Passive social sensing with smartphones: a systematic review

Heng Zhang[1] [iD] · **Ahmed Ibrahim[1]** · **Bijan Parsia[1]** · **Ellen Poliakoff[1]** · **Simon Harper[1]**

Abstract

Smartphones are widely used hubs of personal communication. With their many sensors, they are capable of monitoring social behaviours. Calls, messages, application usage and even face-to-face conversations can be captured by smartphones. These data are then used to study psychological, behavioural and clinical issues, thus providing an objective and continuous perspective on people's social lives. However, compared with questionnaires, researchers are unfamiliar with the potential and challenges of smartphone social sensing. Moreover, previous studies utilised their own instruments and strategies, which caused repetitive work and threatened the generalisability or validity of results. To fill the gap between conceptual and empirical knowledge, we conducted a systematic review on empirical studies that applied passive smartphone social sensing. Forty-seven eligible articles were retrieved from 2,741 results from five major databases. A paradigm of reviewed studies was extracted, and the comprehensive procedures of passive smartphone social sensing, including applied sensors, experiment length and data analysis methods, were also summarised. This information can provide a general reference for researchers interested in applying passive smartphone social sensing. Benefits such as ubiquitousness, unobtrusiveness and personalisability and challenges such as privacy, accuracy and methodology were also reported. Moreover, fundamental research on reasonable sensor frequency choices, standardising features

✉ Heng Zhang
   heng.zhang@manchester.ac.uk

   Ahmed Ibrahim
   ahmed.ibrahim-2@manchester.ac.uk

   Bijan Parsia
   bijan.parsia@manchester.ac.uk

   Ellen Poliakoff
   ellen.poliakoff@manchester.ac.uk

   Simon Harper
   simon.harper@manchester.ac.uk

1   The University of Manchester, Oxford Road, Manchester M13 9PL, UK

and implementing state-of-the-art technologies is recommended to enhance passive smartphone social sensing usability.

1 Introduction

Social activities are essential parts of our daily lives and have significant influence on our mental and physical health. For example, poor-quality social relationships are major risks for depression [1], while positive social behaviour benefits both individuals and society [2]. Typically, experiments on social behaviour consist of three parts: 1) set particular contexts, 2) observe participants' behaviour and 3) retrieve feedback from them [3]. The methods include interviews, questionnaires, voice or video recordings and expert observation [4]. However, these approaches have natural limitations, such as intrusively distorting behaviour and being too costly to perform at the appropriate scale [5].

Unobtrusive, passive sensing (that is, methods for collecting data from participants with minimum ongoing interaction and awareness) potentially avoids many of these problems while capturing data in situ [6]. Passive sensing not only reduces the burden on participants' awareness, cognition and memory, but unobtrusive sensors can also mitigate recall bias and the Hawthorne effect [7]. Moreover, longitudinal studies have become more feasible. They are beneficial for sensitive and stigmatised social experiments in mental health, such as studies in dementia and schizophrenia [8].

In some studies, dedicated research devices have been proposed and examined to obtain social activity data [9]. Compared with the usual carried gadgets such as smartphones and smartwatches [10], they were more obtrusive and noticeable. Nowadays, smartphones have become hubs of personal communication and computing. In 2020, 87% of UK adults owned smartphones [11]. Even for people aged over 55, the ownership rates of smartphones hit 70% in 2020 [12]. It was reported that UK adults spent 2 hours and 34 minutes online on their smartphones on an average day [13]. Social interactions on smartphones, such as calls, messages, emails and social media activity, can be captured naturally; thus, social sensing by smartphones can be less intrusive than any other device.

Additionally, off-the-shelf smartphones are embedded with multiple sensors that are potentially sufficient to capture surrounding offline social contexts. For example, raw data from microphones, the global positioning system (GPS) and accelerometers can be gathered and interpreted as conversation engagement, mobility patterns and the number of encounters to infer social interaction. These data can then be analysed to assess related topics, such as depression, loneliness [8] and work efficiency [2]. Moreover, combined with the capability to store, process and offload data, smartphones can be set up easily to become passive social sensing tools.

Although several surveys and reviews have provided big pictures of how smartphone passive sensing can be utilised in various fields, including healthcare, transportation and behaviour measurement [8, 14–16], they did not explicitly discuss how this technology was applied for social sensing. In this paper, we systematically review the following aspects of passive smartphone social sensing: 1) domains and populations studied: Sect. 3.1, 2) sensors and data collection methods: Sect. 3.2 and Sect. 3.3, 3) sensor data analysis approach after data gathering: Sect. 3.5 and Sect. 3.6, 4) accuracy and performance indexes: Sect. 3.4 and Sect. 3.7 and 5) potential problems and challenges: Sect. 3.9.

Our contributions are threefold:

(1) We developed a paradigm of passive smartphone social sensing out of the reviewed studies, which can serve as a basis for experimental strategies.
(2) We catalogued various details of all phases of existing smartphone social sensing experimentation, which can help inform the design and running of new experiments.
(3) We catalogued the potential and challenges of smartphone social sensing. We also suggested that fundamental research be conducted on sensor frequency choices, feature standardisation and the implementation of state-of-the-art technologies.

2 Article selection

Studies were included if: 1) they were experimental studies on humans, 2) they used sensors embedded on smartphones only, 3) the aim or indirect aim of the sensing was to detect if users of the phones engaged in social interaction (i.e. 'the process of reciprocal influence exercised by individuals over one another during social encounters' [17]) or overall social connectedness, 4) they involved data collection on smartphones and 5) they required minimal user interaction on the smartphones. Studies were excluded if: 1) they were crowd sensing (that is, not measuring aspects of an individual), 2) they used other sensors (such as wearable devices paired with smartphones) and 3) they required participants to put their phones in specific, atypical or atypically fixed positions.

We defined a smartphone as a mobile phone running an operating system (including but not limited to Windows Mobile, Symbian, Android and iOS) in which third-party applications can be installed for data collection purposes. Passive sensing was defined as data collected without user input except the data collected for building ground truths, such as targets for correlation and labels for machine learning. We only included studies using normal, in-the-wild smartphones and excluded studies with additional external sensors. Furthermore, papers that did not aim to infer social interaction were excluded even though they may have studied enabling technologies, such as constructing proximity networks from Bluetooth signals. We included English-language peer-reviewed journal papers and conference proceedings published from January 2000 to October 2020. We chose to start from 2000 because it was the first year Bluetooth was embedded in smartphones [18, 19].

We conducted two searches in computer science and electronics domain-specific databases ACM and IEEE, one search in health domain-specific PubMed and two searches in cross-domain databases Web of Science and ScienceDirect. We gave special consideration to variations of the term 'smartphone'. We did not include domain of application qualifiers because that would have restricted the studies found. We did not include the term 'passive' in the search string because it excluded a number of sentinel studies in our pre-search. We decided to execute the search string without it and filter the results manually.

The search string was: (smartphone OR cellphone OR cell-phone OR 'cell phone' OR 'cellular phone' OR 'mobile phone' OR 'mobile telephone' OR iPhone OR iOS OR Android OR Symbian OR 'Windows phone') AND social AND sensing NOT (crowd OR community)

3 Summary of studies

We followed the PRISMA guideline [20] for the whole review procedure. A total of 47 publications were selected from 2741 non-duplicate results from the five databases mentioned above after title/abstract screening and full-text reading. These results were examined and discussed by the first and second authors according to the inclusion and exclusion criteria. The whole process is shown in Fig. 1. In the snowballing procedure, the total number of citations from a particular publication exceeded 2,000, but most of them did not meet our inclusion criteria. To improve efficiency, we executed our searching string again on those citations to narrow the results. In addition, studies analysing public social datasets not collected by the authors but by other researchers were also considered. The first and second authors worked independently in all inclusion and exclusion processes, so the results are not biassed toward one person. We shared the final included article and discussed the disagreement. The reasons for including and excluding this disagreement were well explained by the two authors so the final agreement could be achieved. As in this case, the probability of error is the number of disagreeing articles (6) divided by the total included articles (47), which is 12.8%.

The majority of these citations were discarded because of irrelevance (e.g. studies on robotics), theoretical scope (e.g. conceptual papers on privacy and algorithms in social sensing), intrusiveness (e.g. they asked participants to label data before the experiment
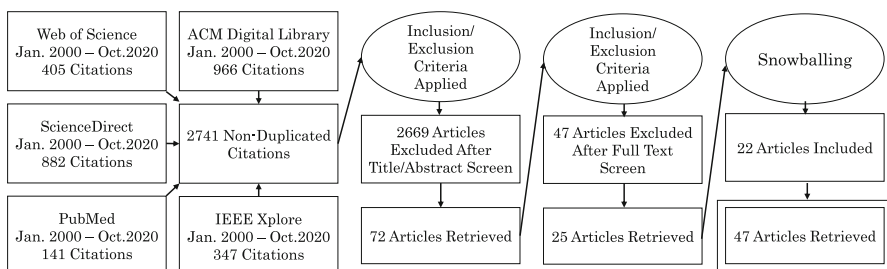

Fig. 1 Diagram of the review process

started) and the involvement of sensors other than smartphones (e.g. fixed locating sensors or wearable sensors). Although some studies used smartphones passively, they were still excluded because their targets were mobility or the contexts of users but not social interaction.

The reviewed studies all followed a standard pattern, which can serve as a paradigm for future studies. An application was installed on participants' smartphones, which collected the designated sensor data throughout the experiment period. The gathered data were transmitted to remote servers or stored locally on the smartphones. Simultaneously, ground truths, such as clinical/psychological scales and self-designed questions, were conducted at the beginning, end or during the experiment at particular frequencies. After the data period, the raw smartphone data were processed, and higher-level features were extracted. Then, various analysis methods were applied to discover the relationships between the ground truths and the smartphone data. The following sections obey the procedure of the paradigm, and we demonstrate the content of each section according to the frequency it was applied in the reviewed studies.

3.1 Domains and participants

Personality was also a favourite topic, catching 13 (28%) of the reviewed studies. Twenty-one (45%) studied human social behaviours, including proximity detection, relationship evolution, etc. Notably, four studies applied smartphone social sensing for disease research, including schizophrenia and bipolar disorder [21–24]. Some studies added interventions based on social sensing. For example, Wahle et al. introduced an intervention for participants to alleviate maladaptive thinking [25].

Most studies (i.e. 46 or 98 %) reported the number of participants as ranging from five to 11,000 with a median of 54. The length of the experiment was explicitly described in 44 (94%) studies. Thirty-five (74%) studies had a fixed study duration, and the study length ranged from two weeks to two years with a median of 70 days.

Although only one study reported that the sample size was determined based on a priori power analysis [26], all selected studies reported the populations of their experiments. Twenty-three (48%) studies used college students and staff, including undergraduates, masters, PhDs and researchers - 10 of which either worked in the same laboratory/university or lived in the same dormitory building [27–36]. Two studies recruited young adults aged 18-21 from the surrounding area of the university [37, 38]. The relationships of the participants in six studies were colleagues, friends or family members [39–44]. The participants of two studies were families consisting of parents and children [45, 46]. Moreover, disease-related papers all had special criteria for recruiting participants [23, 47].

Only 12 studies (23%) explained the providers of the smartphones that the participants used for the experiments. Seven studies (15%) gave smartphones to their participants. Two studies helped the participants migrate to new phones [24, 48]. However, the participants in [49] treated the provided phones as secondary ones. Five studies (11%) installed sensing applications on the users' own smartphones. In particular, Servia et al. asked participants to download the application themselves to take part in the study [50].

Only three of the selected studies reported details of participant withdrawal. One study reported that 50.8% of participants uninstalled the data collection app within the first two weeks, with only one-fifth of the participants left after four weeks [25]. The withdrawal of eight participants (13% of total) was reported by Buck et al. [23] and 14 (29% of total) by Wang et al. [24].

3.2 Sensors

The details of sensors used in all reviewed studies is illustrated in Table 1. GPS was generally employed for the positions in most scenarios, but individually, the locations were obtained through different mechanisms. Two studies referenced participants' visited places by detecting if their smartphones connected to certain Wi-Fi access points [32, 35]. One study collected Wi-Fi, GPS and cellular ID for locations [50]. Two studies combined Wi-Fi and GPS to observe location behaviours [48, 51], and three studies adopted cell tower IDs to interpret participants' rough neighbourhoods [27, 28, 30].

Forty-four studies collected multiple data, and three studies relied on a single sensor (GPS or Bluetooth) [31, 39, 46]. Moreover, there were distinct purposes for each data source. Locations are the significant contexts in social interactions, and they can be gathered from GPS, Wi-Fi and cellular IDs. Messages and calls are two essential ways of communicating through smartphones, but proximity is the primary condition for people to have face-to-face interaction; therefore, Bluetooth is the most common method for sensing if two or more participants are in this situation due to its low power consumption and short distance signal. In addition, a microphone is capable of detecting surrounding sounds. By analysing the raw audio it captures, the condition of whether participants are engaged in conversations or not can be inferred. Application usage, especially social media, is an increasing source of social interaction on smartphones that cannot be ignored. One study even went further by recording Facebook connections and interactions [52]. Nevertheless, the aim of collecting some other sensor data was not directly related to social interactions but for the studies' additional analysis.

The primary considerations for setting the parameters of the sensors were platform limitation and power consumption. Ten of the selected studies described the parameters of some of the social sensors they used, including GPS, Bluetooth, Wi-Fi and microphones. Fifteen studies set fixed sample rates for recording data: 10 [48], 15 [25, 53], 20 [51] or 60 minutes [37, 50] for GPS; 3 [42, 43], 5 [27, 28, 30, 46], 6 [35] or 10 minutes [51] for Bluetooth; 6 [35], 15 [25] or 60 minutes [50]for Wi-Fi; and 3 seconds [51] or 3 minutes [47, 54] for microphones. [53] logged Bluetooth and Wi-Fi whenever the respective events occurred. Three other studies had dynamic sample strategies to balance battery impact and data quality [39–41]. For example, Kiukkonen et al.'s study [55] changed the sampling rates according to known Wi-Fi connections, motion and location status. If accelerometers and GPS showed that participants were moving outdoors, it captured Wi-Fi data every 60 seconds and Bluetooth every 180 seconds. If the smartphones were connected to known Wi-Fi networks, it reduced Wi-Fi to every 120 seconds but increased Bluetooth to every 60 seconds.

Table 1 Sensors and on-device analytics of the reviewed studies

Physical sensor	Studies	On-Device analytics	Studies
GPS	[22, 24–26, 31, 34, 35, 37, 38, 40, 41, 45, 47–53, 56–61]	Call logs	[21–27, 30, 32, 33, 35–38, 41–45, 47, 48, 50, 52–54, 57, 60–65]
Bluetooth	[26–30, 32, 35, 36, 39–44, 46, 48, 49, 51–54, 58, 60, 65]	SMS logs	[22–27, 30, 32, 35–38, 41–43, 45, 47, 48, 50, 52, 57, 58, 62–65][33, 44, 60, 61, 65]
Wi-Fi	[25, 26, 32, 33, 35, 36, 41, 44, 47, 48, 50, 51, 53, 58, 60, 65]	Application usage	[24–27, 32, 33, 42–44, 49, 53, 54, 58, 60–62, 65–67]
Accelerometer	[22, 24, 25, 33, 35, 44, 47, 49, 51, 56, 57, 59, 66–69]	Screen on/off	[21, 22, 33, 44, 56–58][34, 60]
Microphone	[22–24, 47, 49–51, 54, 56, 57, 59, 67–69]	Phone lock/unlock	[24, 27, 33, 56, 58–60, 67]
Ambient light	[22, 24, 33, 49, 56, 57, 66]	Battery status	[26, 33, 44, 53, 57, 58, 60, 68]
Cellular IDs	[27, 28, 30, 44, 50]	System events	[26, 58, 60][25, 34, 35, 53, 61]

3.3 Operating system

The reviewed studies covered all mainstream mobile systems. Thirty-five (77%) studies' data collection platforms were based on the Android system, eight studies on Nokia's Symbian OS, six studies on iOS and two studies on Windows Mobile, whereas one study did not indicate which platform it was implemented on. Five studies implemented their applications both on Android and iOS [51, 54, 56, 59, 67]. All studies using Symbian and Windows Mobile occurred before 2010, as both systems have stopped service.

3.4 Validation measures

Since there were different study targets of passive smartphone social sensing, various methods were applied to validate the collected sensor data. Thirty-six (77 %) studies adopted extra methods for affirmation, including:

(1) External professional clinical or psychological scales: The Big Five personality inventory [70] was commonly used in studies related to personality [26, 42, 52–54, 57–60, 62, 63, 65, 67]. A study measuring anxiety levels [66] applied the state-trait anxiety inventory [71]. A patient health questionnaire [72] was employed to study depression [25, 49]. Mental well-being was measured by the perceived stress scale stress level [73], the flourishing scale [74] and the UCLA loneliness level scale [75] in Wang et al.'s study [49]. Lane et al. [68] treated the Yale physical activity survey [76] and short-form 36 health survey [77] as standard medical instruments. Studies on particular diseases used clinical assessment in their experimental procedures. Bipolar

disorder was diagnosed according to the schedules for clinical assessment of neuropsychiatry [78] in Faurholt et al.'s study [21]. Wang et al. [22] applied the brief psychiatric rating scale [79] to determine schizophrenia. 2) Smartphone-based questions: Smartphone-based questions mean questions directly asked on smartphones or the web. Unlike standardised scales, the authors designed these questions particularly for these studies. These questions were not validated before and were shorter compared with standardised scales. The studies only provided brief descriptions of these questions but not the details. Six studies termed these types of questions ecological momentary assessment (EMA) [22, 24, 47, 49, 49, 56], and one study [26] termed them experience sampling method (ESM). Four studies [27, 44, 50, 61] did not provide a term but asked questions under this concept. 3) Self-designed questionnaires: Two studies [35, 46] designed their questionnaires like smartphone-based ones, but they did not specify if they were conducted on smartphones. 4) In-person sessions: Only one study [38] designated in-person sessions in which the participants 'filled out a number of surveys concerning their health, well-being, trust propensity, and some demographics'. For other studies, parts of the data themselves were regarded as validation. For example, one study considered the weekly meetings of participants as the ground truth to estimate the success rate of smartphone detecting [40]. Bauer and Lukowicz [48] chose a stress period for students - before and after exams - to observe their social behaviour changes. Similarly, Harari et al. [69] monitored a whole academic term to characterise students' sociability.

In addition, these measures were taken at different time intervals. For studies that described the time point of the administration of these measures, professional clinical or psychological scales were typically applied once (at the beginning) by Faurholt et al. [21] or twice (at the beginning and end) by Harai et al. [67] and Wang et al. [49]. Specifically, Buck et al. [23] performed clinical assessments at three-month intervals, and two studies [22, 42] assessed their scales monthly. Pulekar and Agu[65] also applied psychological scales at the beginning and every four hours after. For those with smartphone-based questions, they were required multiple times a day [49, 50, 61], daily [35, 56, 66], every two days [24, 26], three times a week [22, 47] and monthly [44].

3.5 Data processing

Although all data were collected by applications running on smartphones, not all studies implemented their own tools for sensing. Twelve studies deployed platforms developed by others. They had similar functions as collecting smartphone data but used different types of data sources. The StudentLife application [49] was used by [51, 56, 59, 67]. Three studies [22, 23, 47] applied CrossCheck [24]. Funf [63] was administrated by Aharony et al. [44]. Faurholt et al. [21] used MONARCA [80]. Three [52, 64, 64] other studies did not reference their sensing platforms. Twenty-two applications (46.8%) specified that they used remote servers to store data transmitted from smartphones. Three studies (6%) stored their collected records on smartphones. Others did not describe how they aggregated their data. In addition, six studies created thresholds to filter data. Two studies [67, 69] only included days with more than 14

hours of data. Four studies [22, 24, 47, 59] set 19 hours as the minimum amount of data needed per day. In Bati and Singh's study [38], they removed the participants' data if their ground truth surveys were not complete or if the smartphones did not collect sufficient location data.

Ten studies (27 %) emphasised that their data collection procedures were ethically considered. The typical technical strategies were (1) limit the permissions of the application, so sensitive information could not be recorded, e.g. two studies mentioned that the content of messages could not be acquired by their sensing platforms [37, 38]; (2) anonymise identifiable entities like IMEI numbers, Bluetooth/Wi-Fi MAC addresses and call/messages numbers [30, 35, 61], which was usually done by randomisation and one-way hash so that the data could maintain uniqueness but lose traceability; and (3) encrypt secure connections when transferring data from smartphones to servers [49] so that the data could not be intercepted or hijacked by unauthorised parties. One study [55] in particular noted that the rights of participants should be acknowledged: participants should have the power to fully control their data. They designed a website for the participants to view all their records and allowed them to delete some or all of them.

Twenty-nine (62%) studies simply utilised smartphones as raw data collection tools and did not implement any complex algorithms on the smartphones. Their data analyses were conducted elsewhere afterwards. Otherwise, if a study collected microphone audio, it always processed the raw audio on the smartphone and stored the result. Twelve studies implemented algorithms to distinguish current sounds as conversations or voices. Two studies recorded the noise level of the surrounding environment. Furthermore, two studies [33, 68] derived types of activities, such as walking and running, from their applications locally. Similarly, four studies [22, 24, 59, 67] utilised system built-in activity recognition interfaces, including Google's Activity Recognition [81] on Android and Apple's Core Motion [82] on iOS, to obtain activity type inferences.

In addition, five studies [22, 24, 49, 56, 68] combined the data from ambient light, microphone accelerometers and smartphone usage to determine if the participant was asleep. Specifically, if the smartphone was in a dark, silent environment, stayed stationary and was not being used, they inferred that the user was sleeping. Lane et al. [68] also considered recharging events as sleep, since people often recharge their phones overnight. Furthermore, Wahle et al. [25] established several thresholds to provide positive interventions by the smartphones. If the participants stayed home too long, did not make any phone calls or walked less, recommended interactive activities popped up to promote their mental state. Lane et al. [68] also displayed animations to provide passive feedback to users based on their physical and social activities collected by the smartphones.

3.6 Feature construction

Multiple methods were employed to build feature data. They included descriptive statistics, data characteristics, higher-level semantics and other evolving processes, such as creating scores and correlation-based feature selection. All these tasks were

conducted after the data collection period except studies that provided feedback or interventions to participants during the experiment. Usually, if the study had validation measures, the collected data were interpreted and analysed with ground truths to examine its hypothesis. For call logs, SMS logs, Bluetooth, Wi-Fi, conversations/voices and smartphone usage, descriptive statistics were applied by the reviewed studies. Total number, means, variation, standard deviations and frequencies were calculated from plain numbers [25, 26, 37, 42, 43, 45, 50, 52, 56, 58, 59, 62, 64, 65, 67].

Aside from those, more features were formed by the characteristics of the data. Calls were studied as long, short, incoming, outgoing or missed separately [65]. The difference and ratio of the number of these calls were also established as new features [37]. Similarly, messages were divided into sent and received, and character length was also recognised. The reviewed studies also extracted the number of participant contacts from call and SMS logs. Applications were evaluated based on their categories, which were selected manually [26] or according to the classification of the Google Play store [24, 62]. Entropy, which measures diversity, unpredictability or irregularity, was also calculated for calls, SMSs, application usage and Bluetooth [35, 51, 52, 58, 62, 63].

All studies built higher-level semantics from raw values instead of just observing these naive numbers for location data. Some features were formulated by combining the data and their temporal information. For example, in several studies, location points were clustered into places of interest by the length of time of each visit and frequency of visits [27, 31, 37, 45]. Significant positions, such as homes, workplaces and socialisation venues, were recognised by Tsapeli and Musolesi[34]. Features like time spent, distance travelled and number of unique places were calculated for further analysis. Moreover, all the features, including semantic locations, were separated into different hours of days (morning, afternoon, evening and night), weekdays and weekends to discover distinct patterns [23, 27, 36, 57, 57].

Further evolving processes were applied after feature construction. Five studies calculated the correlation coefficient to select subsets of features to train machine learning classifiers [22, 38, 42, 51, 65]. Two studies [37, 43] selected certain features according to their predictive ability with the degree of redundancy [83]. In addition, three studies established their own scores on top of built features. Guo et al. [41] designed a social tie matrix from call, message, Bluetooth and Wi-Fi records. Relaxation scores were accumulated from touches of the screen and the number of messages and calls with different weights [33]. Lane et al. [68] created well-being scores from physical activities, sleep patterns and social interactions.

3.7 Data analysis

After abstracting features from raw data, correlation analysis or machine learning algorithms were usually implemented in the next step to explore the relation between the ground truth and the collected smartphone data. Sixteen studies (34%) performed correlation analysis, such as Pearson's correlation [22, 24, 38, 42, 43, 45, 46, 49, 51, 64], Spearman's correlation [25, 53, 54, 67], the Jaccard similarity coefficient [45], the Kendall correlation [34] and test-retest correlation [69]. In addition, six studies calculated different kinds of coefficients, including the intraclass correlation coefficient

(ICC) [54, 67], within- and between-participants coefficient [47], regression coefficient [23], generalised estimating equations (GEE) coefficient [22] and correlation matrix [56]. In particular, one study [68] used the Levenshtein similarity to compare its self-created scores and survey results, and one study [36] applied the phase slope index (PSI) to measure the temporal information flux between time-series signals.

The typical machine learning algorithms utilised by the reviewed studies were the support vector machine (SVM) [25, 30, 42, 51, 52, 57, 63], random forest [25, 31, 37, 38, 57, 58, 60, 65, 66], regression analysis [21, 24, 35, 37, 60], AdaBoost [37, 38, 65], Naive Bayes [37, 57, 65], neural networks [50, 62], Bayes net [37, 65], probabilistic model [39, 40], hidden Markov model [27], Gaussian mixture model [27], latent Dirichlet allocation [28], exponential random graph model [29], decision tree [65], gradient boosted regression trees (GBRT) [59], Kstar [38], LogitBoost [37] and XGBoost [66]. Five studies attempted different machine learning methods, compared their performances and chose the best alternative[25, 31, 37, 38, 65]. During comparison, two studies also fed particular sets of features (demography only, smartphone only or both) to the algorithm [37, 38], and the results showed that smartphone-based features all outperformed.

3.8 Benefits

Almost all studies illustrated their reasons for applying passive smartphone social sensing and discussed its benefits. Specifically, one study strengthened the advantage of Bluetooth for proximity detection by including its low battery cost, high compatibility in distinct environments, popularity among devices and less privacy sensitivity compared with voice and location [40]. In general, the prevailing incentives were:

(1) Ubiquitousness and unobtrusiveness: Almost every person has a smartphone nowadays, and it is natural to use one for communication habitually [27]. Measures using smartphones did not ask participants to carry extra devices, which could interfere with their normal behaviour [40].
(2) Capability and continuity: Smartphones are equipped with various sensors [35], and they can monitor both the contextual and behavioural information of the participants without interruption over long periods. As such, the researchers were able to observe changes and deviations from a comprehensive and continuous perspective [37, 38, 45, 53].
(3) Personalisation and individualisation: All smartphone data were collected from specific participants, providing the researchers the chance to construct in-depth models for each individual. This was particularly important for health-related studies because dedicated treatments or interventions could be introduced with personalised data [47, 69].

3.9 Problems and challenges

The benefits of passive smartphone social sensing were counterbalanced by issues and challenges. These problems involved different stages of the study, including data col-

lection and analysis. They were summarised from the reviewed studies in the following three categories:

(1) Privacy concerns: Protecting the sensitive information of participants, especially the identifiable parts, was the highest consideration in almost all studies. However, only a few of them described how these procedures were handled in detail. Usually, user identity, such as phone number, MAC address and IMEI, was hashed irreversibly to be anonymised before analysis [30, 35, 38]. In two studies [37, 38], the application was specially designed to require fewer permissions than common ones. Centellegher et al. [45] developed a digital space in which the participants could control and disclose their own data. Some studies also suggested particular methods to protect participants' privacy, such as ignoring individuals but building coarse-grained systems [30], sharing only statistical summaries and inserting random perturbations [84]. For the studies that provided phones to their participants, privacy considerations even threatened the validity of the experiments. Buck et al. [23] reported that some participants refused to use the given phones as their primary ones because they were aware that their activities were being tracked.

(2) Accuracy issues: Although smartphones go with their owners almost everywhere, we cannot fully regard smartphones as users themselves. People can break, lose and neglect to use or charge their phones [22]. Consequently, Eagle and Pentland [27] implemented a forgotten phone classifier by observing if the smartphone was charging, staying in the same place for a long time or reaming idle through missed calls and messages. Most studies relied on Bluetooth as the indicator for face-to-face interaction. Nevertheless, it was not the original intention of this technology, and it had its own technical defects. Some studies reported that all nearby devices could not be detected in a scan, and it was quite noisy [27, 39, 40]. Similarly, it was almost impossible to examine other inferences, such as conversations, locations and activities. Since there was a large amount of data, the collection of ground truths may have been too disruptive for the participants [68]. Due to the different sensor limitations on Android and iOS platforms, the collected data from the two platforms could not be merged correctly; thus, further data analysis had to be applied, which may have destroyed the coherence of the results [54, 59].

(3) Methodology challenges: The actual world is always much more complicated than our assumptions, as various problems may appear during passive smartphone social sensing. One study[30] reported that some participants did not receive any calls or messages during the experiment period, which may have been because the time of study was not long enough. As stated in the accuracy problem, face-to-face interaction was usually inferred from Bluetooth signals within the transceivers' range, but this did not imply that participants were engaged in any form of interaction necessarily [29]. Moreover, as mentioned in the participant section, only one study reported how the number of participants was decided. Most of the participants in the reviewed studies were related to the universities. They were either students, staff, researchers or families, friends and people living around them. The homogeneity of the participants and uncontrolled study designs combined with the small sample sizes threatened the generalisation of the experiment results, which was the most common concern among all the studies [25, 30, 34, 35, 37, 46, 51,

52, 69]. In addition, the ground truths almost all studies relied on were mostly self-assessment-base scales. They had natural deficits, such as subjectivity and recall bias, which meant that they were not the perfect standards [66].

4 Discussion

The reviewed studies illustrated the existing utilisation of passive smartphone social sensing. Although most studies only employed this technique as a novel instrument to investigate human social behaviour, we can still notice its potential applications in health-related research, such as mental health, depression and sleep. Moreover, comprehensive procedures of passive smartphone social sensing were exemplified, and they can provide interested computer science, social and psychology researchers beneficial references. Furthermore, the continuity and unobtrusiveness of smartphones can offer more precise and in-depth monitoring without imposing burdens on participants. The collected phone features showed significantly better performance than traditional demographic measures [37]. Smartphone passive social sensing is a promising technology in the filed not only because of its non-intrusiveness and unobtrusiveness, but also because smartphones are indeed hubs of personal communication. In addition, although not utilised for their original purpose, the variety of sensors embedded in smartphones enable context and environment information retrieval. However, studies are still necessary to confirm the hypothesis that passive smartphone social sensing is more accurate and efficient but less interruptive and troublesome than existing measurements.

4.1 Sensing strategy

All reviewed studies except those that analysed existing datasets implemented particular applications on smartphones for data collection. But only a few reported details within the study or elsewhere about how decisions such as which sensors to monitor and the frequency of sensors were made during development. As for the few who did, the principal consideration was battery consumption. For example, the data collection campaigns [55] analysed in these three studies [39–41] selected optimising power consumption as the basis of application development. They determined sampling rates according to the conditions of the smartphones, such as mobile/stationary and connected to known Wi-Fi. Another example, Meurisch et al. [31] constructed on greedy approach. They gathered as much data as acceptable while considering only privacy and energy consumption. However, none of the reviewed studies chose the parameters of the sensors as the purpose of their study perspective. Indeed, higher granularity data can provide a higher possibility of better model performance intuitively and in practice [85], but simultaneously, the expense it brings, including the privacy, transmission and storage dilemma, cannot be neglected. For example, Eagle and Pentland [27] exhibited data corruption during the collection period. It was caused by continuously writing data from the sensors and the finite number of read-write cycles of the flash memory card. Although no other reviewed studies reported issues on collecting, transferring

and storing huge amounts of data, the passive smartphone social sensing produced was still challenging. It is worth noting that researchers should determine how much data they actually need according to the actual competence of the device and the purpose of the study.

4.2 Privacy

This trade-off is also applicable to the privacy of the participants. Although no study reported participants' unacceptable experiences or violations of privacy, collecting such a high volume of sensitive data on smartphones can certainly involve privacy issues. All studies reported that they obtained appropriate ethical approvals and consent from the participants. But given the novelty, volume and sensitivity of the data gathered, it is unclear whether review boards are in a good position to assess such experiments. Anonymising various identifiers (such as phone numbers) is routine, but it is not clear how to handle the content of social contacts. This can provide a more comprehensive knowledge of participants' social behaviour, especially for those investigating mood, sentiment, mental health and related disorders [86], but extracting it can potentially be very invasive. The participants may have had various attitudes towards different types of sensors. They valued such data, but not the same across all sensors [87]. For example, Predrag et al. studied 24 participants' attitudes towards personal sensing [88]. The results showed that no attention was given to the accelerometer and barometer, but concerns were high for sensitive instruments, such as the microphone and GPS. The participants considered GPS data as 'creepy' and a threat to their physical security. Nearly all of them had a negative attitude towards raw audio. They felt 'too watched and too listened to'. However, recording audio at the necessary frequency for activity inference was more acceptable. The participants in the studies also had different concerns about the length of time the data were kept. In general, they were unwilling to keep raw data from both GPS and microphones; that is, as long as the inference was accomplished, the raw data should not be retained. Context also mattered, e.g. if the participants were concerned about sensitive information at work [89]. The value of the collected sensor data played a role in deciding the acceptability of sensors. For example, runners want to know their workout performance, so raw GPS data would likely be kept for a longer time to analyse routes, pace and distance [88].

Nguyen et al. found that privacy problems were a high concerned for participants, but these considerations were only on the abstract level, as actual everyday tracking technologies like RFID and web records were reported as significantly less concerning [90]. This was probably because, rather than specific sensor-related terms, the participants were more familiar with descriptions used in their daily lives. Higher levels of understanding of the implemented technology can trigger more worries [91]. Nevertheless, there is also a study that found that users who already had sensor-enabled devices were more willing to adopt this monitoring technology [90]. The sharing preferences of collected sensor data also had a hierarchy for different types of contacts. The participants shared more sensor-collected information with strangers than their own family and friends in Prasad et al.'s study [92]. If specific third parties provided enough benefits, the participants were more willing to share. The study also suggested

that users' privacy concerns are not static, and sharing decisions can change over time [92].

Moreover, necessary privacy strategies may be applied by using the preferred technology of target users. An appropriate interface established for users to manage privacy is a reasonable start. Christin et al. tested six graphical privacy interfaces for 80 participants but found no universal preference for the majority of the participants [93]. The users favoured elements with different colours and sizes to visualise privacy protection levels and define their preferred privacy settings. Hence, user ability tests can be conducted before the actual experiment to determine the suitable privacy interface.

4.3 Sample size and integrity

All reviewed studies described the sample sizes and demographics of the participants to some extent, but they did not demonstrate any significant clinical value for health-related research. Although small sample sizes were satisfactory for feasibility or exploratory studies, the accuracy and precision of the statistical results may have been hampered substantially [14]. The majority of the participants were still college students, researchers or related people, which also deteriorated the generalisation of the study. Usually, small-sample-size studies provide opportunities to enhance data integrity [94], but only one of the reviewed studies demonstrated the completeness of its data, which was 85.3% [27]. From the strategies which other studies utilised to filter the data, e.g. counting only the day with at least 15 hours of data as a day to analyse [69], it can be realised that the 100% acquisition of data in passive smartphone social sensing is not always applicable. This may be caused by different reasons, such as application corruption, sensing platforms, storage errors and phone turned-offs. Although no studies reported how data integrity influenced the final results, data loss is always a hidden problem and can possibly affect the quality of analysis.

4.4 Implications for future studies

Reasonable sensing strategy From the reviewed studies, various configurations of sensor application, including types, frequency and combination, were demonstrated. Nevertheless, energy consumption was a major concern for why the reviewed studies deployed these configurations. Certainly, specialised processors added to smartphones have recently made collecting high-frequency data more energy-efficient [8], but limitation considerations, such as platform restrictions, power consumption and participants' privacy, are still necessary and need to be balanced against the amount of data required for good results. From the reviewed studies, the determination of sampling rates seemed instinctive. Most studies simply stated the parameters of the sensors. For example, section 3.4 summarised that the GPS frequencies in the reviewed studies were 10, 15 or 60 minutes. These numbers were given directly by the reviewed studies without many variations. Their primary consideration was battery usage. However, sensor data frequency may have influenced these results. Different resolutions can provide better performance or cause more interference. A nontrivial method for constructing a sensing strategy can be initiating it from the goal of the study, which

may alleviate these issues from the beginning. As a result, a more reasonable choice of sensor can be made, and better results can be achieved. Experiments on different sample rates and combinations of sensors can be conducted to examine which choice is the most efficient in achieving the study's purpose. Although studies exist that try to clarify these issues, such as [95], which explored the necessary Bluetooth signal strength to infer face-to-face proximity in various situations, there are still plenty of unsolved problems in this field. For example, to recover a certain percentage of participants' face-to-face interactions, how frequent should the Bluetooth scan activate, and how does the frequency of Bluetooth scans affect the accuracy of face-to-face recovery? With these verification studies, researchers can make legitimate decisions on which sensors to capture and their sampling rates at the planning stage of the study. These experiments can also contribute to clinical value for health-related studies and lay a solid foundation for sensor usage in passive smartphone social sensing.

Causation and personalisation Although that reviewed studies examined the feasibility and validity of passive social sensing, the most commonly reported results were the correlation coefficient and performance of applied machine learning algorithms without causation explanations. Observational research usually observes individuals directly in natural settings. Therefore, for cohort studies, alternative explanations for results due to confounding may exist [96]. Moreover, researchers may only focus on the designated variables and ignore other possible factors. Two main methodologies have been proposed to control such defects: structural equation modelling and quasi-experimental designs. The former applies multivariate regression, and the latter uses a matching design to exploit the inherent characteristics of observed data [34]. Research also exists that uses collected smartphone features to conduct a quasi-experimental study, which shows the potential for causal studies. Comparing with demographics, smartphones introduce plenty of additional confounding variables, which need to be considered. Intuitively, features such as the number of phone calls and messages can reflect participants' social interactions to some extent. Some studies have also utilised other knowledge to formulate higher-level features; for example, the diversity of calls, messages and GPS based on Shannon's entropy were created [37]. These features are applicable for observation studies, which gives researchers implications and directions for their future work. But strong correlations or classifications with these features do not indicate that they are reliable measures, especially when the demographics and sample sizes are restricted. These issues may threaten the generalisability of the results and applicability of smartphone passive sensing. For example, there was a discussion in a reviewed study on personality that that findings did not match well with previous results [60]. It attributed the difference to the type of data used. Further theoretical investigations or cross-population experiments can be considered, based on current passive smartphone social sensing results. Smartphone features can be treated as items in questionnaires to be validated and reasonably interpreted. Variables constructed from smartphone data may be standardised. Smartphone passive social sensing can be expanded the into the wider field of clinical studies. Using these kinds of sensing technologies, such as smartphones, wearable devices and in-home monitoring, are often termed digital phenotyping in that area. They refer to the 'moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices' [97]. Digital phenotyping has been applied in various disease

research, such as mood disorders [98] and schizophrenia [99], suggesting that digital phenotyping is actionable and potentially useful in future clinical outcomes. However, to our knowledge, none of the digital phenotyping technologies have been approved for clinical usage and few, if any, have been adopted to replace traditional health monitoring [100]. Studies utilising digital phenotyping are often small scale, coarse and unstandardised [101]. Thus, they are insufficient for effective analysis and not suitable for robust identification of clinical signals [102].

To reach clinical validity, several factors contributing to smartphone usage have to be considered for experiment designs. Different age groups may have distinct smartphone interaction patterns. Younger generations who are used to having smartphones may spend much more time on them than the elderly. The assumption that smartphones are always carried may not be applicable to older populations [102]. Machine learning algorithms can also exaggerate this bias. Through the training data from limited populations, models and results may overfit those groups of people, and they cannot be derived for wider communities. However, this does not mean that digital phenotyping cannot be used clinically at all. With the continuity of smartphone sensing, longitudinal observations rather than snapshots can be provided to clinicians. As such, digital phenotyping can be used to explore the mechanisms and behaviours underlying psychiatric disorders rather than the outcomes alone [103].

In addition, although all conclusions drawn from the reviewed studies were in population levels, some researchers have claimed that applying personal models are more efficient than population-based ones [104]. There is also evidence showing that each individual may have unique social patterns revealed by particular characteristics [8]. It is especially effective in mental-health-related studies because dissimilar behavioural indicators of mental health difficulties can be found in different people [105]. It mirrors the N of 1 approach, which offers better efficacy than one-size-fits-all [106]. Therefore, another promising field for passive smartphone social sensing is the personalisation of a model. Each participant can be treated as a singular case, so different features can be analysed to identify which has the most influence. Various cases can also be cross-compared with validation measures to explore their exceptional patterns. Personalisation does not have to be limited to a single person. Similar characteristics, such as age, gender and personality, can be grouped together to generate 'similar user' models [107]. These models can provide opportunities to discover how these characteristics affect social behaviour. They can enlarge the volume of data for specific algorithms if the data from a single participant are not enough [8].

The next phase Sensing platforms are the basis of digital phenotyping, and every sensing platform should have documents, instructions and opportunities for other researchers to use. They can save time from repetitive development so that researchers can deploy their sensor strategies easily on these platforms. As engineering developments, smartphone sensing platforms have rapidly advanced over time [108], and an academic review can be done on these platforms for researchers to know the specifications and choose the appropriate one for their studies. Feature extraction, machine learning approaches and standard and unifying approaches can also boost the analysis of collected data and communications across the community. Smartphone capabilities continue to increase with much greater memory and storage as well as dedicated processor facilities for machine learning applications. These enable data processing,

feature extraction and inference to be performed on the phone instead of on a remote server [8], which can increase user control and potentially provide better privacy. However, smartphone OSs are not aware of such granular considerations and tend to block access to various sensors regardless of whether that data is exfiltrated (after all, even summarised data can compromise privacy). For example, Google emphasised that the accessibility service in Android, which most sensing platform utilised to collect data, should 'only be used to assist users with disabilities', which resulted in sensing applications like AWARE [109] leaving the Google Play Store and distributing on their own. In addition, the popularity of social media applications has changed the role of smartphone communication. People have switched channels from traditional phone calls and messages to video chatting and social media messaging [110], but these new channels are challenging to monitor; therefore, researchers cannot obtain the full picture of participants' social media interaction. How these new methods affect overall social behaviour still need further investigation.

4.5 Limitations

Although the search string considered different variations of the smartphone, full coverage cannot be guaranteed. The two emphasised terms, social and sensing, may not include particular results. The number of included studies was relatively small, even from an extensive search, because of strict inclusion criteria. Studies describing and discussing the procedure of data collection, implementation of platforms, related algorithms and privacy theories were excluded because they did not involve any practical experiments. Mobility, proximity and context-aware studies were also discarded due to their main aims, which were not social sensing. Nonetheless, these papers are still valuable and beneficial from other perspectives of the field. Smartphones in combination with other wearable sensors for passive social sensing are more efficient than smartphone sensing alone, as specialised sensors can capture dedicated types of data, such as heart rates and sleep patterns.

5 Conclusion

Passive smartphone social sensing provides a useful methodology for studying human social behaviour. It can be employed in various social-interaction-related areas, such as colleague cooperation, teaching performance and political opinion propagation [111]. All gathered data are individualised, precise and objective, which can inspire an in-depth understanding of the phenotypic social behaviour of each individual. They can also empower precision feedback or intervention if necessary. However, to achieve these ambitions, issues such as theoretical basis, privacy policy and experiment significance should be further explored.

References

1. Teo AR, Choi HJ, Valenstein M (2013) Social relationships and depression: ten-year follow-up from a nationally representative study. PLoS ONE 8(4):e62396
2. Aral S, Brynjolfsson E, Van Alstyne M (2007) Information, technology and information worker productivity: Task level evidence. National Bureau of Economic Research Cambridge, Mass, USA
3. Vinciarelli A, Pantic M, Bourlard H (2009) Social signal processing: Survey of an emerging domain. Image Vis Comput 27(12):1743–1759
4. Palaghias N, Hoseinitabatabaei SA, Nati M, Gluhak A, Moessner K (2016) A survey on mobile social signal processing. ACM Computing Surveys (CSUR) 48(4):57
5. Groves RM (2004) Survey errors and survey costs, vol 536. John Wiley & Sons
6. Spruijt-Metz D, Nilsen W (2014) Dynamic models of behavior for just-in-time adaptive interventions. IEEE Pervasive Comput 13(3):13–17
7. Simons AM, Beltramo T, Blalock G, Levine DI (2017) Using unobtrusive sensors to measure and minimize hawthorne effects: Evidence from cookstoves. J Environ Econ Manag 86:68–80
8. Cornet VP, Holden RJ (2018) Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 77:120–132
9. Barrat A, Cattuto C, Colizza V, Pinton J-F, Van den Broeck W, Vespignani A (2008) High resolution dynamical mapping of social interactions with active rfid. arXiv preprint arXiv:0811.4170
10. Schmid Mast M, Gatica-Perez D, Frauendorfer D, Nguyen L, Choudhury T (2015) Social sensing for psychology: Automated interpersonal behavior assessment. Curr Dir Psychol Sci 24(2):154–160
11. Ofcom. Do you personally use a smartphone?. statista - the statistics portal, (2019)
12. Statista Ofcom. Smartphone ownership penetration in the united kingdom (uk) in 2008-2020, by demographics. in statista - the statistics portal, (2020)
13. Finder.com. Mobile internet statistics
14. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M-R, Christensen H (2013) Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 15(11):e247
15. Lane ND, Miluzzo E, Hong L, Peebles D, Choudhury T, Campbell AT (2010) A survey of mobile phone sensing. IEEE Commun Mag 48(9):140–150
16. Khan WZ, Xiang Y, Aalsalem MY, Arshad Q (2012) Mobile phone sensing systems: A survey. IEEE Communications Surveys & Tutorials 15(1):402–427
17. Little W, McGivern R, Kerins N (2016) Introduction to Sociology-2nd, Canadian. BC Campus
18. Mobic.com. Ericsson unveils the first bluetooth phone, (2000). Available at http://www.mobic.com/oldnews/2000/06/ericsson_unveils_the_first_bluet.htm
19. Rudström Å, Svensson M, Cöster R, Höök K (2004) Mobitip: Using bluetooth as a mediator of social context. In Ubicomp 2004 Adjunct Proceedings
20. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. Ann Intern Med 151(4):264–269
21. Faurholt-Jepsen M, Busk J, Þórarinsdóttir H, Frost M, Bardram JE, Vinberg M, Kessing LV (2019) Objective smartphone data as a potential diagnostic marker of bipolar disorder. Australian & New Zealand Journal of Psychiatry 53(2):119–128
22. Wang R, Wang W, Aung MSH, Ben-Zeev D, Brian R, Campbell AT, Choudhury T, Hauser M, Kane J, Scherer EA et al (2017) Predicting symptom trajectories of schizophrenia using mobile sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1(3):1–24
23. Buck B, Scherer E, Brian R, Wang R, Wang W, Campbell A, Choudhury T, Hauser M, Kane JM, Ben-Zeev D (2019) Relationships between smartphone social behavior and relapse in schizophrenia: a preliminary report. Schizophr Res 208:167–172
24. Wang R, Aung Min SH, Abdullah S, Brian R, Campbell AT, Choudhury T, Hauser M, Kane J, Merrill M, Scherer EA et al (2016) Crosscheck: toward passive sensing and detection of mental health changes

in people with schizophrenia. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing 435:886–897

25. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S (2016) Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 4(3):e111

26. Schuwerk T, Kaltefleiter LJ, Jiew-Quay A, Hoesl A, Stachl C (2019) Enter the wild: Autistic traits and their relationship to mentalizing and social interaction in everyday life. J Autism Dev Disord 49(10):4193–4208

27. Eagle N, Sandy Pentland A (2006) Reality mining: sensing complex social systems. Pers Ubiquit Comput 10(4):255–268

28. Farrahi K, Gatica-Perez D (2010) Probabilistic mining of socio-geographic routines from mobile phone data. IEEE Journal of Selected Topics in Signal Processing 4(4):746–755

29. Oloritun RO, Madan A, Pentland A, Khayal I (2012) Evolution of social encounters in ad-hoc mobile face-to-face interaction networks. In 2012 International Conference on Social Informatics, IEEE, pp 192–198

30. Zhiwen Yu, Zhou X, Zhang D, Schiele G, Becker C (2013) Understanding social relationship evolution by using real-world sensing data. World Wide Web 16(5–6):749–762

31. Meurisch C, Hussain T, Gogel A, Schmidt B, Schweizer I, Mühlhäuser M (2015) A spatiotemporal approach for social situation recognition. In International Conference on Mobile Computing, Applications, and Services, Springer, pp 309–316

32. Steptoe A, Shankar A, Demakakos P, Wardle J (2013) Social isolation, loneliness, and all-cause mortality in older men and women. Proc Natl Acad Sci 110(15):5797–5801

33. Kostopoulos P, Kyritsis AI, Deriaz M, Konstantas D (2017) Stress detection using smart phone data. In eHealth 360^{circ}. Springer, pp 340–351

34. Tsapeli F, Musolesi M (2015) Investigating causality in human behavior from smartphone sensor data: a quasi-experimental approach. EPJ Data Science 4(1):24

35. Madan A, Moturu ST, Lazer D, Sandy Pentland A (2010) Social sensing: obesity, unhealthy eating and exercise in face-to-face networks. In Wireless Health 2010, ACM, pp 104–110

36. Madan A, Cebrian M, Lazer D, Pentland A (2010) Social sensing for epidemiological behavior change. In Proceedings of the 12th ACM international conference on Ubiquitous computing, pp 291–300

37. Singh VK, Agarwal RR (2016) Cooperative phoneotypes: exploring phone-based behavioral markers of cooperation. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, pp 646–657

38. Bati GF, Singh VK (2018) "trust us": Mobile phone use patterns can predict individual trust propensity. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, ACM, pp 330

39. Do Trinh MT, Gatica-Perez D (2011) Groupus: Smartphone proximity data and human interaction type mining. In 2011 15th Annual International Symposium on Wearable Computers, IEEE, pp 21–28

40. Do TM, Gatica-Perez D (2013) Human interaction discovery in smartphone proximity networks. Pers Ubiquit Comput 17(3):413–431

41. Guo B, Liang Y, Zhiwen Yu, Li M, Zhou X (2016) From mobile phone sensing to human geo-social behavior understanding. Comput Intell 32(2):240–258

42. Chittaranjan G, Blom J, Gatica-Perez D (2011) Who's who with big-five: Analyzing and classifying personality traits with smartphones. In 2011 15th Annual international symposium on wearable computers, IEEE, pp 29–36

43. Chittaranjan G, Blom J, Gatica-Perez D (2013) Mining large-scale smartphone data for personality studies. Pers Ubiquit Comput 17(3):433–450

44. Aharony N, Pan W, Ip C, Khayal I, Pentland A (2011) The social fmri: measuring, understanding, and designing social mechanisms in the real world. In Proceedings of the 13th international conference on Ubiquitous computing, ACM, pp 445–454

45. Centellegher S, De Nadai M, Caraviello M, Leonardi C, Vescovi M, Ramadian Y, Oliver N, Pianesi F, Pentland A, Antonelli F et al (2016) The mobile territorial lab: a multilayered and dynamic view on parents' daily lives. EPJ Data Science 5(1):3

46. Moturu ST, Khayal I, Aharony N, Pan W, Pentland A (2011) Using social sensing to understand the links between sleep, mood, and sociability. In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, IEEE, pp 208–214

47. Buck B, Hallgren KA, Scherer E, Brian R, Wang R, Wang W, Campbell A, Choudhury T, Hauser M, Kane JM et al (2019) Capturing behavioral indicators of persecutory ideation using mobile technology. J Psychiatr Res 116:112–117
48. Bauer G, Lukowicz P (2012) Can smartphones detect stress-related changes in the behaviour of individuals? In 2012 IEEE International Conference on Pervasive Computing and Communications Workshops, IEEE, pp 423–426
49. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, Zhou X, Ben-Zeev D, Campbell AT (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, ACM, pp 3–14
50. Servia-Rodríguez S, Rachuri KK, Mascolo C, Rentfrow PJ, Lathia N, Sandstrom GM (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, pp 103–112
51. Wang H, Marsella S (2017) Assessing personality through objective behavioral sensing. In 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), IEEE, pp 131–137
52. Mønsted B, Mollgaard A, Mathiesen J (2018) Phone-based metric as a predictor for basic personality traits. J Res Pers 74:16–22
53. Schoedel R, Pargent F, Au Q, Völkel ST, Schuwerk T, Bühner M, Stachl C (2020) To challenge the morning lark and the night owl: Using smartphone sensing data to investigate day–night behaviour patterns. European Journal of Personality
54. Harari GM, Müller SR, Stachl C, Wang R, Wang W, Bühner M, Rentfrow PJ, Campbell AT, Gosling SD (2019) Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. Journal of personality and social psychology
55. Kiukkonen N, Blom J, Dousse O, Gatica-Perez D, Laurila J (2010) Towards rich mobile phone datasets: Lausanne data collection campaign. Proc. ICPS, Berlin, 68
56. DaSilva AW, Huckins JF, Wang R, Wang W, Wagner DD, Campbell AT (2019) Correlates of stress in the college environment uncovered by the application of penalized generalized estimating equations to mobile sensing data. JMIR Mhealth Uhealth 7(3):e12084
57. Khwaja M, Vaid SS, Zannone S, Harari GM, Aldo Faisal A, Matic A (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):1–24
58. Stachl C, Quay A, Schoedel R, Gosling SD, Harari GM, Buschek D, Völkel ST, Schuwerk T, Oldemeier M, Ullmann T et al (2020) Predicting personality from patterns of behavior collected with smartphones. Proc Natl Acad Sci 117(30):17680–17687
59. Wang W, Harari GM, Wang R, Müller SR, Mirjafari S, Masaba K, Campbell AT (2018) Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2(3):1–21
60. Stachl C, Au Q, Schoedel R, Buschek D, Völkel S, Schuwerk T, Oldemeier M, Ullmann T, Hussmann H, Bischl B, et al (2019) Behavioral patterns in smartphone usage predict big five personality traits
61. LiKamWa R, Liu Y, Lane ND, Zhong L (2011) Can your smartphone infer your mood. In PhoneSense workshop, pp 1–5
62. Gao S, Li W, Song LJ, Zhang X, Lin M, Lu S (2020) Personalitysensing: A multi-view multi-task learning approach for personality detection based on smartphone usage. In Proceedings of the 28th ACM International Conference on Multimedia, pp 2862–2870
63. de Montjoye Y-A, Quoidbach J, Robic F, Sandy Pentland A (2013) Predicting personality using novel mobile phone-based metrics. In International conference on social computing, behavioral-cultural modeling, and prediction, Springer, pp 48–55
64. Montag C, Błaszkiewicz K, Lachmann B, Andone I, Sariyska R, Trendafilov B, Reuter M, Markowetz A (2014) Correlating personality and actual phone usage. J Individ Differ 35(3):158
65. Pulekar G, Agu E (2016) Autonomously sensing loneliness and its interactions with personality traits using smartphones. In 2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT), IEEE, pp 134–137
66. Fukazawa Y, Ito T, Okimura T, Yamashita Y, Maeda T, Ota J (2019) Predicting anxiety state using smartphone-based passive sensing. J Biomed Inform 93:103151

67. Harari GM, Vaid SS, Müller SR, Stachl C, Marrero Z, Schoedel R, Bühner M, Gosling SD (2020) Personality sensing for theory development and assessment in the digital age. Eur J Pers 34(5):649–669

68. Lane ND, Lin M, Mohammod M, Yang X, Hong L, Cardone G, Ali S, Doryab A, Berke E, Campbell AT et al (2014) Bewell: Sensing sleep, physical activities and social interactions to promote wellbeing. Mobile Networks and Applications 19(3):345–359

69. Harari GM, Gosling SD, Wang R, Chen F, Chen Z, Campbell AT (2017) Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods. Comput Hum Behav 67:129–138

70. John OP, Srivastava S et al (1999) The big five trait taxonomy: History, measurement, and theoretical perspectives. Handbook of personality: Theory and research 2(1999):102–138

71. Spielberger CD (1983) State-trait anxiety inventory for adults

72. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K (2004) Monitoring depression treatment outcomes with the patient health questionnaire-9. Medical care, pp 1194–1201

73. Cohen S, Kamarck T, Mermelstein R (1983) A global measure of perceived stress. J Health Soc Behav 24:385–396

74. Diener E, Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, Biswas-Diener R (2010) New well-being measures: Short scales to assess flourishing and positive and negative feelings. Soc Indic Res 97(2):143–156

75. Russell DW (1996) Ucla loneliness scale (version 3): Reliability, validity, and factor structure. J Pers Assess 66(1):20–40

76. Dipietro L, Caspersen CJ, Ostfeld AM, Nadel ER (1993) A survey for assessing physical activity among older adults. Medicine & Science in Sports & Exercise

77. Jenkinson C, Coulter A, Wright L (1993) Short form 36 (sf36) health survey questionnaire: normative data for adults of working age. BMJ 306(6890):1437–1440

78. Wing JK, Babor T, Brugha TS, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N (1990) Scan: schedules fonr clinical assessment in neuropsychiatry. Arch Gen Psychiatry 47(6):589–593

79. Kopelowicz A, Ventura J, Liberman RP, Mintz J (2008) Consistency of brief psychiatric rating scale factor structure across a broad spectrum of schizophrenia patients. Psychopathology 41(2):77–84

80. Bardram JE, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Kessing LV (2013) Designing mobile health technology for bipolar disorder: a field trial of the monarca system. In Proceedings of the SIGCHI conference on human factors in computing systems pp 2627–2636

81. Google Developers. Google activity recognition api, 2020

82. Apple Developers. Apple ios core motion, 2020

83. Steyerberg EW, Eijkemans MJC, Habbema JDF (1999) Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol 52(10):935–942

84. Mitchell TM (2009) Mining our reality. Science 326(5960):1644–1645

85. Chon Y, Shin H, Talipov E, Cha H (2012) Evaluating mobility models for temporal prediction with high-granularity mobility data. In 2012 IEEE International Conference on Pervasive Computing and Communications, IEEE, pp 206–212

86. Andriotis P, Takasu A, Tryfonas T (2014) Smartphone message sentiment analysis. In IFIP International Conference on Digital Forensics, Springer, pp 253–265

87. Barua D, Kay J, Paris C (2013) Viewing and controlling personal sensor data: what do users want? In International Conference on Persuasive Technology, Springer, pp 15–26

88. Klasnja P, Consolvo S, Choudhury T, Beckwith R, Hightower J (2009) Exploring privacy concerns about personal sensing. In International Conference on Pervasive Computing, Springer, pp 176–183

89. Hayes GR, Poole ES, Iachello G, Patel SN, Grimes A, Abowd GD, Truong KN (2007) Physical, social, and experiential knowledge in pervasive computing environments. IEEE Pervasive Comput 6(4):56–63

90. Nguyen DH, Kobsa A, Hayes GR (2008) An empirical investigation of concerns of everyday tracking and recording technologies. In Proceedings of the 10th international conference on Ubiquitous computing, pp 182–191

91. Beckwith R, Mainwaring S (2005) Privacy: Personal information, threats, and technologies. In Proceedings. 2005 International Symposium on Technology and Society, 2005. Weapons and Wires: Prevention and Safety in a Time of Fear. ISTAS 2005, IEEE, pp 9–16

92. Prasad A, Sorber J, Stablein T, Anthony D, Kotz D (2012) Understanding sharing preferences and behavior for mhealth devices. In Proceedings of the 2012 ACM workshop on Privacy in the electronic society, pp 117–128

93. Christin D, Reinhardt A, Hollick M, Trumpold K (2012) Exploring user preferences for privacy interfaces in mobile sensing applications. In Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia, pp 1–10

94. Leon AC, Davis LL, Kraemer HC (2011) The role and interpretation of pilot studies in clinical research. J Psychiatr Res 45(5):626–629

95. Liu S, Jiang Y, Striegel A (2013) Face-to-face proximity estimationusing bluetooth on smartphones. IEEE Trans Mob Comput 13(4):811–823

96. Carlson MDA, Morrison RS (2009) Study design, precision, and validity in observational studies. J Palliat Med 12(1):77–82

97. Onnela J-P, Rauch SL (2016) Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 41(7):1691–1696

98. Brietzke E, Hawken ER, Idzikowski M, Pong J, Kennedy SH, Soares CN (2019) Integrating digital phenotyping in clinical characterization of individuals with mood disorders. Neuroscience & Biobehavioral Reviews 104:223–230

99. Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela J-P (2018) Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. NPJ digital medicine 1(1):1–9

100. Cohen AS, Schwartz E, Le T, Cowan T, Cox C, Tucker R, Foltz P, Holmlund TB, Elvevåg B (2020) Validating digital phenotyping technologies for clinical use: the critical importance of "resolution". World Psychiatry 19(1):114

101. Faurholt-Jepsen M, Bauer M, Kessing LV (2018) Smartphone-based objective monitoring in bipolar disorder: status and considerations. International journal of bipolar disorders 6(1):1–7

102. Huckvale K, Venkatesh S, Christensen H (2019) Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ digital medicine 2(1):1–11

103. Mohr DC, Zhang M, Schueller SM (2017) Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 13:23–47

104. Estrin D (2014) Small data, where n= me. Commun ACM 57(4):32–34

105. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT (2015) Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 38(3):218

106. Prochaska JO, Redding CA, Evers KE et al (2015) The transtheoretical model and stages of change. Health behavior: Theory, research, and practice. John Wiley & Sons, NY, pp 125–148

107. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T (2015) Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 3(2):e42

108. Wikipedia. Mobile phone based sensing software, 2021

109. Ferreira D, Kostakos V, Dey AK (2015) Aware: mobile context instrumentation framework. Frontiers in ICT 2:6

110. Perrin A (2015) Social media usage. Pew research center 52–68

111. Madan A, Farrahi K, Gatica-Perez D, Sandy Pentland A (2011) Pervasive sensing to model political opinions in face-to-face networks. In International Conference on Pervasive Computing, Springer, pp 214–231