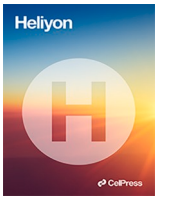
Heliyon 10 (2024) e35472

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Review article

# Use of smartphone sensor data in detecting and predicting depression and anxiety in young people (12–25 years): A scoping review

Joanne R. Beames [a,b,*], Jin Han [a], Artur Shvetcov [a], Wu Yi Zheng [a], Aimy Slade [a], Omar Dabash [a], Jodie Rosenberg [a], Bridianne O'Dea [a], Suranga Kasturi [a], Leonard Hoon [c], Alexis E. Whitton [a], Helen Christensen [d], Jill M. Newby [e]

[a] *Black Dog Institute, University of New South Wales, Sydney, NSW, Australia*
[b] *Center for Contextual Psychiatry, Department of Neurosciences, KU Leuven, Belgium*
[c] *Applied Artificial Intelligence Institute, Deakin University, Melbourne, Australia*
[d] *University of New South Wales, Sydney, NSW, Australia*
[e] *Black Dog Institute and School of Psychology, University of New South Wales, Sydney, NSW, Australia*

ARTICLE INFO

*Keywords:*
Depression
Anxiety
Youth
Phone
Sensing
Machine learning

ABSTRACT

Digital phenotyping is a promising method for advancing scalable detection and prediction methods in mental health research and practice. However, little is known about how digital phenotyping data are used to make inferences about youth mental health. We conducted a scoping review of 35 studies to better understand how passive sensing (e.g., Global Positioning System, microphone etc) and electronic usage data (e.g., social media use, device activity etc) collected via smartphones are used in detecting and predicting depression and/or anxiety in young people between 12 and 25 years-of-age. GPS and/or Wifi association logs and accelerometers were the most used sensors, although a wide variety of low-level features were extracted and computed (e.g., transition frequency, time spent in specific locations, uniformity of movement). Mobility and sociability patterns were explored in more studies compared to other behaviours such as sleep, phone use, and circadian movement. Studies used machine learning, statistical regression, and correlation analyses to examine relationships between variables. Results were mixed, but machine learning indicated that models using feature combinations (e.g., mobility, sociability, and sleep features) were better able to predict and detect symptoms of youth anxiety and/or depression when compared to models using single features (e.g., transition frequency). There was inconsistent reporting of age, gender, attrition, and phone characteristics (e.g., operating system, models), and all studies were assessed to have moderate to high risk of bias. To increase translation potential for clinical practice, we recommend the development of a standardised reporting framework to improve transparency and replicability of methodology.

A significant proportion of young people around the world experience mental health problems [1,2]. The most common psychological disorders in young people are anxiety and depressive disorders [3,4], with 73.3 % of anxiety disorders and 36.9 % of mood disorders emerging by 25 years-of-age [5]. The impact of experiencing mental health problems like anxiety and depression early in life

* Corresponding author. Center for Contextual Psychiatry, Department of Neurosciences, KU Leuven, Leuven, Belgium.
*E-mail address:* joanne.beames@kuleuven.be (J.R. Beames).

https://doi.org/10.1016/j.heliyon.2024.e35472
Received 11 November 2022; Received in revised form 25 July 2024; Accepted 29 July 2024
Available online 30 July 2024

can be severe, causing disruptions across learning and education, social and emotional functioning, and overall health [6–8]. Earlier onset of mental health problems also predicts negative functioning into adulthood, including reduced employment opportunities, relationship problems, as well as a more severe and recurring course of mental disorders [9,10]. Assessment of mental health symptoms has historically been dependant on self-report questionnaires, which can be biased and burdensome to collect. Further, in clinical practice, self-report questionnaires may not be delivered at the right time or frequently enough to capture early deterioration of symptoms, resulting in delayed or inappropriate treatments. To better understand and respond to youth anxiety and depression, there is an urgent need to develop advanced detection and prediction methods that are scalable in mental health research and practice.

Digital phenotyping is a broad term encompassing advanced methodologies that capture real-time moment-to-moment information about people's experiences and behaviours as they go about their daily lives [11,12]. This information can be collected passively using sensors or electronic activity records, with no input from the user (e.g., Global Positioning System or GPS coordinates), or actively, with the user intentionally performing a task or an action (e.g., survey) [11]. Digital phenotyping has demonstrated a range of applied uses including screening and early diagnosis, monitoring for detection of symptoms or relapse, and treatment (e.g., tailoring interventions, monitoring treatment efficacy) [13]. A sensemaking framework proposed by Mohr provides an outline of how digital data can be transformed to provide clinically meaningful insights [14]. The framework includes the following steps: (1) extracting raw sensor data (e.g., GPS data); (2) creating low-level features (e.g., transition time between locations); (3) amalgamating low-level features to create high-level behavioural markers (e.g., activity, social withdrawal); and (4) relating behavioural markers to actual clinical state (e.g., depression/anxiety) [14]. Application of Mohr's digital phenotyping sensemaking framework to empirical passive sensing studies offers a promising way forward in understanding links between features and clinical state [15,16].

Passive data collection is particularly advantageous in mental health and psychiatry. Benefits include the ability to collect many disparate variables concurrently and unobtrusively and the reduction of retrospective biases in reporting [14,17]. Obtaining objective information about emotions and behaviours in real-time provides contextual information about where people are and what they are doing, revealing subtle patterns that are missed by traditional assessments [17]. Smartphones are one device that can be used to passively collect a range of digital data streams. For example, smartphones can continuously record objective digital features of location (e.g., GPS), activity (e.g., accelerometer), social activity or conversations (e.g., microphone), sleep (e.g., light sensor), and phone use (e.g., lock/unlock) [13,14,18]. Using smartphones as the method of data collection for young people facilitates scalability given the availability of this technology [13,19]. Young people are generally open and willing to use their smartphones to manage their own mental health, with some caveats around privacy and use of information [20,21]. However, it is unclear to what extent phone-based passive digital data can facilitate detection and prediction of youth anxiety and depression.

Studies in mental health and psychiatry have focused on digital phenotyping via wearables (e.g., Fitbits) and smartphones in primarily adult samples. Narrative and systematic reviews in adult populations demonstrate that the most collected sensor-based streams include location, accelerometer, and social information, which are used to infer behaviours including sleep, exercise, and social interactions [15,18,22]. These reviews have shown some consistent patterns between digital features from sleep, physical activity, location, and phone use data and adult depression [15,22]. The consensus, however, is that aggregated features might have greater predictive value for mental health than single features [15,22]. Considerably less research using digital phenotyping has been conducted in youth mental health and psychiatry, with only one systematic review focusing specifically on children and adolescent samples under 18 years-of-age [23]. This brief review summarised how studies have combined passive data from wearables and smartphones with active self-report data to better understand a range of paediatric psychopathologies (e.g., anorexia nervosa, attention-deficit hyperactivity disorder, depression, anxiety), treatment efficacy, and preventative measures. The authors concluded that integrating different sources of data may be important for more accurately capturing the emotions and behaviours of youth with psychiatric illnesses. A common finding across the adult and youth digital phenotyping literature is methodological heterogeneity and incomplete reporting [15,18,22,23]. For example, the target samples, types of devices used, and types of digital features extracted or parsed vary considerably across studies. Some studies do not provide sufficient information to allow replication of methodology or statistical analysis techniques. Further inquiry is necessary to synthesise relationships between passive sensing data and specific mental health symptoms or diagnoses in young samples.

Available reviews are limited in some respects. First, the reviews typically examine both wearables and smartphone data collection methods. Focusing on smartphones is important because there are practical limitations of wearables that reduce scalability at a population level (e.g., they are less ubiquitous). Technical differences between smartphones and wearables also introduce an additional source of heterogeneity and bias into the results. For example, there are differences in what these devices are used for in daily life, how they collect data, and how this data are processed and analysed [24,25]. Second, existing reviews do not consider adolescence *and* early adulthood (e.g., from 12 to 25 years-of-age). This gap is problematic because the peak age of onset and emergence of many mental disorders, including depression and anxiety, occurs within this developmental period [5]. Finally, there are no direct comparisons of how digital phenotyping data are used for anxiety and depression in early adolescence and early adulthood. Together, these limitations mean that no clear recommendations have been suggested about how technical architecture and sensing platforms should be designed, or how raw smartphone data should be processed and analysed. The implication is that there is limited understanding about how passive sensing data collected from smartphones is used and analysed, and the types of conclusions that can be made about youth depression and anxiety [15,18,22,23].

## 1. The current scoping review

The current scoping review aims to identify and map the available research to better understand how passive sensing (e.g., GPS, microphone etc) and electronic usage data (e.g., social media, device activity etc) collected via smartphones are used in detecting and

predicting depression and/or anxiety in young people between 12 and 25 years-of-age. A scoping review was deemed most appropriate given the heterogeneity of methodologies and the emerging application of passive data collection in youth mental health research. Extending prior research, we use Mohr et al.'s framework [14] to map sensors (e.g., GPS, accelerometer) with low-level features (e.g., location/activity type), high-level behaviours (e.g., movement/psychomotor activity, avoidance, sleep), and clinical inferences (e.g., depression, anxiety). We summarise how digital data are sampled, how variables are operationalised, and the types of conclusions that can be made about youth anxiety and depression.

We also evaluated the quality of the studies conducted in this field. While conducting this scoping review, it became evident that a suitable quality assessment tool for digital phenotyping studies was not available in the literature. Existing reviews have used various approaches to assess quality, such as combining available tools to increase relevance for different study designs or providing informal descriptions [15,17,18,23,26]. A custom tool is necessary to capture aspects of methodology, reporting, and ethics or privacy requirements that are unique to digital phenotyping studies. Even though a risk of bias assessment is not required for scoping reviews, this project presented an ideal opportunity to develop a novel tool for assessing bias in smartphone digital phenotyping studies. We use this review as a first "test-case" to implement the tool.

## 2. Methods

### 2.1. Protocol and registration

The review protocol was registered on the Open Science Framework (see https://osf.io/6h3a4/). Minor deviations from the protocol in the methodology reflect changes based on familiarisation with the literature and relevance (e.g., extracted data, quality assessment procedure). The scoping review methods were guided by the Joanna Briggs Institute (JBI) Manual for Evidence Synthesis [27] along with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis extension for Scoping Reviews (PRISMA-ScR) [28]. The review is organised according to the framework proposed by Arksey and O'Malley [29].

Step 1 Identifying the Research Question

How is phone sensor data about mobility (i.e., location, activity), social interactions, and sleep used to detect and predict depression and anxiety in young people between 12 and 25 years-of-age? Sub-questions included: (1) What digital features are extracted, combined, and used? (2) What digital features are ubiquitously associated with youth depression and/or anxiety? (3) What analytic approaches are used? (4) What is the quality of studies? and (5) How heterogeneous are the methodologies of available studies?

Step 2 Identifying the Relevant Studies

Final search terms included keywords and MeSH terms (where possible) related to phones, digital phenotyping, depression, anxiety, and young people. Search terms were combined appropriately with Boolean operators and were adapted as appropriate for each database. The final search was conducted in PubMed, PsycINFO, Embase, ACM Digital Library, IEEE Xplore, and Web of Science. These databases were selected given their subject focus on clinical psychology, digital mental health, and computer science/health informatics. Searches were limited to articles published in the English language after 2007. The search was conducted on November 2, 2021. See Tables S1 and S2 in Supplementary Material Appendix A for key concepts and an example search strategy.

Step 3 Study Selection

Specific inclusion criteria are outlined in Table S3 in Supplementary Material Appendix A. Studies were included if they explored relationships between passive sensing data collected via smartphones regarding location, activity, social interactions, and/or sleep (e.g., location, accelerometer, microphone, Bluetooth) and depression and/or anxiety. Studies were excluded if they: (1) did not report on real-time prospective passive sensing data collected by smartphones (e.g., focused on wearable devices or self-report data collection only); (2) did not focus on depression and/or anxiety as the primary outcome, or examined depression and/or anxiety in the context of another mental disorder or physical condition; (3) did not use validated measures of depression and/or anxiety; (4) did not explore relationships between passive sensing data and depression and/or anxiety (or explored relationships in the context of a treatment trial); (5) included adult samples or samples where less than 80 % was aged between 12 and 25 years; (6) were non-empirical, not published in a journal article, non-peer-reviewed, or the full-text could not be accessed; or (7) were qualitative.

Covidence systematic review software [30] was used for screening procedures. Search results from each database were uploaded into Covidence, where duplicates were identified and removed. JRB independently screened the unique titles and abstracts for eligibility. For all records meeting the inclusion criteria, full texts were independently assessed by three reviewers (JRB, WZ, JR). All full texts were screened twice and reasons for exclusion were recorded in Covidence. As required, additional information was sought from study authors to ascertain eligibility. Disagreements were resolved through discussions amongst the three reviewers; a fourth, more senior team member was available for consultation (JN).

Step 5 Charting the Data

Study data was independently extracted by JRB, WZ, AS, JH, OD, and AS into Covidence using a piloted template. JRB checked all data for consistency. The following information was extracted: manuscript details (authors, publication year, discipline/field, study location), study characteristics (study design and setting, secondary analysis of existing dataset and details about dataset), sample characteristics (baseline mental health characteristics, age range and mean, gender, sample size, attrition), self-report measures (validated depression and/or anxiety measures, measurement timepoints), digital data collection (e.g., smartphone operating system, mobile application or platform used, duration of data collection), sensor type, sensor sampling details, low-level features and definitions (e.g., number of locations visited, activity time), high-level behavioural features (e.g., activity, location), analytical methods and results (e.g., type of analysis, purpose of analysis, measures of association reported, summary of non-significant and significant results).

### 2.2. Quality assessment for digital phenotyping studies using smartphones (QA-DPSS)

Items were adapted from available tools [17,26], and revised through consultation with experts in digital psychiatry, mental health research, and computer science. The tool focuses on aspects of methodology and reporting that are unique to digital phenotyping, rather than the overall study design. Evaluated domains include: (1) adequate reporting of digital sampling and data collection; (2) adequate reporting of digital measurements; (3) adequate reporting of digital data quality; (4) adequate reporting of study analysis and results; (5) ethics and safety reporting. The tool aims to provide a judgment on the reproducibility and transparency of digital phenotyping methodology and reporting. See Supplementary Material Appendix B for items, scoring, and overall bias judgments.

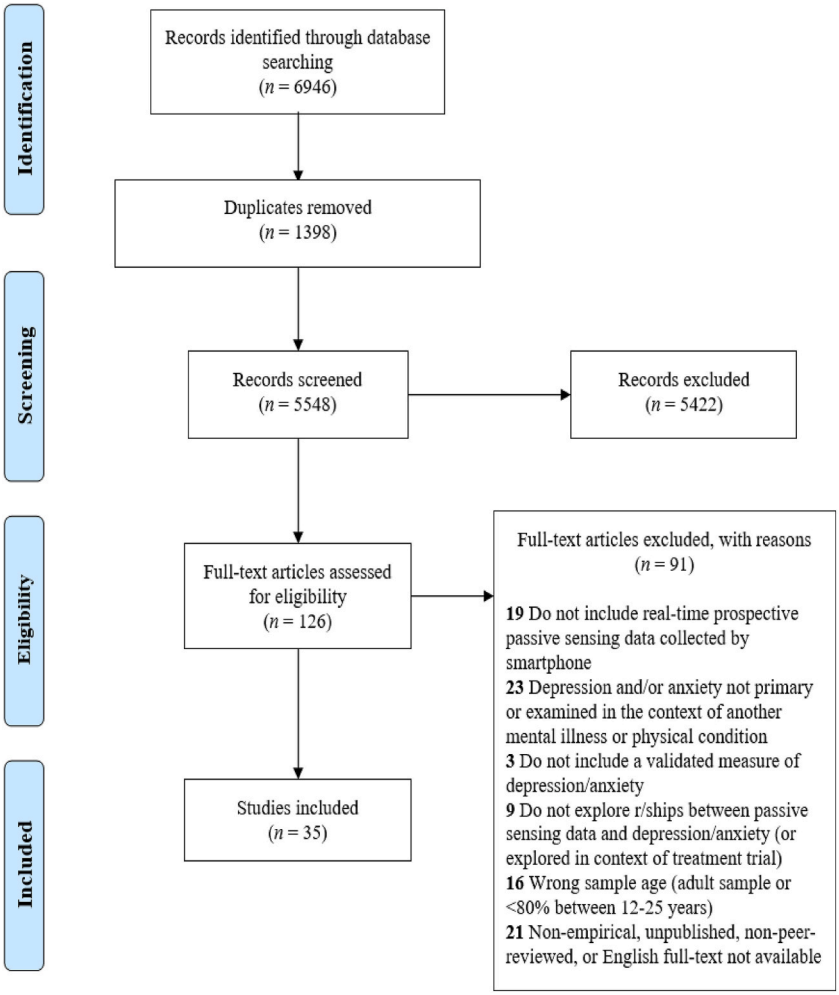



**Fig. 1.** PRISMA flow diagram.

## 3. Results

### 3.1. Study selection

See Fig. 1 for the PRISMA diagram illustrating the study selection process. A total of 6,946 articles were identified, from which 1,398 duplicates were removed. Titles and abstracts of 5,548 articles were screened for eligibility by JRB. 5,422 articles were deemed irrelevant and therefore excluded, leaving 126 articles for full-text review. JRB, WZ, and JR independently screened the full text articles for eligibility. All articles were screened by two reviewers to ensure consistency. Of these articles, 91 were excluded because they did not meet the inclusion criteria. Any disagreements were resolved through discussion and consultation. Screening resulted in 35 original studies being included in the current review.

### 3.2. Study characteristics

Studies primarily used prospective longitudinal observational designs ($n$ = 34, 97.14 %). Of the studies that reported sample characteristics, the median sample size was 72 ($M$ = 105.49, range = 13–816), with an average attrition rate of 21.61 % (range = 0–59 %). Most studies were conducted in the United States of America ($n$ = 28, 80.00 %) with university/college student samples ($n$ = 32, 91.43 %). The mean age was 20.21 (range = 10–30) and participants were predominantly female ($M$ = 57.55 %, range = 20.83–100.00 %). Twenty (57.14 %) studies focused on depression, 6 (17.14 %) on anxiety, and 9 (25.71 %) on a combination of both.

Most studies reported baseline depression and/or anxiety characteristics of the sample ($n$ = 25, 71.43 %). One study required a diagnosis (2.86 %) via structured clinical interview for eligibility. Just over half of the studies involved primary collection of new data sets ($n$ = 18, 51.43 %), with 14 (40.00 %) conducting secondary analyses on existing datasets. In 3 (8.57 %) cases it was unclear. Of the 14 studies using an existing dataset, 9 (64.29 %) used StudentLife, a publicly available dataset of students at Dartmouth university (https://studentlife.cs.dartmouth.edu/). Papers were published between 2014 and 2021, either in the computer data science field ($n$ = 20, 57.14 %) or the psychology, mental health, and medical field ($n$ = 15, 42.86 %). See Table 1 for detailed study characteristics.

Reporting of study characteristics varied across studies: 16 (45.10 %), 8 (22.86 %), and 21 (60.00 %) did not report sufficient details about age, gender, and attrition respectively. Further, 27 (77.14 %) studies failed to report one of these sample characteristics. Computer data science publications were less likely to report these characteristics compared to psychology, mental health, and medical publications (see Fig. 2). This pattern of results might be explained by the fact that computer data science publications were more likely to re-use the same existing datasets ($n$ = 11, 31.43 %) than psychology, mental health, and medical publications ($n$ = 3, 8.57 %). See Table S4 in Supplementary Material Appendix A for a breakdown of sample characteristics not reported.

### 3.3. Self-report depression and anxiety measures

A range of validated measures were used to assess symptoms of depression ($n$ = 11) and anxiety ($n$ = 6). For studies assessing depression, the most common measure was the Patient Health Questionnaire (including different versions e.g., PHQ-9, PHQ-8; $n$ = 16, 55.17 %). The PHQ is used in a relatively high proportion of studies because it is the primary depression measure in StudentLife. For studies assessing anxiety, the most common measure was the Social Interaction Anxiety Scale (SIAS; $n$ = 8, 53.33 %). These measures of depression and/or anxiety were obtained at various time points, with 14 (40.00 %) studies measuring at baseline and post-sensing, 11 (31.43 %) studies measuring at baseline only, and 2 studies measuring at post only (5.71 %). Further, 8 (22.68 %) studies measured self-reports periodically through the sensing period, which was typically obtained using ecological momentary assessments.

### 3.4. Digital data collection

*Phone Specifications.* Given that digital data collection technical details are the same for studies that re-used established datasets, the following section only includes primary studies that collected or used unique data ($n$ = 24[1]). Most of these studies used Android operating systems ($n$ = 10, 41.67 %) or a combination of Android and IOS operating systems ($n$ = 6, 25.00 %); three (20.83 %) used IOS only and five (29.83 %) did not include this technical information. Only six (25.00 %) studies reported phone operating system version for some or all participants in the sample. There was variation in operating system versions across studies: Android ≥4.3 ($n$ = 3), Android ≥4 and IOS ≥8 ($n$ = 1), IOS ≥10 ($n$ = 1), IOS ≥ 4s ($n$ = 1). Most studies ($n$ = 18, 75.00 %) did not report the operating system version, the proportion of which was similar across the computer science field ($n$ = 9, 75.00 %) and the psychology, mental health, and medical field ($n$ = 9, 75.00 %). Most studies utilised participant's own phones ($n$ = 22, 91.67 %) and a purpose-built study app to collect and record participant data ($n$ = 20, 83.33 %). See Table S4 in Supplementary Material Appendix A for a breakdown of phone specifications not reported across the entire dataset.

*Duration.* Digital data collection occurred for an average of 10 weeks, with a minimum of 4 days and a maximum of 40 weeks. Nine (25.71 %) studies collected sensing data for longer than 10-weeks; these studies were primarily focused on depression.

[1] This includes the first published study using the StudentLife dataset, the LifeRhythm dataset, and the DemonicSalmon dataset. Note that in all other sections, data from all studies ($n$ = 35) are descriptively summarised.

**Table 1**
Characteristics of studies included in the scoping review.

| Study | Year | Field | Country | Design | Existing Dataset | Setting | Self-Report Outcome Measures | Self-Report Measurement Time Points | Duration | Baseline Characteristics | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | N (N sensing) | % Attrition | Mental Health Characteristics | Mage (range) | % Female |
| **Depression** | | | | | | | | | | | | | | |
| Ben-Zeev et al. | 2015 | P | USA | L | Y (StudentLife) | University | PHQ-9 | Pre/Post | 10-weeks | 47 (37) | 21.28 | NR | 22.5 (19–30) | 21 |
| Chikersal et al. | 2021 | C | USA | L | Unclear | University | BDI-II | Pre/Post | 16-weeks | 188 (79–110 depending on sensor used) | 26.6 | 14.5 % reported mild (score 14–19), moderate (score 20–28), or severe (score 29–63) depression on the BDI | NR | NR |
| Demasi et al. | 2016 | C | USA | L | N | University | BDI | Pre/Post | 8-weeks | 107 (44) | 59 | $M_{BDI} = 11.5$ at baseline | NR | 61.4 |
| Dissing et al. | 2021 | P | Denmark | L | N | University | MDI | Baseline and approximately 4-months later (3-months after sensing period) | 4-weeks | 816 (816 for baseline analyses; 571 for change score analyses) | 28 | NR | 21.6 (NR) | 23 |
| Elhai et al. | 2018 | P | USA | L | N | University | PHQ-9 | Baseline | 1-week | 68 | 0 | $M_{PHQ\text{-}9} = 5.44$ | 19.75 (18–25) | 64.7 |
| Farhan, Lu, et al. | 2016 | C | USA | L | Y (StudentLife) | University | PHQ-9 | Pre/Post | 10-weeks | 60 (49) | NR | NR | NR | NR |
| Farhan, Yue, et al. | 2016 | C | USA | L | Y (LifeRhythm)[‡] | University | PHQ-9 | Baseline and every two weeks throughout sensing period (used average score) | 32-weeks | 79 | NR | Categorised as 'depressed' or 'not depressed' via initial interview by a clinician based on the Diagnostic and Statistical Manual of Mental Health (DSM-5) and self-reported PHQ-9 scores: 24.05 % depressed; 75.95 % not depressed | NR (18–25) | 73.9 |
| Gerych et al. | 2019 | C | USA | L | Y (StudentLife) | University | PHQ-9 | Post | 10-weeks | 48 | 31.67 | NR | NR | 20.83 |
| Jacobson & Chung | 2020 | C | USA | L | N | University | DASS-21-D | Baseline | 1-week | 31 | NR | DASS-21: 6.45 % moderately depressed; 38.7 % severely depressed; 54-8% very severely depressed (clinical cut-offs NR) | 19.13 (18–27) | 64.52 |

(*continued on next page*)

**Table 1** (*continued*)

| Study | Year | Field | Country | Design | Existing Dataset | Setting | Self-Report Outcome Measures | Self-Report Measurement Time Points | Duration | Baseline Characteristics | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | N (N sensing) | % Attrition | Mental Health Characteristics | Mage (range) | % Female |
| Kim et al. | 2021 | C | South Korea | L | Y (StudentLife) | University | PHQ-9 | Pre/Post | 10-weeks | 38 | NR | PHQ-9: 42.5 % minimal depression (score 1–4); 36.5 % minor depression (score 5–9); 15 % moderate depression (score 10–14); 2.5 % moderately severe depression (score 15–19); 2.5 % severe depression (score 20–27) | NR | NR |
| Li et al. | 2017 | P | USA | L | Y (StudentLife) | University | PHQ-9 | Pre/Post | 10-weeks | 47 | 22 | NR | 22.5 (19–30) | NR |
| Lu et al. | 2018 | C | USA | L | Y (LifeRhythm) [1] | University | QIDS | Once per day throughout sensing period (used normalised average score) | 13-weeks | 103 | NR | Categorised as 'depressed' or 'not depressed' via initial interview by a clinician based on the Diagnostic and Statistical Manual of Mental Health (DSM-5) and self-reported PHQ-9 scores: 37.9 % depressed; 62.1 % not depressed | NR (18–25) | 76.7 |
| Saeb et al. | 2016 | C | USA | L | Y (StudentLife) | University | PHQ-9 | Pre/Post | 10-weeks | 48 | NR | NR | NR | 20.83 |
| Wang, Chen et al. | 2014 | C | USA | L | Y (StudentLife)* | University | PHQ-9 | Pre/Post | 10-weeks | 60 (48) | 31.67 | $M_{PHQ\text{-}9} = 5.6$; 35.42 % minimal depression (score 1–4); 31.25 % minor depression (score 5–9); 12.5 % moderate depression (score 10–14); 2.08 % moderately severe depression (score 15–19); 2.08 % severe depression (score 20–27). | NR | 20.83 |

(*continued on next page*)

**Table 1** (*continued*)

| Study | Year | Field | Country | Design | Existing Dataset | Setting | Self-Report Outcome Measures | Self-Report Measurement Time Points | Duration | Baseline Characteristics | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | N (N sensing) | % Attrition | Mental Health Characteristics | Mage (range) | % Female |
| Wang, Wang, et al. | 2018 | P | USA | L | N | University | PHQ-8, PHQ-4 | Pre/Post (PHQ-8); once per week throughout sensing period (PHQ-4) (used average score) | 18-weeks | 83 | 14.46 | $M_{\text{PHQ-8}} = 6.09$; 19.28 % classified as depressed (PHQ-8 $\geq$ 10) | 20.13 (NR) | 51.8 |
| Ware, Yue, et al. | 2019 | C | USA | L | Y (LifeRhythm) [1] | University | Phase I: PHQ-9 | *Phase I:* Baseline and every two weeks throughout sensing period | *Phase I:* 28-weeks | 79 | NR | Categorised as 'depressed' or 'not depressed' via initial interview by a clinician based on the Diagnostic and Statistical Manual of Mental Health (DSM-5) and self-reported PHQ-9/ QIDS scores. Phase I: 24.05 % depressed; 75.95 % not depressed. Phase II: 37.86 % depressed; 62.14 % not depressed | NR | 73.9 |
| | | | | L | | | Phase II: QIDS | *Phase II:* Baseline and once per week throughout sensing | *Phase II:* 40-weeks | 103 | NR | As above. | NR | 76.7 |
| Xu, Chikersal, Doryab, et al. | 2019 | C | USA | L | N | University | BDI | *Phase I:* Pre/Post | *Phase I:* ≈15-weeks (106 days) | 188 (138) | 14.89 | 82.61 % classified as depressed (BDI >13); 17.39 % as non-depressed (BDI ≤13) | NR | NR |
| | | | | L | | | | *Phase II*: Post | *Phase II:* ≈15-weeks (113 days) | 267 (212) | 11.61 | Baseline data not collected | NR | NR |
| Xu, Chikersal, Dutcher, et al. | 2021 | C | USA | L | Unclear | University | BDI | *Phase I:* Pre/Post | *Phase I:* 16 weeks | 188 (138) | 14.89 | NR | 18.2 (NR) | 58.5 |
| | | | | L | | | | *Phase II:* Post | *Phase II:* 10-weeks | 207 (169) | NR | Baseline data not collected | 18.4 (NR) | 64.1 |
| Yang, Mo, et al. | 2017 | C | USA | L | Y (StudentLife) | University | PHQ-9 | Pre/Post | 10-weeks | 48 | NR | NR | NR | NR |

(*continued on next page*)

**Table 1** (*continued*)

| Study | Year | Field | Country | Design | Existing Dataset | Setting | Self-Report Outcome Measures | Self-Report Measurement Time Points | Duration | Baseline Characteristics | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | N (N sensing) | % Attrition | Mental Health Characteristics | Mage (range) | % Female |
| Yue et al. | 2017 | C | USA | L | Y (LifeRhythm) [‡] | University | PHQ-9 | Baseline and every two weeks throughout sensing period (used average score) | 32-weeks | 79 | NR | Categorised as 'depressed' or 'not depressed' via initial interview by a clinician based on the Diagnostic and Statistical Manual of Mental Health (DSM-5) and self-reported PHQ-9 scores: 24.05 % depressed; 75.95 % not depressed | NR (18–25) | 73.9 |
| **Anxiety** | | | | | | | | | | | | | | |
| Boukhechba, Chow, et al. | 2018 | P | USA | L | N | University | SIAS | Baseline | 2-weeks | 228 | NR | $M_{SIAS} = 29.91$ | 19.43 (NR) | 62 |
| Boukhechba, Huang, et al. | 2017 | C | USA | L | Unclear | University | SIAS | Baseline | 2-weeks | 54 | NR | $M_{SIAS} = 29.67$ | NR | NR |
| Fukazawa et al. | 2019 | C | Japan | L | N | University | STAI | Once per day for 1-month | NR | 20 | NR | NR | NR (20–24) | 25 |
| Gong et al. | 2019 | C | USA | L | N | University | SIAS | Baseline | 2-weeks | 52 | NR | $M_{SIAS} = 35.02$ | 20.5 | 68 |
| Huang et al. | 2016 | P | USA | L | N | University | SIAS | Baseline | 10-days | 18 | NR | $M_{SIAS} = 38.39$ | NR | NR |
| Yang, Tang, et al. | 2021 | C | China | CS | N | University | GAD-7 | Baseline | 2-weeks | 168 | NR | Categorised as 'general anxiety disorder' or 'normal controls' using self-reported GAD-7 scores: 50 % general anxiety disorder subjects with $M_{GAD\text{-}7} = 13.79$; 50 % normal controls with $M_{GAD\text{-}7} = 0.73$ (clinical cut-off NR) | 24.36 (NR) | 58.3 |
| **Depression and Anxiety** | | | | | | | | | | | | | | |
| Boukhechba, Daros, et al. | 2018 | C | USA | L | Y (Demons and Salmon)* | University | DASS-21-D<br>SIAS | Pre/Post | 2-weeks | 72 | NR | Anxiety: $M_{SIAS} = 9.52$<br>Depression: $M_{DASS\text{-}21} = 3.48$ | 19.8 (18–23) | 51.39 |
| Cao et al. | 2020 | P | USA | L | N | Clinical | PHQ-9<br>HAM-D<br>HAM-A | Bi-weekly in-clinic assessments | 8-weeks | 13 (11) | 15.38 | $M_{PHQ\text{-}9} = 12.72$; 72.73 % in the normal-to-mild | 14.93 (12–17) | 84.62 |

(*continued on next page*)

**Table 1** (*continued*)

| Study | Year | Field | Country | Design | Existing Dataset | Setting | Self-Report Outcome Measures | Self-Report Measurement Time Points | Duration | Baseline Characteristics | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | N (N sensing) | % Attrition | Mental Health Characteristics | Mage (range) | % Female |
| | | | | | | | | | | | | range (PHQ-9 ≤ 14); 27.27 % in the moderate-to-severe range (PHQ-9 > 14) | | |
| Chow et al. | 2017 | P | USA | L | N | University | DASS-21-D<br>SIAS | Baseline | 2-weeks | 72 (63) | NR | $M_{SIAS} = 29.9$; ≈16 % likely scored above the mean of a diagnosed sample<br>Depression: $M_{DASS-21} = 3.3$ | 19.8 (18–23) | 37 |
| Jacobson et al. | 2020 | P | USA | L | Y (DemonicSalmon) | University | DASS-21-D<br>SIAS | Baseline | 2-weeks | 72 (59) | NR | $M_{SIAS} = 29.13$; 36 % with clinical levels of social anxiety disorder (SIAS >34) | 19.8 (18–23) | 51 |
| Knight & Bidargaddi | 2018 | P | Australia | L | N | Community | DASS-21-D<br>DASS-21-A | Baseline | 32-weeks | 53 (43) | NR | Depression: $M_{DASS-21} = 10.01$<br>Anxiety: $M_{DASS-21} = 6.47$<br>Stress: $M_{DASS-21} = 10.23$ | 20.7 (18–25) | 77 |
| MacLeod et al. | 2021 | P | Canada | L | N | Clinical and non-clinical | CES-DC<br>SCARED | Baseline | 2-weeks | 161 (122) | 22.36 | Anxiety: $M_{SCARED} = 33.02$; 31.9 % had a lifetime diagnosis of generalised anxiety disorder and 9.8 % had a lifetime diagnosis of social phobia. During the study period, 11.4 % diagnosed with generalised anxiety, and 4.9 % diagnosed with social phobia (diagnostic tool NR).<br>Depression: $M_{CES-DC} = 32.59$; 24.5 % had a lifetime diagnosis of depression. During the study period, 9.8 % diagnosed with depression | 18 (10–21) | 78.60 |

(*continued on next page*)

**Table 1** (*continued*)

| Study | Year | Field | Country | Design | Existing Dataset | Setting | Self-Report Outcome Measures | Self-Report Measurement Time Points | Duration | Baseline Characteristics | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | N (N sensing) | % Attrition | Mental Health Characteristics | Mage (range) | % Female |
| | | | | | | | | | | | | (diagnostic tool NR). | | |
| Melcher et al. | 2021 | P | USA | L | N | University | DASS-21-D<br>PHQ-9<br>HAM-D<br>DASS-21-A<br>GAD-7<br>SIAS | Pre/Post | 4-weeks | 102 (100) | 1.96 | Depression: $M_{PHQ-9}$ = 8.58 (mild range); Anxiety: $M_{GAD-7}$ = 6.50 (mild range) | 20.3 (18–27) | 75 |
| Rozgonjuk et al. | 2018 | P | USA | L | No | University | DASS-21-D<br>PHQ-2<br>DASS-21-A | Baseline (DASS-21); once per day throughout sensing period (PHQ-2) | 1-week | 101 | NR | Depression: 66.34 % normal range; 9.90 % mild range; 16.83 % moderate range, 1 % severe range; 5.94 % extremely severe range on the PHQ-9.<br>Anxiety: 61.39 % normal range; 15.84 % mild range; 10.89 % moderate range, 2.97 % severe range; 8.91 % extremely severe range on the DASS-21. | 19.53 (NR) | 76.20 |
| Shoval et al. | 2020 | P | Israel | L | No | University | BDI-II<br>STAI | Post | 4-days | 40 | NR | Baseline data not collected | 23 (19–30) | 100 |

*Notes*. **Field** – P = Psychology, mental health, or medicine; C = Computer data science. **Design –** L = Longitudinal; CS = Cross-Sectional. **Self-Report Outcome Measures** – PHQ-9 = Patient Health Questionnaire-9; BDI-II = Beck Depression Inventory-Second Edition; BDI = Beck Depression Inventory; MDI = Major Depression Inventory; DASS-21-D = Depression Anxiety Stress Scales-Depression Subscale; QIDS = Quick Inventory of Depressive Symptomatology; PHQ-8 = Patient Health Questionnaire-8; PHQ-4 = Patient Health Questionnaire-4; SIAS = Social Interaction Anxiety Scale; STAI=State-Trait Anxiety Inventory; GAD-7 = Generalised Anxiety Disorder Questionnaire-7; HAM-D = Hamilton Depression Rating Scale; HAM-A = Hamilton Anxiety Rating Scale; DASS-21-D = Depression Anxiety Stress Scales-Anxiety Subscale; CES-DC = Center for Epidemiological Studies Depression Scale for Children; SCARED = Screen for Child Anxiety-Related Emotional Disorders. ***N* (*N* sensing)** – Total number of participants reported in text (number of participants included in sensor analyses, when reported, otherwise assumed to be the same as total *N*). **% Attrition** – Calculated as percent of participants lost to follow-up. ǂ Use same subsets of data. * First published analysis of existing dataset (i.e., StudentLife or Demons/Salmon). **NR** = Not reported.

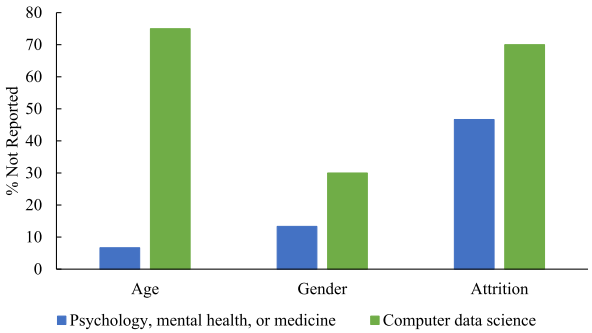



**Fig. 2.** Percentage of publications not reporting sample characteristics (by field).

### 3.5. Sensors, low-level features, and high-level behavioural features

See Table 2 for a summary of sensors and high-level behavioural features. The most common sensors were GPS and/or Wifi association logs ($n = 26$, 74.29 %), followed by accelerometers ($n = 14$, 40.00 %), phone lock/unlock status ($n = 12$, 34.29 %), and call logs ($n = 13$, 37.14 %; see Fig. 3). Low-level features from a range of sensors were used to make inferences about location ($n = 24$, 68.57 %), activity ($n = 15$, 42.86 %), sociability ($n = 19$, 54.29 %), phone use ($n = 12$, 34.29 %), sleep ($n = 10$, 28.57 %), circadian movement ($n = 4$, 11.43 %), orientation ($n = 1$, 2.86 %), and other contextual features ($n = 1$, 2.86 %; see Fig. 4). Here, circadian movement refers to 24-h rhythm in location data. Consistent routines, like leaving and returning home at similar times each day, indicate high circadian movement, while irregular patterns of moving between locations indicate low circadian movement. A similar pattern was observed in feature use across studies focusing on anxiety and/or depression, although circadian movement, phone use, and sleep were more often used to explore depression (see Fig. 5). See Supplementary Material Appendix C for sensors and high-level behaviours used for anxiety and depression.

Location and activity were primarily inferred from GPS/Wifi ($n = 24$, 68.57 %) or accelerometers ($n = 12$, 34.29 %), respectively. Circadian movement was also inferred from GPS/Wifi ($n = 4$, 11.43 %). Sociability was primarily inferred from call logs ($n = 13$, 37.14 %), SMS logs ($n = 7$, 20.00 %), Bluetooth ($n = 4$, 11.43 %), or microphone ($n = 5$, 14.29 %). Sleep was primarily inferred from accelerometers ($n = 5$, 14.29 %), microphone ($n = 3$, 8.57 %), light ($n = 6$, 17.14 %), or phone lock/unlock status ($n = 4$, 11.43 %). These data were typically combined to create an index of sleep. Lock/unlock status ($n = 9$, 25.71 %) and screen status ($n = 6$, 17.14 %) were the main indicators of phone use. Most studies used more than one digital data source (median = 2, range: 1–8), and explored more than one behavioural feature (median = 2, range: 1–5).

There was considerable heterogeneity in which low-level digital features were parsed from sensor data. See Table S5 in Supplementary Material Appendix D for a list.

### 3.6. Statistical analyses

Studies typically used more than one statistical analysis to explore relationships between low-level digital features and depression and/or anxiety. The most common techniques were machine learning methods ($n = 21$, 60.00 %), correlations ($n = 17$, 48.57 %), and statistical regression models ($n = 12$, 34.29 %). Analyses were typically exploratory, with many associations between features investigated and reported.

### 3.7. Correlation and statistical regression results summary

See Tables S6 and S7 in Supplementary Material Appendix E for a comprehensive summary of relevant results, including non-significant results.

*Location and Activity.* Decreased mobility was generally related to increased anxiety and depression symptoms, however there was variation in the specific low-level features that were significant across studies. There was most evidence for a negative relationship between entropy and depression symptoms [31–33] and between location variance and anxiety or depression symptoms [31–35]. There was some evidence that higher depression was related to fewer unique locations visited and greater time spent at home [31,33,34,36]. Further, one study found a significant predictive relationship when combining home-stay data with communication data [37]. In keeping with these results, another study found that higher activity, as indicated by a range of accelerometry descriptive statistics, was negatively correlated with both anxiety and depression [38]. Further, some studies found that mobility features varied across time [39], that there were differences between Android and IOS [31,33,34], and that associations were stronger after fusing data from both operating systems [33]. For anxiety, results were mixed for entropy (i.e., uniformity or volatility of time spent in different locations), cumulative staying time in specific locations, and transition frequency between different locations [35,39,42–44]. For depression, results were mixed for distance travelled, amount of time active/inactive, and moving speed [31–34,36,37,45–49], and few studies examined transitions between locations [32,48], time spent at specific locations other than home [36], or circadian movement [32,48]. It is also important to note that mobility and activity findings often differed between studies that examined anxiety or depression

**Table 2**
Smartphone data collection methods and sensor data features of studies included in the scoping review.

| Study | Year | Phone Source | Operating System (Model) | Application or Platform Name | Behaviour Inference | Sensor Data Features Extracted | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GPS/ Wifi | Accel. | Step Count | Gyro. | Call logs | SMS logs | BT | Mic. | Light | Lock/ Unlock | Screen | Other | N Unique Features |
| **Depression** | | | | | | | | | | | | | | | | | | |
| Ben-Zeev et al. | 2015 | Personal or study phone provided | Android (study phone ≥4.0) | StudentLife | Social | | | | | | | | X | | | | | 8 |
| | | | | | Activity | X | X | | | | | | | | | | | |
| | | | | | Sleep | | X | | | | | | X | X | X | X | | |
| Chikersal et al. | 2021 | Personal | IOS and Android (NR) | AWARE | Location | X | | | | | | | | | | | | 5 |
| | | | | | Social | | | | | X | | X | | | | | | |
| | | | | | Phone Use | | | | | | | | | | X | X | | |
| Demasi et al. | 2016 | Personal | Android (NR) | NR | Activity | | X | | | | | | | | | | | 1 |
| | | | | | Sleep | | X | | | | | | | | | | | |
| Dissing et al. | 2021 | Study phone provided | NR | NR | Social | | | | | X | X | | | | | | X * | 1 |
| Elhai et al. | 2018 | Personal | IOS (≥4S) | Moment | Phone Use | X | | | | | | | | | X | X | | 3 |
| Farhan, Lu et al. | 2016 | Personal or study phone provided | Android (study phone ≥4.0) | StudentLife | Location | X | | | | | | | | | | | | 5 |
| | | | | | Activity | X | | | | | | | | | | | | |
| | | | | | Social | | | | | | | | X | | | | | |
| | | | | | Sleep | | | | | | | | | X | | | | |
| | | | | | Phone Use | | | | | | | | | | X | | | |
| Farhan, Yue et al. | 2016 | Personal | IOS (≥8) and Android (≥4.0) | LifeRhythm | Location | X | | | | | | | | | | | | 2 |
| | | | | | Activity | | X | | | | | | | | | | | |
| Gerych et al. | 2019 | Personal or study phone provided | Android (study phone ≥4.0) | StudentLife | Location | X | | | | | | | | | | | | 1 |
| Jacobson & Chung | 2020 | Personal | Android | MoodTriggers | Location | X | | | | | | | | | | | | 2 |
| | | | | | Social | | | | | X | | | | | | | | |
| Kim et al. | 2021 | Personal or study phone provided | Android (study phone ≥4.0) | StudentLife | Sleep | | | | | | | | | | X | | | 1 |

(*continued on next page*)

**Table 2** (*continued*)

| Study | Year | Phone Source | Operating System (Model) | Application or Platform Name | Behaviour Inference | Sensor Data Features Extracted | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GPS/ Wifi | Accel. | Step Count | Gyro. | Call logs | SMS logs | BT | Mic. | Light | Lock/ Unlock | Screen | Other | N Unique Features |
| Li et al. | 2017 | Personal or study phone provided | Android (study phone ≥4.0) | StudentLife | Location | X | | | | | | | | | | | | 2 |
| | | | | | Social | | | | | | | | X | | | | | |
| Lu et al. | 2018 | Personal | Android and IOS (NR) | LifeRhythm | Location | X | | | | | | | | | | | | 1 |
| Saeb et al. | 2016 | Personal or study phone provided | Android (study phone ≥4.0) | StudentLife | Location | X | | | | | | | | | | | | 1 |
| | | | | | Circadian Movement | X | | | | | | | | | | | | |
| Wang, Chen et al. | 2014 | Personal or study phone provided | Android (study phone ≥4.0) | StudentLife | Location | X | | | | | | X | | | | | | 6 |
| | | | | | Activity | | X | | | | | | | | | | | |
| | | | | | Social | | | | | | | | X | | | | | |
| | | | | | Sleep | | X | | | | | | X | X | X | | | |
| Wang, Wang et al. | 2018 | Personal | IOS and Android (NR) | StudentLife | Location | X | | | | | | | | | | | | 5 |
| | | | | | Activity | | X | | | | | | | | | | | |
| | | | | | Social | | | | | | | | X | | | | | |
| | | | | | Sleep | | X | | | | | | X | X | X | | | |
| | | | | | Phone Use | | | | | | | | | | X | | | |
| Ware, Yue et al. (phase 1 and 2) | 2019 | Personal | IOS and Android (NR) | LifeRhythm | Location | X | | | | | | | | | | | | 1 |
| | | | | | Circadian Movement | X | | | | | | | | | | | | |
| Xu, Chikersal, Doryab et al. (phase 1 and 2) | 2019 | Personal | NR | AWARE | Location | X | | | | | | | | | | | | 4 |
| | | | | | Social | | | | | X | | X | | | | | | |
| | | | | | Phone Use | | | | | | | | | | X | | | |
| | | | | | Circadian Movement | X | | | | | | | | | | | | |
| Xu, Chikersal, Dutcher et al. (phase 1 and 2) | 2021 | Personal | IOS (NR) | AWARE | Location | X | | | | | | | | | | | | 4 |
| | | | | | Social | | | | | X | | X | | | | | | |

(*continued on next page*)

**Table 2** (*continued*)

| Study | Year | Phone Source | Operating System (Model) | Application or Platform Name | Behaviour Inference | Sensor Data Features Extracted | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GPS/ Wifi | Accel. | Step Count | Gyro. | Call logs | SMS logs | BT | Mic. | Light | Lock/ Unlock | Screen | Other | N Unique Features |
| | | | | | Phone Use | | | | | | | | | | X | | | |
| | | | | | Circadian Movement | X | | | | | | | | | | | | |
| Yang, Mo et al. | 2017 | Personal or study phone provided | Android (study phone ≥4.0) | StudentLife | Location | X | | | | | | | | | | | | 5 |
| | | | | | Activity | | X | | | | | | | | | | | |
| | | | | | Social | | | | | | | X | | | | | | |
| | | | | | Phone Use | | | | | | | | | | X | | | |
| | | | | | Other | | | | | | | | X | | X | | | |
| | | | | | Contextual Features | | | | | | | | | | | | | |
| Yue et al. | 2017 | Personal or study phone provided | IOS and Android (NR) | LifeRhythm | Location | X | | | | | | | | | | | | 1 |
| | | | | | Activity | X | | | | | | | | | | | | |
| **Anxiety** | | | | | | | | | | | | | | | | | | |
| Boukhechba, Chow et al. | 2018 | Personal | IOS and Android (NR) | Sensus | Location | X | | | | | | | | | | | | 1 |
| Boukhechba, Huang et al. | 2017 | Personal | Android (NR) | NR | Location | X | | | | | | | | | | | | 2 |
| | | | | | Social | | | | | X | X | | | | | | | |
| Fukazawa et al. | 2019 | Personal | Android (NR) | NR | Activity | | X | | | | | | | | | | | 6 |
| | | | | | Social | | | | | X | X | | | | | | X** | |
| | | | | | Phone Use | | | | | | | | | X | | | | |
| | | | | | Orientation | | | | X | | | | | | | | | |
| Gong et al. | 2019 | Personal | Android (NR) | Sensus | Location | X | | | | | | | | | | | | 4 |
| | | | | | Activity | | X | | | | | | | | | | | |
| | | | | | Social | | | | | X | X | | | | | | | |
| Huang et al. | 2016 | Personal | NR | NR | Location | X | | | | | | | | | | | | 1 |
| Yang, Tang et al. | 2021 | Study phone provided | NR | WeChat applet | Activity | | X | | | | | | | | | | | 1 |
| **Depression and Anxiety** | | | | | | | | | | | | | | | | | | |
| Boukhechba, Daros et al. | 2018 | Personal | Android (≥4.3) | Sensus | Location | X | | | | | | | | | | | | 4 |
| | | | | | Activity | | X | | | | | | | | | | | |
| | | | | | Social | | | | | X | X | | | | | | | |

(*continued on next page*)

**Table 2** (*continued*)

| Study | Year | Phone Source | Operating System (Model) | Application or Platform Name | Behaviour Inference | Sensor Data Features Extracted | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GPS/ Wifi | Accel. | Step Count | Gyro. | Call logs | SMS logs | BT | Mic. | Light | Lock/ Unlock | Screen | Other | N Unique Features |
| Cao et al. | 2020 | Personal | Android (NR) | SOLVD | Location | X | X | | | | | | | | | | | 7 |
| | | | | | Activity | | | X | | | | | | | | | | |
| | | | | | Social | | | | | X | X | | | | | | | |
| | | | | | Sleep | | | | | | | | | X | | | | |
| | | | | | Phone Use | | | | | | | | | | | X | | |
| Chow et al. | 2017 | Personal | Android (≥4.3) | Sensus | Location | X | | | | | | | | | | | | 1 |
| Jacobson et al. | 2020 | Personal | Android (≥4.3) | Sensus | Activity | | X | | | | | | | | | | | 3 |
| | | | | | Social | | | | | X | X | | | | | | | |
| Knight & Bidargaddi | 2018 | Personal | NR | Pre-existing apps on users' phone | Activity | | X | | | | | | | | | | | 1 |
| MacLeod et al. | 2021 | Personal | IOS and Android (NR) | PROSIT | Location | X | | | | | | | | | | | | 5 |
| | | | | | Social | | | | | X | | | | | | | | |
| | | | | | Sleep | | | | | | | | | X | | | | |
| | | | | | Phone Use | | | | | | | | | | X | X | | |
| Melcher et al. | 2021 | Personal | IOS and Android (NR) | mindLAMP | Location | X | | | | | | | | | | | | 4 |
| | | | | | Social | | | | | X | | | | | | | | |
| | | | | | Sleep | | X | | | | | | | | | | | |
| | | | | | Phone Use | | | | | | | | | | | X | | |
| Rozgonjuk et al. | 2018 | Personal | IOS (≥10) | Moment | Phone Use | | | | | | | | | | X | X | | 2 |
| Shoval et al. | 2020 | Personal | Android (NR) | QualityTime (Mobi-days, Inc) | Sleep | | | | | | | | | | | | X*** | Unclear |

*Note.* Some phone specifications and sensors for studies using the same dataset or app (e.g., StudentLife) are assumed to be the same as those reported in the primary paper, when they are not explicitly reported in text. **Application or Platform Name –** PROSIT = Predicting Risk and Outcomes of Social Interactions. **Sensor Data Features Extracted –** GPS = Global Positioning System (GPS/WiFi includes GPS/phone location services, and/or WiFi receivers and/or cell towers); Accel = Accelerometer (includes phone core motion systems); BT = Bluetooth; Total = Total number of features explored in each study. *Social media use & network size. **Social networking via execution status of smartphone applications. ***App-based monitoring of phone activity including type of application, time spent using it, and time of night. Study authors created an index of whether phone was checked at night or not checked at night. **NR** = Not reported.

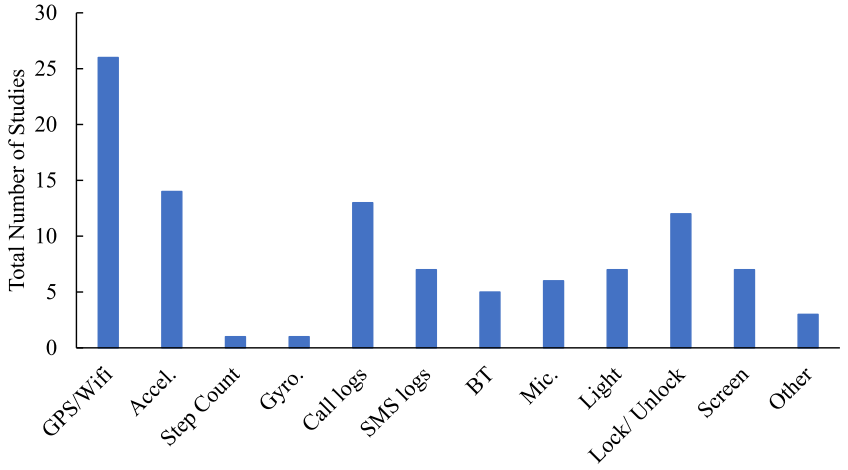



**Fig. 3.** Total number of studies using each source of phone data. *Note.* Accel. = Accelerometer; Gyro. = Gyroscope; BT=Bluetooth; Mic. = Microphone.

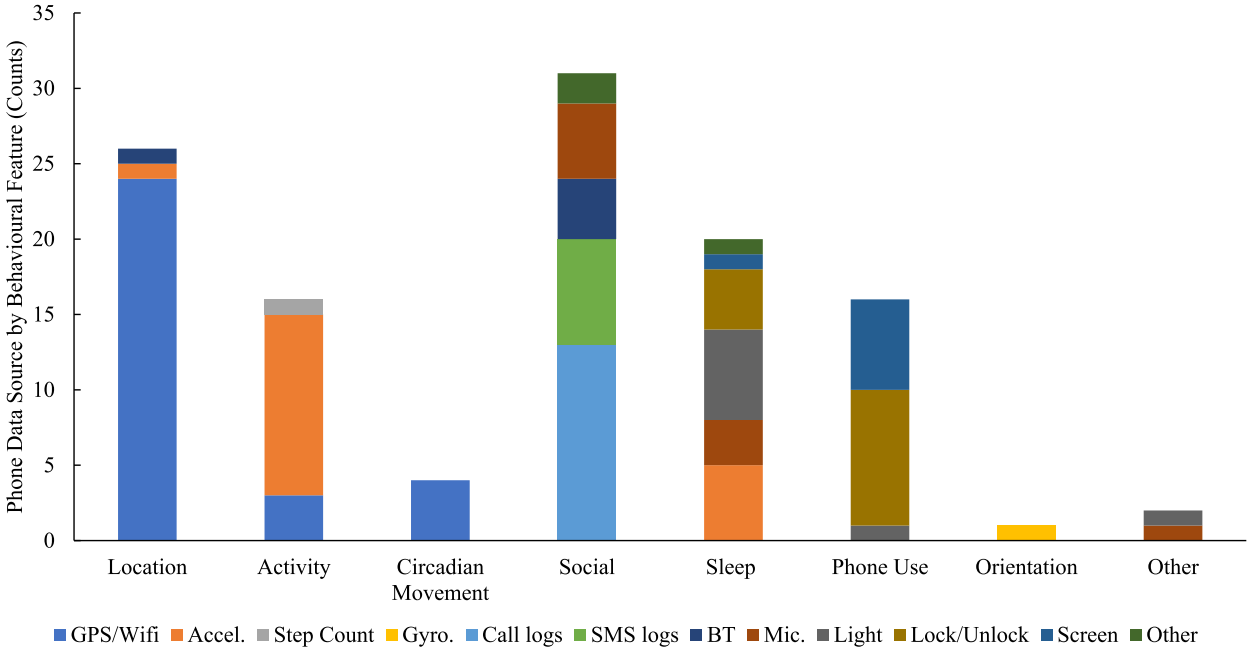



**Fig. 4.** Number of studies using each sensor type to infer high-level behavioural features. *Note.* Accel. = Accelerometer; Gyro. = Gyroscope; BT=Bluetooth; Mic. = Microphone.

only compared to studies that examined both (e.g., see Refs. [38,40,41]).

*Sociability.* Results generally showed that reduced sociability was associated with increased anxiety and depression symptoms. Significant sociability metrics associated with anxiety included fewer calls/texts in public places [42,50] and more motion variations when making calls [50]; those associated with depression included fewer/shorter daily conversations [36,47] and fewer daily co-locations [47]. Studies examining both anxiety and depression showed that relationships with low-level features were similar [35, 49], although one study found that shorter calls were associated with anxiety but not depression [38]. Two studies did not find any significant relationships between most of the low-level features used to infer sociability [41,48]. Other studies found that associations varied across time [37,45] and by gender [51].

*Phone Use.* Most studies did not find significant associations between screen time and anxiety [35,41,52] (see Ref. [49] for an exception) or between screen time and depression [41,49,52]. One study found that higher baseline severity was associated with decreased phone use over the one week monitoring period [53]. Others did find significant associations between phone use/screen time and depression, but the direction of effects were mixed [35,36,48,52]. Mean unlock duration in specific locations (e.g., in dorm

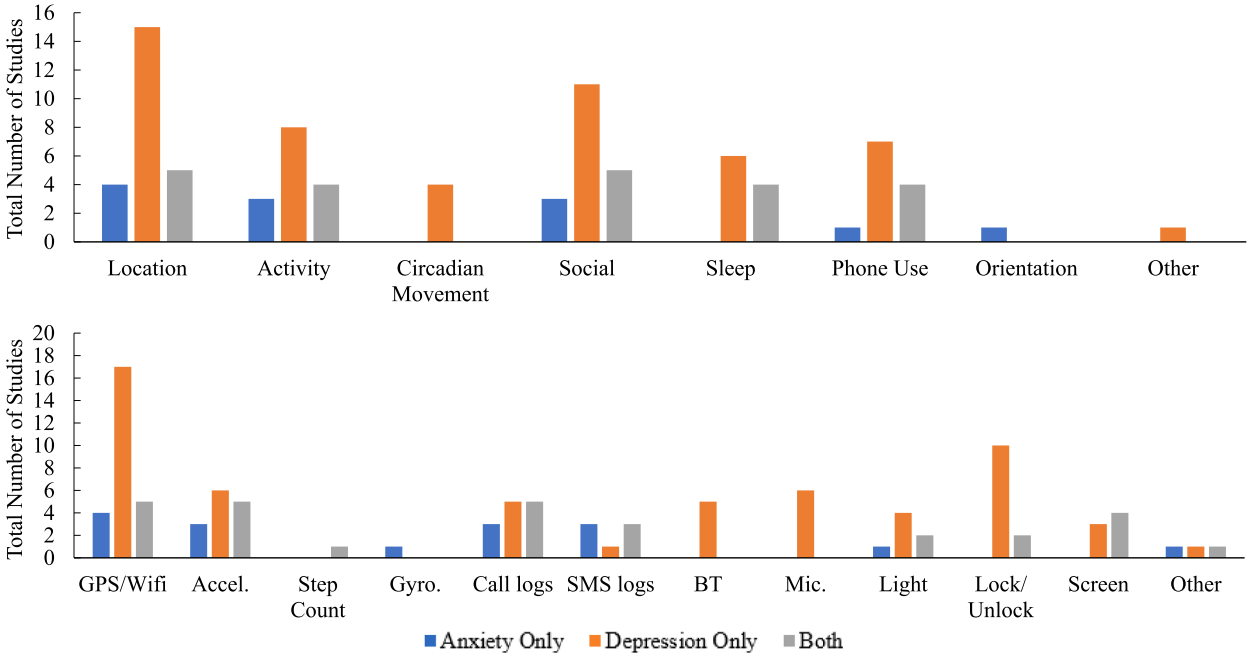



**Fig. 5.** Number of studies inferring high-level behaviours and number of studies using each type of phone sensor. *Note.* Top panel: Number of studies inferring each type of behaviour. Bottom Panel: Number of studies using each type of phone sensor. Accel = Accelerometer; Gyro. = Gyroscope; BT=Bluetooth; Mic. = Microphone.

rooms, at study places) and in general positively associated with depression [36,48], whereas averaged number of screen unlocks negatively associated with depression [52].

*Sleep.* Higher depression was typically associated with reduced duration of sleep [36,47], with one study showing that the relationship varied across time [45]. Higher depression, but not anxiety, was also associated with indices of poor sleep quality such as irregular sleep patterns [41,49]. Similarly, sleep duration irregularity predicted increased depression across time [46]. Anxiety, but not depression, was positively associated with phone checking at night [54] and ambient light intensity [35]. Another study found that including mobility, social interactions, phone use, and sleep-related features significantly improved the fit of models predicting more severe depression and more severe anxiety symptoms [49].

### 3.8. Machine learning results summary

Models using feature combinations (e.g., mobility, sociability, and sleep features) typically had better performance in predicting/detecting anxiety and depression or changes in symptoms than models using single features (e.g., mobility features). For example, a range of sensor data (GPS, accelerometer, steps, call, text, light, and/or screen) can accurately predict social anxiety [55] and depression [34,35,46,56], as well as discriminate anxiety from depression [55]. Another study used unsupervised mining techniques, finding that different combinations of location, sociability, and activity factors were associated with depression [57]. Further, combinations of different mobility features can predict anxiety levels and classify low versus high anxious groups of young people [39,44] and including communication features improves accuracy of these classification models [42]. Other feature combinations also facilitate prediction of anxiety changes, including phone/SMS logs, application execution status, light level, acceleration, and orientation [58]. Most studies focusing solely on anxiety did not identify influential features within best performing models. In comparison, seven studies focusing on depression identified influential features [31,36,48,59–62]. For example, one study found that the "best set" model for predicting depression included Bluetooth, calls, phone usage and steps, and the "best set" model for earliest prediction of change in depression symptoms needed data from weeks 1–2 [59]. Another study identified clusters of behavioural patterns that discriminated between low and high depression scores [60]. For example, participants with low depression scores (cluster 1) tended to have longer conversations, normal sleep patterns, spent less time in a quiet environment, and used their phone less than participants with high depression scores. Other studies explicitly focused on testing novel methodology in the context of youth mental health, such as anomaly detection, ideographically-weighted modelling, multi-task learning, or data fusion techniques [31,33,48,62–65] all of which provided promising results for future machine learning applications.

### 3.9. Quality assessment

See Table 3 for domain ratings for each study using the QA-DPSS. Overall, most studies included in the review were assessed to have

**Table 3**
Quality assessment ratings.

| Study | Domain 1 | Domain 2 | Domain 3 | Domain 4 | Domain 5 | Overall* |
|---|---|---|---|---|---|---|
| Ben-Zeev et al., 2015 | Moderate | Moderate | High | High | Low | High |
| Boukhechba, Chow, et al., 2018 | Moderate | Moderate | High | High | Low | High |
| Boukhechba, Daros, et al., 2018 | Moderate | Moderate | High | High | Low | High |
| Boukhechba, Huang, et al., 2017 | Moderate | Moderate | High | High | High | High |
| Cao et al., 2020 | Moderate | Moderate | High | Moderate | Low | High |
| Chikersal et al., 2021 | Low | Low | Moderate | High | High | High |
| Chow et al., 2017 | Moderate | Low | High | High | High | High |
| Demasi et al., 2016 | Moderate | Low | High | Moderate | High | High |
| Dissing et al., 2021 | High | Low | Moderate | High | High | High |
| Elhai et al., 2018 | High | Moderate | High | Moderate | High | High |
| Farhan, Lu, et al., 2016 | Low | Moderate | Moderate | Moderate | Low | Moderate |
| Farhan, Yue, et al., 2016 | Moderate | Low | High | Moderate | Low | High |
| Fukazawa et al., 2019 | Low | Moderate | High | Moderate | High | High |
| Gerych et al., 2019 | Low | Moderate | Moderate | Moderate | Low | Moderate |
| Gong et al., 2019 | Moderate | Low | High | High | Low | High |
| Huang et al., 2016 | High | Moderate | High | Moderate | High | High |
| Jacobson & Chung 2020 | Low | Moderate | High | Moderate | High | High |
| Jacobson et al., 2020 | Low | Moderate | Moderate | High | High | High |
| Kim et al., 2021 | Low | Low | High | Moderate | Low | High |
| Knight & Bidargaddi 2018 | High | Moderate | Moderate | Moderate | High | High |
| Li et al., 2017 | Moderate | Low | High | Moderate | Low | High |
| Lu et al., 2018 | Moderate | Low | High | High | Low | High |
| MacLeod et al., 2021 | Moderate | Low | Moderate | Moderate | Low | Moderate |
| Melcher et al., 2021 | Moderate | Low | High | High | High | High |
| Rozgonjuk et al., 2018 | High | Low | Low | Moderate | High | High |
| Saeb et al., 2016 | Moderate | Moderate | High | High | Low | High |
| Shoval et al., 2020 | Moderate | Moderate | Moderate | Moderate | High | Moderate |
| Wang, Chen et al., 2014 | Moderate | Moderate | High | High | Low | High |
| Wang, Wang, et al., 2018 | High | Moderate | Low | High | High | High |
| Ware, Yue, et al., 2019 | Low | Moderate | High | Moderate | Low | High |
| Xu, Chikersal, Doryab, et al., 2019 | Moderate | Moderate | High | High | High | High |
| Xu, Chikersal, Dutcher, et al., 2021 | Low | Moderate | Moderate | Moderate | High | Moderate |
| Yang, Mo, et al., 2017 | Low | Low | High | High | Low | High |
| Yang, Tang, et al., 2021 | High | Moderate | Moderate | High | High | High |
| Yue et al., 2017 | Moderate | Moderate | Moderate | High | High | High |

*Note.* *Overall rating only includes Domains 1–4. Domain 1: adequate reporting of digital sampling and data collection. Domain 2: adequate reporting of digital measurements. Domain 3: adequate reporting of digital data quality. Domain 4: adequate reporting of study analysis and results. Domain 5: ethics and safety reporting.

high risk of bias ($n = 30$ 85.71 %), with the remainder assessed to have moderate risk of bias ($n = 5$, 14.29 %). High risk of bias was most prevalent in the reporting of digital data quality (Domain 3; $n = 22$, 62.86 %) and reporting of analyses and results (Domain 4; $n = 18$, 54.29 %). Key contributors to these sources of bias were lack of reporting about the extent of missing data or adequate handling procedures, drop out during digital data collection, and inability to clarify whether analyses were a-priori and reported in full. Indeed, no study referenced a published protocol or registration detailing analyses plans. Further, no study included a power calculation or justification of sample size (Domain 1).

## 4. Discussion

The current scoping review aimed to summarise how phone sensor data has been used in the existing literature to predict and detect depression and anxiety in young people between 12 and 25 years-of-age. In accordance with Mohr's framework, we mapped out what phone sensors were used, what low-level features were extracted/computed, and what higher-level behavioural features were inferred from them. We also summarised analytical techniques and methodological quality, shedding light on reporting standards across disciplines.

### *4.1. What low-level features are extracted, combined, and used?*

Our findings demonstrate that a variety of low-level features were extracted and computed from smartphone sensors to infer behaviours related to youth anxiety and/or depression. For example, for accelerometer data, low-level features ranged from magnitude of acceleration, sum of all active/stationary periods per day, to variation of daily walking activity. Definitions of each low-level feature varied across studies. For example, magnitude of acceleration in one study was defined in terms of several descriptive features (e.g., mean, minimum) [38], as the lowest mean of acceleration at night-time in another study [41], and as specific speeds travelled in another study [58]. Features were also extracted in a wide range of epochs, including per hour, within a pre-defined time window (e.g., 8-h), per day, per week, or across the entire study. Of all sensors, GPS/Wifi data was the most ubiquitous, with most studies using a

cluster-based approach to infer mobility in specific semantic locations. The variation in feature engineering likely reflects the emerging nature of the field and, relatedly, the predominantly exploratory approaches used by researchers.

*StudentLife.* The variability in how data from smartphone sensors have been used to infer behaviours is demonstrated in studies published from the StudentLife dataset ($n = 8$). Studies computed different variables from the same data (e.g., location variance, average staying time, transition frequency, total duration of movement, call network size) and conducted different types of analyses (e.g., correlation, regression, supervised/unsupervised machine learning) with different mental health variables (e.g., pre, post, or pre-post change) over different timescales (e.g., 2-weeks, 10-weeks). For example, one study used an unsupervised machine learning approach that identified three behavioural clusters relating to conversations, sleep, and mobility (GPS, lock/unlock, microphone, light) that differentiated young people with low, medium, and high depression scores [60]. Another study used Support Vector Machines (SVM) to show that changes in sleep patterns can be detected from phone use metrics (lock/unlock) to predict the likelihood of depression [56]. Yet another study used a series of correlations between location and depression, finding that location variance, circadian movement, and entropy were negatively associated with depression [32]. It is difficult to create a unified story of what low-level features are best related to depression from these results. While there are benefits of a publicly available dataset that integrates multiple data sources, it can facilitate ad hoc, atheoretical approaches to data analyses.

### 4.2. What analytic approaches are used?

Analytic approaches included bivariate correlation, statistical regression, and machine learning techniques. Correlation/regressions were used to establish which individual features were related to anxiety and/or depression. Machine learning was used to examine single or combinations of low-level features and their association with anxiety and/or depression or their ability to classify participants into high or low symptom groups. Most studies used supervised approaches (classification, regression), with few using unsupervised approaches (e.g., clustering). Machine learning is a powerful tool for identifying unique combinations of digital features that best predict or detect changes in youth anxiety and depression.

### 4.3. What features are ubiquitously associated with youth depression and/or anxiety?

Consistent with prior research [15,22], combinations of low-level features typically had better performance in predicting and detecting youth anxiety and/or depression, compared to single low-level features. Combinations of low-level features may be more informative than single features given the heterogeneity of clinical phenotypes in anxiety and depression. Although mobility and communication patterns have the most supporting evidence, this likely reflects increased research attention compared to other high-level behaviours. Further exploration of feature combinations has promise for identifying new digital profiles that are temporally and contextually attuned to an individual's daily experiences.

Overall, evidence for the clinical value of phone sensor data is still emerging. For example, some research shows that mobility features can classify depression diagnoses in the absence of self-reported information [34]. There was not enough evidence to identify best predictors, or combinations of predictors, due to heterogeneity in methods and data analytics. It also remains unclear which low-level features are uniquely related to anxiety or depression, and given the lack of studies, whether they can reliably discriminate between these clinical states for an exception, see Ref. [55]. One source of heterogeneity is the operationalisation of low-level features in different studies (and in different analyses within a study). Preliminary support for this explanation comes from the fact that we were unable to identify a clear pattern of significant results even when descriptively comparing studies with shared design characteristics. For example, studies with a longer duration of sensing (i.e., >10-weeks) did not produce a systematically different pattern of results compared to studies with a shorter duration of sensing (≤10-weeks). The absence of a discernible pattern extended to studies that predicted changes in depression or anxiety over time, as well as studies examining correlations at a single time point. Similar to prior work [15], we suggest that our results can be used as a starting point to develop and test theoretically-driven hypotheses to advance the field.

### 4.4. What is the quality of studies?

This scoping review advances the field by developing and performing an initial test-case of a novel quality assessment for studies using smartphones to collect passive sensing data. Consistent with other reviews [15,18], our results indicate that the quality of studies was typically poor. Reporting of digital data quality and reporting of analyses and results were particularly problematic domains. Key contributors to these sources of bias were lack of reporting about the extent of missing data or adequate handling procedures, drop out during digital data collection, and inability to clarify whether analyses were a-priori and reported in full. No study referenced a published protocol or registration detailing analyses plans or provided a sample size justification. Although standard power analyses are not appropriate for machine-based learning analyses, justification for sample size in statistical approaches is important to explain researcher decision-making and facilitate transparency (especially when multiple analytic approaches are being conducted on the same dataset). These limitations threaten reproducibility and transparency, undermining the interpretability of results [15]. Transparency about which indicators are derived, by whom, and why is critical if the field is to offer meaningful contributions to the mental health of young people. In the next section, we demonstrate the importance of transparency by exploring the challenges of collecting sensing data from a range of devices in the field.

*Phone Specifications.* Studies did not adequately describe device confounders relating to hardware and software, which is particularly problematic when leveraging participants' own smartphones (92.00 % of cases). Device hardware can vary significantly. For

example, in the iPhone 13 specification, the gyroscope sensor is described as "Three-axis gyro," whereas in the Pixel 6 specification, the sensor is described as "Gyrometer." The lack of specificity around the specific sensor used and level of calibration in the studies can confound analysis, as the sensors could differ between devices of the different make and model, or possibly within devices of the same make and model with different manufacturing dates. Furthermore, sensing typically posits the assumption that the participant device is undamaged; a participant's device could have suffered from multiple drops which might thusly impede accurate sensor readings if damaged or moved from the impact.

Software differs from hardware in that developers can remotely push new updates to users, assuming connectivity, as opposed to hardware revisions which require issuing a product recall or a new device. In these software updates, there are three levels of abstraction which impact smart device sensing. Namely, the platform Operating System (OS), the Manufacturer-specific OS (MOS), and the study app, as well as their respective versions. The platform OS, such as Android and IOS, can introduce fundamental changes in how the sensor data are collected by the study app, which will have downstream consequences on the results of the study analysis. Above the base Android OS, there is the MOS, where device manufacturers offer tuned versions of base Android OS capabilities exclusive to their devices, such as with the Pixel Extreme Battery Saver mode. The implication of this is that background processes used to collect passive data from participant devices could be halted or ceased in order to preserve the device battery life at a more aggressive setting than with the base Android OS behaviour.

In addition to the device confounders of hardware and software permutations, participants may elect to change permission settings for their study app any time during the study period. This can lead to gaps in the data, thereby affecting the final sample size. Additional reporting is warranted around these specifications to characterise the data, improve confidence of findings reported, and generalisability to the target sample.

### *4.5. Theoretical implications*

Our findings align with Mohr et al.'s layered, hierarchical sensemaking framework of applying personal sensing to mental health [14]. What is needed now is a better understanding of the low-level features used to infer behaviours and, in turn, their relationship to clinical state. Extending the framework, emphasis could be placed on identifying which low-level features are transdiagnostic (i.e., those that relate to anxiety and depression), and which features are discriminatory (e.g., those that are uniquely related to anxiety or depression, or in early identification versus diagnostic detection).

### *4.6. Practical implications*

The lack of transparency and reproducibility highlighted by our review demonstrates the critical need for a standardised reporting instrument that aligns expectations and standards across different fields. The failure to report basic demographic/sample information (i.e., 77.15 % of studies did not report age, gender, and attrition), particularly in the computer science field, and the limited description in feature extraction and analysis overall, has important implications for the interpretation of findings. Along with other researchers in the field [15,22], we recommend the development of a common framework that standardises reporting of sample characteristics, phone specifications (including minimum and maximum OS versions), feature extraction and construction, missing data, analytic plans, and hypothesis testing. Standardised reporting is particularly important given the potential usefulness of exploratory methods to identify novel features or algorithms that better match higher-level behaviour. One generic framework for digital data processing and feature processing has been published for student data (code available on request) [66]. The primary aim of this framework is to facilitate replication of results. Developing a more comprehensive and prescriptive reporting instrument will help to guide future research and facilitate standardisation across different research groups and fields.

### *4.7. Limitations*

The findings of our scoping review must be interpreted in the context of some limitations. First, due to practical reasons, we only included studies that were published in English and we did not include grey literature or unpublished studies. Second, many studies included in our review conducted multiple analyses of data. In these cases, we prioritised findings that were presented as primary in the original study and/or that best aligned with our scoping review aims. It is possible that the findings we have reported are influenced by reporting bias, where we have emphasised significant findings over non-significant findings or selected some findings over others in the interests of brevity. Finally, several studies included in this review have overlapping samples because they leverage existing datasets. This means that some samples are overrepresented and that, when there is some similarity in analyses, some of the feature associations may be duplicated.

## 5. Future research

One promising area of future investigation is establishing feasibility of integrating passive phone sensor data with other types of data, including self-reported mental health status, clinician-rated information, cognitive functioning, ecological momentary assessment, health/medical records, and genetics data. While the current review focused on passive sensor data, other work suggests that combining different sources of data might improve accuracy of capturing emotions and behaviours [23]. This could lead to the development of advanced prediction tools that are more accurate than current indicators of youth mental health. We also developed a quality tool for studies using smartphones to collect digital data. This tool was designed to be relatively brief and can be used in

combination with more traditional tools to capture other issues with design (e.g., confounding and selection of participants, randomisation). Although a formal validation and item review by experts in the field was beyond the scope of our review, we welcome this in future use and iterations of the tool. Another understudied area in digital phenotyping for youth mental health is idiographic analyses. Group-level patterns in sensor data may not accurately reflect individual experiences (e.g., see Ref. [41]). Emphasizing single cases could enhance personalised mental health assessment and interventions, offering a more precise and clinically informative approach.

## 6. Conclusions

Digital phenotyping in youth mental health research is a new and challenging area that combines perspectives from psychiatry, technology, and health informatics [19]. Overall, there is little consensus in the literature about how to extract, combine, and use low-level features from phone sensors. There is emerging evidence that mobility and sociability features are related to youth anxiety and depression, which aligns with well-established clinical phenotypes. Additional research is needed on phone use, sleep, and circadian movement, as well as on exploring both anxiety and depression to identify unique or discriminatory features. We recommend the development of a standardised reporting framework for phone sensing studies in the mental health field to improve transparency and replicability of methodology.

## Ethics declaration

Review and/or approval by an ethics committee and informed consent were not needed for this study because it was a scoping review that involved secondary analyses of existing data.

## Data availability statement

No data was used for the research described in the article.

## CRediT authorship contribution statement

**Joanne R. Beames:** Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Jin Han:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Artur Shvetcov:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Wu Yi Zheng:** Writing – review & editing, Data curation. **Aimy Slade:** Writing – review & editing, Data curation. **Omar Dabash:** Writing – review & editing, Data curation. **Jodie Rosenberg:** Writing – review & editing, Data curation. **Bridianne O'Dea:** Writing – review & editing, Conceptualization. **Suranga Kasturi:** Writing – review & editing. **Leonard Hoon:** Writing – original draft. **Alexis E. Whitton:** Writing – review & editing. **Helen Christensen:** Writing – review & editing, Conceptualization. **Jill M. Newby:** Writing – review & editing, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by Commonwealth of Australia Medical Research Future Fund grant MRFAI000028 Optimising treatments in mental health using AI. The funding body had no role in any aspect of the study design or this manuscript. JRB is funded by a Marie Skłodowska-Curie fellowship (101063326). HC is funded by a NHMRC Senior Principal Research Fellowship 1155614. JN is funded by a NHMRC Investigator Grant 2008839. JH is supported by the Commonwealth Suicide Prevention Research Fund Post-Doctoral Fellowship.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e35472.

## References

[1] C. Kieling, H. Baker-Henningham, M. Belfer, G. Conti, I. Ertem, O. Omigbodun, L.A. Rohde, S. Srinath, N. Ulkuer, A. Rahman, Child and adolescent mental health worldwide: evidence for action, Lancet 378 (2011) 1515–1525, https://doi.org/10.1016/S0140-6736(11)60827-1.

[2] G.V. Polanczyk, G.A. Salum, L.S. Sugaya, A. Caye, L.A. Rohde, Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents, JCPP (J. Child Psychol. Psychiatry) 56 (2015) 345–365, https://doi.org/10.1111/jcpp.12381.

[3] K.R. Merikangas, J.-p. He, M. Burstein, S.A. Swanson, S. Avenevoli, L. Cui, C. Benjet, K. Georgiades, J. Swendsen, Lifetime prevalence of mental disorders in us adolescents: results from the national comorbidity survey replication–adolescent supplement (ncs-a), J. Am. Acad. Child Adolesc. Psychiatry 49 (2010) 980–989, https://doi.org/10.1016/j.jaac.2010.05.017.

[4] D. Lawrence, S. Johnson, J. Hafekost, K. Boterhoven De Haan, M. Sawyer, J. Ainley, S.R. Zubrick, The Mental Health of Children and Adolescents: Report on the Second Australian Child and Adolescent Survey of Mental Health and Wellbeing, Department of Health, Canberra, 2015.

[5] M. Solmi, J. Radua, M. Olivola, E. Croce, L. Soardo, G. Salazar de Pablo, J. Il Shin, J.B. Kirkbride, P. Jones, J.H. Kim, J.Y. Kim, A.F. Carvalho, M.V. Seeman, C. U. Correll, P. Fusar-Poli, Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies, Mol. Psychiatr. 27 (2022) 281–295, https://doi.org/10.1038/s41380-021-01161-7.

[6] J. Breslau, E. Miller, N. Breslau, K. Bohnert, V. Lucia, J. Schweitzer, The impact of early behavior disturbances on academic achievement in high school, Pediatrics 123 (2009) 1472–1476.

[7] D. Johnson, G. Dupuis, J. Piche, Z. Clayborne, I. Colman, Adult mental health outcomes of adolescent depression: a systematic review, Depress. Anxiety 35 (2018) 700–716, https://doi.org/10.1002/da.22777.

[8] U. Jonsson, H. Bohman, L. von Knorring, G. Olsson, A. Paaren, A.L. von Knorring, Mental health outcome of long-term and episodic adolescent depression: 15-year follow-up of a community sample, J. Affect. Disord. 130 (2011) 395–404, https://doi.org/10.1016/j.jad.2010.10.046.

[9] M.C. Dekker, R.F. Ferdinand, N.D. Van Lang, I.L. Bongers, J. Van Der Ende, F.C. Verhulst, Developmental trajectories of depressive symptoms from early childhood to late adolescence: gender differences and adult outcome, JCPP (J. Child Psychol. Psychiatry) 48 (2007) 657–666.

[10] A. Goodman, R. Joyce, J.P. Smith, The long shadow cast by childhood physical and mental problems on adult life, Proc. Natl. Acad. Sci. USA 108 (2011) 6032–6037, https://doi.org/10.1073/pnas.1016970108.

[11] T.R. Insel, Digital phenotyping: technology for a new science of behavior, JAMA 318 (2017) 1215–1216, https://doi.org/10.1001/jama.2017.11295.

[12] J. Torous, M.V. Kiang, J. Lorme, J.-P. Onnela, New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research, JMIR Ment. Health 3 (2016) e16, https://doi.org/10.2196/mental.5165.

[13] K. Huckvale, S. Venkatesh, H. Christensen, Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety, NPJ Digit. Med. 2 (2019) 88, https://doi.org/10.1038/s41746-019-0166-1.

[14] D.C. Mohr, M. Zhang, S.M. Schueller, Personal sensing: understanding mental health using ubiquitous sensors and machine learning, Annu. Rev. Clin. Psychol. 13 (2017) 23–47, https://doi.org/10.1146/annurev-clinpsy-032816-044949.

[15] V. De Angel, S. Lewis, K. White, C. Oetzmann, D. Leightley, E. Oprea, G. Lavelle, F. Matcham, A. Pace, D.C. Mohr, R. Dobson, M. Hotopf, Digital health tools for the passive monitoring of depression: a systematic review of methods, NPJ Digit. Med. 5 (2022) 3, https://doi.org/10.1038/s41746-021-00548-8.

[16] P. Spinazze, Y. Rykov, A. Bottle, J. Car, Digital phenotyping for assessment and prediction of mental health outcomes: a scoping review protocol, BMJ Open 9 (2019) e032255, https://doi.org/10.1136/bmjopen-2019-032255.

[17] T.J. Trull, U.W. Ebner-Priemer, Ambulatory assessment in psychopathology research: a review of recommended reporting guidelines and current practices, J. Abnorm. Psychol. 129 (2020) 56–63, https://doi.org/10.1037/abn0000473.

[18] J. Melcher, R. Hays, J. Torous, Digital phenotyping for mental health of college students: a clinical review, Evid. Base Ment. Health 23 (2020) 161–166, https://doi.org/10.1136/ebmental-2020-300180.

[19] J.-P. Onnela, S.L. Rauch, Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health, Neuropsychopharmacology 41 (2016) 1691–1696, https://doi.org/10.1038/npp.2016.7.

[20] B. Wies, C. Landers, M. Ienca, Digital mental health for young people: a scoping review of ethical promises and challenges, Front. Digit. Health 3 (2021), https://doi.org/10.3389/fdgth.2021.697072.

[21] J. Rooksby, A. Morrison, D. Murray-Rust, Student perspectives on digital phenotyping: the acceptability of using smartphone data to assess mental health, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1–14, https://doi.org/10.1145/3290605.3300655.

[22] D.A. Rohani, M. Faurholt-Jepsen, L.V. Kessing, J.E. Bardram, Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review, JMIR Mhealth Uhealth 6 (2018) e165, https://doi.org/10.2196/mhealth.9691.

[23] M. Nisenson, V. Lin, M. Gansner, Digital phenotyping in child and adolescent psychiatry: a perspective, Harv. Rev. Psychiatr. 29 (2021), https://doi.org/10.1097/HRP.0000000000000310.

[24] M.C. Angelides, L.A.C. Wilson, P.L.B. Echeverría, Wearable data analysis, visualisation and recommendations on the go using android middleware, Multimed. Tool. Appl. 77 (2018) 26397–26448, https://doi.org/10.1007/s11042-018-5867-y.

[25] F. de Arriba-Pérez, M. Caeiro-Rodríguez, J.M. Santos-Gago, Collection and processing of data from wrist wearable devices in heterogeneous and multiple-user scenarios, Sensors 16 (2016), https://doi.org/10.3390/s16091538.

[26] L.F. Ramos-Lima, V. Waikamp, T. Antonelli-Salgado, I.C. Passos, L.H.M. Freitas, The use of machine learning techniques in trauma-related disorders: a systematic review, J. Psychiatr. Res. 121 (2020) 159–172, https://doi.org/10.1016/j.jpsychires.2019.12.001.

[27] M.D.J. Peters, C. Marnie, A.C. Tricco, D. Pollock, Z. Munn, L. Alexander, P. McInerney, C.M. Godfrey, H. Khalil, Updated methodological guidance for the conduct of scoping reviews, JBI Evid Synth 18 (2020) 2119–2126, https://doi.org/10.11124/jbies-20-00167.

[28] A.C. Tricco, E. Lillie, W. Zarin, K.K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M.D.J. Peters, T. Horsley, L. Weeks, S. Hempel, E.A. Akl, C. Chang, J. McGowan, L. Stewart, L. Hartling, A. Aldcroft, M.G. Wilson, C. Garritty, S. Lewin, C.M. Godfrey, M.T. Macdonald, E.V. Langlois, K. Soares-Weiser, J. Moriarty, T. Clifford, Ö. Tunçalp, S.E. Straus, Prisma extension for scoping reviews (prisma-scr): checklist and explanation, Ann. Intern. Med. 169 (2018) 467–473, https://doi.org/10.7326/m18-0850.

[29] H. Arksey, L. O'Malley, Scoping studies: towards a methodological framework, Int. J. Soc. Res. Methodol. 8 (2005) 19–32, https://doi.org/10.1080/1364557032000119616.

[30] Covidence Systematic Review Software, 2023. Melbourne, Australia, www.covidence.org.

[31] J. Lu, C. Shang, C. Yue, R. Morillo, S. Ware, J. Kamath, A. Bamis, A. Russell, B. Wang, J. Bi, Joint modeling of heterogeneous sensing data for depression assessment via multi-task learning, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2 (2018) 1–21, https://doi.org/10.1145/3191753.

[32] S. Saeb, E.G. Lattie, S.M. Schueller, K.P. Kording, D.C. Mohr, The relationship between mobile phone location sensor data and depressive symptom severity, PeerJ 4 (2016) e2537, https://doi.org/10.7717/peerj.2537.

[33] C. Yue, S. Ware, R. Morillo, J. Lu, C. Shang, J. Bi, J. Kamath, A. Russell, A. Bamis, B. Wang, Fusing location data for depression prediction, IEEE Transactions on Big Data 7 (2017) 355–370, https://doi.org/10.1109/UIC-ATC.2017.8397515.

[34] A.A. Farhan, C. Yue, R. Morillo, S. Ware, J. Lu, J. Bi, J. Kamath, A. Russell, A. Bamis, B. Wang, Behavior vs. Introspection: Refining Prediction of Clinical Depression via Smartphone Sensing Data, IEEE, 2016, pp. 1–8, https://doi.org/10.1109/WH.2016.7764553.

[35] J. Cao, T. Anh Lan, S. Banu, A.A. Shah, A. Sabharwal, N. Moukaddam, Tracking and predicting depressive symptoms of adolescents using smartphone-based self-reports, parental evaluations, and passive phone sensor data: development and usability study, JMIR Ment Health 7 (2020), https://doi.org/10.2196/14045.

[36] R. Wang, W. Wang, A. daSilva, J.F. Huckins, W.M. Kelley, T.F. Heatherton, A.T. Campbell, Tracking depression dynamics in college students using mobile phone and wearable sensing, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2 (2018) 1–26, https://doi.org/10.1145/3191775.

[37] T.M.H. Li, C.-T. Li, P.W.C. Wong, J. Cao, Withdrawal behaviors and mental health among college students, Behavioral Psychology/Psicología Conductual: Revista Internacional Clínica y de la Salud 25 (2017) 99–109.

[38] M. Boukhechba, A.R. Daros, K. Fua, P.I. Chow, B.A. Teachman, L.E. Barnes, Demonicsalmon: monitoring mental health and social interactions of college students using smartphones, Smart Health 9–10 (2018) 192–203, https://doi.org/10.1016/j.smhl.2018.07.005.

[39] M. Boukhechba, P. Chow, K. Fua, B.A. Teachman, L.E. Barnes, Predicting social anxiety from global positioning system traces of college students: feasibility study, JMIR Ment Health 5 (2018) e10101, https://doi.org/10.2196/10101.

[40] P.I. Chow, K. Fua, Y. Huang, W. Bonelli, H. Xiong, L.E. Barnes, B.A. Teachman, Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students, J. Med. Internet Res. 19 (2017) e62, https://doi.org/10.2196/jmir.6820.
[41] J. Melcher, J. Lavoie, R. Hays, R. D'Mello, N. Rauseo-Ricupero, E. Camacho, E. Rodriguez-Villa, H. Wisniewski, S. Lagan, A. Vaidyam, J. Torous, Digital phenotyping of student mental health during covid-19: an observational study of 100 college students, J. Am. Coll. Health (2021), https://doi.org/10.1080/07448481.2021.1905650.
[42] M. Boukhechba, Y. Huang, P. Chow, K. Fua, B.A. Teachman, L.E. Barnes, Monitoring social anxiety from mobility and communication patterns, in: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, 2017, pp. 749–753, https://doi.org/10.1145/3123024.3125607.
[43] A. Knight, N. Bidargaddi, Commonly available activity tracker apps and wearables as a mental health outcome indicator: a prospective observational cohort study among young adults with psychological distress, J. Affect. Disord. 236 (2018) 31–36, https://doi.org/10.1016/j.jad.2018.04.099.
[44] Y. Huang, H. Xiong, K. Leach, Y. Zhang, P. Chow, K. Fua, B.A. Teachman, L.E. Barnes, Assessing social anxiety using gps trajectories and point-of-interest data, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2016, pp. 898–903, https://doi.org/10.1145/2971648.2971761.
[45] D. Ben-Zeev, E.A. Scherer, R. Wang, H. Xie, A.T. Campbell, Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health, Psychiatr. Rehabil. J. 38 (2015) 218–226, https://doi.org/10.1037/prj0000130.
[46] O. Demasi, A. Aguilera, B. Recht, Detecting change in depressive symptoms from daily wellbeing questions, personality, and activity, IEEE (2016) 1–8, https://doi.org/10.1109/WH.2016.7764552.
[47] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, A.T. Campbell, Assoc Comp, M. 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, 2014, pp. 3–14, https://doi.org/10.1145/2632048.2632054.
[48] X. Xu, P. Chikersal, A. Doryab, D.K. Villalba, J.M. Dutcher, M.J. Tumminia, T. Althoff, S. Cohen, K.G. Creswell, D.J. Creswell, J. Mankoff, A.K. Dey, Leveraging routine behavior and contextually-filtered features for depression detection among college students, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3 (2019) 1–33, https://doi.org/10.1145/3351274.
[49] L. MacLeod, B. Suruliraj, D. Gall, K. Bessenyei, S. Hamm, I. Romkey, A. Bagnell, M. Mattheisen, V. Muthukumaraswamy, R. Orji, S. Meier, A mobile sensing app to monitor youth mental health: observational pilot study, JMIR Mhealth Uhealth 9 (2021) e20638, https://doi.org/10.2196/20638.
[50] J. Gong, Y. Huang, P.I. Chow, K. Fua, M.S. Gerber, B.A. Teachman, L.E. Barnes, Understanding behavioral dynamics of social anxiety among college students through smartphone sensors, Inf. Fusion 49 (2019) 57–68, https://doi.org/10.1016/j.inffus.2018.09.002.
[51] A.S. Dissing, N. Hulvej Rod, T.A. Gerds, R. Lund, Smartphone interactions and mental well-being in young adults: a longitudinal study based on objective high-resolution smartphone data, Scand. J. Publ. Health 49 (2021) 325–332, https://doi.org/10.1177/1403494820920418.
[52] D. Rozgonjuk, J.C. Levine, B.J. Hall, J.D. Elhai, The association between problematic smartphone use, depression and anxiety symptom severity, and objectively measured smartphone use over one week, Comput. Hum. Behav. 87 (2018) 10–17.
[53] J.D. Elhai, M.F. Tiamiyu, J.W. Weeks, J.C. Levine, K.J. Picard, B.J. Hall, Depression and emotion regulation predict objective smartphone use measured over one week, Pers. Indiv. Differ. 133 (2018) 21–28.
[54] D. Shoval, N. Tal, O. Tzischinsky, Relationship of smartphone use at night with sleep quality and psychological well-being among healthy students: a pilot study, Sleep Health 6 (2020) 495–497, https://doi.org/10.1016/j.sleh.2020.01.011.
[55] N.C. Jacobson, B. Summers, S. Wilhelm, Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors, J. Med. Internet Res. 22 (2020) e16875, https://doi.org/10.2196/16875.
[56] J. Kim, J. Hong, Y. Choi, Automatic depression prediction using screen lock/unlock data on the smartphone, in: 18th International Conference on Ubiquitous Robots (UR), 2021, pp. 1–4, https://doi.org/10.1109/UR52253.2021.9494673.
[57] Z. Yang, X. Mo, D. Shi, R. Wang, Mining relationships between mental health, academic performance and human behaviour, in: 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation, SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2017, pp. 1–8, https://doi.org/10.1109/UIC-ATC.2017.8397544.
[58] Y. Fukazawa, T. Ito, T. Okimura, Y. Yamashita, T. Maeda, J. Ota, Predicting anxiety state using smartphone-based passive sensing, J. Biomed. Inf. 93 (2019) 103151, https://doi.org/10.1016/j.jbi.2019.103151.
[59] P. Chikersal, A. Doryab, M. Tumminia, D.K. Villalba, J.M. Dutcher, X. Liu, S. Cohen, K.G. Creswell, J. Mankoff, D.J. Creswell, M. Goel, A.K. Dey, Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing: a machine learning approach with robust feature selection, ACM Trans. Comput. Hum. Interact. 28 (2021) 1–41, https://doi.org/10.1145/3422821.
[60] A.A. Farhan, J. Lu, J. Bi, A. Russell, B. Wang, A. Bamis, Multi-view bi-clustering to identify smartphone sensing features indicative of depression, in: IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2016, pp. 264–273, https://doi.org/10.1109/CHASE.2016.27.
[61] S. Ware, C. Yue, R. Morillo, J. Lu, C. Shang, J. Bi, J. Kamath, A. Russell, A. Bamis, B. Wang, Predicting depressive symptoms using smartphone data, Smart Health 15 (2020), https://doi.org/10.1016/j.smhl.2019.100093.
[62] X. Xu, P. Chikersal, J.M. Dutcher, Y.S. Sefidgar, W. Seo, M.J. Tumminia, D.K. Villalba, S. Cohen, K.G. Creswell, D.J. Creswell, A. Doryab, P.S. Nurius, E. Riskin, A.K. Dey, J. Mankoff, Leveraging collaborative-filtering for personalized behavior modeling: a case study of depression detection among college students, in: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 5, 2021, pp. 1–27, https://doi.org/10.1145/3448107.
[63] W. Gerych, E. Agu, E. Rundensteiner, Classifying depression in imbalanced datasets using an autoencoder- based anomaly detection approach, in: IEEE 13th International Conference on Semantic Computing (ICSC), 2019, pp. 124–127, https://doi.org/10.1109/ICOSC.2019.8665535.
[64] N.C. Jacobson, Y.J. Chung, Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones, Sensors 20 (2020), https://doi.org/10.3390/s20123572.
[65] M. Yang, J. Tang, Y. Wu, Z. Liu, X. Hu, B. Hu, A behaviour patterns extraction method for recognizing generalized anxiety disorder, in: 2020 IEEE International Conference on E-Health Networking, Application & Services (HEALTHCOM), 2021, pp. 1–4, https://doi.org/10.1109/HEALTHCOM49281.2021.9398995.
[66] A. Doryab, P. Chikarsel, X. Liu, A.K. Dey, Extraction of Behavioral Features from Smartphone and Wearable Data, 2018 *arXiv preprint arXiv:1812.10394*.