sleep period of an individual can be inferred in several ways. Complex techniques use a combination of several smartphone sensors such as accelerometer, light sensor, and microphone to estimate when the person is sleeping. These techniques usually require the use of machine learning algorithms that need a ground truth to be trained—e.g., a sleep-detection wristband based on heart rate. For that reason, some researchers have developed sleep detection algorithms that rely on easier parameters, such as the interaction with the smartphone, to infer the bed and wake times with high accuracy. It is based on the fact that, nowadays, the last thing that most people do before going to sleep and the first thing they do after waking up is to check the phone. Ciman and Wac (Ciman and Wac, 2019) hypothesized that it is possible to understand and estimate the sleep habits of an individual by using only information related to the interaction with the smartphone screen. In particular, they developed the iSenseSleep algorithm, which allows for estimating the sleep duration based on the screen interactions. While not intended to facilitate real-time sleep detection, this algorithm is a great technique to facilitate longitudinal offline assessment using a cost-effective sensor. In this work, the daily sleep duration was computed using a combination of the iSenseSleep algorithm and the dark periods detected with the light sensor to enclose the possible sleeping hours.

The first step of the algorithm is to label the *screen events* during all the study duration. The monitoring platform gathers four types of screen events codified with numeric values, together with the timestamp of the event—0 = screen turned off; 1 = screen turned on; 2 = screen locked; 3 = screen unlocked. Since they are not used in the algorithm, the lock events are removed and those events that occur twice in a row—first register kept. After that, the rest of the events are labeled as *SCREEN_OFF*—code 0, when the screen turns off, regardless if it's deliberate or not—, *SCREEN_ON*—code 1, when the screen is turned on—and *SCREEN_PRESENT*—code 3, when the screen is unlocked and ready to be used. There are some cases in which the sequence of events might be erroneous. For example, it does not make sense that a SCREEN_PRESENT event is followed by a SCREEN_ON one—the screen should always turn on before the unlock. In those cases, if this error is following a SCREEN_OFF event, the event order is reversed; if not, they are removed. Also, if a SCREEN_PRESENT event is surrounded by two SCREEN_OFF events, it is changed to SCREEN_ON, since the screen cannot be unlocked without turning it on.

The second step is to extract the *screen tuples*. It is denoted as a tuple the ensemble of a SCREEN_ON and a SCREEN_OFF event—individual interaction with the smartphone—, which could have a SCREEN_PRESENT event in the middle or not. Figure 5.1 represents the two types of screen tuples. The tuples are extracted using an iterative process, and the timestamp of each event within the tuple is stored.

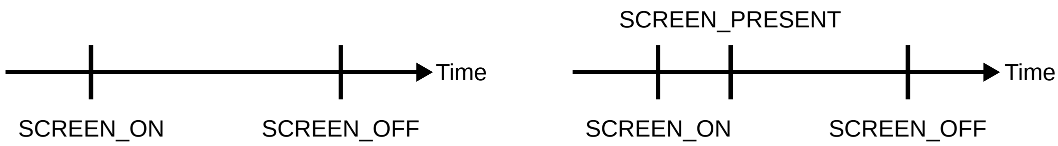


Figure 5.1 Screen interaction tuples extracted using the iSenseSleep algorithm (Ciman and Wac, 2019).

In the third place, the ambient light data is used to extract the *dark bouts* detected. A *dark bout* is defined as a period of time during which the user was not exposed to light and, therefore, the light values detected by the phone are close to zero. To that end, the light data of an individual is resampled to

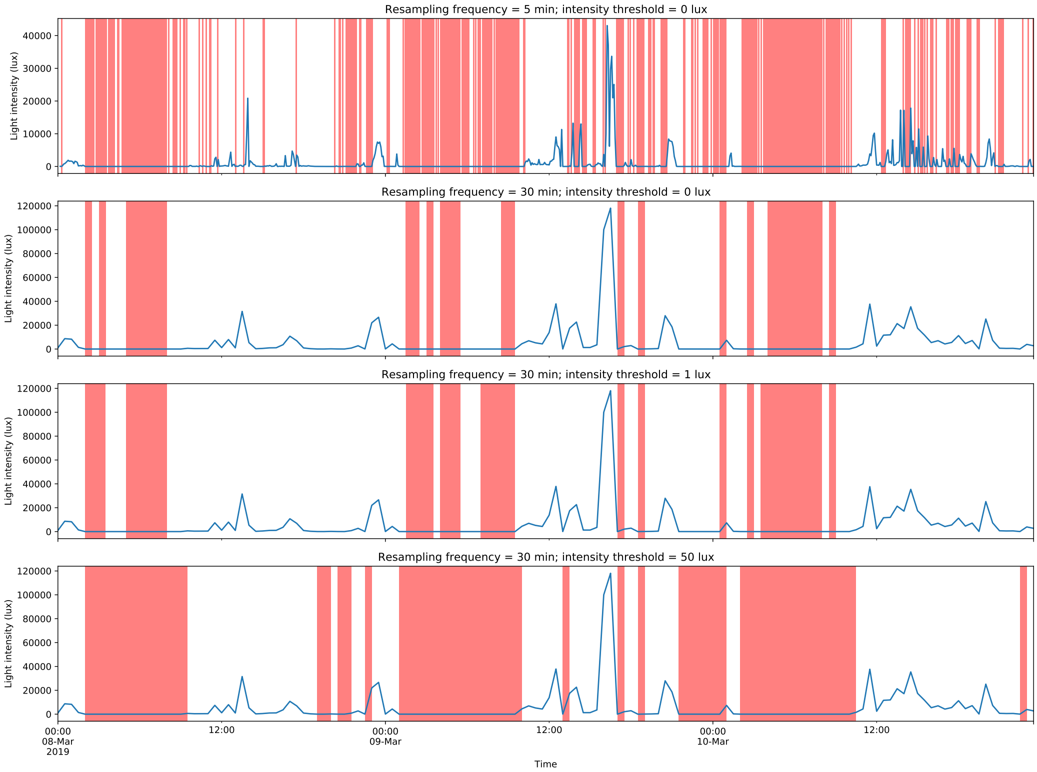


Figure 5.2 Detection of dark bouts for participant P04 during three study days. Each chart shows a combination of a different resampling frequency and intensity threshold. The blue line represents the aggregated light values and the red spans, the detected dark bouts.

a lower sampling frequency, and the periods during which the value keeps under a threshold are selected as dark bouts. By modifying the resampling frequency and the threshold, the sensitivity of the detection can be adjusted. Figure 5.2 shows an example of dark bouts detection for participant P04 during two study days. The blue line represents the aggregated light values, and the red spans, the detected dark bouts. It can be seen that when the resampling frequency is increased, the aggregation causes very short dark periods to surpass the intensity threshold, thus reducing the number of dark bouts detected and just keeping the longer ones, which are the candidates to be sleeping periods. Likewise, more samples are included within the dark bouts when the intensity threshold is increased, thus merging the smaller ones into more homogeneous dark bouts. In this study, the optimal resampling frequency and intensity threshold were empirically established to 30 min and 50 lux, which lead to the most homogeneous dark bouts.

Finally, the dark bouts are used as the enclosing time frames to detect sleeping periods, based on the hypothesis that people mainly sleep within dark environments. According to the iSenseSleep algorithm, the screen tuples separated more than 4 hours are extracted since these are long periods with no interaction with the phone and may be sleeping periods. Using the remaining tuples, the first screen tuple with a SCREEN_PRESENT event within a dark bout is considered the last interaction with the smartphone before going to sleep. The SCREEN_OFF event of that tuple is defined as the *bedtime*. Likewise, the last screen tuple within the dark bout with a SCREEN_PRESENT event is considered the first interaction with the phone after sleep, so the SCREEN_ON event within that tuple is defined as the *waketime*. However, since the individual can check the phone after waking up without unlocking it, the waketime is refined based on the detected phone alarm notifications.

In the end, this algorithm leads to reliable estimation of the wake and bedtimes and, although it may exist some offset between the actual bed and waketimes and the ones detected, the margin of error is acceptable since this methodology does not look for a precise bed and wake times, but an estimation of the daily sleep duration and sleep schedule patterns. Table 5.2 shows an example of some resulting sleep periods detected for two participants.

Table 5.2 Example sleep periods extracted for two study participants.

ID	Bedtime	Waketime	Duration
P03	2019-03-08 00:35:40+01:00	2019-03-08 08:00:01+01:00	0 days 07:24:21
P03	2019-03-09 00:49:08+01:00	2019-03-09 08:00:01+01:00	0 days 07:10:53
P03	2019-03-10 01:27:11+01:00	2019-03-10 08:00:01+01:00	0 days 06:32:50
P03	2019-03-10 22:56:32+01:00	2019-03-11 08:44:54+01:00	0 days 09:48:22
P03	2019-03-10 23:47:07+01:00	2019-03-11 08:00:00+01:00	0 days 08:12:53
...	...	...	...
P10	2019-03-22 01:48:25+01:00	2019-03-22 07:53:09+01:00	0 days 06:04:44
P10	2019-03-23 00:12:01+01:00	2019-03-23 07:29:21+01:00	0 days 07:17:20
P10	2019-03-24 01:21:12+01:00	2019-03-24 07:22:19+01:00	0 days 06:01:07
P10	2019-03-25 01:04:39+01:00	2019-03-25 05:04:01+01:00	0 days 03:59:22
P10	2019-03-26 00:46:18+01:00	2019-03-26 07:23:14+01:00	0 days 06:36:56

Phone usage According to some studies, a great indicator of an individual's social environment can be inferred from the information related to smartphone usage. In this study, several indicators have been extracted from the information gathered by the smartphone, based on recent literature (LiKamWa et al., 2013). The number of phone unlocks per minute and the average time that the screen has been on during each unlock are extracted from the screen interaction events. Similarly, the battery usage information is used to extract the average battery charge and discharge time, the maximum and minimum battery levels reached, and the battery level's mean and standard deviation values. The information about the app usage is used to extract the number of unique apps opened, the rate of apps openings that belong to social media apps, and the number of app openings weighted by the unique apps opened. Finally, the number of notifications coming from social media apps is also extracted. The following apps are considered social media apps: WhatsApp, Telegram, Instagram, Facebook, Twitter, Tuenti, TikTok, Hangouts, and Messenger.

Location In this work, some indicators about the user's location are inferred based on the detected WiFi access points. First, the periods during which the smartphone was connected to a WiFi network are directly considered periods during which the individual was located indoors. Then, those periods with no WiFi connection but with WiFi networks detected are analyzed to determine when the user is more likely to be outdoors. That way, the average time spent indoors by the individual is computed. Additionally, the number of unique WiFi access points connected is extracted as an indicator of the number of indoor places visited, which represents the individual's mobility.

5.4 Data analysis

After the pre-processing stage, two independent datasets are generated: one with the aforementioned behavioral indicators extracted for the whole study period—two-week aggregation—, and the other with the same indicators, but extracted for each selected short-term time frame—in this case, daily indicators. In this section, the two analysis pipelines are described, and each

Table 5.3 List of behavioral indicators extracted from the smartphone sensor data.

Context element represented	Indicator name	Description
Ambient light	light_sum	Sum of the light values
	light_mean	Mean of the light values
	light_std	Standard deviation of the light values
	light_max	Maximum light value during the specified period
	light_min	Minimum light value during the specified period
Sleep	sleep_bedtime[1]	Daily bedtime
	sleep_waketime[1]	Daily waketime
	sleep_duration[2]	Duration of night sleep
Phone usage	screen_unlocks_pm	Average number of screen unlocks per minute
	screen_usage_time_avg	Average duration of the screen usage periods
	apps_events	Number of app openings
	apps_unique	Number of unique apps opened
	apps_social_rate	Rate of apps which are social media
	notif_social	Number of social media app notifications received
Phone battery state	bat_sum	Sum of durations of battery charges
	bat_charge_mean	Mean duration of battery charges
	bat_charge_std	Standard deviation of durations of battery charges
	bat_charge_max	Maximum duration of battery charges
	bat_charge_min	Minimum duration of battery charges
	bat_charge_rate_mean	Mean rate of battery charges
Location	loc_time_indoors	Amount of time spent indoors
	loc_unique_ssid	Unique WiFi networks connected

[1] Only available for daily aggregated data
[2] For the full study data, the average sleep duration is computed—*sleep_time_avg*

dataset is analyzed in its corresponding pipeline. It is important to emphasize that each analysis procedure starts with a feature selection stage. Datasets with data about human behavior and context usually include multiple indicators to describe each facet or aspect from the behavior—e.g., one set of

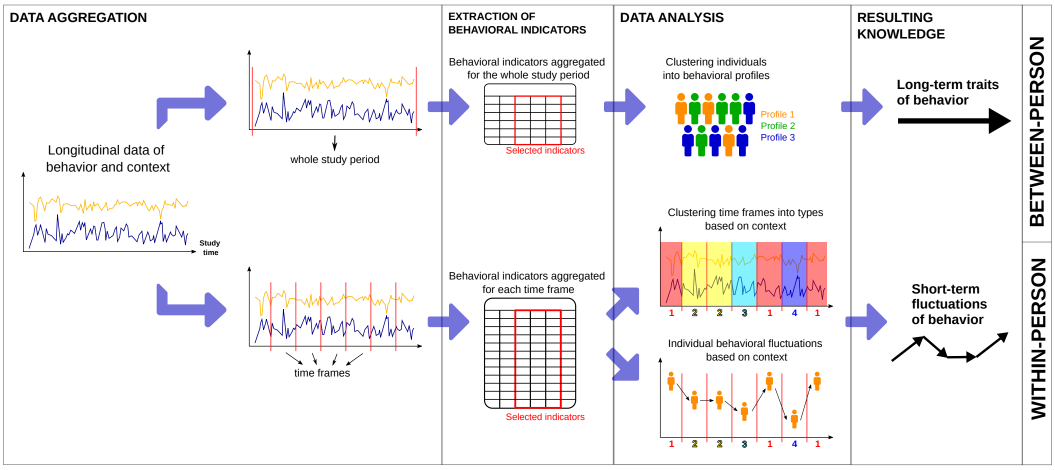


Figure 5.3 Summary of the proposed data analysis methodology.

indicators is extracted to represent mood variability, another set is extracted to represent the ambient light exposure, etc. If some of these indicators are redundant, the performance of the subsequent knowledge extraction techniques applied to the data could be decreased (Guyon et al., 2003). Furthermore, if a smaller subset of indicators is kept, the interpretability of the results may be increased. That is why each analysis pipeline starts with a feature selection step, in which the relations among the indicators are examined, removing those with little to no information supply. Figure 5.3 shows a representation of the two analyses pipelines, illustrating how two types of analyses can be performed using the same starting data.

5.4.1 Between-person analysis: long-term mood profiles

The first analysis pipeline is intended to extract knowledge from the point of view of the long-term mood of the subjects and the differences among them—group-level analysis. Since the initial data should be a set of measures representing the subjects' behavior, mood, and context across the whole study, the first dataset—with indicators extracted for the whole study period—is used. First, in a feature selection step, the relationships among the indicators are explored to find and remove the redundant ones. Then, the subjects are clustered based on the mood indicators. Each cluster represents a long-term mood profile or class, which describes how the subjects within this class broadly behave in terms of mood. The characteristics of each profile are uncovered using statistical tests. Finally, the obtained profiles are compared with the average anxiety values of each subject obtained through the intake questionnaire. This analysis helps to understand how the mood of each subject behaves in the long term and the similarities among the subjects within the same profile. Moreover, the profiles are used to determine whether the different long-term mood behaviors correspond to different arousal zones of optimal performance—according to the IZOF model. Regarding this particular topic, mood profiling has been applied across several contexts, with the sport and exercise domain being the most popular one (Parsons-Smith et al., 2017; Quartiroli, 2018). However, most of the existing research has extracted mood profiles from one-shot mood questionnaires such as POMS (McNair et al., 1992) or BRUMS (Terry, 2003). As no longitudinal mood data is used, the dynamic nature of mood is ignored, thus losing highly relevant information. During this analysis, note that since the mood profiles are extracted based exclusively on the target behavior studied—i.e., mood—, only the mood indicators are used, thus excluding the contextual indicators from the feature selection process.

Feature selection

Several techniques and methods to select the most relevant features from a dataset can be found in the literature. However, when the data represents real-world information—as behavioral, contextual, and affective data do—, its usually complex structure makes it challenging to discern redundant or noisy indicators from those that provide useful but less structured information. It is even more evident when the objective of the analysis is exploratory rather than a prediction or classification algorithm since every piece of data could hold important information. Therefore, before applying a feature selection technique, it is necessary to understand the relationships among the indicators and the data structure. To that end, this methodology employs a feature selection process starting with a data exploration step using a combination of linear correlation measures and Principal Component Analysis (PCA). After the data have been explored and its structure has been understood, the indicators are ranked using an unsupervised feature selection algorithm, and the less informative ones are removed based on the previous exploration. In the following, the two steps of the feature selection stage are described.

Data exploration. First of all, the mood indicators are normalized using standard normalization. Then, a matrix of Pearson's correlations among them is computed to provide a first description of their relationships. A heatmap representing the correlation matrix is depicted in Figure 5.4. It can be seen that the indicators which capture mood variability—SD, RELSTD and RMSSD—are strongly interconnected, with higher correlation values for the indicators of the same affect dimension ($r >= 0.75$, $P < 0.001$, $CI = 0.46$ to 0.89), and slightly lower ones between indicators of different affect dimensions ($r >= 0.59$, $P < 0.01$, $CI = 0.21$ to 0.82). Another indicators with relevant correlations are the mean level of arousal, M_A, and the auto-regressive measure of valence, AR_V, both showing strong negative correlations with all the variability measures of arousal ($r <= -0.51$, $P < 0.05$, $CI = -0.78$ to -0.01). It is also worth mentioning that the measure of correlation among both affect dimensions, CORR, has little associations with the rest of indicators ($|r| <= 0.33$, $P < 0.1$, $CI = -0.67$ to 0.13).

These results show that the indicators that characterize a particular aspect of mood—e.g., mood variability, average mood, etc.—are usually highly related, which could mean that they overlap and provide redundant information. To reaffirm this hypothesis, an additional step is taken towards exploring the data, this time using PCA. This technique is used to transform the data into a reduced set of features known as principal components (PCs), which are generated as linear combinations of the original features. Therefore, each PC is composed of the sum of the original features, each one weighted by a coefficient—called loading. In this analysis, these loadings are used as a measure of the effect of each behavioral indicator on the PCs. If a set of indicators have high loadings on a specific PC, it means that they could be summarized by just one feature—that PC—, so a considerable overlap exists among them.

In this work, this technique is not intended to reduce the dimensionality of the dataset but only to check which mood indicators overlap other ones. The number of PCs to retain is determined using a scree plot. For this analysis, a six-component solution was obtained, which explains 91% of the total variance of the dataset. For better visualization, the PCA solution is rotated

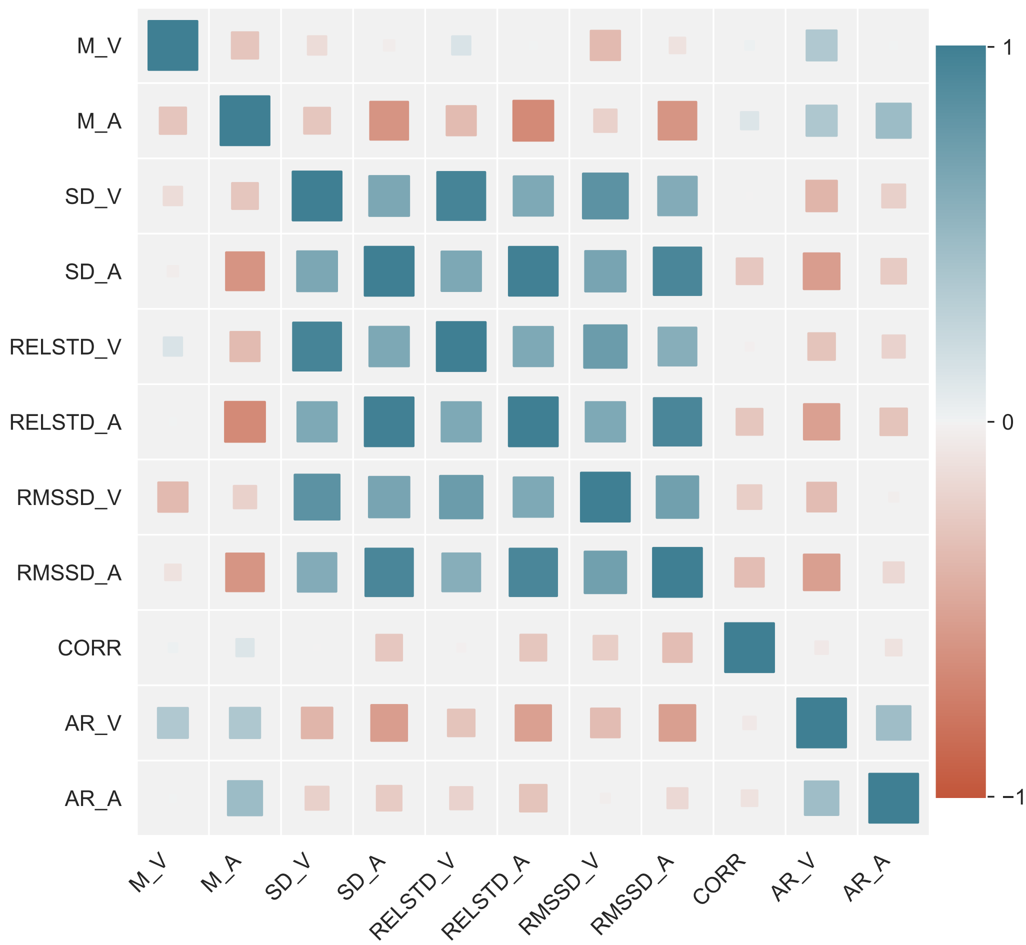


Figure 5.4 Matrix of Pearson's correlation among all the mood indicators extracted from the whole study period data. The colour and the size of each square represents the direction—blue - positive; red - negative—and the intensity of the correlation.

using an orthogonal Varimax rotation (Wu, 2014). This technique maximizes the variance among the indicator loadings, leading to a visualization in which the differences among the loadings are more evident (Figure 5.5).

It can be seen that PC1 and PC2 cover all the mood variability indicators—SD, RELSTD, and RMSSD—of each affect dimension—valence and arousal, respectively. Their loadings over the components are considerably higher than the rest of the indicators, ranging from 0.84 to 0.88 in PC1 and from 0.76 to 0.93 in PC2. The rest of the PCs are only significantly loaded by one indicator each. This result supports the previous findings and reveals that the indicators of mood variability of each affect dimension have appreciable overlap among them. In contrast, the rest of the indicators seem to be

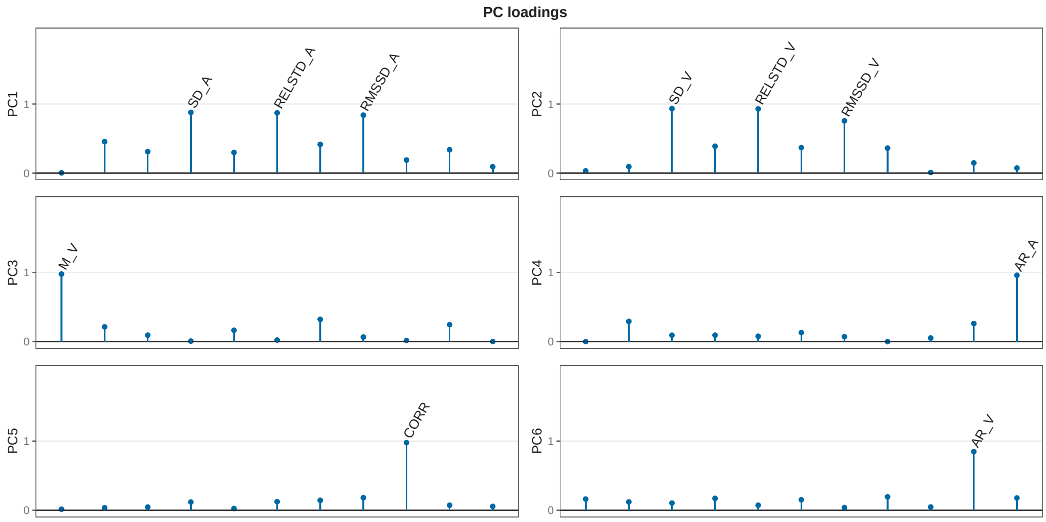


Figure 5.5 Six PCs retrieved from the rotated PCA with each indicator's absolute loading—in the range [0,1]. For a clearer visualization, only the name of the indicators with loadings over 0.6 are displayed.

independent, which means that they provide unique information that cannot be obtained from any other indicator. Therefore, it is necessary to remove redundant indicators, but only from the set of mood variability ones, assessing each dimension independently.

Feature selection. Once the relationships among the indicators have been explored, it is clearer which indicators need for a feature selection process. Choosing a feature selection technique is highly dependent on the type of analysis performed. Since this methodology is intended to analyze the data in an exploratory fashion, there will not generally be labels or scores which could be used to employ supervised feature selection techniques. For that reason, in this work, an unsupervised filter method for feature selection called Laplacian Score (He et al., 2005) is used. This method ranks the indicators based on the amount of information supplied by each one to the dataset. The lower the Laplacian Score of an indicator is, the higher its informativeness will be. After ranking the mood indicators, the rank shown in Table 5.4 is obtained. Based on the previous data exploration, the less informative indicators of mood variability of each affect dimension are removed. In this case, only one indicator per dimension—the one with higher rank—is kept, thus removing *RELSTD_V*, *SD_V*, *RELSTD_A* and *SD_A*.

Table 5.4 Ranking of behavioral indicators extracted from the longitudinal data for the whole study period. Lower scores mean more relevant indicators.

Behavioral indicator	Laplacian score
RELSTD_A	0.351
SD_A	0.362
RMSSD_A	0.397
RMSSD_V	0.450
SD_V	0.489
RELSTD_V	0.537
AR_V	0.560
AR_A	0.563
M_A	0.585
M_V	0.609
CORR	0.747

Extraction of behavioral profiles

Once the most relevant indicators have been selected, they are used to group all the participants into mood profiles. To that end, an agglomerative hierarchical clustering approach (Rokach and Maimon, 2006) is employed. It starts from one-participant groups and recursively merges them into bigger clusters based on their similarity. Therefore, solutions with a different number of clusters can be obtained by modifying the level of aggregation. To perform the clustering, several parameters must be selected based on the data nature: the distance measure among the clustering elements, the elements' linkage criteria, and the clustering evaluation measure. In this work, and based on empirical tests, the Euclidean distance is used to compute the distance among the elements. For the linkage criteria, Ward's method is used since it is more appropriate when the data is less structured, as is the case. Finally, the evaluation measure selected to choose the most appropriate number of clusters is the silhouette coefficient (Rousseeuw, 1987). This measure, which ranges from -1 to 1, represents how similar an element is to the rest of the cluster members—cohesion—compared to the members of the rest of the clusters—separation. Hence, a value of -1 means that the measured element is too much different from the other elements of its cluster—and therefore, it should not belong to that cluster—, while a value of 1 represents an element perfectly matched to its cluster. The silhouette coefficient of a cluster is computed as the average silhouette coefficient of all its elements.

After the clustering is applied, the obtained clusters can be understood as the different mood profiles within the study population. The result is illustrated in Figure 5.6. It shows a dendrogram in which the participants are represented on the X-axis and the aggregation level along the Y-axis. It is represented how, when the aggregation level is increased, the clusters with high similarity are merged, obtaining fewer but bigger clusters. The resulting clustering solutions, their elements, and silhouette values are summarized in Table 5.5.

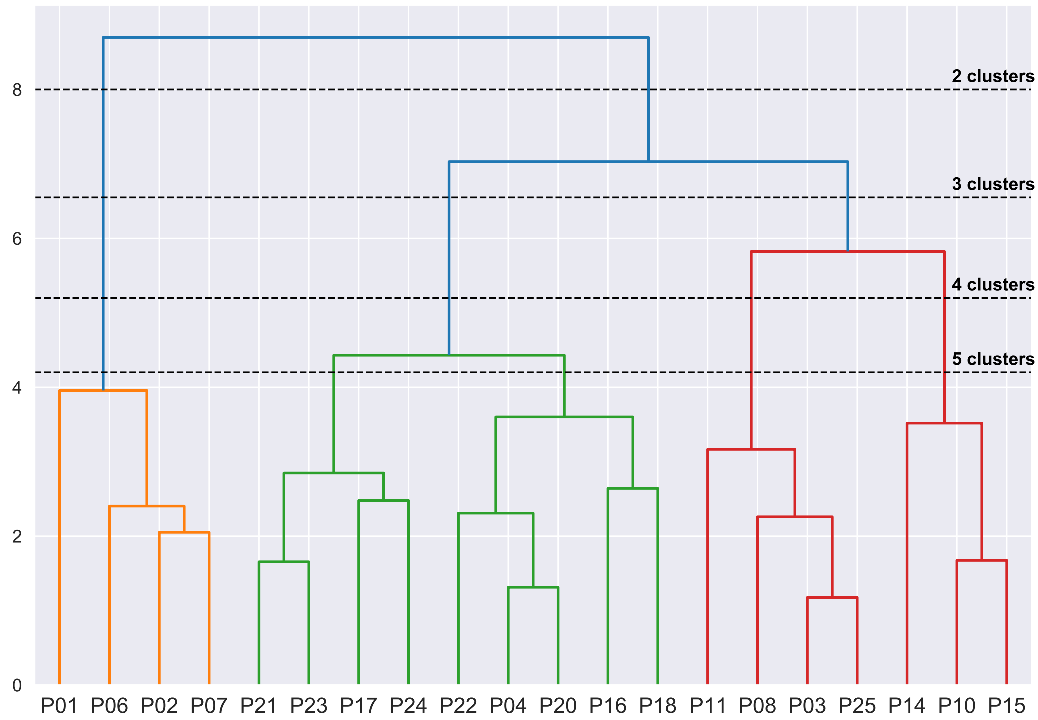


Figure 5.6 Hierarchical clustering result for the whole study data. The participants are represented along the X axis, and the dendrogram shows the agglomerative clusters. The black dotted lines show examples of aggregation levels that lead to solutions of 2, 3, 4 and 5 clusters.

The selection of the appropriate number of clusters is a subjective task. Although evaluation measures—such as silhouette coefficient—can serve as qualitative support, this choice largely depends on the studied problem and the desired granularity of the solution. For example, in this case, although the higher silhouette value is obtained for the two-profile solution—0.266—, one

Table 5.5 Subject distribution among clusters for different number of clusters, along with their average silhouette coefficient.

Number of clusters	Cluster participants	Silhouette coefficient
2	1 → P01, P02, P06, P07	0.266
	2 → P03, P04, P08, P10, P11, P14, P15, P16, P17, P18, P20, P21, P22, P23, P24, P25	
3	1 → P01, P02, P06, P07	0.226
	2 → P04, P16, P17, P18, P20, P21, P22, P23, P24	
	3 → P03, P08, P10, P11, P14, P15, P25	
4	1 → P01, P02, P06, P07	0.233
	2 → P04, P16, P17, P18, P20, P21, P22, P23, P24	
	3 → P03, P08, P11, P25	
	4 → P10, P14, P15	
5	1 → P01, P02, P06, P07	0.193
	2 → P04, P16, P18, P20, P22	
	3 → P17, P21, P23, P24	
	4 → P03, P08, P11, P25	
	5 → P10, P14, P15	

of the profiles contains 80% of the elements, which results in an unbalanced solution. Since a higher number of divisions is preferred for a more granular aggregation, the four-profile solution was chosen. In addition to the higher silhouette value than the three-profile one—0.233 vs. 0.226—, it allows for a greater variety of profiles.

After the mood profiles are obtained, they are characterized based on the mood indicators to understand the affective behavior of each group. First, a Kruskal-Wallis statistic test is used on each indicator to determine which one shows significant differences among two or more profiles. Due to the small sample size, this non-parametric test is used as an alternative to one-way ANOVA. Next, a post-hoc Tukey test is employed to check the mean differences of the indicators' values among the profiles. Therefore, the affective characteristics of each profile can be identified. Table 5.6 summarizes the results of the statistical tests. Table 5.7 contains the traits of each mood profile found in this analysis. Finally, Figure 5.7 provides a visual representation of the indicators' values of each participant and profile.

Table 5.6 Results of the application of Kruskal-Wallis test—adjusted for ties—and post-hoc Tukey test to the selected behavioral indicators of the whole study data. Indicators with p-value below 0.05 are marked in bold. For those indicators, the significant differences in median value among the four behavioral profiles extracted are listed in the last columns.

	Kruskal-Wallis test		**Post-hoc Tukey test**		
Behavioral indicator	H statistic	p-value	Profiles with significant differences	Pairwise mean difference	Adjusted p-value
M_V	13.57	**0.0035**	1 - 4 2 - 3 2 - 4	11.3472 10.9552 16.2116	0.0124 0.0024 0.001
M_A	11.61	**0.0088**	1 - 2 1 - 4 2 - 4 3 - 4	-9.1908 -19.2468 -10.056 -12.6011	0.0263 0.001 0.0286 0.0159
RMSSD_V	13.6	**0.0035**	1 - 2 1 - 4 2 - 4 3 - 4	-9.1908 -19.2468 -10.056 -12.6011	0.0263 0.001 0.0286 0.0159
RMSSD_A	9.92	**0.0193**	1 - 2 1 - 3 2 - 4 3 - 4	13.415 13.3726 -9.0264 -8.9839	0.001 0.001 0.0056 0.0161
CORR	4.8	0.187	-	-	-
AR_V	10.98	**0.0118**	1 - 2 2 - 3	-0.2068 0.201	0.0121 0.0148
AR_A	11.96	**0.0075**	-	-	-

Table 5.7 Long term traits of the behavioral profiles found among the participants.

Profile	Long term traits
1	Very high emotional stability (RMSSD↓↓) and high average activation (M_A↑)
2	Low average happiness (M_V↓) with high variability (RMSSD_V↑ and AR_V↓)
3	High average happiness (M_V↑) but slightly unstable (RMSSD_V↑)
4	Very high average happiness (M_V↑↑) and very low average activation (M_A↓↓)

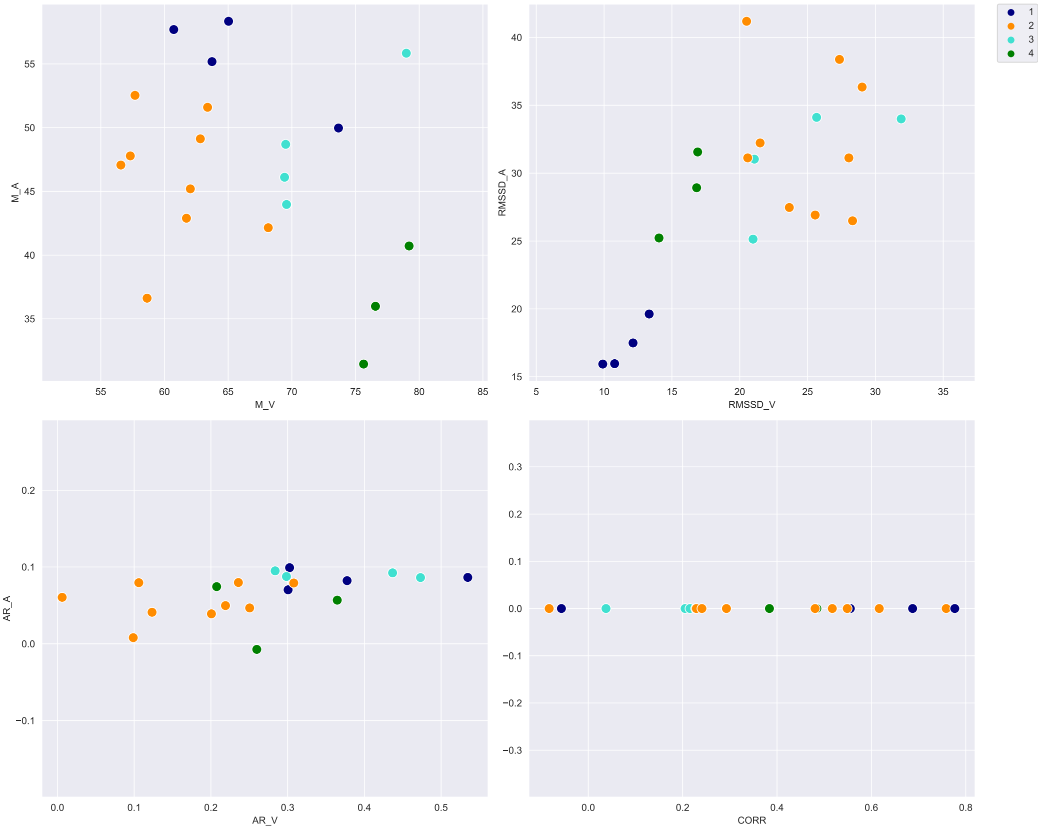


Figure 5.7 Scatterplot of the behavioral indicators of mood, computed for each participant during the whole study period. The colors represent the participants of each behavioral profile.

Relationship between mood profiles and the sport performance

Finally, the mood profiles are compared with the sport level of the participants to check whether the long-term mood tendency of elite athletes—P01, P02, P06, P07, P08, P10, P11, P14, and P15—differs from the one of sedentary people—P03, P04, P16, P17, P18, P20, P21, P22, P23, P24, P25. Table 5.8 summarizes the number of participants of each group within each profile. It can be seen that profiles 1 and 4 belong exclusively to athletes, and profile 2 contains only sedentary participants. On the contrary, participants with mood profile 3 are both athletes and sedentary. Although the sample size is too small to generalize these findings, athletes seem to have different emotional behavior than sedentary people in this specific population. According to the affective traits of each profile (Table 5.7), the two mood profiles of elite athletes—1 and 4—are characterized by relatively higher happiness levels than the profile of sedentary people—2.

Table 5.8 Cross-table with the type of participant and the obtained mood profiles.

		Participant type	
		Athlete	Sedentary
Profile	1	4	0
	2	0	9
	3	2	2
	4	3	0

Regarding the two mood profiles of elite athletes, they are mainly differentiated by the arousal dimension. Profile 1 is characterized by higher emotional stability and a high average level of arousal, meaning that these athletes are, in general, more activated and have fewer mood fluctuations. In contrast, profile 4 shows a very low average arousal and a moderate emotional variability, which results in less activated athletes with slightly more emotional fluctuations. This difference in average arousal could agree with two different activation zones for optimal performance—two IZOF zones. To further explore this hypothesis, the average arousal levels of each profile are compared with the average results of the DASS-21 anxiety scale measured from the intake questionnaire. Athletes of profile 1—high average arousal, mean = 55.29, std = 3.8—show an average DASS score of 2.25 out of 21. Athletes of profile 4—low average arousal, mean = 36.05, std = 4.63—have an average DASS score of 6. Although the DASS score difference might be small in absolute terms, there is a relatively big difference between the average anxiety level of both profiles, and it matches the average arousal levels monitored during the study.

5.4.2 Within-person analysis: short-term mood fluctuations based on daily context

The second analysis pipeline of the methodology is aimed to explore the data from the perspective of each individual and their short-term affective variations based on the context. For that reason, this time, the second dataset—with indicators extracted for each study day—is analyzed. First of all, a feature selection step is performed following the same procedure of the between-person analysis, but this time also includes the context indicators. Then, the daily context data is used to cluster them into groups with similar characteristics. Each cluster represents a specific type of day, during which the contextual situation surrounding the individual had particular characteristics—e.g., a day in which the participant had a lot of sleep time, had little social contact, and spent much time indoors. This specific context situation—i.e., this type of day—could have been experienced by one or multiple participants, just once or several times. The characteristics of each day type are determined using statistical tests. Finally, the short-time mood indicators of each subject are explored to see how the mood of each individual behaves during each previously obtained type of day and how each context makes the mood differ from the expected based on the mood profile.

Feature selection

Data exploration. The procedure followed to remove the less informative features of the dataset takes the same steps as the one performed during the between-person analysis (Section 5.4.1). After normalizing the indicators using standard normalization, a matrix of Pearson's correlations is calculated (Figure 5.8). This provides a first overview of the relationships among the indicators. It can be seen that there are three groups of indicators showing considerable correlation among their elements. The first is the set of indicators of mood variability—SD, RELSTD and RMSSD—, whose strong relation is only significant for those measuring the same affect dimension ($r >= 0.72$, $P < 0.001$, $CI = 0.66$ to 0.78). Second, the set of indicators describing the ambient light exposure. Among them, the standard deviation of light values, light_std, stands out due to its high correlation with the rest ($r >= 0.68$, $P < 0.001$, $CI = 0.62$ to 0.75). The last group is the battery usage indicators. Except for the battery charge rate, the rest show strong correlations among them ($r >= 0.49$, $P < 0.001$, $CI = 0.39$ to 0.58).

Again, these findings emphasize that some groups of indicators characterizing specific aspects of mood, behavior, or context could have potential overlaps among their indicators. However, the relationships found in this analysis are slightly different from those obtained during the between-person analysis due to the different aggregation time of the indicators. To further confirm

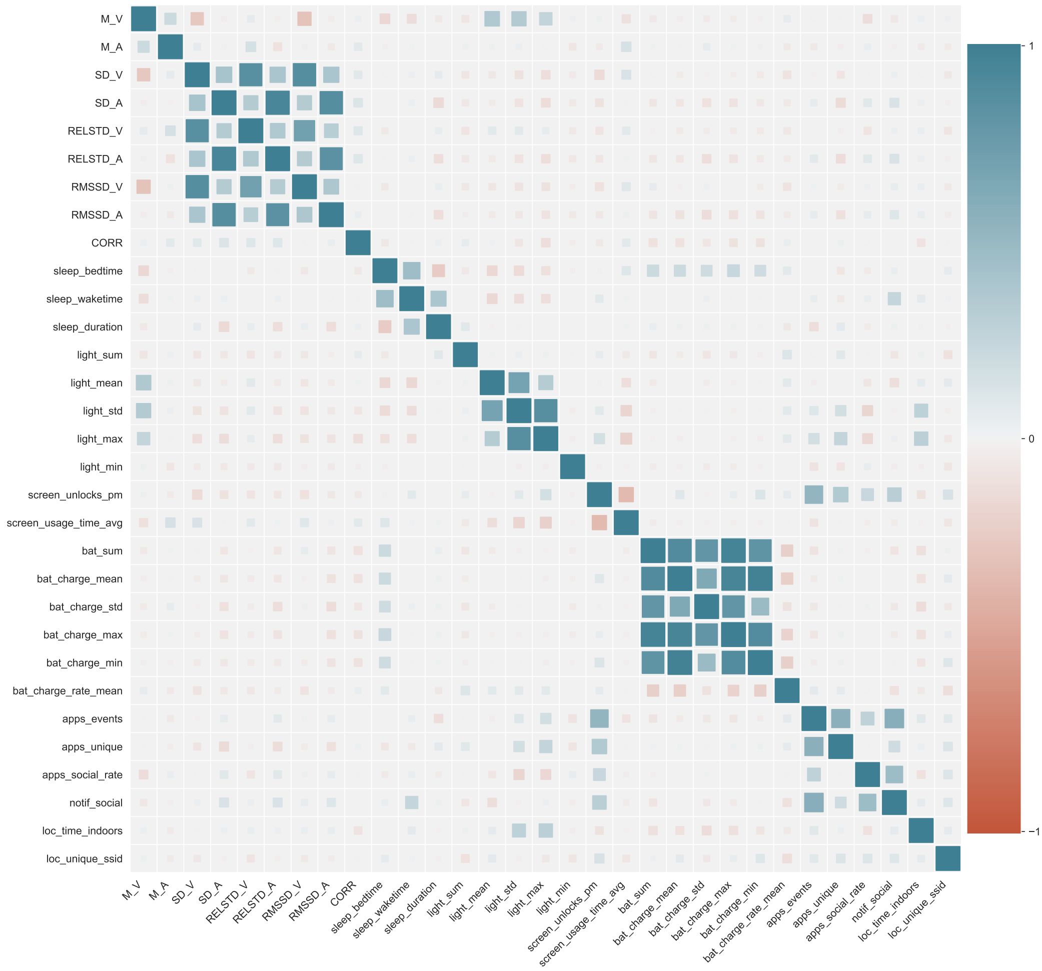


Figure 5.8 Matrix of Pearson's correlation among all the behavioral indicators extracted from short-time—daily—aggregated data data. The colour and the size of each square represents the direction—blue - positive; red - negative—and the intensity of the correlation.

this hypothesis, the data is explored a second time using PCA. In this case, a 14-component solution was obtained using a scree plot. These 14 PCs explain 72% of the total variance of the dataset. The rotated PCA solution is shown in Figure 5.9. A higher number of PCs are only significantly loaded by one indicator, reflecting high feature independence. Regarding those which contain indicator overlap, PC1 covers the indicators of phone battery usage, whose loadings are almost the only ones with a significant value, ranging from 0.79 (bat_charge_std) to 0.99 (bat_charge_max). PC2 and PC3 represent mood variability. Indicators of mood variability of each affect dimension load each of these PCs—arousal and valence, respectively—with values ranging

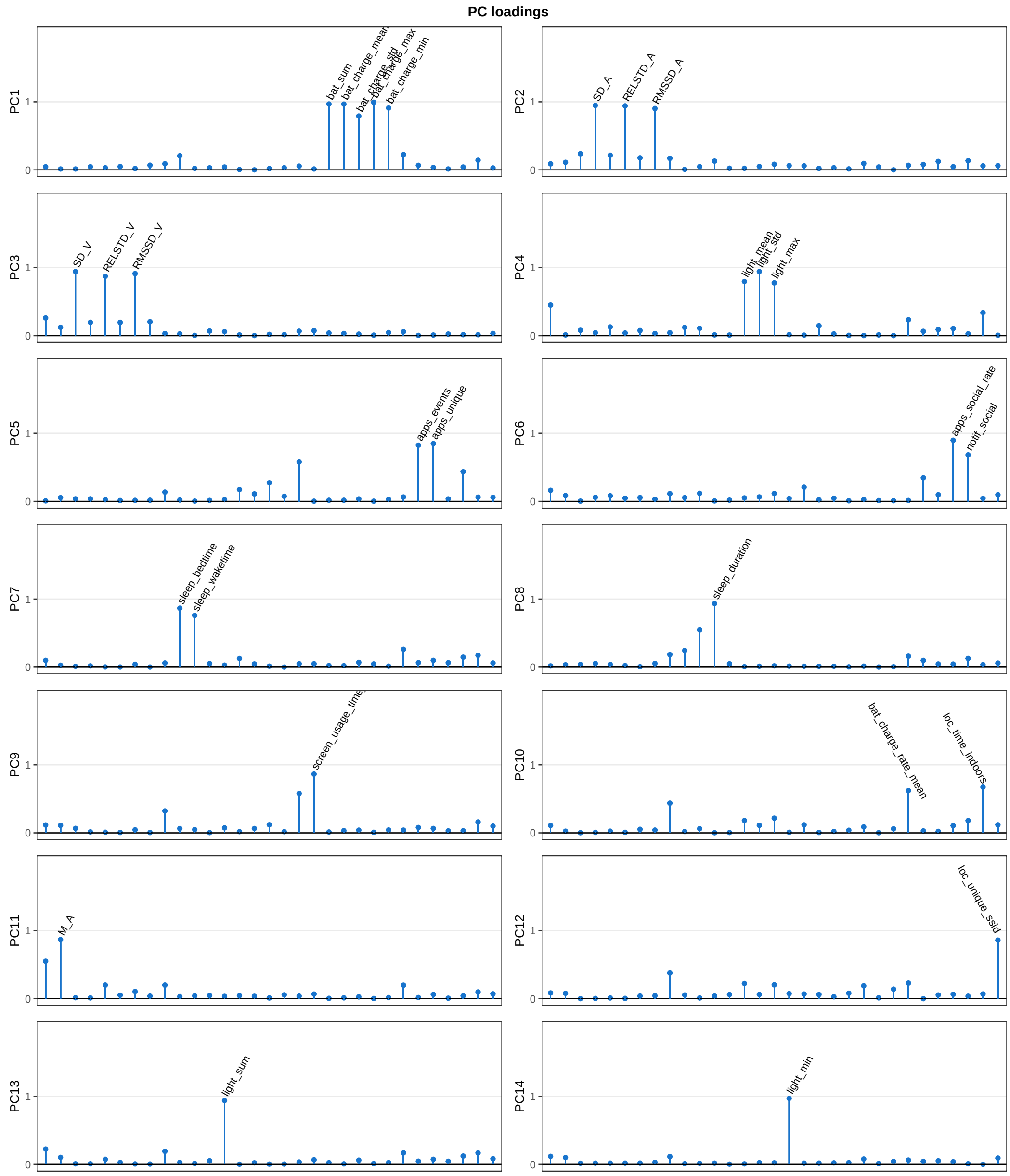


Figure 5.9 14 PCs retrieved from the rotated PCA with each indicator's absolute loading—in the range [0,1]. For a clearer visualization, only the name of the indicators with loadings over 0.6 are displayed.

from 0.87 (RELSTD_V) to 0.95 (SD_A). Ambient light indicators have high loadings in PC4—ranging from 0.78 to 0.94. Finally, PC5 to PC7 and PC10 show little associations among some indicator pairs, and PC8, PC9, and PC11 to PC14 do not exhibit significant relations among the indicators. Again, the sets of indicators needing a feature selection process—phone battery usage, mood variability, and ambient light—agree with the findings of the linear correlation exploration.

Feature selection. Once the data has been explored, and based on the results obtained, the indicators are ranked based on their Laplacian Score. Table 5.9 shows the rank of indicators. Based on the findings of the data exploration, the less informative indicators of each dimension of mood variability, phone usage battery and ambient light exposure are removed: *RELSTD_A*, *RELSTD_V*, *RMSSD_A* and *RMSSD_V*—mood variability—; *bat_charge_min*, *bat_charge_mean*, *bat_charge_std* and *bat_charge_max*—battery—; and *light_min*, *light_std* and *light_mean*—ambient light.

Characterization of context scenarios

Once the most relevant indicators have been selected, the data of all days—regardless of which individual belongs to—is clustered using an agglomerative hierarchical clustering approach. In this case, as the aim is to characterize the context scenarios in which participants were placed during the study, only the indicators related to context are used in this clustering step. The parameters used in the clustering algorithm are the same as in the previous analysis: the distance among elements is computed using Euclidean distance, Ward's method is used as the linkage criteria, and the silhouette coefficient is employed as the evaluation measure.

This dataset contains 266 observations—days—of 14 contextual indicators. After these days are clustered, each cluster can be understood as a particular type of day experienced by one or more individuals within the study population. The resulting dendrogram is illustrated in Figure 5.10, where the days are represented on the X-axis and the aggregation level along the Y-axis.

As in the previous analysis pipeline, the silhouette coefficient of each cluster is used as a supporting measure to select the most appropriate number of clusters. In this case, the silhouette value for solutions from three to eight clusters was found to be very similar—ranging from 0.126 to 0.128. Based on this similarity, the five-cluster solution was chosen because, when the number of clusters is increased, the added clusters were too small—less than three elements—to be relevant. Therefore, five types of days were obtained. The characteristics of each type are uncovered using statistical

Table 5.9 Ranking of behavioral indicators extracted from the longitudinal data for each time frame—day. Lower scores mean more relevant indicators.

	Behavioral indicator	Laplacian score
MOOD	SD_A	0.112
	SD_V	0.126
	RELSTD_A	0.137
	RMSSD_A	0.158
	RMSSD_V	0.161
	RELSTD_V	0.163
	CORR	0.188
	M_V	0.210
	M_A	0.219
CONTEXT	light_max	0.147
	light_std	0.205
	notif_social	0.238
	apps_social_rate	0.267
	light_mean	0.274
	apps_unique	0.276
	screen_unlocks_pm	0.283
	apps_events	0.284
	sleep_waketime	0.298
	loc_unique_ssid	0.319
	screen_usage_time_avg	0.336
	sleep_bedtime	0.336
	sleep_duration	0.390
	bat_charge_std	0.455
	loc_time_indoors	0.474
	light_sum	0.482
	bat_sum	0.499
	bat_charge_max	0.525
	bat_charge_rate_mean	0.529
	bat_charge_mean	0.539
	bat_charge_min	0.541
	light_min	0.826

tests on the context indicators, which show the significant differences among the day types for each indicator. Kruskal-Wallis and post-hoc Tukey tests are used to account for these differences. Although the sample size is larger in this dataset, non-parametric tests are used due to the lack of normality

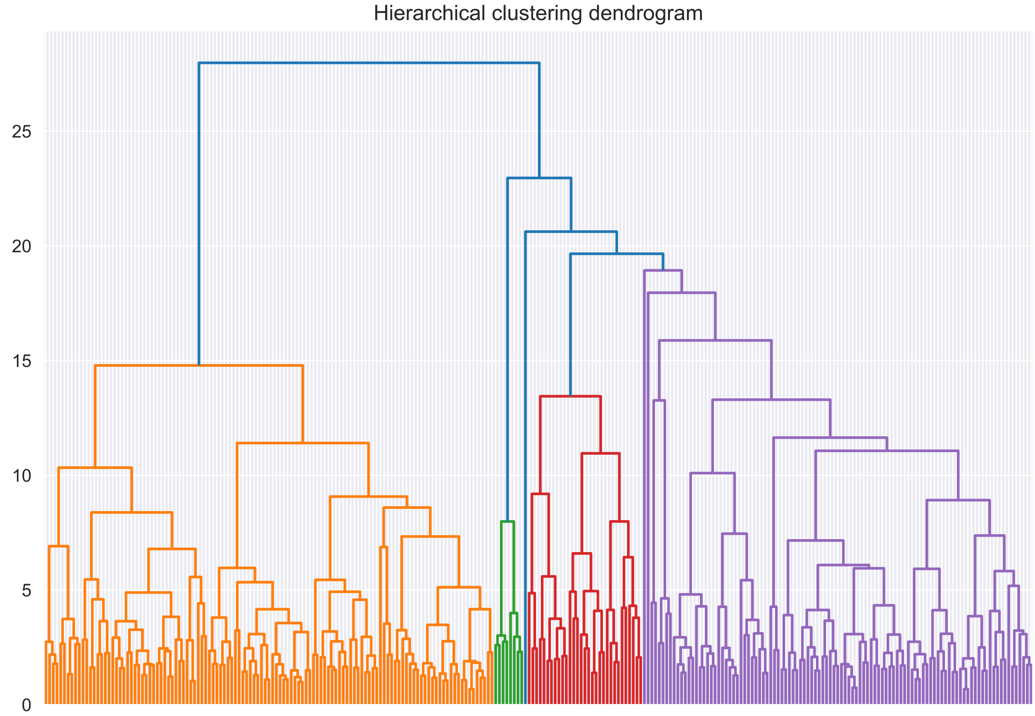


Figure 5.10 Hierarchical clustering result for the daily aggregated data. Each tick along the X axis represents one time frame—day—, and the dendrogram shows the agglomerative clusters. Due to the data density, the X axis labels are not displayed.

in most indicators' data distribution. The description—i.e. the contextual characteristics—of each type of day are summarized in Table 5.10.

Individual affective fluctuations based on context

At this analysis stage, the study days of each participant have been labeled according to their context type found in the previous section. Now, the last step of this analysis pipeline consists of exploring how each individual modifies their affective behavior when exposed to different contexts during their daily life. Each participant is analyzed separately to check if its mood deviates from the expected according to its mood profile—which was identified in Section 5.4.1—when living under specific contextual scenarios.

To do so, the day types experienced by the selected participant during the study period are identified[1]. In order to illustrate the procedure, four par-

[1] Note that all the participants may not have experienced each day type, since the clustering of days was performed using the data of all subjects, to also identify similar days experienced by more than one subject—e.g., P07 did not register any day of type 3.

Table 5.10 Contextual characteristics of the day types found.

Day type	Contextual characteristics
1	Short sleep duration and early wake up time (*sleep_duration*, *sleep_waketime*), a high amount of time spent outdoors (*loc_time_indoors*), and a high number of indoor places visited -not spending much time inside- (*loc_unique_ssid*)
2	Reduced light exposure during the day (*light_sum*), a reduced phone usage (*screen_unlocks_pm*, *app_events*) and a reduced social contact through social media apps (*notif_social*).
3	Short sleep duration and early wake up time (*sleep_duration*, *sleep_waketime*), however this one also shows a high amount of time spent indoors with little to no change of indoor location (*loc_time_indoors*, *loc_unique_ssid*). It also shows an intensive use of the smartphone (*screen_unlocks_pm*, *screen_usage_time_avg*, *apps_events*, *apps_unique*), with considerable social contact (*notif_social*).
4	Large sleep duration (*sleep_duration*), great light exposure during the day (*light_sum*), and a reduced phone usage and social contact (*screen_usage_time_avg*, *notif_social*).
5	Late wake up time (*sleep_waketime*), reduced light exposure (*light_sum*), and a very intensive phone usage for social media contact (*apps_events*, *apps_social_rate*, *notif_social*).

ticipants are analyzed in this work, each one belonging to a different mood profile—P06, P10, P17, and P25. After that, a set of statistical tests are applied to the mood indicators to see if there are significant differences in the participant's mood during each of these day types. Two different statistical tests are used depending on the number of day types experienced by the participant. If only two day types were experienced, the non-parametric Mann-Whitney test is used. If three or more days are compared, a combination of Kruskal-Wallis and post-hoc Tuckey Test is employed. The results of the statistical tests applied to the selected participants are summarized in Table 5.11. Indicators with a *p-value* lower than 0.05—significance level—fluctuate significantly when the participant experiences a change in the context during its daily life.

Finally, the mood of each participant during each time frame is compared to the average mood expected to see based on their mood profile. To that end, Figure 5.11 compares, for each of the selected participants, the values of each mood indicator during each type of day and the values of all the participants of their mood profile. For example, participant P06 has experienced days

Table 5.11 Results of the application of two-sided Mann-Whitney test to the selected behavioral indicators extracted for each time frame of four subjects. The first three rows specify the participants analyzed, their previously obtained long-term behavioral profile and the time frame types experienced by then during the study period. Indicators with p-value below 0.05 are marked in bold.

Participant	P06		P10		P20		P25	
Mood profile	1		4		2		3	
Experienced day types	1,2		1,3		1,5		1,2	
Indicator	**U stat**	**p-value**	**U stat**	**p-value**	**U stat**	**p-value**	**U stat**	**p-value**
M_V	9.0	**0.0306**	1.0	**0.016**	8.0	0.1359	9.0	0.2298
M_A	8.0	**0.0227**	7.0	0.1792	8.0	0.1359	3.0	**0.0322**
SD_V	15.0	0.1362	5.0	0.0923	10.0	0.2234	11.0	0.3558
SD_A	18.0	0.2388	6.0	0.1308	5.0	0.0542	8.0	0.1776
CORR	18.0	0.2388	6.0	0.1308	14.0	0.4663	6.0	0.0978

of type 1 and 2 and belongs to profile 1. Then, the Figure will show, for each mood indicator, a boxplot with the distribution of the indicator values during days type 1, the distribution of the indicator values during days type 2, and the distribution of the values of all participants of profile 1—which is supposed to be the average mood of that participant.

The results obtained for these four participants are described hereafter as an example of the knowledge that could be extracted using this analysis methodology. It can be seen that, as a general rule, the mood of participant P06 matches the expected according to its mood profile—1. However, it deviates from that pattern in days type 2, showing a substantially lower activation. As this participant is an athlete, it could mean that on this type of day, the athlete deviates from its optimal performance IZOF zone. A similar example can be found in participant P10. The behavior shown in days type 1 is considerably different from its typical mood profile—4. While individuals of this profile are expected to show high average happiness values, during these days, participant P10's happiness is substantially lower. In contrast, during days type 3, its affective behavior agrees with the typical of its profile. Participant P20, which experienced days of types 1 and 5, has a mood that matches the expected for its mood profile—2—regardless of the type of day.

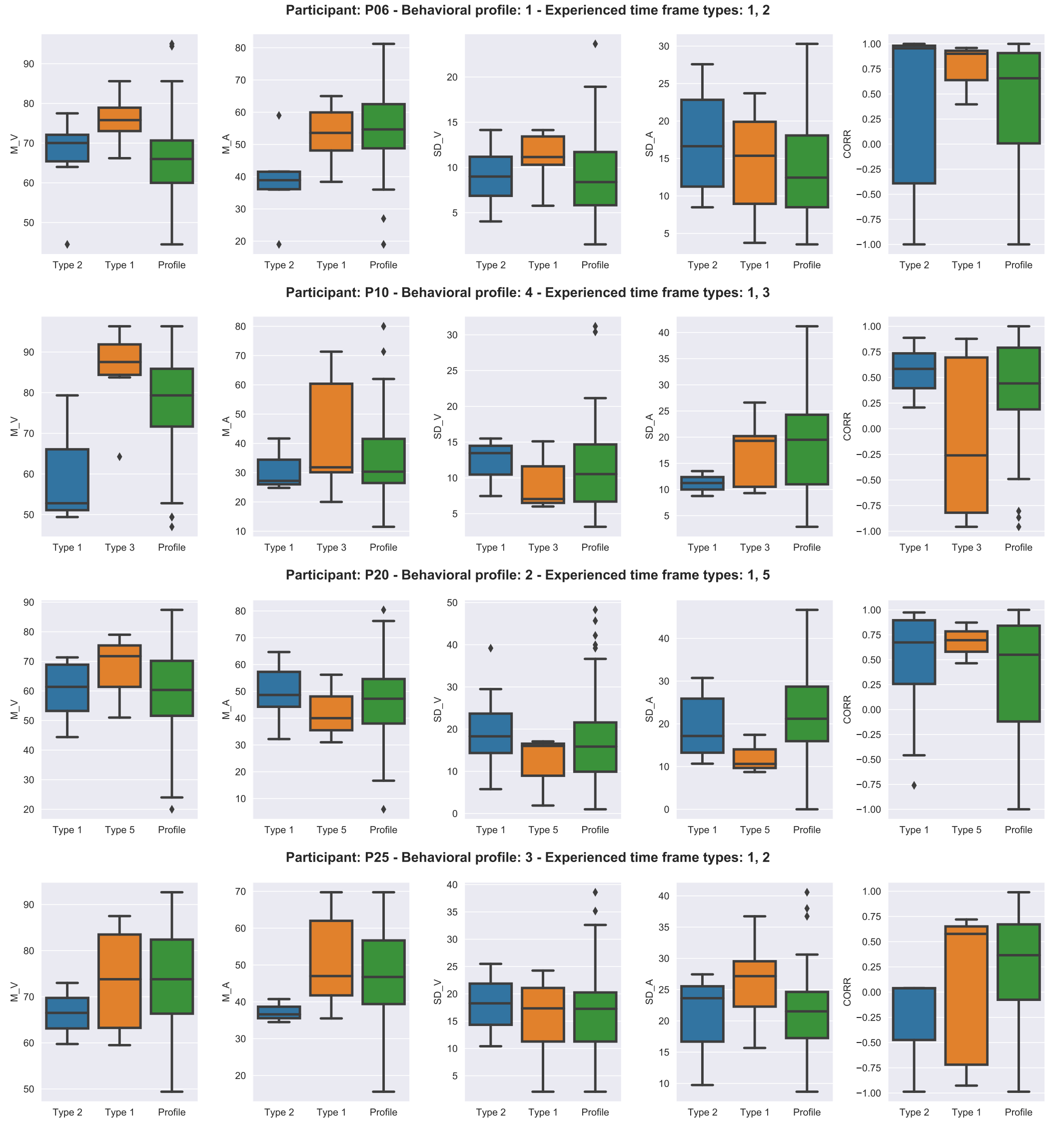


Figure 5.11 Distributions of the participant's value for the selected behavioral indicators. The X axis of the plots show the data represented by each box—blue and orange: indicator values for a specific participant during two time frame types; green: indicator values for all participants within the behavioral profile.

This means that, in theory, its mood is rarely affected by the surrounding context. Finally, participant P25, whose mood profile—3—is characterized by high average happiness, shows a lower value during days type 2, also accompanied by a considerable decrease in its average activation.

5.5 Discussion

Multilevel analysis approach for longitudinal data The vast and rich amount of information provided by longitudinal data offers many possibilities of analysis, which can be approached from multiple perspectives, depending on which aspect of the behavior we want to explore. Multiple research questions can arise from the dynamic nature of longitudinal data: *could people be classified in a finite number of affective lifestyles?*, *does the mood of a particular individual get modulated by the changes in the surrounding context?*, *could specific contextual scenarios trigger a specific mood which causes a decrease in an athlete's performance?*, etc. Each time one of these questions is addressed, the longitudinal data gathered must be pre-processed and analyzed differently, according to some parameters such as the time frame in which the major affective fluctuations occur. Therefore, the same set of longitudinal data can provide knowledge at several levels—thus progressing towards the answers to different research questions—depending on the analysis procedure. In this chapter, it has been illustrated how the proposed analysis methodology provides a unified framework to adapt the analysis of longitudinal data to different research goals based on the same data. It is essential to emphasize the exploratory nature of this methodology. Several parameters like the data aggregation time frame or the number of clusters chosen are highly subjective and could lead to different results. Performing careful data exploration is mandatory before choosing the appropriate parameters, which should be set based on expert knowledge.

The two proposed analysis pipelines provide two different perspectives of the affective dynamics, which are complementary rather than overlapping. The long-term analysis provides information on the long-term traits of each individual's mood by grouping them into mood profiles. It also helps to understand the similarities among multiple individuals. This valuable information could be used to individualize further behavioral interventions, knowing how each person's mood is expected to behave in general terms. This approach extends and improves the concept of mood profiling found in the literature. Existing research in behavioral or mood profiles commonly classifies the individuals in predefined profiles determined by one-shot questionnaires (Parsons-Smith et al., 2017). On the contrary, this methodology is intended to extract the latent profiles within the studied population—which are not known in advance—and characterize them. That way, it opens the gate to find new affective behaviors which fit the target population. Moreover, as the mood profiles are extracted from longitudinal data instead of static ones, they contain temporal aspects—trend over time, variability, etc.—of the mood, which are not subjected to retrospective mood evaluations.

The short-term analysis pipeline provides more detailed information about the short-term mood fluctuations of a single individual. This novel approach entails a new perspective on how to explore mood—and any other target behavior—over time. Although some works can be found which take into account the context and explore its effects on the behavior in the short term, the methodology presented in this work goes one step further by providing a detailed characterization of the context surrounding the individual during each time frame analyzed. Therefore, a much more rich representation of the individual's daily life can be achieved, thus helping to better understand the mood behavior during that particular time frame. Moreover, clustering the time frames into a limited number of types structures a complex concept like the context or the behavior and allows for the comparison among individuals or even among multiple moments within the same individual.

Relationship between mood and sport performance During the between-person analysis, the mood profiles were compared with the sport level of the participants—elite athletes vs. sedentary people. First of all, two profiles were found to be composed exclusively by athletes—profiles 1 and 4—, another one was formed exclusively by sedentary people—profile 2—, while the last one—profile 3—was a mixed profile among athletes and sedentary participants. In particular, a characteristic of the "sport" mood profiles is the high level of average happiness of their members. This contrast is especially evident when compared to the "sedentary" profile, which is characterized by relatively low average happiness. This finding is in line with recent reviews of sports science (Zhang and Chen, 2019), which highlight that the regular practice of physical activity shows a consistent positive relationship with happiness. The existence of two different profiles of elite athletes is also an interesting finding, primarily since both profiles are mainly differentiated in the arousal level. According to the IZOF model (Section 4.3), each athlete has an individual range of arousal level in which the best mood conditions for a good performance during sport practice are achieved. Therefore, based on these results, it could be hypothesized that the athletes belonging to each profile have a different IZOF zone. Although the studied sample is too small to draw generalizable conclusions, this finding show how the presented methodology could be applied to the data of a larger sample of athletes to unveil their IZOF zones with much more complete and objective information of their long-term variations in activation while they are immersed in their daily life and sport practice. This fact signifies a great improvement in how the optimal mood for good performance can be assessed, benefitting from the added information of longitudinal data.

In addition, the within-person analysis provides a more detailed assessment of how the mood of each athlete fluctuates and how it could affect their performance. For example, participant P06 has a mood profile type 1, which indicates that it might have an IZOF zone of high arousal—meaning that the optimal performance of this athlete should be reached when it has a high activation level. During days type 2, characterized by reduced light exposure, phone usage, and social contact through social media apps, this participant experiences a decrease in its average activation level. From the IZOF perspective, this could mean that these behaviors and contextual conditions trigger a mood change for this specific athlete, making its arousal fall out of the IZOF zone. Therefore, during those days, the athlete has not the optimal mood state for a good performance—which does not mean that the athlete could not reach such a good performance, but that the mood does not contribute towards achieving it. Another interesting example is found in participant P10. Its mood profile is number 4, which is related to a low arousal IZOF zone—optimal performance should be reached with low activation levels. From Figure 5.11 it can be seen that the variability of the arousal level during days type 1—which are characterized by short sleep duration, early wake-up time, and high amount of time spent outdoors—is substantially lower—and more constant—than the expected arousal variability of its mood profile. This could mean that, although this athlete does not fall out of its IZOF zone during days type 1, its arousal level during these days is more stable, which could entail less risk of suffering a performance decrease during the day.

Again, these examples are limited by the small sample size and should not be generalized. However, they show how this methodology could be used to perform an individualized and fine-grained exploration of how the mood of each athlete fluctuates according to the context. Moreover, it could open the door to more effective, individualized mood interventions. For example, if these results were further confirmed, a recommendation for participant P06 for ensuring a good performance during a specific competitive event could be to do morning walks under the sunlight and increase the social contact the days close to the competition. Similarly, participant P10 could be advised to wake up early and not spend much time at home the days prior to a competition to ensure a lower variability on its optimal arousal level.

Feature selection with behavioral data This methodology aims to provide a better understanding of the mood and behavior dynamics and, therefore, a smaller subset of indicators facilitates the interpretation of the results obtained. For that reason, including a feature selection step prior to applying knowledge extraction techniques is strongly recommended. However, it is necessary to be careful when using a metric to assess the relevance of each indicator within the dataset. Unsupervised feature selection techniques such as Laplacian Score are not based on any pre-existing label or outcome, which could be used to evaluate the relevance of the indicator. These techniques are usually based on the contribution of each feature to the internal structure of the dataset. For that reason, some indicators could be considered as not relevant—e.g., the valence/arousal correlation, CORR—and may be seen as noisy or redundant. In traditional applications, these indicators may be removed. In fact, if they are removed, the silhouette coefficient of the subsequent clustering is increased, insinuating that a better aggregation is obtained without those indicators.

Nonetheless, in the analysis performed in this work, the CORR indicator does not show a notable correlation with any other one, meaning that it could represent a specific aspect of the mood behavior which is not provided by any other indicator. If we do not include this information, although the clustering result will be “better"—a less complex dataset will be clustered, leading to a better cohesion—, it will not provide a complete picture of the reality which is aimed to be represented. In behavioral research, it is common to use knowledge extraction techniques that are not intended to be *black box* models but to provide understandable information. Therefore, the feature selection process should not aim to optimize the knowledge extraction algorithms. Instead, it should help increase the interpretability of the results obtained, even if it results in a lower value of the evaluation metrics—which simply means that the reality represented through the data is less structured but appropriately represented. For that reason, in this methodology, the feature selection step is based on the exploration of the data and performed individually for each data aggregation time frame.

Additional applications of the methodology This methodology has been designed based on a study designed for a specific topic. Although it serves as an example of the application of the methodology to a particular area of research, it can be applied to any behavior-related casuistry that could be influenced by context. For example, in personality research, the personality states are understood to depend on the situational context, and some studies have even found associations between them (Rüegger et al., 2020). However, there is still a lack of exploration of these associations, and researchers emphasize the need for methodologies that facilitate the exploration of the relationship between personality states and smartphone-sensed contextual data. In this case, the methodology described in this thesis can be used to analyze the fluctuations of the personality states over time depending on the context surrounding the individual and help design individualized interventions to affect personality traits. Similarly, another research topic that could benefit from this methodology is physical activity. Some studies suggest that the context may play a role in influencing the physical or sedentary behavior of individuals (Papini et al., 2020). Therefore, if the contexts in which each person is more predisposed to perform physical activity and the more appropriate activities for these contexts could be identified, more individualized interventions to promote healthy lifestyles could be designed.

Limitations The methodology designed in this chapter is not exempt from certain limitations that motivate future work in this direction. First, the dataset analyzed includes several behavioral and contextual indicators extracted from smartphone-sensed data. However, these indicators may not cover every aspect of the context surrounding the participants during their daily lives, and more smartphone sensors can be included in future works. Also, additional indicators which are not considered in this work could be extracted and included in the analysis. Secondly, the collection of data through smartphones is also subject to some trade-offs inherent to the use of mobile devices. For example, as the dataset has been collected using only Android devices, it could include a slight bias, as the subsequent results may do. For that reason, it is advised to consider the limitations of the data collection tools when interpreting the conclusions drawn from the application of this methodology.

5.6 Conclusions

The extraction of interpretable knowledge from longitudinal data is crucial to understand the dynamic nature of any human behavior—including mood. One of the biggest challenges when exploring longitudinal affective data is to adapt the analysis to its complex structure. Due to its dynamic nature, longitudinal human-related data is inherently nested and can be analyzed at several levels. Studying the differences in mood among a group of individuals—between-person level—provides knowledge on the long-term mood patterns of each individual. On a deeper level, exploring the fluctuations in the mood of each individual separately—within-person level—sheds light on how the mood of that individual changes according to its behavior and the surrounding context.

Current approaches to analyze longitudinal affective data, apart from scarce, do not provide a complete picture from both perspectives. In addition, only a few of them consider the context surrounding the individual during the analysis. For that reason, this chapter presents a novel methodological approach for the study of longitudinal data at both between- and within-person levels. The methodology is designed and described based on the affective, behavioral, and contextual data gathered from 22 individuals during two weeks, half of them elite athletes. It consists of two analysis pipelines aimed at analyzing the data at both group and individual levels, each using the data aggregated at a different time frame—two weeks and one day, respectively.

In the first analysis pipeline, the between-person differences in long-term behavior of the subjects were explored. Using a combination of data exploration, feature selection, and clustering techniques, several mood profiles were identified—each one with particular long-term mood traits—from the two-week aggregated data, and the subjects were classified among them. That way, this analysis pipeline has demonstrated its potential to uncover latent affective profiles among a specific population. In addition, from the four profiles identified, two of them belonged exclusively to athletes and one of them to sedentary participants. The difference between the two types of profiles resides in the average level of happiness, which is considerably higher for the profiles of athletes. This fact agrees with research in sports science, which has demonstrated the relationship between regular physical activity and happiness. The two athletes' profiles also differ in terms of average arousal, which could mean that these athletes have different activation zones for optimal performance—IZOF zones.

The within-person analysis focuses on exploring each individual's behavior fluctuations based on their daily behavior and context. To that end, the daily aggregated contextual data was clustered to find and characterize different types of days according to the individual's behavior and the context experienced during each of these days. Then, the short-term variations of each individual's mood on these types of days were explored. This analysis provides a much richer representation of the individuals' daily lives and facilitates the understanding of how the mood of each person changes depending on the context. Regarding the athletes, a thorough analysis of four participants revealed how this analysis could help to detect in which contexts each athlete is more prone to suffer performance decreases.

In summary, the resulting methodology provides a novel perspective on analyzing longitudinal behavioral and affective data and opens the door to more granular, detailed, and complete explorations of human-based data. In particular, its application to the field of sport performance could be promising since it could help to the early detection of performance decreases and the design of interventions to optimize the mood towards achieving the best performance not only in specific events but also during the daily practice.

6 Conclusions

6.1 Achievements

Understanding the dynamic behavior of mood is a challenging task that has been approached in several ways during the last years. Assessing how the context surrounding individuals affects their mood in the long- and short-terms presents several difficulties during the data collection and the data analysis, which have not been sufficiently addressed. Moreover, the research on the impact of these mood fluctuations on sport performance is scarce and suffers from the lack of advanced knowledge extraction methodologies. Therefore, the goal of this thesis was to investigate how mobile sensing and data science can be employed for the unobtrusive data collection in free-living environments and the extraction of interpretable knowledge from the longitudinal data gathered. In this sense, one of the most relevant global achievements of this thesis is the development of a strong multidisciplinary collaboration involving experimental psychologists, computer scientists, and athletes. We live in the era of data, and the success of behavioral interventions is highly dependent on the quality and interpretability of the information collected from the target population. For that reason, experts from several areas of knowledge must achieve a synergy that paves the way towards more individualized, advanced, and information-driven human-centered research. This work is intended to contribute to traversing that road, providing valuable insights into the development of data collection tools, the deployment of real-world longitudinal experiments, and the individualized analysis of longitudinal data. In the following, the achievement of the objectives established for this thesis is described.

Objective 1: Design and develop a smartphone-based automatic monitoring platform to simultaneously collect affective, behavioral, and contextual information in free-living environments.

Traditional data collection tools employed in sport-related studies of mood are designed to be used in controlled settings, mostly during pre- and post-exercise situations. Nowadays, the extended use of mobile sensing has improved the possibilities for continuous monitoring strategies of mood and context during the daily life of athletes. Nevertheless, existing data collection systems present certain limitations related to the simultaneous capture of contextual, behavioral and affective information, the inability to modify study parameters when the data collection is running, and the difficulty of being used by non-technical experts.

In this work, an integrated multimodal platform aimed at collecting affective and contextual data has been developed. The system employs smartphone-based technologies to gather data in free-living environments, with a combination of self-reported information and unobtrusive sensor-based data. The platform constitutes an end-to-end solution covering all the design and implementation from the data collection smartphone app to the server-backend for study management. The system design follows a modular approach, where the different constituting elements can be extended or modified to address the particular requirements of each study. Moreover, as a novel feature, the platform presents a web-based interface that allows for the real-time flexible management of ESM questionnaires. The interface provides an easy-to-use tool for designing the content and scheduling of ESM self-reports, which can be extended or modified on the fly. This feature constitutes an original approach to mitigating adherence issues during studies and capturing additional information based on the study development.

The system has been evaluated through a pilot study, whose results have been used to assess the validity and usability of the monitoring platform. The system has been proved to be capable of collecting valid data during people's daily lives, and the evaluation results demonstrate the system's feasibility for sampling self-reported mood information and contextual sensor-based data without a significant impact on the daily operation of smartphones. The usability assessment has also led to good acceptability and user-friendliness levels among researchers and end-users. However, despite the promising results of the platform, some difficulties have been faced during the platform implementation, mostly related to the impact of the smartphones' battery optimization systems on the background sensing. Based on these experiences, a set of methodological recommendations for designing data collection experiments using this platform has also been made.

Through these achievements, the first objective of this thesis has been fulfilled since a usable and effective system for affective and contextual data collection in free-living environments has been successfully designed and tested.

Objective 2: Conduct a long-term monitorization experiment within a population of elite athletes and generate a longitudinal dataset with affective, behavioral, and contextual data.

The collection of longitudinal affective, behavioral, and contextual data in free-living environments is a difficult task to accomplish since it requires the development of practical data collection tools and the design of an appropriate study protocol that fits the studied population. For that reason, the number of studies aimed at collecting affective data from athlete populations is scarce and mainly deployed in controlled settings. Moreover, currently, there is currently little to no availability of open datasets with longitudinal data of athletes during their daily lives, which complicates the development of effective analyses to extract knowledge from these data.

In this work, these issues have been addressed by conducting a longitudinal data collection experiment with a mixed population of elite athletes and sedentary people. In the experiment, affective, behavioral, and contextual data are collected during the daily life of the participants using the previously developed monitoring platform. During two weeks, participants reported their mood using ESM self-reports, and contextual data was unobtrusively collected using the smartphone sensors. The experiment resulted in one of the first openly available longitudinal datasets with affective, behavioral and contextual data of elite athletes during their out-of-sport lives.

In addition, the exceptional situation caused by the COVID-19 pandemic motivated another longitudinal experiment to collect the mood fluctuations of the Spanish population during the national lockdown. This experiment was implemented as a citizen science study, opened to all the population countrywide. Participants, who voluntarily joined the study during a maximum of two months, self-reported their mood several times per day using the monitoring platform. This experiment followed a novel strategy of data collection different from the traditional experiments with controlled samples, which entails several difficulties and challenges related to the uncontrolled nature of the sample since participants could join and withdraw from the experiment at will. This experiment also yielded an openly available dataset, one of the first and most extensive datasets released after the COVID-19 outbreak. It provides precious information for the study of the psychological effect of a prolonged lockdown on the population and is aimed at being used by governments and public health stakeholders to design information-driven measures with less psychological impact in future situations.

Prior to designing both experiments, the main methodological considerations that need to be addressed when designing longitudinal studies were investigated and described in this work. Similarly, after the development of both experiments, the lessons learned from these experiences are gathered and discussed to contribute to the safer and more accessible design of future longitudinal experiments by the scientific community. Through these achievements, the second objective of this thesis has been fulfilled since two longitudinal datasets with affective, behavioral, and contextual data have been collected –one of them based on elite athletes–, and the main methodological considerations and lessons learned from these experiments are investigated and discussed.

Objective 3: Develop an individualized analysis methodology to explore the differences in affective behavior among athletes, the contextual factors which modulate their mood, and the effect of their mood on their performance.

The dynamic nature and complex structure of longitudinal data hinder the extraction of interpretable knowledge of the long- and short-term behavioral patterns and their fluctuations. That is a necessary step to understand the contextual elements that modulate the mood of athletes and how these fluctuations affect their performance. However, current approaches to analyze longitudinal behavioral data are scarce and do not provide a complete picture of the characteristics of affective behavior at both between- and within-person levels.

In this work, a multilevel methodology for the analysis of longitudinal behavioral data is designed and presented. The methodology consists of two pipelines to analyze the data at group and individual levels. One of the main contributions of this methodology is that it achieves the analysis at both levels using the same starting data but aggregating them at different time frames. The methodology is tested on the dataset obtained from the first data collection experiment of this work, which leads to a series of insights about the relationship between mood and sport performance. Although not sufficiently generalizable due to the small sample size, these conclusions show the potential of this methodology to the early detection of performance decreases and the design of interventions to optimize the mood towards achieving the best performance of each athlete.

The between-person analysis explores the differences in the long-term affective behavior of the subjects. As a result, they were clustered in a finite number of mood profiles, each one with particular long-term traits. The profiles obtained showed significant differences among the affective behavior of athletes and sedentary subjects in terms of happiness. The two mood profiles of athletes also differ in terms of average arousal, which could be related to a difference in activation zones for optimal performance—IZOF zones. The within-person analysis focuses on the mood fluctuations of each individual based on the context surrounding them. A finite number of day types were characterized based on their contextual characteristics, and the short-term mood behavior of each subject during each of these days was identified. This analysis facilitates the understanding of how each person's mood changes depending on the context. Additionally, a thorough analysis of various athlete subjects revealed the potential of this methodology to detect which contexts trigger moods that could lead to performance decreases or increases.

Through these achievements, the third objective of this thesis has been fulfilled since an analysis methodology to uncover the mood-context-performance relationship has been developed, and it preserves the potential to be generalized and extract interpretable information from other target populations.

6.2 Contributions

In the previous section, it has been proved that the objectives of this thesis have been fulfilled. In the following, the main contributions of this thesis are listed:

- Design, implementation, and validation of a multimodal monitoring platform for the continuous collection of affective, behavioral, and contextual data in free-living environments, using both self-reports and unobtrusive sensor-based data.
- Identification and discussion of the main methodological considerations previous to the design of longitudinal data collection studies.
- Design and implementation of two longitudinal data collection experiments in free-living environments with two different participation strategies: controlled sample and open participation.
- Discussion of the lessons learned from the experience of real-world data collection experiments in free-living environments.
- Design, implementation, and evaluation of a multilevel methodological approach for analyzing the between- and within-person behavioral differences and fluctuations of longitudinal data.
- Generation of preliminary conclusions about the relationship between mood and sport performance using the developed methodology.
- Collection and curation of a longitudinal dataset of mood variations of elite athletes. This dataset includes affective, behavioral, and contextual data of a mixed sample of elite athletes and sedentary people gathered during their daily lives. The dataset is openly available to the research community at https://doi.org/10.17605/OSF.IO/2Z5B6.
- Collection and curation of a longitudinal dataset of the mood fluctuations of the Spanish population during the COVID-19 lockdown. This dataset includes longitudinal affective data of participants country-wide during two months of lockdown in Spain. The dataset, which is aimed at being used to design more effective and psychologically safe healthcare measures in future situations, is openly available to the

research community at https://doi.org/10.17605/osf.io/5CQZK and https://doi.org/10.5281/zenodo.3774526.

6.3 Future directions

Given the novelty and potential of this work, there are several possibilities to continue this line of research. In this section, the most relevant future directions to extend the work presented in this thesis are described.

Integration of new personal sensing devices for the unobtrusive collection of behavioral, affective and contextual data

In order to appropriately represent people's daily lives, the monitored individuals must follow their daily routines as naturally as possible, even when participating in a research study. If they are given any external sensing device such as a wristband, chest strap, or even an additional smartphone for the study, they may be aware that they are being monitored and unconsciously modify their behavior. For that reason, the monitoring platform developed in this work only uses the participants' own smartphones as data collection tools. However, nowadays, the use of new personal devices is continuously increasing. Smartwatches, smart earphones, and intelligent clothes are beginning to become part of our daily equipment. In the next years, these devices are going to change from being rare devices of occasional use to be as widely used as smartphones or regular earphones.

Since many of these devices are directly attached to the user's body, they provide sensing capabilities that surpass the possibilities offered by smartphones. Many physiological and contextual elements such as heart rate, skin conductance, or speech can be monitored using these wearable devices. Therefore, as their use becomes more widespread, there is an excellent opportunity to integrate them into existing platforms and systems of personal sensing. Moreover, they could not only provide additional and more rich sensing information but also help to overcome some limitations related to the use of smartphones which have been discussed in this thesis—for example, the impact of battery optimization systems on the background sensing.

Collection of more extense datasets of athletes

This work is primarily motivated by the fact that mood has a significant impact on the performance of elite athletes. The preliminary results obtained by applying the developed analysis methodology show that the context surrounding the athletes and their behavior during their daily life are highly relevant contributing factors to the modulation of their mood. For that reason, identifying which contexts trigger the optimum mood for a good performance is a crucial task in sports research. However, it could be fascinating to explore the affective behavior not only to optimize the performance towards competitions but also during daily practices. To that end, it is necessary to observe intra-individual variations of mood during much more extended periods than the two weeks covered by the dataset generated in this work. Gathering data of athletes during months or even a whole season allows for capturing a wider variety of context scenarios during different season periods—pre-season, competitive, off-season.

Of course, data collection studies must be carefully designed, based on the methodological considerations presented in this work, to ensure the reliability of the data obtained and keep a reasonable participant burden. However, although technically and practically challenging, the periodic conduction of two- to four-week-long experiments during a whole season can provide invaluable information that could open the gate to a new generation of sports training.

Investigation of the temporal dimension of behavior fluctuations

The developed analysis methodology provides a novel approach for exploring the within-person fluctuations of mood—or any other target behavior—based on the different contexts experienced by the subject. However, the analysis of the dynamic nature of behavior can go a step further and examine the sequencing of these short-time behaviors. It could be helpful to examine not only how each subject behaves under each context but also how these fluctuations are distributed over time. This type of analysis can extend the results yielded by the methodology and trigger new research questions such as whether the effect of specific contexts on the mood could be experienced with a delay of days or weeks, depending on the studied individual. For this kind of analyses, data science and machine learning techniques that account for the causality of events should be applied —for example, Structural Equation Modeling (SEM) based longitudinal models (such as latent growth models, LGM).

If the ultimate goal is the design of behavioral interventions, the study of each subject's behavior must be completely individualized. For that reason, the extensions of the methodology described could be highly valuable for applying the analysis approach not only to behaviors related to well-being but also clinical conditions. With deeper insights about the causality of the events and contexts which trigger significant behavior changes and the direction of those changes, much more effective interventions could be designed to improve the treatment of psychological conditions such as depression or borderline personality disorder (BPD).

7 Conclusiones

7.1 Logros

Durante los últimos años, el complejo reto de entender el comportamiento dinámico del estado de ánimo ha sido abordado de diferentes formas. La tarea de medir el efecto del contexto que rodea a las personas sobre su estado de ánimo a largo y corto plazo, presenta dificultades tanto en la obtención de los datos como en su posterior análisis, las cuales no han sido suficientemente estudiadas. Además, las investigaciones sobre cómo afectan dichas variaciones del estado de ánimo en el rendimiento deportivo son escasas, y sufren de la falta de metodologías avanzadas para extraer conocimiento. Es por esto que el objetivo de esta tesis ha consistido en investigar cómo la monitorización móvil y la ciencia de datos pueden ser utilizadas para obtener datos de forma discreta en entornos no controlados y para extraer conocimiento interpretable de dichos datos. En este sentido, uno de los logros globales más importantes de esta tesis ha sido la formación de una potente colaboración multidisciplinar, implicando a psicólogos experimentales, expertos en ciencia de computadores y atletas. Vivimos en la era de los datos, y el éxito de las intervenciones conductuales depende en gran medida de la calidad e interpretabilidad de la información recogida de la población objetivo. Por este motivo, los expertos de diferentes áreas de conocimiento deben alcanzar una sinergia que allane el camino hacia investigaciones más individualizadas, avanzadas y basadas en la información. Este trabajo pretende contribuir a recorrer ese camino, proporcionando conocimiento valioso en cuanto al desarrollo de herramientas de recogida de datos, la implementación de experimentos longitudinales en el mundo real, y el análisis individualizado de datos longitudinales. A continuación se describen los logros obtenidos para cada uno de los objetivos fijados para esta tesis.

Objetivo 1: Diseñar y desarrollar una plataforma automática de monitorización basada en teléfonos móviles, para recoger simultáneamente información de estado de ánimo, comportamiento y contexto en entornos no controlados.

Las herramientas de captura de datos que han sido tradicionalmente empleadas en estudios relacionados con el deporte y el estado de ánimo, han sido diseñadas para usarse en entornos controlados, mayoritariamente en situaciones pre y post ejercicio. Hoy en día, el uso extendido de monitorización móvil plantea grandes posibilidades a la hora de monitorizar de forma continua el estado de ánimo y el contexto durante la vida diaria de los deportistas. Sin embargo, los sistemas de captura de datos existentes presentan ciertas limitaciones en relación con la captura simultánea de datos de estado de ánimo, comportamiento y contexto, la imposibilidad de modificar parámetros del estudio cuando está en marcha, y la dificultad de uso por parte de expertos no técnicos.

En esta tesis se ha desarrollado una plataforma multimodal integrada, orientada a la captura de datos afectivos y contextuales. El sistema utiliza tecnologías basadas en teléfonos móviles para obtener datos en entornos no controlados, con una combinación de información subjetiva y datos de sensores. La plataforma constituye una solución integrada que cubre todo el diseño e implementación desde la aplicación de recogida de datos para móvil hasta el backend para la gestión de los estudios. El sistema está diseñado siguiendo una arquitectura modular, en la que los diferentes elementos que lo constituyen pueden ser extendidos y modificados para adaptarse a los requisitos específicos de cada estudio. Además, como elemento novedoso, la plataforma contiene una interfaz web que permite la gestión flexible de los cuestionarios lanzados en el teléfono móvil. Esta interfaz proporciona una herramienta para el diseño del contenido y la planificación de cuestionarios, que pueden ser extendidos y modificados en tiempo real. Este elemento supone un novedoso método para mitigar problemas de adherencia durante los estudios, y para capturar información adicional en función de la evolución del estudio.

El sistema ha sido evaluado por medio de un estudio piloto, cuyos resultados han sido utilizados para evaluar la validez y usabilidad de la plataforma de monitorización. En términos de funcionamiento, el sistema ha demostrado ser capaz de recoger datos válidos durante la vida diaria de las personas. Los resultados de la evaluación demuestran la viabilidad del sistema para capturar información subjetiva sobre el estado de ánimo e información objetiva sobre el contexto, sin un impacto significativo en el funcionamiento diario de los teléfonos móviles. La evaluación de usabilidad muestran también una buena aceptación y facilidad de uso por parte de los expertos y los usuarios finales. Sin embargo, a pesar de los prometedores resultados de la plataforma, durante su implementación se han encontrado ciertas dificultades, especialmente relacionadas con el impacto de los sistemas de optimización de batería de los teléfonos móviles en la captura de datos en segundo plano. A partir de estas experiencias, se han realizado una serie de recomendaciones metodológicas para el diseño de futuros experimentos de captura de datos usando esta plataforma.

Mediante estos logros, el primer objetivo de esta tesis se ha completado exitosamente, ya que se ha diseñado y probado un sistema utilizable y efectivo para la captura de datos emocionales y contextuales en entornos no controlados.

Objetivo 2: Llevar a cabo un experimento de monitorización a largo plazo con una muestra de deportistas de élite y generar un conjunto de datos longitudinal con datos de estado de ánimo, comportamiento y contexto.

La recogida de datos longitudinales de estado de ánimo, comportamiento y contexto en entornos no controlados es una tarea difícil de lograr, ya que requiere el desarrollo de herramientas efectivas para captura de datos, y el diseño de un protocolo de investigación que se adapte a la población estudiada. Por ese motivo solo existe un limitado número estudios orientados a la recogida de datos emocionales de deportistas, y la mayoría se desarrollan en entornos controlados. Además, actualmente hay poca o ninguna disponibilidad de conjuntos de datos longitudinales de atletas durante su vida diaria, lo que dificulta el desarrollo de análisis efectivos para extraer conocimiento de dichos datos.

En esta tesis se han abordado estas cuestiones mediante el desarrollo de un experimento longitudinal de recogida de datos con una población mixta de deportistas de élite y personas sedentarias. En el experimento se recogen datos sobre el estado de ánimo, el comportamiento y el contexto que rodea a los participantes durante su vida diaria utilizando la plataforma de monitorización desarrollada anteriormente. Durante dos semanas, los participantes evaluaron su estado de ánimo de forma subjetiva mediante cuestionarios, y se recogieron datos de contexto de forma discreta usando los sensores de sus teléfonos móviles. El experimento dio como resultado uno de los primeros conjuntos de datos longitudinales abiertos con información emocional, comportamental y contextual de deportistas de élite durante su vida extradeportiva.

La situación excepcional provocada por la pandemia de COVID-19 motivó otro experimento longitudinal para recoger datos acerca de las variaciones del estado de ánimo de la población española durante el confinamiento a nivel nacional. Este experimento fue llevado a cabo como un estudio de ciencia ciudadana, abierto a toda la población a lo largo del país. Los participantes, que se unieron de forma voluntaria durante un máximo de dos meses, evaluaron subjetivamente su estado de ánimo varias veces a lo largo del día usando la plataforma de monitorización. Este experimento ha seguido una estrategia de recogida de datos distinta a las utilizadas en estudios tradicionales con muestras controladas, la cual conlleva dificultades y retos inherentes a la naturaleza no controlada de la muestra, ya que los participantes podían unirse y retirarse del estudio a voluntad. El experimento también resultó en un conjunto de datos longitudinal abierto, uno de los primeros y más extensos liberados tras el brote de COVID-19. El conjunto de datos proporciona información valiosa para el estudio del impacto psicológico de una cuarentena prolongada en la población, y está concebido para ser utilizado por los gobiernos y agentes de salud pública para diseñar medidas sanitarias basadas en la información, con menor impacto psicológico en situaciones futuras.

Antes de diseñar los experimentos se han investigado y descrito las principales consideraciones metodológicas que es necesario estudiar cuando se diseñan estudios longitudinales. De forma similar, tras el desarrollo de los experimentos se han recopilado y discutido las lecciones aprendidas en estas experiencias, lo que pretende contribuir al diseño de futuros experimentos longitudinales de forma más segura y accesible por parte de la comunidad científica. Mediante estos logros, el segundo objetivo de esta tesis se ha completado exitosamente, ya que se han recogido dos conjuntos de datos emocionales, comportamentales y contextuales –uno de ellos con deportistas de élite–, y se han investigado y discutido las principales consideraciones metodológicas y lecciones aprendidas de estos experimentos.

Objetivo 3: Desarrollar una metodología de análisis individualizado para explorar las diferencias en el comportamiento emocional de los deportistas, los elementos de contexto que modulan su estado de ánimo, y el efecto del estado de ánimo sobre su rendimiento.

La naturaleza dinámica de los datos longitudinales, así como su compleja estructura, dificultan la obtención de conocimiento interpretable de los patrones de comportamiento y sus variaciones a corto y largo plazo. Este paso es necesario a la hora de entender qué elementos del contexto modulan el estado de ánimo de los deportistas, y cómo estas fluctuaciones afectan a su rendimiento. Sin embargo, los enfoques actuales de análisis de datos longitudinales de comportamiento son limitados y no proporcionan una visión completa de las características del comportamiento emocional a nivel grupal e individual.

En esta tesis se ha diseñado y presentado una metodología multinivel para el análisis de datos longitudinales de comportamiento. La metodología consiste en dos estrategias para analizar los datos en ambos niveles. Una de las mayores contribuciones de esta metodología es que logra ambos tipos de análisis partiendo de los mismos datos, pero agregándolos en una ventana temporal distinta. La metodología ha sido probada en el conjunto de datos obtenido del primer experimento llevado a cabo en este trabajo, lo que ha permitido obtener información acerca de la relación entre el estado de ánimo y el rendimiento deportivo. Aunque no son suficientemente generalizables dada la pequeña muestra del experimento, dichas conclusiones muestran el potencial de la metodología para la detección temprana de disminuciones en el rendimiento y el diseño de intervenciones para optimizar el estado de ánimo de cara a obtener el mejor rendimiento de cada deportista.

El análisis grupal explora las diferencias en el comportamiento afectivo a largo plazo de los sujetos. Como resultado, son agrupados en un número finito de perfiles emocionales, cada uno con características específicas a largo plazo. Los perfiles obtenidos muestran diferencias significativas en comportamiento afectivo entre los deportistas y los paticipantes sedentarios en términos de felicidad. Los dos perfiles emocionales de atletas también difieren en términos de activación promedio, lo que podría estar relacionado con diferentes zonas de activación para el rendimiento óptimo—zonas IZOF. El análisis individual, por su parte, se centra en las fluctuaciones del estado de ánimo de cada individuo en función del contexto que le rodea. Se caracterizó un número finito de tipos de días de acuerdo a sus características de contexto, y se identificó el comportamiento emocional a corto plazo de cada sujeto en cada uno de esos días. Este análisis facilita la comprensión de cómo el estado de ánimo de cada persona varía en función del contexto. Además, un análisis en profundidad de varios sujetos atletas reveló el potencial de la metodología para detectar qué contextos desencadenan estados de ánimo que podrían llevar a disminuciones o mejoras del rendimiento.

Mediante estos logros, el tercer objetivo de esta tesis se ha completado exitosamente, ya que se ha desarrollado una metodología para estudiar la relación estado de ánimo-contexto-rendimiento, la cual conserva el potencial de ser generalizada y extraer información interpretable de otros objetivos de investigación.

7.2 Contribuciones

En la sección anterior se ha descrito cómo los objetivos de esta tesis se han cumplido satisfactoriamente. A continuación, se listan las contribuciones principales de esta tesis:

- Diseño, implementación y validación de una plataforma multimodal de monitorización para la recogida continua de datos de estado de ánimo, comportamiento y contexto en entornos no controlados, utilizando cuestionarios y datos de sensores.
- Identificación y discusión de las principales consideraciones metodológicas previas al diseño de estudios longitudinales de captura de datos.

- Discusión de las lecciones aprendidas de la realización de experimentos reales de recogida de datos en entornos no controlados.
- Diseño, implementación y evaluación de una metodología multinivel para el análisis grupal e individual de las diferencias y fluctuaciones en el comportamiento con datos longitudinales.
- Obtención de conclusiones preliminares sobre la relación entre el estado de ánimo y el rendimiento deportivo utilizando la metodología desarrollada.
- Registro y adecuación de un conjunto de datos longitudinal de variaciones emocionales en deportistas de élite. El conjunto de datos contiene información de estado de ánimo, comportamiento y contexto de deportistas de élite y personas sedentarias, recogida durante su vida diaria. El conjunto de datos está disponible públicamente en https://doi.org/10.17605/OSF.IO/2Z5B6.
- Registro y adecuación de un conjunto de datos longitudinal de las fluctuaciones emocionales de la población española durante el confinamiento por COVID-19. Este conjunto de datos contiene información longitudinal del estado de ánimo de participantes a lo largo de todo el país durante dos meses de confinamiento en España. El conjunto de datos, cuyo objetivo es ser utilizado para diseñar medidas sanitarias más efectivas y psicológicamente seguras en futuras situaciones, está disponible públicamente en https://doi.org/10.17605/osf.io/5CQZK y https://doi.org/10.5281/zenodo.3774526.

7.3 Trabajo futuro

Dada la novedad y el potencial de este trabajo, existe todavía una amplia gama de posibilidades para continuar esta línea de investigación. En esta sección se presentan las principales líneas de trabajo futuro que permiten continuar la investigación desarrollada en esta tesis.

Integración de nuevos dispositivos de monitorización para la captura de datos de comportamiento, estado de ánimo y contexto

Para representar de forma adecuada la vida diaria de las personas, los sujetos monitorizados deben continuar con su rutina de la forma más natural posible, incluso cuando estén participando en un estudio de investigación. Si se les proporciona algún dispositivo de monitorización externo, como una pulsera, una banda para el pecho, o incluso un teléfono móvil adicional, los participantes pueden ser conscientes de que están siendo monitorizados, y modificar su comportamiento de forma inconsciente. Es por ello que la plataforma de monitorización desarrollada en este trabajo utiliza únicamente los teléfonos móviles de los participantes como dispositivos de recogida de datos. Sin embargo, hoy en día el uso de nuevos dispositivos personales está siendo cada vez más común. Dispositivos como relojes inteligentes, auriculares inteligentes o incluso ropa inteligente están comenzando a formar parte de nuestro equipamiento diario. En los próximos años, van a pasar de ser dispositivos de uso ocasional a ser tan utilizados como los teléfonos móviles o los auriculares normales.

Como muchos de estos dispositivos están directamente acoplados al cuerpo del usuario, proporcionan muchas más posibilidades de medida que las que ofrecen los teléfonos móviles. Por ejemplo, pueden medir muchas variables fisiológicas y contextuales tales como el ritmo cardiaco, la conductancia de la piel o las conversaciones del usuario. Por este motivo, conforme su uso vaya siendo más extendido, se abre la posibilidad de integrar estos dispositivos en las plataformas y sistemas de fenotipado digital existentes. Además, no solo podrían proporcionar mayor y más valiosa información, sino también ayudar a contrarrestar algunas de las limitaciones relacionadas con el uso de teléfonos móviles que han sido discutidas en esta tesis—por ejemplo, el impacto de los sistemas de optimización de batería en la monitorización en segundo plano.

Recogida de conjuntos de datos más extensos de atletas

Una de las motivaciones principales de esta tesis es el hecho de que el estado de ánimo tiene un fuerte impacto en el rendimiento de los deportistas de élite. Los resultados preliminares obtenidos aplicando la metodología de análisis desarrollada muestran que el contexto que rodea a los deportistas y su comportamiento durante su vida diaria son factores muy importantes que contribuyen a la modulación de su estado de ánimo. Por este motivo, una tarea fundamental en la investigación deportiva es identificar qué contextos desencadenan el estado de ánimo óptimo para alcanzar un buen rendimiento. Sin embargo, podría ser de enorme interés explorar el comportamiento emocional no solo para optimizar el rendimiento de cara a las competiciones, sino también a los entrenamientos diarios. Para ello es necesario observar las variaciones intra-individuo del estado de ánimo durante periodos de tiempo mucho más extensos que las dos semanas que cubre el dataset generado en esta tesis. Recoger datos de atletas durante meses o incluso una temporada completa permite capturar una mayor variedad de escenarios contextuales durante diferentes periodos de la temporada—pre-temporada, periodo competitivo, post-temporada.

Por supuesto, los estudios de recogida de datos deben ser cuidadosamente diseñados, basándose en las consideraciones metodológicas presentadas en esta tesis, de forma que se asegure la fiabilidad de los datos obtenidos, y el estudio suponga una carga de trabajo razonable para los participantes. Sin embargo, aunque conlleve cierta complejidad técnica y práctica, la realización de forma periódica de experimentos con duraciones de entre dos y cuatro semanas a lo largo de la temporada puede proporcionar información de valor incalculable, que puede abrir la puerta a una nueva generación de entrenamientos deportivos.

Investigación de la dimensión temporal de las fluctuaciones del comportamiento

La metodología de análisis desarrollada proporciona un enfoque novedoso para explorar las fluctuaciones intra-personales del estado de ánimo—o cualquier otro comportamiento—en base a los diferentes contextos experimentados por el sujeto. Sin embargo, el análisis de la naturaleza dinámica del comportamiento puede ir un paso más allá y examinar la secuenciación de esos comportamientos a corto plazo. Podría ser de mucha utilidad examinar no solo cómo cada sujeto se comporta ante cada contexto, sino cómo esas fluctuaciones se distribuyen a lo largo del tiempo. Este tipo de análsis puede extender los resultados obtenidos con la metodología y generar nuevas preguntas de investigación tales como si el efecto de ciertos contextos en el estado de ánimo puede ser experimentado con un retardo de días o semanas, en función del individuo estudiado. Para dichos análisis se deben emplear técnicas de ciencia de datos y *machine learning* que tengan en cuenta la causalidad de los eventos —por ejemplo, modelos longitudinales basados en Modelos de Ecuaciones Estructurales (MES), tales como modelos de crecimiento latente (MCL).

Si el objetivo principal es el diseño de intervenciones de comportamiento, el estudio del comportamiento de cada sujeto debe ser completamente individualizado. Por este motivo, las extensiones de la metodología descritas tienen un alto valor para el uso de la metodología de análisis no solo a comportamientos relacionados con el bienestar, sino también a trastornos clínicos. Con un conocimiento más profundo de la causalidad de los eventos y contextos que provocan cambios en el comportamiento y de la dirección de dichos cambios, se pueden diseñar intervenciones mucho más efectivas para mejorar el tratamiento de trastornos psicológicos tales como la depresión o el trastorno límite de personalidad (TLP).

Bibliography

Abdullah, S., Matthews, M., Murnane, E. L., Gay, G., & Choudhury, T. (2014). Towards circadian computing: "early to bed and early to rise" makes some of us unhealthy and sleep deprived. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 673–684 (cit. on p. 22).

Ahmad, A. R., & Murad, H. R. (2020). The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: Online questionnaire study. *Journal of Medical Internet Research*, *22*(5), e19556 (cit. on p. 77).

Asselbergs, J., Ruwaard, J., Ejdys, M., Schrader, N., Sijbrandij, M., & Riper, H. (2016). Mobile Phone-Based Unobtrusive Ecological Momentary Assessment of Day-to-Day Mood: An Explorative Study. *Journal of Medical Internet Research*, *18*(3), e72 (cit. on p. 19).

Attfield, S., Kazai, G., Lalmas, M., & Piwowarski, B. (2011). Towards a science of user engagement. *Workshop on User Modelling for Web Applications, 09-12 Feb 2011, Hong Kong, China*, 1–8 (cit. on p. 59).

Bados, A., Solanas, A., & Andrés, R. (2005). Psychometric properties of the spanish version of depression, anxiety and stress scales (dass). *Psicothema*, *17*(Número 4), 679–683 (cit. on p. 71).

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Human Computer Interaction*, *24*, 574–594 (cit. on p. 57).

Bardram, J. E. (2020). The CARP Mobile Sensing Framework – A Cross-platform, Reactive, Programming Framework and Runtime Environment for Digital Phenotyping [arXiv: 2006.11904]. *arXiv:2006.11904 [cs]* (cit. on pp. 25–28).

Barrett, L. F., & Barrett, D. J. (2001). An introduction to computerized experience sampling in psychology. *Social Science Computer Review*, *19*(2), 175–185 (cit. on p. 68).

Bauer, D. J., McNeish, D. M., Baldwin, S. A., & Curran, P. J. (2020). Analyzing nested data: Multilevel modeling and alternative approaches. In *The cambridge handbook of research methods in clinical psychology* (pp. 426–443). Cambridge University Press. (Cit. on pp. 8, 95).

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828 (cit. on p. 21).

Blume, C., Garbazza, C., & Spitschan, M. (2019). Effects of light on human circadian rhythms, sleep and mood. *Somnologie*, *23*(3), 147–156 (cit. on p. 102).

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology*, *54*(1), 579–616 (cit. on pp. 5, 17, 50).

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59 (cit. on p. 16).

Brooke, J. (1996). Sus-a quick and dirty usability scale. In *Usability evaluation in industry*. Taylor&Francis. (Cit. on pp. 51, 56).

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. *Lancet*, *395* (cit. on p. 77).

Brunette, W., Sodt, R., Chaudhri, R., Goel, M., Falcone, M., van Orden, J., & Borriello, G. (2012). Open data kit sensors: A sensor integration framework for android at the application-level. *Proceedings of the 10th international conference on Mobile systems, applications, and services*, 351–364 (cit. on p. 27).

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), e55 (cit. on p. 19).

Caine, K. (2016). Local standards for sample size at chi. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 981–992 (cit. on p. 67).

Ciman, M., & Wac, K. (2019). Smartphones as Sleep Duration Sensors: Validation of the iSenseSleep Algorithm. *JMIR mHealth and uHealth*, *7*(5), e11930 (cit. on pp. 73, 102, 103).

Courvoisier, D. S., Eid, M., Lischetzke, T., & Schreiber, W. H. (2010). Psychometric properties of a computerized mobile phone method for assessing mood in daily life. *Emotion*, 115–124 (cit. on p. 19).

Csikszentmihalyi, M., & Larson, R. (2014). Validity and Reliability of the Experience-Sampling Method. In *Flow and the Foundations of Positive Psychology* (pp. 35–54). Springer Netherlands. (Cit. on p. 68).

Davidson, R. (1994). On emotion, mood, and related affective construct. In P. Ekman & R. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 51–55). NY: Oxford University Press. (Cit. on p. 1).

Davis, P. A., & Stenling, A. (2020). Temporal aspects of affective states, physiological responses, and perceived exertion in competitive cycling time trials. *Scandinavian Journal of Medicine & Science in Sports*, *30*(10), 1859–1868 (cit. on pp. 3, 6).

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour*, *3*(5), 478–491 (cit. on p. 100).

Dey, A. K. (2001). Understanding and Using Context. *Personal and Ubiquitous Computing*, *5*, 4–7 (cit. on p. 2).

Dingler, T., & Pielot, M. (2015). I'll be there for you: Quantifying attentiveness towards mobile messaging. *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–5 (cit. on p. 68).

Dogan, E., Sander, C., Wagner, X., Hegerl, U., & Kohls, E. (2017). Smartphone-Based Monitoring of Objective and Subjective Data in Affective Disorders: Where Are We and Where Are We Going? Systematic Review. *Journal of Medical Internet Research*, *19*(7), e262 (cit. on p. 5).

Doron, J., & Gaudreau, P. (2014). A point-by-point analysis of performance in a fencing match: Psychological processes associates with winning and losing streaks. *Journal of Sport and Exercise Psychology*, *36*(1), 3–13 (cit. on p. 2).

Eagle, N., & Pentland, A. (2006). Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing*, *10*(4), 255–268 (cit. on pp. 20, 97).

Ekkekakis, P. (2012). Affect, Mood, and Emotion. *Measurement in sport and exercise psychology*, 321–332 (cit. on pp. 2, 13, 14).

Ferreira, D., Kostakos, V., & Dey, A. K. (2015). AWARE: Mobile Context Instrumentation Framework. *Frontiers in ICT*, *2*, 6 (cit. on pp. 24–27, 32, 45, 63).

Gehlbach, H., & Barge, S. (2012). Anchoring and Adjusting in Questionnaire Responses. *Basic and Applied Social Psychology*, *34*(5), 417–433 (cit. on p. 59).

Gravenhorst, F., Muaremi, A., Bardram, J., Grunerbl, A., Mayora, O., Wurzer, G., Frost, M., Osmani, V., Arnrich, B., Lukowicz, P., & Troster, G. (2015). Mobile phones as medical devices in mental disorder treatment: An overview. *Personal and Ubiquitous Computing*, *19*(2), 335–353 (cit. on pp. 4, 5, 23).

Gu, W., Shangguan, L., Yang, Z., & Liu, Y. (2016). Sleep Hunter: Towards Fine Grained Sleep Stage Tracking with Smartphones. *IEEE Transactions on Mobile Computing*, *15*(6), 1514–1527 (cit. on p. 22).

Guyon, I., Elisseeff, A., & Kaelbling, L. P. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(7-8), 1157–1182 (cit. on p. 108).

Hanin, Y. L. (2000). Individual zones of optimal functioning (izof) model: Emotion-performance relationship in sport. In *Emotions in sport*. Human Kinetics. (Cit. on p. 69).

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. *Perspectives on Psychological Science*, *11*(6), 838–854 (cit. on pp. 4, 8, 9, 20, 24, 67, 73, 95, 96).

Hashemian, M., Knowles, D., Calver, J., Qian, W., Bullock, M. C., Bell, S., Mandryk, R. L., Osgood, N., & Stanley, K. G. (2012). iEpi: An end to end solution for collecting, conditioning and utilizing epidemiologically relevant data. *Proceedings of the 2nd ACM international workshop on Pervasive Wireless Healthcare*, 3–8 (cit. on pp. 24, 27).

He, X., Cai, D., & Niyogi, P. (2005). Laplacian score for feature selection. *Proceedings of the 18th International Conference on Neural Information Processing Systems*, 507–514 (cit. on p. 112).

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Sage Publications, Inc. (Cit. on p. 66).

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *44*(2), 227–239 (cit. on p. 71).

Heron, K. E., & Smyth, J. M. (2010). Ecological Momentary Interventions: Incorporating Mobile Technology Into Psychosocial and Health Behavior Treatments. *British journal of health psychology*, *15*(Pt 1), 1–39 (cit. on p. 59).

Hormuth, S. E. (1986). The sampling of experiences in situ. *Journal of Personality*, *54*(1), 262–293 (cit. on p. 18).

Hossain, S. M., Hnat, T., Saleheen, N., Nasrin, N. J., Noor, J., Ho, B.-J., Condie, T., Srivastava, M., & Kumar, S. (2017). mCerebrum: A Mobile Sensing Software Platform for Development and Validation of Digital Biomarkers and Interventions. *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, 1–14 (cit. on p. 27).

Houben, M., van den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*, 901–930 (cit. on p. 77).

Hsieh, C.-K., Tangmunarunkit, H., Alquaddoomi, F., Jenkins, J., Kang, J., Ketcham, C., Longstaff, B., Selsky, J., Dawson, B., Swendeman, D., Estrin, D., & Ramanathan, N. (2013). Lifestreams: A modular sense-making toolset for identifying important patterns from everyday life. *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, 1–13 (cit. on p. 27).

Huang, M.-H. (2003). Designing website attributes to induce experiential encounters. *Computers in Human Behavior*, *19*(4), 425–442 (cit. on p. 59).

Huckvale, K., Venkatesh, S., & Christensen, H. (2019). Toward clinical digital phenotyping: A timely opportunity to consider purpose, quality, and safety. *npj Digital Medicine*, *2*(1), 1–11 (cit. on p. 65).

Ilda, M., Shrout, P. E., Laurenceau, J.-P., & Bolger, N. (2012). Using diary methods in psychological research. In *Apa handbook of research methods in psychology, vol. 1. foundations, planning, measures, and psychometrics* (pp. 277–305). American Psychological Association. (Cit. on p. 17).

Jain, S. H., Powers, B. W., Hawkins, J. B., & Brownstein, J. S. (2015). The digital phenotype. *Nature biotechnology*, *33*(5), 462–463 (cit. on p. 20).

Kanning, M., Ebner-Priemer, U., & Schlicht, W. (2013). How to investigate within-subject associations between physical activity and momentary affective states in everyday life: A position statement based on a literature overview. *Frontiers in Psychology*, *4*, 187 (cit. on p. 31).

Kanning, M., & Schlicht, W. (2010). Be Active and Become Happy: An Ecological Momentary Assessment of Physical Activity and Mood. *Journal of Sport and Exercise Psychology*, *32*(2), 253–261 (cit. on p. 3).

Katevas, K., Haddadi, H., & Tokarchuk, L. (2014). Poster: SensingKit: A multi-platform mobile sensing framework for large-scale experiments. *Proceedings of the 20th annual international conference on Mobile computing and networking*, 375–378 (cit. on p. 27).

Kim, H., Lee, S., Lee, S., Hong, S., Kang, H., & Kim, N. (2019). Depression Prediction by Using Ecological Momentary Assessment, Actiwatch Data, and Machine Learning: Observational Study on Older Adults Living Alone. *JMIR mHealth and uHealth*, *7*(10), e14149 (cit. on p. 73).

Klasnja, P., Consolvo, S., McDonald, D. W., Landay, J. A., & Pratt, W. (2009). Using mobile & personal sensing technologies to support health behavior change in everyday life: Lessons learned. *AMIA Annual Symposium proceedings*, 338–342 (cit. on p. 20).

Klasnja, P., Harrison, B. L., LeGrand, L., LaMarca, A., Froehlich, J., & Hudson, S. E. (2008). Using wearable sensors and real time inference to understand human recall of routine activities. *Proceedings of the 10th International Conference on Ubiquitous Computing*, 154–163 (cit. on p. 68).

Kumar, D., Jeuris, S., Bardram, J. E., & Dragoni, N. (2021). Mobile and Wearable Sensing Frameworks for mHealth Studies and Applications: A Systematic Review. *2*(1), 28 (cit. on pp. 23–26, 60).

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, *99*(6), 1042–1060 (cit. on p. 2).

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. In *Flow and the foundations of positive psychology*. Wiley Jossey-Bass. (Cit. on p. 17).

Lazarus, R. S. (2000). How emotions influence performance in competitive sports. *The Sport Psychologist*, *14*(3), 229–252 (cit. on p. 2).

Lee, Y., Iyengar, S. S., Min, C., Ju, Y., Kang, S., Park, T., Lee, J., Rhee, Y., & Song, J. (2012). MobiCon: A mobile context-monitoring platform. *Communications of the ACM*, *55*(3), 54–65 (cit. on p. 27).

Li, I., Dey, A., & Forlizzi, J. (2010). A stage-based model of personal informatics systems. *CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 557–566 (cit. on p. 20).

LiKamWa, R., Liu, Y., Lane, N. D., & Zhong, L. (2013). MoodScope: Building a Mood Sensor from Smartphone Usage Patterns. *Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services, Jun; 2013; Taipei, Taiwan*, 25–28 (cit. on pp. 19, 73, 106).

Lim, J. H., Zhan, A., Goldschmidt, E., Ko, J., Chang, M., & Terzis, A. (2012). HealthOS: A platform for pervasive health applications. *Proceedings of the Second ACM Workshop on Mobile Systems, Applications, and Services for Health-Care*, 1–6 (cit. on p. 27).

Ma, Y., Xu, B., Bai, Y., Sun, G., & Zhu, R. (2012). Daily Mood Assessment Based on Mobile Phone Sensing. *2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks*, 142–147 (cit. on pp. 19, 73).

Martinent, G., Gareau, A., Lienhart, N., Nicaise, V., & Guillet-Descas, E. (2018). Emotion profiles and their motivational antecedents among adolescent athletes in intensive training settings. *Psychology of Sport and Exercise*, *35*, 198–206 (cit. on pp. 2, 3, 6).

Mcabe, K. O., Mack, L., & Fleeson, W. (2012). *A guide for data cleaning in experience sampling studies*. Guilford Press, New York, NY, US. (Cit. on pp. 50, 62).

McNair, D. M., Lon, M., & Droppelman, L. F. (1992). *Revised manual for the profile of mood states*. Educational; Industrial Testing Services. (Cit. on p. 109).

McNair, D. M., Lorr, M., & Doppleman, L. F. (1971). *Manual for the profile of mood states*. Educational; Industrial Testing Service San Diego, CA. (Cit. on p. 14).

Meschtscherjakov, A., Weiss, A., & Scherndl, T. (2009). Utilizing emoticons on mobile devices within esm studies to measure emotions in the field. *International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI'09), September 15 - 18, 2009, Bonn, Germany* (cit. on p. 72).

Min, C., Yoo, C., Lee, Y., & Song, J. (2011). Healthopia: Towards your well-being in everyday life. *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, 1–5 (cit. on p. 27).

Min, J.-K., Doryab, A., Wiese, J., Amini, S., Zimmerman, J., & Hong, J. I. (2014). Toss 'n' turn: Smartphone as sleep and sleep quality detector. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 477–486 (cit. on p. 22).

Min, J.-K., Wiese, J., Hong, J. I., & Zimmerman, J. (2013). Mining smartphone data to classify life-facets of social relationships. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 285–294 (cit. on p. 22).

Mitchell, T. R., Thompson, L., Peterson, E., & Cronk, R. (1997). Temporal adjustments in the evaluation of events: The "rosy view". *Journal of Experimental Social Psychology*, *33*(4), 421–448 (cit. on p. 17).

Mitjà, O., Arenas, À., Rodó, X., Tobias, A., Brew, J., & Benlloch, J. M. (2020). Experts' request to the Spanish Government: Move Spain towards complete lockdown. *Lancet*, *395* (cit. on p. 76).

Moccia, L., Janiri, D., Pepe, M., Dattoli, L., Molinaro, M., Martin], V. [, Chieffo, D., Janiri, L., Fiorillo, A., Sani, G., & Nicola], M. [(2020). Affective temperament, attachment style, and the psychological impact of the COVID-19 outbreak: An early report on the Italian general population. *Brain, Behavior, and Immunity*, *725*, 138539 (cit. on p. 77).

Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual Review of Clinical Psychology*, *13*(1), 23–47 (cit. on pp. 20, 21, 73).

Morrison, W., Guerdan, L., Kanugo, J., Trull, T., & Shang, Y. (2018). TigerAware: An Innovative Mobile Survey and Sensor Data Collection and Analytics System. *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, 115–122 (cit. on p. 27).

Moshe, I., Terhorst, Y., Opoku Asare, K., Sander, L. B., Ferreira, D., Baumeister, H., Mohr, D. C., & Pulkki-Råback, L. (2021). Predicting Symptoms of Depression and Anxiety Using Smartphone and Wearable Data. *Frontiers in Psychiatry*, *12*, 625247 (cit. on p. 73).

Müller, S. R., Peters, H., Matz, S. C., Wang, W., & Harari, G. M. (2020). Investigating the Relationships between Mobility Behaviours and Indicators of Subjective Well–Being Using Smartphone–Based Experience Sampling and GPS Tracking [Publisher: SAGE Publications Ltd]. *European Journal of Personality*, *34*(5), 714–732 (cit. on pp. 8, 96, 97).

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry*, *17*(2), 123–132 (cit. on p. 18).

Naragon-Gainey, K. (2019). Affective models of depression and anxiety: Extension to within-person processes in daily life. *Journal of Affective Disorders*, *243*, 241–248 (cit. on p. 19).

Naughton, F., Riaz, M., & Sutton, S. (2016). Response Parameters for SMS Text Message Assessments Among Pregnant and General Smokers Participating in SMS Cessation Trials. *Nicotine & Tobacco Research*, *18*(5), 1210–1214 (cit. on p. 59).

Nicholls, A. R., Backhouse, S. H., Polman, R. C. J., & McKenna, J. (2009). Stressors and affective states among professional rugby union players: Stressors and affect. *Scandinavian Journal of Medicine & Science in Sports*, *19*(1), 121–128 (cit. on p. 3).

Nussbaumer-Streit, B., Mayr, V., Dobrescu, A. I., Chapman, A., Persad, E., Klerings, I., Wagner, G., Siebert, U., Christof, C., Zachariah, C., & Gartlehner, G. (2020). Quarantine alone or in combination with other public health measures to control COVID-19: A rapid review. *Cochrane Database of Systematic Reviews*, *4* (cit. on p. 76).

Odriozola-González, P., Planchuelo-Gómez, A., Irurtia, M. J., & de Luis-García, R. (2020). Psychological effects of the COVID-19 outbreak and lockdown among students and workers of a Spanish university. *Psychiatry Research*, *290*, 113108 (cit. on p. 77).

Papini, N. M., Yang, C.-H., Do, B., Mason, T. B., & Lopez, N. V. (2020). External contexts and movement behaviors in ecological momentary assessment studies: A systematic review and future directions. *International Review of Sport and Exercise Psychology*, 1–31 (cit. on pp. 18, 96, 132).

Parkinson, B. (1996). Emotions are social. *British Journal of Psychology*, *87*(4), 663–683 (cit. on p. 2).

Parsons-Smith, R. L., Terry, P. C., & Machin, M. A. (2017). Identification and Description of Novel Mood Profile Clusters. *Frontiers in Psychology*, *8* (cit. on pp. 109, 128).

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In *Handbook of research methods in personality psychology*. New York, NY: Guilford. (Cit. on p. 20).

Perera, C., Jayaraman, P. P., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2014). MOSDEN: An Internet of Things Middleware for Resource Constrained Mobile Devices [ISSN: 1530-1605]. *2014 47th Hawaii International Conference on System Sciences*, 1053–1062 (cit. on p. 27).

Quartiroli, A. (2018). Cross-Cultural Validation of Mood Profile Clusters in a Sport and Exercise Context. *Frontiers in Psychology*, *9*, 10 (cit. on p. 109).

Rachuri, K. K., Musolesi, M., Mascolo, C., Rentfrow, P. J., Longworth, C., & Aucinas, A. (2010). Emotionsense: A mobile phones based adaptive platform for experimental social psychology research. *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, 281–290 (cit. on p. 67).

Raento, M., Oulasvirta, A., & Eagle, N. (2009). Smartphones: An emerging tool for social scientists. *Sociological Methods & Research*, *37*(3), 426–454 (cit. on pp. 36, 67, 68).

Ranjan, Y., Kerz, M., Rashid, Z., Boettcher, S., Dobson, R. J. B., & Folarin, A. A. (2018). RADAR-base: A Novel Open Source m-Health Platform. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 223–226 (cit. on p. 27).

Rathschlag, M., & Memmert, D. (2013). The Influence of Self-Generated Emotions on Physical Performance: An Investigation of Happiness, Anger, Anxiety, and Sadness. *Journal of Sport and Exercise Psychology*, *35*(2), 197–210 (cit. on p. 2).

Rawassizadeh, R., Tomitsch, M., Wac, K., & Tjoa, A. M. (2013). UbiqLog: A generic mobile phone-based life-log framework. *Personal and Ubiquitous Computing*, *17*(4), 621–637 (cit. on p. 27).

RDA COVID-19 Working Group. (2020). Recommendations and guidelines. *Research Data Alliance* (cit. on p. 83).

Rickard, N., Arjmand, H.-A., Bakker, D., & Seabrook, E. (2016). Development of a Mobile Phone App to Support Self-Monitoring of Emotional Well-Being: A Mental Health Digital Innovation. *JMIR Mental Health*, *3*(4), e49 (cit. on p. 19).

Robazza, C. (2006). Emotion in sport: An izof perspective. In *Literature reviews in sport psychology*. Nova Science Publishers, Inc. (Cit. on pp. 3, 69).

Robazza, C., Pellizzari, M., Bertollo, M., & Hanin, Y. L. (2008). Functional impact of emotions on athletic performance: Comparing the IZOF model and the directional perception approach. *Journal of Sports Sciences*, *26*(10), 1033–1047 (cit. on p. 3).

Rokach, L., & Maimon, O. (2006). Clustering methods. In *Data mining and knowledge discovery handbook* (pp. 321–352). Springer. (Cit. on p. 113).

Roser, M., Ritchie, H., Ortiz-Ospina, E., & Hasell, J. (2020). Coronavirus pandemic (COVID-19). *Our World in Data* (cit. on p. 76).

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 (cit. on p. 113).

Rüegger, D., Stieger, M., Nißen, M., Allemand, M., Fleisch, E., & Kowatsch, T. (2020). How Are Personality States Associated with Smartphone Data? [Publisher: SAGE Publications Ltd]. *European Journal of Personality*, *34*(5), 687–713 (cit. on pp. 2, 73, 132).

Ruiz-Zafra, Á., Benghazi, K., Noguera, M., & Garrido, J. L. (2013). Zappa: An Open Mobile Platform to Build Cloud-Based m-Health Systems. In A. van Berlo, K. Hallenborg, J. M. C. Rodríguez, D. I. Tapia, & P. Novais (Eds.), *Ambient Intelligence - Software and Applications* (pp. 87–94). Springer International Publishing. (Cit. on p. 27).

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178 (cit. on pp. 15, 16).

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493–502 (cit. on p. 16).

Sauro, J. (2011). *A practical guide to the system usability scale*. Createspace Independent Publishing Platform. (Cit. on pp. 51, 57).

Schimmack, U. (2003). Affect Measurement in Experience Sampling Research. *Journal of Happiness Studies*, *4*(1), 79–106 (cit. on p. 69).

Servia-Rodríguez, S., Rachuri, K. K., Mascolo, C., Rentfrow, P. J., Lathia, N., & Sandstrom, G. M. (2017). Mobile Sensing at the Service of Mental Well-being: A Large-scale Longitudinal Study. *Proceedings of the 26th International Conference on World Wide Web - WWW '17*, 103–112 (cit. on pp. 2, 73, 96, 97).

Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., & Agha, R. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery*, *76*, 71–76 (cit. on p. 76).

Stone, A. A., Kessler, R. C., & Haythomthwatte, J. A. (1991). Measuring daily events and experiences: Decisions for the researcher. *Journal of Personality*, *59*(3), 575–607 (cit. on pp. 50, 61, 66).

Tams, S., Hill, K., Ortiz de Guinea, A., Thatcher, J., & Grover, V. (2014). Neurois—alternative or complement to existing methods? illustrating the holistic effects of neuroscience and self-reported data in the context of technostress research. *Journal of the Association for Information Systems*, *15*(10) (cit. on p. 5).

Tangmunarunkit, H., Hsieh, C. K., Longstaff, B., Nolen, S., Jenkins, J., Ketcham, C., Selsky, J., Alquaddoomi, F., George, D., Kang, J., Khalapyan, Z., Ooms, J., Ramanathan, N., & Estrin, D. (2015). Ohmage: A General and Extensible End-to-End Participatory Sensing Platform. *ACM Transactions on Intelligent Systems and Technology*, *6*(3), 38:1–38:21 (cit. on p. 27).

Terry, P. C. (2003). An Overview of Mood and Emotions in Sport. *An overview of mood and emotions in sport. In: 4th Asia-South Pacific Association of Sport Psychology International Congress* (cit. on pp. 3, 109).

Thompson, R. N. (2020). Epidemiological models are important tools for guiding COVID-19 interventions. *BMC Medicine*, *18*(152) (cit. on p. 76).

Tobías, A. (2020). Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up. *Science of The Total Environment*, *725*, 138539 (cit. on p. 76).

Torous, J., 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), e16 (cit. on p. 20).

Vales-Alonso, J., Lopez-Matencio, P., Gonzalez-Castano, F. J., Navarro-Hellin, H., Banos-Guirao, P. J., Perez-Martinez, F. J., Martinez-Alvarez, R. P., González-Jimenez, D., Gil-Castineira, F., & Duro-Fernandez, R. (2010). Ambient Intelligence Systems for Personalized Sport Training. *Sensors*, *10*(3), 2359–2385 (cit. on p. 4).

van Berkel, N., Ferreira, D., & Kostakos, V. (2018). The Experience Sampling Method on Mobile Devices. *ACM Computing Surveys*, *50*(6), 1–40 (cit. on pp. 18, 50, 62, 66–68, 96).

van Berkel, N., Goncalves, J., Koval, P., Hosio, S., Dingler, T., Ferreira, D., & Kostakos, V. (2019). Context-Informed Scheduling and Analysis: Improving Accuracy of Mobile Self-Reports. *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, (51), 1–12 (cit. on pp. 18, 50, 56, 66, 73).

van Berkel, N., Goncalves, J., Lovén, L., Ferreira, D., Hosio, S., & Kostakos, V. (2019). Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. *International Journal of Human-Computer Studies*, *125*, 118–128 (cit. on p. 63).

VanDam, M., & De Palma, P. (2019). A modular, extensible approach to massive ecologically valid behavioral data. *Behavior Research Methods*, *51*(4), 1754–1765 (cit. on pp. 9, 96).

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, *11*(110), 3371–3408 (cit. on p. 21).

Wac, K., & Tsiourti, C. (2014). Ambulatory Assessment of Affect: Survey of Sensor Systems for Monitoring of Autonomic Nervous Systems Activation in Emotion. *IEEE Transactions on Affective Computing*, *5*(3), 251–272 (cit. on p. 13).

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., & Campbell, A. T. (2014). Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3–14 (cit. on p. 67).

Wang, R., Harari, G. M., Hao, P., Zhou, X., & Campbell, A. T. (2015). Smartgpa: How smartphones can assess and predict academic performance of college students. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (cit. on p. 22).

Watson, D., A, C. L., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The panas scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070 (cit. on p. 15).

Wenemark, M., Persson, A., Brage, H. N., Svensson, T., & Kristenson, M. (2011). Applying Motivation Theory to Achieve Increased Response Rates, Respondent Satisfaction and Data Quality. *Journal of Official Statistics*, *27*(2), 393–414 (cit. on p. 59).

Wiskott, L., & Sejnowski, T. J. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural computation*, *14*(4), 715–770 (cit. on p. 21).

Wohlfahrt-Laymann, J., Hermens, H., Villalonga, C., Vollenbroek-Hutten, M., & Banos, O. (2019). Mobilecognitracker: A mobile experience sampling tool for tracking cognitive behaviour. *Journal of Ambient Intelligence and Humanized Computing*, *10*, 2143–2160 (cit. on p. 39).

Wu, A. D. (2014). Varimax loadings. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 6915–6916). Springer Netherlands. (Cit. on p. 111).

Xie, K., Heddy, B. C., & Vongkulluksn, V. W. (2019). Examining engagement in context using experience-sampling method with mobile technology. *Contemporary Educational Psychology*, *59*, 101788 (cit. on p. 18).

Xiong, H., Huang, Y., Barnes, L. E., & Gerber, M. S. (2016). Sensus: A cross-platform, general-purpose system for mobile crowdsensing in human-subject studies. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 415–426 (cit. on p. 27).

Yang, H., & Ma, J. (2020). How an epidemic outbreak impacts happiness: Factors that worsen (vs. protect) emotional well-being during the coronavirus pandemic. *Psychiatry Research*, *289*, 113045 (cit. on p. 77).

Yang, Y. S., Ryu, G. W., & Choi, M. (2019). Methodological Strategies for Ecological Momentary Assessment to Evaluate Mood and Stress in Adult Patients Using Mobile Phones: Systematic Review. *JMIR mHealth and uHealth*, *7*(4), e11215 (cit. on pp. 5, 19, 50, 61, 68, 99).

Yu, H., Lee, L., & Madera, J. M. (2021). Collecting Repeated Data Over Time: Applying Experience Sampling Methodology to the Hospitality Management Context [Publisher: SAGE Publications Inc]. *Cornell Hospitality Quarterly*, *62*(1), 62–75 (cit. on pp. 8, 96).

Zhang, Z., & Chen, W. (2019). A systematic review of the relationship between physical activity and happiness. *Journal of Happiness Studies*, *20*, 1305–1322 (cit. on p. 129).

Zhenyu Chen, Lin, M., Fanglin Chen, Lane, N. D., Cardone, G., Rui Wang, Tianxing Li, Yiqiang Chen, Choudhury, T., & Campbell, A. T. (2013). Unobtrusive sleep monitoring using smartphones. *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, 145–152 (cit. on p. 22).

Zuckerman, M., & Lubin, B. (1985). *Manual for the maacl-r: The multiple affect adjective check list-revised*. Educational; Industrial Testing Service San Diego, CA. (Cit. on p. 14).

Zuckerman, M., & Lubin, B. (1965). *Manual for the multiple affect adjective check list*. Educational; Industrial Testing Service San Diego, CA. (Cit. on p. 14).

Webpages

Bailon, C. (2019). *AWARE Plugin: ESM Flexible*. Retrieved December 21, 2021, from https://github.com/carlos-bailon/com.aware.plugin.esm.flexible. (Cit. on p. 35)

Bailon, C., Goicoechea, C., Sanabria, D., Pomares, H., Damas, M., Banos, O., & Perakakis, P. (2020). *¿Cómo evolucionan nuestras emociones durante una pandemia?* Retrieved December 30, 2021, from https://theconversation.com/como-evolucionan-nuestras-emociones-durante-una-pandemia-150467. (Cit. on p. 89)

Creative Commons. (2021). *Attribution 4.0 international (cc by 4.0)*. Retrieved January 18, 2022, from https://creativecommons.org/licenses/by/4.0/. (Cit. on pp. 74, 83)

Docker, Inc. (2021). *Docker*. Retrieved December 23, 2021, from https://www.docker.com/. (Cit. on p. 40)

EBSCO Industries Inc. (2021). *SPORTDiscus bibliographic database*. Retrieved December 7, 2021, from https://www.ebsco.com/products/research-databases/sportdiscus. (Cit. on p. 3)

EllisLab. (2021). *CodeIgniter*. Retrieved December 23, 2021, from https://codeigniter.com/. (Cit. on p. 42)

Fitbit. (2021). *Fitbit*. Retrieved December 16, 2021, from https://www.fitbit.com. (Cit. on p. 24)

GO FAIR. (2020). *FAIR data principles*. Retrieved January 5, 2022, from https://www.go-fair.org/. (Cit. on p. 83)

Google. (2021). *Flutter*. Retrieved December 16, 2021, from https://flutter.dev/. (Cit. on p. 29)

Hipp, R. D. (2021). *SQLite*. Retrieved December 16, 2021, from https://www.sqlite.org. (Cit. on pp. 24, 37)

Pallets Projects. (2021). *Flask*. Retrieved December 24, 2021, from https://flask.palletsprojects.com/en/2.0.x/. (Cit. on p. 45)

Simple-XML Serialization framework for Java. (2012). Retrieved December 22, 2021, from http://simple.sourceforge.net/. (Cit. on p. 39)

Statista. (2021). *Number of smartphone users from 2016 to 2021*. Retrieved December 20, 2021, from https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/. (Cit. on p. 4)

World Health Organization. (2020). *WHO director-general's opening remarks at the media briefing on COVID-19*. Retrieved March 28, 2020, from https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. (Cit. on p. 76)

List of Figures

List of Tables

List of Listings

Curriculum vitae

Personal Information

Carlos Bailon
Born July 7, 1994, Granada, Spain

Education

2017 – 2022	PhD studies at CASIP research group, Department of Computer Architecture and Technology, University of Granada, Spain
2016 – 2017	MsC Data Science and Computer Engineering at Higher Technical School of Computer and Telecommunication Engineering, University of Granada, Spain
2016 – 2012	BEng Industrial Electronics Engineering, Faculty of Science, 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 (FPU Fellowship, FPU16/04376), Granada, Spain
2019 – 2019	Research Assistant at Biomedical Signals and Systems (BSS) research group, University of Twente, Enschede, Netherlands
2015 – 2017	University of Granada Research Fellowship, Department of Computer Architecture and Technology, University of Granada, Spain

Publications list

International Journals (SCI-indexed)

Bailon, C., Damas, M., Pomares, H., Banos, O. (2021) Multilevel methodological approach to the context-aware analysis of longitudinal behavioral data, **Journal of Ambient Intelligence and Humanized Computing** [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]

Bailon, C., Goicoechea, C., Banos, O., Damas, M., Pomares, H., Correa, A., Sanabria, D., Perakakis, P. (2020) CoVidAffect, real-time monitoring of mood variations following the COVID-19 outbreak in Spain, **Nature Scientific Data**, 365, pp. 1–10

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(5), pp. 3430–3453

Bailon, C., Damas, M., Pomares, H., Banos, O. (2018) SPIRA: an automatic system to support lower limb injury assessment, **Journal of Ambient Intelligence and Humanized Computing**, 10, pp. 2111–2123

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

Bailon, C., Damas, M., Pomares, H., Banos, O. (2018) Intelligent Monitoring of Affective Factors Underlying Sport Performance by means of Wearable and Mobile Technology. In: Proceedings of the 12th International Conference on Ubiquitous Computing and Ambient Intelligence (**UCAmI 2018**), Punta Cana, Dominican Republic, December 4–7

Bailon, C., Damas, M., Pomares, H., Banos, O. (2017) Automatic 2D Motion Capture System for Joint Angle Measurement. In: Proceedings of the 14th International Work-Conference on Artificial Neural Networks (**IWANN 2017**), Cádiz, Spain, June 14

Bailon, C., Damas, M., Pomares, H., Banos, O. (2016) First Approach to Automatic Measurement of Frontal Plane Projection Angle During Single Leg Landing Based on Depth Video. In: Proceedings of the 10th International Conference on Ubiquitous Computing and Ambient Intelligence (**UCAmI 2016**), Las Palmas de Gran Canaria, Spain, November 29–December 2

Dissemination Articles

Bailon, C., Goicoechea, C., Sanabria, D., Pomares, H., Damas, M., Banos, O., Perakakis, P. (2020) ¿Cómo evolucionan nuestras emociones durante una pandemia?, **The Conversation**, https://theconversation.com/como-evolucionan-nuestras-emociones-durante-una-pandemia-150467