Original Paper

Digital Phenotyping of Behavioral Instability as an Early Detection of Clinically Relevant Deterioration in Mental Health: Prospective Observational Multicenter Study

Rafael Salom[1,2], PhD; Xin Ying Leong[3], MSc; Adeline Poncet[3], MSc; José Carlos Aradillas Jaramillo[4], PhD; Álvaro Pico Rada[1,5], MD; Juan Jesús Muñoz García[5], PhD; Helena García-Mieres[4], PhD; Antonio Artés-Rodríguez[3,4,6], PhD

[1]Clínica Nuestra Señora de la Paz, Orden Hospitalaria San Juan de Dios, Madrid, Spain
[2]Department of Psychology, Faculty of Health Sciences, European University of Valencia, Valencia, Spain
[3]Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Leganés, Spain
[4]Evidence-Based Behavior (eB2), Madrid, Spain
[5]Centro de San Juan de Dios, Orden Hospitalaria San Juan de Dios, Ciempozuelos, Madrid, Spain
[6]Centro de Investigación Biomédica en Red, Madrid, Spain

Corresponding Author:

Rafael Salom, PhD
Clínica Nuestra Señora de la Paz, Orden Hospitalaria San Juan de Dios
Calle López de Hoyos, 259
Madrid 28043
Spain
Phone: +34 914 156 000
Email: rafael.salom@sjd.es

Abstract

Background: Early detection of clinically relevant changes remains a major challenge in mental health, as deterioration unfolds through subtle alterations in daily functioning. Advances in digital phenotyping and smartphone-based monitoring offer new opportunities to capture real-world settings. However, existing approaches rely on active self-reports, and few studies have examined whether within-person instability in passively sensed behavior can anticipate clinically recorded relapse events across diagnoses.

Objective: This study aimed to adapt and evaluate an algorithm detecting behavioral changes associated with deterioration using passively collected smartphone data from a prospective, multicenter project investigating smartphone-based behavioral monitoring in adults with mental health conditions, to evaluate its performance in-sample and examine whether instability in behavioral patterns can serve as an early marker of relapse.

Methods: We conducted a prospective, observational, multicenter study including 92 adult participants with mental health conditions (bipolar disorder, personality disorder, substance use disorder, and dual pathology) recruited from 2 specialized mental health centers in Madrid, Spain. Passive behavioral data, including physical activity, mobility, phone use, sleep, and daily routine indicators, were collected using a mobile app over periods ranging from 1 month to 1 year (mean follow-up of 123, SD 94.85 days). Behavioral profiles were generated using heterogeneous mixture models, and changes in behavioral stability were detected using a Bayesian online change-point detection approach. Model hyperparameters and source combinations were selected by grid search and heuristic exploration, and performance was evaluated in-sample by comparing detected change points with relapse events recorded in the electronic health record, using receiver operating characteristic analysis.

Results: Forty relapse events with available passive data were included in the analysis. The best-performing model achieved a mean in-sample area under the curve of 82.99% (SD 2.49%). At a false positive rate (FPR) of 5%, the model reached an average true positive rate (TPR) of 40.12% (95% CI 36.61%-43.63%), increasing to 61.10% (95% CI 56.42%-65.79%) at 10% FPR. Model performance was influenced by the temporal aggregation parameter, with optimal results obtained using minimal accumulation of past observations. Multimodal combinations integrating sleep, activity, and routine structure achieved the highest performance. These operating points correspond to decision thresholds matched at the specified FPR values, and the reported intervals reflect variability across 10 stochastic reruns on the same dataset rather than out-of-sample generalization.

Performance showed run-to-run variability (maximum TPR of 53.58% vs mean TPR of 40.12%, SD 5.66% at 5% FPR), indicating limited stability.

Conclusions: Behavioral instability shows preliminary potential for the early detection of behavioral changes associated with clinically relevant deterioration. Because hyperparameters, source combinations, and thresholds were tuned on the same events used for evaluation, the performance should be interpreted as an in-sample estimate, and external validation is required before clinical application.

JMIR Form Res 2026;10:e98935; doi: 10.2196/98935

Keywords: digital phenotyping; smartphone; mobile health; behavioral monitoring; algorithms; risk assessment; relapse; mental health; early detection; medical informatics

Introduction

Early detection of clinically relevant changes remains a central challenge in mental health [1,2]. In most disorders, progression toward clinical decompensation is not abrupt but gradual, emerging through subtle alterations in daily functioning before consolidating into symptom exacerbations, relapses, or loss of functional stability [3-9]. In practice, these changes often take the form of modifications in daily routines, loss of regularity, fluctuations in activity levels, changes in social interaction, or disruptions in the sleep-wake cycle [10-13]. Detecting these signals at early stages therefore enables more timely intervention, optimization of preventive strategies, and a reduction in the personal and health care burden associated with more advanced stages of deterioration [14-20]. Importantly, clinically significant deterioration is often preceded by a sequence of behavioral and psychological changes, meaning that the observable clinical outcome may represent only the final stage of a longer process unfolding over time [13,15,17].

However, in routine clinical practice, the detection of change often relies on point-in-time assessments which, due to their intermittent nature and vulnerability to recall bias, tend to identify deterioration only once it has become clinically manifest [21,22]. Addressing this gap has become a central objective in mental health research, with digital monitoring and digital phenotyping emerging as key tools to bring everyday life closer to clinical assessment [23-25].

In this context, smartphones provide a particularly suitable platform for continuous behavioral monitoring [25-27]. By accompanying individuals in their daily lives, smartphones enable repeated and ecologically valid measurements that are difficult to obtain through conventional clinical follow-up [24,28-31]. This type of longitudinal monitoring in real-world settings is particularly relevant in mental health, where clinically significant changes often first appear as alterations in daily rhythms and consistency of functioning before evolving into recognizable syndromic episodes [19,23,28,29,32-35]. Such continuous monitoring allows behavioral changes to be captured as they unfold in everyday life, offering temporally fine-grained information that is often inaccessible through retrospective assessments or intermittent clinical evaluations [24,28,36].

From this perspective, it is crucial to shift the focus from absolute levels of behavior to their temporal dynamics, as variability or irregularity in patterns (eg, physical activity, sleep, or device use), even within seemingly normal ranges, may signal processes of clinical destabilization [21,37,38]. A previously stable pattern that becomes inconsistent over time may indicate that an individual is moving away from their usual style of functioning and entering a state of increased clinical vulnerability [39,40]. This approach is particularly pertinent in mental health, where transitions are dynamic and nonlinear and where deviations from an individual's expected pattern may provide more sensitive signals of clinically meaningful change than fixed thresholds or population averages [41-43].

Moreover, an approach centered on behavioral instability provides a transdiagnostic framework, as psychological deterioration often involves shared processes such as disruptions in routines and loss of regularity in daily life [44-46]. These dimensions reflect a broader behavioral organization sensitive to changes in mental state, with relative independence from diagnosis, meaning that digital monitoring focused on instability may detect clinically relevant transitions even in heterogeneous samples [17,46-48].

Methodologically, this approach relies on the integration of active data (subjective and momentary experience) and passive data (continuous monitoring of behavior in naturalistic contexts) obtained through smartphones, enabling a multimodal perspective on individual change [32,38,49,50]. In applied terms, this approach supports the development of individualized monitoring tools capable of detecting deviations from a person's prior behavioral pattern before clinically significant deterioration becomes apparent [23,32,38,51-54]. Through the continuous collection of digital behavioral traces (such as activity levels, mobility patterns, sleep, emotional reports, or device use), these systems provide detailed indicators of everyday functioning that may reveal early signals of emerging psychological change [32,49].

To date, digital tools developed for relapse prevention have largely focused on self-monitoring systems in which individuals periodically record the frequency of their symptoms and receive feedback based on these reports [52,53]. These approaches can promote greater awareness of subjective states and provide useful information for individual monitoring [25,26]. At the same time, advances in mobile technologies have expanded these approaches by enabling the automated collection of behavioral indicators in everyday life contexts [23,24,28,50,55]. Integrating these digital traces

with self-reported information allows monitoring approaches to capture changes in daily functioning without relying exclusively on active user input, opening new possibilities for identifying potentially relevant behavioral transitions [23,26,49,56].

In the present study, we adopt this framework to adapt an algorithm aimed at detecting changes potentially associated with relapses or clinically relevant behavioral transitions. The algorithm integrates passive data collected via smartphones to monitor fluctuations in everyday behavior over time. We start from the premise that instability in daily behavioral patterns may constitute an early marker of emerging change. Accordingly, we examine whether alterations in behavioral stability can help identify clinically relevant transitions before overt deterioration becomes evident. More broadly, this work seeks to advance the use of mobile technologies toward the identification of actionable behavioral signals that may support prevention, monitoring, and personalized intervention in mental health. In this framework, behavioral instability is conceptualized as a proximal digital marker of clinically relevant deterioration, which may precede or coincide with clinically recorded relapse events used as the reference outcome. Throughout this manuscript, these terms are used with distinct meanings: behavioral instability refers to the loss of regularity in an individual's passively sensed daily behavior; a change point refers to the algorithmic detection of a shift in that behavioral pattern; clinically relevant deterioration refers to the underlying clinical process of decompensation; and relapse refers to the clinically recorded event documented in the electronic health record and used as the reference outcome.

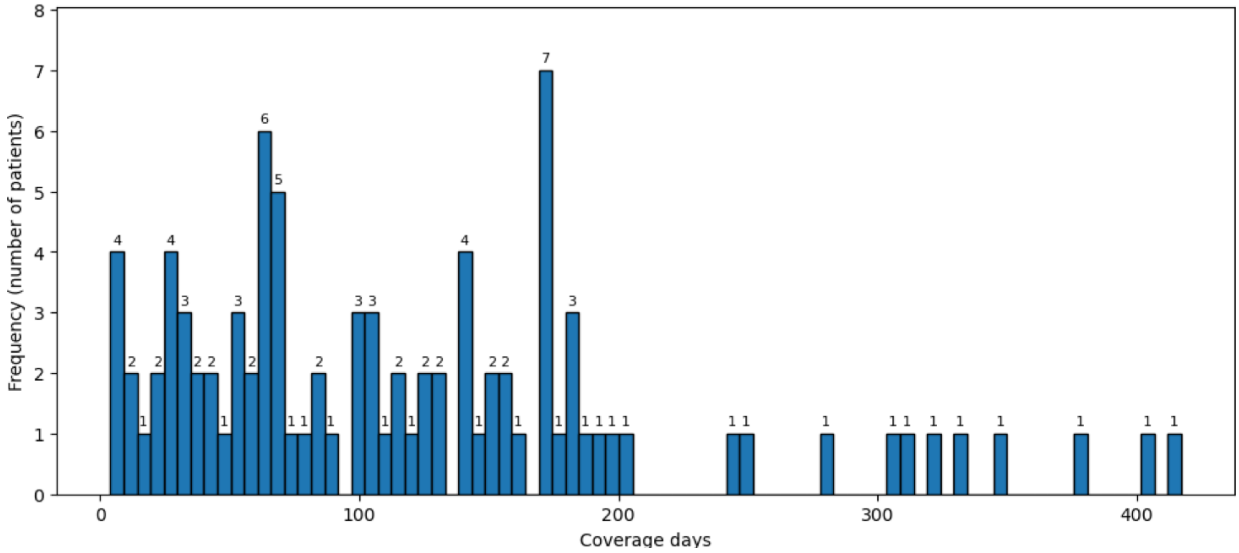
Methods

Study Design

This was a prospective, observational, multicenter clinical study aimed at adapting and evaluating an algorithm for the detection of behavioral changes associated with clinically relevant deterioration based on digital behavioral patterns collected through smartphones. The study sample consisted of 92 participants recruited from 2 centers belonging to the Orden Hospitalaria de San Juan de Dios: Clínica Nuestra Señora de la Paz and Centro San Juan de Dios de Ciempozuelos (Madrid, Spain).

Participants were continuously monitored for a period ranging from 1 month to 1 year between July 2024 and August 2025, with an average monitoring duration of 123 (SD 94.85) days, a median of 104 days, and a range of 4 to 417 days. This variability was partly explained by a subgroup of participants (n=11) with extended follow-up periods exceeding 240 days. The distribution is shown in Figure 1.

Figure 1. Histogram of coverage days.



Follow-up assessments were conducted using the eB2 MindCare mobile app installed on each participant's personal smartphone, enabling the collection of passive behavioral data in real-world settings, complemented by sociodemographic information and clinically recorded relapse events.

Inclusion criteria were as follows: (1) aged 18 years or older; (2) having a clinically confirmed mental health diagnosis (eg, mood disorders, personality disorders (PDs), substance use disorders (SUDs), or other psychiatric conditions characterized by fluctuations in psychological or behavioral functioning) and receiving follow-up care at one of the participating clinical centers; (3) ability to understand and sign informed consent; (4) sufficient fluency in Spanish to comprehend study information; and (5) ownership of a smartphone with internet access and a compatible iOS or Android operating system.

Exclusion criteria included institutionalization or current incarceration, anticipation of imminent imprisonment, lack of a compatible smartphone with internet access, or refusal to install the mobile app required for digital monitoring. Participants were also excluded if they were unable to provide informed consent or presented significant cognitive or decisional impairment that could interfere with their ability to understand the study procedures or comply with the

monitoring protocol, according to the clinical judgment of the professionals responsible for their treatment.

Although a larger number of patients were initially considered (n=99), several were excluded during preprocessing due to a lack of available app data or ambiguity in the linkage between clinical and digital users. The final sample included 92 participants with valid digital records. Additionally, a proportion of participants discontinued their participation before reaching longer monitoring periods, reflecting the challenges inherent to longitudinal studies in real-world psychiatric populations.

Process

Recruitment took place in the mental health services at the 2 participating centers of the Orden Hospitalaria de San Juan de Dios: Clinica Nuestra Señora de la Paz and Centro de San Juan de Dios de Ciempozuelos. Eligibility criteria were initially assessed by the clinical professionals responsible for each patient's care, who systematically reviewed their active caseloads against the inclusion and exclusion criteria and referred potentially eligible patients to the research team. Following this initial screening, the research team conducted the recruitment process, provided detailed information about the study, managed the installation of the mobile app, and carried out all research-related procedures and follow-up activities independently of the clinical care staff. This separation of roles ensured neutrality throughout the recruitment process and helped avoid potential conflicts of interest between clinicians and patients.

After receiving detailed information about the study and signing written informed consent, the eB2 MindCare mobile app, developed by the Department of Signal Theory and Communications at Universidad Carlos III de Madrid, was downloaded and installed on participants' smartphones. Following installation, researchers provided a brief explanation of the app's functionality and the data collection procedures. Participation was entirely voluntary and did not involve any financial or material compensation.

eB2 MindCare is a mobile health app for iOS and Android designed for the continuous, passive monitoring of behavior in real-world settings. Running in the background of the participant's personal smartphone, the app collects raw signals from device sensors and usage logs (eg, accelerometry-derived activity, steps, distance, location-derived presence at home or at work, screen and app use, and sleep-related indicators) and transmits them through a secure app programming interface to a clinical monitoring backend (CMB), where they are aggregated into the daily behavioral indicators described in Table 1. These behavioral indicators are referred to as passive behavioral data sources. The app therefore mediates between raw sensor data and the algorithm by transforming heterogeneous sensor streams into structured, per-day behavioral sources. eB2 MindCare was selected because it enables simultaneous multisource passive data collection on both major mobile platforms, has been used in previous clinical studies of smartphone-based monitoring in psychiatric populations [32], and integrates data collection with the secure clinical infrastructure required for this study.

Table 1. Summary of passive behavioral data sources, including context (digital phenotyping category), coverage percentage, source description, and temporal resolution.

Context and passive source	Coverage[a] (%)	Description	Temporal resolution
Phone			
App usage (A)	75.4	Time spent using the mobile device	Half-hourly
Device unlocks (U[b])	70.8	Number of mobile device unlocks	Hourly
Mobility			
Steps (S[c])	77.0	Number of steps recorded	Half-hourly
Activity summary (V)	65.1	Time spent walking, running, and traveling in a vehicle	Daily
Distance (D)	23.0	Total distance traveled by any mode of transport	Half-hourly
Physical activity (C)	16.5	Time spent practicing a physical activity	Half-hourly
Physical distance (P)	6.8	Distance traveled while practicing a physical activity	Half-hourly
Rhythm			
Sleep (Z)	52.6	Overnight sleep duration	Daily
Home presence (H)	38.4	Binary indicator of presence at home	Half-hourly
Work presence (W)	38.4	Binary indicator of presence at work	Half-hourly
Other			
Sport (R)	99.8	Binary indicator of sport practiced along the day	Daily
Weekend (Q)	99.8	Binary indicator distinguishing weekdays from weekends	Daily

[a]Coverage represents the percentage of available data over the total observation period, and temporal resolution indicates whether values were recorded or computed half-hourly, hourly, or daily.
[b]U: notification unlocks.
[c]S: number of samples (multinomial sampling parameter).

Data Processing and Algorithm Development

Data From eB2 MindCare App

Multiple passive behavioral data sources across different context types were collected using the eB2 MindCare mobile app. Step count, sleep duration/stages, and physical distance are not computed by eB2 MindCare directly from raw sensor signals; instead, the app retrieves these variables from third-party health platform APIs (Google Fit, HealthKit [Apple], or Health Connect [Google], depending on the participant's device and operating system), which perform the underlying sensor fusion and pattern recognition. Mobility-related distance, by contrast, is derived from location data collected directly by the app: the distance between consecutive locations and the classification of a participant as being at home or away are computed using proprietary eB2 algorithms, including a density-based clustering method that identifies stable location clusters (eg, "home") from the stream of incoming location data over time.

App usage (time spent on the mobile device) is likewise a derived variable, computed within the eB2 CMB from raw device event data (eg, screen lock/unlock events) transmitted by the app, using a proprietary eB2 computation method. The CMB, part of the eB2 MindCare ecosystem, receives only such derived/interpreted variables—never raw sensor data that could be used to reconstruct the original sensor signal—and performs data cleaning and formatting for presentation on the clinical monitoring dashboard and, where applicable, feedback to the app.

Due to app permissions and device connectivity, some sources were unavailable for certain participants on specific dates within their observation period, either partially (records with missing values on a given date) or completely (no record for a given date). Loss of contact was identified through indicators such as app uninstallation, revocation of required permissions, or absence of connectivity/data transmission; critically, a genuine "zero" value (eg, zero steps recorded while the device remained active) was distinguished from a "missing/unknown" value (no data received, eg, due to a connectivity gap), so that true inactivity could be separated from monitoring failure. It should be noted that the exclusive use of the registered smartphone by the enrolled participant (including actions such as unlocking the device) is not independently verified by the app and instead relies on the study protocol and participant consent, consistent with standard practice in comparable smartphone-based monitoring studies. Table 1 presents a detailed description of each source used in the study, including its context, total coverage, and temporal resolution.

Data transmission from the device to the server is connectivity-dependent (in real-time when connected, buffered and sent upon reconnection otherwise) rather than following a fixed schedule; however, since relapse risk is estimated on a daily basis, all passive behavioral variables are aggregated at the daily level, and subdaily transmission timing does not affect the analysis or its results.

To ensure training stability and maximize data use, sleep metrics were reduced to duration and start time (converted to days). Continuous variables (app usage, steps, distance, physical distance, and activity summary) were log-transformed via $\log(x + 1)$ to compress extreme values and stabilize variance, and categorical presence/absence and yes/no variables (eg, home, work, sport, and weekend) were binarized for modeling with Bernoulli likelihoods. Negative values, which arise from sensor artifacts rather than genuine physiological signals, were handled at 2 levels of granularity: when a day's entire feature vector was predominantly negative, the day was treated as missing and excluded from the Bayesian online change-point detection (BOCPD) observation sequence, with elapsed time advanced without a run-length update; when only isolated elements of an otherwise valid daily feature vector were negative, they were corrected individually, either set to zero (for count-based variables such as steps and distance, where zero is a meaningful floor) or imputed with the mean (for variables where zero would be implausible or distorting). Missing data were otherwise handled per source.

Data From Electronic Health Record

Relevant clinical data include the clinical diagnosis (addiction, bipolar, or PD), relapse status (true/false), and relapse date. Relapse was operationally defined as a clinically recorded event indicating significant psychological decompensation, documented in the electronic health record by the treating clinical team. Depending on the diagnosis, such events included, for example, psychiatric hospitalization or emergency care, substance use relapse in participants with SUDs, affective episodes in participants with bipolar disorder (BD), and severe behavioral or emotional crises (eg, self-harm or acute decompensation requiring an intensification of care) in participants with PDs. For evaluation purposes, a detected change point was considered a true positive when it fell within a predefined 7-day window before a clinically recorded relapse event.

Clinical records were minimally preprocessed. Dates were standardized to a consistent format, and patient identifiers were matched with app user identifiers, resulting in 92 patients with valid linkages.

For the retained patients, relapse dates were extracted and used as reference events against which the algorithm's predictions were evaluated. There were 40 relapse events within the patients' monitoring periods, distributed among 21 patients.

Behavioral Profiling

Behavioral profiling was performed using the heterogeneous mixture model [57], which is designed to handle different types of data, missing observations, and high-dimensional data. In this study, Gaussian and Bernoulli mixture models were used for numerical and binary variables, respectively. As a result, each day is represented as a probability distribution over K different behavioral profiles. This approach is personalized, as the original passive data measured for each patient have their own latent sequence.

BOCPD

The latent profile distributions generated by the heterogeneous mixture model were used as input to BOCPD [58], an unsupervised probabilistic method for detecting changes in behavioral patterns over time. Since the latent behavioral state is represented as a probability distribution rather than a directly observed profile, the multinomial sampling strategy proposed in [59] was used to generate counting vectors from the latent profile distributions while preserving uncertainty. BOCPD recursively estimates the probability that the current behavior belongs to the existing behavioral pattern and infers a change point when newly observed behavior becomes inconsistent with it. The duration between consecutive change-points is modeled by the run-length variable. A geometric prior with parameter λ controls the expected duration in days of stable behavioral periods. To improve computational efficiency, the maximum run length was limited to a fixed window size, while multinomial sampling provided a tractable approximation of the recursive updates.

Hyperparameter Selection

Experiments were conducted to obtain the most appropriate set of hyperparameters, as detailed in Table 2. Combinations of different hyperparameter values were explored using a grid-search method. In this approach, a discrete set of candidate values is defined for each hyperparameter, and the model is run for all possible combinations within that grid. This allows the identification of the best-performing configuration among the evaluated combinations.

Table 2. Hyperparameters for model configuration.

Hyperparameter	Description
K	Number of profiles generated by the heterogeneous mixture model
S	Number of samples used for constructing the counting vector for each time step
λ	Expected number of days between 2 change-points
window	Number of days or amount of history over which the probability of a change-point is aggregated

In addition, to determine the boundary for detecting behavioral changes associated with relapse events, threshold values were swept from 0 to 1 to produce receiver operating characteristic curves. Each hyperparameter combination was compared by computing its area under the curve (AUC) score.

Source Analysis

To assess the individual contribution of each passive source described in Table 1, one separate experiment was conducted for each source using a fixed hyperparameter configuration. Each experiment was repeated across independent runs to account for the stochasticity of the framework. Then, a search for source combinations with improved performance was carried out. Given the combinatorial complexity of evaluating all possible subsets, a forward-style heuristic exploration strategy was adopted. Initially, pairwise combinations were evaluated, followed by selected triplets, with promising configurations progressively extended by incorporating additional sources. The selection of combinations was guided by domain knowledge and intermediate performance observations, prioritizing sources expected to capture complementary aspects of behavior.

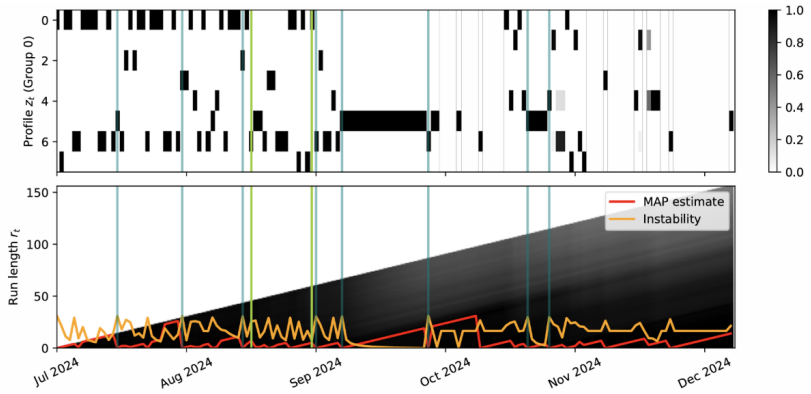
Validation

An indicator of behavioral instability was computed by aggregating and normalizing the probabilities of a change-point within the given window hyperparameter (Table 2). It represents the underlying signal detected by the algorithm. Later, the threshold parameter was applied to this instability to obtain a binary prediction vector, where 1 indicates a detected change point and 0 otherwise. This prediction vector was then compared against clinically recorded relapse events using a tolerance window of 7 days to allow for temporal misalignment. That is, predictions are considered correct only if they fall within the 7 days before the current day where a relapse occurred. Using this procedure, model performance metrics were obtained for evaluation. The metrics considered were the AUC; sensitivity or true positive rate (TPR); false positive rate (FPR), where FPR is equal to 1 minus specificity; and positive predictive value (PPV) or precision, and alarm rate, the proportion of evaluated days for which the model generated an alarm.

To narrow down the solution space, the most promising combinations of sources were selected based on a high TPR at an FPR of less than or equal to 5%. This evaluation criterion aims to prioritize the efficient management of available medical resources by constraining the number of false alarms.

Figure 2 provides a visual narrative of the relapse event detection process for an individual case. The use of behavioral profiling simplifies the visual representation of the discovery of change points by addressing the multidimensional problem. In this example, the model expects the participant to experience a behavioral shift around mid-July of 2024; this is considered a false alarm as no event is actually reported. On the other hand, 3 crises are predicted between mid-August and early September of that year. The first and third do not correspond to real events within the evaluation window and are therefore considered false negatives. However, the second is successfully detected and is hence regarded as a true positive.

Figure 2. Example of detection process: true (green) and predicted (blue) events; heterogeneous mixture model latent profiles (top) and change-point detection (CPD) run-length probability (bottom).



Ethical Considerations

The study adheres to the ethical principles outlined in the Declaration of Helsinki and was approved by the Ethics Committee of the Hospital Universitario 12 de Octubre (protocol code 24/217, approved on May 16, 2024). The collected data were encrypted, pseudoanonymized, and protected through secure authentication mechanisms. Data transmission was carried out via a RESTful (Representational State Transfer) API developed using the Java Spring framework, secured through the Secure Sockets Layer (SSL) protocol, and access was restricted using a token-based authentication policy in compliance with the OAuth 2.0 standard.

Results

Sample Description

As shown in the participant flowchart in Figure 3, a total of 92 participants were retained in the study. Table 3 shows the sample sociodemographic characteristics across all clinical diagnoses.

Figure 3. Participant flow diagram.

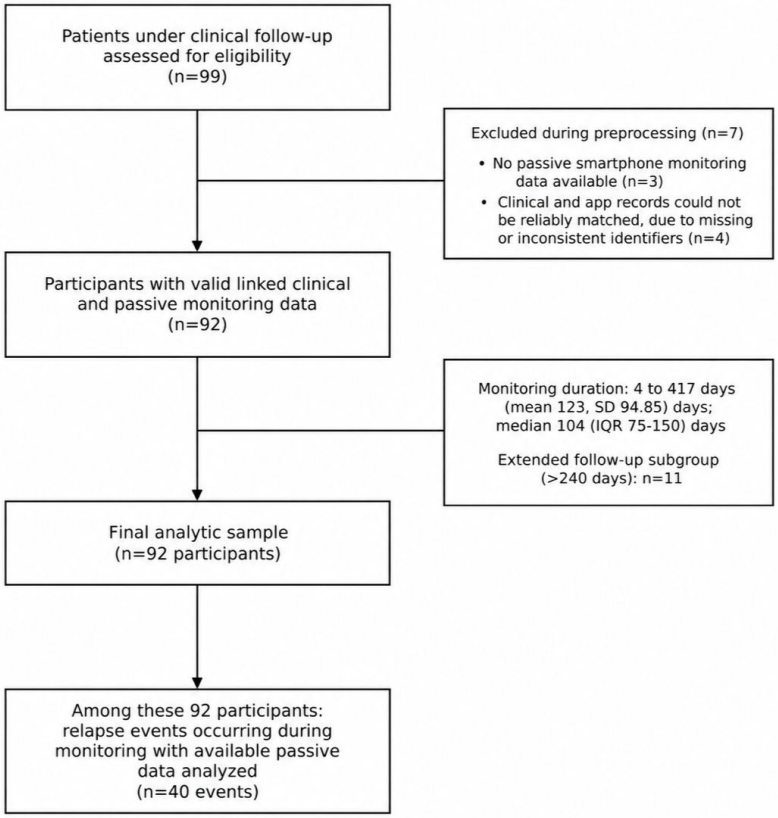


Table 3. Sample characteristics.

Characteristics	Values
Sex assigned at birth, n (%)	
Male	41 (44.6)
Female	51 (55.4)
Age (years), mean (SD)	47.7 (12.2)
Clinical diagnosis, n (%)	
Bipolar disorder	26 (28.3)
Personality disorder	19 (20.7)
Substance use disorder	18 (19.6)
Dual pathology	29 (31.5)
Nationality, n (%)	
Spanish	84 (91.3)
Chilean	2 (2.2)
Cuban	2 (2.2)
Colombian	1 (1.1)
British	1 (1.1)
Peruvian	1 (1.1)
Polish	1 (1.1)
Education level, n (%)	
No formal education	6 (6.5)
Lower secondary	18 (19.6)
Upper secondary	12 (13.0)
Vocational training	26 (28.3)
Undergraduate	24 (26.1)
Postgraduate	5 (5.4)
PhD	1 (1.1)
Socioeconomic level, n (%)	
Low	26 (28.3)
Lower-middle	43 (46.7)
Upper-middle	23 (25)
Relationship status, n (%)	
Single	41 (44.6)
Married	22 (23.9)
Divorced	11 (12.0)
In a relationship	11 (12.0)
Cohabiting	3 (3.3)
Separated	3 (3.3)
Widowed	1 (1.1)
Has children, n (%)	
Yes	58 (63.0)
No	34 (37.0)

Analysis of Passive Data Source Contribution

Individual sources were evaluated to assess whether they provided informative signals for the change-point detection (CPD) model and to establish individual source performance baselines. For each source, the experiment was repeated 10 times to account for the stochasticity of the model. Table 4 reports the mean AUC, SD, and range for the sources in the study. The sources physical activity (C) and physical distance (P) are not reported in the table because the algorithm could not extract enough information from them due to their poor coverage.

Table 4. Individual source performance for the framework[a].

Source	Mean (SD; minimum-maximum)	Range
R[b]	0.788 (0.038; 0.667-0.877)	0.210
Q[c]	0.768 (0.045; 0.619-0.860)	0.242
Z[d]	0.671 (0.014; 0.641-0.721)	0.080
V[e]	0.660 (0.022; 0.613-0.724)	0.111
W[f]	0.635 (0.047; 0.538-0.723)	0.186
A[g]	0.605 (0.030; 0.542-0.676)	0.134
H[h]	0.572 (0.009; 0.550-0.602)	0.052
S[i]	0.534 (0.011; 0.511-0.556)	0.045
D[j]	0.505 (0.005; 0.485-0.518)	0.033
U[k]	0.477 (0.023; 0.440-0.521)	0.082

[a]Reported statistics summarize the AUC distribution across 10 runs for each source.
[b]R: sport.
[c]Q: weekend indicator.
[d]Z: sleep.
[e]V: activity summary.
[f]W: work presence.
[g]A: app usage.
[h]H: home presence.
[i]S: steps.
[j]D: distance.
[k]U: notification unlocks.

The experimental phase built on the individual source results by exploring source combinations that could improve upon the corresponding single-source baselines. In the evaluation of source combinations, work presence (W), sleep (Z), and activity summary (V) consistently formed the core of the configurations achieving the highest AUC values, outperforming their individual source results. These results suggest that these sources capture complementary behavioral dimensions relevant to the detection task. The weekend source (Q) initially appeared to improve performance; however, further analysis indicated that it largely replicated the information provided by source W, suggesting redundancy between these variables.

The inclusion of additional sources, such as app usage (A), steps (S), and distance (D), was associated with further improvements in performance, indicating that combining multiple behavioral signals enhances the model's ability to detect relevant changes. However, this improvement was not unbounded as the inclusion of an excessive number of sources led to a deterioration in performance, suggesting that overly complex combinations may introduce noise or redundant information.

Final Model Performance and Event Detection

Based on the predefined selection criterion of maximizing TPR at 5% FPR, the analysis of the evaluation results identified the source combination ADSVWZ with hyperparameters K=8, S=5, λ=258, and *window*=1 as the best-performing CPD configuration among those evaluated. At an FPR of 5%, this combination achieved an average TPR of 40.12% (SD 5.66 percentage points; run-level 95% CI for the mean 36.61%-43.63%) across 10 stochastic runs (see Figure 2 left), as well as a mean AUC of 82.99% (SD 2.49 percentage points; run-level 95% CI for the mean 81.45%-84.53%). The best trial yielded a maximum TPR of 53.58% as shown in Figure 4 (right panel). The threshold that provided this result falls within the aggregated range between 0.56 and 0.59, since interpolation was necessary to obtain the TPR value at exactly 5% FPR. The difference between the average and maximum TPR at an FPR of 5% suggests variability in the model's performance across trials, indicating limited stability. When interpreting these estimates, it is worth noting that the CIs reflect how much the algorithm's output varied across repeated runs on the same data, rather than sampling uncertainty at the patient level. Additionally, this configuration yielded an average PPV of 0.026 (SD 0.004) over 10 runs, with a mean of 601 (SD 12) alarms, a mean alarm rate of 0.053 (SD 0.001), and a mean lead time of 3.4 (SD 0.4) days. The low value of PPV reflects the strong imbalance between the number of predictable events and the number of days with available data. However, the combination of low PPV and high alarm burden indicates that this result is not yet suitable for clinical deployment without further prospective validation and refinement.

The same was done for an FPR of 10%, for which the average TPR was 61.10% (SD 7.56 percentage points; run-level 95% CI for the mean 56.42%-65.79%), with a decision threshold between 0.15 and 0.22.

The most impactful hyperparameter was the *window* parameter. For a fixed CPD configuration, the AUC scores varied significantly between different *window* values (Figure 4 center), indicating that the model performed better with minimal or no accumulation of previous time steps. While for

the hyperparameters λ, K, and S (Figures 4 and 5), TPRs at low FPRs were similar and AUC variations between different values were modest, the final selected values correspond to those that achieved the maximum AUC. This suggests that these hyperparameters have no significant effect on detection performance.

Figure 4. Receiver operating characteristic (ROC) curves: mean performance, window-based comparison, and low–false positive rate (FPR) detail.

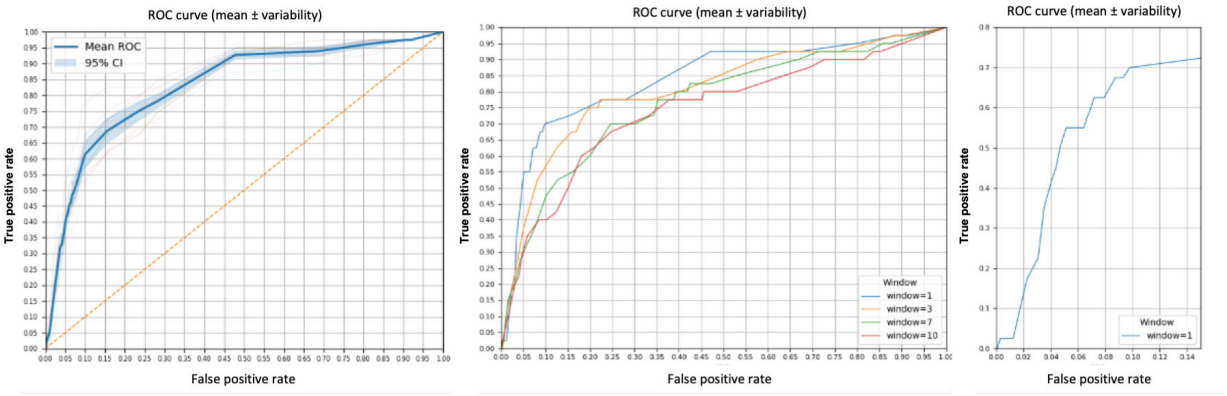
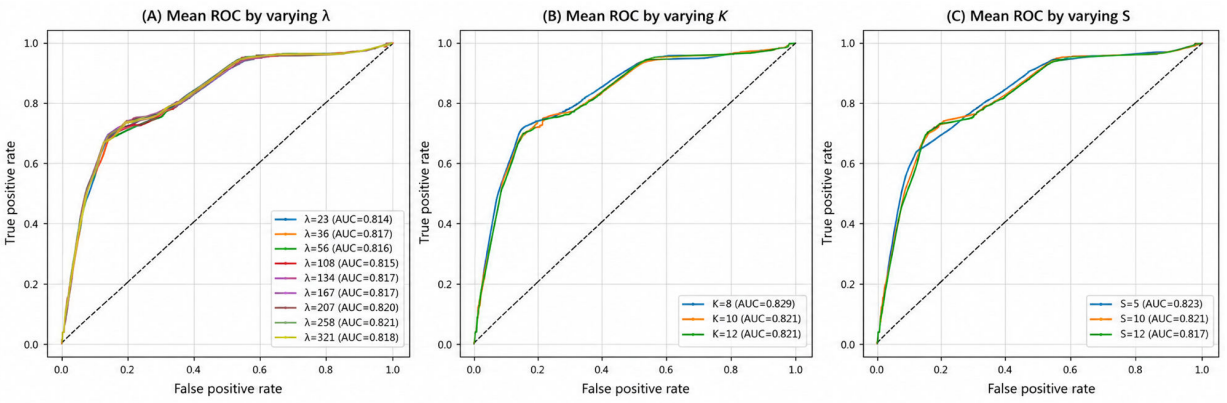


Figure 5. Changes in mean area under the curve (AUC) values when changing hyperparameters λ, number of latent components (K), or number of samples (multinomial sampling parameter; S) with respect to the final combination values. ROC: receiver operating characteristic.



Discussion

The present study aimed to develop, and provide a preliminary in-sample evaluation of, an algorithm capable of detecting behavioral changes associated with clinically relevant deterioration using passive data collected through smartphones. Specifically, we examined whether instability in daily behavioral patterns could serve as an early marker of relapse or clinical transition. The findings provide preliminary support for this hypothesis. The best-performing configuration, based on the combination of sources ADSVWZ, showed encouraging detection performance under restrictive conditions with low false alarm rates: it achieved an average TPR of 40.12% at a 5% FPR and 61.10% at a 10% FPR. In addition, the model showed a mean in-sample AUC of 82.99%, suggesting adequate discriminative ability to distinguish periods preceding relapse from stable behavioral states, although this estimate was obtained on the same data used for model tuning and should therefore be regarded as optimistic

These results suggest that loss of behavioral stability may be detectable prior to clinically recorded relapse events. Therefore, the main contribution of this study lies not only in identifying behavioral changes but also in demonstrating that the temporal dynamics of everyday behavior contain relevant prognostic information. This perspective is consistent with the conceptual framework outlined in the introduction, where mental health deterioration is described as a progressive process characterized by subtle changes in routine, activity, or biological rhythms before consolidating into a clinically manifest episode [21,37,38].

Accordingly, these findings reinforce the idea that, in mental health, deviations from an individual's habitual pattern may be more informative than point-in-time assessments or population-based thresholds [39,40]. In fact, this approach aligns with the shift toward dynamic models of psychopathology, in which fluctuations, loss of regularity, and regime shifts are considered central to understanding clinical evolution [13,15,17]. It is also consistent with the transdiagnostic framework introduced earlier, whereby processes such as behavioral disorganization or disruption of routines reflect states of vulnerability that are relatively independent of specific diagnoses [44-46].

In relation to the prior literature, it is important to note that most studies in digital phenotyping have focused on the association between behavioral variables and specific

symptoms or diagnostic categories. For instance, indicators such as physical activity, sleep, and mobility have been linked to conditions such as depression and BD [20,23]. However, the detection of clinical transitions has received comparatively less attention. In this context, the present study contributes to the field by operationalizing behavioral instability as a dynamic marker and applying an unsupervised approach to identify deviations from an individual's baseline pattern. In doing so, it shifts the focus from descriptive analysis toward risk-oriented monitoring.

Moreover, our findings are consistent with research showing that intraindividual variability may precede clinical change. In particular, fluctuations in daily activity have been shown to anticipate the onset of depressive symptoms in BD [6], while sleep disturbances and loss of regularity have been associated with destabilization processes across different conditions [10-15]. Similarly, studies on suicidal behavior have reported that changes in affective state, sleep, and behavior can precede periods of increased risk [32]. In line with this evidence, our results suggest that signals derived exclusively from passive data can capture this process of destabilization before it translates into a clinically observable event.

From a methodological perspective, the analysis of passive data sources provides additional insights. When evaluated individually, the passive behavioral sources showed heterogeneous performance, indicating that predictive utility cannot be explained solely by data coverage. Some sources with moderate coverage, such as sleep, achieved the highest individual performance, whereas other sources with greater availability contributed comparatively less. These findings suggest that the behavioral information captured by each source plays a central role in its usefulness for CPD.

The individual-source analysis also indicates that no single passive behavioral variable is sufficient to characterize clinically relevant deterioration. Rather, each source captures only a partial aspect of everyday functioning, supporting the rationale for integrating multiple complementary behavioral domains within the proposed framework.

However, when combinations of sources were examined, a different and more clinically meaningful pattern emerged. The best-performing configurations consistently included variables related to sleep, work presence, and overall activity level, which were further enhanced by the addition of sources such as phone usage and step count. More specifically, sleep, work presence, and activity summary formed the core of the most predictive configurations, suggesting that these variables capture complementary and nonredundant aspects of behavioral organization. In contrast, variables such as weekend status appeared to contribute limited additional value due to redundancy with other sources, particularly work presence, highlighting the importance of careful feature selection in multimodal models.

This suggests that instability detection does not rely on a single signal, but rather on the integration of multiple dimensions of daily behavior. In particular, these variables appear to capture complementary aspects of behavioral organization, such as temporal structuring, circadian rhythms, and functional regularity, all of which have been strongly linked to clinical trajectories in mental health [10-14].

At the same time, the observed deterioration in performance when incorporating an excessive number of sources suggests that increasing dimensionality does not necessarily improve predictive accuracy. Instead, overly complex combinations may introduce noise or redundant information, ultimately reducing model sensitivity. This finding emphasizes the importance of identifying parsimonious multimodal configurations that balance informational richness with signal quality. Taken together, these results highlight three key methodological principles for digital phenotyping: (1) sufficient temporal coverage is essential for reliable detection; (2) complementary behavioral domains should be prioritized over redundant signals; and (3) model performance depends on an optimal balance between multimodality and parsimony.

From a clinical perspective, these findings have several implications. First, they support the development of personalized monitoring approaches based on individual trajectories rather than normative comparisons. Second, the system allows detection thresholds to be adjusted according to the clinical context, facilitating adaptation across different healthcare settings. Third, because it relies on passive data, it reduces dependence on active patient participation, which is particularly advantageous in populations with low adherence. Taken together, these characteristics position the approach as a potentially valuable tool for secondary prevention and early relapse detection.

In this sense, the proposed approach aligns with emerging models of personalized medicine that estimate clinical risk based on individual trajectories over time. Previous work has shown that combining multiple sources of information (self-reports, clinical records, and professional assessments) can improve predictive performance [32,33]. It is therefore reasonable to suggest that integrating these sources with passive data, as in the present study, may lead to more accurate and clinically useful systems.

These findings should, however, be interpreted in light of several considerations inherent in this type of research. The heterogeneity of the sample, while consistent with a transdiagnostic and real-world approach, may introduce variability in behavioral patterns. In addition, differences in data availability across sources reflect the practical challenges of passive data collection in naturalistic settings and highlight the need to further optimize data continuity and quality. It is also important to note that not all deviations from baseline necessarily correspond to relapse and thus are best interpreted in conjunction with clinical assessment.

Additionally, the variability observed across model runs suggests limited stability in performance, likely influenced by the stochastic components of the algorithm (eg, behavioral profiling and multinomial sampling). This highlights the importance of further validation in larger samples and more controlled settings.

The findings should be interpreted in light of several considerations, consistent with the developmental nature of this study. As is common in early-stage algorithm development, model selection and performance assessment relied on the same set of relapse events, as no held-out test set or external validation sample was yet available. The reported results therefore characterize the model's behavior within the current sample and constitute a proof of concept rather than an estimate of real-world accuracy or clinical applicability. Similarly, the CIs capture variability across stochastic reruns of the algorithm rather than patient-level sampling uncertainty, offering a useful indication of the model's internal stability. Other factors, including heterogeneity in the clinical definition of relapse across diagnostic groups, temporal dependencies inherent to passive behavioral data, and the exclusion of events with insufficient preceding data, may also have influenced the results and point to concrete refinements for future work. These considerations define a clear roadmap: validation in larger, independent samples is the natural next step to establish whether the promising in-sample performance observed here translates into clinically useful prediction.

Building on these results, the next logical step is to move from detection to intervention. In this regard, a multicenter randomized controlled trial is currently underway to evaluate the effectiveness of a digital relapse prevention plan integrated into a mobile application for individuals with SUDs [60]. This study combines continuous monitoring, automated risk detection, and the real-time activation of personalized interventions, assessing outcomes such as time to relapse, treatment adherence, and emotional regulation. In this way, early detection is directly linked to clinical response.

Overall, the present study provides evidence that behavioral instability derived from passive data constitutes a promising marker for the early detection of clinically relevant changes. Its main contribution lies in capturing emerging transitions in everyday behavior and offering a complementary approach to traditional clinical assessment. These findings reinforce the potential of digital phenotyping as a tool to support a more preventive, personalized, and dynamically informed model of psychiatry.

Acknowledgments

The authors would like to thank all the participants who will take part in this project, as well as the professionals whose involvement and dedication make this work possible.

Generative AI Use

The authors declare that no generative AI tools were used in the generation of any portion of this manuscript.

Funding

The authors declare that this research received financial support. This publication is part of project CPP2022-009537, funded by MCIN/AEI/10.13039/501100011033 and by the European Union–NextGenerationEU/PRTR. This research was also partially funded by the Agencia Estatal de Investigación (AEI) through the Torres Quevedo Programme (grant number PTQ2023-013095 to HGM). AA-R is the director of ELLIS Unit Madrid (European Laboratory for Learning and Intelligent Systems), funded by the Community of Madrid. Additional financial support was provided by MCIN/AEI under grant CPP2023-010721; the "la Caixa" Foundation; and the Community of Madrid through the R and D Activities Program, project IDEA-CM (TEC-2024/COM-89), awarded by the Directorate-General for Research and Technological Innovation under Order 2402/2024 of May 31, 2024. This work also received partial funding from the SMARTomicS project (Prevención de la Conducta Suicida Mediante Intervenciones Terapéuticas Guiadas por Fenotipo Digital y Farmacogenética), funded in part by the Instituto de Salud Carlos III (ISCIII) and the European Union through NextGenerationEU funds under the Recovery and Resilience Facility (RRF/PRTR); grant PID2024-157856NB-I00, funded by MICIU/AEI/10.13039/501100011033 and ERDF/EU; and the Community of Madrid through the project Generalist Medical AI for Clinical Assessment and Risk Evaluation in Suicide Prevention (CARE-CM), reference SYG-2024/COM-870, awarded under the 2024 SINÉRGICOS Call by the Directorate-General for Research and Technological Innovation, pursuant to Order 5022/2025 of October 31, 2025.

Data Availability

The dataset generated and analyzed during this study is available from the corresponding author upon reasonable request. The dataset cannot be shared openly because it consists of sensitive, potentially reidentifiable behavioral and clinical data from a psychiatric population, and the informed consent and ethics approval under which it was collected do not permit unrestricted public release. The eB2 MindCare application is proprietary software; however, the modeling approach (heterogeneous mixture models and Bayesian online change-point detection), the hyperparameter search space, and the final model configuration are fully described in the *Methods* section to support conceptual reproducibility.

Authors' Contributions

Conceptualization: RS, AA-R
Formal analysis: XYL, AP
Funding acquisition: AA-R
Investigation: RS, XYL, AP, JCAJ, APR, JJMG, HG-M, AA-R
Methodology: RS, XYL, AP, AA-R

Project administration: RS, XYL, AP, AA-R
Resources: JCAJ, APR, JJMG, HG-M, AA-R
Supervision: AA-R
Validation: JCAJ, APR, JJMG, HG-M, AA-R
Visualization: RS, XYL, AP
Writing – original draft: RS, XYL, AP
Writing – review & editing: JCAJ, APR, JJMG, HG-M

Conflicts of Interest

The San Juan de Dios authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. HG-M, JCAJ are employees of Evidence-Based Behavior, the company responsible for the development of the digital health product eB2 Mindcare. AA-R is one of the founders of Evidence-Based Behavior. The authors declare that no other financial, commercial, or personal relationships with individuals or organizations exist that could inappropriately influence (bias) the work reported in this study.

References

1. Akinode AO, Ayadi OE, Ezerioha CC, et al. Machine learning approaches for early detection of mental health disorders using wearable devices and big data analytics. Int J Biol Pharm Sci Arch. 2025;10(2):006-023. [doi: 10.53771/ijbpsa.2025.10.2.0077]
2. Mikaeili N, Naeim M, Narimani M. Reimagining mental health with artificial intelligence: early detection, personalized care, and a preventive ecosystem. J Multidiscip Healthc. 2025;18:7355-7373. [doi: 10.2147/JMDH.S559626] [Medline: 41230404]
3. Fusar-Poli P, Salazar de Pablo G, Rajkumar RP, et al. Diagnosis, prognosis, and treatment of brief psychotic episodes: a review and research agenda. Lancet Psychiatry. Jan 2022;9(1):72-83. [doi: 10.1016/S2215-0366(21)00121-8] [Medline: 34856200]
4. Latalova K, Prasko J, Diveky T, Kamaradova D, Grambal A, Velartova H. Prodromes, precipitants, and risk factors for relapse in bipolar disorder. Neuro Endocrinol Lett. 2012;33(6):619-625. [Medline: 23160231]
5. Lawrie SM, O'Donovan MC, Saks E, Burns T, Lieberman JA. Towards diagnostic markers for the psychoses. Lancet Psychiatry. Apr 2016;3(4):375-385. [doi: 10.1016/S2215-0366(16)00021-3] [Medline: 27063388]
6. Ortiz A, Halabi R, Alda M, et al. Day-to-day variability in activity levels detects transitions to depressive symptoms in bipolar disorder earlier than changes in sleep and mood. Int J Bipolar Disord. Apr 2, 2025;13(1):13. [doi: 10.1186/s40345-025-00379-6] [Medline: 40175826]
7. Proudfoot J, Doran J, Manicavasagar V, Parker G. The precipitants of manic/hypomanic episodes in the context of bipolar disorder: a review. J Affect Disord. Oct 2011;133(3):381-387. [doi: 10.1016/j.jad.2010.10.051] [Medline: 21106249]
8. Subramanian K, Sarkar S, Kattimani S, Philip Rajkumar R, Penchilaiya V. Role of stressful life events and kindling in bipolar disorder: converging evidence from a mania-predominant illness course. Psychiatry Res. Dec 2017;258:434-437. [doi: 10.1016/j.psychres.2017.08.073] [Medline: 28870645]
9. Vafaie N, Kober H. Association of drug cues and craving with drug use and relapse: a systematic review and meta-analysis. JAMA Psychiatry. Jul 1, 2022;79(7):641-650. [doi: 10.1001/jamapsychiatry.2022.1240] [Medline: 35648415]
10. Basquin L, Maruani J, Leseur J, et al. Study of the different sleep disturbances during the prodromal phase of depression and mania in bipolar disorders. Bipolar Disord. Aug 2024;26(5):454-467. [doi: 10.1111/bdi.13429] [Medline: 38653574]
11. Liang L, Hou WK, Li TW, Liu H, Goodwin R, Lee TMC. Latent profiles and transitions of daily routine disruptions are associated with severity of symptoms of anxiety and depression. Leis Sci. Nov 17, 2025;47(8):1642-1657. [doi: 10.1080/01490400.2023.2228311]
12. Meyer N, Harvey AG, Lockley SW, Dijk DJ. Circadian rhythms and disorders of the timing of sleep. Lancet. Sep 24, 2022;400(10357):1061-1078. [doi: 10.1016/S0140-6736(22)00877-7] [Medline: 36115370]
13. Schreuder MJ, Hartman CA, George SV, et al. Early warning signals in psychopathology: what do they tell? BMC Med. Oct 14, 2020;18(1):269. [doi: 10.1186/s12916-020-01742-3] [Medline: 33050891]
14. Crouse JJ, Carpenter JS, Song YJC, et al. Circadian rhythm sleep-wake disturbances and depression in young people: implications for prevention and early intervention. Lancet Psychiatry. Sep 2021;8(9):813-823. [doi: 10.1016/S2215-0366(21)00034-1] [Medline: 34419186]
15. Helmich MA, Olthof M, Oldehinkel AJ, Wichers M, Bringmann LF, Smit AC. Early warning signals and critical transitions in psychopathology: challenges and recommendations. Curr Opin Psychol. Oct 2021;41:51-58. [doi: 10.1016/j.copsyc.2021.02.008] [Medline: 33774486]

16. Lee J, Barnett P, Goldsmith LP, et al. Early interventions for first onset of symptoms of mental health conditions: an umbrella review of systematic reviews. BMC Med. Jan 29, 2026;24(1):116. [doi: 10.1186/s12916-026-04617-1] [Medline: 41606589]
17. Nelson B, McGorry PD, Wichers M, Wigman JTW, Hartmann JA. Moving from static to dynamic models of the onset of mental disorder: a review. JAMA Psychiatry. May 1, 2017;74(5):528-534. [doi: 10.1001/jamapsychiatry.2017.0001] [Medline: 28355471]
18. Pedrero EJ, Ruiz JM, Verdejo A, Llanero M, Ambrosio E. Neurociencia y Adicción [Book in Spanish]. Sociedad Española de Toxicomanías; 2011. ISBN: 978-84-615-0572-2
19. Scott EM, Carpenter JS, Iorfino F, et al. Early intervention, prevention, and prediction in mood disorders: tracking multidimensional outcomes in young people presenting for mental health care. In: Baune BT, editor. Personalized Psychiatry. Academic Press; 2020:39-62. [doi: 10.1016/B978-0-12-813176-3.00005-5]
20. Smagula SF, Zhang G, Gujral S, et al. Association of 24-hour activity pattern phenotypes with depression symptoms and cognitive performance in aging. JAMA Psychiatry. Oct 1, 2022;79(10):1023-1031. [doi: 10.1001/jamapsychiatry.2022.2573] [Medline: 36044201]
21. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J. Sep 2015;38(3):218-226. [doi: 10.1037/prj0000130] [Medline: 25844912]
22. Guliyev C, İnce-Guliyev E, Ögel K. Predictors of relapse to alcohol and substance use: are there any differences between 3 and 12 months after inpatient treatment? J Psychoactive Drugs. 2022;54(4):358-367. [doi: 10.1080/02791072.2021.1976887] [Medline: 34553671]
23. Bufano P, Laurino M, Said S, Tognetti A, Menicucci D. Digital phenotyping for monitoring mental disorders: systematic review. J Med Internet Res. Dec 13, 2023;25(1):e46778. [doi: 10.2196/46778] [Medline: 38090800]
24. Doherty K, Balaskas A, Doherty G. The design of ecological momentary assessment technologies. Interact Comput. May 17, 2020;32(3):257-278. [doi: 10.1093/iwcomp/iwaa019]
25. Miralles I, Granell C, García-Palacios A, et al. Enhancing in vivo exposure in the treatment of panic disorder and agoraphobia using location-based technologies: a case study. Clin Case Stud. Apr 2020;19(2):145-159. [doi: 10.1177/1534650119892900]
26. Alòs F, Puig-Ribera A. Uso de wearables y aplicaciones móviles (mHealth) para cambiar los estilos de vida desde la práctica clínica en atención primaria: una revisión narrativa [Article in Spanish]. Atención Primaria Práctica. Dec 2021;3:100122. [doi: 10.1016/j.appr.2021.100122]
27. Kazdin AE, Rabbitt SM. Novel models for delivering mental health services and reducing the burdens of mental illness. Clin Psychol Sci. Apr 2013;1(2):170-191. [doi: 10.1177/2167702612463566]
28. Beckjord E, Shiffman S. Background for real-time monitoring and intervention related to alcohol use. Alcohol Res. 2014;36(1):9-18. [doi: 10.35946/arcr.v36.1.02] [Medline: 26258996]
29. Beiwinkel T, Kindermann S, Maier A, et al. Using smartphones to monitor bipolar disorder symptoms: a pilot study. JMIR Ment Health. Jan 6, 2016;3(1):e2. [doi: 10.2196/mental.4560] [Medline: 26740354]
30. Shiffman S. Conceptualizing analyses of ecological momentary assessment data. Nicotine Tob Res. May 2014;16 Suppl 2(Suppl 2):S76-S87. [doi: 10.1093/ntr/ntt195] [Medline: 24323571]
31. Wray TB, Merrill JE, Monti PM. Using ecological momentary assessment (EMA) to assess situation-level predictors of alcohol use and alcohol-related consequences. Alcohol Res. 2014;36(1):19-27. [doi: 10.35946/arcr.v36.1.03] [Medline: 26258997]
32. Barrigon ML, Romero-Medrano L, Moreno-Muñoz P, et al. One-week suicide risk prediction using real-time smartphone monitoring: prospective cohort study. J Med Internet Res. Sep 1, 2023;25(1):e43719. [doi: 10.2196/43719] [Medline: 37656498]
33. Chih MY, Patton T, McTavish FM, et al. Predictive modeling of addiction lapses in a mobile health application. J Subst Abuse Treat. Jan 2014;46(1):29-35. [doi: 10.1016/j.jsat.2013.08.004] [Medline: 24035143]
34. Miklowitz DJ, Price J, Holmes EA, et al. Facilitated integrated mood management for adults with bipolar disorder. Bipolar Disord. Mar 2012;14(2):185-197. [doi: 10.1111/j.1399-5618.2012.00998.x] [Medline: 22420594]
35. Papachristou E, Oldehinkel AJ, Ormel J, et al. The predictive value of childhood subthreshold manic symptoms for adolescent and adult psychiatric outcomes. J Affect Disord. Apr 1, 2017;212:86-92. [doi: 10.1016/j.jad.2017.01.038] [Medline: 28157551]
36. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess. Dec 2009;21(4):486-497. [doi: 10.1037/a0017074] [Medline: 19947783]

37. Heinz A, Deserno L, Zimmermann US, Smolka MN, Beck A, Schlagenhauf F. Targeted intervention: computational approaches to elucidate and predict relapse in alcoholism. Neuroimage. May 1, 2017;151:33-44. [doi: 10.1016/j.neuroimage.2016.07.055] [Medline: 27480622]
38. Kadirvelu B, Bellido Bel T, Freccero A, Di Simplico M, Nicholls D, Faisal AA. Digital phenotyping for adolescent mental health: feasibility study using machine learning to predict mental health risk from active and passive smartphone data. J Med Internet Res. Feb 4, 2026;28:e72501. [doi: 10.2196/72501] [Medline: 41637624]
39. Kandola A, Lewis G, Osborn DPJ, Stubbs B, Hayes JF. Depressive symptoms and objectively measured physical activity and sedentary behaviour throughout adolescence: a prospective cohort study. Lancet Psychiatry. Mar 2020;7(3):262-271. [doi: 10.1016/S2215-0366(20)30034-1] [Medline: 32059797]
40. Shin J, Bae SM. A systematic review of location data for depression prediction. Int J Environ Res Public Health. May 29, 2023;20(11):5984. [doi: 10.3390/ijerph20115984] [Medline: 37297588]
41. Elovainio M, Kuula L, Halonen R, Pesonen AK. Dynamic fluctuations of emotional states in adolescents with delayed sleep phase-A longitudinal network modeling approach. J Affect Disord. Nov 1, 2020;276:467-475. [doi: 10.1016/j.jad.2020.07.050] [Medline: 32741749]
42. Paterson JL, Dorrian J, Ferguson SA, et al. Changes in structural aspects of mood during 39–66 h of sleep loss using matched controls. Appl Ergon. Jan 2011;42(2):196-201. [doi: 10.1016/j.apergo.2010.06.014] [Medline: 20659729]
43. Ryan O, Dablander F, Haslbeck JMB. Toward a generative model for emotion dynamics. Psychol Rev. Mar 2025;132(2):416-441. [doi: 10.1037/rev0000513] [Medline: 40146601]
44. Berrouiguet S, Ramírez D, Barrigón ML, et al. Combining continuous smartphone native sensors data capture and unsupervised data mining techniques for behavioral changes detection: a case series of the evidence-based behavior (eB2) study. JMIR Mhealth Uhealth. Dec 10, 2018;6(12):e197. [doi: 10.2196/mhealth.9472] [Medline: 30530465]
45. Dalgleish T, Black M, Johnston D, Bevan A. Transdiagnostic approaches to mental health problems: current status and future directions. J Consult Clin Psychol. Mar 2020;88(3):179-195. [doi: 10.1037/ccp0000482] [Medline: 32068421]
46. Wise T, Robinson OJ, Gillan CM. Identifying transdiagnostic mechanisms in mental health using computational factor modeling. Biol Psychiatry. Apr 15, 2023;93(8):690-703. [doi: 10.1016/j.biopsych.2022.09.034] [Medline: 36725393]
47. Fox CA, Teckentrup V, Donegan KR, et al. Cognitive arbitration between candidate dimensions of psychopathology. Mol Psychiatry. Mar 2026;31(3):1634-1647. [doi: 10.1038/s41380-025-03297-2] [Medline: 41094062]
48. Wigman JTW, van Os J, Thiery E, et al. Psychiatric diagnosis revisited: towards a system of staging and profiling combining nomothetic and idiographic parameters of momentary mental states. PLoS One. 2013;8(3):e59559. [doi: 10.1371/journal.pone.0059559] [Medline: 23555706]
49. Rashid Z, Folarin AA, Zhang Y, et al. Digital phenotyping of mental and physical conditions: remote monitoring of patients through RADAR-Base platform. JMIR Ment Health. Oct 23, 2024;11(1):e51259. [doi: 10.2196/51259] [Medline: 39441952]
50. Sükei E, Norbury A, Perez-Rodriguez MM, Olmos PM, Artés A. Predicting emotional states using behavioral markers derived from passively sensed data: data-driven machine learning approach. JMIR Mhealth Uhealth. Mar 22, 2021;9(3):e24465. [doi: 10.2196/24465] [Medline: 33749612]
51. Oudin A, Maatoug R, Bourla A, et al. Digital phenotyping: data-driven psychiatry to redefine mental health. J Med Internet Res. Oct 4, 2023;25(1):e44502. [doi: 10.2196/44502] [Medline: 37792430]
52. Rajagopalan A, Shah P, Zhang MW, Ho RC. Digital platforms in the assessment and monitoring of patients with bipolar disorder. Brain Sci. Nov 12, 2017;7(11):150. [doi: 10.3390/brainsci7110150] [Medline: 29137156]
53. Rickard N, Arjmand HA, Bakker D, Seabrook E. Development of a mobile phone app to support self-monitoring of emotional well-being: a mental health digital innovation. JMIR Ment Health. Nov 23, 2016;3(4):e49. [doi: 10.2196/mental.6202] [Medline: 27881358]
54. Sathyanarayana A, Joty S, Fernandez-Luque L, et al. Sleep quality prediction from wearable data using deep learning. JMIR Mhealth Uhealth. Nov 4, 2016;4(4):e125. [doi: 10.2196/mhealth.6562] [Medline: 27815231]
55. Riadi I, Kervin L, Teo K, Churchill R, Cosco TD. Digital interventions for depression and anxiety in older adults: protocol for a systematic review. JMIR Res Protoc. Dec 23, 2020;9(12):e22738. [doi: 10.2196/22738] [Medline: 33355095]
56. Gunsilius CZ, Heffner J, Bruinsma S, et al. SOMAScience: a novel platform for multidimensional, longitudinal pain assessment. JMIR Mhealth Uhealth. Jan 12, 2024;12:e47177. [doi: 10.2196/47177] [Medline: 38214952]
57. Moreno-Muñoz P, Romero-Medrano L, Moreno A, Herrera-López J, Baca-García E, Artés-Rodríguez A. Passive detection of behavioral shifts for suicide attempt prevention. arXiv. Preprint posted online on Nov 14, 2020. [doi: 10.48550/arXiv.2011.09848]
58. Adams RP, MacKay DJC. Bayesian online changepoint detection. arXiv. Preprint posted online on Oct 19, 2007. [doi: 10.48550/arXiv.0710.3742]

59. Romero-Medrano L, Moreno-Muñoz P, Artés-Rodríguez A. Multinomial sampling of latent variables for hierarchical change-point detection. J Sign Process Syst. Feb 2022;94(2):215-227. [doi: 10.1007/s11265-021-01705-8]
60. Salom R, Pico Rada Á, Muñoz García JJ, García-Mieres H, Artés-Rodríguez A. Digital relapse prevention plan for substance use disorders: study protocol for a multicentre randomised controlled trial. BMJ Health Care Inform. Nov 24, 2025;32(1):e101808. [doi: 10.1136/bmjhci-2025-101808] [Medline: 41290251]

Abbreviations

AUC: area under the curve
BD: bipolar disorder
BOCPD: Bayesian online change-point detection
CMB: clinical monitoring backend
CPD: change-point detection
eB2: Evidence-Based Behavior
FPR: false positive rate
PD: personality disorder
PPV: positive predictive value
RESTful: Representational State Transfer
SSL: Secure Sockets Layer
SUD: substance use disorder
TPR: true positive rate

Edited by Alicia Stone; peer-reviewed by Arushi Singh, Myron Pulier, Yihan Hu; submitted 22.Apr.2026; initial comments to author 29.Jun.2026; final revised version received 09.Sep.2026; accepted 15.Oct.2026; published 20.Nov.2026

Please cite as:
Salom R, Leong XY, Poncet A, Aradillas Jaramillo JC, Pico Rada Á, Muñoz García JJ, García-Mieres H, Artés-Rodríguez A Digital Phenotyping of Behavioral Instability as an Early Detection of Clinically Relevant Deterioration in Mental Health: Prospective Observational Multicenter Study
JMIR Form Res 2026;10:e98935
URL: https://formative.jmir.org/2026/1/e98935
doi: 10.2196/98935

© Rafael Salom, Xin Ying Leong, Adeline Poncet, José Carlos Aradillas Jaramillo, Álvaro Pico Rada, Juan Jesús Muñoz García, Helena García-Mieres, Antonio Artés-Rodríguez. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.