

# Smartphone-Based Digital Phenotyping Across Health Conditions: A Scoping Review

Arlen Dumas, Joanne Hokayem, Georgia Goodman, Krishna Venkatasubramanian, Peter Chai

Submitted to: Journal of Medical Internet Research  
on: September 15, 2025

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

## *Table of Contents*

---

<b>Original Manuscript</b> .....	<b>5</b>
<b>Supplementary Files</b> .....	<b>41</b>
0.....	41
<b>Figures</b> .....	<b>42</b>
Figure 1.....	43
Figure 2.....	44
Figure 3.....	45
<b>Multimedia Appendixes</b> .....	<b>46</b>
Multimedia Appendix 1.....	47

Preprint  
JMIR Publications

# Smartphone-Based Digital Phenotyping Across Health Conditions: A Scoping Review

Arlen Dumas<sup>1</sup>; Joanne Hokayem<sup>2</sup>; Georgia Goodman<sup>3</sup>; Krishna Venkatasubramanian<sup>1</sup> PhD; Peter Chai<sup>2</sup> MD

<sup>1</sup> Department of Computer Science and Statistics University of Rhode Island Kingston US

<sup>2</sup> The Fenway Institute Boston US

<sup>3</sup> UCLA Health Los Angeles US

## Corresponding Author:

Krishna Venkatasubramanian PhD

Department of Computer Science and Statistics  
University of Rhode Island  
9 Greenhouse Road  
Kingston  
US

## Abstract

**Background:** Smartphone-based digital phenotyping uses built-in sensors and usage patterns to passively capture behavioral and environmental data relevant to a person's health. This practice has been applied extensively in the context of mental health and chronic disease management.

**Objective:** This review aims to synthesize digital phenotyping articles that are solely smartphone-based. That is, they use data exclusively from onboard sensors on smartphones to characterize specific health conditions.

**Methods:** We conducted a scoping review of English-language, peer-reviewed articles published between 2012 and 2024 in Google Scholar, IEEE Xplore, ACM Digital Library, and PubMed using terms such as "mobile sensing" and "digital phenotyping." Eligible articles used onboard smartphone sensors to assess health, allowed validation with other devices, and went beyond self-report (e.g., ecological momentary assessments). We excluded articles that focused solely on algorithm development or data capture without linking results to a health context.

**Results:** The search identified 111 articles, of which 65 articles met the inclusion criteria. Most articles described observational and used passive sensing. Mental health conditions were the most frequently examined health conditions. These included: depression (n = 17), stress or anxiety (n = 14), bipolar disorder (n = 11) and schizophrenia (n = 8). Other, relatively less common conditions included: substance use disorders (n = 7), Parkinson's disease (n = 4), and sleep apnea (n = 2). The smartphone sensor data used in these articles were: screen state or device usage logs (n = 35), GPS (n = 34), accelerometer (n = 26), microphone data (n = 18), call logs (n = 21), app usage statistics (n = 15), WiFi connectivity (n = 5), Bluetooth (n = 4), battery state (n = 3), and keystroke dynamics (n = 2). Ground-truth measurements relied on validated clinical scales (e.g., PHQ-9, GAD-7, YMRS, PSQI) (n = 41), followed by ecological momentary assessments (n = 18), clinician-confirmed diagnoses (n = 9), and physiological measures, such as polysomnography (n = 3). In terms of study focus, articles could be categorized into monitoring symptoms (n = 43), diagnostic applications (n = 15), and intervention strategies (n = 7).

Overall in these articles presented several methodological gaps. These included: (1) inconsistent reporting of sensor streams, (2) limited data quality descriptions, and (3) lack of standard in outcome validation. These gaps underscore the need for standardized reporting and greater data availability to improve reproducibility and comparability in the smartphone-based digital phenotyping space.

**Conclusions:** This review shows that smartphone-based digital phenotyping is a versatile tool for understanding health-related behaviors across diverse conditions. Most research targets mental health, but work also extends to other disorders. Future work should expand applications, standardize reporting, improve rigor, and develop shared datasets to advance the field. Clinical Trial: N/A

(JMIR Preprints 15/09/2025:84146)

DOI: <https://doi.org/10.2196/preprints.84146>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [JMIR Publications](#)

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in [JMIR Publications](#)

## Original Manuscript

Preprint  
JMIR Publications

# Smartphone-Based Digital Phenotyping Across Health Conditions: A Scoping Review

Arlen Dumas<sup>1</sup>, Joanne Hokayem<sup>2</sup>, Georgia Goodman<sup>3</sup>, Krishna Venkatasubramanian<sup>1</sup>, Peter R. Chai<sup>2</sup>

<sup>1</sup>Department of Computer Science and Statistics, The University of Rhode Island, Kingston, RI, USA

<sup>2</sup>The Fenway Institute, Boston, MA, USA

<sup>3</sup>UCLA Health, Los Angeles, CA, USA

ardumas@uri.edu; jhokayem2000@gmail.com; ggoodman96@gmail.com; krish@uri.edu; pchai@bwh.harvard.edu Corresponding author

Krishna Venkatasubramanian, krish@uri.edu

## Abstract

### Background:

Smartphone-based digital phenotyping uses built-in sensors and usage patterns to passively capture behavioral and environmental data relevant to health and has been applied extensively in mental health and chronic disease contexts.

### Objective:

This review synthesizes studies that use *smartphone-based digital phenotyping*, defined as approaches that rely exclusively on onboard smartphone sensors to characterize specific health conditions. To our knowledge, this work provides the most comprehensive cross-condition synthesis of smartphone-based digital phenotyping to date, spanning mental health, physical health, and substance use disorders, and highlighting common practices, gaps, and opportunities for future research.

### Methods:

We conducted a scoping review of English-language, peer-reviewed articles published between 2012 and 2025 in Google Scholar, IEEE Xplore, ACM Digital Library, and PubMed using terms such as “mobile sensing” and “digital phenotyping.” Eligible articles used onboard smartphone sensors to assess health and went beyond self-report. Studies that relied on auxiliary sensing modalities outside of smartphones or did not use digital phenotyping to characterize one or more well-defined health conditions were excluded.

### Results:

We performed a descriptive synthesis of study characteristics, sensors, and health domains. Of 111 articles identified, 65 met inclusion criteria. Most studies were observational and relied on passive sensing. Sample sizes ranged from fewer than 10 to over 18,000 participants, with a median of 52. Mental health conditions were most frequently examined, including depression ( $n = 16$ ), bipolar disorder ( $n = 11$ ), stress or anxiety ( $n=10$ ), and schizophrenia ( $n=8$ ). Less commonly studied conditions included substance use disorders ( $n = 7$ ), Parkinson’s disease ( $n = 4$ ), and sleep apnea ( $n = 2$ ). Sensor streams varied widely and included diverse passive smartphone data sources capturing mobility, communication, device usage, environmental context, and user interaction patterns. Ground-truth

measurements most commonly relied on validated clinical scales (e.g., PHQ-9, YMRS, PSQI) ( $n = 41$ ), followed by ecological momentary assessments ( $n = 18$ ), clinician-confirmed diagnoses ( $n = 9$ ), and physiological measures such as polysomnography ( $n=3$ ). Across studies, recurring methodological limitations included incomplete or inconsistent sensor descriptions, limited reporting of data quality (e.g., sampling rates, missingness), and heterogeneous validation practices. These issues limit comparability and reproducibility and underscore the need for clearer reporting standards and greater data availability.

#### Conclusions:

This scoping review provides the first comprehensive synthesis of smartphone-only digital phenotyping studies spanning mental health, physical health, and substance use disorders. Unlike prior reviews that focus on specific conditions or incorporate auxiliary sensing technologies, this work maps behavioral associations derived exclusively from smartphone sensors across a broad range of health domains. The primary contribution of this review lies in its consolidation of behavioral associations observed across studies, enabling researchers to correlate new findings to the existing evidence base and identify opportunities for replication, extension, or clinical translation. Collectively, these findings highlight both the promise of smartphone-based digital phenotyping in real-world settings and the need for improved standardization to support translation into clinical and public health applications.

Keywords: digital phenotyping; smartphones; digital health; scoping review

## Introduction

Since their introduction in 1994, smartphones have penetrated almost every aspect of human life globally [1]. These devices are ubiquitous in many cultures and have transformed access to the internet and communication with others. Like many other sectors, healthcare has been directly impacted by the use of smartphones [2, 3]. Patients can now conduct virtual visits with their healthcare providers [4], have access to a wide range of sources of medical information, and can access health records on smartphones [3]. Recent survey data in the US suggests that 62% of individuals would like to access their medical records through mobile-facing apps, and in 2022, 48% of patients accessed medical records via the internet at least once [5]. Additionally, increased use of wearable and other health-related tracking devices have positioned smartphones as a centralized hub for health data and behavior tracking [6, 7]. This data suggests that individuals access and use smartphones to assess and modulate individual health, positioning these devices as important platforms to interact with individuals [6, 8].

One potential strategy to leverage smartphones to understand health behavior is through digital phenotyping, a strategy that leverages data from mobile computing devices that most people now have with them (e.g., smartphones, smartwatches, rings, and other wearable devices) to capture a person's behavior and through them understand any health conditions present [9]. More specifically, digital phenotyping is the collection and analysis of data generated through human interaction with mobile computing devices. It encompasses: (1) *passive input* collected from sensors on a smartphone, such as continuous GPS monitoring, accelerometry data and (2) *active input*, such as ecological momentary assessments (EMAs) completed by the user of the smartphone. By integrating such passive and active inputs, digital phenotyping aims to explore behaviors that correlate with specific health conditions [10].

Smartphone-based digital phenotyping is a form of digital phenotyping in which

smartphones are the only devices used for collecting sensor data from individuals. Smartphones offer a unique advantage due to their widespread use, enabling researchers to collect a rich, high-volume dataset that can be used to assess behaviors (also known as *phenotypes*) in the context of a specific health condition in an unobtrusive way [11]. Data that has been captured via smartphones can be analyzed through a variety of methods to provide insight into a variety of mental and physical health conditions [12]. Compared to traditional approaches of monitoring and diagnosing complex health conditions (e.g., schizophrenia and bipolar disorder) that require considerable human interventions, smartphone-based digital phenotyping offers an opportunity to passively track these conditions over time [13]. The other potential advantage of smartphone-based digital phenotyping is the opportunity to democratize diagnoses of various health conditions simply by assessing smartphone use parameters in context of the user's known actions [14]. Despite its promise, multiple challenges exist to operationalize smartphone-based digital phenotyping, account for variability in data and generalize findings across patient populations [9].

This scoping review aims to comprehensively describe the current state of smartphone-based digital phenotyping work. Our scope is more expansive than existing scoping reviews as it covers any and all health conditions that have been explored as part of smartphone-based phenotyping. By exploring the breadth of health conditions studied by smartphone-based digital phenotyping, we can: (1) show the versatility of the smartphone-based digital phenotyping to address the needs of whole variety of health conditions and (2) obtain a broader understanding of how smartphone-based digital phenotyping can be used across different clinical contexts. Further, in this work, in keeping with the spirit of digital phenotyping, we describe the relationship between smartphone data and behaviors that are exhibited in the presence of specific medical conditions for the works reviewed. We avoid going into algorithmic or machine learning methods in this scoping review as that can sometimes promote the discovery without understanding mindset when looking at prior work. As the majority of the studies reviewed are observational, the relationships reported between smartphone-derived behaviors and health conditions should be interpreted as associational rather than causal.

This scoping review aims to synthesize the existing literature on smartphone-only digital phenotyping. Specifically, we focused on studies that *exclusively* use smartphone sensors to characterize health conditions. Our objectives are to consolidate evidence across health domains, describe how behavioral signals are captured and assessed using different smartphone sensor streams, and to identify recurring methodological limitations and gaps in existing literature. By taking a cross-condition perspective, this review clarifies where evidence is concentrated, where it remains sparse, and how future research and better support replication, comparability, and translation into clinical and public health settings.

## Methods

We conducted a scoping review to understand the current state of smartphone-based digital phenotyping. No review protocol was registered for this study. Methods adhered to PRISMA-ScR guidelines [15]. [The search strategy was reported in accordance with PRISMAS guidelines \[16\].](#) We systematically queried Google Scholar, IEEE Xplore, ACM Digital Library, and PubMed for peer-reviewed manuscripts published between 2012 and October 2025. The revised search strategy yielded approximately 3,700 records across the four databases prior to screening. Search terms included "mobile sensing", "digital phenotyping", "smartphone sensing", "personal sensing" and "smartphone assessment.". Terms were developed through discussion among the study team consisting of experts in digital phenotyping, digital health technologies, behavioral medicine, and computer science. A sample PubMed search string is as

follows: ("digital phenotyping" OR "smartphone sensing" OR "mobile sensing") AND (mental OR behavioral OR health OR monitoring). Full search strategies for each database are provided in Appendix A. In addition to database searching, we used backward reference screening of included articles to identify additional studies; we did not search study registries or gray literature. The search strategy followed PRISMA-S guidelines and was informed by prior reviews in digital phenotyping and mobile sensing. We did not utilize any standardized search filters.

All records retrieved from databases were exported or saved in PDF format from their respective databases and managed manually. Prior to formal screening, records were manually reviewed to remove duplicate entries and studies that were clearly outside the scope of the review. A substantial proportion of retrieved records were excluded at this stage due to the use of wearable or auxiliary sensors, non-health-related applications, review articles, study protocols, or a primary focus on algorithmic development rather than health characterization. Following this initial filtering step, 111 records remained and proceeded to formal title and abstract screening. This approach is consistent with PRISMA 2020 guidance and is reflected in the PRISMA flow diagram.

Screening followed a two-stage process consistent with PRISMA-ScR [15]. Two authors (AD and JH) independently screened a total of 111 retrieved PDF records for relevance to smartphone-based sensing and health-related outcomes. A total of 43 studies were identified from PubMed, 21 from IEEE Xplore, 26 from ACM digital library, and 21 from Google Scholar (total = 111). After this initial title and abstract review, the same reviewers examined the full text of the remaining studies to confirm that they met inclusion criteria, which required the use of smartphone-based sensing and a defined health-related outcome.

Screening was performed manually, and any disagreements between reviewers were resolved through discussion until consensus was reached. No formal inter-rater reliability statistic (e.g., Cohen's  $k$ ) was calculated, as the screening process was exploratory and descriptive in nature. Reasons for exclusion at each stage, including the use of auxiliary sensors, EMA-only studies, or the absence of health-related outcomes, are summarized in Figure 1.

We included only English-language peer reviewed manuscripts that 1) primarily used smartphone-based sensor data as the input for digital phenotyping studies, 2) described the use of smartphone-based digital phenotyping with relation to a pre-described health condition. We excluded 1) non-English language literature, 2) conference abstracts, white papers or other policy papers, 3) manuscripts that described digital phenotyping through the use of smartphones as a platform to exclusively perform ecological momentary assessments or other surveys without using on-board smartphone sensor data.

To find studies for this scoping review we first determined the eligibility criteria for our studies. For a work to be considered within our scope for this scoping review, they must meet *all* three of the inclusion criteria listed below:

- The studies *must* use digital phenotyping approaches that *exclusively* use smartphone-based sensor data as *input*.
- The studies *must not* include approaches that *exclusively* use smartphones as a method for manual data entry, such as EMAs, surveys, etc.
- The studies *must* use digital phenotyping explicitly in the characterization of one or more well-defined health conditions. This includes studies that looked at monitoring the general well-being of an individual.

We did not include studies in this scoping review that solely focused on algorithmic advances or enhanced methods to capture data without analyzing it in the context of specific health conditions.

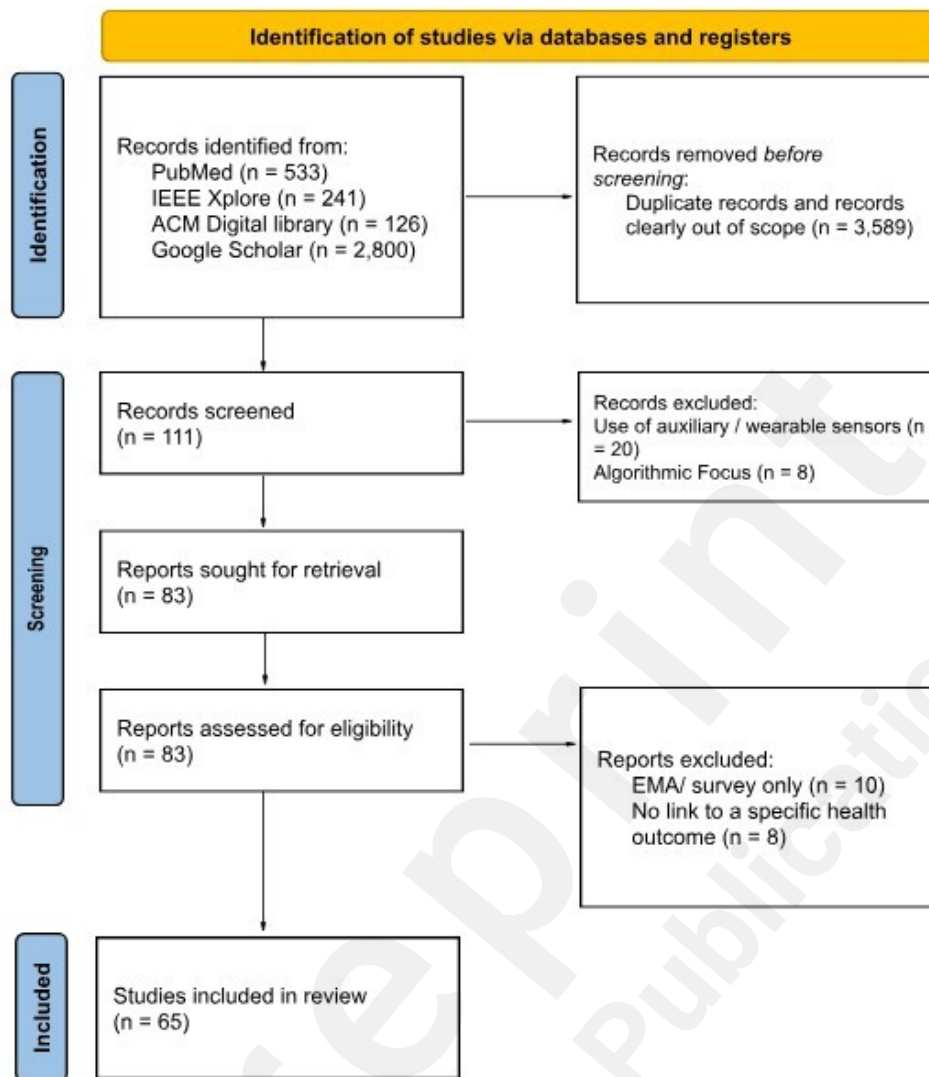


Figure 1: PRISMA flow diagram illustrating the screening and selection process for studies included in this scoping review. The initial search across four databases (Google Scholar, IEEE Xplore, ACM Digital Library, and PubMed) identified 111 records published between **January 1, 2012 and October 31, 2025**. After removal of non-relevant records during title and abstract screening, 65 full-text articles met the inclusion criteria for smartphone-based digital phenotyping using on-device sensor data and a defined health-related outcome. Studies were excluded for reasons including: use of auxiliary sensors or wearables as primary data sources, EMA-only designs without sensor data, absence of a health-related condition or outcome, or non-peer-reviewed publication type. The final set of 65 studies formed the basis of the analysis presented in this review.

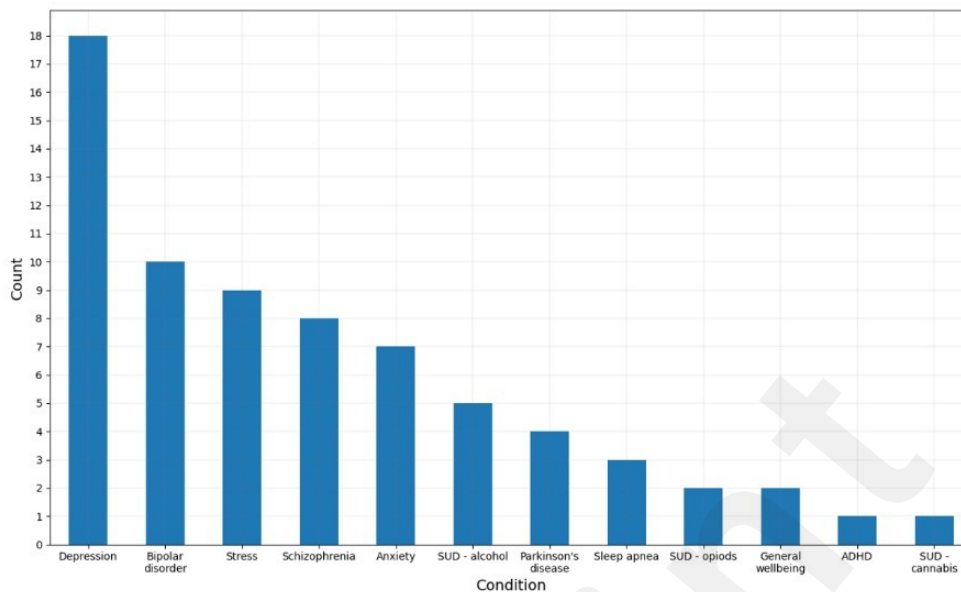


Figure 2: This bar graph depicts the counts of health conditions considered by the studies in this scoping review. For studies that cover multiple health conditions are counted for each condition (i.e., studies that consider both anxiety and depression are present in the counts for both conditions). Here, "SUD" stands for substance use disorder.

Figure 1 depicts a PRISMA flowchart of the review process. Using the keywords listed above, we identified 111 studies, of which 46 did not meet the inclusion criteria. This resulted in 65 studies that form the basis of this review. Figure 2 shows the frequency of the health conditions represented. The most commonly studied condition was depression ( $N = 18$ ), followed by bipolar disorder ( $N = 10$ ) and stress ( $N = 9$ ). The least represented conditions were substance use disorder (cannabis) ( $N = 1$ ) and ADHD ( $N = 1$ ).

Before we proceed further, a note about terminology, in the rest of the article use the terms *study* and *studies* interchangeably. Since each of the 65 studies we looked at in this review reported on one study, the two terms are synonymous in the context of this study.

## Results

Before we delve into the details of the findings of our scoping review, we quickly present an overview of prior scoping reviews in the digital phenotyping space and argue why our work is different. We then delve into the findings of our scoping review.

### Prior scoping reviews

There are existing reviews that have explored the utilization of digital phenotyping. Our search yielded 21 prior reviews. We classify these reviews based on the types of sensing technology used: smartphone, smartwatches, and auxiliary sensors. We define auxiliary sensors to be any sensing device outside of smartphones and smartwatches such as smart textiles or sensors worn on the body. A brief overview of the reviews reported on in this section can be found in Table 1, which we now elaborate on. Earlier reviews predominantly examined mental health outcomes such as depression, bipolar disorder, and stress, whereas more recent work expanded to neurological, physiological, and multimodal applications. The final row (gray) highlights the present scoping review, which uniquely synthesizes 65 smartphone-only studies spanning nine

## health domains

Citation	# Studies	Sensing modalities used	Year	Area of interest
[17]	119	Smartphones	2018	Health and well-being
[18]	18	Smartphones	2019	Bipolar disorder
[19]	224	Smartwatches	2020	Pre-surgical risk assessment, Post-surgical outcome prediction
[20]	62	Smartphones, Smartwatches, Auxiliary sensors	2021	Bipolar disorder
[21]	92	Smartphones, Smartwatches, Auxiliary sensors	2021	Depression, Bipolar disorder, Anxiety, Stress
[22]	51	Smartphones, Smartwatches	2022	Depression
[23]	46	Smartphones, Smartwatches, Auxiliary sensors	2022	Machine learning practices
[24]	31	Smartphones, Smartwatches	2022	Available datasets and smartphone applications
[25]	81	Smartphones, Auxiliary sensors	2023	Parkinson's Disease, Alzheimer's
[26]	11	Smartphones, Smartwatches, Auxiliary sensors	2023	Anxiety
[14]	40	Smartphones	2023	Schizophrenia, Depression, General well-being
[27]	29	Smartphones, Smartwatches	2023	Mood disorders, Schizophrenia-spectrum
[28]	47	Smartphones	2023	Social behavior sensing
[29]	59	Smartphones, Smartwatches	2023	Depression, Anxiety, Stress, Schizophrenia, Mood, General well-being
[30]	35	Smartphones	2024	Youth depression and anxiety
[31]	24	Smartphones	2024	Machine learning practices
[32]	40	Smartphones	2024	Depression, Stress, Anxiety
[33]	112	Smartphones	2025	Digital phenotyping pipeline
[34]	16	Smartphones, Smartwatches, Auxiliary sensors	2025	Stress monitoring
this scoping review	65	Smartphones	2024	Depression, Bipolar disorder, Schizophrenia, Anxiety, Stress, General well-being, Substance use, Sleep apnea, Parkinson's disease

Table 1: Summary of existing digital phenotyping surveys (N=19). This table summarizes prior reviews and reviews examining digital phenotyping across different sensing modalities and health domains. Each citation lists the number of studies covered, primary sensing technologies (e.g., smartphones, smartwatches, or auxiliary sensors), publication year, and main health focus.

About a third of the reviews ( $N = 7$ ) focus on exclusively using smartphone for digital phenotyping. Three of these reviews focus on describing phenotyping in the context of specific health conditions: depression [31], bipolar disorder [18], and a combination of depression, stress, and anxiety [32]. The remaining reviews include: a scoping review of 35 studies on depression and anxiety in youth [30], a review of 47 studies using smartphones for social behavior sensing [28], a large review of 112 studies characterizing digital phenotyping pipelines [33], and a systematic review of 119 studies on smartphone-based health sensing [17]. Additionally, [27] reports on 29 clinical studies using smartphones for psychiatric relapse monitoring.

Another third of the reviews ( $N = 6$ ) cover a combination of smartphone and smartwatch-based digital phenotyping. Similar to the previous category, two of the studies focus on

phenotyping work addressing mental health conditions: depression [22] and depression, anxiety, stress, schizophrenia, mood, and general well-being [29]. One review covers a purely smartwatch-based digital phenotyping to assess pre-surgical risk and post-surgical care [19].

The last review in this category reports on publicly available applications and datasets for digital phenotyping that utilize both smartphones and smartwatches [24].

The remaining reviews ( $N = 8$ ) report on digital phenotyping using a combination of smartphones, smartwatches, and auxiliary sensors. Like the prior two sections, the reviews in this category focused on mental health conditions: Alzheimer's [25]; depression, bipolar disorder, anxiety, and stress [21]; bipolar disorder [20]; and anxiety [26]. [23] and [14] focus on the use of machine learning to interpret phenotyping data, while [27] examines combined sensing modalities in psychiatric relapse studies. Most recently, Kallio et al. [34] provide a review of approximately 16 studies focused on continuous stress monitoring in cognitively demanding work environments using smartphones, smartwatches, and ambient sensors. Their review emphasizes real-world deployment, unobtrusiveness, and user acceptance, expanding the digital phenotyping literature into long-term occupational health monitoring.

Most of these reviews ( $N = 15$ ) cover research that use more sensing modalities than those available on smartphones. Those that do cover purely smartphone-based digital phenotyping ( $N = 7$ ) are scoped in terms of specific mental health conditions, youth populations, or derived behavioral features.

Our review is different from these prior reviews for three reasons: (1) it focuses exclusively on smartphone-based digital phenotyping; (2) its scope is much more expansive, in that it covers any and all health conditions (and not just mental health conditions); and (3) it focuses on describing the relationships between behaviors captured through phenotyping in the context of specific health conditions and not on algorithms used to extract the behaviors.

## Findings of this scoping review

Our search to understand the current state of smartphone-based digital phenotyping identified 111 records. After title/abstract screening, 65 studies met inclusion criteria and proceeded to full-text review; all 65 were included in the synthesis. Across the 65 included studies, we conducted a descriptive synthesis to summarize the sample sizes, study durations, and participant populations. The median sample size across all studies was 52 participants (IQR = 26 - 126), with a median study duration of 8 weeks (IQR = 3 weeks - 6 months). Studies commonly used multiple different measurements for ground-truth. Clinical rating scales (e.g., PHQ-9, HAM-D, YMRS, GAD-7, PANSS) were used in 41 studies (63%), ecological momentary assessments in 18 (28%), clinician-confirmed diagnoses in 9 (14%), and physiological validation measures (e.g., polysomnography or gait laboratory testing) in 3 (5%). Each of these studies were additionally described by the study setting. Controlled laboratory studies are conducted under precise conditions. Naturalistic "daily life" studies capture behavior in real-world environments. Hybrid study designs intentionally combine both settings within the same study [35].

One way of examining smartphone-based digital phenotyping research is by considering the conditions that studies target. These works can be grouped into six broad categories: mood disorders, stress-related disorders, schizophrenia, disorders from the use of psychoactive substances, nervous system disorders, and general well-being management.

In all of these efforts, the smartphone sensors were used to collect a variety of contextual information about the individual, such as call/text logs, phone/app usage, location, accelerometer, and microphone data. These were used to extract phenotypes, such as social functioning, physical activity, movement patterns, and sleep patterns, which were associated

with the specific health condition being considered. These phenotypes are then analyzed using some form of machine learning algorithm to answer questions about specific health conditions. To build such models, knowing the ground truth of the health condition of the participants is essential. This ground truth was usually assessed using ecological momentary assessments (EMA) or an appropriate scale based on the condition being tracked. In our description below, we do not focus on the intricacies of the machine learning algorithms used, as they are largely similar across most of the studies. Rather, as much as possible, we focus on the explanations that these models offer about the relationship between sensing, phenotypes, and the health condition being considered.

For each health condition we discuss in this section, we categorize the work done in one or more of the following *focus categories*: monitoring, diagnosis, and intervention. *Monitoring* involves collecting and analyzing smartphone-based digital phenotypes to track the health condition being considered. *Diagnosis* uses smartphone-based digital phenotypes to identify the presence of the health condition or risks thereof. *Intervention* involves implementing strategies or treatments often as a response to monitoring a health condition. Not every health condition has work in all three categories.

## Mood disorders

Mood is defined as a pervasive and sustained feeling that is experienced internally [52]. It impacts nearly all aspects of a person's behavior externally. Mood disorders are described by marked disruptions in emotions. For example, experiencing severe lows called depression or highs called hypomania or mania [52]. Though there are a variety of mood disorders, smartphone-based digital phenotyping work in the context of mood disorders has targeted only two major conditions in this regard: depression and bipolar disorder, which we now describe.

### Depression

Depression (otherwise known as depressive disorder and major depressive disorder) involves low moods and a loss of pleasure or interest in activities over a prolonged period of time [53]. Over the years, studies have shown that passive sensor data collected from smartphones can be associated with depression [36, 37, 38]. Much of the digital phenotype effort to explore and characterize depression has focused on four types of phenotyping efforts: tracking social interactions, tracking location diversity, vocal characteristics, and tracking sleep quality. For these approaches, the ground truth was collected via EMA or self-reported using pertinent scales, such as the Personal Health Questionnaire (PHQ) [54], the Center for Epidemiological Studies-Depression Scale (CES-D) [55], the Children's Depression Rating Scale-Revised (CDRS-r) [56], and the Clinical Global Impressions-Severity Scale (CGIS) [57]. Phenotyping work done in the depression space were primarily focused on the monitoring of depression with some work in the diagnosis focus category. Table 2 provides a summary of the studies reviewed in this section.

Citation	Condition	Phenotypes considered	Sensors used	Ground truth	# Part.	Duration	Focus	Study Setting
[36]	Depression	Sociability, Movement patterns	Application use, Gyroscope	PHQ-9	558 Adults	10.7 days avg.	Monitoring	Naturalistic setting (app store recruitment)
[37]	Depression	Movement patterns	GPS, Application logs, Call logs, Text logs,	PHQ-9	38 Adults	2 years	Monitoring	Naturalistic setting (Spain)

[38]	Depression	Movement	IMU GPS, Weather, Light	DASS, PANAS-X	31 Adults	2 weeks	Monitori ng	Naturalistic setting (USA)
[39]	Depression	Sociability	Application use, Call logs, GPS, Battery, Screen state	CES-D	120 Adults	1 month	Monitori ng	Naturalistic setting (China)
[40]	Depression	Activity, Sociability	GPS, Screen state	PHQ-9	40 Adults	2 weeks	Monitori ng	Naturalistic setting (USA)
[41]	Depression	Sleep, Activity, Sociability	Call logs, Accelerometer, Light, GPS, Screen state	PHQ-9	47 Adults with diabetes	20 weeks	Monitori ng	Naturalistic setting (India)
[42]	Depression	Sociability, Activity, Movement	Call/Text logs, GPS, Accelerometer, Light, Screen State	PHQ-9, HAM- D, HAM-A	13 Adolesc ents	8 weeks	Monitori ng	Naturalistic setting (USA)
[43]	Depression	Sociability, Mobility	GPS, Gyroscope, Call/Text logs	CDRS-R	24 Adolesc ents	8 weeks	Monitori ng	Naturalistic setting (South Korea)
[44]	Depression	Activity, Movement	GPS	PHQ-9	48 College students	10 weeks	Monitori ng	Naturalistic setting (USA)
[45]	Depression	Movement patterns	GPS	PHQ-8	28 Adults	2 mont hs	Monitori ng	Naturalistic setting (app store recruitment)
[46]	Depression	Activity, Sociability	Accelerometer, GPS, WiFi, Call/Text logs	PHQ-9	126 Adults	9 mont hs	Monitori ng	Naturalistic setting (app store recruitment)
[47]	Depression	Movement patterns	Accelerometer, GPS	PHQ-9	33 Adults	11 weeks	Monitori ng	Naturalistic setting (Bangladesh)
[48]	Depression	Activity, Sleep disruption	GPS, Accelerometer, Logs, Screen state	Self- reported mood	100 College students	28 days	Monitori ng	Naturalistic setting (USA)
[49]	Depression	Sociability	Microphone, Text, GPS	PHQ-9	302 College students	6 mont hs	Monitori ng, Diagnosis	Hybrid setting (USA)
[50]	Depression	Psychological stress	Microphone	BDI	1814 Adults	2 years	Monitori ng, Diagnosis	Naturalistic setting (Japan)
[51]	Depression	Sociability	Microphone, Call/Text logs, GPS	PHQ-9, GAD- 7	70 College students	28 weeks	Diagnosis	Hybrid setting (USA)

Table 2: Summary of smartphone-based digital phenotyping studies focused on depression (N=16). Studies examined depressive symptoms and mood-related behaviors through passive smartphone sensing. Common phenotypes included sociability, mobility, activity, and sleep patterns. Sensors most frequently used were GPS, accelerometer, screen state, and communication logs, with PHQ-9 as the most common ground-truth measure.

**Monitoring:** One of the most common ways of characterizing depression and depressive episodes is by looking at changes in social interaction. The social interaction of a person can be approximated by measuring the number of incoming calls a person receives as well as the amount of time people spend on social media apps [39]. It was found that a lower volume of incoming calls and higher frequency of social media app use was indicative of depression, indicating that individuals who were depressed were not communicating with others and were instead turning to their phones as a potential form of self-soothing [39]. Similar findings on the presence of inverse correlation between number and length of phone calls, text messages, and overall phone use patterns and depression were found in four different studies [40, 41, 42, 49]. However, in one study the authors looked at an adolescent population in South Korea (as

opposed to adults in the aforementioned work) and found that those with depression received more calls compared to the control participants [43]. Though the study did find that a person spending more time on smartphones may be associated with depression as above.

Phone call and text logs as a proxy for social connection were not the only way to infer and characterize depression. Another common approach was to look at location features (based on the GPS and IMU). It has repeatedly been found that movement patterns may be a similarly strong proxy to mental state [44, 45]. A case in point is a study that looked at individuals with spinal cord injuries. For such individuals there was a direct correlation between a decrease in the number of locations visited (presumably due to their condition) with depression [58]. This was found to be true not just for adults but also for younger populations as well [59, 42]. A similar pattern was also reported in other studies that showed that the variance, number of location clusters, and location entropy all decreased in relation to an increased depression scores [44, 45]. When the location data was analyzed by making assumptions about the location context, such as duration of time away from home, the information could be used to provide interventions for individuals with depression. [46]. Beyond just diversity of location, in one study the researchers found that physical activity in general was associated with depression, with lower physical activity (in terms of distance variance from any starting point, greater resting time, greater time using the phone in a lying position, fewer steps, less exercise, and greater staying at home) were associated with increased depression [47].

A few studies have also looked at using sleeping patterns to characterize depression. Measuring sleep quality from smartphones, unlike wearable devices (e.g. smartwatches) is not straightforward. Consequently, researchers often make broad assumptions about the daily routines of a typical person and check to see if smartphone usage patterns can detect a break in this routine. For instance, in one of these studies a higher number of instances of “screen on” states recorded from 12am-6am was interpreted as indicating a lower sleep quality or disturbed sleep schedule in individuals [41]. This pattern of being up at night was found to be associated with the presence of depression. In a similar study on college students, it was found that higher variance in sleep patterns to be associated with depression [48].

Diagnosis: In recent years, an interesting and different approach to detecting the presence of depression in individuals has been explored using vocal characteristics. In two different studies, the authors looked at changes in vocalization and tone based on features such as vocal fray, vocal jitter, and monotony measured via unscripted and scripted audio recordings to determine the presence of depression [49, 51]. A similar study also analyzed voice features during phone calls to determine the presence of depression [50]. In one of the aforementioned studies, the researchers tried to extend their vocal characteristics-based depression models to the most extreme cases where individuals are experiencing suicidal ideation[51]. They found such models were not sufficient and required additional context to predict suicidal risk [51].

## Bipolar Disorder Research

Bipolar disorder is a mental health condition that causes individuals to experience extreme mood swings, typically between emotional highs (known as mania or hypomania) and lows (depression) [71]. There are several types of bipolar and related disorders, not all of which require for both manic and depressive episodes to be present, but all of which require clear changes in mood, energy, and activity levels [71]<sup>1</sup> With respect to smartphone-based digital

---

<sup>1</sup> Though very similar to the phenotypes used in characterizing depression, the way they are used for characterizing bipolar disorder is somewhat different. As discussed in the prior section, we know that it is possible to identify depression through a variety of methods, but it is important to note that depressive episodes in bipolar

phenotyping, considerable work has also focused on using exploring such bipolar disorder. These efforts have primarily looked at four types of phenotyping efforts: tracking phone usage, tracking location diversity and routine, and tracking sleep quality. Further, the ground truth for bipolar disorder was collected via EMA or self-reported using a combination of scales such as HAMD (Hamilton Depression Rating Scale) [72] and YMRS (Young Mania Rating Scale) [73]. Phenotyping work in bipolar disorder was primarily focused in the monitoring category with some work for both diagnosis and intervention. Table 3 provides a summary of research surveyed in this section.

Citation	Condition	Phenotypes considered	Sensors used	Ground truth	# Part.	Duration	Focus	Study Setting
[60]	Bipolar Disorder	Physical activity, Sociability	GPS, Cell tower pings, Accelerometer, Call logs, Text logs	Daily self-reported mood, HAM-D, YMRS	13 Adults with BD	12 months	Monitoring	Naturalistic setting (Germany)
[61]	Bipolar Disorder	Sociability	Accelerometer, GPS, Call data	HAM-D, YMRS	10 Adults with BD	12 weeks	Monitoring	Naturalistic setting (Austria)
[62]	Bipolar Disorder	Physical activity, Sociability	Screen state, Text logs, Call logs, Cell tower pings	HAM-D, YMRS	29 Adults with BD	12 weeks	Monitoring	Naturalistic setting (Denmark)
[63]	Bipolar Disorder	Sociability	Call logs, Text logs	Self-reported mood, YMRS, HAM-D	26 Adults with BD, 12 controls	28 days	Monitoring	Naturalistic setting (USA)
[64]	Bipolar Disorder	Movement, Activity, Sociability	Accelerometer, GPS, WiFi, Bluetooth, Call logs	Self-reported mood, HAM-D, MSS, ADS	10 patients with BD	in- 8 weeks	Monitoring	Naturalistic setting (Austria)
[65]	Bipolar Disorder	Activity, Movement patterns	GPS, WiFi, Cell tower pings	QIDS-SR	29 Adults with BD, 20 controls	3 months	Monitoring	Naturalistic setting (United Kingdom)
[66]	Bipolar Disorder	Movement, Sleep Disruption	GPS	ASRM, DAS-21	159 Adults with BD	1 year	Monitoring	Hybrid setting (Taiwan)
[67]	Bipolar Disorder	Activity, Sociability, Movement	Accelerometer, Microphone, Battery, Screen state, Logs, Calls	SRM-5	7 Adults with BD	4 weeks	Monitoring	Naturalistic setting (Germany)
[68]	Bipolar Disorder	Sociability	Call/Text logs, Microphone, Screen state, Accelerometer, Cell tower pings	Self-reported mood	28 Adults with BD	12 weeks	Monitoring, Diagnosis	Naturalistic setting (29)
[69]	Bipolar Disorder	Sleep, Activity	Accelerometer, Light sensor	PSQI	22 Adults with BD, 23	3 weeks	Intervention	Hybrid setting (Austria)

disorder may not be as intense or last as long [71]. In addition, mixed states (having depressive symptoms and manic symptoms at the same time) and manic episodes are just as important to recognize in individuals with bipolar disorder, meaning that while some techniques may be useful between both disorders, they are not the same disorder and cannot be treated as such.

					control s			
[70]	Bipolar Disorder	Sociability, Stress, Sleep	Accelerometer, WiFi, Logs, Screen state, Bluetooth, Microphone	Self-reported mood, HAM-D, YMRS	18 Adults with BD	5 mont hs	Monitoring, Diagnosis, Interventio n	Naturalistic setting (app store recruitment)

Table 3: Summary of smartphone-based digital phenotyping studies focused on bipolar disorder (N=11). Studies used passive smartphone sensors to assess behavioral and physiological markers associated with bipolar disorder. Commonly examined features included sociability, activity levels, and sleep-related changes. Ground-truth measures primarily relied on clinical scales such as HAM-D, YMRS, and QIDS-SR.

**Monitoring:** Like tracking depression, several efforts in recent years have focused on tracking a bipolar individual's mood over time. This information can be used to determine if a person is having a depressive or manic episode. In this regard, one of the common phenotypes is *smartphone use pattern* [61, 62, 70, 68]. In one study, the researchers used features of phone calls themselves, such as total length, average call length, number of unique calls, total number of calls, etc. along with sounds emanating during a phone call to determine the current mood/change in moods of bipolar individuals [61]. The study indicated the presence of some kind of association between bipolar states and these phone usage features but those exact associations were not clear. In two other similar studies it was found that severe depressive symptoms of bipolar disorder were associated with: longer the "screen on" state per day, more received calls per day, fewer outgoing calls per day, fewer answered incoming calls per day, and less movement [62, 63]. On the other hand, manic symptoms showed the exact opposite pattern as described above.

Another approach to characterizing bipolar disorder has been to look at movement/physical activity using smartphone-based GPS data [64, 65, 66]. In another study, researchers found that physical activity changes as mood states transition from one to another [64]. When transitioning from a manic state to a baseline normal state, physical activity decreases, but it increases when transitioning from a depressed state to baseline [64]. In a study focused on quantifying depressive states in those with bipolar disorder, the researchers derived a variety of features from GPS data from an individual's smartphone and were able to use it to predict acute depressive states with high accuracy [65].

**Diagnosis:** In the context of smartphone-based for bipolar disorder diagnosis researchers looked at vocal patterns along with self reported surveys on a variety of mood and sleep along with smartphone usage metadata from individuals with bipolar disorder to detect manic and depressive states [68]. Overall, it was found that while it was possible to classify bipolar states on voice features alone, combining different passively sensed and self-report data with voice features creates a more robust model in terms of detecting subtler manifestations of bipolar states (e.g., hypomania as opposed to mania) [68]. Another potential indicator of a bipolar person's mood is in patterns of app use on the smartphone by a bipolar individual [70]. In particular, the way that bipolar individuals interact with their phones has a strong correlation to what moods they're experiencing at the time. For instance, in one study, the researchers found that increased social interaction is associated with a decrease in self-reported stress level, irritability, and mixed mood. This indicated that increased social interaction is a good marker for normal social functioning outside of mood swings. Further, patterns in average number of apps being used can be to detect the presence of manic or depressive episodes with a higher number of apps used being positively associated with self-reported stress and a decrease in the average number of apps used was suggested to indicate a depressive episode [70].

**Intervention applications:** Finally, sleep disturbances have been shown to be early warning

sign of manic or depressive episodes in individuals with bipolar disorder [69]. As stated above, smartphone use (screen on/off, accelerometer readings) can provide insight into lower sleep quality and satisfaction of sleep, alluding to potential sleep disturbances [70]. In a different study, the researchers showed that a smartphone phone-based app could be a valid tool for measuring sleep onset and wake-up times, to measure sleep disturbances. This information can then be used for the management of the depressive symptoms of bipolar disorder [69].

## Stress-related disorders

Stress is the perception of an individual that their environmental demands tax or exceed their adaptive capacity [74]. Stress-related disorders are maladaptive responses to short- or long-term exposures to physical or emotional stressors. Smartphone-based phenotyping work for stress-related disorders can be broadly categorized into two groups: anxiety disorders, and stress quantification. We describe these below. Table 4 provides a summary of reviewed work in this section.

### Anxiety disorder

Anxiety disorders usually involves a persistent feeling of anxiety or dread, which can interfere with daily life. People living with such disorders experience frequent anxiety for extended periods of time [75]. For these efforts, the ground truth was collected via EMA or self-reported using pertinent scales, such as General Anxiety Disorder Scale (GAD-7) [76]. There were relatively fewer studies in this space, all of which focused on monitoring.

Monitoring: In one study, researchers showed that social function derived based on basic statistics collected about phone call made, received, missed, diversity of callers, etc. can be used to detect the presence of social anxiety disorder in an individual [77]. Other studies found that lack of social connection in terms of lower call use and increased social media application use [78], and use of phones in the dark [79] were associated with higher levels of anxiety disorders. Further, days of the week may also be important in recognizing anxiety, as self-reported anxiety appears to increase towards the end of the weekend and decrease towards the end of the week, possibly aligning with perceived stress or uncertainty about workloads during the week [79]. Overall, the results for characterizing anxiety disorders were very similar to characterizing depression, such that discerning between one and the other may be challenging. For instance, researchers showed a weak and positive relationship between the predicted social anxiety disorder symptom and depression when capturing social functioning of an individual based on phone call usage statistics [77].

### Stress Quantification

Psychological stress is another condition that has been targeted as part of smartphonebased digital phenotyping research. Though experiencing psychological stress is not an uncommon condition for most individuals, prolonged or significant stress has been linked to a variety of physical and mental illnesses [80]. Prolonged stress may create a space for burgeoning mental illnesses such as depression and anxiety to flourish [80, 81, 53]. Characterizing stress typically relies on the fusion of a variety of behavioral information about an individual, such as their movement patterns, physical activity, social interaction, and sleep patterns, much like works described in the section above. Phenotyping work related to stress were in the monitoring focus category.

Monitoring: In looking at monitoring stress, researchers looked at automatic stress detection in completely unconstrained environments. The study showed that it was possible to extract

time and frequency domain features from accelerometer data to characterize routine and associate these features with self-reported stress levels [82]. In a similar study, the researchers explored more complex fusion models where the the relationship between self-reported daily stress and sensor-derived activity changes, speech duration, changes in location, and sleep patterns were associated with changes in participants' stress levels over the course of the study [80]. In a similar work, the authors found that speech duration, movement information (measured using GPS and WiFi) and sleep duration were associated with self-reported daily stress levels while their activity information (measured using accelerometer) was associated with changes in their subjective loneliness [81]. Another approach to measuring stress proposed the fusion of information from call and text logs (this included observing broader patterns on from calls/text, such as percentage of calls done during the night; percentage of initiated calls during the night; text response rate; text response latency; and the percentage of initiated texts) combined with proximity information (collected from Bluetooth signals from neighboring devices) along with weather information (temperatures, precipitation, humidity, visibility, etc.) to accurately predict the presence of stress [83].

All these aforementioned works focused on work-related stress. Another group of individuals whose stress has been studied in some detail are college students. In one study the researchers developed an app called StudentLife for continuously sensing the day-to-day impact of college workload on stress, sleep, activity, mood, sociability, mental well-being and academic performance among a single class of 48 students across a 10 week term [84]. Analysis of the data collected shows that students at the start the term had high positive affect and conversation levels; low stress; and healthy sleep and daily activity patterns. However, as the term progressed and the workload increased, stress appreciably increased while positive affect, sleep, conversation, and activity dropped off [84]. A followup study looked at a more general approach to stress quantification in college students utilizing a geo-fencing approach on the GPS data from the smartphones carried by students [85]. The various spaces students frequented were designated as "study spaces" and "party locations" and in each of these locations the student's behavior is recorded based on smartphone activity and audio collected from the phone's microphone. The results indicated that students who reported decreasing stress levels throughout the term (self-reported) obtained higher GPAs [85].

Finally, researchers used keyboard pressure as a proxy for detecting the presence of stress. Their idea was that the increase in muscle tension in response to an individual experiencing stress would lead to increased keyboard pressure when typing. In this study, the researchers asked individuals to recall a stressful and relaxing event in the recent past and write about it using an iPhone with a modified QWERTY keyboard. They then logged the timing, position, and pressure values associated with each keyboard interaction. Subsequently, the differences in typing pressure across relaxed and stressful expressive writing tasks were associated with differences in self-reported stress [86].

## Delusional disorders

A delusion is a false belief based on an inaccurate interpretation of an external reality despite evidence to the contrary [95]. In the context of delusional disorders, smartphone-based phenotyping work has exclusively focused on schizophrenia. Schizophrenia is a significant and often life-altering mental disorder characterized by *episodes*, which include persistent delusions and/or hallucinations, disorganized thinking and behaviors, and instances of agitation or *negative symptoms*. The term negative symptoms describes a lessening or absence of normal behaviors and functions related to motivation and interest, or verbal/emotional expression, which can manifest as blunted affect, reduction in speech, goal-directed activities, motivation, and experience of pleasure [96]. We summarize our findings in this space below. Table 4

provides a summary of reviewed research in this section.

## Schizophrenia

Typically, smartphone-based digital phenotyping schizophrenia research has focused on identifying predictors of negative symptoms. This is done by tracking data such as: call patterns, location, physical activity, and sleep quality. The ground truth in most of these studies was collected via EMA. Phenotyping work with respect to schizophrenia space were in both monitoring and diagnosis focus categories.

Monitoring: Relatively few works have looked into applying phenotyping to people with schizophrenia. In one study, the researchers use patterns in incoming and outgoing calls and their duration, incoming and outgoing text messages, and capturing the frequency and duration of nearby voices captured through a smartphone microphone to determine exactly how much social activity an individual with schizophrenia was engaging in. Overall the researchers reported that reductions in number and duration of outgoing calls, as well as number of text messages were associated with relapses [87].

In examining negative symptoms of schizophrenia, researchers explored the feasibility of clustering GPS location as a proxy for social functioning and determining where individuals with schizophrenia are spending their time. Subsequently, they added contextual labels to the location clusters to attain a general understanding of social functioning for a specific individual [88]. In a similar study, researchers looked at GPS-derived indicators of location to determine if they could be used as a proxy for negative symptoms. They found that GPS-derived estimations of average distance traveled over time, distance from home, and the percentage of GPS samples at home were found to be highly associated with negative symptoms. Essentially, they found that less mobility was associated with greater negative symptom severity - possibly due to diminished motivation [89].

Diagnosis: Methods outside of GPS-based locations have also been explored to determine the level of social functioning among people with schizophrenia. In this case, routines and disturbances within those routines are seen as indicators of a relapse and may be used in the future as a marker for potential intervention [91]. Researchers have looked at smartphone sensor data (e.g., call/text logs, phone screen status, accelerometer, and GPS information) as a proxy for social functioning to determine the relationship between measuring social rhythm (using phenotyping) and schizophrenia. They found that increased stability in social rhythm and routine was associated with improved symptom scores in schizophrenia. This also implies that using phenotyping data related to sociability, mobility, and physical activity as we have described above in detecting depression, anxiety, and bipolar disorder can all be associated with identifying negative symptoms of schizophrenia [90].

In another study involving adults with schizophrenia, the researchers looked at mobility prior to a relapse and found that visual hallucinations were associated with a change in sleeping patterns and the number of voice and non-voice sounds captured, while auditory hallucinations decreased with new places visited and an increase in call duration [92]. Further, compared to individuals who did not require clinical intervention, there was a significantly higher rate of anomalies in the two weeks directly preceding the relapse. In a similar study the researchers looked at movement information using purely accelerometry data from the smartphone to identify negative symptoms of schizophrenia. The findings of the study indicate that smartphone accelerometers, which are markers of movement and vigor (measured via accelerometry) hold promise in detecting negative symptoms of schizophrenia [93]. In an additional study, the researchers explored the use of accelerometer data to estimate sleep duration and sleep quality

in individuals with schizophrenia. The results showed that these attributes were reliable to indicate the presence of sleep disturbances, which can be early warning signs for psychosis and relapse [94].

Citation	Condition	Phenotypes considered	Sensors used	Ground truth	# Part.	Duration	Focus	Study Setting
[77]	Anxiety	Sociability, Movement patterns	Accelerometer, Call logs, Text logs	SIAS, DASS-21, PANAS	59 college students	2 weeks	Monitoring	Naturalistic setting (USA)
[78]	Anxiety	Sociability, Physical activity	App log, Power level, Screen, GPS, Call/Text logs	SIAS, UCLA-LS	127 adults	30 days	Monitoring	Naturalistic setting (China)
[79]	Anxiety	Sociability, Physical activity	Light, Accelerometer, Gyroscope, App logs	STAI	20 healthy adults	30 days	Monitoring	Hybrid setting (Japan)
[80]	Stress	Movement, Activity, Sociability, Psychological stress	Accel., GPS, Mic, WiFi, Call/Text logs, App logs	Mood report	self-30 employees	8 weeks	Monitoring	Naturalistic setting (Italy)
[81]	Stress	Sleep, Activity, Sociability	Mic, WiFi, Light sensor	GPS, PSS, PHQ-9, UCLA-LS	47 young adults	10 weeks	Monitoring	Naturalistic setting (USA)
[82]	Stress	Physical activity, Sociability	Accelerometer	OLBI	30 employees	8 weeks	Monitoring	Naturalistic setting (Italy)
[83]	Stress	Sociability	Call/Text logs, Bluetooth	Stress report	self-117 graduate students	7 months	Monitoring	Naturalistic setting (USA)
[84]	Stress	Activity, Sociability, Sleep, Movement	Accel., Mic, Light, GPS, Bluetooth	PHQ-9, UCLA-LS	PSS, 48 students	10 weeks	Monitoring	Naturalistic setting (USA)
[85]	Stress	Activity, Sociability, Stress, Movement	GPS, WiFi, Mic, Accelerometer	GPA, Stress/Mood self-report	48 students	10 weeks	Monitoring	Naturalistic setting (USA)
[86]	Stress	Psychological stress	Keyboard pressure	Mood self-report, Event context	11 adults	1 lab session	Diagnosis	Controlled laboratory setting (USA)
[87]	Schizophrenia	Sociability	Call logs, Text logs, Microphone	BPRS, Hospital relapse records	61 adults with SZ	1 year	Monitoring	Naturalistic setting (USA)
[88]	Schizophrenia	Activity, Sociability	Call logs, Text logs, Microphone	Social function diary	5 adults with SZ	5 days	Monitoring	Naturalistic setting (location not stated)
[89]	Schizophrenia	Movement patterns	GPS	Self-reported activity/context	85 adults with SZ, 58 controls	7 days	Monitoring	Naturalistic setting (USA)
[90]	Schizophrenia	Sociability	GPS, Accel., Screen, Call/Text logs	PHQ-9, GAD-7, reported symptoms	45 adults with SZ, 43 controls	3 months	Diagnosis	Naturalistic setting (USA)
[91]	Schizophrenia	Movement, Sociability	Accel., GPS, Call/Text logs, Screen, Battery	Symptom report	17 adults with SZ	3 months	Diagnosis	Naturalistic setting (USA)
[92]	Schizophrenia	Movement, Activity, Sociability, Sleep	Accel., GPS, App logs, Light, Call/Text logs	Mood self-report	21 adults with SZ	8 months	Diagnosis	Naturalistic setting (USA)
[93]	Schizophrenia	Activity level	Accelerometer	Self-reported activity	50 adults with SZ, 70 controls	6 days	Diagnosis	Naturalistic setting (USA)
[94]	Schizophrenia	Sleep quality	GPS, Screen	PSQI, lab sleep	17 SZ	3	Diagnosis	Naturalistic setting

	a		Call/Text logs, assessment Accelerometer		months		(USA)
--	---	--	---	--	--------	--	-------

Table 4: Summary of smartphone-based digital phenotyping studies focused on anxiety, stress, and schizophrenia (N=18). Studies span sociability, movement/activity, sleep, and psychological stress phenotypes using signals such as accelerometer, GPS, microphone, app logs, and communication metadata. Ground-truth measures include standardized scales, self-reports, and clinical or laboratory validations.

## Disorders from the use of psychoactive substances

A substance use disorder (SUD) is a mental disorder in which a person experiences an inability to control their use of psychoactive substances such as alcohol, cannabis, opioids, etc. [104]. Understanding these substance use disorders is an emerging aspect of smartphone-based digital phenotyping. We group the effort in this space into two broad categories: alcohol use and cannabis and opioid use. As with the previous efforts at phenotyping, these approaches use sensor data available from smartphones and associate behaviors such as social functioning, movement patterns, and physical activity to psychoactive substance use. Table 5 provides a summary of research surveyed in this section. Phenotyping work in substance use disorders were in both monitoring and diagnosis focus categories for both alcohol and cannabis and opioid use.

### Alcohol use

One of the main areas of interest in smartphone-based phenotyping around the use of psychoactive substance is in studying alcohol use. These efforts take two basic forms: determining the level of intoxication and predicting cravings. Phenotyping work related to alcohol use falls under the monitoring and diagnosis categories.

**Monitoring:** Using phenotyping to determine how intoxicated an individual is has been the most common form of research in this space. One notable study in this domain looked at a wide range of mobile phone sensor data (e.g., accelerometry, keyboard usage, screen state, call/text logs, battery status etc. ) from young adults with a history of hazardous drinking over a period of a month. It found that time of day, day of week, screen duration, frequency of keyboard presses had significant associations with heavy drinking, indicating behavioral changes during these episodes [97].

In a slightly different study, the researchers looked at alcohol craving in individuals with alcohol-associated liver disease via phone call status, movement patterns, phone screen state, and keystroke information. They found that the level of craving was directly proportional to loneliness, stress, and anxiety as measured primarily through a decrease in the entropy of their movement patterns [98].

**Diagnosis:** A couple of other efforts looked at blood alcohol levels based on gait information inferred from smartphone sensors [99, 100]. The researchers conducted these studies in an uncontrolled environment where it was not clear where the phone was located on the individual's body. These works utilized accelerometer and accelerometer and gyroscope, respectively. Overall, they found some positive correlation between gait measured through smartphone sensors and the level of intoxication of the individual.

### Cannabis and opioid use

Not much work has been done when it comes to characterizing cannabis and opioid use using smartphone-based phenotyping. The extant work in this space can be grouped into the same two

categories as the alcohol case: determining the level of intoxication and predicting cravings. Similar to alcohol use, phenotyping work related to cannabis and opioid use fell under the monitoring and diagnosis categories.

**Monitoring:** For cannabis, researchers have looked at identifying slowed psychomotor function as a result of consuming cannabis. They measured psychomotor function in an individual based on their movement pattern and physical activity measured using GPS, accelerometer, call logs. They found a smaller travel boundary and lower activity changes at times when the individual felt “high.” Additionally, stronger body movements indicated by changes in accelerometer may be indicative of impaired motor control or unsteady movements typically associated with higher levels of intoxication [101].

**Diagnosis:** In a study related to the context of opioid use, the researchers explored the feasibility of using smartphone data to monitor respiratory changes indicative of an impending opioid overdose. They showed that they could convert a phone microphone into a short-range active sonar using frequency shift to measure chest motion and respiration. This work was done in a controlled operating room environment of the operating room, where overdose conditions were simulated, this method showed promising accuracy in identifying respiratory depression [102].

Finally, the increased interest in utilizing smartphones to aid in the understanding and treatment of substance use disorders has created an opportunity to assess substance craving as reported in one study. By looking at the past five hours of solely GPS-based movement data from individuals with opioid-use disorder and correlating it with environmental exposure (measured as a function of affluence of the locations they visited), the researchers could predict craving behaviors up to 90 minutes in the future for both heroin and cocaine. They were even able to predict stress in the individuals (indicated through EMA responses) using this information [103].

Citation	Condition	Phenotypes considered	Sensors used	Ground truth	# Part.	Duration	Focus	Study Setting
[97]	SUD – alcohol use	Movement, Sociability	Accelerometer, Battery, Call/Text logs, Light, WiFi, Bluetooth, GPS, Gyroscope, Screen	Self-reported alcohol use	30 Adults	28 days	Monitoring	Naturalistic setting (location unclear)
[98]	SUD – alcohol use	Movement, Sociability	Accelerometer, GPS, Screen, App logs, Call/Text logs, Keystroke data	Mood and craving self-report	24 Adults	30 days	Monitoring	Naturalistic setting (USA)
[99]	SUD – alcohol use	Physical activity	Accelerometer	Alcohol use self-report	7 Adults	2 weeks	Diagnosis	Naturalistic setting (USA)
[100]	SUD – alcohol use	Physical activity	Accelerometer, Gyroscope	Urine-based BAC	65 Adults	1 session (5 hrs)	in-lab Diagnosis	Naturalistic setting (USA)
[101]	SUD cannabis use	Activity, Movement, Sociability	GPS, Accelerometer, Call logs	Acute cannabis intoxication self-report	57 Adults	30 days	Monitoring	Naturalistic setting (USA)
[102]	SUD opioid use	Respiratory changes	Microphone	Simulated overdose events	209 injection events	3–5 mins post-event	Diagnosis	Hybrid (USA) setting
[103]	SUD opioid use	Movement patterns	GPS	Drug craving self-report	189 Adults with OUD	16 weeks	Diagnosis	Naturalistic setting (USA)

Table 5: Summary of smartphone-based digital phenotyping studies focused on substance use

disorders (N=7). Studies examined behavioral and physiological markers related to alcohol, cannabis, and opioid use. Phenotypes primarily involved movement, sociability, and respiratory features derived from accelerometer, GPS, microphone, and app log data. Ground-truth measures included self-reported use, craving reports, and laboratory validation of blood alcohol concentration (BAC).

## Sleep disorders

Smartphone-based digital phenotyping research has also specifically looked at sleep disorders. However, it is somewhat limited. The main focus in phenotyping for sleep disorders has looked at sleep apnea. Sleep apnea is a common condition in which a person's breathing will stop and start again throughout the night [114]. This is a potentially serious disorder that not only impacts sleep quality, but it may also lead to increased cardiovascular risk and lead to premature death if not treated correctly [114]. Smartphone-based approaches primarily monitor breathing and snoring patterns using the smartphone microphone. The ground truth for these studies was established by recruiting individuals with previously diagnosed sleep apnea or by conducting polysomnography as part of the study. Table 6 provides a summary of sleep disorder research surveyed in this section.

### Sleep apnea

Traditionally, measuring the presence of sleep apnea requires an overnight stay at a sleep lab in order for a formal diagnosis to take place [115]. This stringent requirement may discourage individuals from actively pursuing treatment if they are unsure of the outcome of a sleep study or may perceive it to be too much of a hassle. As a result smartphone-based digital phenotyping aims to remain as unobtrusive as possible. Phenotyping work for sleep apnea was primarily in the diagnosis focus category.

Diagnosis: In one study, the researchers proposed a method that used accelerometer data and compared movements to a threshold to determine the extent of sleep movement. This was combined with snoring related sound data from the smartphone microphone to determine the quality of sleep, which were divided into three levels of apnea: no apnea, mild apnea, and severe apnea [105]. In a similar study, researchers monitored breathing patterns of a person using a smartphone app placed at the bedside. Their approach includes collecting information on breathing, the sound of snoring, and breathing pauses to determine the presence of sleep apnea and hypopnea, by parsing out white noise and various sounds that occur during the night [106].

The team that developed the means to determine feasibility of using smartphone microphone to monitor respiratory changes indicative of an impending opioid overdose also developed a similar approach that turned the microphone into a short-range active sonar to detect chest and abdomen movements. The changes in chest and abdomen movement detected from the reflected sound waves then allowed for apnea detection, which was then shown to be strongly associated with polysomnography results for central apnea, hypopnea, and obstructive apnea [102].

## Neurological disorders

A neurological disorder is any disorder of the nervous system. Examples of symptoms include paralysis, muscle weakness, poor coordination, loss of sensation, seizures, confusion, pain, and altered levels of consciousness [116]. From a smartphone-based digital phenotyping standpoint, work on neurological disorders has primarily focused on characterizing Parkinson's disease. Table 6 provides a summary of the research reviewed in this section.

## Parkinson's disease

Parkinson's disease is a movement disorder of the nervous system that progressively worsens over time. As neurons in the brain weaken and die, individuals with the disorder begin to suffer with tremors, impaired balance, stiffness in the limbs and trunk of the body, and difficulty with movement [117]. Like sleep apnea, phenotyping work for Parkinson's was also primarily in the diagnosis focus category.

**Diagnosis:** An aspect of Parkinson's that has been explored in the smartphone-based phenotyping context is the detection of real-time freezing of gait (stopping suddenly while walking) [107]. In a similar study, 20 individuals with Parkinson's were asked to perform a video-recorded Timed Up and Go (TUG) test with and without dual-tasks while carrying a smartphone. The video and accelerometer recordings from the smartphone were synchronized in order to assess the reliability of the FOG detection system in comparison to clinician ratings with reasonable results [117]. In a similar study, the researchers looked at the accuracy of smartphone-based gait analysis using the built-in triaxial gyroscope and accelerometer and validated it against two heel contact-based measurements. These findings highlighted that smartphone-based gait analysis could serve as an alternative to conventional gait analysis methods (e.g., foot-switch systems or sensor-embedded walkways), particularly when those methods are inconvenient (e.g., cost-prohibitive) [108].

Smartphones have been evaluated as a way to measure hand tremor in Parkinson's disease patients [109]. Tremors are a common symptom of Parkinson's disease. In this study, accelerometer and gyroscope were utilized to detect and measure hand tremors, which were then associated with clinical scores in the hand tremor components of the Unified Parkinson's Disease Rating Scale (UPDRS). The results suggest relatively strong correlation between the patients' UPDRS hand tremor scores and the quantitative measurement derived from smartphones [109]. In another similar study, the goal was to develop a diagnostic test to differentiate between Parkinson's disease and essential tremor based on time-frequency differences using smartphone-based accelerometer data without the use of neuroimaging techniques [110]. In this case, the smartphone was placed over the back of the outstretched hand of a person with Parkinson's disease while at rest. The results demonstrated that not only was the phone-based approach effective, it was also able to diagnose the presence of Parkinson's disease and essential tremor in several participants each, who had not been officially diagnosed [110].

Citation	Condition	Phenotypes considered	Sensors used	Ground truth	# Part.	Duration	Focus	Study Setting
[105]	Sleep apnea	Sleep disruption Sleep patterns	Microphone	PSG taken prior to study	55 samples	One Eight hour sleep period	Diagnosis	Hybrid setting (location not stated)
[106]	Sleep apnea	Sleep disruption Sleep patterns	Microphone	PSG Score	55 sleep samples	One Eight hour sleep period	Diagnosis	Controlled laboratory setting (location not stated)
[107]	Parkinson's Disease	Tremors, Freezing of Gait (FOG)	Accelerometer	Timed Up and Go (TUG) test	20 Adults with PD	One in-lab assessment	Diagnosis	Controlled laboratory setting (Location not stated)
[108]	Parkinson's Disease	Freezing of Gait (FOG)	Accelerometer, Gyroscope	Heel-mounted foot-switch, Pressure sensor mat	12 Adults with PD, 12 healthy controls	One in-lab assessment	Diagnosis	Controlled laboratory setting (Singapore)
[109]	Parkinson's Disease	Tremors	Accelerometer, Gyroscope	UPDRS	23 Adults with PD	One in-lab assessment	Diagnosis	Controlled laboratory setting (Greece)

[110]	Parkinson's Disease		Tremors	Accelerometer	Diagnosis of PD or ET, reevaluation after one year	17 Adults with PD, 16 adults with ET, 12 healthy controls	One in-lab assessment	Diagnosis	Controlled laboratory setting (Location not stated)
[111]	ADHD		Loss of concentration	Text logs	Self-reported symptoms	2 Adults with ADHD, 2 healthy controls	10 weeks	Monitoring	Naturalistic setting (USA)
[112]	General Wellbeing		Movement patterns, Physical activity level, Sociability	Microphone, Accelerometer, GPS, Text logs, Call logs	Self-reported mood	18,000 Adults	3 years	Monitoring	Naturalistic setting (app store recruitment)
[113]	General Wellbeing		Physical activity level, Sleep disruption, Sociability	Accelerometer, Microphone, Screen Battery level	Self-reported physical and mental state/wellbeing	27 Adults	19 days	Intervention	Naturalistic setting (app store recruitment)

Table 6: Summary of smartphone-based digital phenotyping studies focused on sleep disorders, neurological disorders, ADHD, and general wellbeing (N=10). Studies in these domains used smartphone sensors to assess sleep quality, neurological motor symptoms, attentional patterns, and general wellbeing. Common sensor modalities included microphones, accelerometers, and gyroscopes, with validation against PSG, clinical scales, or self-reported measures.

## Other efforts

A small amount of work has also been done in characterizing Attention Deficit Hyperactivity Disorder (ADHD) and also assessing the general mental health and well-being of individuals [111, 112, 113]. Table 6 provides a summary of the research covered in this section. Phenotyping work for ADHD was primarily focused on monitoring, whereas general wellbeing had work in both monitoring and diagnosis focus categories.

**Monitoring:** With respect to ADHD, researchers collected text data and associated it to weekly self-report ADHD symptoms scores. The ground truth for the ADHD symptoms were self-reported by the participants. Overall, the results of the study indicated that specific ADHD inattention, sluggish cognitive tempo and hyperactivity symptoms could be predicted fairly accurately using the smartphone data collected from college students in a pilot study [111].

Further, smartphone-based phenotyping also focused on general well-being. These studies aimed to associate self-reported mental well-being with routine captured through smartphone. This was a particularly large and longitudinal study. Overall, the study found that positive or negative affect for an individual could be predicted using passive sensing data (such as GPS, phone/call logs, accelerometer, and microphone data). These results suggest that such approaches could reduce the need for self-report [112].

**Intervention:** In the one intervention study, researchers deployed an app called BeWell [113]. BeWell monitored user behavior, such as sleep, physical activity, and social interaction based on accelerometer, screen on/off, charging, and microphone data to give individuals a wellbeing score. The wellbeing scores range between 0 and 100 and are calculated for each of the three dimensions (sleep patterns, physical activity, and social interaction). A score of 100 indicates the person is matching or exceeding recommended guidelines. The idea is to use these scores to eventually provide feedback to individuals to maintain their wellbeing over time [113].

## Sensor Modalities Across Health Domain

Finally, the reviewed studies showed wide variation in *sensor use* across health domains. While some investigations relied on a single data stream (e.g., the work in [103] that exclusively used GPS to track Bipolar Disorder or the research in [93] that assessed the activity level of adults with Schizophrenia using only the smartphone accelerometer), others combined multiple sources such as GPS, accelerometry, and microphone data to capture richer behavioral context. As shown in Figure 3, *GPS and accelerometer* emerged as the two most common Data streams. Accelerometers appeared in input 58% ( $N = 38$ ) of studies, making them the most frequently used modality overall. GPS followed close behind appearing in 53% ( $N = 35$ ) studies. Communication logs were also common with call/text logs appearing roughly 44% ( $N = 29$ ) of studies. These streams formed the core of most smartphone-based digital phenotyping work due to their ability to intuitively capture mobility, activity, and daily routines. Other smartphone sensing modalities were used far less often. Microphone-based sensing appeared in 27% ( $N = 18$ ) studies and was fairly evenly spread between health conditions, indicating the potential for widespread use across conditions. Screen state data also appeared in 27% ( $N = 18$ ) studies, typically as a proxy for phone engagement and sleep-wake cycles. Gyroscope data was used relatively infrequently, appearing in just 10% ( $N = 7$ ) studies. Environmental and proximity sensing (ambient light, WiFi, cell tower pings, app usage, and Bluetooth) appeared sparingly. At most these streams appeared in 15% ( $N = 10$ ) of studies in the case of ambient light and just 6% ( $N = 4$ ) of studies in the case of cell tower pings. Battery percentage was used surprisingly infrequently with only 7% ( $N = 5$ ) studies utilizing this stream. We speculate this may be due to internal differences for how Android and iOS devices may collected data but this is not made clear by the authors who elect to use (or not use) the battery state as a data stream. Keystroke dynamics appeared the least out of all steams being used in just 3% ( $N = 2$ ) studies. This is likely due to keystroke dynamics being used for specific, tailored tasks as opposed to sensing broad behavioral changes.

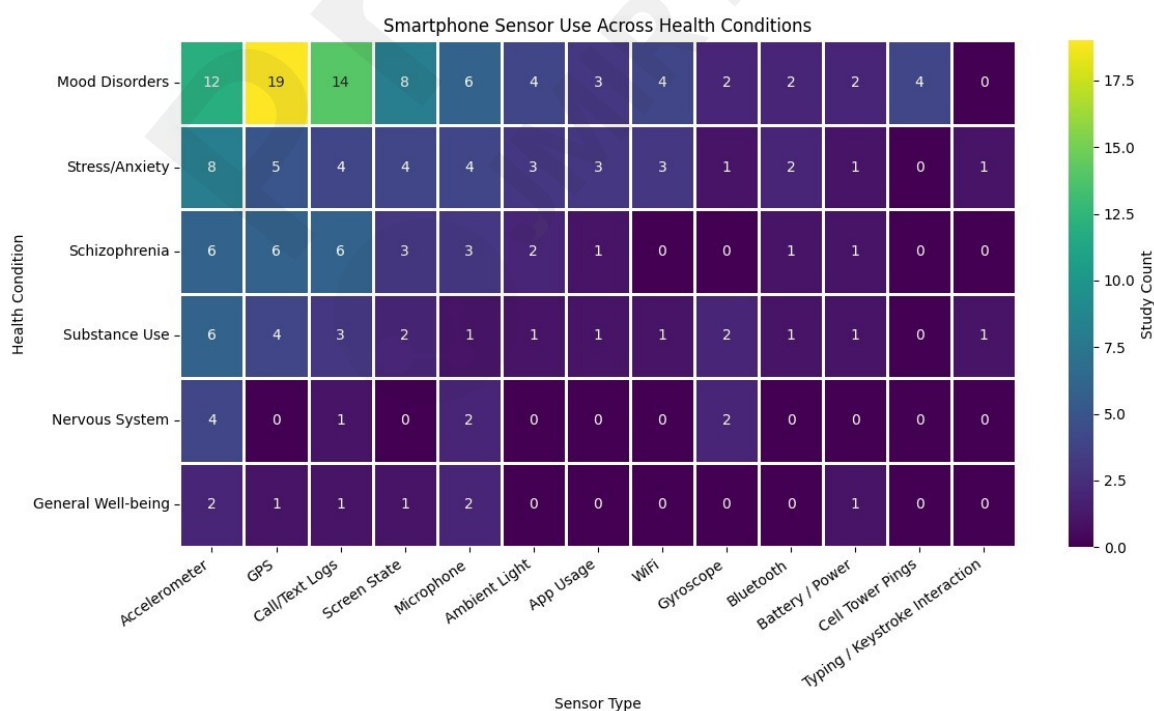


Figure 3: Heatmap illustrating the frequency with which different smartphone sensor modalities

were employed across six major health domains in the reviewed literature. Brighter and warmer-colored cells indicate more frequent usage of a given sensor type within that domain.

## Discussion

This scoping review synthesized research that used *smartphone-based digital phenotyping* to understand health-related conditions. Most studies in our review ( $N = 45$ ) used smartphone-based digital phenotyping for helping with *mental health conditions*. Within the wider health scope, depression and bipolar disorder are of particular interest to researchers due to the ability to effectively track behavioral and mood fluctuations over time through the use of passive smartphone sensors. In contrast, relatively few studies explored other domains such as substance use ( $N = 7$ ), neurological disorders ( $N = 4$ ), or sleep disorders ( $N = 7$ ).

When classified by focus, about 63.5% of the studies ( $N = 41$ ) were conducted with the intent to *monitor*, which we have defined as tracking symptoms, behaviors, and environmental contexts over time. A smaller subset of these studies 23.1% ( $N = 15$ ) looked at phenotypes for the purpose of *diagnosis*, which involves using smartphone-derived features to identify the presence or severity of health conditions. Finally, a few studies totaling 3% ( $N = 2$ ) implemented or evaluated *intervention*-based approaches in which smartphone-derived behaviors were used to inform treatment strategies or real-time interventions.

Collectively, these findings demonstrate that while smartphone-based phenotyping has some potential across monitoring, diagnostic, and intervention-focused studies for a variety of health conditions, there are several aspects of this research domain that could be improved, which we discuss below.

## Methodological inconsistencies and gaps

We found that 15.3% ( $N = 10$ ) of included research studies lacked a clear description of which sensor streams were being used in their phenotyping work and why. This lack of detail about the streams take several forms: (1) instances where the sensor streams are mentioned as being in scope but are not mentioned in the collection process described [65, 46, 36, 111, 37]; (2) instances where a variety of sensor streams are measured from participants but only a subset are used in the phenotyping work [94, 82]; and (3) instances where the type of sensor stream collected to capture a phenotype is not even specified [67]. These limitations hinder continued iteration on the described work and suggests the need for guidelines for how to report data surrounding smartphone-based digital phenotyping work to enhance reproducibility, learn from successes and failures of groups conducting this work and ultimately advance these tools towards broader impacts. Each of these instances prevents us from fully understanding the scope of the work being done and hinders future research from properly expanding upon the findings of these works. We recommend at the minimum, research groups should describe the types of smartphones used in the study, their operating system, the specific sensor streams and the strategies in which the data was obtained. This methodology could benefit others who seek to leverage smartphone-based digital phenotyping to detect and understand disease trajectory.

These reporting issues extend beyond sensor specifications. Despite all studies using smartphones as the sole sensing device, few provided sufficient technical or procedural detail to allow for replication. Only 9.2% ( $N = 6$ ) of the studies quantitatively reported missingness at the stream level [65, 46, 36, 59, 37, 94]. Similarly, only 6.1% ( $N = 4$ ) of the studies explicitly reported sampling rates or duty cycles [46, 84, 85, 87] with all other studies leaving the frequency of data capture and transmission unclear. The most common detail reported was phone OS with 33.8% ( $N = 22$ ) of the studies stating the phone operating system used in data collection, indicating whether the study employed Android, iOS, or mixed devices [112, 80, 88,

61, 92, 91, 47, 45, 82, 87, 67, 68, 65, 46, 84, 85, 59, 36, 111, 37, 89, 42]. Finally, only 10.7% ( $N = 7$ ) of the studies discussed participant compliance or attrition, offering information about adherence, dropout, or data completeness at the participant level [87, 68, 65, 46, 84, 36, 37].

Machine learning (ML) models are commonly used in smartphone-based digital phenotyping work to determine the presence of specific health conditions when monitoring or diagnosing them. Of the 65 studies we assessed as part of this work, about 90.7% ( $N = 59$ ) utilized ML models, while about 9.3% ( $N = 6$ ) did not [86, 105, 110, 108, 48, 69]. For our discussion on the use of ML models we will be noting statistics as being out of 59 so as to not unfairly include studies that do not utilize machine learning. About 69.4% ( $N = 41$  out of 59) of the studies that utilized ML models in their smartphone-based phenotyping work did delve deeper in to how these models were learning and what derived features may mean in the context of a given health outcome. However, we found that a substantial number about 30.5% ( $N = 18$  out of 59) of the studies did not engage with ML models beyond the perfunctory [102, 99, 100, 66, 63, 39, 36, 106, 111, 79, 118, 83, 82, 107, 109, 88, 50, 47]. These studies reported how accurate these models are, with respect to detecting the presence of a health condition in a dataset. While this kind of analysis is important to understanding if models are working, they often eschewed any other kind of analysis on why the model is making its decision. We found that this emphasis on how accurate an ML model performed did not provide any insight into the validity of the models themselves or an understanding of a person's behavior being used as phenotypes for characterizing a specific health condition. We believe this points to a tendency to *discover without understanding* in a substantial portion of the studies that utilized ML approaches, is a cause for numerous errors and bad science [119].

We believe that when ML models are discussed, an effort should be made in exploring the various features used to determine the presence of specific health conditions in addition to reporting on the global accuracy of detection. Reporting on the features allows us, at the very least, to sanity check if the model is making decisions based on known behavioral patterns that individuals with specific health conditions exhibit. But more importantly, such an analysis has the potential to provide insights into newer ways of identifying these health conditions. This would advance the science of detecting specific health conditions and thus truly leverage the pattern matching capabilities of ML models.

Finally, we found that despite the amount of work in smartphone-based digital phenotyping, there is a lack of publicly available datasets in this space. The primary issue with not having any data available is that researchers are required to spend a significant amount of time recruiting participants, collecting data, and cleaning data before they perform any kind of analysis. If data was to be made available, innovation and further research would be significantly easier. Only 10.7% ( $N = 7$ ) studies discuss data availability. Of the studies that provide access to their data, two utilize the Dartmouth StudentLife dataset which is openly available [85, 84]. About 4.6% ( $N = 3$ ) of the studies indicate that the data is available upon request to the authors [43, 89, 69]. One study indicates that the information is individually identifiable and it must be requested through the institutional ethics board [90]. Only 1.5% ( $N = 1$ ) of the studies provides a working link to the GPS data collected during the study [44]. This suggests that researchers should consider strategies to provide data for other teams to advance smartphone-based digital phenotyping as an innovative strategy to address multiple different health conditions. In this process, researchers should consult with their respective institutional ethics boards and ethicists to understand the implications of data sharing, especially for groups whose phenotyping data can be perceived as being more sensitive (e.g., specific geolocation and text message data). In some cases, it may be necessary to restrict data access for certain streams of phenotyping data to protect the confidentiality of participants. At the same time, considerations surrounding consent during enrollment for future data sharing to advance the field should be encouraged among research teams. We encourage authors to disclose how to access the data. In the case that data is

not readily accessible and must be requested through specific channels, we feel that authors should state this. In the case that the data is not available under any circumstances, we feel that authors should disclose this as well.

In this scoping review we provide a comprehensive synthesis of smartphone-based digital phenotyping research across a broader range of health conditions than has been examined in prior surveys. By focusing exclusively on data collected from smartphone sensors, this work differs from existing reviews that either incorporate auxiliary sensing modalities or limit their scope to a small number of predominantly mental health conditions. From a real-world perspective, these results underscore the potential of smartphones as scalable, unobtrusive tools for monitoring, diagnosis, and intervention across mental health, physical health, and substance use disorders while also highlighting the need for improved reporting practices and validation to support responsible deployment.

## Limitations

This scoping review has a few of limitations, though we do not believe these affected the findings in this study. First, we excluded non-English and non-peer-reviewed literature, which may have resulted in missing relevant work conducted in other regions or early-stage research not yet published in academic journals. Second, we only searched Google Scholar, IEEE Xplore, ACM Digital Library, and PubMed databases for identifying the studies for this scoping review. Again, searching other databases could have revealed more studies in this space, which we did not include in this review. Third, as a scoping review we aimed to map available literature to our three given focus categories rather than assessing the quality or strength of each individual study included. As an additional aspect to this point, studies included in this scoping review vary widely in multiple aspects including the length of the study, the number of participants, and study setting, which makes direct comparison between studies challenging. Finally, most studies in this review were observational and conducted in relatively small samples, often within specific demographic groups (e.g., students). Due to this, the behavioral associations identified should be interpreted as correlational rather than causal, and their applicability to broader or more diverse populations remains an open question.

Despite these limitations, this scoping review provides a comprehensive overview of smartphone-based digital phenotyping across health domains and highlights important gaps that future work must address, particularly around reporting standards, validation practices, and data sharing.

## Conclusion

In this survey we examined the work done in the domain of smartphone-based digital phenotyping in determining the presence of various health conditions. Our scope was much more expansive than prior surveys in this space and includes any and all health conditions that have been explored using smartphone-based phenotyping work. Further, and most importantly, in this work we focus on describing the relationships between behaviors exhibited in the context of specific health conditions and not on algorithms used to extract these behaviors. Overall, based on a survey of 65 studies, we found that smartphone-based digital phenotyping is a powerful and informative tool that can give us insights into a variety of health conditions. Though most of the work has focused on applying smartphone-based digital phenotyping to mental health issues, such as depression, bipolar disorder, and stress, some effort has also gone into phenotyping sleep and neurological conditions.

Smartphone-based digital phenotyping offers substantial promise for understanding health-

related behaviors in everyday life. Despite this promise, continued progress depends on addressing key issues surrounding data quality, reproducibility, and reporting standards. Further, as smartphone-based phenotyping scales to larger and more heterogeneous populations, privacy and consent challenges will become increasingly important. Ensuring ethical collection, sharing, and use of sensitive behavioral data will be essential to maintaining participant trust and supporting responsible growth of this research domain. Smartphones are no longer simple communication tools but they are emerging as primary instruments for continuous health monitoring.

## Acknowledgments

The authors would like to thank the reviewers for providing excellent comments that were very helpful in improving the quality of this scoping review.

## Funding

This work was supported by the National Institutes of Health grant DP2DA056107. The funders had no role in decision to publish or preparation of the manuscript.

## Conflicts of Interest

The authors declare they have no competing interests.

## Data Availability

No data was created for this scoping review.

## Authors' Contributions

Authors contributions are not disclosed at the time of review to preserve the anonymity of the authors. Conceptualization: Arlen Dumas, Krishna Venkatasubramanian, and Peter R. Chai

Data curation: Arlen Dumas and Joanne Hokayem

Formal analysis: Arlen Dumas, Joanne Hokayem

Funding acquisition: Peter R. Chai

Investigation: Arlen Dumas and Krishna Venkatasubramanian

Methodology: Arlen Dumas, Krishna Venkatasubramanian, and Peter R. Chai

Project administration: Georgia R. Goodman and Peter R. Chai Resources: Peter R. Chai

Supervision: Krishna Venkatasubramanian and Peter R. Chai

Writing – original draft: Arlen Dumas, Krishna Venkatasubramanian, and Peter R. Chai

Writing – review and editing: Arlen Dumas and Krishna Venkatasubramanian

## Abbreviations

- ADHD – Attention Deficit Hyperactivity Disorder

- BAC – Blood Alcohol Concentration
- EMA – Ecological Momentary Assessment
- GPS – Global Positioning System
- ML – Machine Learning
- PHQ – Patient Health Questionnaire
- SUD – Substance Use Disorder

## References

- [1] Center PR. Mobile Technology and Home Broadband 2021; 2021. Accessed: 202505-07. Available from: <https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/>.
- [2] Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. *P & T: a peer-reviewed journal for formulary management*. 2014;39(5):356-64. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4029126/>.
- [3] Farrington R, et al. Mobile health technology and cancer care: recent advances and future directions. *JMIR Cancer*. 2024. Available from: <https://pubmed.ncbi.nlm.nih.gov/39413226/>.
- [4] Karim SA, et al. Telemedicine and the transformation of health care delivery during the COVID-19 pandemic: The case of radiation oncology. *JAMA*. 2021;325(13):1323-4. Available from: <https://jamanetwork.com/journals/jama/fullarticle/2775696>.
- [5] of the National Coordinator for Health IT O. Trends in Individuals' Access, Viewing, and Use of Online Medical Records and Other Technology for Health Needs: 2022; 2023. Accessed: 2025-0507. Available from: <https://www.healthit.gov/data/data-briefs/trends-individuals-access-viewing-and-use-online-medical-records-and-other>. [6] Walsh CG, et al. Digital phenotyping: the next frontier in mental health care. *npj Digital Medicine*. 2023;6(1). Available from: <https://www.nature.com/articles/s41746-023-00971-z>.
- [7] Chan SR, Torous JB, Hinton L, Yellowlees PM. The use of smartphones for health research. *Academic Medicine*. 2017;92(2):157-60. Available from: [https://journals.lww.com/academicmedicine/fulltext/2017/02000/the\\_use\\_of\\_smartphones\\_for\\_health\\_research.15.aspx](https://journals.lww.com/academicmedicine/fulltext/2017/02000/the_use_of_smartphones_for_health_research.15.aspx). [8] Davidson TM, et al. Digital health tools and mobile applications for mental health. *Journal of Medical Internet Research*. 2017;19(4):e101. Available from: <https://www.jmir.org/2017/4/e101/>.
- [9] Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. *Neuropsychopharmacology*. 2021;46(1):45-54.
- [10] Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annual review of clinical psychology*. 2017;13:23-47.
- [11] Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*. 2016;11(6):838-54.
- [12] Insel TR. Digital phenotyping: technology for a new science of behavior. *Jama*. 2017;318(13):1215-6.

- [13] Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*. 2016;41(7):1691-6.
- [14] Lee K, Lee TC, Yefimova M, Kumar S, Puga F, Azuero A, et al. Using digital phenotyping to understand health-related outcomes: A scoping review. *International Journal of Medical Informatics*. 2023:105061.
- [15] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*. 2018;169(7):467-73.
- [16] Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews*. 2021;10(1):39.
- [17] Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of Biomedical Informatics*. 2018;77:120-32.
- [18] Fraccaro P, Beukenhorst A, Sperrin M, Harper S, Palmier-Claus J, Lewis S, et al. Digital biomarkers from geolocation data in bipolar disorder and schizophrenia: a systematic review. *Journal of the American Medical Informatics Association*. 2019;26(11):1412-20.
- [19] Jayakumar P, Lin E, Galea V, Mathew AJ, Panda N, Vetter I, et al. Digital phenotyping and patient-generated health data for outcome measurement in surgical care: a scoping review. *Journal of Personalized Medicine*. 2020;10(4):282.
- [20] Saccaro LF, Amatori G, Cappelli A, Mazziotti R, Dell'Osso L, Rutigliano G. Portable technologies for digital phenotyping of bipolar disorder: A systematic review. *Journal of affective disorders*. 2021;295:323-38.
- [21] Hilty DM, Armstrong CM, Luxton DD, Gentry MT, Krupinski EA. A scoping review of sensors, wearables, and remote monitoring for behavioral health: uses, outcomes, clinical competencies, and research directions. *Journal of Technology in Behavioral Science*. 2021;6:278-313.
- [22] De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. *NPJ digital medicine*. 2022;5(1):3.
- [23] Dlima SD, Shevade S, Menezes SR, Ganju A. Digital Phenotyping in health using machine learning approaches: scoping review. *JMIR Bioinformatics and Biotechnology*. 2022;3(1):e39618.
- [24] Mendes JP, Moura IR, Van de Ven P, Viana D, Silva FJ, Coutinho LR, et al. Sensing apps and public data sets for digital phenotyping of mental health: systematic review. *Journal of medical Internet research*. 2022;24(2):e28735.
- [25] Alfalahi H, Dias SB, Khandoker AH, Chaudhuri KR, Hadjileontiadis LJ. A scoping review of neurodegenerative manifestations in explainable digital phenotyping. *npj Parkinson's Disease*. 2023;9(1):49.
- [26] Dobson R, Li LL, Garner K, Tane T, McCool J, Whittaker R, et al. The use of sensors to detect anxiety for in-the-moment intervention: scoping review. *JMIR Mental Health*. 2023;10(1):e42611.
- [27] Bufano G, Maisto D, Jirsa V, et al. Digital phenotyping for monitoring mental disorders: systematic review. *JMIR Mental Health*. 2023;10:e45706.
- [28] Zhang Y, Kumar S, Smith L, et al. Passive social sensing with smartphones: a systematic review. *ACM Computing Surveys*. 2023;55(9):1-30.
- [29] Moura I, Teles A, Viana D, Marques J, Coutinho L, Silva F. Digital phenotyping of mental health using multimodal sensing of multiple situations of interest: A systematic literature review. *Journal of Biomedical Informatics*. 2023;138:104278.
- [30] Beames J, Green MJ, Werner-Seidler A. Use of smartphone sensor data in detecting and predicting depression and anxiety in young people (12–25 years): A scoping review.

JMIR Mental Health. 2024;11:e47206.

- [31] Leaning IE, Ikani N, Savage HS, Leow A, Beckmann C, Ruhé HG, et al. From smartphone data to clinically relevant predictions: A systematic review of digital phenotyping methods in depression. *Neuroscience & Biobehavioral Reviews*. 2024;105541.
- [32] Choi A, Ooi A, Lottridge D, et al. Digital Phenotyping for Stress, Anxiety, and Mild Depression: Systematic Literature Review. *JMIR mHealth and uHealth*. 2024;12(1):e40689.
- [33] Linardon J, Chapman A, et al. Smartphone digital phenotyping in mental health disorders: A review of raw sensors utilized, machine learning processing pipelines, and derived behavioral features. *Psychiatry Research*. 2025;327:115593.
- [34] Kallio J, Vildjiounaite E, Tervonen J, Bordallo López M. A Survey on Sensor-Based Techniques for Continuous Stress Monitoring in Knowledge Work Environments. *ACM Trans Comput Healthcare*. 2025 May;6(3). Available from: <https://doi.org/10.1145/3720551>.
- [35] Booth BM, Mundnich K, Feng T, Nadarajan A, Falk TH, Villatte JL, et al. Multimodal human and environmental sensing for longitudinal behavioral studies in naturalistic settings: Framework for sensor selection, deployment, and management. *Journal of medical Internet research*. 2019;21(8):e12832.
- [36] Choudhary S, Thomas N, Ellenberger J, Srinivasan G, Cohen R, et al. A machine learning approach for detecting digital behavioral patterns of depression using nonintrusive smartphone data (complementary path to patient health questionnaire-9 assessment): prospective observational study. *JMIR Formative Research*. 2022;6(5):e37736.
- [37] Berrouiguet S, Ramírez D, Barrigón ML, Moreno-Muñoz P, Camacho RC, Baca-García E, et al. Combining continuous smartphone native sensors data capture and unsupervised data mining techniques for behavioral changes detection: a case series of the evidence-based behavior (eB2) study. *JMIR mHealth and uHealth*. 2018;6(12):e9472.
- [38] Jacobson NC, Chung YJ. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. *Sensors*. 2020;20(12):3572.
- [39] Wang Y, Ren X, Liu X, Zhu T, et al. Examining the correlation between depression and social behavior on smartphones through usage metadata: empirical study. *JMIR mHealth and uHealth*. 2021;9(1):e19046.
- [40] Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of medical Internet research*. 2015;17(7):e4273.
- [41] Sarda A, Munuswamy S, Sarda S, Subramanian V, et al. Using passive smartphone sensing for improved risk stratification of patients with depression and diabetes: cross-sectional observational study. *JMIR mHealth and uHealth*. 2019;7(1):e11041.
- [42] Cao J, Truong AL, Banu S, Shah AA, Sabharwal A, Moukaddam N, et al. Tracking and predicting depressive symptoms of adolescents using smartphone-based self-reports, parental evaluations, and passive phone sensor data: development and usability study. *JMIR mental health*. 2020;7(1):e14045.
- [43] Kim JS, Wang B, Kim M, Lee J, Kim H, Roh D, et al. Prediction of diagnosis and treatment response in adolescents with depression by using a smartphone app and deep learning approaches: usability study. *JMIR formative research*. 2023;7(1):e45991.
- [44] Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ*.

- 2016;4:e2537.
- [45] Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In: Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing; 2015. p. 1293-304.
- [46] Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S, et al. Mobile sensing and support for people with depression: a pilot trial in the wild. *JMIR mHealth and uHealth*. 2016;4(3):e5960.
- [47] Masud MT, Mamun MA, Thapa K, Lee D, Griffiths MD, Yang SH. Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. *Journal of biomedical informatics*. 2020;103:103371.
- [48] Melcher J, Lavoie J, Hays R, D’Mello R, Rauseo-Ricupero N, Camacho E, et al. Digital phenotyping of student mental health during COVID-19: an observational study of 100 college students. *Journal of American College Health*. 2023;71(3):736-48.
- [49] Tlachac M, Flores R, Reisch M, Kayastha R, Taurich N, Melican V, et al. StudentSADD: Rapid mobile depression and suicidal ideation screening of college students during the coronavirus pandemic. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. 2022;6(2):1-32.
- [50] Higuchi M, Nakamura M, Shinohara S, Omiya Y, Takano T, Mitsuyoshi S, et al. Effectiveness of a voice-based mental health evaluation system for mobile devices: prospective study. *JMIR formative research*. 2020;4(7):e16455.
- [51] Tlachac M, Flores R, Toto E, Rundensteiner E. Early mental health uncovering with short scripted and unscripted voice recordings. In: *Deep Learning Applications, Volume 4*. Springer; 2022. p. 79-110.
- [52] Sekhon S, Gupta V. Mood disorder; 2020. Accessed: 2024-06-02. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK558911/>.
- [53] World Health Organization. Depressive disorder (depression); 2023. Accessed: 2024-05-27. Available from: <https://www.who.int/news-room/fact-sheets/detail/depression>.
- [54] Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*. 2001;16(9):606-13.
- [55] Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement*. 1977;1(3):385-401.
- [56] Poznanski EO, Grossman JA, Buchsbaum Y, Banegas M, Freeman L, Gibbons R. Children’s depression rating scale–revised. *Journal of the American Academy of Child Psychiatry*. 1995.
- [57] Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (edgmont)*. 2007;4(7):28.
- [58] Mercier HW, Hamner JW, Torous J, Onnela JP, Taylor JA. Digital phenotyping to quantify psychosocial well-being trajectories after spinal cord injury. *American journal of physical medicine & rehabilitation*. 2020;99(12):1138-44.
- [59] MacLeod L, Suruliraj B, Gall D, Bessenyei K, Hamm S, Romkey I, et al. A mobile sensing app to monitor youth mental health: observational pilot study. *JMIR mHealth and uHealth*. 2021;9(10):e20638.
- [60] Beiwinkel T, Kindermann S, Maier A, Kerl C, Moock J, Barbian G, et al. Using smartphones to monitor bipolar disorder symptoms: a pilot study. *JMIR mental health*. 2016;3(1):e4560.
- [61] Grünerbl A, Muaremi A, Osmani V, Bahle G, Oehler S, Tröster G, et al. Smartphonebased recognition of states and state changes in bipolar disorder patients.

- IEEE journal of biomedical and health informatics. 2014;19(1):140-8.
- [62] Faurholt-Jepsen M, Vinberg M, Frost M, Debel S, Margrethe Christensen E, Bardram JE, et al. Behavioral activities collected through smartphones and the association with illness activity in bipolar disorder. *International journal of methods in psychiatric research*. 2016;25(4):309-23.
- [63] Ryan KA, Babu P, Easter R, Saunders E, Lee AJ, Klasnja P, et al. A smartphone app to monitor mood symptoms in bipolar disorder: Development and usability study. *JMIR mental health*. 2020;7(9):e19476.
- [64] Grünerbl A, Oleksy P, Bahle G, Haring C, Weppner J, Lukowicz P. Towards smart phone based monitoring of bipolar disorder. In: *Proceedings of the Second ACM Workshop on Mobile Systems, Applications, and Services for HealthCare*; 2012. p. 1-6.
- [65] Palmius N, Tsanas A, Saunders KE, Bilderbeck AC, Geddes JR, Goodwin GM, et al. Detecting bipolar depression from geographic location data. *IEEE Transactions on Biomedical Engineering*. 2016;64(8):1761-71.
- [66] Tseng YC, Lin ECL, Wu CH, Huang HL, Chen PS. Associations among smartphone app-based measurements of mood, sleep and activity in bipolar disorder. *Psychiatry Research*. 2022;310:114425.
- [67] Abdullah S, Matthews M, Frank E, Doherty G, Gay G, Choudhury T. Automatic detection of social rhythms in bipolar disorder. *Journal of the American Medical Informatics Association*. 2016;23(3):538-43.
- [68] Faurholt-Jepsen M, Busk J, Frost M, Vinberg M, Christensen EM, Winther O, et al. Voice analysis as an objective state marker in bipolar disorder. *Translational psychiatry*. 2016;6(7):e856-6.
- [69] Fellendorf FT, Hamm C, Dalkner N, Platzer M, Sattler MC, Bengesser SA, et al. Monitoring sleep changes via a smartphone app in bipolar disorder: practical issues and validation of a potential diagnostic tool. *Frontiers in Psychiatry*. 2021;12:641241.
- [70] Alvarez-Lozano J, Osmani V, Mayora O, Frost M, Bardram J, Faurholt-Jepsen M, et al. Tell me your apps and I will tell you your mood: correlation of apps usage with bipolar disorder state. In: *Proceedings of the 7th international conference on pervasive technologies related to assistive environments*; 2014. p. 1-7.
- [71] National Institute of Mental Health. Bipolar Disorder; 2024. Accessed: 2024-05-28. Available from: <https://www.nimh.nih.gov/health/topics/bipolar-disorder>.
- [72] Hamilton M. Development of a rating scale for primary depressive illness. *British journal of social and clinical psychology*. 1967;6(4):278-96.
- [73] Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *The British journal of psychiatry*. 1978;133(5):429-35.
- [74] Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. *Jama*. 2007;298(14):1685-7.
- [75] National Institute of Mental Health. Anxiety Disorder; 2024. Accessed: 2024-05-29. Available from: <https://www.nimh.nih.gov/health/topics/anxiety-disorders>.
- [76] Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine*. 2006;166(10):1092-7.
- [77] Jacobson NC, Summers B, Wilhelm S. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. *Journal of medical Internet research*. 2020;22(5):e16875.
- [78] Gao Y, Li A, Zhu T, Liu X, Liu X. How smartphone usage correlates with social anxiety and loneliness. *PeerJ*. 2016;4:e2197.
- [79] Fukazawa Y, Ito T, Okimura T, Yamashita Y, Maeda T, Ota J. Predicting anxiety state using smartphone-based passive sensing. *Journal of biomedical informatics*. 2019;93:103151.

- [80] Maxhuni A, Hernandez-Leal P, Morales EF, Sucar LE, Osmani V, Mayora O. Unobtrusive stress assessment using smartphones. *IEEE Transactions on Mobile Computing*. 2020;20(6):2313-25.
- [81] Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-Generation Psychiatric Assessment: Using Smartphone Sensors to Monitor Behavior and Mental Health (vol 38, pg 218, 2015). *Psychiatric Rehabilitation Journal*. 2015;38(4):313-3.
- [82] Garcia-Ceja E, Osmani V, Mayora O. Automatic stress detection in working environments from smartphones' accelerometer data: a first step. *IEEE journal of biomedical and health informatics*. 2015;20(4):1053-60.
- [83] Bogomolov A, Lepri B, Ferron M, Pianesi F, Pentland A. Daily stress recognition from mobile phone data, weather conditions and individual traits. In: *Proceedings of the 22nd ACM international conference on Multimedia*; 2014. p. 477-86.
- [84] Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. 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. p. 3-14.
- [85] Wang R, Harari G, Hao P, Zhou X, Campbell AT. SmartGPA: how smartphones can assess and predict academic performance of college students. In: *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*; 2015. p. 295-306.
- [86] Exposito M, Hernandez J, Picard RW. Affective keys: Towards unobtrusive stress sensing of smartphone users. In: *proceedings of the 20th international conference on human-computer interaction with Mobile devices and services adjunct*; 2018. p. 139-45.
- [87] Buck B, Scherer E, Brian R, Wang R, Wang W, Campbell A, et al. Relationships between smartphone social behavior and relapse in schizophrenia: a preliminary report. *Schizophrenia research*. 2019;208:167-72.
- [88] Difrancesco S, Fraccaro P, Van Der Veer SN, Alshoumr B, Ainsworth J, Bellazzi R, et al. Out-of-home activity recognition from GPS data in schizophrenic patients. In: *2016 IEEE 29th international symposium on computer-based medical systems (CBMS)*. IEEE; 2016. p. 324-8.
- [89] Depp CA, Bashem J, Moore RC, Holden JL, Mikhael T, Swendsen J, et al. GPS mobility as a digital biomarker of negative symptoms in schizophrenia: a case control study. *NPJ digital medicine*. 2019;2(1):108.
- [90] Henson P, Barnett I, Keshavan M, Torous J. Towards clinically actionable digital phenotyping targets in schizophrenia. *npj Schizophrenia*. 2020;6(1):13.
- [91] Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela JP. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology*. 2018;43(8):1660-6.
- [92] Wang R, Aung MS, Abdullah S, Brian R, Campbell AT, Choudhury T, et al. CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In: *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*; 2016. p. 886-97.
- [93] Strauss GP, Raugh IM, Zhang L, Luther L, Chapman HC, Allen DN, et al. Validation of accelerometry as a digital phenotyping measure of negative symptoms in schizophrenia. *Schizophrenia*. 2022;8(1):37.
- [94] Staples P, Torous J, Barnett I, Carlson K, Sandoval L, Keshavan M, et al. A comparison of passive and active estimates of sleep in a cohort with schizophrenia. *NPJ schizophrenia*. 2017;3(1):37.
- [95] Joseph SM, Siddiqui W. Delusional disorder; 2023. Accessed: 2024-06-02. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK539855/>.

- [96] Correll CU, Schooler NR. Negative symptoms in schizophrenia: a review and clinical guide for recognition, assessment, and treatment. *Neuropsychiatric disease and treatment*. 2020;519-34.
- [97] Bae S, Ferreira D, Suffoletto B, Puyana JC, Kurtz R, Chung T, et al. Detecting drinking episodes in young adults using smartphone-based sensors. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*. 2017;1(2):1-36.
- [98] Wu T, Sherman G, Giorgi S, Thanneeru P, Ungar LH, Kamath PS, et al. Smartphone sensor data estimate alcohol craving in a cohort of patients with alcohol-associated liver disease and alcohol use disorder. *Hepatology Communications*. 2023;7(12):e0329.
- [99] Arnold Z, Larose D, Agu E. Smartphone inference of alcohol consumption levels from gait. In: *2015 International Conference on Healthcare Informatics*. IEEE; 2015. p. 417-26.
- [100] Li R, Balakrishnan GP, Nie J, Li Y, Agu E, Grimone K, et al. Estimation of blood alcohol concentration from smartphone gait data using neural networks. *IEEE Access*. 2021;9:61237-55.
- [101] Bae SW, Chung T, Islam R, Suffoletto B, Du J, Jang S, et al. Mobile phone sensorbased detection of subjective cannabis intoxication in young adults: A feasibility study in real-world settings. *Drug and alcohol dependence*. 2021;228:108972.
- [102] Nandakumar R, Gollakota S, Sunshine JE. Opioid overdose detection using smartphones. *Science translational medicine*. 2019;11(474):eaau8914.
- [103] Epstein DH, Tyburski M, Kowalczyk WJ, Burgess-Hull AJ, Phillips KA, Curtis BL, et al. Prediction of stress and drug craving ninety minutes in the future with passively collected GPS data. *NPJ digital medicine*. 2020;3(1):26.
- [104] for Disease Control C, Prevention. Substance Use; 2024. <https://wwwnc.cdc.gov/travel/yellowbook/2024/additional-considerations/substance-use>.
- [105] Alqassim S, Ganesh M, Khoja S, Zaidi M, Aloul F, Sagahyroon A. Sleep apnea monitoring using mobile phones. In: *2012 IEEE 14th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE; 2012. p. 443-6.
- [106] Tiron R, Lyon G, Kilroy H, Osman A, Kelly N, O'Mahony N, et al. Screening for obstructive sleep apnea with novel hybrid acoustic smartphone app technology. *Journal of Thoracic Disease*. 2020;12(8):4476.
- [107] Capecci M, Pepa L, Verdini F, Ceravolo MG. A smartphone-based architecture to detect and quantify freezing of gait in Parkinson's disease. *Gait & posture*. 2016;50:2833.
- [108] Ellis RJ, Ng YS, Zhu S, Tan DM, Anderson B, Schlaug G, et al. A validated smartphone-based assessment of gait and gait variability in Parkinson's disease. *PLoS one*. 2015;10(10):e0141694.
- [109] Kostikis N, Hristu-Varsakelis D, Arnaoutoglou M, Kotsavasiloglou C. Smartphonebased evaluation of parkinsonian hand tremor: Quantitative measurements vs clinical assessment scores. In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE; 2014. p. 906-9.
- [110] Barrantes S, Sánchez Egea AJ, González Rojas HA, Martí MJ, Compta Y, Valldeoriola F, et al. Differential diagnosis between Parkinson's disease and essential tremor using the smartphone's accelerometer. *PloS one*. 2017;12(8):e0183843.
- [111] Ware S, Knouse LE, Draz I, Enikeeva A. Predicting ADHD Symptoms Using Smartphone Sensing Data. In: *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*; 2022. p. 503-5.
- [112] Servia-Rodríguez S, Rachuri KK, Mascolo C, Rentfrow PJ, Lathia N, Sandstrom GM. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In:

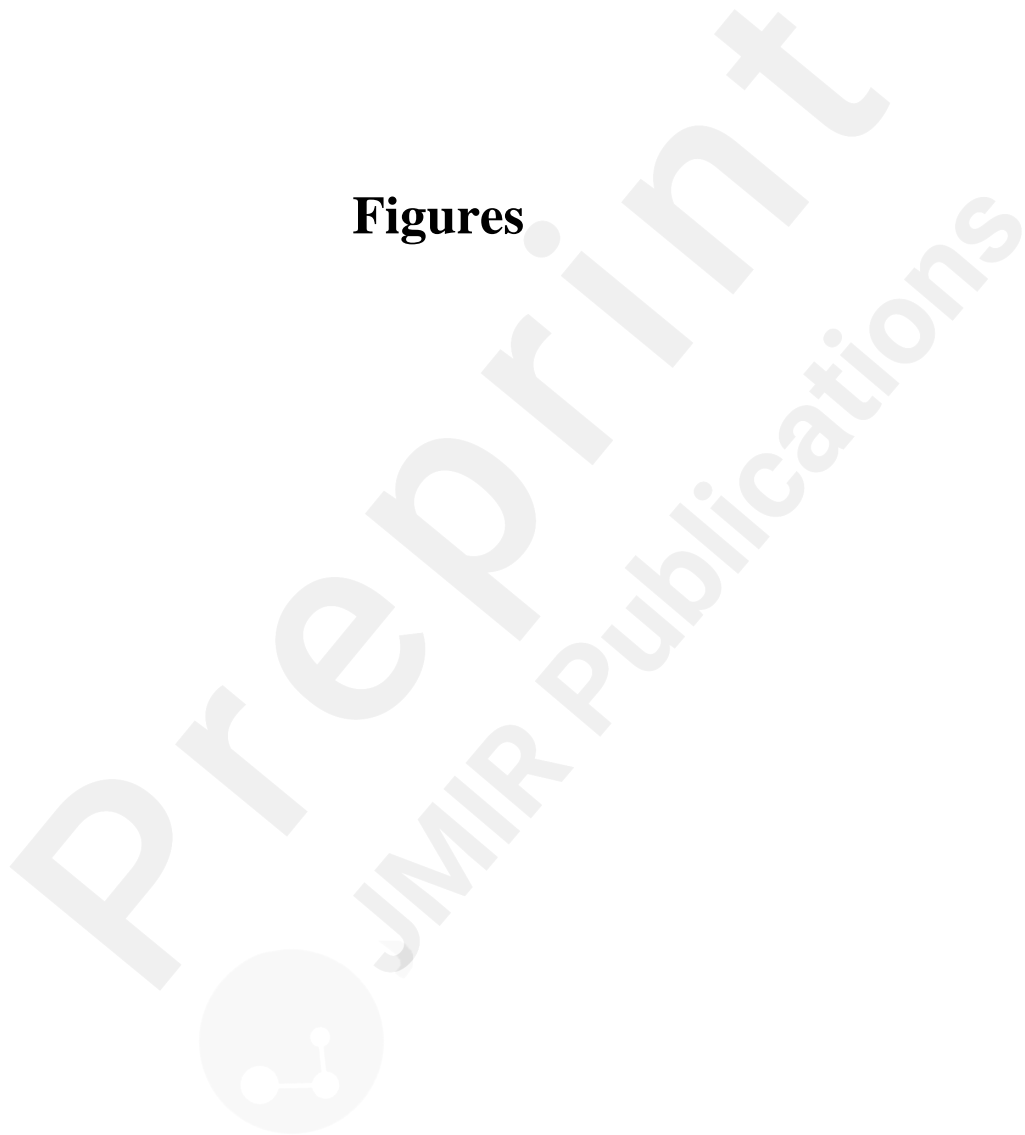
- Proceedings of the 26th international conference on world wide web; 2017. p. 103-12.
- [113] Lane ND, Lin M, Mohammad M, Yang X, Lu H, Cardone G, et al. Bewell: Sensing sleep, physical activities and social interactions to promote wellbeing. *Mobile Networks and Applications*. 2014;19:345-59.
- [114] Mayo Clinic Staff. Sleep apnea - Symptoms and causes; 2023. Available from: <https://www.mayoclinic.org/diseases-conditions/sleep-apnea/symptoms-causes/syc-20377631>.
- [115] American Psychiatric Association. What Are Sleep Disorders?; 2024. Available from: <https://www.psychiatry.org/patients-families/sleep-disorders/what-are-sleep-disorders>.
- [116] Thakur KT, Albanese E, Giannakopoulos P, Jette N, Linde M, Prince MJ, et al. Neurological disorders. *Disease control priorities*. 2016;4:87-107.
- [117] National Institute of Neurological Disorders and Stroke. Parkinson's Disease; 2024. Available from: <https://www.ninds.nih.gov/health-information/disorders/parkinsons-disease>.
- [118] Charron E, White A, Carlston K, Abdullah W, Baylis JD, Pierce S, et al. Prospective acceptability of digital phenotyping among pregnant and parenting people with opioid use disorder: A multisite qualitative study. *Frontiers in Psychiatry*. 2023;14:1137071.
- [119] Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learningbased science. *Patterns*. 2023 Aug. Publisher: Elsevier. Available from: [https://www.cell.com/patterns/abstract/S2666-3899\(23\)00159-9](https://www.cell.com/patterns/abstract/S2666-3899(23)00159-9).

## Supplementary Files

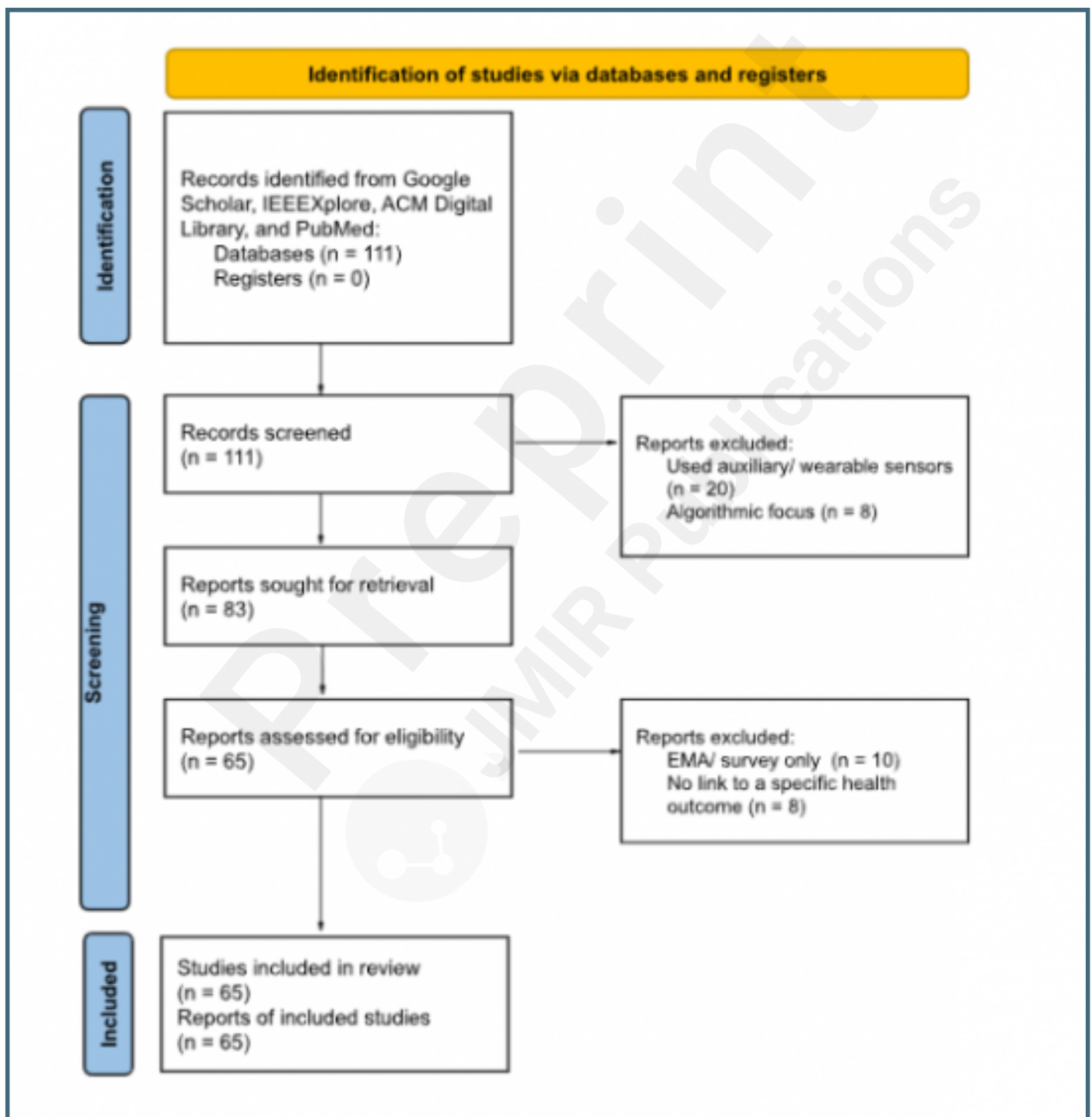
PRISMA for abstracts.

URL: <http://asset.jmir.pub/assets/559afa47d28d2c1c18b6a3cfd00fec90.docx>

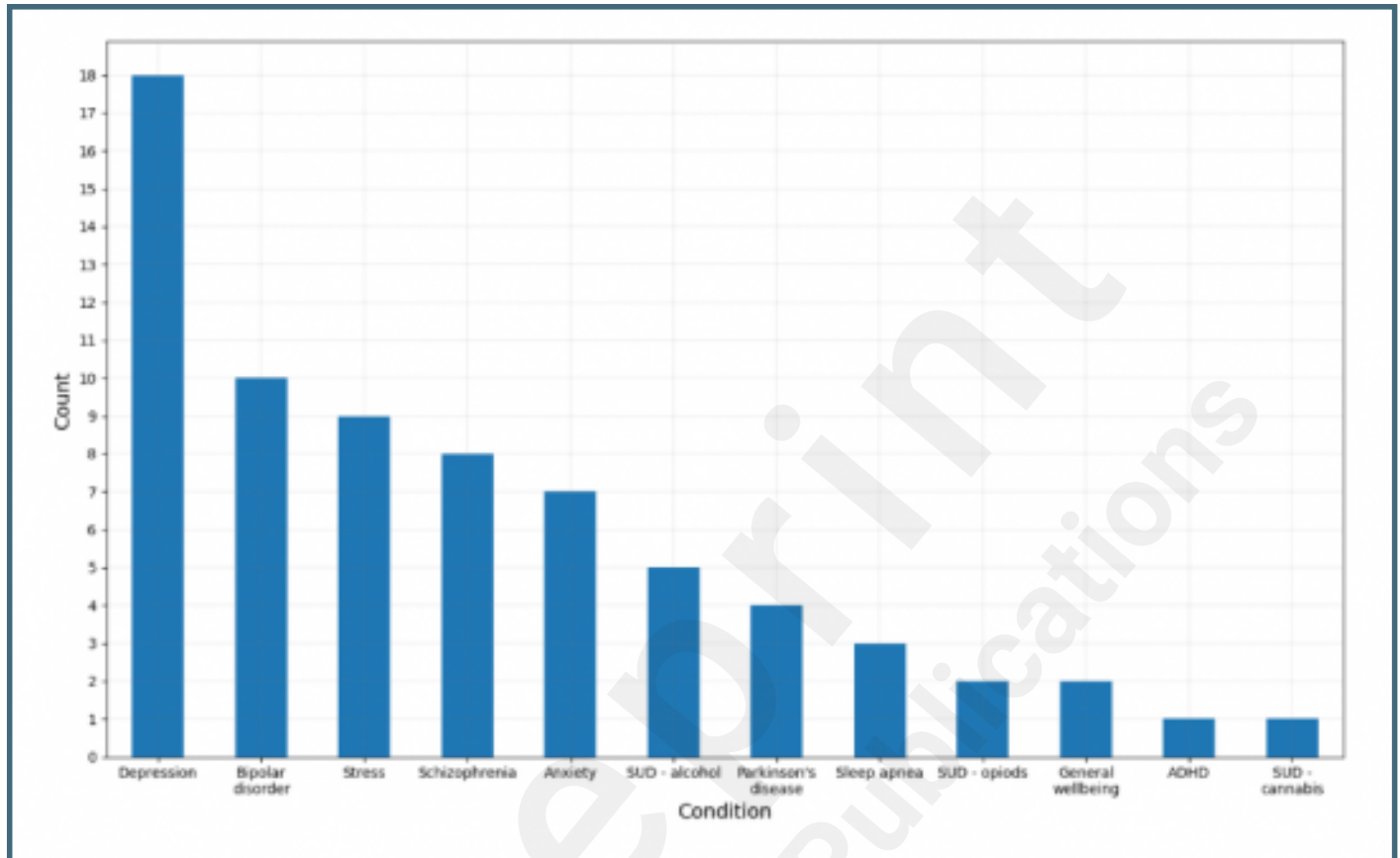
## Figures



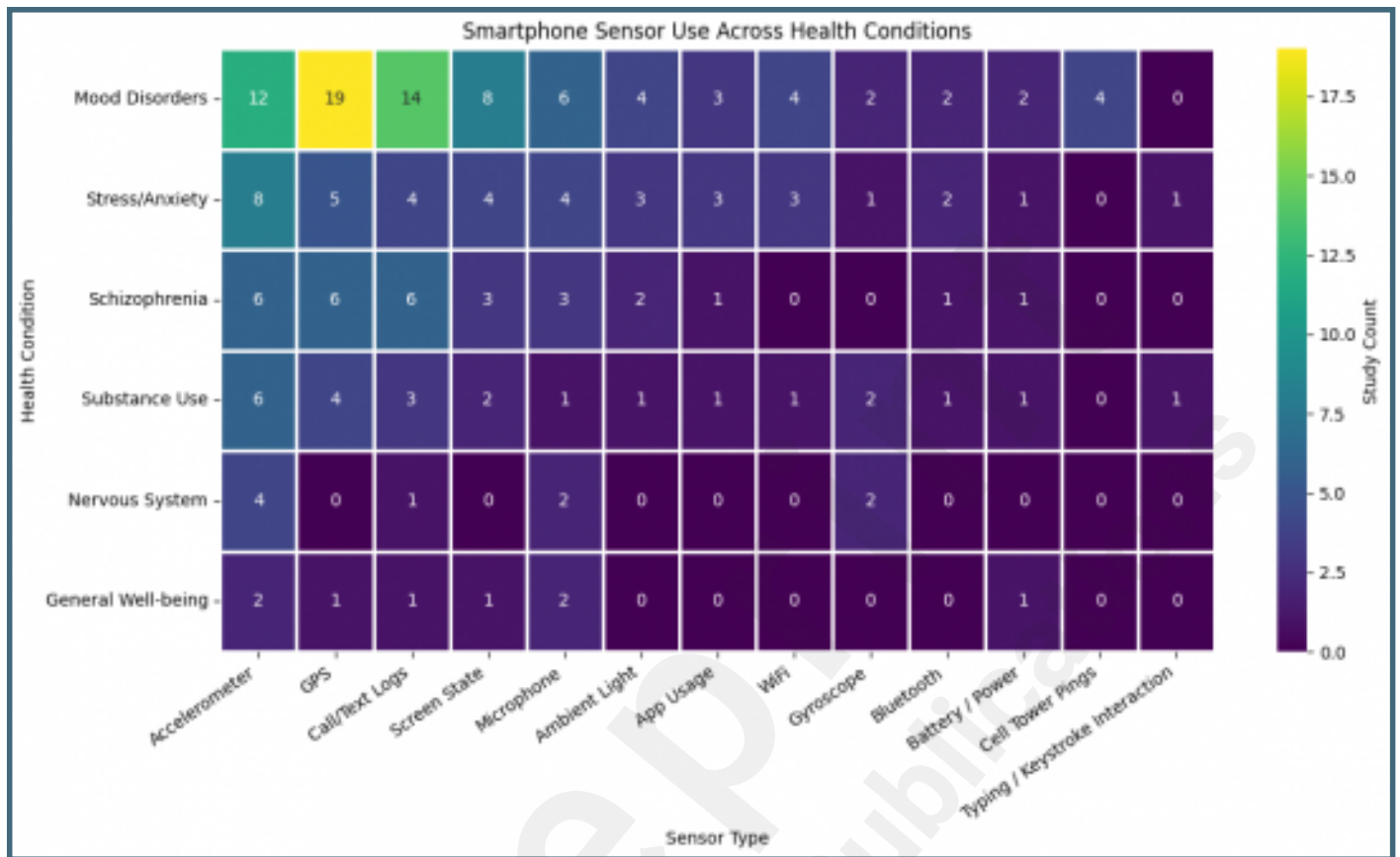
PRISMA flow diagram illustrating the screening and selection process for papers included in this scoping review. The initial search across four databases (Google Scholar, IEEE Xplore, ACM Digital Library, and PubMed) identified 111 records published between 2012 and October 2025. After removal of non-relevant records during title and abstract screening, 65 full-text articles met the inclusion criteria for smartphone-based digital phenotyping using on-device sensor data and a defined health-related outcome. Papers were excluded for reasons including: use of auxiliary sensors or wearables as primary data sources, EMA-only designs without sensor data, absence of a health-related condition or outcome, or non-peer-reviewed publication type. The final set of 65 papers formed the basis of the analysis presented in this review.



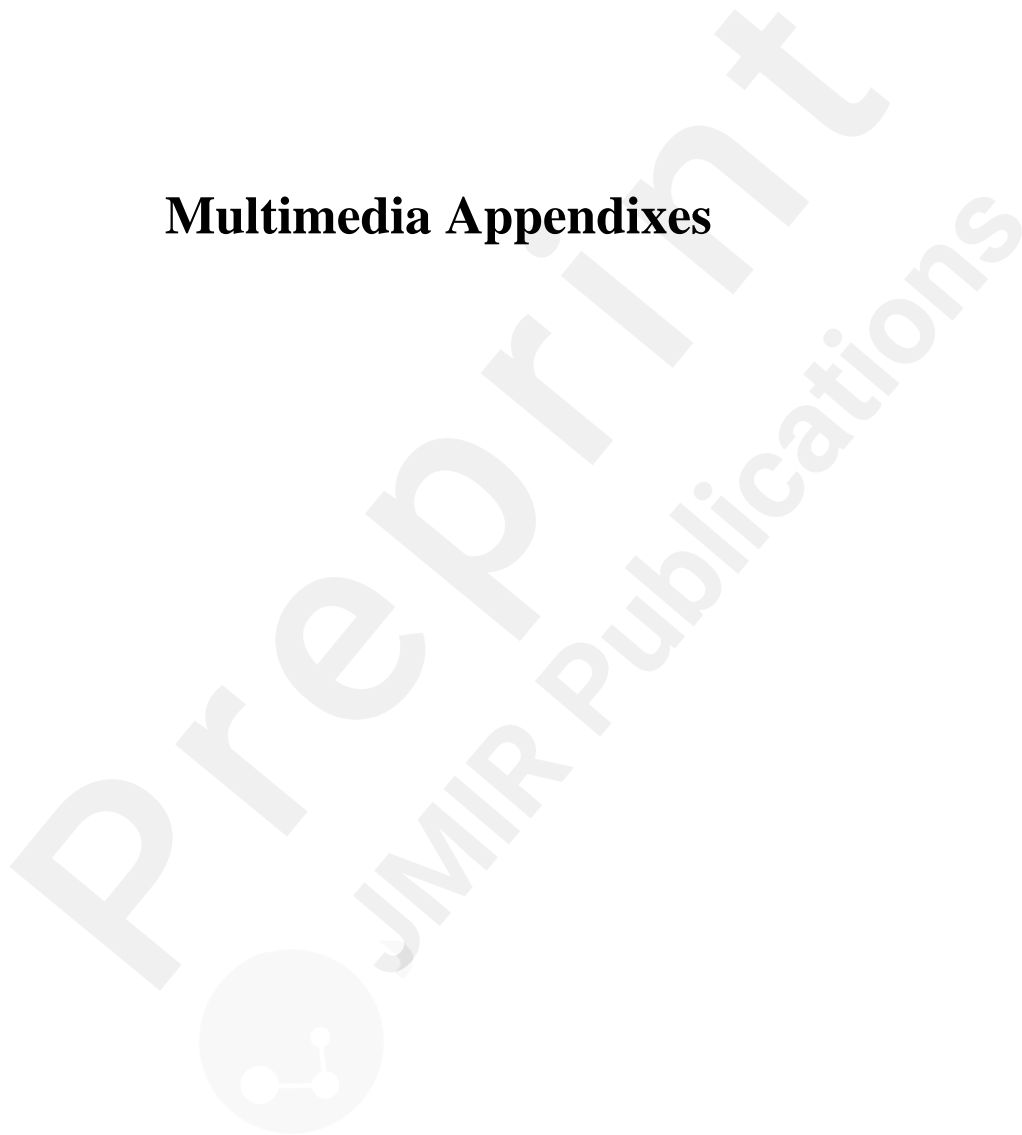
This bar graph depicts the counts of health conditions considered by the papers in this survey. For papers that cover multiple health conditions are counted for each condition (i.e., papers that consider both anxiety and depression are present in the counts for both conditions). Here, "SUD" stands for substance use disorder.



Heatmap illustrating the frequency with which different smartphone sensor modalities were employed across six major health domains in the reviewed literature. Brighter and warmer-colored cells indicate more frequent usage of a given sensor type within that domain.



## Multimedia Appendixes



Search strategies.

URL: <http://asset.jmir.pub/assets/bc674fc059191d92649ffe8a3d5d781c.docx>

