DiversityOne: A Multi-Country Smartphone Sensor Dataset for Everyday Life Behavior Modeling

MATTEO BUSSO*, ANDREA BONTEMPELLI, and LEONARDO JAVIER MALCOTTI, University of Trento, Italy
LAKMAL MEEGAHAPOLA, ETH Zurich, Switzerland
PETER KUN, IT University of Copenhagen, Denmark
SHYAM DIWAKAR and CHAITANYA NUTAKKI, Amrita Vishwa Vidyapeetham, India
MARCELO DARIO RODAS BRITEZ, University of Trento & FBK, Italy
HAO XU and DONGLEI SONG, Jilin University, China
SALVADOR RUIZ CORREA and ANDREA-REBECA MENDOZA-LARA, Instituto Potosino de Investigación Científica y Tecnológica, Mexico
GEORGE GASKELL, SALLY STARES, and MIRIAM BIDOGLIA, London School of Economics and Political Science, UK
AMARSANAA GANBOLD and ALTANGEREL CHAGNAA, National University of Mongolia, Mongolia
LUCA CERNUZZI and ALETHIA HUME, Universidad Católica "Nuestra Señora de la Asunción", Paraguay
RONALD CHENU-ABENTE, ROY ALIA ASIKU, and IVAN KAYONGO, University of Trento, Italy
DANIEL GATICA-PEREZ, Idiap Research Institute & EPFL, Switzerland
AMALIA DE GÖTZEN, Aalborg University, Denmark
IVANO BISON, University of Trento, Italy
FAUSTO GIUNCHIGLIA, University of Trento, Italy

Understanding everyday life behavior of young adults through personal devices, e.g., smartphones and smartwatches, is key for various applications, from enhancing the user experience in mobile apps to enabling appropriate interventions in digital health apps. Towards this goal, previous studies have relied on datasets combining passive sensor data with human-provided annotations or self-reports. However, many existing datasets are limited in scope, often focusing on specific countries primarily in the Global North, involving a small number of participants, or using a limited range of pre-processed sensors. These limitations restrict the ability to capture cross-country variations of human behavior, including the possibility of studying

*Corresponding author

Authors' Contact Information: Matteo Busso, matteo.busso@unitn.it; Andrea Bontempelli; Leonardo Javier Malcotti, University of Trento, Italy; Lakmal Meegahapola, ETH Zurich, Switzerland; Peter Kun, IT University of Copenhagen, Denmark; Shyam Diwakar; Chaitanya Nutakki, Amrita Vishwa Vidyapeetham, India; Marcelo Dario Rodas Britez, University of Trento & FBK, Italy; Hao Xu; Donglei Song, Jilin University, China; Salvador Ruiz Correa; Andrea-Rebeca Mendoza-Lara, Instituto Potosino de Investigación Científica y Tecnológica, Mexico; George Gaskell; Sally Stares; Miriam Bidoglia, London School of Economics and Political Science, UK; Amarsanaa Ganbold; Altangerel Chagnaa, National University of Mongolia, Mongolia; Luca Cernuzzi; Alethia Hume, Universidad Católica "Nuestra Señora de la Asunción", Paraguay; Ronald Chenu-Abente; Roy Alia Asiku; Ivan Kayongo, University of Trento, Italy; Daniel Gatica-Perez, Idiap Research Institute & EPFL, Switzerland; Amalia de Götzen, Aalborg University, Denmark; Ivano Bison, University of Trento, Italy; Fausto Giunchiglia, University of Trento, Italy, fausto.giunchiglia@unitn.it.

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.
© 2025 Copyright held by the owner/author(s).
ACM 2474-9567/2025/3-ART1
https://doi.org/10.1145/3712289

model generalization, and robustness. To address this gap, we introduce *DiversityOne*, a dataset which spans eight countries (China, Denmark, India, Italy, Mexico, Mongolia, Paraguay, and the United Kingdom) and includes data from 782 college students over four weeks. DiversityOne contains data from 26 smartphone sensor modalities and 350K+ self-reports. As of today, it is one of the largest and most diverse publicly available datasets, while featuring extensive demographic and psychosocial survey data. DiversityOne opens the possibility of studying important research problems in ubiquitous computing, particularly in domain adaptation and generalization across countries, all research areas so far largely underexplored because of the lack of adequate datasets.

Important Links

Dataset Catalog: https://livepeople.disi.unitn.it
***DiversityOne* Website:** https://datascientia.disi.unitn.it/projects/diversityone/

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**; *Empirical studies in ubiquitous and mobile computing*; *Smartphones*; *Mobile computing*.

Additional Key Words and Phrases: datasets, diversity, social practices, mobile sensing, smartphone sensing, wellbeing, health, generalization

ACM Reference Format:
Matteo Busso, Andrea Bontempelli, Leonardo Javier Malcotti, Lakmal Meegahapola, Peter Kun, Shyam Diwakar, Chaitanya Nutakki, Marcelo Dario Rodas Britez, Hao Xu, Donglei Song, Salvador Ruiz Correa, Andrea-Rebeca Mendoza-Lara, George Gaskell, Sally Stares, Miriam Bidoglia, Amarsanaa Ganbold, Altangerel Chagnaa, Luca Cernuzzi, Alethia Hume, Ronald Chenu-Abente, Roy Alia Asiku, Ivan Kayongo, Daniel Gatica-Perez, Amalia de Götzen, Ivano Bison, and Fausto Giunchiglia. 2025. *DiversityOne*: A Multi-Country Smartphone Sensor Dataset for Everyday Life Behavior Modeling. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 1, Article 1 (March 2025), 49 pages. https://doi.org/10.1145/3712289

1 Introduction

Human behavior, routines, habits, and social practices are deeply interwoven with the events, places, interactions, and technologies that compose our everyday lives. Each of these elements contributes to larger behavioral patterns influenced by our continuous engagement with pervasive devices. Our emotional experiences fluctuate daily, shaped by factors such as sleep quality [64], which, in turn, impact our physical activity, forming both recurring weekly patterns [119] and longer-term habits [53]. Everyday actions, such as eating and drinking, also have significant effects on our health [6, 98], with additional factors like interactions with technology playing important roles in everyday life [26].

Smartphones, as prevalent devices in daily life, influence our behavior, often in nuanced ways that can be both positive [74] and negative [46]. Their use is linked to our internal states [14, 32], and understanding how people engage with their smartphones could provide valuable insights into human behavior. Prior research has shown that smartphone sensors can infer a range of behavioral aspects in young adults, including mood [80, 105], depression [131], activities [3, 54], social context [57, 77], and eating or drinking episodes [4, 6, 98, 114]. This makes smartphone sensor data valuable for improving app design, supporting user well-being, and creating more effective interventions. However, there are profound variations in daily behaviors across different cultural and geographic backgrounds [34][1]. Ignoring these variations has often hindered model generalization [65, 80], limiting the effectiveness of behavior inference models when applied in diverse real-world settings.

[1]We operationalize the concept of culture through the theory of social practices outlined in Section 2.4.1. This concept recognizes that cultures are complex and multidimensional, often coexisting or overlapping within the same geographic areas [134]. Additionally, it acknowledges that certain regularities emerge from specific socio-cultural contexts or from the communities of practice that develop locally. The use of terms like "country," "culture," and "geographic region" throughout the paper reflects this distinction.

Despite the richness of smartphone sensor data, a critical challenge persists: the lack of datasets that capture behavioral diversity across different cultural and geographic contexts. Existing datasets and research often draw from specific populations typically in the Global North [65, 80], and may not fully account for the cultural, environmental, and social norms that shape daily routines and smartphone usage patterns in other regions [90]. This limits the development of machine learning models capable of generalizing across different populations. For instance, eating habits [116], sleep routines [23, 112], and social interactions [50, 89] vary widely between countries, affecting how individuals interact with their smartphones and how other daily life aspects are inferred from sensor data [76, 90]. Consequently, models trained on data from one country or region may perform poorly when applied to another. This underscores a fundamental challenge in generalizing machine learning models to real-world applications [3, 80]. Addressing this challenge requires comprehensive datasets that not only capture a variety of sensor data but also integrate cultural, social, and geographical diversity. Such datasets would enable researchers to study behavioral differences and similarities across populations, to test model robustness in diverse contexts beyond typical experimental settings, and to investigate how to combine model generalization and adaptation to local data to capture local and individual behaviors.

The *DiversityOne* dataset, developed as part of the large-scale European project *"WeNet - The Internet of US"*[2] [85], provides rich, cross-country data for studying everyday life behavior through smartphone sensors and self-reports. *DiversityOne* spans data from college students in eight countries: China (Jilin University, **JLU**), Denmark (Aalborg University, **AAU**), India (Amrita Vishwa Vidyapeetham, **AMRITA**), Italy (University of Trento, **UNITN**), Mexico (Instituto Potosino de Investigación Científica y Tecnológica, **IPICYT**), Mongolia (National University of Mongolia, **NUM**), Paraguay (Universidad Católica "Nuestra Señora de la Asunción", **UC**), and the United Kingdom (London School of Economics and Political Science, **LSE**). The study involved more than 18,000 students, of which 782 agreed to participate in an intensive longitudinal survey of four weeks. Based on the innovative `iLog app` [136], adopted for the data collection, *DiversityOne* includes raw data from 26 smartphone sensors, such as accelerometers, gyroscopes, and GPS, as well as derived information like notification interactions, app usage, activities, and step counts, during these four weeks. These data streams are organized into six category bundles (connectivity, environment, motion, position, app usage, and device usage), allowing modular downloads for ease of use. Additionally, participants provided multiple in-situ self-reports each day, through time diaries, detailing their activities, locations, social contexts, and moods, together with daily reports on sleep quality and daily expectations. This combination of self-reported data and sensor data provides a comprehensive view of behavior, making *DiversityOne* a valuable resource for studying the impact of smartphone technology across various contexts in both the Global North and the Global South [6, 49, 64, 92]. Additionally, DiversityOne enables more computational social science-oriented studies focused on human behavior (see, e.g., [9, 46, 140]).

The *DiversityOne* dataset supports the development of machine learning models capable of inferring user behaviors, with potential applications in domain generalization, domain adaptation, transfer learning, and self-supervised learning. It is a valuable research resource in ubiquitous computing, human-computer interaction, and machine learning, offering a broader scale than previously available datasets. Furthermore, *DiversityOne* adheres to the European General Data Protection Regulation (GDPR) [1], with ethical approvals from all participating institutions. The study design and data collection processes were developed by an interdisciplinary team of computer scientists, social scientists, interaction designers, ethicists, and legal experts to meet the highest standards. This extensive dataset has been used in a few initial publications by different authors, showing its various uses [3, 40, 63, 77, 80, 82, 83], and demonstrating its potential for reuse in studies aiming to understand behaviors across countries through smartphone sensing. The paper describes three contributions:

Contribution 1: We publicly release *DiversityOne*, one of the largest and most geographically diverse datasets combining questionnaires from 18K+ participants and passive smartphone sensor data and self-reports from 782

[2] https://doi.org/10.3030/823783.

participants across eight countries, covering both the Global North and the Global South. The dataset includes raw sensor data from 26 modalities, providing researchers with a flexible resource for analyzing and developing machine learning models across multiple domains, such as behavior recognition, cross-cultural studies, and multimodal time-series modeling. This dataset facilitates explorations of topics like model generalization and domain adaptation across countries, with detailed self-reports on demographic and psychosocial variables, adding depth for nuanced, multifaceted analyses that can be conducted independently or integrated with sensor data.

Contribution 2: We provide an in-depth description of the study design and data collection, which was tailored to enable effective data gathering across multiple countries, each with unique languages, privacy regulations, and cultural contexts (Section 3). To manage the diversity and sensitivity of the collected data, the dataset was consolidated under rigorous privacy and ethical standards, hosted in a GDPR-compliant secure environment, and respecting the ownership rights of all contributors.

Contribution 3: We synthesize key insights, lessons learned, and recommendations derived from some initial studies leveraging *DiversityOne*, highlighting unique opportunities for further studies in smartphone sensing, behavioral modeling, and machine learning. We provide the insights under two broader themes—Lessons learned from (i) study design and cross-country data collection (Section 6.1) and (ii) multi-country sensor data analysis (Section 6.2). Our 12 recommendations, which we see as actionable, aim to provide ideas for future research, enhancing the adaptability and accuracy of smartphone-based behavioral studies in diverse global contexts.

The remainder of the paper is organized as follows: Section 2 provides background and context for the methodology used to create *DiversityOne*, reviewing existing datasets for modeling everyday behaviors through smartphone sensors alongside various forms of ground truth. Section 3 details the methodology and protocol adopted to collect diverse, cross-country data, with specific attention to variations in cultural contexts of student's daily routines and data privacy regulations. Section 4 presents key results to validate the data collection approach and the dataset's quality. Section 5 outlines the resources available for accessing and using the dataset, highlighting compliance with privacy and copyright standards. Section 6 explores potential use cases for the dataset, provides insights and recommendations, emphasizing its value in studying behavioral patterns across different regions and advancing machine learning applications. Lastly, Section 7 summarizes the paper's contributions and impacts.

2 Background and Related Work

2.1 Smartphone Sensing for Behavior Modeling

Smartphone sensing data is widely used to train machine learning models to infer user behaviors in real-world environments [81]. Developing these models involves continuously collecting data from sensors like accelerometers, gyroscopes, GPS, Bluetooth, and app usage logs to create labeled datasets. These datasets serve as ground truth for various behaviors, including daily activities, social interactions, mental states, sleep, and dietary habits [81]. These labeled datasets fuel machine learning models, from classic approaches (e.g., random forest, decision tree, support vector machines) to deep learning techniques, enabling the detection of behavioral patterns and predictions on new, unseen data. Once trained, these models are deployed in-the-wild, running on users' smartphones to provide real-time insights into daily routines, mental well-being, and contextual changes, all without requiring active input from the user. Traditionally, such models have relied heavily on continuous sensing modalities—such as movement patterns and physical proximity—to infer user behavior. More recently, however, interaction sensing modalities (e.g., app usage, typing events, and notification interactions) have gained significance, providing rich insights into users' engagement patterns and internal states.

2.2 Continuous and Interaction Sensing Modalities

Smartphone sensors can be broadly classified into two types [81]: continuous sensing and interaction sensing, both essential for understanding nuances of everyday life.

Continuous sensing involves passive data collection without direct user interaction. For instance, accelerometers and gyroscopes detect movement patterns (e.g., walking, running, or inactivity) that offer insights into physical activity levels and types, relevant for understanding behavioral patterns, energy expenditure, and emotional states, such as depression or anxiety indicated by prolonged inactivity or erratic movements [16, 32]. Step count, derived from inertial sensors, further enhances activity monitoring, revealing daily routines and physical well-being. Additionally, proximity sensors and Bluetooth signals can indicate social contexts by identifying close proximity to others, which helps assess social isolation or engagement [69, 81]. GPS and WiFi data are indispensable for determining semantic locations, enabling insights into whether users are at home, at work, or engaged in recreational activities—each of which correlates with mental states or other behavioral patterns [80, 98, 105].

Interaction sensing captures smartphone user engagement, providing insights into attention, productivity, and emotional states. App usage, for example, can infer attention spans, productivity, and even stress levels, as excessive or reduced engagement with certain apps may be linked to anxiety or depressive symptoms [51]. Studies have shown that interactions with social media, in particular, are tied to mental well-being [14]. Similarly, frequent typing events, rapid touch interactions, or ignoring notifications can signal heightened stress or distraction [121]. Notifications, app-opening frequencies, and related metrics also provide important context for attention spans and can identify potentially compulsive behaviors, such as frequent app-checking, which may indicate underlying mental health concerns [18, 81].

Combining continuous and interaction sensing modalities offers a comprehensive view of behavior, allowing machine learning models to capture nuanced and expressive aspects with a high level of detail. Currently available public smartphone sensing datasets, however, often lack this level of depth and diversity in sensing modalities, limiting the ability to fully understand human behavior at scale. Even if many modalities are present, most publicly available datasets only offer pre-processed versions of data, limiting the use of such datasets for diverse purposes. Therefore, there is an unmet need for raw smartphone sensor datasets that integrate continuous and interaction sensing modalities, offering richer insights into behavioral patterns.

2.3 Effect of Country Diversity on Sensing and Self-reported Ground Truth

Together, continuous and interaction-sensing modalities build a holistic picture of user behavior, capturing both passive and longitudinal patterns and immediate interactions that correlate with mental and emotional states. However, sensor data and self-reported behaviors used as ground truth for training machine learning models can vary significantly across countries, influenced by cultural, social, and environmental differences [90].

For example, sensor data like accelerometer readings or step counts may reflect distinct physical activity patterns based on country-specific factors, such as urban infrastructure, transportation habits, and climate [59, 120, 138]. In countries where walking or cycling is common, step counts may show higher physical activity levels, while regions reliant on driving will have lower activity levels. Similarly, proximity sensor and Bluetooth data capture different social interaction patterns, reflecting cultural norms around personal space, social gatherings, and work environments [57, 61, 88, 104]. In collectivist cultures, for instance, users may exhibit more frequent close proximity with others, while individualistic cultures may show more solitary sensor readings [117, 118]. Cultural attitudes towards technology could also shape app usage and notification interactions [8]. Social media may dominate app engagement in some countries, whereas others might have more work-focused or communication-restricted technology use [22, 91]. Beyond sensing modalities, the way behaviors are reported as ground truth—such as moods, stress, context, or activity types—varies culturally [80, 84, 103]. Cultural norms can influence self-reported stress or mood, leading to different labels for similar sensor data patterns across countries [82, 84, 100]. In machine learning terms, such cross-country variations introduce challenges related to data covariate shift and label shift, both of which complicate generalization [5, 66].

Despite these evident variations, most prior studies overlook country-level differences, relying on homogeneous datasets from specific regions, often within the Global North [81, 90]. This oversight leads to models that struggle to generalize effectively to diverse populations and cultural contexts. In some instances, even though cross-country data were available, analysis has not focused on country differences [105]. Without datasets that capture country-specific data and behavioral patterns, machine learning models are likely to produce incorrect inferences when applied outside the regions in which they were trained. To improve model robustness, globally diverse datasets are essential for exploring behavioral differences across populations, allowing models to account for the variations in behaviors and sensor data across countries and cultures. Addressing this gap is one of the main objectives of the *DiversityOne* dataset.

Furthermore, collecting large-scale, passive smartphone sensing data with accurate labels is challenging due to significant costs, time, and effort, especially when collecting data from diverse international participants [132]. Datasets in this field typically involve fewer than 100 participants (see Table 1) since managing continuous data streams, conducting longitudinal data collection, and providing accurate behavioral labels make scaling these studies complex. Scaling up to engage hundreds of participants requires intensive recruitment, sustained user engagement, and technical infrastructure to handle continuous sensor data collection over extended periods. This is partly why the field of ubiquitous computing (ubicomp) lags behind fields like computer vision or natural language processing, where collecting millions of labeled images, videos, or text is relatively more straightforward and cost-effective, enabling faster progress. However, data collection across countries presents additional challenges, as cultural, behavioral, and infrastructure differences influence how people use smartphones, complicating the labeling process. Unlike images, videos, or text, which can be labeled through standardized methods or crowdsourcing, passive sensing data requires detailed, context-specific labeling of behaviors like mood, activity, and social interaction, often relying on time-consuming self-reports. Cross-country collection further complicates matters with logistical issues such as language differences, varying ethical standards, and compliance with different data privacy laws. These challenges make large-scale, diverse data collection costly and time-intensive. The *DiversityOne* dataset's achievement in collecting data from 782 participants demonstrates the commitment and resources required to overcome these hurdles.

2.4 Sociological Theory and Approaches to Study Design

2.4.1 Cultural Diversity as Social Practices. The concept of culture has been extensively examined in social sciences. Previously, culture was viewed as a cohesive framework that influenced attitudes and practices through socialization [113]. However, recent studies suggest that culture is fragmented and diverse among social groups [75]. This shift in perspective sees culture as complex structures of quasi-rules that individuals can use strategically [11, 106] and highlights the need to analyze relationships among various cultural influences, linking them to distinct phenomena and calling for a more nuanced psychological understanding [19].

The study design of the *DiversityOne* dataset focuses on a particular definition of culture, deriving from the social practice theory [10, 11, 29, 38, 39, 47, 94, 99, 128]. This theory frames human behavior as composed of daily practices that produce societal outcomes and influence individual skills and mindsets, contributing to social structure and culture. According to Shove et al. [108], social practices consist of three key components:

- **Material**: The physical objects or resources, such as a car or a membership, that enable the execution of a particular practice.
- **Competence**: The knowledge, skills, and abilities that make a certain practice possible.
- **Meaning**: The cultural and symbolic elements that give significance to social practices, motivating individuals to perform them in alignment with societal norms.

For instance, environmental consciousness may motivate individuals to choose public transportation, aligning with behaviors like cycling, waste sorting, or adopting a vegetarian lifestyle. These motivations, shaped by material

access (such as bike availability or recycling facilities), personal competence, and meaningful commitment to sustainability, collectively define behaviors that, when widely adopted within a community, become recognized social practices [96, 107].

Thus, there is a distinction between the practitioner (individual) and the social practice (community). While interconnected, social practices exist independently and are established at the social level. Indeed, Røpke [96] noted that practices consist of recognizable, interconnected elements that individuals reproduce, with new members being constantly recruited. In this sense, individuals are "carriers of practices" recruited based on their backgrounds, not merely choosing practices by utility [94], and the distribution of practices often reflects social inequality, with varying cultural perceptions of what constitutes ethical practice. Participation in a practice leaves lasting effects, such as knowledge and skills, which facilitate future involvement in that practice in a path-dependent process [96]. Finally, diversity is socially recognized, and practices are inherently social, resembling one another across different contexts [94]. In other words, social practices exhibit regularity - models of how certain daily practices are typically and habitually performed in (a considerable part) of a society [58].

This perspective enhances our understanding of human behavior in real-world environments, where the activities recorded by sensors may be limited in terms of quantity, quality, and context representation. Identifying whether an individual belongs to a community of practice and their "career" within it enables the inference of important characteristics, such as the nature of their activities—distinguishing, for example, between an athlete's training session or recreational activity—the tools utilized, such as appropriate clothing and supportive technologies, and the likelihood of recurring behavioral patterns over time. This understanding is particularly valuable in situations where data may be limited or less accurate.

Furthermore, gaining insight into the context in which these practices occur allows for culturally informed comparisons. This encompasses both the individual perspective, which considers the factors that may facilitate or impede engagement in a practice or access to a community, and the social perspective, which examines how the reference community typically interacts with that practice.

2.4.2 Capturing Social Practices. Researchers utilize various methodologies to effectively capture behaviors and lifestyles over time. In European countries, for example, the Harmonized European Time Use Surveys (HETUS) measure time spent on various activities. At the same time, the Experience Sampling Methodology (ESM) focuses on the interoceptive, or internal, aspects of behavior [25, 86]. Time-use diaries, commonly used in HETUS, are intensive longitudinal surveys where participants self-report activity sequences over a day, detailing the frequency and duration of their actions to reveal intricate social patterns. Participants typically complete these diaries at regular intervals, tracking each activity, which enables a detailed, objective view of daily routines and interactions [110]. In contrast, ESM prompts participants to report their thoughts and behaviors over days or even months, detailing their personal points of view on their daily experiences. A comprehensive review by Van Berkel et al. [123] highlights the strengths of ESM in capturing these aspects.

Our approach combines these two perspectives, namely time-use diaries and ESM with smartphone sensor data. For every sensor pattern that indicates an action or habit—such as accelerometer data, Movement Activity Labels, or GPS positions—researchers can reconstruct the individual's social and personal context. A study review explores this combination of continuous sensing and interaction sensing with self-report data for the well-being of young adults [81]. Here, self-report data can serve as ground truth events. Conversely, in scenarios like interactive classification in the wild [9], algorithms such as skeptical learning can use sensor information to validate user annotation. Furthermore, this approach enables the researcher to observe human behavior from the individual point of view within the community of practice, specifically her understanding and interpretation of the context in which the activity occurs (see, e.g., [140]).

Table 1. Public available datasets for activity and context recognition in the wild. Duration is expressed in days. The number of pilot sites indicates whether the countries are in the Global North (N) and/or Global South (S). (+) Sensors are divided between smartphones and (+) other devices. No '+' means only smartphone sensors. (*) The GLOBEM data collection was done during a semester for 10 weeks each year, for 4 different years.

Datasets	**Coverage**			**Purpose**	
	Sample	**Days**	**Sites**	**Self-reports**	**Sensors**
MDC (2013) [71]	185	365	1 N	Location	12+14
StudentLife (2014) [125]	48	70	1 N	Sleep-related	10
ExtraSensory (2017) [122]	60	7	1 N	Activity	8+2
Real-life HAR (2020) [35]	19	28	1 N	.	4
ContextLabeler (2021) [15]	3	14	1 N	Activity	18
Qwantify (2022) [127]	242	7	1 N	Desire, emotion, well-being	.
LifeSnaps (2022) [133]	71	120	4 N	Location, Mood, Step Goal	0+23
ETRI (2022) [24]	22	28	1 N	Activity, Location, Relation, Mood	10+4
SmartUnitn2 (2018) [72]	158	28	1 N	Activity, Location, Relation, Mood	28
LAUREATE (2023) [70]	42	91	1 N	Activities, Health	0+6
GLOBEM (2023) [131]	534	280*	4 N	Standard scales (personality, physical, mental and social well-being)	~10
EgoADL (2024) [111]	30	5	1 N	Automated labels	4+2
DiversityOne	**782**	**28**	**8 (3N,5S)**	**Activity, Location, Relation, Mood**	**26**

2.5 Currently Available Public Smartphone Sensing Datasets

Several datasets that leverage smartphone and smartwatch sensors for activity and context recognition have been developed across various research fields. Table 1 compares key publicly available datasets, highlighting their sample size, duration, data collection sites, self-reported questionnaires or annotations via intensive longitudinal survey, and number of sensors used, both collected via smartphone and other devices. While these datasets contribute valuable insights, most are constrained by small sample sizes and short data collection durations, which limit their ability to capture daily routines and behavioral patterns. For instance, although datasets like MDC [71], StudentLife [125], and Real-life HAR [35] extend beyond short-term studies, they still lack the diversity required to observe routine behaviors that often necessitate more extended observation periods and cross-country differences.

Datasets that focus solely on continuous sensor data, such as Real-life HAR [35], are limited in capturing the complexity of human activities because they do not include annotations reflecting users' experiences or contexts. Moreover, a growing trend involves using experience sampling methodologies (ESM), where users frequently self-report their experiences, to capture behaviors and contextual nuances. Datasets like the Qwantify app [127] provide valuable insights into emotions, desires, and well-being. Still, sensor data in these collections often remain secondary, serving mainly as context rather than central to the analysis. Hence, the most valuable datasets combine sensor data with self-reports about the collected data. Such datasets can be further divided into those with in-situ (reports provided about the current moment) and retrospective (reports provided about past periods or days) self-reports [81]. A notable example of retrospective self-reporting is the work of Krumm and Rouhana [68], which uses time diaries (i.e., similar to the HETUS approach described in Section 2.4.2) for labeling. However, the lack of direct user interaction during labeling can lead to inaccuracies and contextual

inconsistencies. In contrast, datasets like StudentLife [125], ExtraSensory [122], and ContextLabeler [15] rely on in-situ self-reports from users. Still, they are often limited to specific activities and do not adhere to a standardized reference, which introduces social and cognitive biases.

GLOBEM [131] presents a large-scale dataset that enables cross-dataset generalization analysis, which is particularly valuable for examining behavioral patterns across different periods and university environments. Its longitudinal nature, spanning over four years (ten weeks of data collection each year), allows researchers to assess how models perform across diverse academic settings over time. However, the dataset predominantly focuses on the USA, limiting its applicability for cross-country studies. Hence, while GLOBEM provides essential insights, its geographic concentration contrasts with the more globally diverse scope of *DiversityOne*, which includes participants from eight countries across the global north and south, offering a broader foundation for generalization and cross-country behavior modeling. Moreover, *DiversityOne* provides more fine-grained and raw sensor data and frequent in-situ self-reports, giving researchers much flexibility and depth in their analysis.

Among existing datasets, LifeSnaps [133] is the most comparable to *DiversityOne* in terms of countries of data collection. However, despite the richness, the dataset falls short in crucial areas such as sample size, diversity of data collection countries, and breadth of sensor coverage. LifeSnaps focuses primarily on four European countries and only has a sample size of 71 participants, whereas *DiversityOne* spans both the global north and south, including a broader cultural and geographic range, with 782 participants. This comprehensive scope makes *DiversityOne* a unique resource for understanding everyday life behavior across diverse countries, surpassing other available datasets in terms of depth and scale.

2.6 The *DiversityOne* Dataset

Our dataset aims at making a valuable contribution by differing significantly from previous works in several key aspects, as outlined below. First, as described in Section 2.4.2, our dataset methodology follows best practices from earlier studies by measuring behaviors and psychosocial traits using standardized and validated methods. We propose a combination of self-reported annotations and sensor data from smartphones. This integration enables a nuanced understanding of social practices, considering routine behaviors alongside contextual data about both physical actions and mental states. While many current datasets aim to capture such elements, few achieve the same depth of blending multiple methodologies with sensor data. This multifaceted approach allows for a more accurate, holistic view of human behavior, providing fine-grained labels essential for training robust machine learning models that reflect the complexity of everyday life. Second, to the best of our knowledge, there are currently no publicly available smartphone datasets specifically tailored to examining social practices (see Section 2.4.1). This innovative framework allows researchers to enrich their understanding of individuals by situating them within a broader community of profiles that share similar competencies, materials, and meanings. For instance, it distinguishes between those engaged in professional activities and those pursuing recreational interests, highlighting each group's distinct personalities and values. This nuanced approach encourages researchers to explore the diversity of daily activities in a way that emphasizes the cultural context of behaviors, thereby creating a more comprehensive understanding of how social practices shape individual actions and interactions.

Finally, our sample size and the breadth of data collected internationally, with a particular emphasis on regions in the global south —underrepresented in prior research— is rather unique. In addition to our theoretical and methodological approach, these characteristics enable a more comprehensive exploration of human behavior, shedding light on the diversity among people and their similarities. In summary, by integrating diverse cultural and socio-demographic factors, *DiversityOne* facilitates a deeper exploration of behavior modeling, personalization,

and cross-cultural adaptation at a more nuanced level. Moreover, incorporating these factors fosters cross-cultural and multidisciplinary analysis of human behavior rather than promoting it solely as a machine learning benchmark dataset [87, 93].

3 Methodology

The design and collection of this dataset posed the challenge of ensuring consistent data comparability between different countries while mapping social practices in a culturally sensitive and diversity-aware manner. Behaviors and practices vary considerably according to time of day and location and are shaped by different cultural and social factors. Local cultural variations also impact ethical and privacy considerations, influencing the adoption of data collection protocols, legal frameworks, and public perceptions. Furthermore, the diversity of participating organizations and the resources available in each context significantly shaped the data collection process.

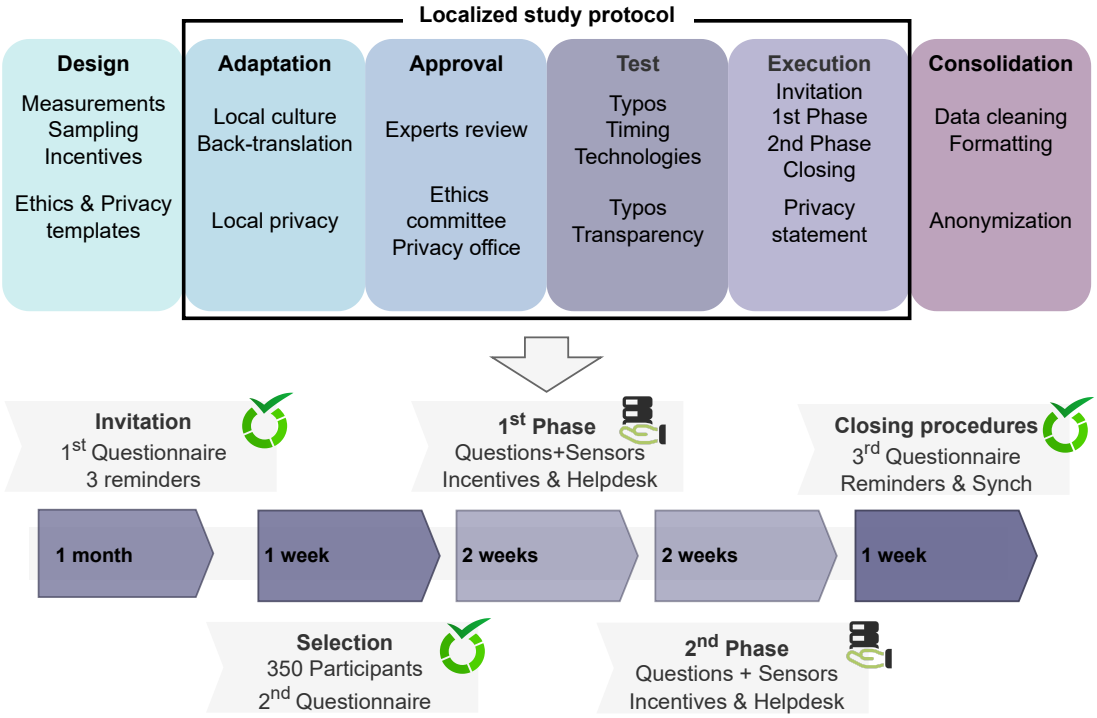


Fig. 1. Study set-up and data collection process. Invitation, Selection, and Closing procedures were done using the LimeSurvey platform. The 1st Phase and 2nd Phase were done using the iLog mobile app.

Figure 1 shows the study set-up and the survey protocol, as applied consistently across all sites, with minor adjustments which still maintained comparability (refer to Section 3.6). The protocol was developed with an adaptive approach which allowed to address local perspectives and requirements. A team of experts, including computer scientists, social scientists, interaction designers, incentive designers, ethicists, and legal representatives, developed the study design, survey protocol, and procedures to ensure compliance with ethics and privacy

standards. The experiment partners then adapted these materials in a way to align with local contexts. Each adapted version was validated by the initial team or an authorized body to ensure functionality, adherence to standards, and comparability.

For the study design, UNITN and LSE identified the need to measure human behavior over time by adopting both questionnaires and an intensive longitudinal survey approach [60, 79] enabled by an adapted version of the iLog app [136][3]. The content of the questions and the technology were tailored to address local needs (refer to Sections 3.1 and 3.2). The questions from the iLog app and the questionnaires were subsequently translated into multiple languages, as detailed in Section 3.1. Furthermore, this type of intensive longitudinal survey often faces challenges related to participant dropout, which can impact the quality of the results. According to [27], dropout in longitudinal studies can originate from *(i)* failure to locate the research unit, *(ii)* failure to contact the potential respondent, and *(iii)* failure to obtain cooperation.

Moreover, repeated assessments can become burdensome, potentially leading to decreased compliance over time. However, according to Wrzus and Neubauer [129], factors such as the total number of assessments or the frequency of assessments do not impact compliance. Indeed, the burden on respondents can be mitigated, as participants who agree to participate are already willing to adhere to demanding schedules. There is a consensus that offering monetary incentives can enhance participation rates. Wrzus and Neubauer [129] also discusses how it is possible to hypothesize that including break options could improve compliance, although their meta-analysis did not find evidence to support this idea.

A survey protocol was developed in response to these considerations. This protocol involved sending invitations on the same day to all students enrolled at various universities via their institutional email accounts, addressing the issues of locating research units, and contacting potential respondents. We tackled the challenge of obtaining cooperation by ensuring the participants received all the necessary information in advance and incentivizing participation in the one-month survey with monetary compensation. Adjustments were made to the invitation language, sample selection, and incentive amounts to better align with local requirements, as detailed in Section 3.3.

Data collection involved handling personal data across both European and non-European countries. European legislation mandates that any processing of data related to European citizens or conducted using tools based in Europe must comply with the General Data Protection Regulation (GDPR). Consequently, UNITN prepared procedures and documentation to ensure GDPR compliance. These processes were then adapted for different contexts to ensure effective data management, clear communication with participants, and adherence to privacy requirements where necessary (refer to Section 3.5).

The following sections outline the strategy for adapting measurement tools, privacy protocols, and methods to collect high-quality data for studying diversity in human behavior and social practices over time. Each section addresses both the general approach and specific local adaptations.

3.1 Questionnaires about Demographics and Social Practices

The administered questionnaires measured various aspects of social practices and identified communities of practice relevant to students. Given the wide range of social practices, we focused on a specific set, including social relationships (both online and offline), cultural and sports activities, daily commuting, and shopping and cooking habits. The questionnaire items were designed to capture elements of competence, materiality, and meaning associated with these practices (see Section 2). Each question and scale thus provide foundational information on components of social practices. For example, a profile of an individual's shopping and cooking habits can be built by examining the types of food purchased (materials), cooking skills (competencies, e.g., buying convenience

[3]iLog is an app designed to collect intensive longitudinal survey data (in the form of daily self-reports or hourly time diary self-reports) along with passive smartphone sensor data. The initial version of the app, made available by Zeni et al. [135, 136], was modified with new features to facilitate the collection of the *DiversityOne* dataset.

foods *vs.* cooking daily), and motivational values (meaning, e.g., hedonism *vs.* universalism). This combination of components yields a nuanced measure of specific social practices, allowing cross-country comparisons to identify diverse practices. Due to the extensive number of questions, we divided the questionnaire into three parts and administered it throughout the data collection, as mentioned in Figure 1.

(1) The **First Questionnaire** was sent via an invitation email to all students at each university participating in the study. This questionnaire covered demographic information, cultural interests, and leisure activities, as well as questions about online and offline social interactions. It also included standard scales for measuring personality traits using the **Big Five Inventory** [28] and attitudes toward values measured through the **Basic Values Survey** [48].
(2) The **Second Questionnaire** was administered only to participants involved in the sensor and survey data collection. It primarily explored specific social practices such as daily commuting, cooking, grocery shopping, and physical activities. Additionally, a second set of personality questions based on the **Jungian Scale for Personality Types** [13, 62, 78, 126] was included, along with a second set of values questions based on the **Human Values Survey** [101, 102].
(3) The **Third Questionnaire** was administered to the same participants as the second questionnaire. It gathered feedback on the user's experience with the data collection app and assessed a **Multiple Intelligence Scale** [115].

The structure of the questions reflects the goal of measuring different components of social practices. We adopted standard scales to assess aspects related to meaning—often intangible and challenging to quantify—while additional questions aimed at understanding the other two components of social practices, namely competencies and materials. These components were evaluated through questions about specific practices, such as commuting, sports, cultural activities, and cooking and shopping habits. To administer the questionnaires, we used *LimeSurvey* [73], a tool widely used in social sciences to protect user privacy. The full questionnaires, translations and adaptations are available in the data catalog, see Section 5.

3.1.1 Questionnaire Translation, Testing, and Adaptation. Each experiment site received the English version of the questionnaire. This version was later translated into the local language and adapted to the local context. We used a validated translated questionnaire whenever available for the standard scales mentioned above. Then, in case of absence, translations were completed by an expert translator and validated through panel and back-translation processes [20] by independent translators. Adaptations were made to tailor specific questions and answer options to each site. This was necessary for questions such as field of study, which varies depending on the educational programs offered by each institution, and accommodation, which differs based on personal choices, available facilities, and cultural norms. For example, at NUM, there is substantial migration to the capital for work and study, intertwining with students' educational journeys and influencing their choices. At JLU, all students live on the university campus, contrasting with other institutions that offer various accommodation options, including university-provided, public, or private housing. Further adaptations addressed unique local needs. At JLU, response options related to nationality were modified to exclude countries whose sovereignty is not recognized by the local government. Additionally, JLU adapted the list of social media and applications list to include those popular in both Western and Asian markets.

Certain items related to sexuality ("To have sexual relationships; to obtain sexual pleasure") and religiosity ("To believe in God as the savior of humanity; to fulfill the will of God") from the Basic Values Survey [48] were also modified. The content was deemed inappropriate in some locations, such as AMRITA and UC, and alternative options like "love" were suggested. For the religiosity item, an option with other religious and spiritual figures, such as Buddha, was explicitly provided for AMRITA and JLU. Finally, aware of the cross-cultural variation of personality traits in the case of JLU, the locally developed personality trait scale [141] was adopted in addition to the Western version, thus facilitating a comparison.

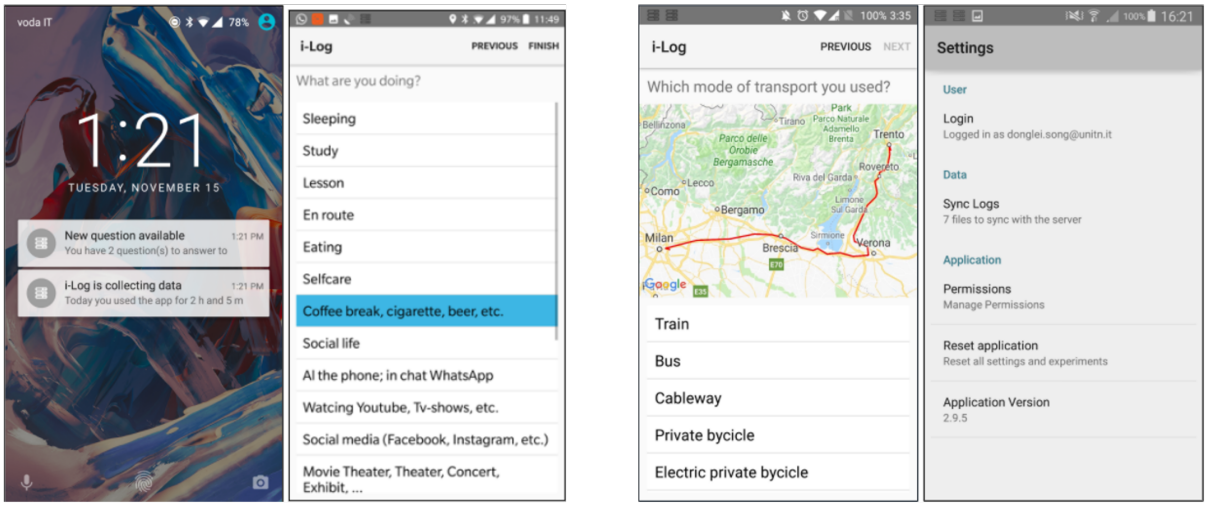


Fig. 2. The iLog app adopted for intensive longitudinal survey and sensor data collection.

After translation and adaptation, the questionnaires were tested locally. They were distributed to approximately 30 participants, including project partners and university students. This allowed to validate the translations and assess completion times.

3.2 Intensive Longitudinal Surveys and Passive Sensor Data Collection

Temporal aspects are essential for understanding how everyday life behavior evolves over time. Intensive longitudinal surveys are valuable for observing human behavior as it changes, especially when paired with detailed sensor data from everyday devices. Thus, we adopted the `iLog` [136], an App that effectively collects this information through user interactions and data from smartphone sensors[4]. Figure 2 shows the application screenshots. The app's ability to gather user self-reports and passive sensor data makes it unique (see [67, 97, 125] for other similar tools). This dual capability significantly enhances traditional time diary methods [45, 110], mainly when structured [55] to enable real-time user responses. In addition, the application enters power-saving mode when the smartphone reaches around 8% of the phone battery charge, thus limiting cases of non-response due to the smartphone being turned off.

3.2.1 Diary Methods and Daily Routines. Daily diaries, as a form of self-reports, are foundational for studying daily human behavior. Following standards set by HETUS and ATUS, three types of daily diaries were implemented[5]:

- **Morning and Evening Diaries**: The first diary collects information at the start and end of each day. At 08:00 AM, participants received two qualitative questions about sleep quality and daily expectations, listed in Table 10. At 10:00 PM, participants responded to questions in Table 11 regarding *(a)* a rating of their day, *(b)* any challenges encountered, *(c)* how they addressed these challenges, and *(d)* reflections on the COVID-19 pandemic.

[4]The use of iLog and collected data in various experiments has been described in [41, 42, 44, 46, 135]. At the time of this study, only Android devices were supported. Currently, the new app version runs on both Android and iOS devices.
[5]Answer options were supplemented with explanatory notes, not included here for readability. Complete documentation is available in the catalog (see Section 5).

- **Time Diaries**: The second time diary follows the HETUS standard (Table 2), covering main activities and moods[6]. Throughout the study, participants were asked to answer four brief questions every thirty minutes during the first two weeks and every hour during the last two weeks:
  - *What are you doing?*: Participants select from 34 activities, such as sleeping, eating, and working.
  - *Where are you?*: The current location is identified from 26 categories, including home, workplace, university, and restaurant.
  - *Who is with you?*: Participants indicate their company, with options like "alone," "with my partner," and "with friends."
  - *What is your mood?*: Participants rate their mood on a five-point scale ranging from happy to sad.

  If a participant reports engaging in "eating," "traveling," or "doing sport," they are prompted to answer additional questions (Table 13):
  - For eating, participants select foods and drinks from 20 categories, such as rice, potatoes, meat, and beer (adapted from [36]).
  - For sports, participants specify the type of sport from nine options, such as jogging, running, or water sports.
  - For traveling, participants state *(a)* the purpose, selecting from seven categories (e.g., study, social life), and *(b)* the mode of transport from 16 options (e.g., car, bus).
- **Snacks Diary**: As listed in Table 14, these questions document food and drink consumption outside of main meal times. These prompts appear every two hours, allowing participants to select multiple items.

To reduce the participant burden [31, 95], which can lead to non-response missing data or wrong answers [7], notification breaks were incorporated (see Table 12). These breaks allow participants to pre-fill answers or pause notifications temporarily. For example, participants can select the "Go to sleep" option in the evening to silence notifications for six hours or pause notifications for two hours during classes or sports. Moreover, each notification expired only after 12 hours, allowing participants to choose not to respond immediately. They could accumulate a maximum of 24 unanswered notifications. After this limit, the application deleted the oldest notifications and marked them as "Expired" in the dataset.

3.2.2 Sensor Data. During the study, iLog collected data from all available smartphone sensors, spanning continuous and interaction sensing modalities (see Section 2), at a very high level of granularity to capture the breadth of daily behaviors, which can be studied through various analytical techniques. Appendix A.2 details the sensors and their collection frequency or whether they are event-based. The collected sensors support research in the following areas:

- **Connectivity**: includes Bluetooth Low-Energy, Bluetooth Normal, cellular network, connected WiFi network, and discovered WiFi Networks, measuring connections to nearby devices.
- **Environment**: includes ambient temperature, light, pressure, and relative humidity sensors, which are hardware-based and available only on compatible devices.
- **Motion**: captures participants' movements in different contexts and includes accelerometer, activities, gravity, gyroscope, step counter, and step detector.
- **Position**: includes location, magnetic field, and proximity sensors, tracking physical positioning and interaction with the device.
- **App Usage**: tracks social media and app interactions, including headset plug, music playback, notifications, and running applications.

[6]The questions design aligns with the definition proposed in [43, 72] concerning a model of the situational context of a person which can be applied in organizing and reasoning about massive streams of sensor data and annotations. This alignment expands the dataset's potential, enabling its adoption in fields such as knowledge graph modeling and engineering.

Table 2. Time diaries collecting contextual information about the participant every half hour.

A3. What are you doing?	**A4. Where are you?**
(1) Sleeping	(1) Home apartment, room
(2) Personal care	(2) Home garden, patio, courtyard
(3) Eating **(go to A3c, Table 13)**	(3) Relatives Home
(4) Cooking, Food preparation & management	(4) House (friends others)
(5) Study/work group	(5) Classroom, Laboratory
(6) Lecture, seminar, conference, university meeting	(6) Classroom, Study hall
(7) Did not do anything special	(7) University Library
(8) Rest/nap	(8) Other university places
(9) Break	(9) Canteen
(10) Walking	(10) Other Library
(11) Travelling **(go to A3a1, A3a2, Table 13)**	(11) Gym, swimming pool, Sports centre…
(12) Social life	(12) Grocery Shop
(13) Happy Hour, Drinking, Party	(13) Supermarket…
(14) Phone/Video calling	(14) Street markets
(15) In chat on Internet or reading, sending e-mail	(15) Shops, shopping centers, indoor markets, other shops
(16) Surfed or seeking, reading information via Internet	(16) Café, pub, bar
(17) Social media (Facebook, Instagram, etc.)	(17) Restaurant, pizzeria, Street food vendor
(18) Watching TV, video, YouTube, etc.	(18) Movie Theatre Museum…
(19) Listening to music	(19) In the street
(20) Reading a book, periodicals, news, etc.	(20) Public Park/Garden
(21) Movie Theatre Concert…	(21) Countryside, mountain, hill, beach
(22) Entertainment Exhibit, and Culture	(22) Workplace, office
(23) Others Entertainment and Culture	(23) Weekend home or holiday apartment
(24) Arts	(24) Hotel, guesthouse, camping site
(25) Hobbies	(25) Another indoor place
(26) Games	(26) Another outdoor place
(27) Freetime study	
(28) Sport **(go to A3b, Table 13)**	
(29) Voluntary work and participatory activities	
(30) Household and family care	
(31) Grocery Shopping	
(32) Other Shopping	
(33) Work	
(34) Other	

A5. With whom are you?	**A6a. What is your mood?**
(1) Alone	(1) 😃
(2) Friend(s)	(2) 🙂
(3) Relative(s)	(3) 😑
(4) Classmate(s)	(4) 😕
(5) Roommate(s)	(5) ☹️
(6) Colleague(s)	
(7) Partner	
(8) Other	

- **Device Usage**: tracks device interaction patterns, i.e., airplane mode, battery charge, battery level, doze, ring mode, screen status, touches, and user presence sensors.

3.2.3 Intensive Longitudinal Survey Translation, Testing, and Adaptation. Like the previous questionnaires conducted with LimeSurvey, the intensive longitudinal self-report questions and responses were translated into local languages. In addition to the local coordinators, approximately 15 participants downloaded and used the iLog app for two weeks to assess user experience and technical stability in a separate test. We performed adaptations to accommodate specific local interests and needs. Some partners were interested in investigating eating behaviors and eating disorders. Thus, additional questions were added to cover food consumption and dietary routines, balanced with standard time diary items, to minimize invasiveness. Regional variations in response options were also included to reflect local food and beverage consumption. Additionally, IPICYT opted to collect the notification at a half-hour frequency for the entire month.

Further technical adaptations were made to the iLog app code provided by [135]. For instance, iLog typically relies on Google services to receive notifications from the server. However, these services were unavailable in certain regions, such as at JLU, or in areas with limited cellular network coverage. An "offline" version of iLog was developed to overcome this issue. Here, the notification schedules are downloaded at the beginning using WiFi connections and displayed on participants' smartphones at planned times, removing the dependency on Google services or data connections. Participants can then synchronize their collected data via WiFi as needed. Initially designed for JLU, the offline version was also implemented at AMRITA and IPICYT for their data collections.

3.3 Target Population and Sampling Strategy

We focused on a specific target population to ensure effective comparisons across study sites: *college students*. Globally, students tend to have similar routines, such as attending classes, studying, working part-time, and participating in extracurricular activities. This shared lifestyle makes it easier to compare social practices, revealing commonalities and differences across diverse contexts—i.e., the "unity in diversity" of their behaviors.

The sampling strategy was designed with specific criteria to represent the student population. Except for JLU and IPICYT, all registered students at each university were invited to participate in the study and complete the first questionnaire. From the participant pool, individuals with incompatible smartphones (e.g., Android OS below 5.0) or low survey attendance were excluded. The sample was designed to be balanced by gender, age, and academic department to minimize routine variation due to scheduling, community characteristics, and demographics. Thus, the dataset is comparable to existing studies, such as [125], and aligns with similar studies that follow this methodology and collect comparable variables [46, 72].

3.3.1 Recruitment Strategy. Students at each pilot site received an invitation to their institutional email accounts, followed by four monthly reminders. Recruitment strategies differed at JLU and IPICYT, employing snowball sampling via WeChat and WhatsApp. After this initial phase, 350 students were selected according to predefined sampling criteria to ensure a minimum of 250 active participants, accounting for typical longitudinal survey dropout rates of 30% to 70% [52]. In the pilot sites where achieving this threshold proved challenging, all students who completed the questionnaire and owned compatible smartphones were invited to participate.

The selected students received a second email with attachments, including instructions for downloading the iLog app, survey information, and a privacy statement detailing data processing methods. This email also included the second questionnaire. Participants were given approximately five days to download the app and complete the registration, receiving additional official communications halfway through the survey (along with the third questionnaire) and at the study's conclusion. The helpdesk provided daily personalized support. Table 4 summarizes the sampling and recruitment strategy outcomes.

Table 3. Incentives were paid upon completing at least 85% of the notifications after the first and second two-week periods. Additionally, daily and final prizes were awarded through random selection. Each pilot site adapted the incentive protocol based on local practices from similar studies they had conducted previously.

	Payments		**Daily**	**Final Prizes**	
	1st Phase	**2nd Phase**	**Prizes**	**1st Phase**	**2nd Phase**
JLU	100 rmb	100 rmb	1 of 20 rmb	3 of 88 rmb	3 of 88 rmb
AAU	150 kr	150 kr	5 of 40 kr	3 of 800 kr	3 of 1200 kr
AMRITA	-	-	-	-	-
UNITN	20 €	20 €	5 of 5 €	3 of 100 €	3 of 150 €
IPICYT	-	-	-	-	-
NUM	10k MNT	10k MNT	5k MNT	100k MNT	150k MNT
UC	25k GS	25k GS	10 vouchers	1 restaurant voucher	1 restaurant voucher
LSE	-	-	-	£150 (1/50)	£150 (1/50)

3.3.2 Recruitment Adaptations. Specific adaptations were made to accommodate local needs. During the invitation phase, some modifications were implemented at JLU and IPICYT. At JLU, the survey invitation was posted on social channels and in various WeChat groups, reaching about 5,000 students. Reminders were shared on these channels rather than being sent directly to participants. At IPICYT, recruitment was conducted through direct participant engagement, with reminders delivered verbally and via messages in a WhatsApp group created specifically for the study. Additionally, at IPICYT, the selection phase occurred before the first questionnaire was distributed.

3.4 Incentive Strategy

A group of experts designed the incentive strategy. It was informed by sociological literature, particularly the work of Singer and Ye [109], who extensively reviewed types of incentives (monetary and non-monetary) and their effectiveness in reducing respondent bias and dropout rates. The review suggests that monetary incentives are generally more effective than gifts, and prepaid incentives outperform random prizes. Based on these findings, an incentive strategy was developed, focusing primarily on fixed payments. Table 3 reports the incentive strategy adopted in the different pilots. While no incentives were adopted for the invitation questionnaire, in the first and second phases of iLog data collection, selected students received monetary rewards for completing at least 85% of notifications. Additionally, daily prizes and three final prizes were awarded in both phases.

3.4.1 Incentives Adaptation. The incentives were adjusted according to the cost of goods in each country. In addition, they were tailored to local cultures and circumstances. Thus, in the case of the LSE, it was decided to offer a prize of £150 drawn randomly for every 50 participants, aligning with the incentive methods proposed at the same institute. Similarly, at AMRITA, no monetary incentives were implemented; instead, participation certificates were provided, reflecting the local culture of engagement in science. This strategy was particularly successful for questionnaires but did not yield the same results for the intensive longitudinal survey, as discussed in Section 6.3. This lack of success may also be attributed to the novelty of the approach. Lastly, at IPICYT, no monetary incentives were used either. Instead, there was a focus on social engagement by creating a WhatsApp group, which facilitated multiple meetings and exchanges of interests and opinions between participants and researchers.

3.5 Ethics and Privacy

Addressing privacy in cross-country research requires adapting to diverse regulations, social norms, and individual attitudes toward privacy [2, 17, 33]. Privacy management requires careful consideration of ethical and legal aspects tailored to each context. Consequently, each university's principal investigator or legal deputy was designated as data controller, thus covering a legal and ethical role of responsibility. Specific data processing responsibilities were also delegated to UNITN, which managed data collection and management, as outlined in Section 3.7.

From an ethical perspective, the data controller in each country appointed a local ethics committee to support and approve project activities. If universities or research centers lacked an ethics committee, the local institution designated a suitable one. To facilitate this process, UNITN, in collaboration with its ethics committee and privacy office, developed templates and documentation that were then adapted to meet local requirements.

Given that the iLog app operates within Europe, the methodology adopted a Eurocentric approach, benefiting from the comprehensive framework provided by the GDPR, which governs personal data processing. UNITN supplied all necessary documentation—such as privacy statements, information protocols, and additional guidelines—beginning with GDPR standards and extending them with clauses as required by local regulations. In essence, regulations in each country that do not conflict with the GDPR were considered complementary.

This approach ensured that each country's implementation adhered to procedures and documentation based on European legislation while accommodating local laws. For example, in the case of IPICYT, any participant request for data deletion must be fulfilled within three days—a timeframe stricter than GDPR requirements and incorporated into the local documentation and protocols.

3.6 Data Collection

The academic calendar and various aspects of university life greatly influence students' routines, often more so than other factors. To ensure the ecological validity of the study, data collection at all sites was carried out during the academic semester. This approach guarantees that participants likely adhered to similar academic routines as their peers in other countries. The questionnaires, longitudinal time diaries, and passive sensor data were collected consistently across the eight pilot sites. The data collection process spanned approximately ten weeks, divided into the following phases (see Figure 1):

- **Invitation**: The first phase involved sending an email with a survey description, an invitation to complete the first questionnaire, and information on the second part of the survey. This invitation was followed by three weekly reminders for students who had not yet completed the survey.
- **Selection**: At this stage, a subset of eligible participants was selected for the second part of the survey. Selection criteria included consent to personal data processing and possession of a compatible Android smartphone.
- **First Phase**: This phase began with distributing the second questionnaire to the selected participants, followed by a reminder after one week. Along with the second questionnaire, participants received an email with instructions on downloading the iLog app, along with a brief user manual.
- **Second Phase**: During this final data collection phase, the third and final questionnaire was sent, followed by a reminder after one week. In this phase, the frequency of iLog time diary prompts was reduced.
- **Closing Procedures**: At the end of the survey, participants received a final email with instructions on app uninstallation and a last reminder, where necessary, to complete the second and third questionnaires.

Daily reports were generated throughout the data collection to monitor iLog survey responses and quickly identify potential issues. Based on these reports, local field supervisors contacted inactive participants every three days to provide support as needed. Participants were also contacted with the results of the daily prize draws (see Section 3.4). Moreover, due to timing issues, primarily related to technical challenges at JLU and delays in

ethics committee approvals at AMRITA and IPICYT, data collection in these countries began several months later than initially planned.

3.7 Data Management

As discussed in Section 3.5, ethical and legal considerations led to establishing protocols to ensure data ownership and privacy compliance, including copyright management, data cleaning, and anonymization, described below.

3.7.1 Data Ownership and Copyright. At the end of each data collection, UNITN, acting as the data processor, securely transferred the entire raw dataset—including personal data—to the designated local data controllers. UNITN maintained a copy of the non-anonymous data for further preparation, thus sharing a processed copy of the cleaned dataset with all data controllers and researchers involved in refining and analyzing the data. Towards the end of the project, each data controller licensed their cleaned and anonymized dataset to UNITN, establishing shared management and distribution policies, partly described in Section 5. This procedure ensured local data controllers retained ownership while facilitating legally compliant data sharing for research purposes.

3.7.2 Data Cleaning and Anonymization. To ensure GDPR compliance and achieve high data quality, the dataset underwent *Data Cleaning* and *Anonymization*. Data cleaning included standardizing the dataset and addressing non-random missing data due to technical errors. Variable and value labels were adapted for clarity, making the dataset understandable to third-party researchers. On the other hand, Anonymization aimed to prevent individual identification through (i) Personal Data Anonymization, (ii) Network Anonymization, and (iii) GPS Anonymization.

Regarding personal data anonymization, all personal information, such as email addresses, home addresses, names, and surnames, was removed from the three dataset types (online questionnaire, time diaries, and sensor data). Unique identifiers were assigned to ensure consistency across these datasets. Considering Network Anonymization, potential re-identification sources, such as the WiFi network connected to the smartphone, cellular networks, and available WiFi networks, were anonymized. Relevant sens r file columns were anonymized using irreversible hash functions applied to WiFi network names. Regarding GPS Anonymization, high-precision GPS data were intentionally truncated to maintain anonymity while preserving scientific usability. Anonymization included identifying points of interest (POIs) where participants spent significant time and providing generalized location data. Two versions of this dataset, named *RoundDown* and *POI*, were created. To further protect privacy, a single research institution can access only one of these datasets to minimize re-identification risks associated with combining the two versions.

It is important to note that, despite these measures, the dataset could still allow re-identification if cross-referenced with other datasets. For this reason, the dataset is not publicly available online and can only be accessed under specific conditions (see Section 5).

4 Validation

Our dataset provides a rich and diverse array of information, making it possible to extract valuable insights into students' behavior across different pilot sites. Below, we present key descriptive statistics[7] and focus on crucial dimensions that highlight study participation and the diversity observed in the students' everyday lives.

4.1 Overall Study Engagement

Study participation is described according to the protocol outlined in Section 3.6. As shown in Table 4, during the invitation phase, 18,628 students responded to the initial questionnaire, with peak participation at the UNITN, AMRITA, and NUM pilot sites, each registering over 3,000 responses. Most sites saw more than 1,000 students

[7]The statistics are compatible with those reported in previous studies analyzing the dataset, as they applied data filtering.

Table 4. Number of participants for each of the three questionnaires and data collection. "iLog signed" and "iLog data" columns show the total number of students who logged into the iLog app and those who actively contributed with data, respectively. Additionally, demographic information of the participants involved in the iLog data collection is reported.

Country	Acronym	1st QU	2nd QU	3rd QU	iLog signed	iLog data	μ Age (σ)	% women
China	JLU	1,007	69	41	54	44	19.4 (2.2)	61
Denmark	AAU	412	16	15	27	20	28.2 (6.3)	58
India	AMRITA	4,183	141	38	64	45	21.4 (2.9)	29
Italy	UNITN	5,692	287	215	263	241	22.1 (3.2)	56
Mexico	IPICYT	40	9	11	21	21	25.2 (4.1)	33
Mongolia	NUM	3,972	214	152	231	201	20.0 (3.1)	65
Paraguay	UC	1,342	33	25	43	28	23.3 (5.1)	61
UK	LSE	1,980	143	45	88	66	24.6 (5.0)	66
Total		**18,628**	**912**	**542**	**782**	**666**		

participate, except for IPICYT and AAU. The lower participation at these sites is likely due to the recruitment strategies used at IPICYT and the communication channels adopted at AAU. In the first case, participants may not have encouraged their acquaintances to take part in the survey, as in the usual snowball sampling procedures. In contrast, in the second case, students might not have been used to checking their emails or visiting the institutional website, so they had not received the communications.

After the invitation phase, up to 350 students from each site were randomly selected for participation in the iLog data collection. This selection process considered factors such as gender and the students' enrolled departments to ensure a representative sample. This approach aimed to address potential biases in the data that might arise from differences in lesson schedules across departments, which could influence students' daily behavior.

Among the invited participants, 782 students installed the iLog app. Table 4 details the number of students who signed into the app and those who actively provided data. The number of participants varies across universities, with the highest participation rates observed at NUM and UNITN and the lowest at AAU and IPICYT. These differences are also evident in the second and third questionnaires, distributed exclusively to previously selected participants. The uneven participation across pilot sites could be due to several factors, such as smartphone compatibility with the iLog app, individual inclinations and preferences, and variations in the research protocols and engagement methods (detailed in Section 6.3). However, the dataset remains consistent and suitable for high-quality analysis, as shown in Section 6, presenting a unique opportunity for methodological exploration and analysis. These analyses can examine participant profiles in relation to survey participation, as well as response rates and response times (as outlined below), offering valuable insights for optimizing future data collection.

Having introduced participant engagement at various stages, Figure 3 and Figure 4 illustrate the response rates and sensor data provision during the intensive longitudinal survey, as discussed in the following sections.

4.2 Response Rate in Time Diaries

Figure 3 shows the number of responses to the daily questions outlined in Section 3.2.1. We expected approximately 1400 responses per participant throughout the study, including morning and evening questions, time diaries, and snack questions. In all pilots (except AMRITA), 40% of participants provided an average of 20 responses daily, equating to roughly ten hours of annotated sensor data each day. Notably, the daily response rate is significantly higher at UNITN and NUM, where over 150 and 130 participants, respectively, provided more than 20 responses

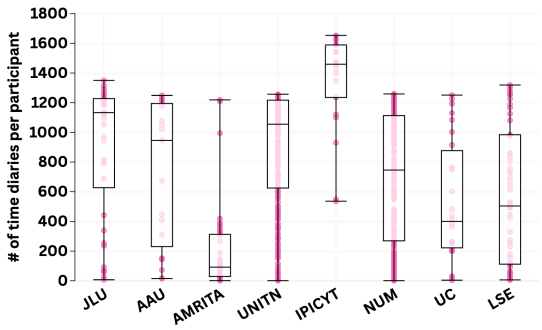


Fig. 3. Distribution of participants based on the number of daily diaries completed. Each dot represents a participant.

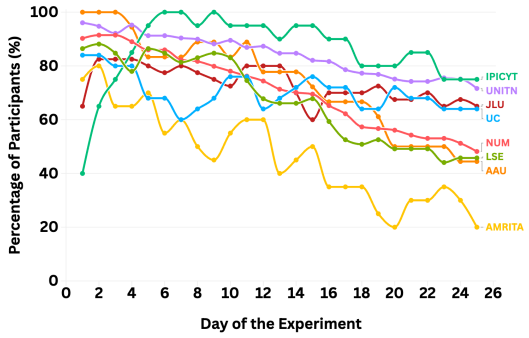


Fig. 4. Percentage of participants at each pilot site providing sensor data for each day.

per day. JLU and IPICYT also had high participation rates; however, the latter's increased response rate was partly due to the study configuration (see Section 3.2), which included more questions than other pilot sites.

Variations in response rates can be attributed to several factors, including adaptations to the survey protocol mentioned above, as well as technical issues and participant characteristics. Some participants did not receive all notifications correctly due to smartphone settings, connectivity issues, or server management. While these technical problems affected only a small fraction of the data, they may have impacted participant engagement.

Furthermore, the data collection process still required the participants to be considerably careful and organized, despite measures such as break options and maintaining active notifications for 12 hours (see Section 3.2). Personal factors, such as interest and curiosity (like the appeal of tracking one's activities to increase self-awareness), as well as tendencies toward procrastination or stress, may have influenced response consistency.

Ultimately, each student's consistency of participation mainly varied due to dropout effects. While some students ceased involvement after a few days, others maintained consistent engagement throughout the data collection period.

4.3 Sensor Data Provision

Sensor data collection adhered to state-of-the-art methods, ensuring that studies based on this dataset can achieve high replicability. We also present details on dropout rates in Figure 4, along with the percentage of participants who provided data compared to the total participants with at least one recorded event, shown in Figure 5[8].

Figure 4 illustrates the percentage of participants providing sensor data over time with respect to the total number of participants who contributed data. The most notable decrease occurs at AMRITA, where participation drops from 80% on the first day of data collection to less than 25% by the final day. On average, the dropout rate across sites is approximately 35%. The initial increase observed at JLU and IPICYT results from participants who joined the data collection after the start date.

For each sensor, Figure 5 shows the percentage of participants who contributed in relation to all participants with at least one recorded event (see also the "iLog data" column in Table 4). A lower percentage does not necessarily indicate low data quality. For instance, the number of participants reporting data from on-change sensors, such as airplane mode or music playback, is influenced by the actual usage of these smartphone features. Sensors collected at fixed intervals, like the accelerometer and application usage, tend to have a higher number of participants with data. In contrast, sensors like pressure capture data from a smaller subset of participants.

[8]A comprehensive description of the sensors and their collection frequency is available in Tables 15 and 16.

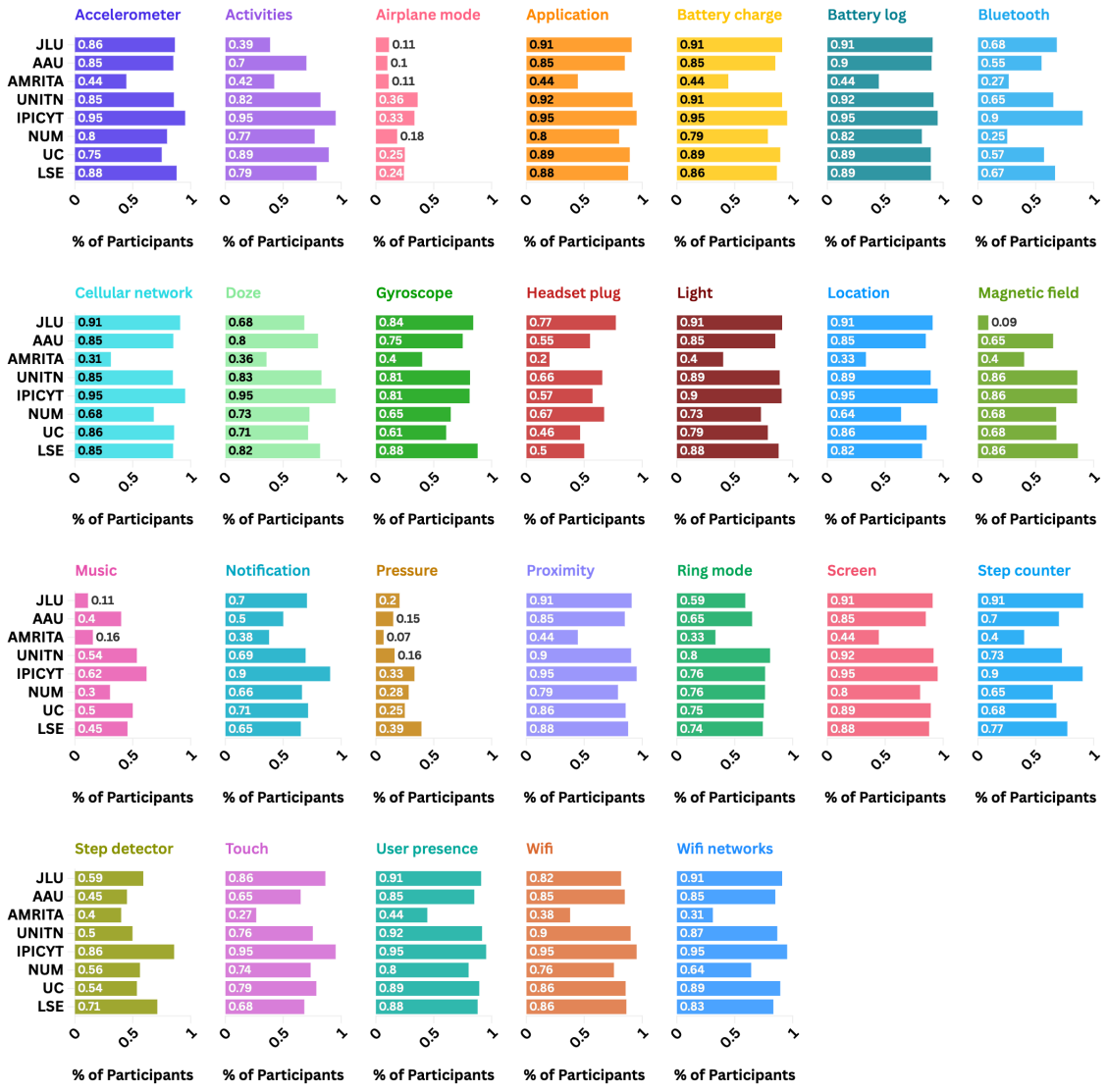
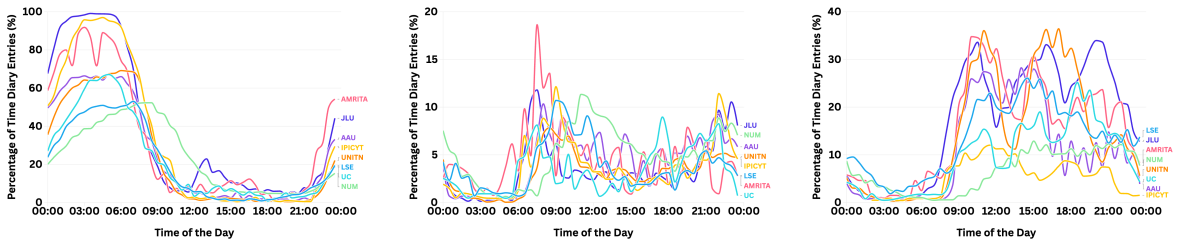


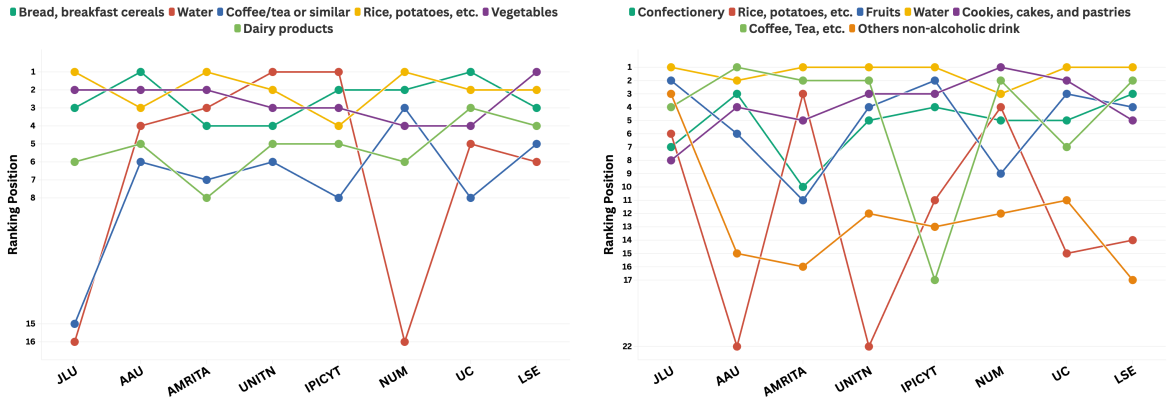
Fig. 5. Average percentage of participants that provided sensor data for each site.

In addition to dropouts, variations in sensor data provision may be attributed to application or device issues and the lack of support for specific hardware sensors on some devices. Participants could also disable any sensor directly from the app at any time for privacy considerations. For instance, a participant could disable GPS tracking if visiting a location they preferred to keep private (e.g., a place of worship or a hospital). Additionally, some sensors may have been disabled to conserve battery life. Together, these factors contribute to the observed instances of missing data. Hence, although this privacy feature resulted in some missing data, we believe this



(a) Sleeping (b) Eating (c) Studying

Fig. 6. Proportion of time diary reports during different hours of the day.



(a) Foods (b) Snacks

Fig. 7. Ranking comparison of the three most consumed foods and snacks.

design significantly contributed to attracting a higher number of participants who felt comfortable providing data over an entire month.

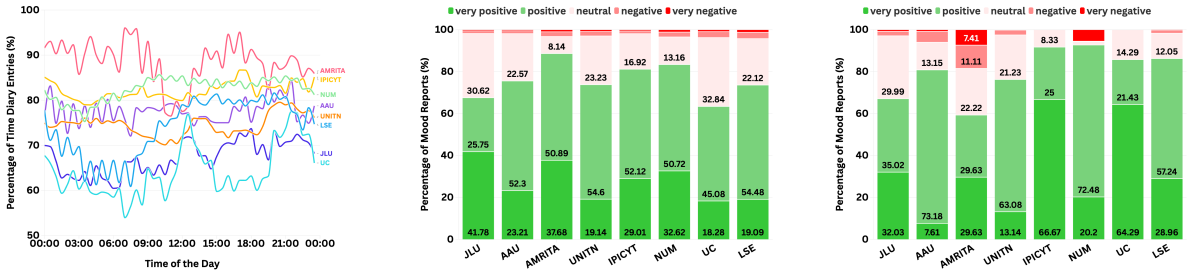
4.4 Diversity in Everyday Life

To highlight behavioral variations, we analyze the distribution of primary daily activities—sleeping, eating, and studying—across different pilot sites. For instance, Figure 6 displays these differences. Notably, Figure 6a shows distinct sleeping patterns between midnight and 7:00 AM, with many students from JLU resting in the afternoon, despite similar waking times to other pilot sites. Comparing these patterns with study hours (Figure 6c) reveals that study-related annotations account for over 30% of activities at 8:00 PM in JLU, compared to just 7% in IPICYT. It is also worth noting that sleep-related annotations were primarily captured using an app feature that allowed participants to define sleep time windows for several hours before returning to regular self-reporting in the morning.

The timing of eating activities also varies among universities. As shown in Figure 6b, AMRITA reports eating spikes around 8:00 AM and 10:00 PM, while NUM peaks around 11:00 AM and UC around 6:00 PM. Differences in eating behaviors are further highlighted in Figure 7, which ranks the top three foods and drinks reported at each site. For example, water is the most consumed drink at UNITN, while it ranks only 16th at NUM and JLU (refer to

Table 5. Ranking of each institution's five most used mobile applications.

	1st	2nd	3rd	4th	5th
JLU	WeChat	QQ - Tencent	Taobao	Bilibili	Google Messages
AAU	Youtube	Facebook Messenger	Spotify	Google Chrome	Google Maps
AMRITA	Android settings	Whatsapp	Google Chrome	Android settings	Youtube
UNITN	Whatsapp	Google Gmail	Youtube	Google Chrome	Android settings
IPICYT	Facebook	Whatsapp	Google quick search	Google Chrome	Android settings
NUM	Facebook Messenger	Facebook	Youtube	Android settings	Google Play Store
UC	Google Chrome	Whatsapp	Instagram	Android settings	Youtube
LSE	Whatsapp	Google Chrome	Android settings	Google Play Store	Google Gmail



(a) Percentage of positive mood reports over the day for each site.

(b) Mood distribution when the participants are at home.

(c) Mood distribution when the participants are at the university.

Fig. 8. Mood ratings distribution.

Figure 7a). Interestingly, water consistently appears among the top-reported drinks during snack breaks across all sites, as illustrated in Figure 7b.

Mood reporting adds an intriguing layer to this study, allowing researchers to investigate how mood ratings vary in similar contexts across different universities (what was captured is, actually, a facet of mood called Valence). This has been discussed in detail by Meegahapola et al. [80], who used this dataset for one of their studies). Figure 8a illustrates changes in positive mood over the day. While trends appear similar across locations, absolute percentages vary, with UC reporting the lowest mood levels and AMRITA the highest. Participants' personal rating criteria also influence these differences. Furthermore, mood differences are noticeable depending on location: distinct distributions of mood ratings are evident when participants are at home (e.g., apartments, gardens, or relatives' homes) compared to university spaces (e.g., classrooms and libraries), as shown in Figure 8b and Figure 8c.

Research questions regarding phone usage and time spent on activities such as internet browsing, chatting, or gaming can be examined by analyzing application usage. Table 5 ranks applications based on the number of participants using them. As expected, social networking and messaging apps are the most common across universities, with JLU students, for instance, favoring local applications such as WeChat.

This section offers only an initial glimpse into the potential analyses and research questions that can be pursued by combining and reshaping the data. A more comprehensive discussion can be found in Section 6.

Table 6. The size of each bundle of data in Gigabytes, with uncompressed sizes in brackets, and bundle names as they appear in the catalog are presented. Synchronic and diachronic interaction refers to questionnaires and time diaries, respectively.

Bundle name	**JLU**	**AAU**	**AMRITA**	**UNITN**	**IPICYT**	**NUM**	**UC**	**LSE**
App-usage	< 0.1 (0.3)	< 0.1 (0.2)	< 0.1 (0.2)	0.3 (2.8)	< 0.1 (0.4)	0.1 (2.3)	< 0.1 (0.3)	< 0.1 (0.7)
Connectivity	< 0.1 (2.5)	0.1 (5.7)	< 0.1 (0.4)	0.6 (26.6)	< 0.1 (1.2)	< 0.1 (3.7)	< 0.1 (2.8)	0.4 (16.8)
Device-usage	< 0.1 (0.4)	< 0.1 (0.3)	< 0.1 (0.2)	0.5 (4.4)	< 0.1 (0.2)	0.2 (2.3)	< 0.1 (0.4)	< 0.1 (0.9)
Environment	0.6 (5.6)	0.3 (2.6)	0.1 (1.4)	5.2 (44.5)	0.6 (6.4)	3.3 (35.2)	0.5 (5.4)	1.5 (12.1)
Motion	1.2 (12.4)	1.9 (18.0)	1.3 (13.2)	18.0 (174.8)	4.0 (36.8)	17.1 (161.3)	1.8 (19.9)	7.0 (61.3)
Position	0.1 (1.7)	0.6 (5.2)	0.6 (5.5)	12.7 (104.8)	1.3 (10.2)	6.8 (54.7)	0.8 (7.6)	2.9 (23.0)
Synchronic int.	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Diachronic int.	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Total	2.1	2.9	2.1	37.3	6.0	27.7	3.3	12.0

5 Dataset and Code Availability

Due to the variety of sources and the sensitive nature of the content, this dataset is made available in a secure environment that complies with GDPR regulations. The main entry point documentation can be accessed in the dataset catalog:

Link to the dataset catalog

https://livepeople.disi.unitn.it

The catalog simplifies the process of finding and selecting datasets of interest by providing metadata and materials to explore their features. The filter options allow to restrict the search on *DiversityOne* data. It also ensures privacy and copyright compliance, fully respecting the ownership of the authors who contributed to its collection. In addition to the catalog, the webpage available at https://datascientia.disi.unitn.it/projects/diversityone/ collects further information. The descriptions of downloadable datasets, associated documentation, and procedures for dataset requests are detailed below.

5.1 Downloadable Datasets and Bundles

The resources are organized and made available separately considering the full dataset size (approximately 94GB in Parquet format) and GDPR's minimization principle—which mandates that data must be adequate, limited, and relevant for analysis. Researchers can request access to basic datasets from a specific pilot site (e.g., time diaries collected at UNITN) or a combination of datasets from multiple pilot sites. To streamline dataset selection, we have created thematic bundles that group data commonly used together for main research purposes. For example, the motion bundle includes all motion sensor data relevant to activity recognition studies, while another bundle, combining questionnaires, time diaries, and location data, is tailored for studying social interactions. The catalog lists both datasets containing one single sensor and bundles. Table 6 reports the available bundles (Section 3.2 details the sensors they contain), and Appendix A.2 reports the complete list of sensors in each bundle. Any combination of bundles and single sensors can be downloaded. All datasets are provided in Parquet format[9], an efficient storage format with high compression. Access to the entire *DiversityOne* dataset is also available upon request. No source code needs to be made available for the release of this dataset. Future benchmarked

[9]Apache Parquet https://parquet.apache.org/

datasets derived from our raw dataset will be made available, respecting privacy and copyright, along with code for pre-processing and machine learning-based modeling.

5.2 Metadata and Documentation

Each single sensor dataset and bundle is accompanied by metadata and comprehensive documentation that outline content, size, format, and other relevant information. The catalog user can search the datasets through the metadata values such as the acronym, data collection location, and type of bundle or dataset. The metadata for each dataset and bundle includes: *(i)* A technical report detailing the data collection process; *(ii)* Dataset metadata and a codebook with summary statistics; *(iii)* Data collection information and links to related projects, including articles published using the dataset; *(iv)* Documentation and procedures for dataset requests.

The catalog enhances the findability, accessibility, and reusability of the dataset in several ways. First, it provides an efficient search method for specific data within a secure institutional environment. Second, it includes concise descriptions that guide users through the available resources, along with detailed information on each dataset, including variable values and labels. Lastly, the technical report and dataset descriptions enable users to analyze the dataset and replicate the data collection process for their own research needs.

5.3 Dataset Request Process

We outline the procedure to request the data. First, the user retrieves the identifiers of the bundle or single sensor data of interest from the metadata shown on the catalog for each sensor and bundle. Second, the metadata also includes a link to the request form, designed together with legal and privacy experts to comply with GDPR and privacy regulations. Interested users affiliated with a research institution can request bundles and datasets by listing the identifiers selected in the previous step and presenting a research proposal. Upon completing the form, users submit their request to the catalog manager using the email address specified in the metadata. Detailed eligibility criteria and required information are available on the catalog website. Third, once approved, users must sign a Terms and License Agreement. Key licensing terms include: *(i)* datasets are used exclusively for research purposes; *(ii)* redistribution of the datasets is prohibited; *(iii)* datasets cannot be publicly shared (e.g., on a website); and *(iv)* any attempt to reverse engineer any portion of the data or to re-identify the participants is strictly forbidden and could constitute unlawful processing of personal data. Finally, the catalog manager sends the instructions for downloading the dataset.

6 Discussion

6.1 Data Collection – Lessons Learned and Recommendations for the Future

In Section 3, we detailed our core design decisions for gathering cross-country *DiversityOne* dataset. Building on those design considerations, this section focuses on the broader lessons learned during implementation, compares our experiences with other local data collection practices, and discusses how to adapt these insights for future studies. We consolidate all recommendations into Table 7.

Collecting smartphone data across diverse cultural and regulatory contexts offered an invaluable opportunity to observe how different countries and research teams navigate ethical, logistical, and cultural considerations. Our collaboration revealed several common threads shared with local researchers as follows. First, *Cultural Adaptation:* while our study required adjustments in survey content, incentive structures, and communication styles, local researchers often faced similar challenges. For instance, local studies had to modify language and question framing to ensure sensitivity to cultural or religious norms. Our experiences echo these practices but highlight the importance of designing a standardized core protocol that can be flexibly adapted without sacrificing data comparability. Second, *Privacy and Regulatory Compliance:* many local research groups reported ongoing adaptations to meet international guidelines (e.g., GDPR) while conforming to national regulations. Our approach—having

Table 7. Summary Recommendations for Future Cross-Country Smartphone-Based Data Collection.

Recommendation #1: Core Protocol Design Develop a standardized core protocol that allows flexible local adaptations. Consult with local experts early to validate instruments, and pilot-test before large-scale deployment.
Recommendation #2: Cultural Sensitivity Involve local stakeholders to ensure cultural norms are respected. Conduct small-scale pilots to refine sensitive questions, and leverage cultural experts for precise translations.
Recommendation #3: Privacy and Ethics Create a modular privacy framework compliant with international (e.g., GDPR) and local regulations. Use adaptable templates and regularly consult legal experts to monitor changes.
Recommendation #4: Data Anonymization Adopt multi-level anonymization (e.g., personal identifiers, location data) to balance privacy with data utility. Offer multiple data versions (e.g., "RoundDown" GPS vs. "POI" GPS) to accommodate different re-identification risks.
Recommendation #5: Incentive Strategies Tailor incentives to local economic conditions and cultural preferences. Collaborate with local partners to choose monetary vs. non-monetary rewards for maximum engagement.
Recommendation #6: Technological Adaptation Design or select survey tools that can operate both online and offline. Plan for restricted services (e.g., Google) and ensure solutions work in diverse mobile ecosystems.
Recommendation #7: Participant Engagement Maintain transparent communication through notifications, reminders, and helpdesks. Offer support channels that address technical or procedural issues quickly without biasing responses.

a core GDPR-compliant framework and then tailoring it to meet local legal expectations—aligned with similar multi-country studies in the literature. This comparison highlights the benefits of centralized documentation that can be adjusted to local requirements. Third, *Technological Infrastructure and Offline Solutions:* the need to accommodate varying degrees of internet connectivity and restricted services was not unique to our study; many local researchers rely on hybrid models (online/offline) for data gathering. Our development of an "offline" version of the `iLog` app is congruent with other best practices in low-connectivity settings, such as storing data locally and syncing at scheduled intervals.

Hence, Table 7 brings together our main recommendations and the current reflections. Each item links directly to a stage of the study—from protocol design to data anonymization—and highlights how researchers can adapt these suggestions to their own contexts. By integrating our field experiences, we observed that successful cross-country data collection depends on a core set of adaptable protocols. We hope these consolidated recommendations will guide future multi-country studies in designing, implementing, and scaling smartphone-based data collection responsibly and effectively.

Table 8. A summary of previous studies that utilized the *DiversityOne* dataset is presented in the table. The acronyms used are as follows: Target/s Inferred refers to the primary aspect that was focused in the particular study; CS indicates that country-specific models were trained; MC denotes that multi-country models were trained by combining data from multiple countries; CA refers to the examination of a country-agnostic setting, focusing on the generalization performance of models across countries; and PER indicated that model personalization was examined; SL refers to supervised learning; FT/RT refers to transfer learning with fine-tuning or re-training for classic machine learning models to achieve personalization; and UDA refers to unsupervised domain adaptation.

	Venue and Year	**Target/s Inferred**	**CS**	**MC**	**CA**	**PER**	**Methods**
Bouton–Bessac et al. [12]	PervasiveHealth '22	Complex Daily Activities	*x*	✓	*x*	✓	SL, FT/RT
Meegahapola et al. [80]	IMWUT '23	Mood	✓	✓	✓	✓	SL, FT/RT
Assi & Meegahapola et al. [3]	CHI '23	Complex Daily Activities	✓	✓	✓	✓	SL, FT/RT
Kammoun et al. [63]	ICMI '23	Social Context when Eating	✓	✓	*x*	✓	SL, FT/RT
Girardini et al. [40]	EPJ DataScience '23	Daily routines and Covid	✓	*x*	*x*	✓	SL, FT/RT
Meegahapola et al. [82]	IMWUT '24	Mood, Social Context	✓	*x*	✓	*x*	SL, UDA
Mader & Meegahapola et al. [77]	CHI '24	Social Context	✓	✓	✓	✓	SL, FT/RT

6.2 Dataset analysis – Lessons Learned and Recommendations for the Future

The analysis of *DiversityOne* across various studies has revealed critical insights into the influence of local context, country-specific nuances, and multi-modal data streams on the effectiveness of mobile sensing models. Presented below are key lessons learned from these studies, each accompanied by a recommendation for future research.

6.2.1 Country-specific Models over Multi-country Models. Studies consistently demonstrate that models trained on country-specific data outperform those trained on multi-country datasets. For tasks such as mood inference [80] and complex daily activity recognition [3], country-specific models achieved higher accuracy, with AUROC scores ranging from 0.76 to 0.98 when using partially personalized, hybrid models. Conversely, generic models not adapted to specific countries often struggled when applied to new regions, failing to capture unique data distributions that reflect localized contexts. This gap emphasizes the critical importance of country-specific patterns and nuances in daily routines and sensor data, which vary widely and directly influence data interpretation. These findings are in line with the findings of Khawaja et al. [65].

Moreover, these insights shed light on limitations faced by previous multi-country sensing studies, such as those conducted by Servia-Rodriguez et al. [105]. By pooling data from multiple countries, these models averaged out region-specific patterns, reducing predictive power and capacity to capture unique, localized behaviors. For example, in the study by Meegahapola et al. [80], location and movement data were highly predictive of mood in certain countries but held minimal relevance in others, underscoring regional differences in daily routines and smartphone usage. This diversity indicates that a one-size-fits-all approach overlooks essential cultural and contextual factors crucial for accurate behavior inference, reinforcing the need for adaptable, localized models.

Computer vision models provide a valuable contrast here. Convolutional neural networks (CNNs) and vision transformers excel at capturing features in images—textures, edges, and spatial hierarchies—regardless of geographic origin. These models generalize effectively across multi-country visual datasets by learning local and global image features, excelling when regional content differences are less impactful on performance. However, smartphone time-series data captures complex, region-specific behavioral nuances that current time-series models struggle to interpret consistently across geographies. Developing models that can capture local nuances in behavioral data while generalizing across regions remains crucial for advancing global smartphone sensing applications.

6.2.2 Value of Hybrid Model Personalization. Typically, models developed for various prediction tasks are tested on a different population to evaluate their performance, reflecting real-world scenarios. These are referred to as population-level models. However, prior work [3, 80, 139] in ubiquitous computing has shown that such models often underperform, especially for highly subjective or complex inferences. Therefore, the population-level model should either be fine-tuned or re-trained for the target individual using a small portion of that individual's data (known as hybrid models), or, if sufficient data are available and it is computationally feasible, a model can be trained solely on the target individual's data (fully personalized models). Research using *DiversityOne* dataset has shown that even modest personalization using hybrid models enhances performance in behavioral inference tasks such as mood detection [80], complex activity recognition [3], and social context inference [77]. This aligns with findings from other datasets, such as those studied by Zhang et al. [139], which further validate the value of hybrid models. Hybrid models, which combine general population data with user-specific features, improve accuracy by balancing broad patterns with individual nuances. Fully personalized models can yield the highest accuracy but require extensive user-specific data, making them resource-intensive and less scalable. Hybrid models offer a practical compromise, delivering strong performance with fewer data demands per user, better accommodating individual behavioral variations.

For example, in mood detection, hybrid models incorporating individual app usage and movement data can capture unique variations, improving accuracy compared to purely generic models. This hybrid approach demonstrates that while general models detect broader trends, they often miss critical individual subtleties that partially personalized models can capture, providing a practical strategy for improving model relevance and precision.

6.2.3 Advancing Techniques in Multimodal Integration for Mobile Sensing. Integrating diverse data streams—such as activity, location, Bluetooth, app usage, device use, and WiFi— enhances behavioral inference by capturing a broader spectrum of contextual and activity patterns. Existing research using *DiversityOne* has largely relied on foundational multimodal techniques, often without leveraging advanced integration methods tailored to the high-dimensional, diverse data sources typical in mobile sensing. A recent study applying an unsupervised domain adaptation model, M3BAT [82], explored multimodal settings, demonstrating that domain adaptation can help mitigate distributional shifts across modalities, achieving up to a 12% improvement in AUC on specific classification tasks. However, much remains to be done to develop models capable of handling the complexities of multimodal data in mobile sensing, particularly when accounting for regional, cultural, and individual behaviors.

Smartphone sensor data is shaped by regional routines, cultural practices, and individual habits, requiring models to adapt dynamically to context-dependent patterns across data streams. Temporal transformers and adversarial training approaches tailored to multimodal data offer promising directions, though these are not yet widely applied in mobile sensing. Future research could benefit from multimodal architectures that use independent branches for each sensor modality, allowing each data type to be processed in context while adapting to shifting behavioral patterns. The next generation of multimodal models in mobile sensing should aim to improve upon initial efforts by adopting advanced architectures that facilitate robust, context-aware fusion of diverse data sources. Such models would enable richer, more generalized applications, supporting domains like health monitoring, behavior analysis, and activity recognition across diverse populations.

6.2.4 The Challenge of Distribution Shift. In cross-country smartphone sensing, distribution shifts in data impact the performance of models trained in one region when applied in another. Recent studies highlight that sensor data distributions vary widely between countries, introducing substantial challenges for model generalization. For instance, Meegahapola et al. [82] observed high Cohen's-d values, up to 1.0, across various sensing modalities such as location and step count. These large effect sizes reveal pronounced shifts in data distributions that hinder a model's ability to transfer effectively across geographic contexts. For example, GPS patterns that indicate

Table 9. Summary Recommendations for Future Research Based on the Dataset Analysis.

Recommendation #8: Country-Specific vs. Multi-Country Models Country-specific or regional models could be better for initial deployment in diverse regions, as compared to multi-country models. Tailoring models to capture unique behavioral patterns and cultural nuances in countries enhances accuracy without full personalization.
Recommendation #9: Hybrid Model Personalization Adopt a hybrid strategy for model personalization by combining general population data with user-specific inputs to capture both local and individual behavioral nuances. This approach achieves a balance between broad generalization and user-specific relevance, enhancing model performance while managing data demands effectively.
Recommendation #10: Advanced Multimodal Integration Develop advanced multimodal integration techniques for behavioral models in mobile sensing. Multi-branch architectures may provide an effective way forward by facilitating context-aware fusion of diverse data sources.
Recommendation #11: Addressing Distribution Shifts Implement domain adaptation techniques that account for cross-country distribution shifts in mobile sensing. Consider using multi-branch adversarial training to tailor model responses to region-specific patterns, improving generalization across diverse data distributions.
Recommendation #12: Label Quality Prioritize objective labels for tasks where label quality critically impacts model performance. For subjective data, explore advanced preprocessing techniques such as outlier regularization, attention mechanisms, and weighted averaging to reduce noise, thereby improving label consistency and model accuracy.

social activities in one country may reflect commuting in another, necessitating region-specific adaptations to accurately interpret similar data.

While substantial progress has been made in addressing distribution shifts in computer vision and natural language processing [37, 124], domain adaptation in multimodal mobile sensing is still an emerging field [21, 82, 130]. As mentioned earlier, the M3BAT framework by Meegahapola et al. employed a multi-branch adversarial training approach to tackle distribution shifts in multimodal sensor data, achieving performance gains of up to 12% in AUC for classification tasks and a 0.13 MAE reduction in regression tasks. By creating separate branches for features with varying distribution shifts, M3BAT tailored adaptation strategies to each modality, enhancing model robustness. Moreover, Wu et al. [130] also applied domain adaptation for multimodal data, however, without considering the multimodality. Xu et al. [131] used a domain generalization algorithm, again without considering the multimodality. However, all the above methods are still nascent, addressing only a subset of multimodal data shifts.

6.2.5 Effect of Label Quality on Model Performance. The quality of labels has been shown to profoundly influence model performance after transer learning, especially in mobile sensing applications where labels can vary widely in terms of subjectivity [9, 130, 137]. A study using *DiversityOne* [82] revealed that models trained on subjective labels, such as self-reported mood, tend to exhibit lower consistency and reliability compared to models trained

on more objective labels, such as social context. Although both labels are silver standard [10] because they are self-reported, social context is less subjective and mood is more subjective, leading to various behaviors for models. This difference suggests that, where possible, researchers should aim to use objective and gold standard labels [11] to enhance model reliability and accuracy. However, for tasks that inherently rely on subjective input—like mood, depression, stress or perceived energy levels—it is crucial to develop preprocessing methods that can transform subjective data into more standardized, reliable labels. Techniques such as regularizing outliers, employing weighted averages for conflicting reports, or integrating machine learning methods like attention mechanisms to focus on higher confidence data points can help mitigate the impact of noise in subjective labels.

By integrating these lessons with advanced modeling techniques, researchers can enhance the robustness and scalability of mobile sensing applications in diverse cultural and geographic contexts. These recommendations aim to guide future work toward more accurate, context-aware models, ultimately supporting broader real-world adoption in areas such as health monitoring, behavior prediction, and social context inference.

6.3 Limitations

Although we observed a high response rate for questionnaires at many pilot sites, participation in the intensive longitudinal survey varied between countries. For example, UNITN and NUM had exceptionally high participation numbers, while AMRITA faced a notably high dropout rate, with some cases reporting completion rates below 40%. Despite achieving strong results in certain areas of data collection during the challenges of the COVID-19 pandemic, we have identified several critical reflections on the study's progress to better contextualize the disparities observed in the collected data.

First, although the methodology included a local adaptation phase, communication challenges arose during the design and preparation stages. These issues likely stemmed from the involvement of experts outside the social sciences and/or ubiquitous computing, which may have disadvantaged those less familiar with questionnaire design or intensive longitudinal surveys [56]. To foster a more inclusive and diversity-aware approach in future data collections involving interdisciplinary teams, it would be beneficial not only to share and adapt data collection materials but also to provide methodological guidance that informs design strategies across fields, as mentioned in Section 6.1.

Second, it is evident that the recruitment strategy yielded varying results across pilot sites. The decision to use a uniform approach—sending an email invitation to all students at each institute—was based on methodological considerations regarding sampling strategies, data standardization, and comparability. However, cultural differences and distinct socialization processes led to divergent response rates. For example, while some sites recorded over 5,000 responses to the invitation questionnaire, others received very few. This suggests that future communication campaigns should better account for cultural diversity by utilizing a range of channels and enriching promotional materials, possibly incorporating videos, interviews, or participant testimonials to engage broader audiences.

Third, although sensor data collection remained consistent across countries, participants outside of AMRITA submitted a high volume of sensor data meeting state-of-the-art standards (see [3]). Nevertheless, high dropout rates persisted throughout data collection, a common issue in intensive longitudinal surveys. Various factors

[10] In machine learning, the term "silver-standard" refers to labels or annotations that, due to uncertainty, bias, or other sources of noise, offer lower accuracy or reliability than "gold-standard" labels. Despite these limitations, silver-standard labels remain valuable and are commonly employed in inference tasks. They may be derived from a variety of less rigorous methods, including automated algorithms, self-reports, surrogate measures, or indirect observations. [30, 82, 130]

[11] In machine learning, "gold-standard" refers to the most accurate and reliable ground truth labels or annotations. They are often derived from highly trusted methods—such as expert manual annotations, precise measurements, or comprehensive and well-established criteria. However, producing these gold-standard labels can be challenging, time-consuming, and costly, as it typically requires specialized expertise, rigorous data collection, and meticulous validation [30, 82, 130].

contribute to this problem, many of which relate to incentives (discussed further below) and respondent burden. For instance, delays in providing support (especially if app-related issues arose) and communication challenges likely impacted participants' continued engagement. Although break options were designed to alleviate respondent burden (outlined in Section 3.2), these were not always utilized, perhaps due to expectations for plug-and-play apps or an overwhelming amount of materials, including privacy documentation. To address this, a single repository or website where participants could easily access all relevant information, supported by tutorials and videos, would likely improve usability and engagement, which is to be explored in future work.

Fourth, disparities in the proposed incentives appear to have influenced data collection success in some contexts. This finding aligns with Singer's [109] assertion that fixed monetary compensation generally outperforms gifts or prize draws in motivating participation unless no incentive is provided. Therefore, it is advisable to offer fixed monetary incentives or a set of equivalent alternatives, such as university credits combined with vouchers for students. Additionally, exploring incentives such as digital badges, personalized messages, or gamification elements may be worth considering to sustain long-term engagement.

Finally, the COVID-19 pandemic presented an unforeseen circumstance that invalidated some initial design choices, complicating efforts to ensure ecological validity. In response, we adjusted the survey as needed and conducted a follow-up study at the end of 2021 to examine the pandemic's impact on student's lives through the lens of smartphone sensors, providing the possibility of comparative insights for post-pandemic contexts in future studies, even though the later dataset is not released as part of *DiversityOne*. This second will be released soon, hopefully within 2024. [40] studies the effects of Covid-19 by comparing a subset of DiversityOne, limited to UniTN, with a dataset collected earlier on, in 1918, on the same population.

7 Conclusion

Human behavior varies not only by time of day but also across individuals, countries, and cultures, introducing layers of complexity that most existing smartphone sensing and self-report datasets fail to capture. Current datasets are often limited to specific countries in the Global North, small participant samples, or a narrow set of sensors, which restricts their ability to reflect cross-country behavioral variations and nuanced sensor data patterns. To address this gap, we introduce *DiversityOne*, a dataset designed to combine passive smartphone sensor data with self-reported labels at multiple granularities. *DiversityOne* is built on a robust interdisciplinary methodology that incorporates standards from computer science, behavioral sciences, social practice theory, and intensive longitudinal surveys via time diaries. This approach supports machine learning models in inferring various everyday aspects, including mood, activities, social context, and eating habits. Data were collected across eight countries from 782 college students over four weeks, encompassing 26 smartphone sensors and more than 350,000 self-reports, positioning *DiversityOne* as one of the largest and most diverse datasets of its kind, that is publicly available. Additionally, *DiversityOne* provides raw sensor data, offering researchers flexibility in processing and usage for diverse analytical purposes. As a significant contribution of this paper, *DiversityOne* is publicly released with a comprehensive description of the study design and data collection process, highlighting key insights, lessons learned, and future research directions toward diversity- and privacy-aware studies. We hope this dataset will serve as a valuable resource for the community, enabling studies across multiple disciplines, including ubiquitous computing, human-computer interaction, and machine learning.

8 Author contributions

The names order is by the contribution of the institution and, inside each institution, by the contribution of the individuals. As such, the order of names does not necessarily reflect the importance of the contribution of single individuals. The roles of the authors are presented by their initials as follows:

- *Study management*: F.G., I.B., A.D.G., M.B.;

- *Study design*: F.G., I.B., G.G., A.D.G., M.B., R.C.A., L.M., D.G.P.;
- *Technical support*: M.R., M.B., R.C.A., L.J.M.;
- *Data Collection*: M.B., R.C.A., M.R., A.D.G, P.K., A.G., A.C., G.G., S.S., M.B., L.C., A.H., H.X., D.S., S.D., C.N., S.R.C., A.R.M.;
- *Data Preparation and correction*: A.B., R.A.A., I.K., R.C.A., I.B., M.B., D.G.P., L.M.;
- *Data analysis*: A.B., L.M.;
- *Writing the manuscript*: M.B., A.B., L.M., D.G.P., F.G..

Acknowledgments

This research has received funding from the European Union's Horizon 2020 FET Proactive project "WeNet - The Internet of us", grant agreement No. 823783 (https://doi.org/10.3030/823783). We deeply thank all the volunteers across the world for their participation in the study. We thank the anonymous reviewers for their valuable feedback. We acknowledge the use of ChatGPT and Grammarly as tools for grammar refinement. University of Trento, Jilin University, National University of Mongolia and Universidad Católica "Nuestra Señora de la Asunción" have participated to the work described in this paper as part of the DataScientia initiative (https://datascientia.eu/).

References

[1] 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). 88 pages. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

[2] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514.

[3] Karim Assi, Lakmal Meegahapola, William Droz, Peter Kun, Amalia De Götzen, Miriam Bidoglia, Sally Stares, George Gaskell, Altangerel Chagnaa, Amarsanaa Ganbold, et al. 2023. Complex daily activities, country-level diversity, and smartphone sensing: A study in Denmark, Italy, Mongolia, Paraguay, and UK. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–23.

[4] Sangwon Bae, Denzil Ferreira, Brian Suffoletto, Juan C Puyana, Ryan Kurtz, Tammy Chung, and Anind K Dey. 2017. Detecting drinking episodes in young adults using smartphone-based sensors. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 2 (2017), 1–36.

[5] Steffen Bickel, Michael Brückner, and Tobias Scheffer. 2009. Discriminative learning under covariate shift. *Journal of Machine Learning Research* 10, 9 (2009).

[6] Joan-Isaac Biel, Nathalie Martin, David Labbe, and Daniel Gatica-Perez. 2018. Bites 'n'bits: Inferring eating behavior from contextual mobile data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–33.

[7] Ivano Bison, Haonan Zhao, and Fausto Giunchiglia. 2024. What Impacts the Quality of the User Answers when Asked about the Current Context? *arXiv preprint arXiv:2405.04054* (2024).

[8] Madeline Renee Bombardi-Bragg. 2017. *Exploring App Users' Experiences with App Notifications.* Master's thesis. Colorado State University.

[9] Andrea Bontempelli, Stefano Teso, Fausto Giunchiglia, and Andrea Passerini. 2020. Learning in the Wild with Incremental Skeptical Gaussian Processes. In *IJCAI*.

[10] Pierre Bourdieu. 1977. *Outline of a Theory of Practice.* Number 16. Cambridge University Press.

[11] Pierre Bourdieu. 1990. *The logic of practice.* Stanford university press.

[12] Emma Bouton-Bessac, Lakmal Meegahapola, and Daniel Gatica-Perez. 2022. Your Day in Your Pocket: Complex Activity Recognition from Smartphone Accelerometers. In *International Conference on Pervasive Computing Technologies for Healthcare*. Springer, 247–258.

[13] Isabel Briggs-Myers and Peter B Myers. 1995. Gifts differing: Understanding personality type. (1995).

[14] Teodora Sandra Buda, Mohammed Khwaja, and Aleksandar Matic. 2021. Outliers in smartphone sensor data reveal outliers in daily happiness. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–19.

[15] Mattia Giovanni Campana and Franca Delmastro. 2021. ContextLabeler dataset: Physical and virtual sensors data collected from smartphone usage in-the-wild. *Data in brief* 37 (2021), 107164.

[16] Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 1293–1304.

[17] Rafael Capurro. 2005. Privacy. An intercultural perspective. *Ethics and information technology* 7 (2005), 37–47.

[18] Andrew D Carlo, Reza Hosseini Ghomi, Brenna N Renn, and Patricia A Areán. 2019. By the numbers: ratings and utilization of behavioral health mobile applications. *NPJ digital medicine* 2, 1 (2019), 54.

[19] Karen A Cerulo. 2010. Mining the intersections of cognitive sociology and neuroscience. *Poetics* 38, 2 (2010), 115–132.

[20] Eun-Seok Cha, Kevin H Kim, and Judith A Erlen. 2007. Translation of scales in cross-cultural research: issues and techniques. *Journal of advanced nursing* 58, 4 (2007), 386–395.

[21] Youngjae Chang, Akhil Mathur, Anton Isopoussu, Junehwa Song, and Fahim Kawsar. 2020. A systematic study of unsupervised domain adaptation for robust human-activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–30.

[22] Cecilia Cheng, Yan-ching Lau, Linus Chan, and Jeremy W Luk. 2021. Prevalence of social media addiction across 32 nations: Meta-analysis with subgroup analysis of classification schemes and cultural values. *Addictive behaviors* 117 (2021), 106845.

[23] Benjamin Y Cheung, Kosuke Takemura, Christine Ou, Anne Gale, and Steven J Heine. 2021. Considering cross-cultural differences in sleep duration between Japanese and Canadian university students. *PLoS One* 16, 4 (2021), e0250671.

[24] Seungeun Chung, Chi Yoon Jeong, Jeong Mook Lim, Jiyoun Lim, Kyoung Ju Noh, Gague Kim, and Hyuntae Jeong. 2022. Real-world multimodal lifelog dataset for human behavior study. *ETRI Journal* 44, 3 (2022), 426–437.

[25] Mihaly Csikszentmihalyi, Mihaly Csikszentmihalyi, and Reed Larson. 2014. Validity and reliability of the experience-sampling method. *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi* (2014), 35–54.

[26] Vedant Das Swain, Koustuv Saha, Hemang Rajvanshy, Anusha Sirigiri, Julie M Gregg, Suwen Lin, Gonzalo J Martinez, Stephen M Mattingly, Shayan Mirjafari, Raghu Mulukutla, et al. 2019. A multisensor person-centered approach to understand the role of daily activities in job performance with organizational personas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–27.

[27] Edith D de Leeuw and Peter Lugtig. 2014. Dropouts in longitudinal surveys. *Wiley STATSREF: Statistics reference online* (2014), 1–6.

[28] M Brent Donnellan, Frederick L Oswald, Brendan M Baird, and Richard E Lucas. 2006. The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment* 18, 2 (2006), 192.

[29] Hubert Dreyfus. 1991. Being-in-the-World: A Commentary on Heidegger's Being and Time. (1991).

[30] Amanda Dy, Ngoc-Nhu Jennifer Nguyen, Seyed Hossein Mirjahanmardir, Dimitrios Androutsos, Melanie Dawe, Anthony Fyles, Wei Shi, Fei-Fei Liu, Susan Done, and April Khademi. 2023. Domain Adaptation using Silver Standard Labels for Ki-67 Scoring in Digital Pathology A Step Closer to Widescale Deployment. In *Medical Imaging with Deep Learning*.

[31] Gudrun Eisele, Hugo Vachon, Ginette Lafit, Peter Kuppens, Marlies Houben, Inez Myin-Germeys, and Wolfgang Viechtbauer. 2022. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment* 29, 2 (2022), 136–151.

[32] Jon D Elhai, Mojisola F Tiamiyu, Justin W Weeks, Jason C Levine, Kristina J Picard, and Brian J Hall. 2018. Depression and emotion regulation predict objective smartphone use measured over one week. *Personality and individual differences* 133 (2018), 21–28.

[33] Charles Ess. 2005. " Lost in translation"?: Intercultural dialogues on privacy and information ethics (Introduction to special issue on privacy and data privacy protection in Asia). *Ethics and Information Technology* 7, 1 (2005), 1.

[34] Nancy Foner, Jan Willem Duyvendak, and Philip Kasinitz. 2020. Introduction: Super-diversity in everyday life. In *Super-Diversity in Everyday Life*. Routledge, 1–16.

[35] Daniel Garcia-Gonzalez, Daniel Rivero, Enrique Fernandez-Blanco, and Miguel R. Luaces. 2020. A Public Domain Dataset for Real-Life Human Activity Recognition Using Smartphone Sensors. *SENSORS* 20, 8 (2020). doi:10.3390/s20082200

[36] Daniel Gatica-Perez, Joan-Isaac Biel, David Labbe, and Nathalie Martin. 2019. Discovering Eating Routines in Context with a Smartphone App. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) *(UbiComp/ISWC '19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 422–429. https://doi.org/10.1145/3341162.3349297

[37] Chunjiang Ge, Rui Huang, Mixue Xie, Zihang Lai, Shiji Song, Shuang Li, and Gao Huang. 2023. Domain adaptation via prompt learning. *IEEE Transactions on Neural Networks and Learning Systems* (2023).

[38] Anthony Giddens. 1979. *Central problems in social theory: Action, structure, and contradiction in social analysis*. Vol. 241. Univ of California Press.

[39] Anthony Giddens. 1984. *The Constitution of Society*. Cambridge: Polity Press.

[40] Nicolò Alessandro Girardini, Simone Centellegher, Andrea Passerini, Ivano Bison, Fausto Giunchiglia, and Bruno Lepri. 2023. Adaptation of student behavioural routines during Covid-19: a multimodal approach. *EPJ Data Science* 12, 1 (2023), 55.

[41] Fausto Giunchiglia, Enrico Bignotti, and Mattia Zeni. 2017. Human-Like Context Sensing for Robot Surveillance. *International Journa"l of Semantic Computing* 12, 01 (2017), 129–148.

[42] Fausto Giunchiglia, Enrico Bignotti, and Mattia Zeni. 2017. Personal context modelling and annotation. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 117–122.

[43] Fausto Giunchiglia, Xiaoyue Li, Matteo Busso, and Marcelo Rodas-Britez. 2022. A context model for personal data streams. In *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*. Springer, 37–44.

[44] Fausto Giunchiglia, Mattia Zeni, and Enrico Big. 2018. Personal Context Recognition via Reliable Human-Machine Collaboration. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, IEEE, 379–384. doi:10.1109/PERCOMW.2018.8480307

[45] Fausto Giunchiglia, Mattia Zeni, Enrico Bignotti, and Wanyi Zhang. 2018. Assessing Annotation Consistency in the Wild. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, IEEE, 561–566.

[46] Fausto Giunchiglia, Mattia Zeni, Elisa Gobbi, Enrico Bignotti, and Ivano Bison. 2017. Mobile social media and academic performance. In *International conference on social informatics*. Springer, Cham, Springer, Cham, 3–13.

[47] Erving Goffman. 1975. *Asylums: Essays on the social situation of mental patients and other inmates*. Harmondsworth: Penguin.

[48] Valdiney V Gouveia, Taciano L Milfont, and Valeschka M Guerra. 2014. Functional theory of human values: Testing its content and structure hypotheses. *Personality and Individual Differences* 60 (2014), 41–47.

[49] Andreas Grammenos, Cecilia Mascolo, and Jon Crowcroft. 2018. You are sensing, but are you biased? a user unaided sensor calibration approach for mobile sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–26.

[50] Sonia Gsir. 2014. *Social interactions between immigrants and host country populations: a country-of-origin perspective*. Position Paper 2014/02. Migration Policy Centre, INTERACT. https://cadmus.eui.eu/handle/1814/31243

[51] Yonas Deressa Guracho, Susan J Thomas, and Khin Than Win. 2023. Smartphone application use patterns for mental health disorders: a systematic literature review and meta-analysis. *International Journal of Medical Informatics* 179 (2023), 105217.

[52] Kristin Gustavson, Tilmann von Soest, Evalill Karevold, and Espen Røysamb. 2012. Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. *BMC public health* 12 (2012), 1–11.

[53] Gabriella M Harari, Sandrine R Müller, and Samuel D Gosling. 2020. 19 Naturalistic Assessment of Situations Using Mobile Sensing Methods. *The Oxford handbook of psychological situations* (2020), 299.

[54] Harish Haresamudram, Irfan Essa, and Thomas Plötz. 2021. Contrastive predictive coding for human activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–26.

[55] Mattias Hellgren. 2014. Extracting more knowledge from time diaries? *Social Indicators Research* 119, 3 (2014), 1517–1534.

[56] Paula Helm, Amalia de Götzen, Luca Cernuzzi, Alethia Hume, Shyam Diwakar, Salvador Ruiz Correa, and Daniel Gatica-Perez. 2023. Diversity and neocolonialism in Big Data research: Avoiding extractivism while struggling with paternalism. *Big Data & Society* 10, 2 (2023), 20539517231206802.

[57] Tania Karina Hernández-Heredia, Cesar Fabián Reyes-Manzano, Diego Alonso Flores-Hernández, Gabriel Ramos-Fernández, and Lev Guzmán-Vargas. 2024. Proximity Sensor for Measuring Social Interaction in a School Environment. *Sensors* 24, 15 (2024), 4822.

[58] Georg Holtz. 2014. Generating social practices. *Journal of Artificial Societies and Social Simulation* 17, 1 (2014), 17.

[59] ICLEI - Local Governments for Sustainability. 2024. Urban transport and climate change. https://sustainablemobility.iclei.org/urban-transport-and-climate-change/ Accessed on November 01, 2024.

[60] Masumi Iida, Patrick E Shrout, Jean-Philippe Laurenceau, and Niall Bolger. 2023. Using intensive longitudinal methods in psychological research. (2023).

[61] Loes HC Janssen, Bart Verkuil, Andre Nedderhoff, Lisanne AEM van Houtum, Mirjam CM Wever, and Bernet M Elzinga. 2024. Tracking real-time proximity in daily life: A new tool to examine social interactions. *Behavior Research Methods* (2024), 1–16.

[62] Carl G Jung. 1971. Psychological types. Collected works of CG Jung. *Tr. HG Baynes. Rev. RFC Hull. Princeton: Princeton UP* (1971).

[63] Nathan Kammoun, Lakmal Meegahapola, and Daniel Gatica-Perez. 2023. Understanding the Social Context of Eating with Multimodal Smartphone Sensing: The Role of Country Diversity. In *Proceedings of the 25th International Conference on Multimodal Interaction*. 604–612.

[64] Maryam Khalid, Elizabeth B Klerman, Andrew W McHill, Andrew JK Phillips, and Akane Sano. 2024. SleepNet: Attention-Enhanced Robust Sleep Prediction using Dynamic Social Networks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 1 (2024), 1–34.

[65] Mohammed Khwaja, Sumer S Vaid, Sara Zannone, Gabriella M Harari, A Aldo Faisal, and Aleksandar Matic. 2019. Modeling personality vs. modeling personalidad: In-the-wild mobile data analysis in five countries suggests cultural impact on personality models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.

[66] Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, et al. 2021. Wilds: A benchmark of in-the-wild distribution shifts. In *International conference on machine learning*. PMLR, 5637–5664.

[67] Frauke Kreuter, Georg-Christoph Haas, Florian Keusch, Sebastian Bähr, and Mark Trappmann. 2020. Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review* 38, 5 (2020), 533–549.

[68] John Krumm and Dany Rouhana. 2013. Placer: semantic place labels from diary data. In *Proceedings of ACM - UbiComp*. 163–172.

[69] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. 2010. A survey of mobile phone sensing. *IEEE Communications magazine* 48, 9 (2010), 140–150.

[70] Matias Laporte, Martin Gjoreski, and Marc Langheinrich. 2023. LAUREATE: A Dataset for Supporting Research in Affective Computing and Human Memory Augmentation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (2023), 1–41.

[71] J. K. Laurila, Daniel Gatica-Perez, Jan Blom, Olivier Bornet, Trinh-Minh-Tri Do, O. Dousse, Julien Eberle, and Markus Miettinen. 2013. From Big Smartphone Data to Worldwide Research: The Mobile Data Challenge. *Pervasive and Mobile Computing* 9, 6 (Dec. 2013), 752–771.

[72] Xiaoyue Li, Marcelo Rodas-Britez, Matteo Busso, and Fausto Giunchiglia. 2022. Representing habits as streams of situational contexts. In *International Conference on Advanced Information Systems Engineering*. Springer, 86–92.

[73] LimeSurvey Project Team / Carsten Schmitz. 2012. *LimeSurvey: An Open Source survey tool.* LimeSurvey Project, Hamburg, Germany. http://www.limesurvey.org

[74] Yanqing Lin, Yong Liu, Wenjie Fan, Virpi Kristiina Tuunainen, and Shengli Deng. 2021. Revisiting the relationship between smartphone use and academic performance: A large-scale study. *Computers in Human Behavior* 122 (2021), 106835.

[75] Omar Lizardo, Robert Mowry, Brandon Sepulvado, Dustin S Stoltz, Marshall A Taylor, Justin Van Ness, and Michael Wood. 2016. What are dual process models? Implications for cultural analysis in sociology. *Sociological Theory* 34, 4 (2016), 287–310.

[76] Olatz Lopez-Fernandez, Daria J Kuss, Lucia Romo, Yannick Morvan, Laurence Kern, Pierluigi Graziani, Amélie Rousseau, Hans-Jürgen Rumpf, Anja Bischof, Ann-Kathrin Gässler, et al. 2017. Self-reported dependence on mobile phones in young adults: A European cross-cultural empirical survey. *Journal of behavioral addictions* 6, 2 (2017), 168–177.

[77] Aurel Ruben Mäder, Lakmal Meegahapola, and Daniel Gatica-Perez. 2024. Learning About Social Context from Smartphone Data: Generalization Across Countries and Daily Life Moments. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[78] David Dennis Lee Mascarenas. 2016. A Jungian based framework for Artificial Personality Synthesis.. In *EMPIRE@ RecSys*. 48–54.

[79] Daniel McNeish, David P Mackinnon, Lisa A Marsch, and Russell A Poldrack. 2021. Measurement in intensive longitudinal data. *Structural equation modeling: a multidisciplinary journal* 28, 5 (2021), 807–822.

[80] Lakmal Meegahapola, William Droz, Peter Kun, Amalia De Götzen, Chaitanya Nutakki, Shyam Diwakar, Salvador Ruiz Correa, Donglei Song, Hao Xu, Miriam Bidoglia, et al. 2023. Generalization and personalization of mobile sensing-based mood inference models: an analysis of college students in eight countries. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 6, 4 (2023), 1–32.

[81] Lakmal Meegahapola and Daniel Gatica-Perez. 2021. Smartphone Sensing for the Well-Being of Young Adults: A Review. *IEEE Access* 9 (2021), 3374–3399. doi:10.1109/ACCESS.2020.3045935

[82] Lakmal Meegahapola, Hamza Hassoune, and Daniel Gatica-Perez. 2024. M3BAT: Unsupervised Domain Adaptation for Multimodal Mobile Sensing with Multi-Branch Adversarial Training. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–30.

[83] Andrea Mercado, Alethia Hume, Ivanno Bison, Fausto Giunchiglia, Amarsanaa Ganbold, and Luca Cernuzzi. 2023. Social interactions mediated by the Internet and the Big-Five: a cross-country analysis. *arXiv preprint arXiv:2309.10681* (2023).

[84] Batja Mesquita and Nico H Frijda. 1992. Cultural variations in emotions: a review. *Psychological bulletin* 112, 2 (1992), 179.

[85] L. Michael, I. Bison, M. Busso, L. Cernuzzi, A. de Götzen, S. Diwakar, K. Gal, S. Ganbold, G. Gaskell, D. Gatica-Perez, J. Heesen, D. Miorandi, N. Osman, S. Ruiz-Correa, L. Schelenz, A. Segal, C. Sierra, H. Xu, and F. Giunchiglia. 2025. Towards Open Diversity-Aware Social Interactions. *ArXiv – under submission* (2025).

[86] I. Myin-Germeys and P. Kuppens. 2022. *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies (2nd ed.).* Center for Research on Experience Sampling and Ambulatory Methods Leuven.

[87] Will Orr and Edward B Kang. 2024. AI as a Sport: On the Competitive Epistemologies of Benchmarking. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1875–1884.

[88] Laura Ozella, Daniela Paolotti, Guilherme Lichand, Jorge P Rodríguez, Simon Haenni, John Phuka, Onicio B Leal-Neto, and Ciro Cattuto. 2021. Using wearable proximity sensors to characterize social contact patterns in a village of rural Malawi. *EPJ Data Science* 10, 1 (2021), 46.

[89] Giancarlos Parady, Andreas Frei, Matthias Kowald, Sergio Guidon, Michael Wicki, Pauline van den Berg, Juan-Antonio Carrasco, Theo Arentze, Harry Timmermans, Barry Wellman, et al. 2021. A comparative study of social interaction frequencies among social network members in five countries. *Journal of Transport Geography* 90 (2021), 102934.

[90] Le Vy Phan, Nick Modersitzki, Kim Karen Gloystein, and Sandrine Müller. 2022. Mobile Sensing Around the Globe: Considerations for Cross-Cultural Research. (2022).

[91] Jacob Poushter, Caldwell Bishop, and Hanyu Chwe. 2018. Social media use continues to rise in developing countries but plateaus across developed ones. *Pew research center* 22 (2018), 2–19.

[92] Xin Qin, Yiqiang Chen, Jindong Wang, and Chaohui Yu. 2019. Cross-dataset activity recognition via adaptive spatial-temporal transfer learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–25.

[93] Inioluwa Deborah Raji, Emily M Bender, Amandalynne Paullada, Emily Denton, and Alex Hanna. 2021. AI and the everything in the whole wide world benchmark. *arXiv preprint arXiv:2111.15366* (2021).

[94] Andreas Reckwitz. 2002. Toward a theory of social practices: A development in culturalist theorizing. *European journal of social theory* 5, 2 (2002), 243–263.

[95] Sindre Rolstad, John Adler, and Anna Rydén. 2011. Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis. *Value in Health* 14, 8 (2011), 1101–1108. doi:10.1016/j.jval.2011.06.003

[96] Inge Røpke. 2009. Theories of practice—New inspiration for ecological economic studies on consumption. *Ecological economics* 68, 10 (2009), 2490–2497.

[97] Jason D Runyan, Timothy A Steenbergh, Charles Bainbridge, Douglas A Daugherty, Lorne Oke, and Brian N Fry. 2013. A smartphone ecological momentary assessment/intervention "app" for collecting real-time data and promoting self-awareness. *PloS one* 8, 8 (2013), e71325.

[98] Darshan Santani, Florian Labhart, Sara Landolt, Emmanuel Kuntsche, Daniel Gatica-Perez, et al. 2018. DrinkSense: Characterizing youth drinking behavior using smartphones. *IEEE Transactions on Mobile Computing* 17, 10 (2018), 2279–2292.

[99] Theodore Schatzki and Karin Knorr Cetina. 2001. The Practice Turn in Contemporary Theory. (2001).

[100] Philip Schmidt, Attila Reiss, Robert Dürichen, and Kristof Van Laerhoven. 2019. Wearable-based affect recognition—A review. *Sensors* 19, 19 (2019), 4079.

[101] Shalom H Schwartz. 1994. Are there universal aspects in the structure and contents of human values? *Journal of social issues* 50, 4 (1994), 19–45.

[102] Shalom H Schwartz, Gila Melech, Arielle Lehmann, Steven Burgess, Mari Harris, and Vicki Owens. 2001. Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of cross-cultural psychology* 32, 5 (2001), 519–542.

[103] Nicu Sebe, Ira Cohen, and Thomas S Huang. 2005. Multimodal emotion recognition. In *Handbook of pattern recognition and computer vision*. World Scientific, 387–409.

[104] Vedran Sekara and Sune Lehmann. 2014. The strength of friendship ties in proximity sensor data. *PloS one* 9, 7 (2014), e100915.

[105] Sandra Servia-Rodríguez, Kiran K Rachuri, Cecilia Mascolo, Peter J Rentfrow, Neal Lathia, and Gillian M Sandstrom. 2017. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In *Proceedings of the 26th international conference on world wide web*. 103–112.

[106] William H Sewell Jr. 1992. A theory of structure: Duality, agency, and transformation. *American journal of sociology* 98, 1 (1992), 1–29.

[107] Elizabeth Shove and Mika Pantzar. 2005. Consumers, producers and practices: Understanding the invention and reinvention of Nordic walking. *Journal of consumer culture* 5, 1 (2005), 43–64.

[108] Elizabeth Shove, Matt Watson, and Mika Pantzar. 2012. The dynamics of social practice: Everyday life and how it changes. (2012).

[109] Eleanor Singer and Cong Ye. 2013. The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science* 645, 1 (2013), 112–141.

[110] Pitirim Aleksandrovich Sorokin and Clarence Quinn Berger. 1939. *Time-budgets of human behavior*. Vol. 2. Harvard University Press.

[111] Ke Sun, Chunyu Xia, Xinyu Zhang, Hao Chen, and Charlie Jianzhong Zhang. 2024. Multimodal Daily-Life Logging in Free-living Environment Using Non-Visual Egocentric Sensors on a Smartphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 1 (2024), 1–32.

[112] Sunday Citizen. 2023. 30 Sleep Habits From Around the World. https://stacker.com/health/30-sleep-habits-around-world Accessed on November 01, 2024.

[113] Ann Swidler. 1986. Culture in action: Symbols and strategies. *American sociological review* (1986), 273–286.

[114] Edison Thomaz. 2016. *Automatic eating detection in real-world settings with commodity sensing*. Ph. D. Dissertation. Georgia Institute of Technology, Atlanta, GA, USA.

[115] Kirsi Tirri and Petri Nokelainen. 2008. Identification of multiple intelligences with the Multiple Intelligence Profiling Questionnaire III. *Psychology Science* 50, 2 (2008), 206.

[116] Taylor Tobin. 2018. Here's what time people in different countries stop eating at night. https://www.businessinsider.nl/dinner-times-around-the-world-2018-11 Accessed on November 01, 2024.

[117] Harry C Triandis. 2001. Individualism-collectivism and personality. *Journal of personality* 69, 6 (2001), 907–924.

[118] Elise Trumbull, Carrie Rothstein-Fisch, Patricia M Greenfield, and Blanca Quiroz. 2001. *Bridging cultures between home and school: A guide for teachers*. Routledge.

[119] Vincent WS Tseng, Michael Merrill, Franziska Wittleder, Saeed Abdullah, Min Hane Aung, and Tanzeem Choudhury. 2016. Assessing mental health issues on college campuses: Preliminary findings from a pilot study. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: adjunct*. 1200–1208.

[120] U.S. Environmental Protection Agency. 2024. Climate Change Impacts on Transportation. https://www.epa.gov/climateimpacts/climate-change-impacts-transportation Accessed on November 01, 2024.

[121] Zahra Vahedi and Alyssa Saiphoo. 2018. The association between smartphone use, stress, and anxiety: A meta-analytic review. *Stress and Health* 34, 3 (2018), 347–358.
[122] Yonatan Vaizman, Katherine Ellis, and Gert Lanckriet. 2017. Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE pervasive computing* 16, 4 (2017), 62–74.
[123] Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The experience sampling method on mobile devices. *ACM - CSUR* (2017).
[124] Mei Wang and Weihong Deng. 2018. Deep visual domain adaptation: A survey. *Neurocomputing* 312 (2018), 135–153.
[125] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. 3–14.
[126] Douglass J Wilde. 2011. *Jung's personality theory quantified.* Springer Science & Business Media.
[127] Christine D Wilson-Mendenhall, Paul Condon, Wendy Hasenkamp, George MacKerron, Karen S Quigley, and Lisa Feldman Barrett. 2022. The Qwantify app dataset: A remote experience sampling study of desire, emotion, and well-being. *Frontiers in Psychology* 13 (2022), 1054292.
[128] Ludwig Wittgenstein and GEM Anscombe. 1953. Philosophical investigations. (1953).
[129] Cornelia Wrzus and Andreas B Neubauer. 2023. Ecological momentary assessment: A meta-analysis on designs, samples, and compliance across research fields. *Assessment* 30, 3 (2023), 825–846.
[130] Yu Wu, Dimitris Spathis, Hong Jia, Ignacio Perez-Pozuelo, Tomas I Gonzales, Soren Brage, Nicholas Wareham, and Cecilia Mascolo. 2023. UDAMA: Unsupervised Domain Adaptation through Multi-discriminator Adversarial Training with Noisy Labels Improves Cardio-fitness Prediction. In *Machine Learning for Healthcare Conference*. PMLR, 863–883.
[131] Xuhai Xu, Xin Liu, Han Zhang, Weichen Wang, Subigya Nepal, Yasaman Sefidgar, Woosuk Seo, Kevin S Kuehn, Jeremy F Huckins, Margaret E Morris, et al. 2023. GLOBEM: Cross-Dataset Generalization of Longitudinal Human Behavior Modeling. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–34.
[132] Sofia Yfantidou, Marios Constantinides, Dimitris Spathis, Athena Vakali, Daniele Quercia, and Fahim Kawsar. 2023. Beyond accuracy: a critical review of fairness in machine learning for mobile and wearable computing. *arXiv preprint arXiv:2303.15585* (2023).
[133] Sofia Yfantidou, Christina Karagianni, Stefanos Efstathiou, Athena Vakali, Joao Palotti, Dimitrios Panteleimon Giakatos, Thomas Marchioro, Andrei Kazlouski, Elena Ferrari, and Šarūnas Girdzijauskas. 2022. LifeSnaps, a 4-month multi-modal dataset capturing unobtrusive snapshots of our lives in the wild. *Scientific Data* 9, 1 (2022), 663.
[134] Nira Yuval-Davis. 2004. *Gender and nation.* Routledge.
[135] Mattia Zeni, Ivano Bison, Britta Gauckler, Fernando Reis, and Fausto Giunchiglia. 2020. Improving time use measurement with personal big collection - the experience of the European Big Data Hackathon 2019. *Journal of Official Statistics* (2020).
[136] Mattia Zeni, Ilya Zaihrayeu, and Fausto Giunchiglia. 2014. Multi-device activity logging. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. 299–302.
[137] Mattia Zeni, Wanyi Zhang, Enrico Bignotti, Andrea Passerini, and Fausto Giunchiglia. 2019. Fixing Mislabeling by Human Annotators Leveraging Conflict Resolution and Prior Knowledge. *IMWUT* 3, 1 (2019).
[138] Zero Hour Climate. 2024. Inadequate Urban Planning Leading to Car Dependency & Climate Change. https://www.zerohourclimate.org/inadequate-urban-planning-leading-to-car-dependency-climate-change/ Accessed on November 01, 2024.
[139] Panyu Zhang, Gyuwon Jung, Jumabek Alikhanov, Uzair Ahmed, and Uichin Lee. 2024. A Reproducible Stress Prediction Pipeline with Mobile Sensor Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 3 (2024), 1–35.
[140] Wanyi Zhang, Qiang Shen, Stefano Teso, Bruno Lepri, Andrea Passerini, Ivano Bison, and Fausto Giunchiglia. 2021. Putting human behavior predictability in context. *EPJ Data Science* 10, 1 (2021), 42.
[141] Xintong Zhang, Meng-Cheng Wang, Lingnan He, Luo Jie, and Jiaxin Deng. 2019. The development and psychometric evaluation of the Chinese Big Five Personality Inventory-15. *PloS one* 14, 8 (2019), e0221621.

A APPENDIX

A.1 Time Diaries

This section provides a detailed outline of all the questions and answer options delivered through iLog. We report the diaries sent at fixed intervals throughout the study in the main text (Table 2). Additionally, two time diaries were sent at the start and end of each day. The morning questions listed in Table 10 inquire about the sleep quality from the previous night and the participant's expectations for the upcoming day. The evening questions, in Table 11, ask participants to evaluate their day, mention any problems they encountered, describe how they addressed them, and report any challenges due to the COVID-19 pandemic.

To reduce the answering burden, participants could deactivate receiving questions for a fixed period by selecting one of the reasons listed in Table 12.

The time diary questions were sent to participants every thirty minutes during the first two weeks and every hour in the final two weeks, asking about the activity being performed, location, social context, and mood. Additional questions were activated based on the activity selected in question *A3* in Table 2 to gather more details. Specifically, when participants indicated they were traveling, they were asked about the purpose and mode of transportation (*A3a1* and *A3a2* in Table 13, respectively). Additional questions also prompted the type of food and drinks when participants were eating (*A3c* in Table 13) and the type of sport when engaging in physical activities (*A3b* in Table 13).

Every two hours, questions in Table 14 collect information about snacks and drinks outside the main meal periods.

Table 10. Morning questions sent at 8:00 AM.

A1. How would you rate your sleep quality last night?	**A2. How do you expect your day to be?**
(1) 😃 very good	(1) 😃
(2) 🙂 fairly good	(2) 🙂
(3) 😑	(3) 😑
(4) 🙁 fairly bad	(4) 🙁
(5) ☹️ very bad	(5) ☹️

Table 11. Evening questions sent at 10:00 PM.

A7. How was your day?	**A8. Did you have any problem at [college (weekdays)] today?**	**A9. What was the problem you had?**
(1) 😃 (2) 🙂 (3) 😑 (4) 🙁 (5) ☹️	(1) Yes (2) No	open-ended question

A10.Were you able to solve the problem (alone or with help of someone)?	**A11. Is there anything that you would have liked to do today that was not possible because of the Covid-19 virus?**
open-ended question	open-ended question

Table 12. List of motivations to suspend question notifications for a fixed number of hours.

Break options
(36) Others
(37) I will participate in sports activities
(38) I have a work/study meeting
(39) I am at the cinema/theatre/hospital/church
(40) I am starting classes/lessons/lab
(41) I will go to sleep

A.2 Sensors Description

The smartphone sensors used by iLog are categorized as follows:

- *Hardware (HW)*, physical sensors of the device that detect and respond to physical environment, e.g., accelerometer, gyroscope and GPS;

Table 13. In-depth questions that appear when certain options are selected in the question "What are you doing?"

A3a1. And you travel to/from or related to:	**A3a2. How are you moving?**	**A3b. What kind of sports activity?**	**A3c. Select the main food & drink you ate** (Multiple choices)
(1) study (2) social life (3) shopping and services (4) other leisure (5) work (6) changing locality (7) other or unspecified travel purpose	(1) on foot (2) by bike (3) by bus/tram (4) by metro/subway/underground (5) by train (6) by e-scooter (7) by car (8) by car as passenger (9) by car sharing (10) by moped, motorbike (11) by moped, motorbike as passenger (12) by motorboat (13) by airplane (14) by taxi/Uber (15) other private transport modes (16) other public transport modes	(1) Walking, Trekking, and hiking (2) Jogging and running (3) Cycling, skiing, and skating (4) Ball games (5) Gymnastics and Fitness (6) Water sports (7) Other or unspecified sports or indoor activities (8) Other or unspecified sports or outdoor activities (9) Productive exercise (e.g., hunting, fishing, picking berries, mushrooms, or herbs)	(1) Bread, steamed buns and/or breakfast cereals (2) Rice, potatoes, beans, pasta, noodles, dumplings, etc. (3) Vegetables (4) Fruits (5) Meat (6) Fish (7) Processed meat (ham, bacon, sausages) (8) Dairy products (Plain or low-fat milk, yogurt, cheese) (9) Soya-based food (milk, yogurt, tofu) (10) Pastries and sweets (11) Snack/sandwiches (chips…) (12) Water (13) Soda (14) Coffee/tea or similar (15) Others non-alcoholic drink (16) Beer (17) Wine (18) Liquor (19) Other alcoholic drink (20) Other food

- *Software (SW)*, software component collecting events from the operating system and software, for instance, the Wifi the phone is connected to.

Table 15 reports the collection frequency of each sensor. The possible frequency values are:

- *on change* means that the value of the sensor is recorded only when the current value changes (along with a timestamp of when it happened);
- *up to X samples per second* means that a maximum of X events are recorded for each second (the actual frequency in the data might slightly diverge from the reported one due to the device's inaccuracy);
- *once every Y* means a new event is generated once every Y (also in this case the actual frequency approximates the reported one).

In Table 16, we briefly describe each sensor and collected variable. The sensors are grouped by category, which corresponds to the distribution bundles (see Section 5).

Table 14. Additional questions related to food and drinks.

A6b. In the last two hours did you have any snacks or drinks (except breakfast, lunch, and dinner). (administered at hours 02, 04, 06, 10, 12, 15, 17, 19, 22, 24) (Multiple choices)	**A6c. Select the food & drink taken as snack. If you had more than one snack in the last two hours, only focus on the most recent one.** (Multiple choices)
(1) No (2) Yes, between now and 30 minutes ago **(go to A6c)** (3) Yes, between 0.5 and 1 hour ago **(go to A6c)** (4) Yes, between 1 and 1.5 hours ago **(go to A6c)** (5) Yes, between 1.5 and 2 hours ago **(go to A6c)**	(1) Confectionery (Candy, Chocolate, etc) (2) Cookies, cakes, and pastries (3) Bars (Energy bar, etc.) (4) Crackers/biscuits (5) Seeds, nuts, grains, legumes (6) Savory snacks (Chips, Tapas, Pizza, Nachos, Snack mix, deep frying) (7) Sandwiches (Sandwich, Hamburgers, Hot dogs, Bagel) (8) Frozen (Ice cream, Milkshake, etc.) (9) Bread, steamed buns and/or breakfast cereals (10) Rice, potatoes, beans, pasta, noodles, dumplings, etc. (11) Vegetables (12) Fruits (13) Dairy products (milk, yogurt, cheese) (14) Soya-based food (milk, yogurt, tofu) (15) Meat (16) Fish (17) Processed meat (ham, bacon, sausages) (18) Water (19) Soda (20) Coffee/tea or similar (21) Others non-alcoholic drink (22) Beer (23) Wine (24) Spirit (25) Others alcoholic drink (26) Other food

Table 15. List sensors. The type column reports HW for hardware sensors, and SW for software sensors.

No	Type	Name	Frequency
1	SW	Bluetooth Devices	Once every minute
2	SW	Cellular network info	Once every minute
3	SW	WIFI Network Connected to	On change
4	SW	WIFI Networks Available	Once every minute
5	HW	Light	up to 10 samples per second
6	HW	Pressure	up to 10 samples per second
7	HW	Accelerometer	up to 10 samples per second
8	HW	Gyroscope	up to 10 samples per second
9	SW	Movement Activity Label	Once every 30 seconds
10	SW	Step Counter	up to 10 samples per second
11	SW	Step Detection	On change
12	HW	Location	Once every minute
13	HW	Magnetic Field	up to 10 samples per second
14	SW	Proximity	up to 10 samples per second
15	SW	Headset Status [ON/OFF]	On change
16	SW	Music Playback	On change
17	SW	Notifications received	On change
18	SW	Running Applications	Once every 5 seconds
19	SW	Airplane Mode [ON/OFF]	On change
20	SW	Battery Charge [ON/OFF]	On change
21	SW	Battery Level	On change
22	SW	Doze Mode [ON/OFF]	On change
23	SW	Ring mode [Silent/Normal]	On change
24	SW	Touch event	On change
25	SW	Screen Status [ON/OFF]	On change
26	SW	User Presence	On change

Table 16. Description of the collected ilog sensors.

Sensor	Description
Connectivity	
Bluetooth	List of discovered Bluetooth normal or low energy devices around the smartphone containing the following information: • *name*: user-friendly name of the remote device; • *address*: hardware MAC Address of the device; • *bondstate*: whether the remote device is connected; • *rssi*: Received Signal Strength Indicator; • *class code* and *class tag*: Bluetooth class of the device (e.g., phone or computer), and the class describes the characteristics and capabilities of the device (e.g., audio and telephony).
Cellular Network	Information of the cellular network to which the phone is connected to: • *cellid*: identifier of the cell; • *dbm*: signal strength; • *type*: type of the cell, possible values are lte, wcdma, gsm and cdma.
WiFi Event	Returns information related to the WIFI network to which the phone is connected; if connected, it also reports the WIFI network ID. Additional features are: • *ssid*: (Service Set Identifier) ID or unique identifier of a digital network (Wi-Fi or WLAN); • *bssid*: (Basic Service Set Identifier): sequence of characters that define a wireless computer network configured to communicate directly with each other; • *isconnected*: return whether the phone is connected to the WIFI.
WiFi Networks Event	Returns all WIFI networks detected by the smartphone. Additional features are: • *address*: is a unique identifier assigned to a network interface controller for use as a network address in communications within a network segment; • *capabilities*: list of capabilities supported by the network, e.g., WPA and WPS; • *frequency*: the WiFi frequency bands include 2.4 GHz and 5 GHz; • *name*: the name assigned to the WiFi network • *rssi*: Received Signal Strength Indicator is an estimated measure of signal strength that indicates how effectively a device can receive signals from any wireless access point or Wi-Fi router. It provides insight into the quality and reliability of the connection, often measured in decibels (dBm) to represent signal strength. The RSSI value range is between 0 and -100, where 0 signifies stronger and more stable connections.
Motion	

Accelerometer	Measures the acceleration to which the phone is subjected and captures it as a 3D vector. The unit is m/s^2.
Gyroscope	Measures the rotational forces to which the phone is subjected and it captures it as a 3D vector. The unit is rad/s.
Activities	It reports the user's activity recognized by the Google Activity Recognition API. The recognized activities are *in vehicle*, *on bicycle*, *on foot*, *running*, *still*, *tilting*, *walking* and *unknown*. The sensor reports a confidence score between 0 and 100, which represents the likelihood that the user is performing the activity.
Step Counter	It counts the total number of steps performed by the user (while carrying the phone) since the phone was powered on.
Step detector	An event is triggered each time the user takes a step.
Position	
Location	It provides the geographic coordinates of the phone: • *latitude*: latitude in degrees; • *longitude*: longitude in degrees; • *altitude*: the altitude in meters; • *accuracy*: estimated horizontal accuracy radius in meters; • *speed*: current speed of the phone in m/s; • *provider*: the source of the coordinates, i.e., GPS (hardware sensor in the devices), network (based on the WiFi network the phone is connected to) and passive (retrieve the location from other applications that already requested it); • *bearing*: horizontal direction of travel and it's the angle with respect to the north that is being faced.
Magnetic field	Reports the ambient magnetic field along the three sensor axes at the phone location.
Proximity	Measures the distance between the user's head and the phone. Depending on the phone, it may be measured in centimeters (i.e., the absolute distance) or as labels (e.g., 'near', 'far').
App usage	
Headset status	Indicates whether the headphones are connected to the phone.
Music Playback	Returns whether music is being played on the phone using the default music player from the operating system. Track information is not collected.

Notifications	It generates an event every time the phone receives a notification and when it is dismissed by the user: • *identifier*: the unique identifier of notification within the application that generated it • *isclearable*: whether the notification can be canceled when the user clears the notifications; • *isongoing*: the notification refers to an event that is ongoing, e.g., a phone call; • *package*: package name of the application. • *status*: whether the notification is posted or dismissed.
Applications	Reports the name of the application (or application package) currently running in the foreground of the phone.
Device usage	
Airplane Mode	Returns whether the phone's airplane mode is on or off. When off, all the connectivity features of the phone are turned off. Airplane mode also conserves battery life by reducing power-consuming background activities. It's accessible through the quick settings menu on most devices.
Battery Charge	Returns whether the phone is on charge: • *source*: type of power source connected to the device; possible values are USB, AC charge, wireless power source or unknown. • *status*: whether the device is being charged
Battery Monitoring Log	Returns the phone's battery level: • *level*: current battery level between 0 and 100; • *scale*: maximum level of the battery represented as a value between 0 and 100.
Doze Mode	Returns whether the phone's doze mode is on or off. Doze mode is a low-power state that a phone enters after a period of inactivity to conserve battery. In this mode, background processes and network access are restricted, allowing only essential tasks, such as high-priority notifications or alarms, to function periodically. This helps significantly reduce battery consumption while the device is idle.
Ring Mode	Reports the current ring status of the phone. When set to *normal*, the smartphone rings on incoming calls, messages and notifications by producing audible alerts. Other statuses are *vibrate* (where the phone vibrates instead of ringing) or *silent* (where all sounds are muted). Ring mode is often managed through the device's sound settings.
Touch event	It generates an event each time the user touches the screen.
Screen status	Returns whether the phone's screen is on or off.
User Presence	Detects when the user is present near the phone, for example, when the user unlocks the screen.
Enviroment	

Light	This component detects ambient light level around the phone, and it is measured in illuminance (lux). This sensor helps the device adjust the screen brightness automatically, optimizing visibility while conserving battery life. For example, in bright conditions, the screen brightness increases for better readability, whereas in darker environments, it dims to reduce eye strain and save power. Light sensors are commonly located near the top of the device, often beside the front-facing camera.
Pressure	It measures the ambient air pressure to which the phone is subjected in hPa or mbar.

A.3 Dataset Files Structure

The folder structure of the dataset is

```
. dataset root
|- Site_Trento_ITA
   |- Diachronic-Interactions
      - timediaries.parquet
   |- Synchronic-Interactions
      |- matching.csv
      |- survey1.parquet
      |- survey2.parquet
      |- survey3.parquet
   |- Sensors
      |- App-usage
         |- application.parquet
         |- headsetplug.parquet
         |- music.parquet
         |- notification.parquet
      |- Connectivity
         |- bluetooth.parquet
         |- cellularnetwork.parquet
         |- wifi.parquet
         |- wifinetworks.parquet
      |- Device-usage
         |- airplanemode.parquet
         |- doze.parquet
         |- touch.parquet
         |- batterycharge.parquet
         |- ringmode.parquet
         |- userpresence.parquet
         |- batterymonitoringlog.parquet
         |- screen.parquet
      |- Environment
         |- light.parquet
         |- pressure.parquet
      |- Motion
         |- accelerometer.parquet
         |- gyroscope.parquet
         |- stepdetector.parquet
         |- activitiespertime.parquet
         |- stepcounter.parquet
      |- Position
         |- location_poi.parquet
```

```
        |- magneticfield.parquet
        |- location_rd.parquet
        |- proximity.parquet
|- Site_Copenhagen_DEN
   |- Diachronic-Interactions
      |-...
   |- Synchronic-Interactions
      |-...
   |- Sensors
      |-...
...
```

The file `matching.csv` maps the identifier of the participants in the sensor data with the identifier used in the surveys.