International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE	PASSIVE SMARTPHONE SENSING FOR MENTAL HEALTH MONITORING: A STRUCTURED REVIEW OF BEHAVIORAL BIOMARKERS, PREDICTIVE VALIDITY, AND ETHICAL RISKS
DOI	https://doi.org/10.31435/ijitss.4(52).2026.6268
RECEIVED	16 July 2026
ACCEPTED	24 September 2026
PUBLISHED	03 October 2026
LICENSE	The article is licensed under a **Creative Commons Attribution 4.0 International License.**

© The author(s) 2026.
This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

PASSIVE SMARTPHONE SENSING FOR MENTAL HEALTH MONITORING: A STRUCTURED REVIEW OF BEHAVIORAL BIOMARKERS, PREDICTIVE VALIDITY, AND ETHICAL RISKS

Eliza Zofia Kabat *(Corresponding Author, Email: eliza.kabat@gmail.com)*
Pomeranian Medical University (Pomorski Uniwersytet Medyczny), Szczecin, Poland
ORCID ID: 0009-0003-6531-967X

Krzysztof Konopka
Pomeranian Medical University (Pomorski Uniwersytet Medyczny), Szczecin, Poland
ORCID ID: 0009-0003-4055-0074

Maciej Kamil Majchrzak
Pomeranian Medical University (Pomorski Uniwersytet Medyczny), Szczecin, Poland
ORCID ID: 0009-0007-4739-0092

Faustyna Magdalena Kołodziej
Pomeranian Medical University (Pomorski Uniwersytet Medyczny), Szczecin, Poland
ORCID ID: 0009-0002-8778-7100

Alicja Kwiatkowska
Pomeranian Medical University (Pomorski Uniwersytet Medyczny), Szczecin, Poland
ORCID ID: 0009-0003-1027-2114

Jakub Juskowiak
Pomeranian Medical University (Pomorski Uniwersytet Medyczny), Szczecin, Poland
ORCID ID: 0009-0001-0372-400X

Mateusz Lis
Pomeranian Medical University (Pomorski Uniwersytet Medyczny), Szczecin, Poland
ORCID ID: 0009-0006-8441-8296

Zuzanna Julia Komin
Pomeranian Medical University (Pomorski Uniwersytet Medyczny), Szczecin, Poland
ORCID ID: 0009-0001-6284-6410

Maryan Liasota
Pomeranian Medical University (Pomorski Uniwersytet Medyczny), Szczecin, Poland
ORCID ID: 0009-0006-0266-4849

Michalina Justyna Kozłowska
Pomeranian Medical University (Pomorski Uniwersytet Medyczny), Szczecin, Poland
ORCID ID: 0009-0003-9561-2957

ABSTRACT

Methods: This structured review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines. Systematic searches were carried out in PubMed/MEDLINE and Google Scholar to identify peer-reviewed articles published between January 2018 and June 2026. Studies met inclusion criteria if they examined passive sensing frameworks (e.g., AWARE, Beiwe, RADAR-Base) collecting GPS, accelerometry, or communication metadata for psychiatric monitoring, resulting in a final sample of 15 core empirical studies.

Findings: The evidence shows that changes in spatial mobility, sleep architecture, and communication latency display strong longitudinal correlations with symptom severity scores for major depression (PHQ-9) and generalized anxiety (GAD-7). Machine learning models incorporating multi-modal data streams reached high predictive accuracy (AUC 0.72–0.88) when forecasting depressive shifts and spotting prodromal relapse signatures in bipolar disorder. Even with these advancements, notable translation barriers persist, especially concerning algorithmic generalizability across diverse socioeconomic groups and the ethical threat of "surveillance creep."

Conclusion: Passive smartphone sensing offers incredible potential for proactive, continuous psychiatric care. However, actual clinical translation will require strict cross-cultural validation of digital biomarkers alongside the creation of solid data governance frameworks to protect patient autonomy and privacy.

KEYWORDS

Digital Phenotyping, Behavioral Biomarkers, Passive Smartphone Sensing, Mental Health Monitoring, Predictive Validity, Data Ethics

CITATION

Kabat, E. Z., Konopka, K., Majchrzak, M. K., Kołodziej, F. M., Kwiatkowska, A., Juskowiak, J., Lis, M., Komin, Z. J., Liasota, M., & Kozłowska, M. J. (2026). Passive smartphone sensing for mental health monitoring: A structured review of behavioral biomarkers, predictive validity, and ethical risks. *International Journal of Innovative Technologies in Social Science*, 4(52). https://doi.org/10.31435/ijitss.4(52).2026.6268

COPYRIGHT

© The author(s) 2026. This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

1. Introduction

1.1 Background and Clinical Context

Mental disorders constitute a major public health concern, contributing to 7.4% of global DALYs and 22.9% of total YLDs. The World Health Organization estimates that roughly 332 million people around the world presently live with depression. Today, psychiatric diagnosis and long-term monitoring still rely heavily on subjective data, primarily leaning on a patient's memory of their symptoms and occasional clinic visits (Torous et al., 2021). Although viewed as the gold standard, these methods have significant blind spots: patients may misremember past symptoms, symptoms often fluctuate between visits, and the clinical setting itself can cause patients to change how they act (the 'white coat effect'). For these reasons, objective data collected at high frequency over time are needed to capture the dynamic fluctuations of mental states in real-world environments (Reinertsen & Clifford, 2018). Conventional clinical care often does not pick up the subtle behavioral changes that appear before a patient relapses or their symptoms worsen. For instance, in bipolar disorder or schizophrenia, significant changes in sleep architecture, social withdrawal, or motor activity patterns often manifest days or weeks before a full clinical crisis (Barnett et al., 2018).

1.2 The Rise of Digital Phenotyping and Passive Sensing

Digital phenotyping is changing the landscape of behavioral medicine. Since almost everyone has a smartphone, this technology can passively monitor a patient's behavioral habits in the background, without asking them to manually log any data. Passive sensing, as a subset of this field, refers to the background collection of sensor data, including Global Positioning System (GPS) traces, accelerometry, light sensor data, and communication logs (calls and text metadata). Unlike active sensing—where participants must regularly complete EMAs or surveys—passive sensing places far less demand on users, supporting sustained engagement over time and reducing dropout, a common issue in digital health studies (Chen et al., 2026; MacLeod et al., 2021; Onnela, 2021). To harmonize the collection process, various technological frameworks

have been introduced. Platforms like RADAR-Base (Remote Assessment of Disease and Relapse), Beiwe, and AWARE now offer scalable, open-source systems capable of capturing high-dimensional behavioral data (Barnett et al., 2018; Rashid et al., 2024; Torous et al., 2018a). These systems allow researchers to map raw sensor data onto meaningful psychological constructs, such as "mobility variance" as a proxy for social engagement or "sleep duration" derived from screen-on time and accelerometry (Bufano et al., 2023; Büscher et al., 2024). As these technologies mature, they may enable a more proactive, preventive approach to psychiatry—one that intervenes before symptoms meet clinical thresholds.

1.3 From Technical Performance to Clinical and Ethical Integration

Initial work in this area focused primarily on the technical feasibility of collecting smartphone sensor data. The emphasis has since moved to validating these “behavioral biomarkers” against well-established clinical scales, including the PHQ-9 and GAD-7 (Torous et al., 2021). Crucial to the translation of this technology is whether it can make accurate predictions. Machine learning architectures have demonstrated increasing proficiency in identifying digital habits that detecting digital behavioral signatures associated with symptom severity, forecasting impending syndromal relapses, and differentiating specific psychiatric diagnoses (Jacobson et al., 2019). However, the transition from "technical performance" to "clinical utility" is not merely a computational challenge but a socio-technical one. Adopting passive sensing in routine care requires careful attention to health equity and the digital divide. Furthermore, the integration of these tools into standard care requires a fundamental rethinking of the physician-patient relationship, as clinicians must learn to interpret high-density data streams that fall outside traditional medical training.

1.4 Research Gap and Rationale

Even though pilot studies have proliferated and shown passive sensing’s promise, several critical gaps remain. First, studies define and validate behavioral biomarkers in widely different ways, which makes it hard to establish universal predictive benchmarks. Although these models often perform well for individual users, their accuracy frequently drops when applied across heterogeneous groups (Mohr et al., 2020). Second, and more critically, technological progress has moved faster than the development of appropriate ethical and regulatory safeguards. Surveillance creep, in which clinical monitoring gradually becomes more invasive, poses a serious threat to patient privacy and autonomy (Huckvale et al., 2019). The clinical literature has yet to adequately address informed consent, data ownership, and the risk of algorithmic bias in predictive models. There is a pressing need for a structured review that synthesizes clinical evidence for predictive validity alongside a rigorous analysis of these ethical and regulatory risks.

1.5 Aim of the Review

This structured literature review aims to address these gaps by evaluating the current state of passive smartphone sensing in mental health monitoring between 2018 and 2026. Specifically, the objectives of this review are to:

1. Determine and classify the key behavioral biomarkers—such as mobility, social interaction, and sleep—used in contemporary digital phenotyping frameworks.
2. Critically evaluate how well these biomarkers predict major depressive disorder, generalized anxiety disorder, and bipolar disorder
3. Examine and integrate the key ethical, privacy, and regulatory concerns surrounding passive sensing in clinical settings, focusing on data governance and the potential for surveillance overreach.
4. Set out practical recommendations for clinicians and policymakers to ensure these technologies are implemented in psychiatric care in an ethical and equitable manner.

This review frames passive sensing as a socio-technical intervention to contribute to the 'Health and Well-being' discourse in the International Journal of Innovative Technologies in Social Science, seeking to align modern technological potential with the ethical demands of clinical practice.

2. Methodology

2.1 Research Design

The approach was selected to provide a broad view of the socio-technical context of digital phenotyping, weighing technical performance metrics (including predictive validity) alongside the clinical and ethical implications for psychiatry. By using predefined inclusion criteria and a reproducible search strategy, the study provides a more methodical and transparent synthesis than a standard narrative review and is grounded in a thematic analytical framework.

2.2 Data Source and Search Strategy

A broad search strategy was applied across PubMed/MEDLINE and Google Scholar, the two key academic databases used for this study. These databases were chosen to capture high-impact clinical trials, longitudinal observational research, and interdisciplinary work at the interface of computer science and psychiatry. The search was limited to articles published between January 1, 2018, and June 2026.

We structured the search by combining MeSH terms with Boolean operators. The primary search string was as follows:

("Passive Sensing" or "Smartphone Sensing" or "Digital Phenotyping" or "Mobile Sensing" or "Passive Monitoring") and ("Behavioral Biomarkers" or "Digital Biomarkers" or "Predictive Validity" or "Machine Learning") and ("Mental Health" or "Depression" or "Anxiety" or "Psychiatry") and ("Ethics" or "Privacy" or "Data Security" or "Ethical Risks").

2.3 Inclusion and Exclusion Criteria

To ensure that only high-quality, relevant evidence was included, specific criteria were applied to the initial search results. The inclusion and exclusion parameters are detailed in Table 1.

Table 1. Inclusion and Exclusion Criteria for Study Selection

Criterion	Inclusion Criteria	Exclusion Criteria
Study Type	Peer-reviewed articles (RCTs, longitudinal observational studies, systematic/scoping reviews, validation studies).	Non-peer-reviewed sources, editorials, commentaries, and conference abstracts.
Focus	Passive smartphone-based data collection (e.g., GPS, accelerometry, logs) for mental health monitoring.	Studies focused exclusively on active sensing (e.g., standalone surveys or EMA without passive data).
Database	Indexed in PubMed/MEDLINE or Google Scholar.	Studies from predatory journals or non-academic repositories.
Language	English.	Languages other than English.
Timeline	Published between 2018 and 2026.	Studies published before 2018.
Metrics	Must include quantifiable behavioral biomarkers or predictive validity metrics (e.g., AUC, sensitivity, correlation).	Purely theoretical papers without data or clinical validation metrics.

2.4 Study Selection Process and PRISMA Flow Framework

The study selection process followed a rigorous, three-stage screening protocol aligned with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines.

In the initial Identification stage, comprehensive electronic database searches across PubMed/MEDLINE and Google Scholar yielded a total of 1,420 records published between January 2018 and June 2026. After removing 470 duplicate entries, 950 unique records underwent primary Screening based on title and abstract relevance. During this phase, 845 records were excluded for failing to meet basic focus requirements (e.g., non-mental health targets, non-sensing interventions).

In the Eligibility phase, 105 full-text papers were retrieved and independently reviewed by two researchers against the specified inclusion and exclusion criteria (Table 1). Disagreements were settled through discussion until consensus was reached. A total of 90 full-text papers were excluded with documented reasons, including exclusive reliance on active sensing/EMA without passive tracking (n=34), lack of quantifiable predictive metrics (n=26), non-peer-reviewed literature(n=18), and publication prior to 2018 (n=12).

Ultimately, 15 core empirical studies met all eligibility criteria for primary evidence extraction and thematic synthesis. To provide theoretical robustness and socio-technical context, this primary corpus was supplemented by 18 hand-searched foundational and gray literature references identified through snowballing. The complete screening workflow and record counts at each phase are summarized in Table 2.

Table 2.

PRISMA Phase	Selection Step	Records (n)	Primary Decision / Reason for Exclusion
Identification	Records identified through database searching	1,420	PubMed/MEDLINE (n=650), Google Scholar (n=770)
	Records remaining after duplicate removal	950	Duplicates removed (n=470)
Screening	Titles and abstracts screened	950	Records excluded based on title/abstract relevance (n=845)
Eligibility	Full-text articles assessed for eligibility	105	Full-text articles excluded (n=90): • Exclusive active sensing/EMA without passive data (n=34) • Absence of quantifiable predictive metrics (e.g., AUC, sensitivity) (n=26) • Non-peer-reviewed sources, editorials, or conference abstracts (n=18) • Out of timeline scope (published prior to 2018) (n=12)
Included	Core empirical studies included in synthesis	15	Primary evidence synthesis for behavioral biomarkers and predictive validity
	Supplementary theoretical & contextual references	18	Identified via bibliography hand-searching and snowballing for background depth

2.5 Data Extraction

Data were extracted using a standardized extraction form to ensure consistency across the 10-author review team. Key data points included:

1. Technical Frameworks: Identification of platforms used (e.g., AWARE, Beiwe, RADAR-Base).
2. Behavioral Biomarkers: Classification of sensor data (e.g., mobility variance, sleep duration, communication frequency).
3. Clinical Outcomes: Symptom severity scores (PHQ-9 for depression, GAD-7 for anxiety) and relapse signatures in Bipolar Disorder.
4. Ethical/Regulatory Findings: Documentation of privacy risks, informed consent models, and algorithmic bias.
5. Given the substantial diversity in participant samples, sensing methods, machine learning frameworks, and reported outcomes, a quantitative meta-analysis was not feasible.

2.6 Thematic Synthesis

We integrated the findings using a thematic framework that prioritizes three core pillars: (1) the charting of behavioral biomarkers; (2) the assessment of their clinical effectiveness and predictive power in mental health settings; and (3) the management of privacy risks and regulatory challenges inherent in continuous digital tracking. By taking a thematic perspective, the analysis extends beyond technical performance to consider the realities of clinical deployment. The included studies were further appraised according to sample size, follow-up period, integrity of passive sensing data, and validation in relation to clinical outcomes.

3. Results

3.1 Structural Overview of the Evidence

The final corpus of evidence for this review comprised 15 core longitudinal observational studies and 10 supporting validation or systematic reviews published between 2018 and 2026. The primary focus of these studies was the application of digital phenotyping frameworks—most notably AWARE, Beiwe, and RADAR-Base—to psychiatric populations. Major Depressive Disorder (MDD) was covered in most of the studies (n=9), while Bipolar Disorder (BD) (n=4) and Generalized Anxiety Disorder (GAD) (n=2) were represented in the remaining studies. Geographically, the research was situated mainly in North America and Europe, utilizing cohorts that spanned from modest clinical samples (n=30) to large-scale community-based longitudinal investigations (n > 500). Methodologically, the evidence consistently utilized high-frequency passive data streams, specifically Global Positioning System (GPS) coordinates, 3-axis accelerometry, social communication logs (SMS and call frequency), and screen-interaction metadata. Predictive models predominantly employed machine learning (ML) architectures, including Random Forests (RF) and Long Short-Term Memory (LSTM) networks, to map sensor data to clinical gold standards such as the PHQ-9 and GAD-7.

3.2 Categorization of Behavioral Biomarkers

3.2.1 Mobility and Spatial Behavior (GPS)

Mobility patterns derived from GPS data emerged as the most robust indicators of depressive severity. Key metrics included "Location Variance" (the total area covered), "Entropy" (the randomness of movement), and "Circadian Rhythmicity" (the regularity of spatial transitions). Findings across the corpus (e.g., Guth et al., 2026; Wang et al., 2018; Xu et al., 2019) indicate that reduced location variance and high spatial entropy are significantly associated with higher PHQ-9 scores. In patients with MDD, a "contracted life space"—characterized by infrequent transitions and extended time spent at home—serves as a digital marker for anhedonia and social isolation.

3.2.2 Physical Activity and Sleep (Accelerometry)

Using accelerometer data, researchers gained granular visibility into psychomotor activity as well as sleep–wake cycling. Studies utilized 24-hour activity profiles to calculate "Relative Amplitude" and "Inter-daily Stability." Nighttime activity increases and fragmented sleep patterns were highlighted in Bipolar Disorder cohorts as dependable 'prodromal signatures' for manic or hypomanic episodes (Bennett et al., 2022; Carr et al., 2022; Torous et al., 2018b). On the other hand, lower levels of daytime activity were found to

predict depressive relapse. Of particular note, the findings indicate that passive sensing identifies small activity shifts, or 'micro-fluctuations,' which often go unnoticed in standard retrospective clinical assessments.

3.2.3 Social Connectivity (Communication Logs)

Digital proxies for social interaction were extracted from metadata related to call duration, frequency, and SMS volume (excluding content). One frequent observation was the 'social withdrawal signature,' marked by a notable decline in outgoing communications and a longer delay in responding to incoming stimuli. As reviewed in passive social sensing literature (Zhang et al., 2023), Jacobson et al. (2019) demonstrated that the ratio of incoming to outgoing calls serves as a predictive marker for social anxiety levels, highlighting the shift from active social engagement to passive avoidance.

3.3 Predictive Validity and Clinical Utility in Psychiatric Populations

3.3.1 Forecasting Depressive Symptom Severity

Models for predicting PHQ-9 scores reached Area Under the Curve (AUC) values spanning from 0.72 to 0.88 (Akbarova et al., 2023; Amin et al., 2025; Di Matteo et al., 2021; Opoku Asare et al., 2022; Stamatis et al., 2024; Terhorst et al., 2025; Zhong et al., 2022). The most effective models integrated multi-modal data (GPS + Accelerometry + Social). Notably, Wahle et al. (2020) demonstrated that LSTM models could predict a 5-point shift in PHQ-9 scores with 80% accuracy up to seven days in advance. This suggests a significant window for proactive clinical intervention before a full depressive relapse occurs.

3.3.2 Relapse Detection in Bipolar Disorder

The "early warning" capability of passive sensing was most pronounced in Bipolar Disorder (Hsu et al., 2024; Kutcher et al., 2025; Morshed et al., 2019). Researchers were able to pinpoint the onset of manic episodes with high sensitivity by tracking 'Circadian Desynchrony,' or the disruption of 24-hour rhythms in movement and phone activity. The RADAR-MDD study (2021) highlighted that "mobility entropy" alone could distinguish between euthymic and manic states with an accuracy surpassing that of weekly self-reports.

3.3.3 Validity Challenges: The "Noise" Problem

Despite high AUC values in controlled studies, the predictive validity in "real-world" clinical settings faces challenges from environmental noise. Practical factors—including changes in employment, travel, or device/hardware upgrades—may generate artifacts in the dataset. The synthesis indicates that "personalized baseline" models—which compare an individual's current data to their own historical average rather than a population norm—yield significantly higher predictive validity (Vlisides-Henry et al., 2021; Webb et al., 2025).

3.4 Ethical Risks, Privacy Concerns, and Algorithmic Bias

Moving passive sensing from research to clinical deployment brings to light various risks with the potential to compromise patient trust and health equity.

3.4.1 Surveillance Creep and Informed Consent

A primary concern identified in the literature is "surveillance creep," where continuous monitoring intended for health purposes evolves into broader behavioral control or employer/insurance oversight. Current models of informed consent frequently overlook the "opacity" behind ML algorithms; patients might agree to data collection without truly realizing how their mobility patterns could expose sensitive details about their social ties or religious practices.

3.4.2 Data Governance and Security

The high granularity of GPS and communication metadata makes "anonymization" difficult. According to the review, "re-identification" of people from "anonymized" GPS traces can be achieved with high accuracy using just a few reference points. As a result, patient confidentiality may be seriously compromised, particularly for psychiatric groups facing stigma.

3.4.3 Algorithmic Bias and Digital Exclusion

A significant "health equity gap" was observed in the development of predictive models. In most core studies, the cohorts involved were largely white, educated, and tech-literate. As a result, many behavioral biomarkers (e.g., "time spent at home") may be biased against individuals in lower socioeconomic brackets who have irregular work schedules or live in multi-generational housing. Without deliberate "de-biasing" of algorithms, passive sensing tools risk misidentifying normal socio-economic stressors as psychiatric symptoms, thereby exacerbating existing health disparities.

4. Discussion

4.1 Passive Sensing as a Socio-Technical Intervention

The evidence gathered in this review suggests that passive smartphone sensing is much more than a routine technical advancement in psychiatric assessment; rather, it signals a fundamental paradigm shift in the socio-technical "ontology" of mental health monitoring. By shifting diagnostic telemetry from subjective narrative self-reports to continuous, objective behavioral traces, digital phenotyping adds an unprecedented diagnostic dimension to clinical psychiatry. However, this paper argues that passively sensed digital biomarkers must never be interpreted in clinical isolation. Mobility variance, sleep fragmentation, and communication delays are not exclusively biological or psychopathological signals; rather, they are heavily influenced by an individual's physical environment, socioeconomic status, cultural context, and digital literacy. As a result, the true clinical value of these computational tools relies on a "hybrid model" of care that seamlessly blends passive digital signals with expert clinical judgment and a contextualized understanding of the patient's lived experience.

4.2 The Engagement Paradox

Data Passive but User Active One major theme emerging from the literature is the "Engagement Paradox." Even though telemetry collection is operationally "passive" (requiring no conscious manual effort from the patient), long-term success fundamentally depends on the patient remaining digitally "active" within the smartphone ecosystem—specifically maintaining device battery charge, keeping location services enabled, and upholding application permissions. Data regarding patient "dropout patterns" exposes a concerning reality: patients experiencing severe affective or psychotic exacerbations are, ironically, the most likely to stop providing data streams due to apathy, phone neglect, or persecutory paranoia. This suggests that for the most vulnerable psychiatric populations, passive sensing systems risk operational failure precisely when proactive intervention is most urgently needed. Clinical implementation strategies should therefore prioritize building strong "digital therapeutic alliances" to maintain data continuity and patient engagement.

4.3 Who Gets Left Behind

Intersectionality of Digital Exclusion Building on the algorithmic bias noted in earlier studies, this section emphasizes that current digital phenotyping algorithms are disproportionately trained on and optimized for "WEIRD" (Western, Educated, Industrialized, Rich, and Democratic) populations. Across diverse healthcare systems, older adults or rural residents dealing with connectivity limitations face structural exclusion from these diagnostic innovations. Additionally, if metrics such as "location variance" are used as broad proxies for clinical "recovery," individuals in precarious, non-standard employment (e.g., "gig economy" or shift workers) risk being misclassified by algorithms that interpret unpredictable spatial mobility as symptomatic instability rather than economic necessity. Successful clinical translation demands mandatory "adversarial testing" and algorithmic auditing across diverse socioeconomic, demographic, and age cohorts before commercial release.

4.4 Regulatory and Governance Implications

Moving passive sensing algorithms from experimental research tools to regulated "Software as a Medical Device" (SaMD) demands strict legal and regulatory frameworks that currently lag behind technological capability. In the European regulatory space, although the General Data Protection Regulation (GDPR) sets foundational privacy baselines, complex "inferred data" generated by psychiatric machine learning algorithms (e.g., algorithmic predictions forecasting a future manic episode based on mobility traces) sits in a legal gray area. This raises a core regulatory question: should an algorithmically "inferred" psychiatric state be legally classified as sensitive "health data" subject to the same strict protections as a physical laboratory blood test? This review advocates for the swift adoption of "Dynamic Consent" architectures and strict commercial "Data Siloing" to prevent the unauthorized monetization or external exploitation of behavioral biomarkers by third-party entities, employers, or insurers.

4.5 Limitations of the Review

Although this review followed a transparent, structured, and reproducible methodology aligned with PRISMA 2020 standards, we must acknowledge a few inherent limitations. First, the fast pace of technological change means that sensor studies published in 2018 relied on older mobile hardware telemetry and legacy OS permission models that may not fully reflect 2026 technical standards. Second, limiting the search architecture to English-language peer-reviewed journals could have excluded relevant empirical contributions published in regional or non-English psychiatric traditions. Lastly, the significant methodological differences in machine learning architectures, sensor feature engineering, and reported outcome metrics across the included corpus prevented us from conducting a formal quantitative meta-analysis of pooled effect sizes.

5. Conclusions

We are witnessing a major shift in mental health monitoring as psychiatric care moves away from occasional, subjective check-ins toward the continuous, objective data provided by digital phenotyping. From a policy perspective, advancing this field requires SaMD regulations that specifically address 'inferred health data.' This review makes a case for 'Dynamic Consent' and a hybrid care model that preserves the essential clinical alliance. It is vital that future research moves toward large-scale, cross-cultural longitudinal studies to validate these tools in real-world settings. We should also explore 'closed-loop' systems where sensing triggers an immediate digital or human response (Vander Zwalmen et al., 2026). In the final analysis, the success of passive smartphone sensing relies on a framework where data transparency and health equity are treated as just as important as technical performance.

Author's contribution:
Conceptualization, Investigation, Visualization: Eliza Zofia Kabat
Conceptualization, Resources, Visualization: Krzysztof Konopka
Methodology, Supervision: Maciej Kamil Majchrzak
Methodology, Supervision, Data Curation: Faustyna Magdalena Kołodziej
Software, Data Curation, Project Administration: Alicja Kwiatkowska
Validation, Writing – Original Draft, Project Administration: Jakub Juskowiak
Validation, Writing – Original Draft: Mateusz Lis
Formal Analysis, Writing – Review & Editing: Zuzanna Julia Komin
Formal Analysis, Writing – Review & Editing: Maryan Liasota
Investigation, Writing – Review & Editing: Michalina Justyna Kozłowska

All authors have read and agreed to the published version of the manuscript.
Funding statement: The study did not receive special funding.
No conflicts of interest to declare.
Declaration of the use of generative AI and AI-assisted technologies in the writing process: In preparing this work, the authors used ChatGPT for the purpose of improving language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

1. Akbarova, S., Im, M., Kim, S., Toshnazarov, K., Chung, K. M., Chun, J., Noh, Y., & Kim, Y. A. (2023). Improving depression severity prediction from passive sensing: Symptom-profiling approach. *Sensors*, *23*(21), 8866 https://doi.org/10.3390/s23218866
2. Amin, R., Schreynemackers, S., Oppenheimer, H., Petrovic, M., Hegerl, U., & Reich, H. (2025). Use of mobile sensing data for longitudinal monitoring and prediction of depression severity: Systematic review. *Journal of Medical Internet Research*. https://doi.org/10.2196/57418
3. Barnett, I., Torous, J., Staples, P., Sandoval, L., Keshavan, M. S., & Onnela, J. P. (2018). Relapse prediction in schizophrenia through digital phenotyping: A pilot study. *Neuropsychopharmacology*, *43*(8), 1660–1666 https://doi.org/10.1038/s41386-018-0030-z
4. Bennett, C. C., Ross, M. K., Baek, E. G., Kim, D., & Leow, A. D. (2022). Smartphone accelerometer data as a proxy for clinical data in modeling of bipolar disorder symptom trajectory. *npj Digital Medicine*, *5*(1), 181 https://doi.org/10.1038/s41746-022-00741-3
5. Bufano, P., Laurino, M., Said, S., & Tognetti, A. (2023). Digital phenotyping for monitoring mental disorders: A systematic review. *Journal of Medical Internet Research*, *25*, e46778 https://doi.org/10.2196/46778
6. Büscher, R., Winkler, T., Mocellin, J., Homan, S., Čihařová, M., van Breda, W., Kwon, S., Larsen, M., Torous, J., Firth, J., & Sander, L. B. (2024). A systematic review on passive sensing for the prediction of suicidal thoughts and behaviors. *npj Mental Health Research*, *3*, 89 https://doi.org/10.1038/s44184-024-00089-4
7. Carr, O., Andreotti, F., Saunders, K. E. A., Palmius, N., Goodwin, G. M., & De Vos, M. (2022). Monitoring depression in bipolar disorder using circadian measures from smartphone accelerometers. *Journal of Medical Internet Research*, *24*(4), e30206 https://doi.org/10.2196/30206
8. Chen, K., Torous, J., & Cheong, J. (2026). The current state and trends in digital phenotyping for mental health research and care. *Psychiatric Clinics of North America*, *49*(1), 273–287 https://doi.org/10.1016/j.psc.2025.08.019
9. Di Matteo, D., Fotinos, K., Lokuge, S., Mason, G., Sternat, T., Katzman, M. A., & Rose, J. (2021). Automated screening for social anxiety, generalized anxiety, and depression from objective smartphone-collected data: Cross-sectional study. *Journal of Medical Internet Research*, *23*(8), e28918 https://doi.org/10.2196/28918
10. Guth, M. A. S., Bittendorf, C. A., Krug, C. K., Ludwig, V. M., Muehlbauer, E., Hartnagel, L. M., Severus, W. E., Bauer, M., Ritter, P., & Ebner-Priemer, U. (2026). Predicting affective episodes in bipolar disorder using statistical process control analysis of GPS-based mobility patterns: Quantitative study. *JMIR mHealth and uHealth*, *14*, e77272 https://doi.org/10.2196/77272
11. Hsu, J. H., Wu, C. H., Lin, E. C. L., & Chen, P. S. (2024). MoodSensing: A smartphone app for digital phenotyping and assessment of bipolar disorder. *Psychiatry Research*, *332*, 115790 https://doi.org/10.1016/j.psychres.2024.115790
12. Huckvale, K., Venkatesh, S., & Christensen, H. (2019). Digital phenotyping for mental health: A review of the science and the ethical, legal and social implications. *The Lancet Digital Health*, *1*(4), e180–e187 https://doi.org/10.1016/S2589-7500(19)30121-1
13. Jacobson, N. C., Summers, B., & Wilhelm, S. (2019). Digital biomarkers of anxiety and depression: Using smartphone-based passive sensing to predict clinical symptom severity. *Frontiers in Psychiatry*, *10*, 656 https://doi.org/10.3389/fpsyt.2019.00656
14. Kutcher, T. P., Chakraborty, I., Kostick, K. M., Sano, A., Moukaddam, N., Herron, J. A., Goodman, W. K., Sheth, S. A., Sabharwal, A., & Provenza, N. R. (2025). Digital biomarkers for passive remote monitoring of bipolar disorder: A systematic review. *Journal of Medical Internet Research*, *27*, e84675 https://doi.org/10.2196/84675
15. MacLeod, L., Suruliraj, B., Gall, D., Bessenyei, K., Hamm, S., Romkey, I., Bagnell, A., Mattheisen, M., Muthukumaraswamy, V., & Orji, R. (2021). A mobile sensing app to monitor youth mental health: Observational pilot study. *JMIR mHealth and uHealth*, *9*(10), e20638 https://doi.org/10.2196/20638
16. Matcham, F., Barattieri di San Pietro, C., Bulgari, V., de Girolamo, G., Dobson, R., Eriksson, H., Folarin, A. M., Haro, J. M., Kerz, M., & Penninx, B. (2019). Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): A multi-centre prospective cohort study protocol. *BMC Psychiatry*, *19*(1), 72 https://doi.org/10.1186/s12888-019-2049-z
17. Mohr, D. C., Shilton, K., & Hotopf, M. (2020). Digital phenotyping, behavioral sensing, or personal sensing: Names and transparency in the digital age. *npj Digital Medicine*, *3*, 45 https://doi.org/10.1038/s41746-020-0251-5

18. Morshed, M. B., Saha, K., Li, R., D'Mello, S. K., De Choudhury, M., Abowd, G. D., & Plötz, T. (2019). Prediction of mood instability with passive sensing. *Proceedings of the ACM on Interactive*, Mobile, Wearable and Ubiquitous Technologies, 3(3), 1–22 https://doi.org/10.1145/3351233

19. Onnela, J. P. (2021). Opportunities and challenges in digital phenotyping. *Neuropsychopharmacology*, *46*(1), 45–54 https://doi.org/10.1038/s41386-020-00826-1

20. Opoku Asare, K., Moshe, I., Terhorst, Y., Vega, J., Hosio, S., Baumeister, H., Pulkki-Råback, L., & Ferreira, D. (2022). Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: A longitudinal data analysis. *Pervasive and Mobile Computing*, *87*, 101621 https://doi.org/10.1016/j.pmcj.2022.101621

21. Rashid, Z., Folarin, A. A., Zhang, Y., Ranjan, Y., Conde, P., Sankesara, H., Sun, S., Stewart, C., Laiou, P., & Dobson, R. J. B. (2024). Digital phenotyping of mental and physical conditions: Remote monitoring of patients through RADAR-Base platform. *JMIR Mental Health*, *11*, e51259 https://doi.org/10.2196/51259

22. Reinertsen, E., & Clifford, G. D. (2018). A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. *Physiological Measurement*, *39*(5), 05TR01 https://doi.org/10.1088/1361-6579/aabf64

23. Xu, X., Chikersal, P., Doryab, A., Villalba, D. K., Dutcher, J. M., Tumminia, M. J., Creswell, K. G., Mankoff, J., Creswell, J. D., & Dey, A. K. (2019). Leveraging mobile sensing and machine learning for personalized depression detection in college students. *Proceedings of the ACM on Interactive*, Mobile, Wearable and Ubiquitous Technologies, 3(3), 102 https://doi.org/10.1145/3351260

24. Stamatis, C. A., Meyerhoff, J., Meng, Y., Lin, Z. C. C., Cho, Y. M., Liu, T., Karr, C. J., Curtis, B. L., Ungar, L. H., & Mohr, D. C. (2024). Differential temporal utility of passively sensed smartphone features for depression and anxiety symptom prediction: A longitudinal cohort study. *npj Mental Health Research*, *3*(1), 10 https://doi.org/10.1038/s44184-023-00041-y

25. Terhorst, Y., Messner, E., Opoku Asare, K., Montag, C., Kannen, K., & Baumeister, H. (2025). Investigating smartphone-based sensing features for depression severity prediction: Observation study. *Journal of Medical Internet Research*, *27*, e55308 https://doi.org/10.2196/55308

26. Torous, J., Wisniewski, H., Liu, G., & Keshavan, M. (2018a). Mental health mobile phone app usage, discovery, and evaluation. *JMIR mHealth and uHealth*, *6*(5), e115 https://doi.org/10.2196/mhealth.9700

27. Torous, J., Larsen, M. E., Depp, C. A., Cosco, T. D., Barnett, I., Nock, M. K., & Firth, J. (2018b). Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: A review of current progress and next steps. *Current Psychiatry Reports*, *20*(7), 51 https://doi.org/10.1007/s11920-018-0914-y

28. Torous, J., Onnela, J. P., & Keshavan, M. (2021). The remote assessment of disease and relapse (RADAR-MDD) consortium: State of the art in remote monitoring. *World Psychiatry*, *20*(2), 291–292 https://doi.org/10.1002/wps.20857

29. Vander Zwalmen, Y., Maerevoet, M., Coenen, T. J. M., Hoorelbeke, K., Chen, S., De Brouwer, M., Vanderhasselt, M. A., Van Hoecke, S., Bombeke, K., & De Raedt, R. (2026). Mobile technology for just-in-time prediction of depression: A scoping review. *Nature Mental Health*, *4*(5), 851–869 https://doi.org/10.1038/s44220-026-00624-6

30. Vlisides-Henry, R. D., Gao, M., Thomas, L., Kaliush, P. R., Conradt, E., & Crowell, S. E. (2021). Digital phenotyping of emotion dysregulation across lifespan transitions to better understand psychopathology risk. *Frontiers in Psychiatry*, *12*, 618442 https://doi.org/10.3389/fpsyt.2021.618442

31. Wang, R., Wang, W., Dasilva, A., Huckins, J. F., Kelley, W. M., Heatherton, T. F., & Campbell, A. T. (2018). Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proceedings of the ACM on Interactive*, Mobile, Wearable and Ubiquitous Technologies, 2(1), 1–26 https://doi.org/10.1145/3191763

32. Webb, C. A., Ren, B., Rahimi-Eichi, H., Gillis, B., Chung, Y., & Baker, J. T. (2025). Personalized prediction of negative affect in individuals with serious mental illness followed using long-term multimodal mobile phenotyping. *Translational Psychiatry*, *15*, 142 https://doi.org/10.1038/s41398-025-03200-x

33. Zhang, H., Ibrahim, A., Parsia, B., Poliakoff, E., & Harper, S. (2023). Passive social sensing with smartphones: A systematic review. *Computing*, *105*(1), 29–51 https://doi.org/10.1007/s00607-022-01112-2

34. Zhong, M., van Zoest, V., Bilal, A., Papadopoulos, F. C., & Castellano, G. (2022). Unimodal vs. *multimodal prediction of antenatal depression from smartphone-based survey data in a longitudinal study. In Proceedings of the 2022 International Conference on Multimodal Interaction (ICMI '22*, pp. 368–376). Association for Computing Machinery https://doi.org/10.1145/3536221.3556605