Internet Interventions 44 (2026) 100934

Contents lists available at ScienceDirect

Internet Interventions

journal homepage: www.elsevier.com/locate/invent

Smartphone-based digital phenotyping for detection of high-risk depression and anxiety in Korean community settings

Ah Young Kim [a,b], Seonmin Kim [b], Jisu Lee [c], Youngwoong Han [a], Heon-Jeong Lee [d], Chul-Hyun Cho [d,e,*]

[a] Medical Information Research Section, Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea
[b] Department of Biomedical Informatics, Korea University College of Medicine, Seoul, Republic of Korea
[c] Batoners Inc, Daegu, Republic of Korea
[d] Department of Psychiatry, Korea University College of Medicine, Seoul, Republic of Korea
[e] Department of Medical Education, Korea University College of Medicine, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Digital phenotyping
Smartphone sensing
Passive and active data
Depression
Anxiety
High-risk

ABSTRACT

Background: Smartphones generate continuous behavioral signals such as mobility and activity patterns, offering scalable opportunities for monitoring mental health in community settings. Digital phenotyping approaches that integrate passive sensing with brief self-report measures may enable early identification of individuals at high risk for depression and anxiety without reliance on additional wearable devices.

Method: We prospectively evaluated a smartphone-based digital phenotyping framework in 455 community-dwelling adults in Korea who contributed 28 days of passive Global Positioning System and accelerometer data, daily self-report microsurveys, and weekly PHQ-9/GAD-7 assessments for screening high-risk depression and anxiety. Machine learning models were compared across active-only, passive-only, and combined feature sets. After applying predefined coverage criteria (≥60% passive-data coverage and ≥ 60% corresponding active-data availability), 277 participants were included in the depression cohort and 275 in the anxiety cohort.

Results: Passive features capturing mobility, activity regularity, and sleep-related behaviors were derived, and machine learning models were trained using active-only, passive-only, and combined feature sets. For depression, combined models achieved the best performance, with AUCs ranging from 0.77 to 0.83 and APs ranging from 0.86 to 0.91 across classifiers. Similar patterns were observed for anxiety, with AUCs up to 0.86 and APs up to 0.95. Ablation analyses identified robust deployment conditions relevant to clinical screening, including tolerance to missing data and short look-back windows.

Discussion: These findings support the practical utility of smartphone-based digital phenotyping pipelines that integrate passive behavioral signals with brief self-reports for scalable screening for high-risk depression and anxiety in real-world environments, and they may inform future just-in-time mental health intervention systems.

1. Introduction

Depression and anxiety disorders remain leading contributors to global years lived with disability (Collaborators, 2022; Szücs et al., 2025). Existing screening approaches—reliant on clinic visits (Institute of Medicine Forum on Drug Discovery, 2010; Niv et al., 2007) or intermittent self-report questionnaires (Villarreal-Zegarra et al., 2023)—fail to provide continuous monitoring or timely intervention (Huang et al., 2024; Onnela and Rauch, 2016; Saddichha et al., 2014). Modern smartphones, however, continuously generate rich background streams of location and accelerometer data, offering an ideal platform for high-frequency mental health monitoring via "digital phenotyping" and for combining passive sensing with lightweight active self-report data (Leaning et al., 2024; Teo et al., 2019; Unützer and Park, 2012; Zhang et al., 2021).

Previous studies have demonstrated the feasibility of smartphone- and wearable-based digital phenotyping under controlled or semi-controlled conditions. For example, a two-week Purple Robot study

* Corresponding author at: Department of Psychiatry, Biomedical Informatics, and Medical Education, Korea University College of Medicine, 73 Goryeodae-ro, Seongbuk-gu, Seoul, 02841, Republic of Korea.
E-mail address: david0203@korea.ac.kr (C.-H. Cho).

https://doi.org/10.1016/j.invent.2026.100934
Received 21 January 2026; Received in revised form 22 March 2026; Accepted 23 March 2026

extracted normalized entropy and circadian-rhythm features from Global Positioning System (GPS) and phone-usage data, achieving a correlation of $r = -0.63$ with Patient Health Questionnaire-9 (PHQ-9) scores and 86.5% accuracy in distinguishing participants with PHQ-9 $\geq$ 5 (Saeb et al., 2015). However, this preliminary study was limited to correlational associations in a small, nonrepresentative sample using self-reported symptoms, precluding causal inference. Similarly, smartwatch-based research monitored total motor activity, heart rate metrics, and sleep–wake ratios in patients with psychotic-spectrum disorders for up to fourteen months, uncovering considerable links with Positive and Negative Syndrome Scale scores (Kalisperakis et al., 2023). Large-scale wearable analyses across 800 first-year physicians and 50,000+ monitoring days revealed bidirectional associations between internal circadian misalignment and mood changes, while a thirteen-week study of 249 patients with moderate-to-severe depression showed that physiological sleep metrics and self-reported sleep disturbances predicted distinct symptom domains (Lee et al., 2024).

A smartphone-only sensing strategy represents a pragmatic design choice that prioritizes clinical feasibility and scalability over theoretical optimization. Although wearable devices can provide richer physiological data streams (Boer et al., 2023), smartphone adoption is substantially higher in the general population, whereas lower wearable adoption may introduce selection bias toward more technologically engaged and higher socioeconomic groups (Gelles-Watnick, 2024). From a technical perspective, prior digital phenotyping studies can be broadly grouped into passive-only sensing approaches, active or self-report-based approaches, and multimodal frameworks that combine both (Huckvale et al., 2019; Onnela and Rauch, 2016). Passive-only models reduce participant burden and support continuous behavioral monitoring, but they may miss subjective symptom context and depend heavily on the quality and continuity of passive sensor streams (Choi et al., 2024). Active-only approaches capture self-reported internal states more directly, yet they are more vulnerable to missingness, response fatigue, and declining adherence over time (Moura et al., 2023). Recent systematic reviews further indicate that smartphone sensors can capture behavioral patterns associated with stress, anxiety, and depression, while also highlighting substantial heterogeneity across studies in sensor selection, derived behavioral features, machine-learning pipelines, and reporting practices (Choi et al., 2024; Linardon et al., 2025; Mendes et al., 2022). GPS, in particular, has been identified as one of the most frequently used sensing modalities in digital phenotyping studies, but the broader literature still lacks sufficient methodological standardization and direct head-to-head comparisons of active-only, passive-only, and combined smartphone-based configurations under deployment-relevant conditions (Linardon et al., 2025).

Despite these advances, most existing systems face three major limitations. First, they depend on wearable devices, whose adoption lags behind smartphones, and which are often removed during sleep, restricting continuous data collection (Chandrasekaran and Moustakas, 2025; Shandhi et al., 2024; Veney, 2025). Second, they impose burdensome self-report schedules (Gromatsky et al., 2020; Kim et al., 2019; Magallón-Neri et al., 2016). Third, they lack robust strategies for handling real-world missingness: both smartphones and wearables frequently suffer gaps owing to improper device placement, transmission failures, battery-saving or charging intervals, revoked permissions, and hardware–software heterogeneity (Bähr et al., 2022; Bladon et al., 2025; Du et al., 2020). Observation windows that are too short risk missing transient behavioral shifts, whereas overly long windows introduce heterogeneity that destabilizes predictions. High self-report frequencies further increase participant burden and dropout risk, undermining large-scale deployment (Bladon et al., 2025; Hasselhorn et al., 2022). Together, these limitations point to a need for deployment-oriented evaluations that explicitly compare passive-only, active-only, and combined configurations under realistic data-availability and observation-window constraints.

In this study, we addressed these challenges by collecting naturalistic, real-world data from community-dwelling participants using commercial Android and iOS smartphones. In particular, we systematically compared active-only, passive-only, and combined feature configurations under real-world data-availability constraints. We evaluated a smartphone-only framework for detecting high-risk depression and anxiety by fusing passively captured accelerometer and GPS data with lightweight self-report microprompts. As shown in Fig. 1A, the device continuously analyzes these signals to identify risk-relevant states and inform future just-in-time intervention workflows without additional hardware. We systematically explored the design space (Fig. 1A–C) by defining missing-data tolerance, optimizing the balance between observation-window length and self-report frequency, and compressing sensor inputs into a concise digital phenotype feature set. The final pipeline aggregated weekday and weekend statistics and validated performance with lightweight classifiers, supporting the feasibility of scalable screening and future intervention workflows using smartphone-based sensing.

By embedding clinically actionable analytics directly into everyday smartphones, this approach reduces reliance on specialized wearables or continuous clinician oversight, enabling scalable, low-cost early warning systems for personalized mental health care and population surveillance.

2. Methods

2.1. Recruitment and eligibility

This observational study recruited 455 community-dwelling adults in Korea between April 2024 and June 2025 through community outreach and hospital-based campaigns at medical centers nationwide (Shin et al., 2025). Eligible participants were adults aged 19–59 years who self-identified as Korean, were capable of providing valid informed consent, owned a compatible smartphone (iOS 16.4+ or Android 18.2+), and had basic digital literacy. Individuals were excluded if they had cognitive impairment preventing consent, diagnosed neuropsychiatric conditions that could interfere with participation (e.g., epilepsy, seizure disorders, schizophrenia, psychosis, intellectual disability, or dementia), or were currently enrolled in other interventional studies. Participants contributed 4 weeks of passive smartphone sensing data and daily active self-reports, with weekly PHQ-9 and GAD-7 assessments used to define depression and anxiety risk labels. The study was approved by the Institutional Review Board at Korea University Anam Hospital (approval no. 2023AN0506), and all participants provided electronic informed consent. The study was registered with the Clinical Research Information Service (KCT0009183; February 20, 2024).

2.2. Data collection and weekly outcome labels

After consent, participants installed the PixelMood application (Batoners Inc., South Korea) on their Android or iOS smartphones. During the 4-week study period, PixelMood continuously collected passive smartphone sensor data, including triaxial accelerometer recordings at 1 Hz and GPS latitude/longitude coordinates at 15-min intervals. In parallel, active self-report data were collected through two mechanisms: (1) a daily microprompt assessing mood and energy on a 7-point scale (-3 to $+3$) and anxiety on a 4-point scale (0 to 3); and (2) weekly PHQ-9 and GAD-7 questionnaires beginning 1 week after enrollment and repeated thereafter at weekly intervals ($\pm$2 days).

Weekly PHQ-9 and GAD-7 scores were used to define the primary binary prediction targets. High-risk depression was defined as PHQ-9 $\geq$ 10, and low-risk depression as PHQ-9 $<$ 10. High-risk anxiety was defined as GAD-7 $\geq$ 10, and low-risk anxiety as GAD-7 $<$ 10. These weekly binary outcomes served as the model labels. In this study, PHQ-9 $\geq$ 10 was used as a pragmatic screening threshold to define a PHQ-9-defined high-risk depressive state rather than a formal clinical diagnosis of depression. We therefore interpret the depression label as a

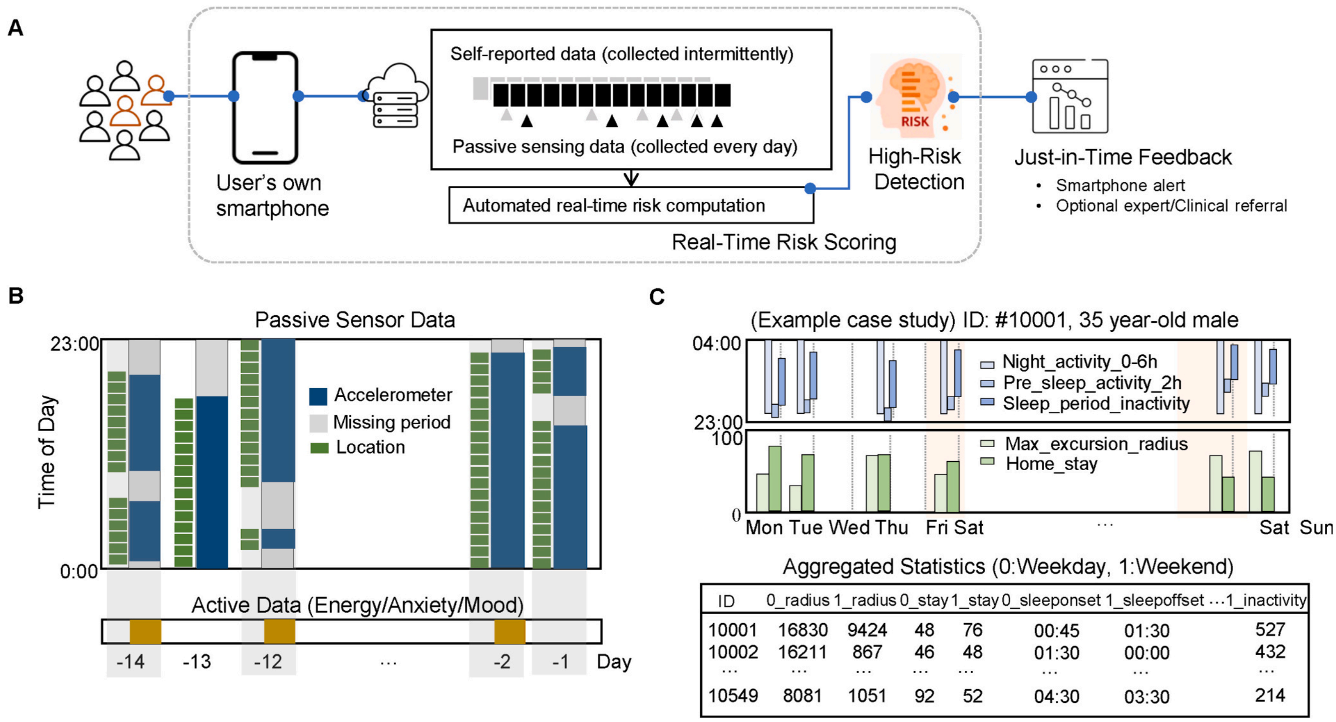

Fig. 1. Framework for real-time detection of high depression and anxiety risk from smartphone sensing data.
A: Participants' smartphones intermittently collected self-reported items (mood, energy, anxiety) and continuously streamed passive accelerometer and GPS signals. These data are processed to compute risk scores that may inform future just-in-time feedback and clinical decision workflows. B: Example 14-day timeline illustrating passive sensor coverage (blue = accelerometer, green = GPS, grey = missing) alongside intermittently collected active self-report data. C: Feature-engineering pipeline: day-level behavioral metrics (such as nocturnal activity 0–6 h, pre-sleep activity ≥2 h, sleep-period inactivity, maximum excursion radius, and home stay) are aggregated into weekday and weekend statistics, yielding the input matrix for downstream machine learning models.
Abbreviations: GPS, Global Positioning System; h, hours.

questionnaire-defined screening target, not as diagnostic ascertainment. To support secure and privacy-preserving study operations, adherence and data integrity were monitored in real time through an encrypted web portal without access to personal communications or in-app content. All encrypted study data were stored on Amazon Web Services and were accessible only to authorized study personnel. At the end of week 4, participants uninstalled the application.

2.3. Observation window construction and label alignment

Participants completed the PHQ-9 and GAD-7 surveys 1 week after enrollment (±2 days) and weekly thereafter, with each survey date defined as the label date. For each label date t, we constructed retrospective observation windows covering the days immediately preceding that label date (Supplementary Fig. S1A). In the ablation analyses, we evaluated look-back windows ranging from 8 to 28 days in 2-day increments. For the primary analyses, we selected a 14-day look-back window on the basis of ablation performance.

Within each observation window (days t–L through t–1), passive sensor features and active self-report features were extracted and then paired with the PHQ-9 or GAD-7 label obtained on day t. Each analytic instance therefore consisted of features summarized from a retrospective observation window and paired with the corresponding weekly binary risk label at day t. To ensure true out-of-sample evaluation, we used a participant-level splitting throughout model development and testing. All repeated observations from a given participant were assigned to a single fold, such that the participant's entire data vector appeared either in the training set or in the test fold, but not both. This design prevented information leakage across folds and ensured that performance reflected generalization to unseen participants.

2.4. Passive feature extraction

To capture behavioral and circadian signals relevant to depression and anxiety, we derived seven passive digital phenotyping features from smartphone GPS and accelerometer data: excursion radius, home-stay ratio, nocturnal activity, pre-sleep activity, sleep-period inactivity, digital sleep onset (DSO), and digital wake-up time (DWU). DSO was defined as the time when motion stopped, detected between 20:00 on the previous day and 12:00 (noon) on the current day, and DWU was defined as the time when motion resumed within the same interval. These accelerometer-derived timing variables were treated as digitally inferred behavioral proxies for sleep-wake timing rather than validated measures of physiological sleep onset or awakening.

These features were calculated for each valid day and then summarized within each look-back window for model input. Specifically, daily values were aggregated separately for weekdays and weekends using the mean and variance, thereby capturing both central tendency and day-to-day variability within each observation window. Detailed definitions and units are provided in Supplementary Table S3.

2.5. Data quality, missingness handling, and window eligibility

Raw accelerometer and GPS streams were first aligned to each participant's retrospective observation window preceding each PHQ-9 or GAD-7 label date (Supplementary Fig. S1). Daily passive-data coverage was defined as the ratio of actual collection time to the expected 24-h recording duration. Days with less than 50% expected passive-data coverage were excluded from feature extraction and treated as invalid days.

To ensure minimum temporal representation within each retrospective window, each analytic window was required to include at least

7 valid weekdays and at least 1 valid weekend day. For the primary analyses, participant-level inclusion additionally required at least 60% passive-data coverage across eligible days, at least 60% corresponding active-data availability, and sufficient data to construct features over the selected 14-day look-back window. These criteria were used to define the main analytic configuration reported in the Results.

To assess robustness under incomplete self-report input, we simulated active-data masking within each look-back window by randomly masking daily mood, energy, and anxiety responses at proportions of 0%, 20%, 40%, 60%, and 80%. Masked entries were treated as missing and were not imputed; feature extraction relied only on observed data. In the ablation analyses, we systematically varied three parameters: passive-data coverage threshold (30%–70%, in 10% increments), active-data missingness tolerance (0–0.8, in 0.2 increments), and observation-window length (8–28 days, in 2-day increments). These analyses were designed to quantify the trade-off among data completeness, sample retention, and predictive performance under deployment-relevant conditions.

2.6. Analytical sample

Of the 455 community volunteers who installed the study application, yielding 12,146 person-days of potential monitoring data, 10 were excluded because they did not provide either passive sensor data or active survey data.

For the depression analyses, participants whose passive sensor coverage averaged less than 60% of eligible days and/or who lacked corresponding active self-reports were excluded, resulting in the removal of 168 participants. The final depression cohort comprised 277 participants contributing 2753 person-days, including 222 low-risk participants (Android: 73, iOS: 147; 2313 days) and 55 high-risk participants (Android: 20, iOS: 35; 439 days).

For the anxiety analyses, application of the same criteria excluded 170 participants, leaving 275 participants contributing 2713 person-days. The final anxiety cohort consisted of 219 low-risk participants (Android: 73, iOS: 146; 2274 days) and 56 high-risk participants (Android: 20, iOS: 35; 439 days). The resulting exclusion proportion reflects the trade-off between minimum data-quality requirements and sample retention under real-world monitoring conditions.

2.7. Model training, evaluation, and ablation analysis

We trained Random Forest, Gradient Boosting, and XGBoost classifiers to predict weekly high-risk depression (PHQ-9 $\geq$ 10) and high-risk anxiety (GAD-7 $\geq$ 10) using three feature sets: active-only, passive-only, and combined. Active-only models used self-report-derived features, passive-only models used sensor-derived features, and combined models used both modalities together.

Model performance was evaluated using participant-level five-fold cross-validation, with all repeated observations from the same participant assigned to a single held-out test fold. This evaluation framework ensured that no overlapping participant data appeared in both training and testing and that the resulting performance estimates reflected generalization to unseen individuals rather than repeated observations from the same participant.

Performance was assessed using the area under the receiver operating characteristic curve (AUC), average precision (AP), accuracy, recall, and F1-score. Unless otherwise specified, performance values are reported as mean $\pm$ standard deviation across participant-level cross-validation folds. These values represent fold-to-fold variability and should not be interpreted as observation-level inferential confidence intervals. All ablation settings described in Section 2.5 were evaluated using the same participant-level five-fold cross-validation framework. For each ablation setting, the same training, testing, and metric-computation procedure was repeated to enable direct comparison across deployment-relevant parameter configurations.

3. Results

3.1. Risk group comparison by passive features

Fig. 2 shows clear behavioral and circadian distinctions between low- and high-risk groups for both depression (Fig. 2A) and anxiety (Fig. 2B). For mobility, low-risk participants consistently demonstrated a wider excursion radius during weekdays—often exceeding 80 km—whereas high-risk individuals typically traveled less than 25 km per day and remained largely homebound. Although overall movement decreased on weekends, the gap between groups persisted.

Home-stay ratios followed a U-shaped weekly pattern in both subgroups, declining from Monday to Saturday and rising again on weekends. In the depression group, high-risk participants consistently spent more time at home during weekdays (e.g., ~59% on Monday vs ~56% among low-risk), but these differences diminished by Saturday and Sunday, when both groups converged around 55–57%. In the anxiety subgroup, weekday differences were less stable, showing some overlap across days but again converging on weekends. Nocturnal activity was higher in high-risk participants across both disorders, averaging approximately 2.10–2.40 $\times$ 10^3 m/s^2 between midnight and 6 AM, compared with ~1.95–2.20 $\times$ 10^3 m/s^2 among low-risk participants. In depression, this corresponded to a stable separation of ~150–200 m/s^2, reflecting more fragmented and restless sleep. In anxiety, the separation was less consistent but still showed higher nocturnal activity overall.

High-risk participants also exhibited greater accelerometer activity during the 2 h preceding sleep onset (~560–640 m/s^2 vs ~500–560 m/s^2 in low-risk), suggesting more difficulty winding down before sleep. This pattern was particularly evident on weekdays in depression but less stable on weekends and in anxiety. Nighttime inactivity duration—a proxy for continuous sleep—was longer among low-risk participants (~310–325 min per night) than among high-risk participants (~275–300 min per night), with both groups converging on weekends. Overall, high-risk participants, especially those with depression, displayed restricted mobility, greater home confinement, and more disrupted sleep–wake rhythms than their low-risk counterparts, whereas anxiety showed similar but more variable trends.

Fig. 3 illustrate contrasts in digitally inferred sleep–wake timing between risk groups for both depression and anxiety. In the depression subgroup (Fig. 3A), the low-risk group (blue) demonstrated earlier and more consistent bedtimes, with a median DSO of approximately 01:30 and a steep cumulative slope. In contrast, the high-risk group (orange) showed a median onset closer to 02:30 and a heavier right tail, reflecting more frequent very late bedtimes (after 04:00). The DWU distribution also diverged: low-risk participants typically awoke between 06:30 and 07:00, whereas high-risk participants clustered closer to 08:00–08:30 and displayed a more gradual cumulative increase, indicating both delayed and variable wake times. Similar patterns were observed in the anxiety subgroup (Fig. 3B). Low-risk participants exhibited earlier and more tightly clustered DSO and DWU distributions, whereas high-risk participants displayed delayed timing and greater variability, with right-skewed onset distributions and flatter cumulative wake curves. Together, these findings highlight circadian disruption in both high-risk depression and anxiety subgroups, characterized by later and less consistent sleep–wake timing relative to their low-risk counterparts.

3.2. High-risk detection performance

We evaluated high-risk depression and anxiety detection across the three classifiers and feature sets defined in the Methods. Unless otherwise specified, all performance values are reported as mean $\pm$ standard deviation across participant-level cross-validation folds.

For high-risk depression detection (Fig. 4A), models trained on the combined features consistently achieved the highest AUCs across classifiers. For Random Forest, the AUC improved from 0.72 $\pm$ 0.11 with active-only features and 0.65 $\pm$ 0.04 with passive-only features to 0.82

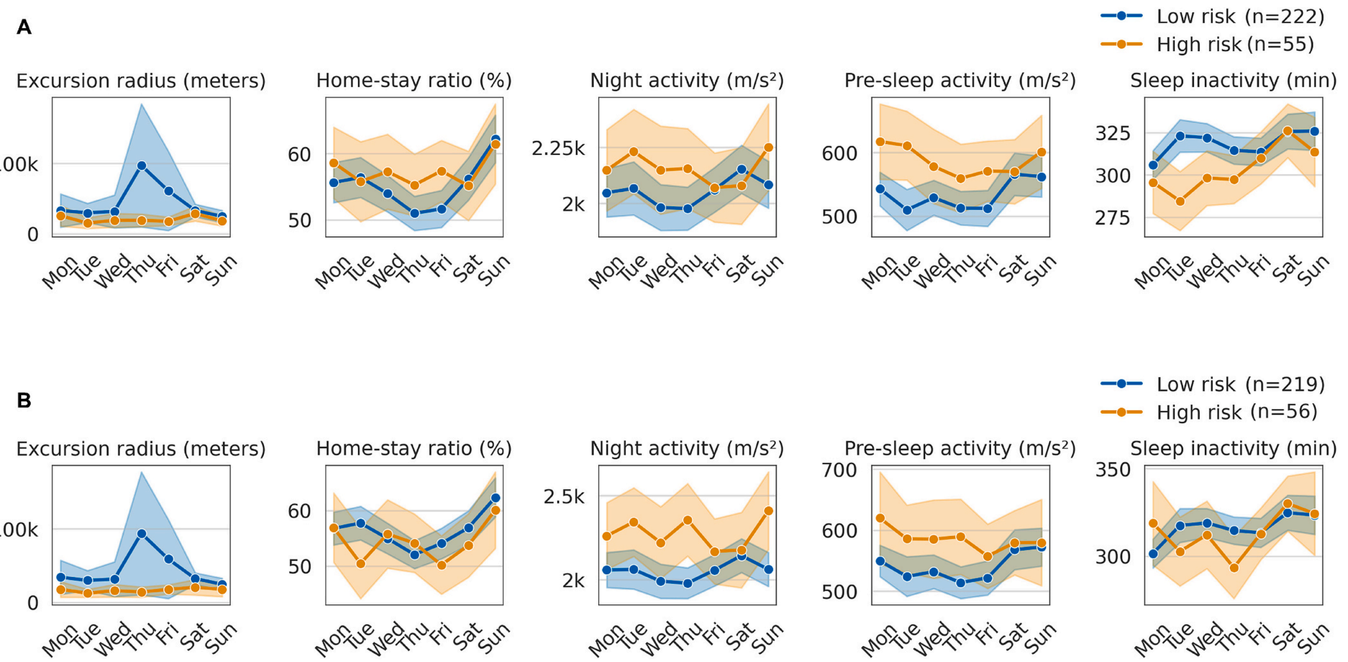

Fig. 2. Weekly trajectories of passive digital-phenotyping features in low-risk vs. high-risk subgroups for depression and anxiety.
A: Depression subgroup (low-risk $n = 222$; high-risk $n = 55$). B: Anxiety subgroup (low-risk $n = 219$; high-risk $n = 56$). Day-of-week profiles are shown for five behavioral and circadian metrics derived from smartphone accelerometer and GPS data: excursion radius, home-stay ratio, nocturnal activity (0:00–06:00), pre-sleep activity (2 h before sleep onset), and sleep-period inactivity. Solid lines denote group means, and shaded areas represent 95% confidence interval around each mean. Blue represents the low-risk group, and orange represents the high-risk group, highlighting persistent differences in mobility and sleep patterns across weekdays and weekends.

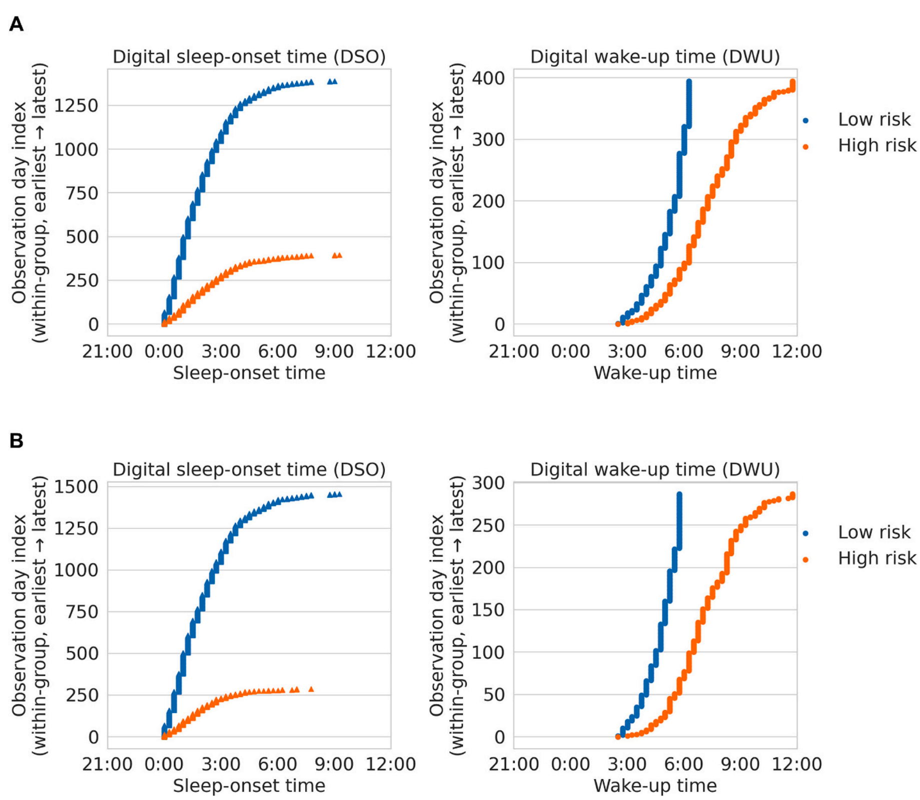

Fig. 3. Cumulative distributions of digital sleep–wake timing in depression and anxiety subgroups.
A: Digital Sleep-Onset Time (DSO) and Digital Wake-up Time (DWU) plotted as within-group cumulative indices (earliest to latest) for low-risk (blue) and high-risk (orange) depression cohorts. B: Digital Sleep-Onset Time (DSO) and Digital Wake-up Time (DWU) plotted as within-group cumulative indices (earliest to latest) for low-risk (blue) and high-risk (orange) anxiety cohorts.

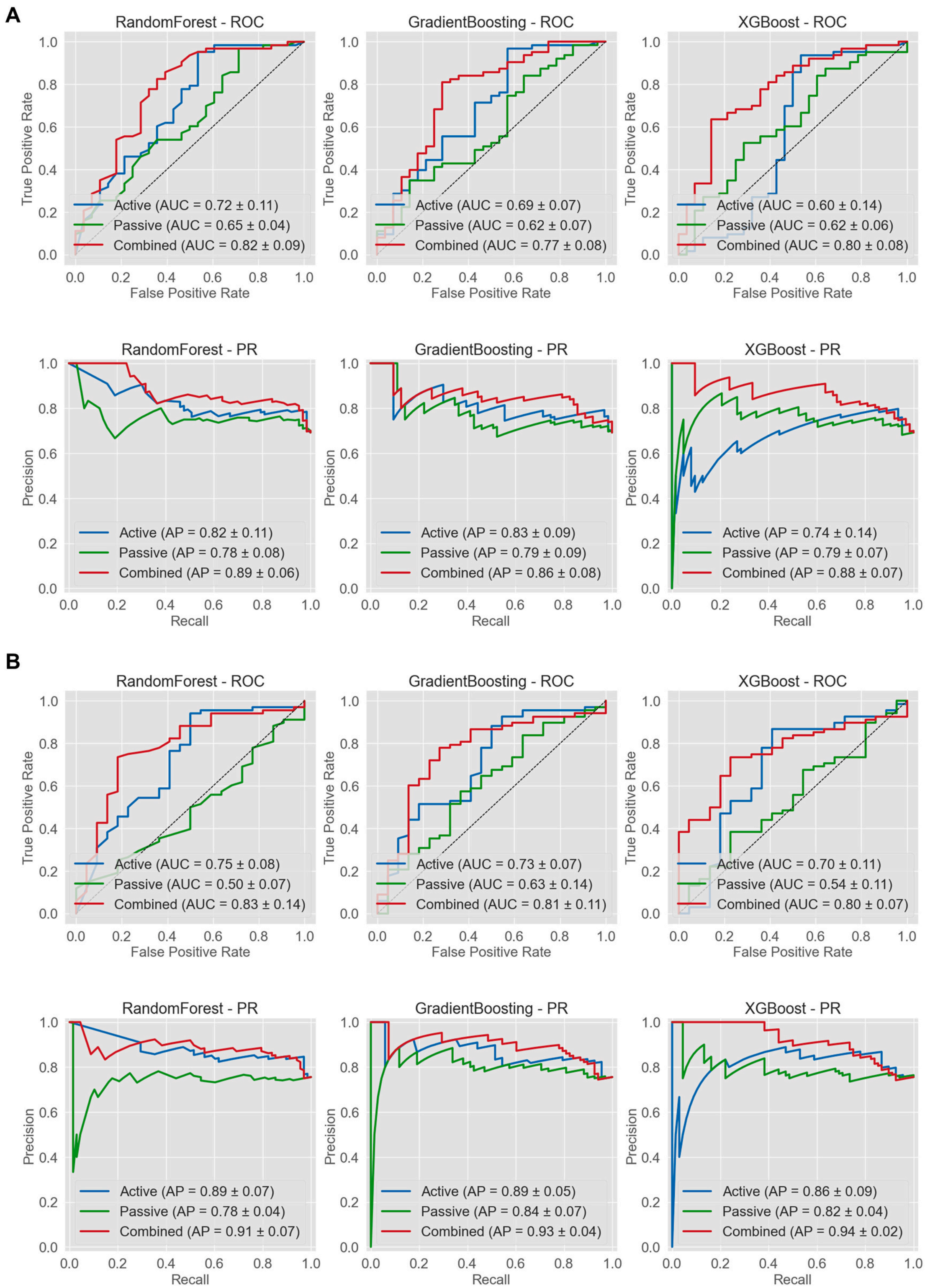

Fig. 4. Classification performance of high-risk depression and anxiety detection models across active, passive, and combined smartphone sensing modalities. A: Receiver Operating Characteristic (ROC) and Precision–Recall (PR) curves for depression classification across three classifiers—Random Forest (left), Gradient Boosting (center), and XGBoost (right). B: ROC and PR curves for anxiety classification using the three classifiers. Each classifier was trained on: (i) *active features* only (blue; self-reported micro-surveys), (ii) *passive features* only (green; GPS and accelerometer data), or (iii) *combined features* (red; fusion of active and passive inputs). The Area Under the Curve (AUC) and Average Precision (AP) are reported as mean ± standard deviation across participant-level cross-validation folds for each classifier and feature set.

$\pm$ 0.09 with combined features. Similarly, the Gradient Boosting model achieved an AUC of 0.69 $\pm$ 0.07 with active-only features and 0.62 $\pm$ 0.07 with passive-only features, increasing to 0.77 $\pm$ 0.08 with combined features. XGBoost also achieved its best AUC with combined features (0.80 $\pm$ 0.08). Precision-recall analysis showed parallel gains. In the Random Forest model, AP increased from 0.82 $\pm$ 0.11 (active-only) and 0.78 $\pm$ 0.08 (passive-only) to 0.89 $\pm$ 0.06 (combined). Gradient Boosting and XGBoost also reached their highest AP values with combined features (0.86 $\pm$ 0.08 and 0.88 $\pm$ 0.07, respectively). Detailed classification metrics are provided in Supplementary Table S4. Because average precision is influenced by class prevalence, the AP values should be interpreted relative to the positive-class base rate in this cohort rather than as prevalence-independent measures of discrimination.

Similar trends were observed for the detection of high-risk anxiety (Fig. 4B). Combined features yielded the highest AUCs across classifiers, improving Random Forest performance from 0.77 $\pm$ 0.08 (active-only) and 0.54 $\pm$ 0.10 (passive-only) to 0.86 $\pm$ 0.11 (combined). Gradient Boosting improved from 0.72 $\pm$ 0.07 (active-only) and 0.63 $\pm$ 0.14 (passive-only) to 0.79 $\pm$ 0.12 (combined), whereas XGBoost improved from 0.70 $\pm$ 0.11 (active-only) and 0.54 $\pm$ 0.11 (passive-only) to 0.80 $\pm$ 0.07 (combined). Precision-recall analyses reinforced these findings, with AP values reaching 0.95 $\pm$ 0.05 for Random Forest, 0.93 $\pm$ 0.05 for Gradient Boosting, and 0.94 $\pm$ 0.02 for XGBoost under the combined-feature setting. Comprehensive performance metrics for anxiety models are presented in Supplementary Table S5.

Across both the depression and anxiety subgroups, combined active and passive features consistently improved model performance relative to either modality alone. The gains in recall and F1-score highlight an enhanced ability to identify high-risk individuals without missing true cases. For the primary analyses reported here, we adopted a final analytic configuration requiring at least 60% passive-data coverage, at least 60% corresponding active-data availability, and a fourteen-day look-back window. These settings—identified through ablation analyses—yielded robust and stable performance across classifiers (Supplementary Tables S4–5; Fig. 4).

3.3. Ablation analyses: robustness to missingness

Using the passive-coverage, active-missingness, and observation-window settings defined in Methods 2.5 and 2.7, we evaluated model robustness across a range of deployment-relevant data-availability conditions.

For the depression models (Supplementary Fig. S2A), performance (AUC) declined at both very low and very high passive coverage thresholds, with optimal results observed at intermediate levels (50–60%). This suggests that moderate passive data completeness offers the best balance between information richness and sample retention. Model performance remained stable up to 80% of artificially masked entries, indicating that the models retained a strong discriminative ability even under substantial active data loss. Longer observation periods ($\geq$20 days) improved the AUC, although gains plateaued thereafter, reflecting diminishing returns from extended historical data.

The corresponding ablation analyses for high-risk anxiety detection (Supplementary Fig. S2B) showed similar patterns. Model performance peaked at intermediate passive coverage thresholds (approximately 60%), whereas both very low and high thresholds were associated with reduced stability. Under active-data masking, performance remained robust up to approximately 60–70% missingness, although declines were steeper beyond this point compared to depression. Performance was more variable at shorter observation windows but stabilized beyond approximately 16–20 days, after which further improvements were minimal.

Altogether, these ablation analyses demonstrate that detection models for both depression and anxiety are robust to moderate levels of passive and active data missingness, and that a 14–20-day look-back window provides an optimal trade-off between predictive performance and data efficiency.

4. Discussion

4.1. Main findings and practical contributions

This study supports the feasibility and clinical potential of smartphone-based digital phenotyping for detecting high-risk depression and anxiety in real-world community settings. Rather than introducing a novel sensing modality, the present work systematically evaluated a smartphone-only framework under deployment-relevant conditions, including varying levels of passive and active data availability and different observation-window settings. Across classifiers, fusing passive sensor streams with brief self-report microsurveys yielded robust discriminative performance, with combined-model AUCs ranging from 0.77 to 0.83 for depression and from 0.79 to 0.86 for anxiety. A central challenge—data incompleteness arising from device nonuse, permissions changes, and technical issues—was directly evaluated. Models retained useful discriminative performance under realistic passive and active data gaps.

The need for scalable screening is underscored by pandemic-era increases in depression (27.6%) and anxiety (25.6%) (Bufano et al., 2023) and the limits of episodic, clinic-based assessments that miss symptom dynamics (Huckvale et al., 2019). From a practical standpoint, the consistent performance gains observed when combining active and passive features suggest that multimodal input may improve screening performance beyond either modality alone. However, because the active features and weekly labels were both partly derived from self-reported symptom information, the observed gain should not be interpreted as arising solely from independent contributions of passive sensing.

In addition, although reporting AP alongside AUC is informative under class imbalance, AP is influenced by the prevalence of the positive class and therefore should be interpreted in the context of the cohort-specific base rate. These reported ranges reflect observed performance across classifiers rather than inferential confidence intervals.

4.2. Clinical interpretation and relation to prior work

The identified digital behavioral signatures were aligned with established clinical patterns while providing quantifiable behavioral markers in daily life. High-risk participants showed reduced mobility, increased home confinement, disrupted sleep-wake cycles, and elevated nocturnal activity—patterns consistent with psychomotor retardation, social withdrawal, and circadian dysregulation typical of depression and anxiety disorders (Jang et al., 2024; Lim et al., 2024; Onnela and Rauch, 2016; Song et al., 2024). These differences were stronger on weekdays than on weekends, suggesting that symptom-related behavioral disruption may become more detectable under structured daily demands such as work, commuting, and regular social obligations. This pattern is clinically relevant because it indicates that the predictive value of smartphone-derived behavioral features may depend not only on symptom severity itself but also on the degree to which everyday routines expose or constrain behavioral dysfunction. Taken together, these findings support the interpretability of smartphone-derived digital phenotypes as scalable behavioral markers for high-risk depression and anxiety in community settings.

4.3. Implications for scalable digital mental health systems

Missing data remains a central barrier to translating digital phenotyping into clinical use. Few studies have quantified tolerance thresholds or developed robust analytical strategies for deployment-oriented settings (Onnela and Rauch, 2016). Our ablation analyses showed that models maintained useful discriminative performance despite up to

60–80% of active data missing, addressing a key limitation highlighted in prior reviews (Mendes et al., 2022). Commercial smartphones and wearables often face gaps related to device placement, battery management, or permission loss. In this context, our results suggest that clinically relevant performance may still be attainable under imperfect real-world data conditions without relying on idealized complete-data assumptions.

The optimal parameters—60% passive coverage, 60% active data tolerance, and a 14-day look-back window—offer practical deployment guidance, balancing performance with feasibility and mitigating participant burden (Hasselhorn et al., 2022). Our results compare favorably with earlier smartphone and wearable studies: the Purple Robot study ($r = -0.63$, $n = 28$) (Saeb et al., 2015) and large-scale wearable studies (Jacobson et al., 2020) achieved similar effect sizes but required specialized devices, whereas our smartphone-only approach demonstrated comparable accuracy using ubiquitous technology. Systematic reviews have identified missing data and limited external validation as key sources of bias in digital phenotyping (Leaning et al., 2024). In that context, our participant-independent validation framework and deployment-oriented ablation analyses help clarify how a smartphone-only system may perform under realistic use conditions within this cohort. For clarity, the performance ranges and parameter values reported here describe observed analytic settings and model behavior, not inferential uncertainty bounds.

Continuous, objective risk assessment using smartphones could enable population-level screening and early identification of individuals requiring clinical attention. Integration with electronic health records may inform future just-in-time interventions, personalized feedback, and remote monitoring to support continuity of care (Huckvale et al., 2019). Smartphone-based assessments could also reduce clinician burden and expand reach to underserved populations, though successful implementation will require workflow integration, provider training, and clear protocols for risk-threshold alerts and false positives.

4.4. Strengths, limitations, and future directions

Several strengths support translation: participant-independent validation, systematic ablation analyses yielding practical thresholds, and weekday-weekend aggregation capturing real-world behavioral rhythms. By computing only aggregated behavioral summaries, the system minimizes privacy risks while maintaining analytic utility (Mohr et al., 2020), aligning with data-minimization principles.

Limitations include the homogeneous Korean adult sample (ages 19–59), limiting generalizability across cultures and usage patterns. Although participant-independent validation reduced within-person data leakage and supported out-of-sample evaluation at the participant level, the absence of external or temporal validation means that transportability across populations, smartphone platforms, and clinical settings remains uncertain. The short-term observational design precludes causal inference, and reliance on self-report measures (PHQ-9 and GAD-7) may introduce circularity. An additional interpretive limitation concerns the relationship between the active feature set and the weekly questionnaire-based outcomes. The active features included daily self-reported mood, energy, and anxiety ratings, whereas the depression and anxiety labels were derived from weekly self-report measures (PHQ-9 and GAD-7). Accordingly, the relatively strong performance of the active-only models and the gains observed with combined models should be interpreted with some caution, as both the active features and the weekly labels included self-reported symptom information collected on different schedules. This consideration may be particularly relevant for anxiety, for which passive-only performance was weaker and less consistent across classifiers. Future work should further examine modality-specific contributions and the practical role of passive sensing alongside repeated self-report. In addition, because the PHQ-9 references symptoms over the preceding 2 weeks whereas it was administered weekly in our study, adjacent PHQ-9 assessments were not fully temporally independent. This feature of the design may have complicated interpretation of week-to-week changes and should be considered when interpreting the PHQ-9-based depression label. An additional limitation concerns the interpretation of the accelerometer-derived sleep-wake features. DSO and DWU were operationalized as smartphone-based markers of when motion stopped and resumed, and should therefore be interpreted as digitally inferred behavioral proxies rather than validated measures of physiological sleep onset or awakening. Unlike wrist-worn actigraphy, smartphone accelerometer data depend on phone placement and use behavior, so a participant putting the phone down does not necessarily indicate sleep onset, and resumed phone-related motion does not necessarily indicate awakening. Accordingly, these features are best understood as scalable, behaviorally informative proxies of sleep-wake timing rather than established sleep measures. An additional practical limitation is that a substantial proportion of initially recruited participants did not meet the minimum usable-data criteria and were therefore excluded from the final analyses. This should be interpreted not only as sample attrition but also as a real-world implementation barrier for smartphone-based monitoring. In our study setting, usable data depended on a multi-step data-collection pipeline, including sustained participant engagement, completion of active daily reports, continuous on-device sensing, stable background execution, retention of sensor permissions, successful local recording, and transmission and synchronization with the study server. Accordingly, data loss could arise from both behavioral sources, such as participant disengagement and intermittent nonuse, and technical sources, such as operating-system restrictions, battery optimization, connectivity problems, and synchronization failures. The imbalance between high- and low-risk groups and device heterogeneity further constrains robustness. Future studies should evaluate strategies to improve end-to-end data capture, retention, and long-term engagement under real-world deployment conditions.

This work supports the transition toward P4 medicine (predictive, preventive, personalized, participatory) by showing that smartphone-based digital phenotyping can complement conventional assessment (Zhang et al., 2025). The convergence of ubiquitous technology and mental-health analytics offers important opportunities for scalable screening and monitoring. By demonstrating clinically meaningful performance within this cohort and identifying practical deployment parameters, this study provides an empirical basis for future development of scalable, low-cost digital mental health screening embedded in everyday technology.

5. Conclusion

This study demonstrated that smartphone-based digital phenotyping can achieve clinically meaningful performance for screening high-risk depression and anxiety in naturalistic settings using a pragmatic approach that prioritizes implementation feasibility over theoretical optimization. The present work systematically evaluated a smartphone-only framework combining passive behavioral monitoring with minimal active self-report under realistic conditions of missing data and short observation windows. The results suggest that fused active and passive features can support scalable screening for high-risk depression and anxiety while also providing practical guidance on deployment-relevant operating thresholds. These findings offer empirical support for smartphone-based digital phenotyping as a feasible component of scalable, low-cost mental health screening and future intervention workflows embedded in everyday technology.

CRediT authorship contribution statement

Ah Young Kim: Writing – original draft, Writing – review & editing, Formal analysis, Investigation, Data curation, Visualization. **Seonmin Kim:** Investigation, Data curation. **Jisu Lee:** Investigation, Data curation. **Youngwoong Han:** Formal analysis, Investigation. **Heon-Jeong**

Lee: Supervision, Resources. **Chul-Hyun Cho:** Conceptualization, Methodology, Supervision, Project administration, Writing – review & editing.

Funding

This work was supported by grants from the Institute of Information & Communications Technology Planning & Evaluation (IITP) (grant number: RS-2023-00224823) and the National Research Foundation (NRF) of Korea (grant numbers: NRF-2021R1A5A8032895, NRF-2022M3C1B6080866, and RS-2024-00422599) funded by the Ministry of Science and Information and Communications Technology, Republic of Korea. The funding source played no role in the study design, data collection, data analysis, interpretation, or the writing of the manuscript.

Declaration of competing interest

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

Acknowledgements

We thank all study participants for their time and commitment, and we acknowledge the support of clinical and community partners who assisted with recruitment and data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.invent.2026.100934.

References

Bähr, S., Haas, G.-C., Keusch, F., Kreuter, F., Trappmann, M., 2022. Missing data and other measurement quality issues in mobile geolocation sensor data. Soc. Sci. Comput. Rev. 40 (1), 212–235. https://doi.org/10.1177/0894439320944118.

Bladon, S., Eisner, E., Bucci, S., Oluwatayo, A., Martin, G.P., Sperrin, M., Ainsworth, J., Faulkner, S., 2025. A systematic review of passive data for remote monitoring in psychosis and schizophrenia. npj Digit. Med. 8 (1), 62. https://doi.org/10.1038/s41746-025-01451-2.

Boer, C.D., Ghomrawi, H., Zeineddin, S., Linton, S., Kwon, S., Abdullah, F., 2023. A call to expand the scope of digital phenotyping. J. Med. Internet Res. 25 (1), e39546. https://doi.org/10.2196/39546.

Bufano, P., Laurino, M., Said, S., Tognetti, A., Menicucci, D., 2023. Digital phenotyping for monitoring mental disorders: systematic review. J. Med. Internet Res. 25 (1), e46778. https://doi.org/10.2196/46778.

Chandrasekaran, R., S. .M., T, Moustakas, E., 2025. Usage trends and data sharing practices of healthcare wearable devices among US adults: cross-sectional study. J. Med. Internet Res. 27 (1), e63879. https://doi.org/10.2196/63879.

Choi, A., Ooi, A., Lottridge, D., 2024. Digital phenotyping for stress, anxiety, and mild depression: systematic literature review. JMIR Mhealth Uhealth 12 (1), e40689. https://doi.org/10.2196/40689.

Collaborators, G.B.D.M.D., 2022. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 9 (2), 137–150. https://doi.org/10.1016/S2215-0366(21)00395-3.

Du, J., Hu, M., Zhang, W., 2020. Missing data problem in the monitoring system: a review. IEEE Sensors J. 20 (23), 13984–13998. https://doi.org/10.1109/JSEN.2020.3009265.

Gelles-Watnick, R., 2024. Americans' Use of Mobile Technology and Home Broadband. Pew Research Center. https://www.pewresearch.org/internet/2024/01/31/americans-use-of-mobile-technology-and-home-broadband/.

Gromatsky, M., Sullivan, S.R., Spears, A.P., Mitchell, E., Walsh, S., Kimbrel, N.A., Goodman, M., 2020. Ecological momentary assessment (EMA) of mental health outcomes in veterans and servicemembers: a scoping review. Psychiatry Res. 292, 113359. https://doi.org/10.1016/j.psychres.2020.113359.

Hasselhorn, K., Ottenstein, C., Lischetzke, T., 2022. The effects of assessment intensity on participant burden, compliance, within-person variance, and within-person relationships in ambulatory assessment. Behav. Res. Methods 54 (4), 1541–1558. https://doi.org/10.3758/s13428-021-01683-6.

Huang, S., Wang, Y., Li, G., Hall, B.J., Nyman, T.J., 2024. Digital mental health interventions for alleviating depression and anxiety during psychotherapy waiting lists: systematic review. JMIR Mental Health 11 (1), e56650. https://doi.org/10.2196/56650.

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

Institute of Medicine Forum on Drug Discovery, 2010. Clinical trials in depression. In: Transforming Clinical Research in the United States: Challenges and Opportunities: Workshop Summary. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK50890/.

Jacobson, N.C., Summers, B., Wilhelm, S., 2020. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. J. Med. Internet Res. 22 (5), e16875.

Jang, S., Sun, T.h., Shin, S., Lee, H.-J., Shin, Y.-B., Yeom, J.W., Park, Y.R., Cho, C.-H., 2024. A digital phenotyping dataset for impending panic symptoms: a prospective longitudinal study. Sci. Data 11 (1), 1264. https://doi.org/10.1038/s41597-024-04147-6.

Kalisperakis, E., Karantinos, T., Lazaridi, M., Garyfalli, V., Filntisis, P.P., Zlatintsi, A., Efthymiou, N., Mantas, A., Mantonakis, L., Mougiakos, T., Maglogiannis, I., Tsanakas, P., Maragos, P., Smyrnis, N., 2023. Smartwatch digital phenotypes predict positive and negative symptom variation in a longitudinal monitoring study of patients with psychotic disorders. Front. Psychol. 14, 1024965. https://doi.org/10.3389/fpsyt.2023.1024965.

Kim, H., Lee, S., Lee, S., Hong, S., Kang, H., Kim, N., 2019. Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: observational study on older adults living alone. JMIR Mhealth Uhealth 7 (10), e14149. https://doi.org/10.2196/14149.

Leaning, I.E., Ikani, N., Savage, H.S., Leow, A., Beckmann, C., Ruhé, H.G., Marquand, A. F., 2024. From smartphone data to clinically relevant predictions: a systematic review of digital phenotyping methods in depression. Neurosci. Biobehav. Rev. 158, 105541. https://doi.org/10.1016/j.neubiorev.2024.105541.

Lee, M.P., Kim, D.W., Fang, Y., Kim, R., Bohnert, A.S.B., Sen, S., Forger, D.B., 2024. The real-world association between digital markers of circadian disruption and mental health risks. npj Digit. Med. 7 (1), 355. https://doi.org/10.1038/s41746-024-01348-6.

Lim, D., Jeong, J., Song, Y.M., Cho, C.-H., Yeom, J.W., Lee, T., Lee, J.-B., Lee, H.-J., Kim, J.K., 2024. Accurately predicting mood episodes in mood disorder patients using wearable sleep and circadian rhythm features. npj Digit. Med. 7 (1), 324. https://doi.org/10.1038/s41746-024-01333-z.

Linardon, J., Chen, K., Gajjar, S., Eadara, A., Wang, S., Flathers, M., Burns, J., Torous, J., 2025. Smartphone digital phenotyping in mental health disorders: a review of raw sensors utilized, machine learning processing pipelines, and derived behavioral features. Psychiatry Res. 348, 116483. https://doi.org/10.1016/j.psychres.2025.116483.

Magallón-Neri, E., Kirchner-Nebot, T., Forns-Santacana, M., Calderón, C., Planellas, I., 2016. Ecological momentary assessment with smartphones for measuring mental health problems in adolescents. World J. Psychiatry 6 (3), 303–310. https://doi.org/10.5498/wjp.v6.i3.303.

Mendes, J.P.M., Moura, I.R., Ven, P.V.d., Viana, D., Silva, F.J.S., Coutinho, L.R., Teixeira, S., Rodrigues, J.J.P.C., Teles, A.S., 2022. Sensing apps and public data sets for digital phenotyping of mental health: systematic review. J. Med. Internet Res. 24 (2), e28735. https://doi.org/10.2196/28735.

Mohr, D.C., Shilton, K., Hotopf, M., 2020. Digital phenotyping, behavioral sensing, or personal sensing: names and transparency in the digital age. npj Digit. Med. 3 (1), 45. https://doi.org/10.1038/s41746-020-0251-5.

Moura, I., Teles, A., Viana, D., Marques, J., Coutinho, L., Silva, F., 2023. Digital phenotyping of mental health using multimodal sensing of multiple situations of interest: A systematic literature review. J. Biomed. Inform. 138, 104278. https://doi.org/10.1016/j.jbi.2022.104278.

Niv, N., Cohen, A.N., Mintz, J., Ventura, J., Young, A.S., 2007. The validity of using patient self-report to assess psychotic symptoms in schizophrenia. Schizophr. Res. 90 (1), 245–250. https://doi.org/10.1016/j.schres.2006.11.011.

Onnela, J.-P., Rauch, S.L., 2016. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 41 (7), 1691–1696. https://doi.org/10.1038/npp.2016.7.

Saddichha, S., Al-Desouki, M., Lamia, A., Linden, I.A., Krausz, M., 2014. Online interventions for depression and anxiety – a systematic review. Health Psychol. Behav. Med. 2 (1), 841–881. https://doi.org/10.1080/21642850.2014.945934.

Saeb, S., Zhang, M., Karr, C.J., Schueller, S.M., Corden, M.E., Kording, K.P., Mohr, D.C., 2015. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J. Med. Internet Res. 17 (7), e4273. https://doi.org/10.2196/jmir.4273.

Shandhi, M.M.H., Singh, K., Janson, N., Ashar, P., Singh, G., Lu, B., Hillygus, D.S., Maddocks, J.M., Dunn, J.P., 2024. Assessment of ownership of smart devices and the acceptability of digital health data sharing. npj Digit. Med. 7 (1), 44. https://doi.org/10.1038/s41746-024-01030-x.

Shin, Y.-B., Kim, A.Y., Kim, S., Shin, M.-S., Choi, J., Lee, K.L., Lee, J., Byun, S., Kim, S., Lee, H.-J., Cho, C.-H., 2025. Development of prediction models for screening depression and anxiety using smartphone and wearable-based digital phenotyping: protocol for the Smartphone and Wearable Assessment for Real-Time Screening of Depression and Anxiety (SWARTS-DA) observational study in Korea. BMJ Open 15 (6), e096773. https://doi.org/10.1136/bmjopen-2024-096773.

Song, Y.M., Jeong, J., Reyes, A.A.d.l., Lim, D., Cho, C.-H., Yeom, J.W., Lee, T., Lee, J.-B., Lee, H.-J., Kim, J.K., 2024. Causal dynamics of sleep, circadian rhythm, and mood symptoms in patients with major depression and bipolar disorder: insights from longitudinal wearable device data. eBioMedicine 103. https://doi.org/10.1016/j.ebiom.2024.105094.

Szücs, A., van der Lubbe, S.C.C., Arias de la Torre, J., Valderas, J.M., Hay, S.I., Bisignano, C., Morgan, B.W., Acharya, S., Adnani, Q.E.S., Apostol, G.L.C., Aslam, M. S., Asri, Y., Aung, Z.Z., Aurizki, G.E., Baig, A.A., Bermudez, A.N.C., Cenderadewi, M., Danpanichkul, P., Efendi, F., Estrada, C.A.M., Fauk, N.K., Fridayani, N.K.Y., Garcia, F.B., Hasan, F., Ibrahim, U.I., Ilyas, M.F., Ismail, N.E., Jamaluddin, J., Jamora, R.D.G., Jonas, J.B., Kanmodi, K.K., Khan, K.U., Khotimah, F.H., Kim, Y.J., Kurniasari, M.D., Kustanti, C.Y., Kusuma, D., Laksono, T., Lau, J., Loreche, A.M., Ma, Z.F., Maravilla, J.C., Marzo, R.R., Maude, R.J., Porntaveetus, T., Rahmawaty, S., Rodriguez, A., Roy, B., Salami, A.A., Samodra, Y.L., Sreeramareddy, C.T., Subramaniyan, V., Sukaew, T., Sulistiyorini, D., Sumpaico-Tanchanco, L.B.C., Wicaksana, A.L., Wong, Y.J., Younis, M.Z., Zuniga, Y.M.H., Murray, C.J.L., Santomauro, D.F., Ng, M., 2025. The epidemiology and burden of ten mental disorders in countries of the Association of Southeast Asian Nations (ASEAN), 1990–2021: findings from the Global Burden of Disease Study 2021. Lancet Public Health 10 (6), e480–e491. https://doi.org/10.1016/S2468-2667(25)00098-2.

Teo, J.X., Davila, S., Yang, C., Hii, A.A., Pua, C.J., Yap, J., Tan, S.Y., Sahlén, A., Chin, C. W.-L., Teh, B.T., Rozen, S.G., Cook, S.A., Yeo, K.K., Tan, P., Lim, W.K., 2019. Digital phenotyping by consumer wearables identifies sleep-associated markers of cardiovascular disease risk and biological aging. Commun. Biol. 2, 361. https://doi.org/10.1038/s42003-019-0605-1.

Unützer, J., Park, M., 2012. Strategies to improve the management of depression in primary care. Prim. Care 39 (2), 415–431. https://doi.org/10.1016/j.pop.2012.03.010.

Veney, B., 2025. Smartphone Statistics in 2025: Key Trends and Insights You Need to Know. Internet Speed Test | TestMySpeed.com. https://www.testmyspeed.com/insights/smartphone-statistics-you-need-to-know.

Villarreal-Zegarra, D., Barrera-Begazo, J., Otazú-Alfaro, S., Mayo-Puchoc, N., Bazo-Alvarez, J.C., Huarcaya-Victoria, J., 2023. Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9, PHQ-8, PHQ-2) and General Anxiety Disorder Scale (GAD-7, GAD-2) for depression and anxiety diagnosis: a cross-sectional study in a Peruvian hospital population. BMJ Open. https://doi.org/10.1136/bmjopen-2023-076193.

Zhang, Y., Folarin, A.A., Sun, S., Cummins, N., Bendayan, R., Ranjan, Y., Rashid, Z., Conde, P., Stewart, C., Laiou, P., Matcham, F., White, K.M., Lamers, F., Siddi, S., Simblett, S., Myin-Germeys, I., Rintala, A., Wykes, T., Haro, J.M., Penninx, B.W.J.H., Narayan, V.A., Hotopf, M., Dobson, R.J.B, RADAR-CNS Consortium, 2021. Relationship between major depression symptom severity and sleep collected using a wristband wearable device: multicenter longitudinal observational study. JMIR Mhealth Uhealth 9 (4), e24604. https://doi.org/10.2196/24604.

Zhang, Y., Wang, J., Zong, H., Singla, R.K., Ullah, A., Liu, X., Wu, R., Ren, S., Shen, B., 2025. The comprehensive clinical benefits of digital phenotyping: from broad adoption to full impact. npj Digit. Med. 8 (1), 196. https://doi.org/10.1038/s41746-025-01602-5.