



# Connecting physical activity with context and motivation: a user study to define variables to integrate into mobile health recommenders

Ine Coppens<sup>1</sup> · Toon De Pessemier<sup>1</sup> · Luc Martens<sup>1</sup>

Received: 8 April 2022 / Accepted in revised form: 15 May 2023 / Published online: 24 June 2023  
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

## Abstract

In this paper, we aim to improve existing health recommender systems by defining relevant contextual and motivational variables to recommend physical activities and collect appreciation feedback. Current health recommenders do not sufficiently include users' context and motivational theory when personalizing health suggestions. To bridge these gaps, we conducted a 21-day longitudinal user study with 36 participants using our Android app with collected sensor data and Ecological Momentary Assessments to collect daily activities, mood, and motivation. This study resulted in a dataset of 724 activities. Two approaches to determine feature relevance were followed: variable importances analysis on 40 input variables, and statistical analysis of mean differences in outcome variables across contexts. Our findings suggest recommending activity duration, intensity, location, and type by incorporating: company, situation (e.g., free time or work), happiness, calmness, energy level, physical complaints, and motivation. As such, we propose opportunities for future health recommenders to integrate these data with contextual pre-filtering techniques, extended with our suggestions for automatically collected weather, location types, step count, and time. We also propose to use mood and motivation as appreciation feedback to focus on user well-being and boost motivation.

**Keywords** Physical activity · mHealth · Recommender system · Context · Motivation · Ecological momentary assessments

---

✉ Ine Coppens  
ine.coppens@ugent.be

Toon De Pessemier  
toon.depessemier@ugent.be

Luc Martens  
luc.l.martens@ugent.be

<sup>1</sup> imec - WAVES, Ghent University, iGent - Technologiepark 126, 9052 Ghent, Belgium

## 1 Introduction

Physical inactivity is a risk factor for noncommunicable diseases, such as cardiovascular diseases, cancer, and diabetes (World Health Organization 2020b). In addition to physical health, physical activity (PA) also contributes to mental health (Penedo and Dahn 2005; Young et al. 2016; Stamatakis et al. 2019; Biddle et al. 2019). The World Health Organization (WHO) advises a minimum of 150 min of moderate intensity PA per week for adults (World Health Organization 2020a). However, 44.2% of the adult population in Belgium in 2018 (Drieskens et al. 2018), 28.6% in the European Union in 2013 (Gerovasili et al. 2015), and 27.5% globally in 2016 (World Health Organization 2020b) did not achieve this weekly minimum. Therefore, motivating people to engage in more PA would reduce numerous negative health outcomes.

Today's technology offers us several possibilities to improve people's health. Mobile health (mHealth) aims at health improvement supported by mobile technologies, such as smartphone apps (Miyamoto et al. 2016). However, users often stop using these mHealth apps and digital interventions (Miyamoto et al. 2016; Cheung et al. 2019). To gain effectiveness and retain users, personalization and context could be added to the interventions (Miyamoto et al. 2016). One way to achieve this, is to generate personalized health recommendations with Recommender System (RS) algorithms.

RSs generate personalized and useful suggestions by predicting user preferences and helping users with decision-making in the information overload (Ricci et al. 2022). They are widely used in websites, e-commerce, and online entertainment (Ricci et al. 2022), but can also be of great use to personalize interventions in the health domain (Hors-Fraile et al. 2018) and in mHealth. In this way, a health recommender system (HRS) aims to motivate users to change their behavior by proposing healthier suggestions, tailored to their profile and characteristics (De Croon et al. 2021), and by automatically generating personalized health information (Hors-Fraile et al. 2018).

People have to be motivated to execute the recommendations they receive, which is not always the case (Cheung et al. 2019). They also have to be motivated to maintain the healthier behavior in the long term for sustained improvements in their health (Miyamoto et al. 2016). However, current research in HRSs does not include enough motivational theory and behavior change theories (Hors-Fraile et al. 2018). We support that HRSs for motivating behavior change in PA should be solidly based on scientific evidence (Hors-Fraile et al. 2018).

Additionally, Ricci et al. (2022) explain that integrating contextual information in the recommendations is important because user preferences might differ across different contexts. In these context-aware RSs, "context" is used to describe the setting in which the recommendations are generated (De Croon et al. 2021), such as the user's mood, company (Hussein et al. 2014), weather, and time (Baltrunas and Ricci 2014). However, existing HRSs do not include sufficient contextual features (Hors-Fraile et al. 2018). Moreover, to the best of our knowledge, limited research exists of which relevant contextual factors to integrate in HRSs for PAs, while this has been more thoroughly studied in RSs for music (Odić et al. 2013; Baltrunas and Amatriain 2009), movies (Baltrunas and Ricci 2014; Zheng et al. 2013), and other mobile recommenders (Ricci

et al. 2022). By not utilizing the full potential of the available context, existing HRSs might miss opportunities to increase personalization.

The main contribution of this study is an overview of relevant contextual and motivational features to collect and integrate in an HRS (Odić et al. 2013). As such, we define two tasks of an HRS in which these features can contribute: (1) predicting PA recommendations and (2) collecting *appreciation feedback* from its users, which indicates how much they liked the consumed PAs and with which the RS algorithm can generate future PA recommendations. The data were collected in a longitudinal user study using an Android app installed on the participants' smartphone. With this app, we collected a large amount of manually provided information using Ecological Momentary Assessments (EMAs) to collect real-time data in participants' daily environment (Liao et al. 2016), and automatically collected information from smartphone sensors and online services. To the best of our knowledge, this is the first research that collects and combines this large amount of data, tackling the lack of context and motivation in current HRSs (Hors-Fraile et al. 2018).

The paper is structured as follows. Related work is summarized in Sect. 2, together with our two research questions. Next, Sect. 3 describes the Methods of our user study and two analysis approaches. Firstly, using statistical analyses, we determine whether the items are experienced or the users behave significantly different across contexts (Zheng et al. 2013), explained in 3.2. Secondly, we investigate the variable importances using Random Forests, a technique used for data exploring and understanding (Verikas et al. 2011), described in 3.3. The results of the analyses are covered in Sect. 4, and discussed in Sect. 5. Lastly, Sect. 6 presents the conclusion of this research.

## 2 Related work

Personalization and motivation for PA are well-researched topics. Some studies focus on personalizing training programs for a specific goal, such as training for a marathon by improving personal finish-times and providing sufficient recovery (Smyth et al. 2021). Others focus on improving people's general physical health, such as weight loss (Gasparetti et al. 2020), calorie loss (Rabbi et al. 2015), and establishing behavior change (op den Akker et al. 2014). Some studies argue that both food and PA should be incorporated in an HRS for adopting healthy habits (Alcaraz-Herrera et al. 2022; Rabbi et al. 2015). In this paper, however, the focus is only on PA behavior, which we connect to people's motivation and context.

### 2.1 Motivational features

Since health interventions often experience high dropout rates (Cheung et al. 2019), it is essential that mHealth systems focus on motivating them for behavior change (Nurmi et al. 2020). Moreover, it is important that the systems are grounded in scientific evidence and motivational theory (Hors-Fraile et al. 2018).

A theory of motivation is the Self-Determination Theory (SDT) of Ryan and Deci (2000) that distinguishes between autonomous motivation, in which people experience

volition, and controlled motivation, in which people feel pressured. When people are autonomously motivated, they tend to have greater long-term persistence and sustained behavior change (Deci and Ryan 2008). The SDT continuum of Ryan and Deci (2000) contains the following regulation styles, ordered from most self-determined to least self-determined: *intrinsic motivation*, *integrated regulation*, and *identified regulation* for autonomous motivation, and *introjected regulation* and *external regulation* for controlled motivation (Deci and Ryan 2008). In contrast to these regulatory styles, *amotivation* refers to the absence of motivation (Deci and Ryan 2008).

Other apps and studies implement motivational theory to motivate for behavior change for PA, such as promoting autonomous motivation in the interventions in Nurmi et al. (2020) and defining motivational profiles in Fukuoka et al. (2018). Other studies aim to motivate by using personalization, such as the tailoring techniques of op den Akker et al. (2014), the tailoring of messages in Kamphorst et al. (2014), the tailoring of interventions based on personality traits in Ndulue et al. (2022), and implementing just-in-time adaptive interventions (JITAI) to provide the right type of support at the right time in Nahum-Shani et al. (2018).

In our study, we chose to collect motivational features based on the momentary motivation for PA and on SDT-based questionnaires. We selected the Behavioural Regulation in Exercise Questionnaire (BREQ) which is based on the SDT and measures motivation for exercises across the six regulatory styles (Mullan et al. 1997; Cid et al. 2018).

## 2.2 Contextual features

Smartphones offer us the potential to automatically collect data from its sensors, such as the GPS or accelerometer. Several applications and projects use a variation of sensor sources to collect contextual data, such as the environmental context (e.g., weather in Costa et al. (2017)), temporal context (e.g., seasonal patterns in Pontin et al. (2021)), social context (e.g., available PA buddies based on location in Dharia et al. (2018)), and other spatial context (e.g., greenspace and recreational locations as suggested by Pontin et al. (2021)). Additionally, they can automatically sense activities (e.g., stationary, running, and walking using the smartphone's accelerometer in Rabbi et al. (2015)), or integrate step count (e.g., detected by wearables in Gasparetti et al. (2020)). To the best of our knowledge, previous research mostly only focuses on one or two collected contextual features.

Previous work also mentions the need for additional research about collected contextual data. Firstly, Pontin et al. (2021) argue that more location data collected by smartphones should be connected to the PA behaviors. Secondly, Turrisi et al. (2021) emphasize analyzing the effect on PA from specific weather characteristics, such as cloud cover, humidity, and atmospheric pressure, because their influence is often disregarded (Brocherie et al. 2015). Lastly, the paper of Polignano et al. (2021) focuses on collecting data from emotions to predict what a user would like when having those feelings (e.g., whether a user wants to confirm their negative mood when feeling sad). Polignano et al. (2021) describe how these emotions can be integrated as contextual variables in a context-aware RS, and make it emotion-aware.

To collect emotion data in our study, we use EMAs that ask users about their daily mood, based on previous studies. Firstly, in the EMA study of Asselbergs et al. (2016), mood was self-reported with a one-dimensional and a two-dimensional measure. The latter was based on the circumplex model of affect of Russell (1980), which maps affect in a circle defined by the dimensions *arousal-sleepiness* and *pleasure-displeasure*. Correspondingly, we used *energy level* and *happiness*. Secondly, the EMA study of Pórarinsdóttir et al. (2019) integrates stress. Similarly, we added *calmness level* to measure this stress dimension. In addition to these mental components, we added *physical complaints* as a fourth dimension, as we hypothesize that this could be a reason why people do not want to have PA.

### 2.3 Defining relevant features

Ricci et al. (2022) emphasize that contextual factors must be relevant, and that relevance can vary across different domains (e.g., user location is not relevant in every domain). To determine which contextual information is relevant, and which can thereby contribute to the recommendation process, Odić et al. (2013) propose a methodology with statistical tests. In these tests, they investigate which factors result in significant differences in ratings that users assign to consumed items (Odić et al. 2013). In addition to statistics, another approach is to use machine learning for feature selection to identify relevant context data (Ricci et al. 2022), such as variable importances from Random Forests (Verikas et al. 2011).

Integrating the relevant contextual factors, the recommendation process can use the contextual pre-filtering, post-filtering, or modeling strategy (Ricci et al. 2022). In the pre-filtering strategy, the contextual information is used to filter the data before being given to the RS (Ricci et al. 2022). Since certain contextual factors, such as bad weather, work environment, or night time, can rule out some PAs, such as outdoor PA, we chose to focus on the pre-filtering approach in this paper. Applying this pre-filtering approach, we expect RSs for PA to not suffer as hard from the sparsity problem, which is the difficulty to calculate accurate recommendations when little consumption data are available (Ricci et al. 2022), because people engage daily in multiple PAs.

Multiple practical approaches can be used in pre-filtering. Firstly, *item-splitting* splits items depending on their context, such as day of the week (weekend or week day) and companion in Baltrunas and Ricci (2014). Secondly, *micro-profiling* splits the user profile into several sub-profiles, such as the same user in different time spans in Baltrunas and Amatriain (2009). Lastly, *UI splitting* is a combination of these two, such as the combination of item and user splitting in Zheng et al. (2013). In these splitting approaches, the statistical significance of the difference in means of the ratings across the different contexts can be analyzed with statistical tests, such as T-tests or other statistical metrics (Zheng et al. 2013).

#### 2.3.1 Relevance in the user's decision for physical activity

We investigate which contextual data are relevant, which means they have an effect on the participant's choice of PA duration, intensity, type, and to either have PA indoors

or outdoors. Knowing which contextual data have an effect on the user's decision, can help HRSs predict suited activities. Therefore, the first research question refers to how an HRS can provide useful recommendations to its users, and is stated as:

*RQ1: What are the relevant contextual and motivational features for an HRS to predict useful suggestions for PA duration, intensity, location, and type?*

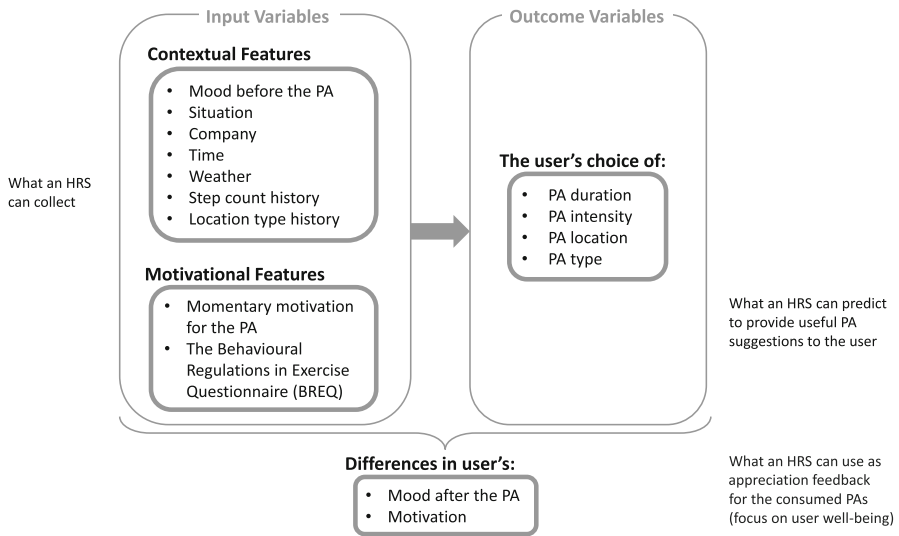
We expect differences in the choice of PA type and location depending on the weather. Outdoor activities are less likely when the weather is bad (Hussein et al. 2014), such as too high temperatures and rain (Wagner et al. 2019; Turrisi et al. 2021), and cloud cover and wind that can contribute to wind chill (Brocherie et al. 2015). Additionally, we hypothesize that PA duration and intensity increase with motivation because people with higher autonomous motivation are more likely to engage in more PA (Nurmi et al. 2020; Ryan and Deci 2000). We also expect that all PA outcome variables are dependent on mood, since consumption can differ across feelings (Polignano et al. 2021). Lastly, we presume that time of day, time spent in every location type, and steps already taken that day will influence PA behavior for the rest of the day because this is indication of how much PA was already done, similar to Gasparetti et al. (2020), and because everyone has a maximum PA volume per day (Pontin et al. 2021; Pelliccia et al. 2020).

### 2.3.2 Relevance in the user's appreciation for the activity

The ratings on the recommended items are the explicitly expressed preferences of the user for the consumed item (Ricci et al. 2022). In our user study, we do not ask the participants to assign a rating to their consumed PA to indicate how much they liked it. Rather, we ask them about their motivation for having engaged in this PA, and their mood after the PA. Because user preferences and item consumption might vary according to how the users feel, the rating evaluation alone does not represent whether the user liked the item or not (Polignano et al. 2021). Therefore, Polignano et al. (2021) emphasize that the specific mood at that moment should be integrated as a contextual variable. As such, our study focuses on the motivation and mood of the user after having engaged in the PA, as an expression of their appreciation or preference feedback for the item. Therefore, our second research question assumes collecting mood and motivation as RS feedback, and analyzing their significant differences across contexts:

*RQ2: What are the relevant contextual features of physical activities that influence mood and motivation?*

We expect differences in people's mood after the PA because PA has positive influence on people's mental health (Biddle et al. 2019; Polignano et al. 2021). We also expect differences in motivation depending on the company and situation. Firstly, because PA performance is better in team than alone (Feltz et al. 2011). Secondly, because motivation is higher when there is less pressure (Ryan and Deci 2000), such as free time PA, than during work, household tasks, and active transport (Wanner et al. 2012).



**Fig. 1** An overview of the variables, methods, and analyses shows the different contextual and motivational variables that were collected on the left. On the right, the outcome variables represent what an HRS can recommend (used in RQ1). At the bottom, motivation and mood represent what an HRS can use as appreciation feedback for its recommendations (used in RQ2)

### 3 Methods

An overview of the collected contextual and motivational variables is presented in Fig. 1. To collect these data, an Android app was designed and developed in Android Studio to the needs of this study. This app was named MoodApp, referring to the emoji present in its main screen, as shown in Fig. 2. Additionally, questionnaires were provided before and after the 21-day app study in MoodApp, which we refer to as the pre-test and post-test questionnaire.

The pre-test questionnaire included the BREQ to measure the different types of regulations for PA based on the SDT (Mullan et al. 1997). In our study, a selection of two questions for each of all six types were used (Markland and Tobin 2004; Wilson et al. 2006; Cid et al. 2018), and are elaborated in Appendix A. The post-test questionnaire contained questions that we designed ourselves, and selected questions from the Usefulness, Satisfaction, and Ease of use questionnaire for Reliability and Validity (USE) (Gao et al. 2018) to test our self-developed app, as summarized with their results in Figs. 7 and 8.

Between the pre- and post-test questionnaires, participants were asked to use the app for 21 days in their daily life following the principle of EMAs. EMAs are used to monitor participants in their typical environments to measure their real-time data (Liao et al. 2016). In this study, EMAs were used by asking the participants to self-report their mood and PA details in MoodApp before and after every PA, as seen on the left in Fig. 2. To increase user-friendliness, we chose to use emoji to assess people's mood. Phan et al. (2019) demonstrate that emoji are suited to measure affect. Following their

The figure displays two screenshots of the MoodApp interface. The left screenshot shows the main mood selection screen. It features four rows of emoji scales for 'Current mood', 'Stressed', 'Tired', and 'Lots of physical complaints'. Each row has seven emoji options ranging from sad to happy. Below these are radio buttons for 'When are you entering this information?' with options: 'Before the activity', 'After the activity' (selected), and 'End of the day'. There is also a note '(= no activities performed today)'. At the bottom, there are icons for information, location, and a star, along with a 'SUBMIT' button. The right screenshot shows a pop-up form titled 'How much time did you spend on "Running"?'. It includes a time selector showing 23 hours and 15 minutes, and 0 hours and 20 minutes. Below this are dropdown menus for 'What was the experienced intensity?' (Moderate), 'What was your motivation?' (Highly motivated), and 'Why/when did you do the activity?' (In my free time). The 'Specifications of the activity' section has radio buttons for 'indoors', 'outdoors', 'alone', 'with buddy', 'really easy', 'easy', 'medium', 'hard', and 'really hard'. An 'OK' button is at the bottom right.

**Fig. 2** Two screens of MoodApp in which participants were asked to enter their mood before and after every PA for 21 days using four rows of emoji (left), and the details of the chosen activity in the input fields (right)

future guidelines, we chose to extend our scale to a 7-point Likert scale, and a scale that can be mapped on “strongly disagree” to “strongly agree” from negative to positive affect (Phan et al. 2019). As such, we created our own emoji that range from open mouth oriented downwards (negative) to upwards (positive), with a neutral expression in the center.

Next, the participants could select whether they entered the information before or after the activity, or at the end of a day without any PA. When they selected *After the activity*, they provided more detailed information about the activity in a pop-up, shown on the right of Fig. 2. We asked them to add any activity that raises the heart rate, which includes light intensities as well, such as walking in the supermarket, with a minimum duration of five minutes. In this way, we follow World Health Organization (2020a) that defines PA as any kind of bodily movement, and we expected that it would be uncommon to have days without any PA at all.

In Table 1, an overview of all the manually collected data in MoodApp using EMAs is listed. The PA details were based on the description of PA as FITT: frequency, intensity (and corresponding metabolic equivalents (MET) from the Compendium of Physical Activities (Ainsworth et al. 2011)), time (duration), and type (Pelliccia et al.



**Table 1** The manually entered data in the EMAs included information about the PA, the motivation for it, and its context. The participants answered these at every PA submit. Later, durations were also converted to categorical values and intensities to numerical (based on the Compendium of Physical Activities (Ainsworth et al. 2011))

Variable	Values
Timing	Categorical: before or after the activity, or at the end of a day without any activities
PA duration	Numerical: minutes spent on the PA (converted to categorical: 0-9, 10-29, 30-59, 60-89, or $\geq 90$ min)
PA intensity	Categorical: light, moderate, or vigorous (converted to numerical: MET value ( $kcal.kg^{-1}.h^{-1}$ ))
PA location	Categorical: indoors or outdoors
PA type	Categorical: chosen from a list, or a new one (pre-processed into 21 types as elaborated in Appendix B)
Momentary PA motivation	5-point Likert scale for motivation for the PA of that moment: extremely motivated, highly motivated, medium motivated, a little motivated, or not motivated at all
Mood	7-point Likert scale, both before and after the PA: happiness, calmness level, energy level, and physical complaints
Situation	Categorical: <i>when</i> or <i>why</i> the PA was performed: free time, active transport, during work, or household task (based on Caspersen et al. (1985); Wanner et al. (2012))
Company	Categorical: alone or with a buddy

2020). Lastly, Table 2 shows an overview of the automatically collected context data and their sources.

### 3.1 Data collection and filtering

This study was targeted towards healthy adults (18 years or older) with an Android smartphone. They were recruited via the Sona System of Ghent University, which is a platform for research participation. Additionally, advertisement for the study was posted in a Facebook group for paid studies.

The study started on the 1st of November, 2021 for all participants. It ended when exactly 21 days of data were collected for that participant, which had to be completed before the end of the first week of December, 2021. Participants were asked to submit their information at least once a day: either before and after a PA, or at the end of a day without any activities. To motivate people to submit daily, an incentive of 30 euros was provided for those who submitted on 21 consecutive days, with a maximum of two consecutive days without any activities. Additionally, MoodApp sent out a notification

**Table 2** The automatically collected contextual data were collected using the smartphone's clock, GPS sensor, and accelerometer. Online services were used to request weather and location type at the user's coordinates

Variable	Values	Source
Time	Hour of the day Time of day (e.g., morning, evening) Day of the week Weekend or week day	Device's clock
Weather	Description (e.g., wind, clouds, rain) Cloud cover (%) Temperature (°C) Perceived temperature (°C) Wind speed (m/s) Humidity (%) Pressure (hPa)	Device's GPS + online API service of OpenWeather: <a href="https://openweathermap.org/">https://openweathermap.org/</a>
Step count (between the previous and current submit)	Amount of steps taken Steps per minute	Device's accelerometer
Location type (amount of minutes spent per type, between the previous and current submit)	Building Road Transportation Commerce Outdoors Health Education Travel	Device's GPS + online API service of OpenCage GmbH: <a href="https://opencagedata.com/">https://opencagedata.com/</a>

**Table 3** The descriptive statistics of the PA amounts, step count, and answers to the motivation questions show that the participants had on average 321 min of PA per week, but only 2633 steps per day were detected on average

Variable		Participants (N = 36)					$\alpha$
		N	Mean	SD	Min	Max	
Minutes PA / week	All intensities	36	321	198	89	1068	
	Light	36	155	150	3.75	858	
	Moderate	36	157	123	13.33	490	
	Vigorous	13	25	16	2.50	65	
	Free time	36	219	129	30	600	
	Active transport	31	70	38	15	176	
	Household	19	78	47	15	185	
	During work	12	317	257	30	720	
Amount of PA / week		36	5.97	2.23	2.00	11.33	
Amount of steps / day		24	2633	1828	393	6148	
1 PA session	MET value	36	3.99	1.83	2.00	14.70	
	Duration	36	54.63	64.57	5.00	540.00	
Motivation	Amotivation	36	0.67	0.80	0.00	3.00	.65
	External	36	0.89	0.81	0.00	3.00	.74
	Introjected	36	1.99	1.14	0.00	3.50	.75
	Identified	36	2.81	0.86	0.50	4.00	.67
	Integrated	36	1.69	1.35	0.00	4.00	.93
	Intrinsic	36	2.50	1.15	0.00	4.00	.92
	Momentary	36	3.19	1.04	1.00	5.00	

several times per day from 10 am to 10 pm to remind the participants to use the app, with fixed 2-hour intervals, or longer after a recent submit.

In total, 78 participants installed MoodApp, but only 61 started to submit information in the app. During data cleaning and preprocessing, we took into account our criteria communicated to the participants to remove data based on the requirements for this study. An overview of the criteria, omitted participants, and removed data points are listed in Appendix B.

The resulting total sample size is 36. These 36 participants created collectively 1427 data points in the dataset, which is approximately 40 measurements per participants. 625 of these data points were submitted before the PA, 724 after, and only 78 were submitted when a participant indicated that there was no PA that day. The descriptive statistics of this dataset are shown in Table 3. The N-column displays the amount of participants who have at least one value for the corresponding variable and thereby contribute to the calculated descriptive statistics values. These values contain standard deviations often larger than half of the mean, which indicates a large spread in the data. Table 3 also contains the answers to the BREQ questions and corresponding Cronbach's alpha values of each regulatory style. Of these six regulatory styles, the autonomous motivation styles show higher means than the controlled styles.

### 3.2 Statistical analysis

When the study was finished, two structures of datasets were obtained. In the dataset with the answers to the pre- and post-test questionnaires, each row represents a participant. In contrast, the other dataset with the activities of the 21-day study had to be analyzed differently, because multiple data points per participants are spread over multiple rows. Both datasets were preprocessed using Python (version 3.9.7) in Anaconda (version 2.1.1).

For the statistical analysis of both datasets, IBM SPSS Statistics Version 28 was used. Following Zheng et al. (2013), statistical metrics were used to determine significant differences across contexts to determine relevant features. In the longitudinal dataset, the data were correlated because they were repeated measures within individuals. Therefore, we had to conduct the analysis with multilevel modeling with two levels and user ID as the subject variable (Heck et al. 2013).

In IBM SPSS Statistics, this analysis was executed with a two-level Generalized Estimating Equation (GEE). Following Pekár and Brabec (2018), the working correlation matrix of the GEE was set to  $AR(1)$  because this autoregressive model of the first order is preferred for regularly repeated measurements on the same user. This study had on average 40 measurements per person. Since mood was scored with seven possible answers and motivation with five, they can be seen as a Likert scale with 5 or more categories, making them usable as continuous variables in a regression analysis (Norman 2010).

In addition to the statistical significance using the p-value, we also analyzed the effect size. A low p-value does not guarantee a practical significance of the effect (Sun et al. 2010) and does not show the size of the effect (Kim 2017). The effect size complements the p-value as it quantifies the size of association strength (Sun et al. 2010). We analyzed the effect size using Cramer's V (Kim 2017). For  $df = 1$ , a Cramer's V value of .10 indicates a small effect, .30 a medium effect, and .50 a large effect (Kim 2017). However, interpreting these effect sizes should not only be done using these cut-off values, but also by comparing the effect sizes within and between studies (Sun et al. 2010).

Statistical analyses were conducted on the dependent variables PA duration, PA intensity, PA location, mood, and motivation. Due to the large amount of PA type categories and several with only a few instances, as elaborated in Appendix B, we did not perform a GEE with PA type as the dependent variable.

### 3.3 Variable importance analysis

A large amount of data were collected in this study. Thus, many statistical analyses with several combinations of input and output variables were possible. Moreover, some of these data have many dimensions and are less structured, such as the time spent in each of the different detected location types by OpenCage, and the six regulation types of the BREQ. Therefore, our approach to determine the relevance of all input data together was to determine their variable importances obtained from a Random Forest (Verikas et al. 2011).

Variable importance evaluations from Random Forests are widely used for exploring and understanding multidimensional data (Verikas et al. 2011). Random Forests are a machine learning technique that use a combination of tree predictors and a variable importance measure, based on the permutation importance measure proposed by Breiman (2001). These importance measures provide a measure of the contribution of each variable to predict the outcome variable (El Haouij et al. 2019). A more detailed description of the variable importance measure can be found in El Haouij et al. (2019). Although this approach does not take into account the multilevel structure of our dataset, possible correlations between the input variables, and the direction of the relations, this did allow us to explore the relative relevance of each feature. As such, we explored which features are more relevant than the others, and applied statistical analysis on the most relevant ones.

In total, four Random Forest Classifiers were constructed in this study: for the categorical outcome variables PA duration, intensity, location, and type, using the scikit-learn package (<https://scikit-learn.org/>) version 1.0.2. In this way, the input variables, as depicted in Fig. 1, were included in the Random Forest Classifier. Using Randomized Search with 5-fold cross validation that randomly samples a fixed number of parameter combinations, we determined the most optimal hyperparameters for the Random Forest Classifiers.

Before running the Randomized Search, all input variables had to be preprocessed in Python to ensure they could be interpreted correctly and no data were missing. Firstly, we connected the mood before the PA to the corresponding record after the PA, since we wanted to interpret how the mood was before engaging in the PA. Similarly, since location type and step count history were both measured between the previous and current submit, their values represent the activity during the PA. Thus, location type and step count history before the PA were also connected to the corresponding record after the PA.

Next, we looked at missing data records. Through EMAs, values for situation, company, and motivation were always provided, and none of these features were missing. Because momentary motivation and mood after the PA were measured with a 5-point and 7-point Likert scale respectively, these could be included as numerical input variables. However, not all users provided mood data before every PA, and mood scores before the PA were imputed with the corresponding user's mean. For the automatically collected data, all time features were always present, but not all smartphones provided GPS or accelerometer data. Therefore, the missing numerical weather characteristics and step count variables were imputed with the general mean, and missing minutes spent in location types with zero. Missing categorical features were imputed with the most common category: weather descriptions with "Clouds" and current location type categories with "building".

Lastly, numerical dummy variables were created for all categorical input variables since these cannot be entered in a Random Forest: free time/during work/active transport/household task, buddy/alone, weekend/week day, morning/noon/afternoon/evening, rain/clouds/clear, and location type currently indoors/outdoors (based on current location type). For the six BREQ regulation styles, the mean per construct was calculated for each participant, and added to every submitted PA record of the corresponding participant.

As a result of this preprocessing, 40 input variables were obtained. Using the code from Listing 1 in Appendix C, we performed the Randomized Search with these 40 input variables on the Random Forest Classifiers for each of the four outcome variables. To avoid overfitting, we generalized by limiting the tree depth to 4. The resulting hyperparameters, elaborated in Appendix C, were put in the Random Forest Classifiers, and feature importances of all 40 input variables were computed.

## 4 Results

### 4.1 The choice of activity duration

A Random Forest (train accuracy =.55, test accuracy =.59) was built with all 40 input variables and duration as categorical outcome variable, of which the conversion is detailed in Table 1. Figure 3 shows relevance of each input feature with respect to the outcome variable PA duration by ranking them according to their relative importance measures (El Haouij et al. 2019).

Situation and company, both categorical variables, ranked in the top ten of the relative variable importances for the PA duration. As such, we put these separately as independent variables in a GEE with the numerical PA duration as the dependent variable, which has values for all 724 records. Firstly, PA in free time significantly predicted a longer duration than active transport ( $p < .001$ , Cramer’s  $V = .29$ ) and

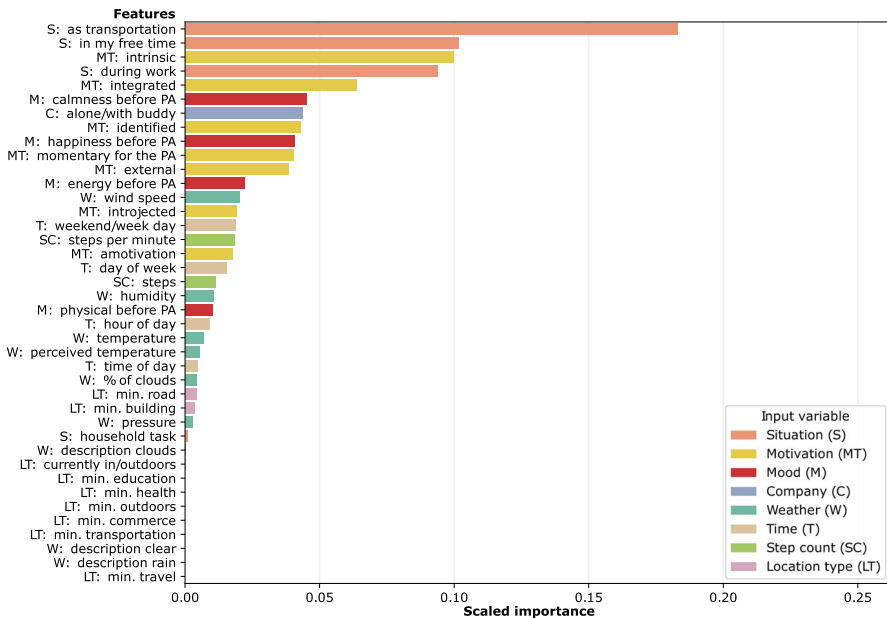


Fig. 3 The variable importances on the outcome variable PA duration show that the top ten most relevant features to predict the user’s PA duration include the situation, company, and some dimensions of mood and motivation

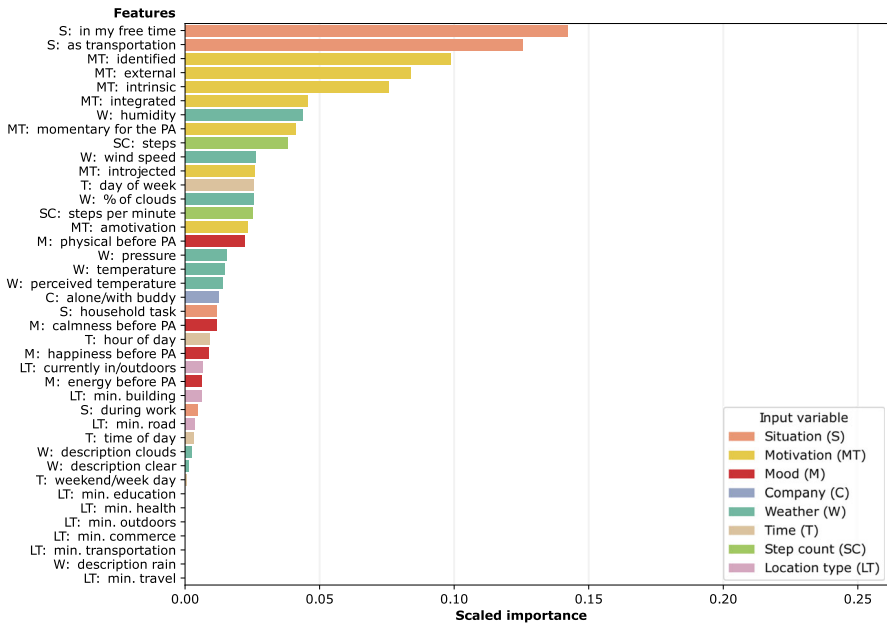


Fig. 4 The variable importances on the outcome variable PA intensity show that the most relevant feature to predict PA intensity is the situation, followed by some of the regulatory styles of the BREQ

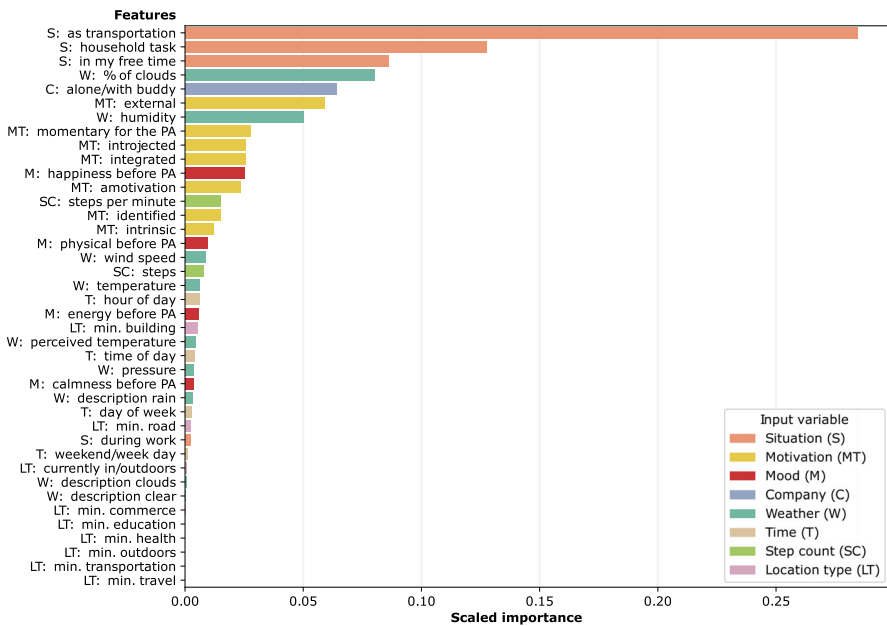


Fig. 5 The variable importances on the outcome variable PA location show that the situation determines the most whether the PA will be performed indoors or outdoors, followed by the percentage of cloud cover

than household tasks ( $p < .05$ , Cramer's  $V = .11$ ), but shorter than during work ( $p < .001$ , Cramer's  $V = .14$ ). Secondly, having a buddy resulted in a significantly longer duration ( $p < .05$ , Cramer's  $V = .10$ ) than doing PA alone. We also ran the GEE with the independent variable momentary motivation for the PA. For every point scored higher on momentary motivation, the duration was on average 6.7 min higher ( $p < .05$ , Cramer's  $V = .08$ ).

## 4.2 The choice of activity intensity

Figure 4 shows the features importances of the Random Forest (train accuracy = .68, test accuracy = .61) with the intensity category (light, moderate, or vigorous) as the outcome variable.

We analyzed the effect of situation as dependent variable using a GEE with the numerical MET intensity of all 724 submitted PAs. PA in free time significantly predicted a MET value higher than active transport ( $p < .05$ , Cramer's  $V = .12$ ), than during work ( $p < .001$ , Cramer's  $V = .13$ ), and than household tasks ( $p < .001$ , Cramer's  $V = .24$ ). Similar to the low relevance of company in Fig. 4, there was no significant effect found on MET intensity when having a buddy compared to exercising alone ( $p > .05$ , Cramer's  $V = .02$ ). Lastly, for every point scored higher on the momentary motivation for PA, the MET value was on average .35 higher ( $p < .001$ , Cramer's  $V = .16$ ).

## 4.3 The choice between indoor or outdoor physical activity

A Random Forest (train accuracy = .77, test accuracy = .70) was built with PA location as the outcome variable. The results of the sorted feature importances are shown in Fig. 5. Weather was requested at the time of submit, but only for smartphones with access to the GPS sensor, providing us with weather characteristics for only 660 PA records. Since we expected weather to have the biggest influence on PA location, we investigated the effect of these characteristics on the user's decision for indoor or outdoor PA with SPSS.

We created five binary logistic GEEs because location is a dichotomous variable (Austin and Merlo 2017), with each of the weather characteristics, and with indoors as the reference category. The results of these five models are shown in Table 4. Separately, clouds, wind, and humidity had a significant effect on the PA location ( $p < .05$ ), while perceived temperature and pressure did not ( $p > .05$ ). The effect sizes for clouds, wind, and humidity were small, showing a limited association on the decision for outdoor PA (Sun et al. 2010).

## 4.4 The choice of the type of physical activity

For all the 724 submitted activities, participants either selected an activity in the app, or typed one themselves. After categorization by PA name, 21 PA types were obtained, as elaborated in Appendix B. The type of PA was put as the categorical outcome variable in the Random Forest Classifier (train accuracy = .52, test accuracy = .48). Since a



**Table 4** The results of the binary logistic GEEs with different weather characteristics as independent variables show that cloud cover, wind speed, and humidity significantly predicted outdoor PA: higher odds for outdoor PA were expected with lower cloud cover, lower wind speeds, or less humidity

Parameter	$\beta$	CI		Wald $\chi^2$	df	p	exp( $\beta$ )	Cramer's V
		Lower	Upper					
(Intercept)	1.761	1.091	2.431	26.524	1	<.001	5.816	
Clouds	-.014	-.021	-.008	17.943	1	<.001	.986	.16
(Intercept)	1.142	.586	1.698	16.229	1	<.001	3.133	
Wind	-.111	-.221	.000	3.874	1	.049	.895	.08
(Intercept)	3.988	1.273	6.703	8.285	1	.004	53.946	
Humidity	-.039	-.072	-.006	5.304	1	.021	.962	.09
(Intercept)	.747	-.032	1.526	3.529	1	.060	2.110	
Perceived temperature	.003	-.088	.095	.005	1	.944	1.003	.00
(Intercept)	-8.626	-21.813	4.560	1.644	1	.200	<.001	
Pressure	.009	-.004	.022	1.930	1	.165	1.009	.05

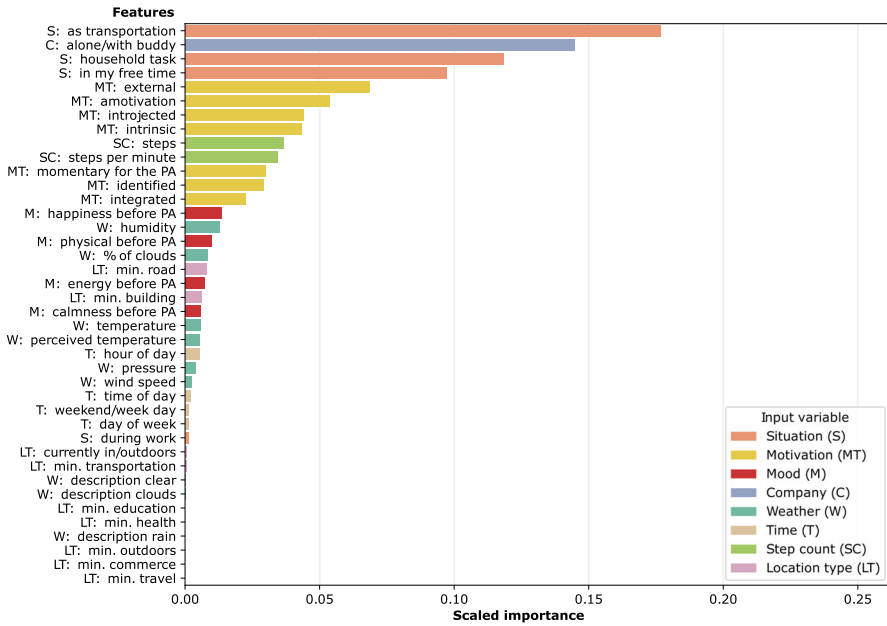


Fig. 6 The variable importances on outcome variable PA type show that the situation and the company are the most relevant predictors of the type of PA, followed by the regulatory styles of the BREQ

Random Forest Classifier cannot handle categories with less than two instances, we removed the two instances with “Football” and “Squash.” The results are depicted in Fig. 6.

4.5 Differences in mood across contexts

Every time participants submitted their information in the EMAs, filling in the mood dimensions was mandatory. In total, this resulted in 1349 mood records before and after PA to investigate the effect of PA on mood.

The results of the GEE with *before* as the reference category, are shown in Table 5. They indicate that physical complaints after performing a PA neither significantly improved, nor deteriorated ( $p > .05$ ). In contrast, there was a significant positive effect for happiness, calmness level, and energy level, meaning that these three mood dimensions improved on average after engaging in PA.

We conducted additional GEEs with situation and company as the independent variables to measure their effect for the 724 measurements of mood after PA. For situation, we selected free time as the reference category, which resulted in a significant effect with small effect size, compared to all three other situations, on happiness ( $p < .05$ , Cramer’s  $V = .13$ ), calmness ( $p < .05$ , Cramer’s  $V = .12$ ), and energy level ( $p < .001$ , Cramer’s  $V = .15$ ), but not on physical complaints ( $p > .05$ , Cramer’s  $V = .08$ ). This means that happiness, calmness, and energy level were higher after PA performed in free time than in the other situations. For company, having a buddy only had a

**Table 5** The result of the GEE shows that timing had a significant effect on happiness, calmness, and energy level. The GEE resulted in a significant increase in value of these three mood dimensions after the PA compared to before

Parameter	$\beta$	CI		Wald $\chi^2$	df	p	Cramer's V
		Lower	Upper				
<b>Happiness</b>							
(Intercept)	5.175	4.982	5.369	2739.597	1	<.001	
After the PA (reference = before)	.378	.263	.494	41.153	1	<.001	.17
<b>Calmness level</b>							
(Intercept)	5.152	4.873	5.430	1314.032	1	<.001	
After the PA (reference = before)	.299	.144	.454	14.343	1	<.001	.10
<b>Energy level</b>							
(Intercept)	4.823	4.593	5.052	1697.684	1	<.001	
After the PA (reference = before)	.287	.125	.449	12.032	1	<.001	.09
<b>Physical complaints</b>							
(Intercept)	5.461	5.174	5.748	1393.548	1	<.001	
After the PA (reference = before)	.087	-.007	.181	3.321	1	.068	.05

significant effect with small effect size on calmness level: having PA with a buddy resulted in .22 points higher on calmness level ( $p < .05$ , Cramer's  $V = .11$ ), which means that this improved people's calmness level after the PA.

#### 4.6 Differences in motivation across contexts

When participants submitted their PA, their momentary motivation for that PA was asked using a 5-point Likert scale, together with the situation and company that PA was performed in. In total, 724 PA records were analyzed. Firstly, Table 6 shows the results of the analysis of the categorical variable situation on the continuous variable motivation. A significant difference was found for all three situations compared to free time: on average, motivation in free time was .54 higher than active transport, .55 than during work, and .88 than household tasks. Secondly, the effect of company on motivation is presented in Table 7. This Table shows that motivation score was on average .35 higher when having a buddy compared to engaging in PA alone.

#### 4.7 The physical activity habits and app experience of the participants

The questions from the post-test questionnaires were answered by all 36 participants, and measured from 0 (disagree strongly) to 6 (agree strongly), using a 7-point Likert scale with neutral value 3. Their responses are visualized in Figs. 7 and 8. The histograms in Fig. 7 show that having good weather (94%, 34/36), spare time (94%, 34/36), a link to something useful (83%, 30/36), nature surroundings (75%, 27/36), an indoor workout space (75%, 27/36), and a workout buddy (75%, 27/36) have the most agreeing answers. Additionally, the amount of answers higher than the neutral value of 3 in the histograms in Fig. 8 show that 89% (32/36) found the app easy to use, 86% (31/36) thought it was user friendly, and 83% (30/36) were prepared to submit their information.

## 5 Discussion

### 5.1 Defining relevant contextual and motivational features for predicting useful physical activities

As seen on the right in Fig. 1, we focused on the task of HRSs to predict PA duration, intensity, location, and type to recommend useful PAs. To improve HRSs' prediction of these outcome variables, our first research question aimed to define relevant contextual and motivational features, to be integrated into context-aware techniques, such as pre-filtering (Ricci et al. 2022).

Firstly, for **PA duration**, our results confirmed that people engaged significantly longer in PA when they had higher motivation, performed the activities in free time, and had a buddy. Compared to free time PA, effect sizes were medium for PA as active transport and small for PA in the other situations, company, and motivation. The variable importances also confirm this relevance of situation on the duration. Moreover,

**Table 6** The GEE with situation as independent variable shows there was a significant difference in motivation for the different situations for PA, compared to PA in free time

Parameter	$\beta$	CI		Wald $\chi^2$	df	p	Cramer's V
		Lower	Upper				
(Intercept)	3.463	3.256	3.670	1074.465	1	<.001	
Situation (reference = Free time)							
Active transport	-.540	-.747	-.332	26.017	1	<.001	.19
During work	-.547	-.852	-.242	12.357	1	<.001	.13
Household task	-.876	-1.182	-.569	31.385	1	<.001	.21

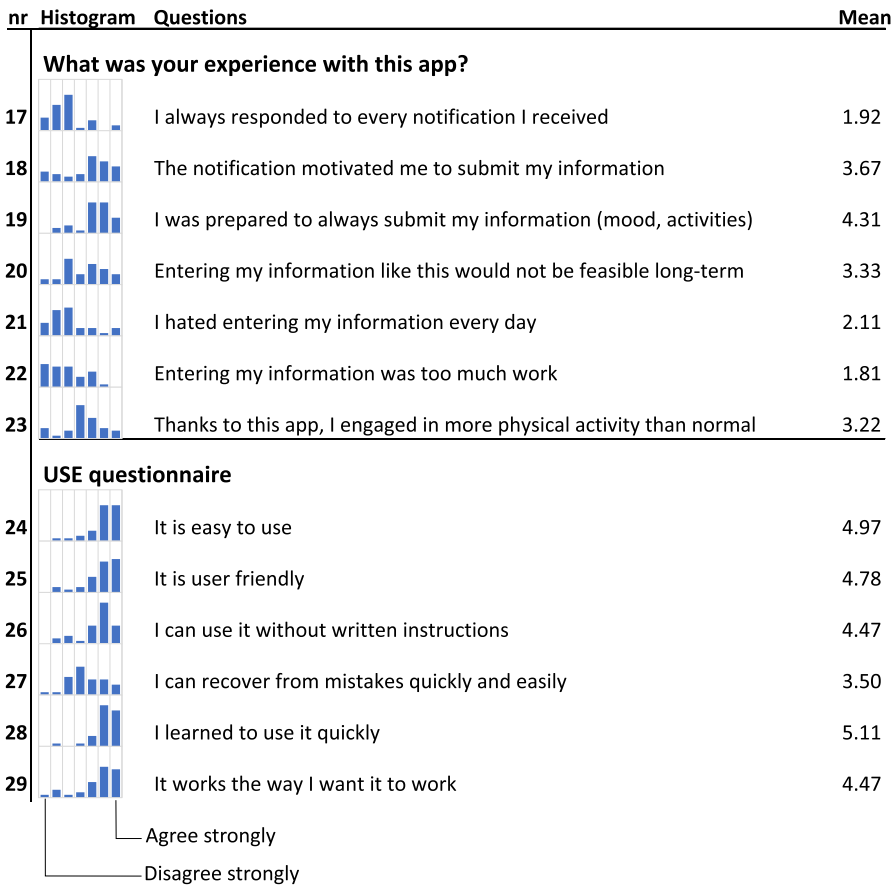
**Table 7** The GEE with company as independent variable shows that having a buddy for PA significantly predicted a higher motivation

Parameter	$\beta$	CI		Wald $\chi^2$	df	p	Cramer's V
		Lower	Upper				
(Intercept)	3.063	2.820	3.307	607.041	1	<.001	
With buddy (reference = alone)	.346	.163	.529	13.717	1	<.001	.14



Fig. 7 Part 1: histograms and means of our own questions about people’s motivation for PA, the influence on their mood, and whether they would like to engage more in it

the relevances show that mood before the PA also had an effect on the duration, especially for calmness and happiness. Additionally, the two most autonomous forms of motivation, intrinsic and integrated regulation, were both in the top 5 of the variable importances. This can be explained by the fact that higher autonomous motivation is linked to more PA (Nurmi et al. 2020; Ryan and Deci 2000). Other contextual features scored lower on feature importances, such as weather, which can be explained because not all PAs are executed outdoors.



**Fig. 8** Part 2: histograms and means of our own questions about the experience with the app, and of the USE questionnaire (Gao et al. 2018)

Therefore, we argue that HRSs should collect information about the mood, motivation, situation, and company using EMAs as input variables for the RS. As such, HRSs can adjust their duration recommendations to the user's mood at that time (Polignano et al. 2021). However, we suggest deriving the durations also from other information, such as the duration of previous PAs of that user and the WHO guidelines (World Health Organization 2020a).

Secondly, both effect size and relative feature importance of company on **PA intensity** were low, suggesting its association is less strong than on PA duration (Sun et al. 2010). Although our results show that intensity was mainly dependent on the situation, we assume this was because the situation mainly determined the type of PA, and therefore also the intensity. Additionally, people tended to have more vigorous PA when their motivation was higher, and physical complaints was more relevant than the other mood dimensions.



As such, we suggest that HRSs collect motivation and physical complaints data for suggesting PA intensity. Nonetheless, since intensity is inversely related to duration (Pelliccia et al. 2020), we suggest that HRSs also take into account the duration. Moreover, World Health Organization (2020a) recommends to gradually increase intensity and duration over time, and Pelliccia et al. (2020) recommend to adapt to the user's age, genetic, and fitness.

Thirdly, suggesting **PAs indoors or outdoors** was also mainly influenced by the situation, which makes sense since active transport is executed outdoors. The statistical analysis showed that cloud cover, wind speed, and humidity had a significant effect on PA location with small effect sizes, while perceived temperature and pressure did not. This suggests that higher values for the cloud percentage, wind speed, and humidity predicted lower odds for engaging in PA outdoors with a stronger association than the other weather characteristics (Sun et al. 2010).

Nonetheless, we suggest integrating all these weather characteristics in an HRS, supporting the need for more weather data on PA behavior (Turrisi et al. 2021; Pontin et al. 2021). Referring to Brocherie et al. (2015), temperature in particular should be taken into account for outdoor PA to avoid health risks. Additionally, despite the low relevance of time, we assume that time of day is also important to determine the location in case some people do not want outdoor PA when it is dark outside (Pontin et al. 2021).

Lastly, situation and company scored highest on the relative variable importances for **PA type**. This can be explained that some PAs can only be executed in certain situations (e.g., cleaning) or with a buddy (e.g., squash). Although this is an obvious relationship, our results confirmed that it is definitely useful to incorporate the situation and company in an HRS. Therefore, we believe this is of great importance for the RS because when having this information about the user, the RS can suggest more useful recommendations for suitable PAs. The regulation scores also scored highly, suggesting that there could be differences in chosen PA types across people's motivation.

## 5.2 Defining relevant contextual features for mood and motivation as appreciation feedback

For our second research question, we chose to analyze statistical differences in mood after PA and in motivation, as shown at the bottom of Fig. 1. Mood and motivation were chosen because we argue they are important as appreciation estimation of the performed PAs (Polignano et al. 2021), and can be used as feedback for the HRS to focus on user well-being.

The first significant differences in mood were analyzed between before and after the PA. The different mood dimensions happiness, energy level, and calmness level improved after engaging in PA, while this was not the case for physical complaints. Therefore, the increase in mood could serve as a form of feedback to the generated recommendations. For example, whether or not the recommended PA increased the user's energy level or caused more stress. Following the research of Polignano et al. (2021), we argue that integrating mood in the rating feedback of the RS is important

to measure how that item matches the mood of the user. Similarly, we argue that integrating the user's motivation to engage in the PA is equally important as feedback. In this way, instead of only maximizing the ratings of the PA consumptions, the RS focuses on maximizing the user's mood and motivation for PA, thereby increasing the user's well-being. As such, we argue that an RS should maximize on a combination of these variables. For example, the average of the rating and the motivation could serve as an aggregated feedback score. In addition, more advanced aggregation methods can be implemented. For example, the *average without misery* method, which is used for aggregating different users' ratings in group recommendations, can exclude low scores before calculating the average, providing a high average feedback score for that item (Ricci et al. 2022).

Next, significant differences were analyzed across situations. On average, people's motivation was higher when performing PA in their free time, which was expected following the SDT (Ryan and Deci 2000). Moreover, people's happiness, calmness, and energy level were significantly higher after free time PA than the other situations. Lastly, people's motivation and calmness level were significantly higher when PA was performed with a buddy. However, effect sizes were small, meaning that the effect of situation or company does not have a large association strength (Sun et al. 2010). Nonetheless, we argue that context-aware RSs should take these contextual factors into account in pre-filtering, such as *item-splitting* (Baltrunas and Ricci 2014), *micro-profiling* (Baltrunas and Amatriain 2009), or *UI splitting* (Zheng et al. 2013).

### 5.3 Limitations and future work

Across the results of the different outcome variables of RQ1, there were several contextual features that consistently score low on variable importance: time, minutes spent per location type, and step count before the PA. Although our statistical analyses for both RQ1 and RQ2 showed significant effects, their effect sizes were small according to Kim (2017), suggesting that association strength of the input on the outcome variables is limited (Sun et al. 2010).

These limitations can be caused by inaccuracies in automatically collected data on smartphones. These data are highly dependent on whether the users carry the smartphone with them, whether hardware for location and step count services is present, and whether permissions for these services are allowed. For example, the average detected daily step count of 2633 was lower than expected, compared to a study of Saint-Maurice et al. (2020), where the average daily step count was 9124. However, in the study by Saint-Maurice et al. (2020), the participants wore an accelerometer on the hip for, on average, 14.4 h per day. This was not the case in our study since step count was measured with the smartphone's accelerometer. A better alternative would be to collect data from wearables, which can be worn by the user the whole time, resulting in a more accurate step count (Althoff et al. 2017). However, we decided to create an app that is accessible, can reach many people, and therefore only required a smartphone. Moreover, wearables can suffer from incomplete data due to short battery life and users who do not always wear them (Gasparetti et al. 2020).

Furthermore, only 31 participants' smartphones delivered data from the GPS coordinates and 24 from step count, probably due to missing hardware or disallowed permissions. Moreover, most location types were detected as "building" and "road". We suspect that GPS coordinates were not accurate enough to distinguish roads from buildings, since these are often close to each other, and from more specific types, such as outdoors. Since the study was executed during only one month, the weather also did not change as much, resulting in limited conclusions of weather as contextual variable. Therefore, additional user studies are necessary in which a larger variety of weather conditions and location types are collected, as previously also proposed by Pontin et al. (2021).

Despite the low relative variable importances score of location type and step count history, we still consider their potential. In the study of Gasparetti et al. (2020), the amount of detected steps of the day represents the user's current state, which is used to recommend a range of steps in the near future. Similarly, the amount of minutes spent in every location type (e.g., 8 h in a building) and the steps taken (e.g., 1000) can define the user's state (e.g., very inactive day). Future HRSs can then predict PAs based on that state. Another option could be that the service detects a location type nearby that is ideal for a walk, such as a park, to suggest taking a walk there.

Another cause of the small effect sizes could be the limited amount of data points per PA category. For example, only 12 had PA for work, only 13 had PA of vigorous intensity, and there is only one data point for the PA type "Football". Inconsistencies could also be caused by people who overestimate their PA durations and intensities when they self-report PA, because of social desirability bias (Pontin et al. 2021) and their recall ability (Sylvia et al. 2014). In the EMAs of this study, however, participants were asked to enter their PA immediately afterwards, which could result in more accurately reported durations because they record PA real-time (Sylvia et al. 2014).

Furthermore, there could be other variables that we have not collected and analyzed, but that have an important effect on the outcome variables. Situated in the SDT, Deci and Ryan (2008) describe that motivation and behavior are dependent on the satisfaction of the psychological needs competence, autonomy, and relatedness, but also on people's long-term life goals, mindfulness, and vitality. Additionally, following psychological theories, Wang et al. (2021) explain that decisions to engage in activities are regulated by the user's current context (such as the time and weather, as we analyzed in this study), but also by the satisfaction of the user's goals and their subjective utility. Thus, future work could integrate additional variables relating to personal goals and whether or not the user recently already engaged in that activity, which could lower the utility value of that activity (Wang et al. 2021).

Lastly, we emphasize that effect sizes should also be compared between studies (Sun et al. 2010). As discussed in the meta-analysis of Mertens et al. (2022), effect sizes of behavioral interventions are generally small to medium. As such, the small effect sizes reported in this study can also be mainly due to this common occurrence, implying that future study designs might also not observe large effect sizes.

Another limitation is that not all motivational regulation scores of the BREQ had acceptable internal consistency scores. Since we only used two questions per construct to keep the app user friendly, we suggest incorporating all items of the BREQ in future studies. Additionally, the BREQ measures people's motivation for exercise, and

exercise is only a subset of PA, since PA is defined as any bodily movement that includes every kind of activity (World Health Organization 2020a). Therefore, we suggest to use a different SDT-based questionnaire that is focused on all types of PA to measure the six regulation styles. Furthermore, we included a calmness level measurement in our mood emoji that does not differentiate between the additional dimension of dominance to distinguish between irritated and nervous, as proposed by Mehrabian (1996). Future EMAs and HRSs could incorporate this additional dimension to assess mood more thoroughly.

We also asked the participants whether it was feasible to enter their manual features in the app. The post-test questionnaires showed that more than 83% participants were prepared to always submit their information in the app, and found the app user friendly. This means that the user burden was limited in our app designed for EMAs, and that it is feasible to collect user data. However, 50% answered that entering information in the app would not be feasible long-term. It could be that users experience a large burden of longer studies with these types of EMAs in apps. Moreover, 20 participants did not fill in the post-test questionnaire, which means they dropped out during the study.

Participants' burden with these types of apps could be reduced by omitting questions to be answered manually for variables that have limited effect on the outcome variables or rank low in the variable importances. However, the effect size and relevance of some manual variables differs across outcome variables, such as the company on PA intensity compared to on PA type. As such, future developers should carefully consider which variables to include. Moreover, user friendliness would increase when some manual contextual data would be automatically collected. For example, company could be detected automatically with Bluetooth or GPS to scan nearby users. Future apps could also automatically estimate situation, based on the current hour, day, and previously entered PAs (e.g., this user cycles as transportation every week day at 8 am) or the agenda (e.g., work activities planned on Monday afternoon). Because time can be collected and connected to the situation automatically, we argue that HRSs should integrate time, despite its low variable importances.

Finally, we emphasize avoiding unnecessary data collection, such as personal data that do not contribute to personalization. For example, the raw GPS coordinates, age, and sex of the users, are classified as highly sensitive to the General Data Protection Regulation (GDPR) legislation, and should only be collected when useful for the HRS. Therefore, we argue that it is not because data can be collected, that they should, and future work should always investigate the relevance of all collected data.

## 6 Conclusion

The goal of this study was to improve HRSs by defining which context and motivation variables can be usefully integrated into an HRS for PAs. The focus was on integrating these features in the recommendation and feedback processes of the HRS. As such, this research innovatively combined a large amount of variables, based on motivational theory and evidence-based PA guidelines. To collect these data, a longitudinal study with 36 participants was conducted with an Android app. We followed two approaches

to define the relevant features: using statistical analyses with GEE to detect significant differences in outcome variables across contexts, and using variable importance analyses with Random Forests to determine the relevance of 40 input variables.

Our results can be used as a guidance for developers of context-aware HRSs, by improving them in two tasks. Firstly, an HRS can generate useful PA recommendations for PA duration, intensity, location, and type using our proposed relevant features in contextual pre-filtering: the user's current company, situation, happiness, calmness, energy level, physical complaints, momentary motivation, and motivational BREQ score. Additionally, we suggested automatically collecting the following contextual data to further personalize PA predictions: weather and time of day to estimate whether outdoor PA should be suggested, location type and step count to estimate the user's PA state, and time of day and day of week to automatically estimate the user's current situation. Secondly, an HRS can collect the user's motivation for the PA and mood after the PA as appreciation feedback for the PA, which have shown significant differences across situation and company. In this way, we argue that integrating these variables in the HRS will provide a focus on user well-being, motivate users for behavior change, and thereby improve people's health.

**Author Contributions** Ine Coppens wrote the main manuscript text and prepared all the figures. All authors reviewed the manuscript.

**Data availability** The datasets generated and analyzed during the current study are not publicly available due to privacy restrictions because they contain sensitive personal data, but are available in restricted form from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** This study was funded by Ghent University.

**Ethical approval** The research was approved by the Ethical Committee of the Faculty of Psychology and Educational Sciences of Ghent University (<https://www.ugent.be/pp/en/research/ec>) on the 25th of October, 2021.

**Accordance** The methods of this study were carried out in accordance with the relevant guidelines and regulations, as discussed with the Ethical Committee and Data Protection Officers of Ghent University.

**Informed consent** All participants were older than 18 years and provided their informed consent before they were granted access to the app, and thereby the study.

## Appendix A Selected questions from the behavioural regulation in exercise questionnaire (BREQ)

To keep the app user friendly, a selection of questions from the BREQ was chosen. However, some questions of the same BREQ construct are almost exactly the same (e.g., "I exercise because others say I should" and "I exercise because other people say I should" in external regulation (Cid et al. 2018)). Therefore, we chose two questions per construct, based on avoiding similar questions, and with their internal consistency checked using Cronbach's alpha. This resulted in the following 12-item questionnaire:

- Amotivation: “*I don’t see why I should have to exercise*”, and “*I think that exercising is a waste of time*” (Markland and Tobin 2004)
- External regulation: “*I exercise because others say I should*”, and “*I feel under pressure from others to exercise*” (Wilson et al. 2006)
- Introjected regulation: “*I feel guilty when I don’t exercise*”, and “*I feel a failure when I haven’t exercised*” (Wilson et al. 2006)
- Identified regulation: “*I value the benefits of exercise*”, and “*It’s important to me to exercise regularly*” (Wilson et al. 2006)
- Integrated regulation: “*I consider exercise to be a part of my identity*”, and “*I consider exercise a fundamental part of who I am*” (Wilson et al. 2006)
- Intrinsic regulation: “*I exercise because it’s fun*”, and “*I enjoy my exercise sessions*” (Wilson et al. 2006)

## Appendix B Data preprocessing criteria

Participants were informed about the study criteria at the beginning of the study. Based on these criteria, the following 25 participants were excluded during data preprocessing of the collected data:

- The pre- and post-test questionnaires should be filled in: 20 participants failed to fill in the post-test questionnaire
- 21 different days of one or more submits should be sent between the 1st of November, 2021 and the end of the first week of December, 2021: 1 participant submitted until the 11th of December which was too late
- Of those 21 days, there should be no more than two consecutive days with a submit without any PA (because every kind of PA counts, such as walking in the supermarket): 4 participants submitted data with the option “no activities performed today” more than two consecutive days

This preprocessing resulted in 36 participants who met our criteria. For these participants, their submitted data points were also preprocessed following these steps (with the amount of affected data points in parentheses):

- Remove “no activities performed today” data point if this was submitted on the same day when a PA record of that same user was also submitted (49)
- Remove data point if submitted less than 30 s after previous submit of that same user (3)
- Outliers for step count: if the majority of the detected step counts of a participant does not have a value, we assume a malfunctioning accelerometer and mark all step counts of that participant as missing (3 participants)
- Outliers for PA duration were determined based on the 99.5 percentile, resulting in three data points to be removed: one with a duration of 15 h, and two with a duration of 23 h (3 points in total)
- Remove data points with manually typed PA types involving only sedentary activities, since these are not classified as PA by World Health Organization (2020a): “reading” (5), “Zoom conversation” (4), and “date” (1)

Removal of these data points resulted in a total of 1427 valid submits. Since people could type their activities themselves, spelling errors were adjusted and remaining PA types were manually re-categorized:

- “Walking” for manually entered “went shopping” (2), “groceries” (3), “shopping” (1), and “going to a market” (1)
- “Padel” for manually entered “padellen” (1), and “paddelen” (1)
- “Fitness” for manually entered “Exercise in gym” (2), “gym” (8), and “elliptical” (2)
- “Weight training” for manually entered “Lifting (fitness)” (10)
- “Dancing” for manually entered “dancing (latin and ballroom)” (7), and “dance workout” (1)
- “Workout” for manually entered “Exercising” (1)
- “Active socially” for manually entered “chiro geven” (3), and “scouts” (1) (both Dutch names of youth organizations)

The resulting 21 PA types with their corresponding total amount of instances were: Walking (314), Cycling (142), Dancing (35), Running (43), Swimming (2), Cleaning (58), Boxing (4), Power training (8), Football (1), Yoga (20), Weight training (15), Fitness (22), Workout (10), Working (4), Gardening (10), Padel (6), Horseback riding (2), Active socially (4), Stretching (20), Squash (1), and Bootcamp (3).

## Appendix C Random forest classifier hyperparameters

**Listing 1** Python code for computing optimal hyperparameters using RandomizedSearchCV

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import RandomizedSearchCV

n_estimators = [int(x) for x in np.linspace(start = 100,
                                          stop = 700,
                                          num = 30)]

max_features = ['auto', 'sqrt', 'log2']
max_depth = [1, 2, 3, 4]
min_samples_split = [2, 5, 10]
min_samples_leaf = [1, 2, 4, 10]
bootstrap = [True, False]

random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}

rfc = RandomForestClassifier()
rf_random = RandomizedSearchCV(estimator = rfc,
                              param_distributions = random_grid,
                              n_iter = 300,
                              cv = 5,
                              verbose = 2,
                              random_state = 42,
                              n_jobs = -1,
                              scoring = "recall")

rf_random.fit(X_train, y_train)
```

The resulting most optimal hyperparameters for all four Random Forest Classifiers were:

- n estimators: 617
- min samples split: 5
- min samples leaf: 10
- max features: ‘auto’

- max depth: 3
- bootstrap: False

## References

- Ainsworth, B., Haskell, W., Herrmann, S., et al.: 2011 compendium of physical activities: a second update of codes and met values. *Med. Sci. Sports Exerc.* **43**(8), 1575–1581 (2011). <https://doi.org/10.1249/MSS.0b013e31821ece12>
- Alcaraz-Herrera, H., Cartlidge, J., Toumpakari, Z., et al.: Evorecsys: evolutionary framework for health and well-being recommender systems. *User Model. User-Adap. Inter.* (2022). <https://doi.org/10.1007/s11257-021-09318-3>
- Althoff, T., Sosič, R., Hicks, J.L., et al.: Large-scale physical activity data reveal worldwide activity inequality. *Nature* **547**(7663), 336–339 (2017). <https://doi.org/10.1038/nature23018>
- Asselbergs, J., Ruwaard, J., Ejdys, M., et al.: Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. *J. Med. Internet Res.* **18**(3), e72 (2016). <https://doi.org/10.2196/jmir.5505>
- Austin, P.C., Merlo, J.: Intermediate and advanced topics in multilevel logistic regression analysis: multilevel logistic regression. *Stat. Med.* **36**(20), 3257–3277 (2017). <https://doi.org/10.1002/sim.7336>
- Baltrunas, L., Amatriain, X.: Towards time-dependant recommendation based on implicit feedback (2009)
- Baltrunas, L., Ricci, F.: Experimental evaluation of context-dependent collaborative filtering using item splitting. *User Model. User-Adap. Inter.* **24**(1–2), 7–34 (2014). <https://doi.org/10.1007/s11257-012-9137-9>
- Biddle, S.J., Ciacconi, S., Thomas, G., et al.: Physical activity and mental health in children and adolescents: an updated review of reviews and an analysis of causality. *Psychol. Sport Exerc.* **42**, 146–155 (2019). <https://doi.org/10.1016/j.psychsport.2018.08.011>
- Breiman, L.: Random Forests. *Mach. Learn.* **45**(1), 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
- Brocherie, F., Girard, O., Millet, G.: Emerging environmental and weather challenges in outdoor sports. *Climate* **3**, 492–521 (2015). <https://doi.org/10.3390/cli3030492>
- Caspersen, C., Powell, K., Christenson, G.: Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep.* **100**, 126–131 (1985)
- Cheung, K.L., Durusu, D., Sui, X., et al.: How recommender systems could support and enhance computer-tailored digital health programs: a scoping review. *Digit. Health* **5**, 1–19 (2019). <https://doi.org/10.1177/2055207618824727>
- Cid, L., Monteiro, D., Teixeira, D., et al.: The behavioral regulation in exercise questionnaire (breq-3) portuguese-version: Evidence of reliability, validity and invariance across gender. *Front Psychol.* (2018). <https://doi.org/10.3389/fpsyg.2018.01940>
- Costa, A., Heras, S., Palanca, J., et al.: Using argumentation schemes for a persuasive cognitive assistant system. In: Criado Pacheco, N., Carrascosa, C., Osman, N., et al. (eds.) *Multi-Agent Systems and Agreement Technologies*, vol. 10207, pp. 538–546. Springer International Publishing, Cham (2017). [https://doi.org/10.1007/978-3-319-59294-7\\_43](https://doi.org/10.1007/978-3-319-59294-7_43)
- De Croon, R., Van Houdt, L., Htun, N.N., et al.: Health recommender systems: systematic review. *J. Med. Internet Res.* (2021). <https://doi.org/10.2196/18035>
- Deci, E., Ryan, R.: Self-determination theory: a macrotheory of human motivation, development, and health. *Can. Psychol. Psychol. Can.* **49**, 182–185 (2008). <https://doi.org/10.1037/a0012801>
- Dharia, S., Eirinaki, M., Jain, V., et al.: Social recommendations for personalized fitness assistance. *Pers. Ubiquit. Comput.* **22**, 245–257 (2018). <https://doi.org/10.1007/s00779-017-1039-8>
- Drieskens, S., Gisle, L., Charafeddine, R., et al.: Gezondheidsenquête 2018: Levensstijl. Samenvatting van de resultaten. *Tech. rep.*, Sciensano (2018)
- El Haouij, N., Poggi, J.M., Ghozi, R., et al.: Random forest-based approach for physiological functional variable selection for driver’s stress level classification. *Stat. Methods Appl.* **28**(1), 157–185 (2019). <https://doi.org/10.1007/s10260-018-0423-5>
- Feltz, D., Kerr, N., Irwin, B.: Buddy up: The kohler effect applied to health games. *J. Sport Exer. Psychol.* **33**, 506–526 (2011). <https://doi.org/10.1123/jsep.33.4.506>
- Fukuoka, Y., Lindgren, T.G., Mintz, Y.D., et al.: Applying natural language processing to understand motivational profiles for maintaining physical activity after a mobile app and accelerometer-based



- intervention: the mPED randomized controlled trial. *JMIR Mhealth Uhealth* **6**(6), e10,042 (2018). <https://doi.org/10.2196/10042>
- Gao, M., Kortum, P., Oswald, F.: Psychometric evaluation of the use (usefulness, satisfaction, and ease of use) questionnaire for reliability and validity. *Proc. Human Factors Ergon. Soc. Ann. Meet.* **62**, 1414–1418 (2018). <https://doi.org/10.1177/1541931218621322>
- Gasparetti, F., Aiello, L., Quercia, D.: Personalized weight loss strategies by mining activity tracker data. *User Model. User-Adap. Inter.* **30**, 447–476 (2020). <https://doi.org/10.1007/s11257-019-09242-7>
- Gerovasilis, V., Agaku, I.T., Vardavas, C.I., et al.: Levels of physical activity among adults 18–64 years old in 28 European countries. *Prev. Med.* **81**, 87–91 (2015). <https://doi.org/10.1016/j.ypmed.2015.08.005>
- Heck, R.H., Thomas, S.L., Tabata, L.N.: *Multilevel and Longitudinal Modeling with IBM SPSS*, 2nd edn. Routledge, New York (2013). <https://doi.org/10.4324/9780203701249>
- Hors-Fraile, S., Rivera, O., Schneider, F., et al.: Analyzing recommender systems for health promotion using a multidisciplinary taxonomy: a scoping review. *Int. J. Med. Inform.* **114**, 143–155 (2018). <https://doi.org/10.1016/j.ijmedinf.2017.12.018>
- Hussein, T., Linder, T., Gaulke, W., et al.: Hybreed: a software framework for developing context-aware hybrid recommender systems. *User Model. User-Adap. Inter.* **24**, 121–174 (2014). <https://doi.org/10.1007/s11257-012-9134-z>
- Kamphorst, B., Klein, M., Wissen, A.: Autonomous E-coaching in the wild: Empirical validation of a model-based reasoning system. pp. 725–732 (2014)
- Kim, H.Y.: Statistical notes for clinical researchers: Chi-squared test and Fisher’s exact test. *Restor. Dent. Endod.* **42**(2), 152 (2017). <https://doi.org/10.5395/rde.2017.42.2.152>
- Liao, Y., Skelton, K., Dunton, G., et al.: A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: An adapted strobe checklist for reporting ema studies (cremas). *J. Med. Internet Res.* (2016). <https://doi.org/10.2196/jmir.4954>
- Markland, D., Tobin, V.: A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. *J. Sport Exer. Psychol.* **26**, 191–196 (2004). <https://doi.org/10.1123/jsep.26.2.191>
- Mehrabian, A.: Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in Temperament. *Curr. Psychol.* **14**(4), 261–292 (1996). <https://doi.org/10.1007/BF02686918>
- Mertens, S., Herberz, M., Hahnel, U.J.J., et al.: The effectiveness of nudging: a meta-analysis of choice architecture interventions across behavioral domains. *Proc. Natl. Acad. Sci.* **119**(1), e2107346,118 (2022). <https://doi.org/10.1073/pnas.2107346118>
- Miyamoto, S., Henderson, S., Young, H., et al.: Tracking health data is not enough: A qualitative exploration of the role of healthcare partnerships and mhealth technology to promote physical activity and to sustain behavior change. *JMIR Mhealth Uhealth* **4**, e5 (2016). <https://doi.org/10.2196/mhealth.4814>
- Mullan, E., Markland, D.A., Ingledew, D.K.: A graded conceptualisation of self-determination in the regulation of exercise behaviour: Development of a measure using confirmatory factor analytic procedures. *Pers. Individ. Differ.* **23**, 745–752 (1997). [https://doi.org/10.1016/S0191-8869\(97\)00107-4](https://doi.org/10.1016/S0191-8869(97)00107-4)
- Nahum-Shani, I., Smith, S.N., Spring, B.J., et al.: Just-in-time adaptive interventions (JITAs) in mobile health: key components and design principles for ongoing health behavior support. *Ann. Behav. Med.* **52**(6), 446–462 (2018). <https://doi.org/10.1007/s12160-016-9830-8>
- Ndulue, C., Oyebode, O., Iyer, R.S., et al.: Personality-targeted persuasive gamified systems: exploring the impact of application domain on the effectiveness of behaviour change strategies. *User Model. User-Adap. Inter.* **32**(1–2), 165–214 (2022). <https://doi.org/10.1007/s11257-022-09319-w>
- Norman, G.: Likert scales, levels of measurement and the “laws” of statistics. *Adv. Health Sci. Educ. Theory Pract.* **15**, 625–32 (2010). <https://doi.org/10.1007/s10459-010-9222-y>
- Nurmi, J., Knittle, K., Ginchev, T., et al.: Engaging users in the behavior change process with digitalized motivational interviewing and gamification: Development and feasibility testing of the precious app. *JMIR Mhealth Uhealth* **8**(1), e12,884 (2020). <https://doi.org/10.2196/12884>
- Odić, A., Tkalčić, M., Tasić, J.F., et al.: Predicting and detecting the relevant contextual information in a movie-recommender system. *Interact. Comput.* **25**(1), 74–90 (2013). <https://doi.org/10.1093/iwc/iws003>
- Op den Akker, H., Jones, V., Hermens, H.: Tailoring real-time physical activity coaching systems: a literature survey and model. *User Model. User-Adap. Inter.* **24**(5), 351–392 (2014). <https://doi.org/10.1007/s11257-014-9146-y>

- Pekár, S., Brabec, M.: Generalized estimating equations: a pragmatic and flexible approach to the marginal GLM modelling of correlated data in the behavioural sciences. *Ethology* **124**(2), 86–93 (2018). <https://doi.org/10.1111/eth.12713>
- Pelliccia, A., Sharma, S., Gati, S., et al.: 2020 ESC Guidelines on sports cardiology and exercise in patients with cardiovascular disease: The Task Force on sports cardiology and exercise in patients with cardiovascular disease of the European Society of Cardiology (ESC). *Eur. Heart J.* **42**(1), 17–96 (2020). <https://doi.org/10.1093/eurheartj/ehaa605>
- Penedo, F., Dahn, J.: Exercise and well-being: a review of mental and physical health benefits associated with physical activity. *Curr. Opin. Psychiatry* **18**, 189–193 (2005). <https://doi.org/10.1097/00001504-200503000-00013>
- Phan, W.M.J., Amrhein, R., Rounds, J., et al.: Contextualizing interest scales with emojis: implications for measurement and validity. *J. Career Assess.* **27**(1), 114–133 (2019). <https://doi.org/10.1177/1069072717748647>
- Polignano, M., Narducci, F., de Gemmis, M., et al.: Towards emotion-aware recommender systems: an affective coherence model based on emotion-driven behaviors. *Expert Syst. Appl.* **170**(114), 382 (2021). <https://doi.org/10.1016/j.eswa.2020.114382>
- Pontin, F., Lomax, N., Clarke, G., et al.: Socio-demographic determinants of physical activity and app usage from smartphone data. *Soc. Sci. Med.* **284**(114), 235 (2021). <https://doi.org/10.1016/j.socscimed.2021.114235>
- Rabbi, M., Pfammatter, A., Zhang, M., et al.: Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. *JMIR Mhealth Uhealth* **3**, e42 (2015). <https://doi.org/10.2196/mhealth.4160>
- Ricci, F., Rokach, L., Shapira, B. (eds.): *Recommender Systems Handbook*. Springer, US, New York, NY, (2022). <https://doi.org/10.1007/978-1-0716-2197-4>
- Russell, J.A.: A circumplex model of affect. *J. Pers. Soc. Psychol.* **39**(6), 1161–1178 (1980). <https://doi.org/10.1037/h0077714>
- Ryan, R., Deci, E.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* **55**, 68–78 (2000). <https://doi.org/10.1037/0003-066X.55.1.68>
- Saint-Maurice, P.F., Troiano, R.P., Bassett, D.R., et al.: Association of daily step count and step intensity with mortality among US adults. *JAMA* **323**(12), 1151 (2020). <https://doi.org/10.1001/jama.2020.1382>
- Smyth, B., Lawlor, A., Berndsen, J., et al.: Recommendations for marathon runners: on the application of recommender systems and machine learning to support recreational marathon runners. *User Model. User-Adap. Inter.* (2021). <https://doi.org/10.1007/s11257-021-09299-3>
- Stamatakis, E., Gale, J., Bauman, A., et al.: Sitting time, physical activity, and risk of mortality in adults. *J. Am. Coll. Cardiol.* **73**, 2062–2072 (2019). <https://doi.org/10.1016/j.jacc.2019.02.031>
- Sun, S., Pan, W., Wang, L.L.: A comprehensive review of effect size reporting and interpreting practices in academic journals in education and psychology. *J. Educ. Psychol.* **102**(4), 989–1004 (2010). <https://doi.org/10.1037/a0019507>
- Sylvia, L.G., Bernstein, E.E., Hubbard, J.L., et al.: Practical guide to measuring physical activity. *J. Acad. Nutr. Diet.* **114**(2), 199–208 (2014). <https://doi.org/10.1016/j.jand.2013.09.018>
- Þórarinsdóttir, H., Faurholt-Jepsen, M., Ullum, H., et al.: The validity of daily self-assessed perceived stress measured using smartphones in healthy individuals: cohort study. *JMIR mHealth uHealth* **7**(8), e13,418 (2019). <https://doi.org/10.2196/13418>
- Turrisi, T.B., Bittel, K.M., West, A.B., et al.: Seasons, weather, and device-measured movement behaviors: a scoping review from 2006 to 2020. *Int. J. Behav. Nutr. Phys. Act.* **18**(1), 24 (2021). <https://doi.org/10.1186/s12966-021-01091-1>
- Verikas, A., Gelzinis, A., Bacauskiene, M.: Mining data with random forests: a survey and results of new tests. *Pattern Recogn.* **44**(2), 330–349 (2011). <https://doi.org/10.1016/j.patcog.2010.08.011>
- Wagner, A.L., Keusch, F., Yan, T., et al.: The impact of weather on summer and winter exercise behaviors. *J. Sport Health Sci.* **8**(1), 39–45 (2019). <https://doi.org/10.1016/j.jshs.2016.07.007>
- Wang, S., Zhang, C., Kröse, B., et al.: Optimizing adaptive notifications in mobile health interventions systems: reinforcement learning from a data-driven behavioral simulator. *J. Med. Syst.* **45**(12), 102 (2021). <https://doi.org/10.1007/s10916-021-01773-0>
- Wanner, M., Goetschi, T., Martin-Diener, E., et al.: Active transport, physical activity, and body weight in adults a systematic review. *Am. J. Prev. Med.* **42**(5), 493–502 (2012). <https://doi.org/10.1016/j.amepre.2012.01.030>

- Wilson, P., Rodgers, W., Loitz, C., et al.: “it’s who i am... really!” the importance of integrated regulation in exercise contexts I. *J. Appl. Biobehav. Res.* **11**, 79–104 (2006). <https://doi.org/10.1111/j.1751-9861.2006.tb00021.x>
- World Health Organization: WHO guidelines on physical activity and sedentary behaviour. World Health Organization (2020a)
- World Health Organization: World health statistics 2020: monitoring health for the SDGs, sustainable development goals. World Health Organization (2020b)
- Young, D.R., Hivert, M.F., Alhassan, S., et al.: Sedentary behavior and cardiovascular morbidity and mortality: a science advisory from the American heart association. *Circulation* **134**, e262–e279 (2016). <https://doi.org/10.1161/CIR.0000000000000440>
- Zheng, Y., Burke, R., Mobasher, B.: The role of emotions in context-aware recommendation. *Decis. RecSys.* **45**, 45 (2013). <https://doi.org/10.13140/2.1.2660.1769>

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Ine Coppens** is a PhD student at Ghent University. She received the MSc degree in Information Engineering Technology at Ghent University in July 2018 for which she conducted her master thesis research on collecting implicit feedback for video recommendations using emotion recognition. In July 2020, she obtained the MSc degree of Complementary Studies in Business Economics of the same university. This master’s thesis was focused on increasing motivation with sustainability tips. In 2020, she started her PhD research on applying recommender system algorithms in mobile applications to improve people’s health and motivate them to have more physical activity.

**Toon De Pessemier** received the MSc degree in computer science engineering at Ghent University in 2006. In 2013, he obtained a PhD on his research “improved online services by personalized recommendations and optimal quality of experience parameters”. In 2015, he obtained a post-doc scholarship for his research “Design and validation of innovative techniques to assist people in the decision making process by analyzing data of user behavior and contextual features”. From 2018, he is part-time professor and teaches Recommender Systems. Another research topic that he investigates is scheduling of production environments.

**Luc Martens** received the MSc degree in electrical engineering from the Ghent University, Belgium in July 1986. From September 1986 to December 1990 he was a research assistant at the Department of Information Technology (INTEC) of the same university. During this period, his scientific work was focused on the physical aspects of hyperthermic cancer therapy. His research work dealt with electromagnetic and thermal modelling and with the development of measurement systems for that application. This work led to the PhD degree in December 1990. Since 1991, he manages the wireless & cable research group at INTEC. This group is since 2004 part of the Minds institute and since April 1993 he is Professor at Ghent University. His experience and current interests are in modelling and measurement of electromagnetic channels, of electromagnetic exposure e.g. around telecommunication networks and systems such as cellular base station antennas, and of energy consumption in wireless networks. He is author or co-author of more than 300 publications in the domain of electromagnetic channel predictions, dosimetry, exposure systems and health and wireless communications.