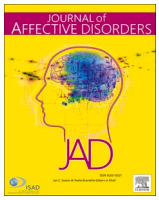
Journal of Affective Disorders 391 (2025) 119972

Contents lists available at ScienceDirect

Journal of Affective Disorders

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



Research paper

Early warning signals of bipolar relapse: Investigating critical slowing down in smartphone data

Vera M. Ludwig [a,*], Carl A. Bittendorf [b], Iris Reinhard [c], Andreas B. Neubauer [d], Eva Mennigen [a], Esther Mühlbauer [a], Wolfram E. Severus [e,f], Michael Bauer [a], Ulrich W. Ebner-Priemer [b,g]

[a] Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Carl Gustav Carus, Dresden University of Technology, Dresden, Germany
[b] Mental mHealth Lab, Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany
[c] Core Facility Biostatistics, Central Institute of Mental Health, University of Heidelberg, Medical Faculty Mannheim, Mannheim, Germany
[d] Institute of Psychology, RWTH Aachen University, Aachen, Germany
[e] Asklepios Klinik Nord-Ochsenzoll, Hamburg, Germany
[f] Technical University of Dresden, Dresden, Germany
[g] Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, University of Heidelberg, Medical Faculty Mannheim, Mannheim, Germany

ARTICLE INFO

Keywords:
Bipolar disorder
Digital phenotyping
Early warning signals
Critical slowing down
Prediction
Passive sensing

ABSTRACT

Background: Early warning signals (EWS) based on dynamical systems theory, such as increased autocorrelation (AR) and variance, may indicate impending mood episodes in bipolar disorder (BD). This study examines whether smartphone-based digital phenotyping can detect these signals before depressive and (hypo)manic episodes.

Methods: We analyzed smartphone sensor and self-report data from 29 BD patients (16 female, mean age of 43,97 ± 11.9 years) over one year, totaling 10,587 patient days and including 30 depressive and 20 (hypo)manic episodes in 22 patients. 26 bi-weekly expert interviews per patient established daily disease status. AR, variance and moving averages were calculated from passively assessed activity, communication, and smartphone-usage, as well as self-reported sleep parameters. Multilevel logit models assessed whether these measures could predict pre-episode weeks of depression or mania compared to euthymic days. Receiver operating characteristics (ROC) curves evaluated clinical utility.

Results: Several parameters significantly predicted pre-episode weeks, but no single robust predictor emerged. Manic episodes were best predicted by altered AR and variance in activity-related measures, while sleep parameters predicted both manic and depressive transitions. Latent factors combining multiple parameters showed stronger predictive potential than individual variables. However, ROC analyses revealed that even the best predictors did not meet predefined clinical utility thresholds.

Conclusions: Smartphone-based digital phenotyping holds promise for early detection of BD mood episodes. However, predictive accuracy remains below clinically useful levels. Future research should refine parameters, explore machine-learning approaches, and optimize analytical frameworks to realize the full potential of relapse prediction in BD.

1. Introduction

Detecting prodromal signs of relapse or recurrence into depression or mania is crucial for effective bipolar disorder (BD) management (Goodwin et al., 2016), as early recognition enables timely interventions to prevent full-blown episodes. While early studies suggested individuals with BD can detect their individual prodromal signs fairly well (Jackson et al., 2003), recent research indicates that 28 % to 50 % fail to do so before recurrence (Goossens et al., 2010; Mantere et al., 2008). Clinical and sociodemographic factors - such as being in a manic phase or older age – further reduce this ability (de Assis da Silva et al., 2015; Perich et al., 2013).

Given these challenges, digital phenotyping has emerged as a promising approach for continuous, objective monitoring of psychopathology (Insel, 2018). Leveraging smartphone-based data enables real-time tracking of behavioural indicators to monitor and predict

* Corresponding author.
E-mail address: Vera.Ludwig@ukdd.de (V.M. Ludwig).

https://doi.org/10.1016/j.jad.2025.119972
Received 14 April 2025; Received in revised form 22 July 2025; Accepted 24 July 2025
Available online 28 July 2025

symptoms (Brietzke et al., 2019; Ikäheimonen et al., 2024). Digital phenotyping comprises *active* monitoring – self-report measures such as ecological momentary assessment, EMA (Stone and Shiffman, 1994), experience sampling method, ESM (Csikszentmihalyi and Larson, 1987), or ambulatory assessment (Trull and Ebner-Priemer, 2009) – and *passive* monitoring, which continuously collects sensor data (e.g. GPS, accelerometer), without active patient input, enhancing ecological validity (Ebner-Priemer et al., 2009).

Despite the initial enthusiasm and rapid expansion of research on digital phenotyping, no smartphone-based monitoring system has proven robust for clinical use (de Azevedo Cardoso et al., 2024; Ortiz et al., 2021). Several methodological limitations contribute to this gap. First, mood episodes are rare in existing studies due to small sample sizes and short monitoring periods, making it difficult to gather sufficient episode data. Second, mood assessments often lack adequate temporal resolution, relying on infrequent clinical interviews (e.g., every two months) or cut-off scores in self-rating tools, limiting detection of dynamic changes before relapse. Third, most studies rely on dimensional self-ratings rather than clearly defined expert-rated transitions between clinical states (e.g. from euthymia to depression) based on DSM/ICD diagnostic criteria (Ebner-Priemer et al., 2020; Ebner-Priemer and Santangelo, 2020). These limitations highlight the need for high-quality longitudinal data and refined analytical frameworks to unlock the full potential of digital phenotyping for relapse prediction (Salvi et al., 2021).

One theoretical framework that has gained attention for anticipating recurrences in affective disorders is dynamical systems theory (Scheffer et al., 2024b; Van De Leemput et al., 2014). Originating in mathematics and successfully applied across different scientific fields such as ecology, climate research, and medicine (Dakos et al., 2023; Trefois et al., 2015); this framework suggests that critical transitions in complex systems – such as shifts in mood states – are preceded by a gradual loss of resilience, descriptively termed 'critical slowing down'. This phenomenon reflects a system's declining ability to recover from perturbations, resulting in both more extreme fluctuations and a slower return to equilibrium (Scheffer et al., 2012, 2024a). Statistically, this manifests as increased autocorrelation (AR) and increased variance in system-specific measures (Dablander et al., 2022; Dakos et al., 2012). Although increased autocorrelation (rigidity) and increased variance (extreme fluctuations) may appear conceptually opposed, they often co-occur as a system approaches a critical transition. This reflects a weakening of stabilizing forces: perturbations not only persist longer (temporal inertia) but also tend to grow larger (instability), resulting in both heightened persistence and variability. Given its broad applicability, critical slowing down has been proposed as an early warning sign (EWS) for mood episodes in BD (Nunes et al., 2022) and has shown promise in a simulation study (Bayani et al., 2017).

However, empirical investigations of critical slowing down preceding mood transitions in BD have yielded mixed results and suffer from the methodological limitations mentioned above. First, many studies analyze few mood episodes (e.g., n = 15 Bos et al., 2022; n = 8 Jakobsen et al., 2024; Kunkels et al., 2021; n = 1 Nakamura et al., 2013; n = 0 Curtiss et al., 2019), limiting statistical power. Second, mood episodes are frequently defined by simple cut-offs in self-report questionnaires (Bos et al., 2022; Kunkels et al., 2021; Nakamura et al., 2013), rather than categorical diagnoses derived from expert interviews. Third, even when expert assessments are used, they are often conducted only monthly (Curtiss et al., 2019) or bi-monthly (Jakobsen et al., 2024), which may be too infrequent to capture the precise onset of a transition. Consequently, while some studies emphasize the promise of EWS, reporting increased AR and/or variance before most affective episodes (Curtiss et al., 2019; Jakobsen et al., 2024; Kunkels et al., 2021; Nakamura et al., 2013), others find no such patterns (Ludwig et al., 2024), or raise concerns about high false-positive/negative rates and inter-individual heterogeneity (Bos et al., 2022).

Previous studies have examined critical slowing down in self-reported mood or affect captured via EMA (Bos et al., 2022; Ludwig et al., 2024), physical activity using wrist-worn actigraphy (Jakobsen et al., 2024; Kunkels et al., 2021; Nakamura et al., 2013), or a combination of both (Curtiss et al., 2019). However, for automated recurrence or relapse prediction without active patient participation, research must focus on detecting EWS in passive smartphone-based data.

In sum, empirical evidence of critical slowing down as a precursor to mood changes in BD remains inconclusive, partly because of methodological inconsistencies. The limited number and imprecise assessment of mood episodes highlight the need for studies providing expert ratings, high temporal precision, and prolonged assessment periods to acquire sufficient high-resolution validated pre-episode data. Additionally, evaluating smartphone-based, mostly passive parameters is key for clinical applicability. Given the successful application of critical slowing down to capture complex systems' shifts across fields, further exploration in BD recurrence prediction is warranted.

Thus, the aim of this exploratory study is to investigate whether indicators of critical slowing down (EWS: rising AR and variance) in smartphone-based digital phenotyping measures can predict emerging mood episodes in patients with BD. We will (1) assess whether EWS predict the weeks before depressive or hypomanic/manic episodes ("pre-episode weeks") compared to euthymia, (2) compare EWS to simple mean values of each parameter to determine added predictive value of EWS, (3) evaluate if aggregating multiple predictors to latent EWS improves the prediction of pre-episode weeks, and (4) explore the clinical utility of our predictive markers.

2. Methods

2.1. Participants and procedures

Data were drawn from the BipoSense study (Ebner-Priemer et al., 2020; Ludwig et al., 2024), which tracked BD patients for one year via the movisensXS app (movisens GmbH, Karlsruhe, Germany) using daily active and continuous passive data collection (for an overview of the study design compare Fig. 1). Participants were recruited between November 2014 and June 2016 from the University Hospital Dresden outpatient clinic and media outreach. Study details, including inclusion and exclusion criteria can be found in the supplementary materials. The study was conducted in accordance with the Declaration of Helsinki and received ethics approval from the Technical University of Dresden (EK-Nr.: 26012014). Participants were compensated €35 per month for their time, travel, and potential phone plan costs.

The present study is a secondary analysis of the BipoSense dataset (Ebner-Priemer et al., 2020), which originally aimed to demonstrate the feasibility of year-long, high-frequency smartphone-based monitoring in bipolar disorder and to overcome common methodological limitations in digital phenotyping. While the parent study modeled same-day associations between latent behavioural and symptom variables, the current study investigates whether early warning signals (EWS) in these behavioural domains predict emerging mood episodes. We retained the validated latent constructs (activity, sleep, communicativeness) but applied a distinct, prediction-oriented analytic framework using multilevel logit models.

2.2. Measures

2.2.1. Expert interviews

A trained clinical psychologist conducted bi-weekly interviews (alternating between in-person and telephone) using the SCID-I section A for affective episodes (First et al., 2015), assessing mood status over the preceding 14 days ("ground truth"). Mood episodes were coded categorically, with subthreshold symptoms labeled as euthymia. Dimensional symptom scales (e.g. Young Mania Rating Scale (Young et al., 1978), Bech-Rafaelsen Mania Scale (Bech et al., 1979) and Montgomery-Asberg Depression Inventory (Montgomery and Asberg,

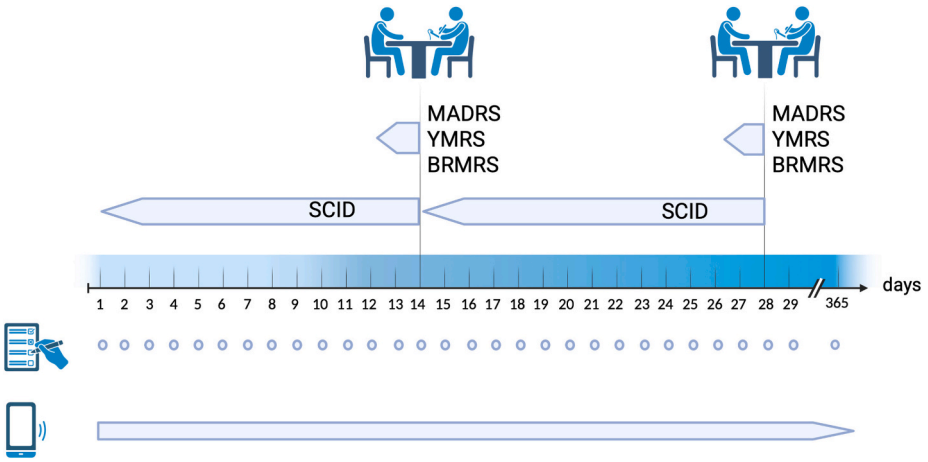


Fig. 1. BipoSense study design.
Over the course of one year, patients participated in 26 bi-weekly expert interviews. During each interview, affective status over the past 14 days was assessed categorically using the Structured Clinical Interview for DSM Disorders (SCID), Section A. Dimensional symptom severity over the past two days was evaluated using standardized questionnaires: the Montgomery-Åsberg Depression Rating Scale (MADRS), Young Mania Rating Scale (YMRS), and Bech-Rafaelsen Mania Rating Scale (BRMRS). Additionally, participants completed a daily electronic diary (e-diary) and had passive smartphone data continuously recorded.

1979)) were administered, but are not part of this manuscript.

2.2.2. Digital phenotyping

The movisensXS app (https://www.movisens.com/en/products/movisensxs/) is a research-grade experience sampling platform designed for ambulatory assessment. In our study, it continuously collected passive smartphone-sensor data and prompted daily electronic diary (e-diary) entries for the entire study period. The passive data collection included GPS location, accelerometer-based physical activity, screen on/off status, and communication behavior (e.g., call logs). All data were encrypted on the device and securely transmitted to a GDPR-compliant server infrastructure.

For active data collection, the app prompted participants to complete daily e-diaries on mood, sleep and medication intake, adapted from the validated ChronoRecord tool (Bauer et al., 2004, 2008). These entries were prompted hourly from 8 pm to midnight and required an average of 3 min to complete.

2.3. Sample characteristics

A total of 112 patients were contacted; 53 were screened and 31 were included. One patient dropped out after three weeks due to the study smartphone (Android) incompatibility with their other devices (iOS), and another's private phone (Huawei Technologies Co Ltd., Shenzen, China) impeded complete data extraction, leaving 29 patients (55 % female, mean age 43,97 ± 11.9). Thus, 29 of the 31 enrolled participants were included in the final analyses, corresponding to a 6,5 % exclusion rate post-inclusion. Seventeen had BD I and 12 had BD-II with an average history of 7.07 depressive, 2.97/2.76 (hypo)manic episodes (details Table 1).

A total of 10,584 patient days were recorded (mean 365.07 days per participant). 22 patients (75.86 %) experienced a total of 30 depressive and 20 (hypo)manic episodes during the study period, resulting in a mean number of 2.27 episodes/year for patients experiencing relapse or recurrence. 16 patients provided at least one depressive episode, 11 at least one (hypo)manic episode and 5 patients provided at least one episode of each polarity. Seven patients (24.14 %) did not experience any mood episode during the study period. Bi-weekly interviews were 97 % complete (n = 726) with 25 missed due to illness, absence, or scheduling issues. Daily e-diaries were 89 % complete (n = 9,433), with 11 % missing due to forgetfulness, switched-off phones, low battery, or app issues. Passive sensing data were available for 99 % of patient days, with 1 % (n = 152) completely missing due to phone issues (Ebner-Priemer et al., 2020; Mühlbauer, 2020).

Table 1
Sociodemographic and clinical baseline data.

	N (%)	Mean (SD)	Range
Female	16 (55.2 %)		
Male	13 (44.8 %)		
Age		43.97 (11.9)	25–70
Ethnicity: Caucasian	29 (100 %)		
Bipolar I	17 (58.62)		
Bipolar II	12 (41.38)		
Lifetime depressive episodes		7.07 (5.61)	2–30
Lifetime hypomanic episodes		2.97 (3.78)	0–15
Lifetime manic episodes		2.76 (3.48)	0–10
Lifetime hospital admissions due to BD		3.59 (3.70)	0–15
Psychopharmacological medication class*			
Lithium	15 (51.7 %)		
Atypical antipsychotics	10 (34.5 %)		
Anticonvulsants	11 (37.9 %)		
Antidepressants	8 (27.6 %)		
Psychopathology scores			
MADRS score		2.62 (3.05)	0–10
YMRS score		1.41 (2.13)	0–8
BRMRS score		0.72 (1.34)	0.6

Values are presented as N (%) for categorical variables and mean (SD) for dimensional variables. SD = Standard deviation; BD = Bipolar Disorder, MADRS = Montgomery Asberg Depression Rating Scale; YMRS = Young Mania Rating Scale; BRMRS = Bech Rafaelsen Mania Rating Scale.* Patients may be taking more than one medication.

2.4. Data preparation

2.4.1. Defining mood status

Bi-weekly SCID interviews determined daily mood status (euthymic, depressed, or manic, with hypomanic episodes grouped under mania) for the past 14 days. Pre-episode weeks were classified as 'early' (days −14 to −8) and 'late' (days −7 to −1), stratified by episode polarity. Each occurring mood episode was included in the study, this means a patient with several mood episodes, regardless of mood polarity, contributed several pre-episode periods to the study.

2.4.2. Digital phenotyping variables

The movisensXS app passively recorded smartphone usage data (e.g. calls, activities, and position tracking via GPS, WiFi and cell tower signals), and generates data entries at irregular time intervals, e.g. a missed call, an

activity classified as walking on foot or a GPS coordinate. The data are automatically generated by an algorithm of the movisensXS software (https://docs.movisens.com/movisensXS/mobile_sensing/#features-library-version-version). Given the virtually infinite number of mobile sensing parameters and their derivatives (Ebner-Priemer et al., 2020; Ebner-Priemer and Santangelo, 2020), we deliberately restricted our analyses to a theory-driven set of predictors (i.e. available smartphone-parameters from movisensXS that most closely relate to psychopathological symptoms) that had been used in previous publications by the same authors. This strategy was chosen to reduce the risk of α-error inflation and to minimize the likelihood of model overfitting (Ebner-Priemer et al., 2020; Mühlbauer et al., 2018). All passive sensing parameters were aggregated daily to align with ground truth assessments. Sleep was assessed daily via e-diaries, generating the variables "hours asleep" and "time got up". The complete code for data pre-processing, statistical analyses and visualization is available on GitHub (https://github.com/CarlBittendorf/BipoSense). A detailed description of the computation of all parameters can be found in the supplementary material.

2.4.3. Latent variables

Three latent constructs - sleep, activity and communicativeness - were derived using a latent multilevel structural equation model that empirically combined relevant mobile sensing parameters (Ebner-Priemer et al., 2020: Communication: Outgoing calls, Incoming missed calls, Outgoing not reached calls, Minutes call duration, number of unique conversation partners; Active: Steps, Minutes in vehicle, Minutes on foot, Minutes still, Kilometers fast, Kilometers slow; Sleep: Hours asleep, Time got up). Estimated factor scores for each latent variable were extracted daily for each participant and used for further analyses.

2.4.4. Autocorrelation (AR)

Lag-1 AR was computed using a 14-day moving window with variables detrended via local polynomial regression (Ludwig et al., 2024). The fitted values from this regression were subtracted from the observed values to obtain detrended values, which were used in subsequent analyses. Missing values were not imputed. If fewer than seven valid (k, $k-1$) pairs were available within the 14-day window, the AR for that day (day 14) was set missing.

2.4.5. Variance

Variance was computed using a 14-day moving window, following the same approach as the AR calculation. Specifically, for each day k, the variance was calculated from data covering the previous 14 days (k-13 to k). Missing values were not imputed, and if fewer than seven complete data points were available, the variance for that day was set to missing. To address skewness in the variance distribution, we applied the natural logarithm (ln) to all variance values. Since the logarithm is undefined for zero, we first added + 1 to each variance before transformation. Because the range of the latent variables had values below 1 (range variance communication: 0.0–0.749; active: 0.0–0.812; sleep: 0.0–0.211), the resulting odds ratios (ORs) were misleadingly high, as they reflected changes based on a full-unit increase in variance. To obtain more interpretable ORs, we multiplied the variances of the latent variables by 10, making ORs correspond to 0.1-unit increase instead.

2.4.6. Moving averages

In addition to analyzing the dynamics of the measures, we aimed to incorporate information on their absolute levels, as it has been hypothesized, depending on the system under investigation, EWS might not add much predictive value above other indicators such as the mean (Dablander et al., 2022). To ensure compatibility with AR and variance, we computed moving averages of the raw values, using an identical 14-day moving window. Specifically, for each day k, the moving average was calculated as the arithmetic mean of the observed values from the preceding 13 days (k-13 to k), including the current day. If there were less than 7 valid entries in the 14-day period, the moving average was set to missing for that day. This was repeated for each day, days with seven or more valid entries were computed, those with less set missing.

2.4.7. Receiver operating characteristics-curves

Receiver operating characteristics (ROC) curves were constructed to assess predictive performance by systematically varying the alarm threshold for each variable, stratified by mood polarity and pre-episode week.

2.5. Analyses

To account for the nested structure of the variables, generalized linear mixed models, in particular multilevel logit models, were fitted for AR, variance and average of each digital phenotyping parameter as the independent variable (fixed main effect) and mood status (early/late pre-episode week vs. euthymic) as the binary dependent variable. This approach allowed us to focus specifically on within-subject effects by comparing mood status (early/late pre-episode weeks vs. euthymic) within each individual. This strategy helps mitigate the impact of inter-individual heterogeneity, even though participants contributed varying numbers of episodes (some with multiple episodes, others with none, contributing only euthymic days).

To account for α-inflation (20 variables × 2 weeks of comparisons x 3 methods × 2 episode polarities), we applied Bonferroni-Holm corrections. For our main analyses, we applied 6 corrections for AR, variance and moving averages, separated for depression and mania, adjusting for 40 parameters each time (20 variables × 2 weeks). For comparison we computed another stricter correction, adjusting for all depressive and all manic parameters together (120 per polarity), these results can be found in the supplementary materials (Table S1) alongside the uncorrected results (Table S2). Analyses were conducted using the Julia programming language (Bezanson et al., 2017).

3. Results

3.1. Multilevel models

Table 2 summarizes results for early warning signals (EWS), moving averages, and latent factors differentiating early (−14 to −8 days) and late (−7 to −1 days) pre-episode weeks of depressive and manic episodes from euthymia. We analyze these predictors separately by episode type to identify polarity-specific patterns.

3.1.1. Depression

3.1.1.1. Predicting Early and Late Pre-Episode weeks of Depression with EWS (AR/variance). Table 2 summarizes the results of EWS in smartphone-based parameters for pre-episode detection. The first four columns represent the results for early and late pre-episode weeks for AR and variance respectively.

The *AR* of four and the variance of six variables significantly differentiated being in a pre-episode week vs. being euthymic after correcting for multiple testing. Specifically, the AR of 'outgoing not reached calls', 'unique conversation partners', 'minutes display on' and 'time got up' all significantly predicted pre-episode weeks. Notably, all effects were positive, indicated by a positive unstandardized regression coefficient b, meaning increased AR was associated with a higher likelihood of being in a pre-episode week. However, no parameter showed significant effects for both, early and late pre-episode weeks. *Variance* of 'Minutes in vehicle', 'kilometers slow', 'outgoing calls', 'hours asleep', 'time got up', and 'minutes phone inactive' significantly differentiated pre-episode weeks. The variance of 'hours asleep' predicted early as well as late pre-episode weeks. All observed effects were positive, except for the variance of 'minutes in vehicle', which was significant, but negative.

These findings indicate that EWS based on variance may be more

Table 2
Results of the multilevel logit models comparing autocorrelation, variance and moving averages as predictors for being in a pre-episode week vs. being euthymic, adjusted for multiple testing.

		Depression vs. Euthymia						Mania vs. Euthymia					
		Autocorrelation		Variance		Moving Averages		Autocorrelation		Variance		Moving Averages	
		Early pre-episode week	Late pre-episode week	Early pre-episode week	Late pre-episode week	Early pre-episode week	Late pre-episode week	Early pre-episode week	Late pre-episode week	Early pre-episode week	Late pre-episode week	Early pre-episode week	Late pre-episode week
		b/OR/ p-value	b/OR/ p-value	b/OR/ p-value	b/OR/ p-value	b/OR/ p-value	b/OR/ p-value	b/OR/ p-value	b/OR/ p-value	b/OR/ p-value	b/OR/ p-value	b/OR/ p-value	b/OR/ p-value
Activity	Steps	0.09 1.01 1.0	-0.53 0.59 1.0	-0.04 0.96 1.0	0.07 1.07 1.0	0.00 1.00 0.070	0.00 1.00 0.119	0.53 1.70 1.0	**1.30** **3.69** **0.028**	**-0.17** **0.84** **0.003**	0.05 1.05 1.0	**0.00** **1.00** **<0.001**	**0.00** **1.00** **<0.001**
	Minutes in vehicle	0.20 1.23 1.0	0.60 1.82 0.741	**-0.15** **0.86** **0.028**	-0.02 0.98 1.0	0.00 1.00 1.0	0.00 1.00 1.0	0.00 1.00 0.992	-0.21 0.81 1.0	**-0.22** **0.80** **0.021**	0.03 1.03 1.0	-0.01 0.99 1.0	0.00 1.00 1.0
	Minutes on foot	0.32 1.37 1.0	-0.54 0.58 1.0	-0.04 0.96 1.0	0.16 1.17 0.618	0.01 1.01 0.089	**0.02** **1.02** **0.043**	0.34 1.41 1.0	1.00 2.71 0.482	0.23 0.79 0.193	0.20 0.82 1.0	**0.03** **1.03** **<0.001**	0.01 1.00 1.0
	Minutes still	-0.08 0.92 1.0	0.27 1.31 1.0	-0.01 0.99 0.869	0.07 1.07 1.0	0.00 1.00 1.0	0.00 1.00 1.0	0.66 1.93 1.0	**1.39** **4.02** **0.012**	0.09 1.09 1.0	0.14 1.15 1.0	**0.00** **1.00** **<0.001**	0.00 1.00 1.0
	Kilometers total	0.06 1.06 1.0	0.82 2.26 0.338	-0.09 0.92 1.0	-0.02 0.98 1.0	0.01 1.01 0.207	0.01 1.01 0.087	0.17 1.18 1.0	-0.17 0.84 1.0	**0.22** **1.25** **0.001**	**0.24** **1.27** **0.001**	0.01 1.01 0.996	0.01 1.01 0.284
	Kilometers fast	0.17 1.18 1.0	0.83 2.30 0.248	-0.13 0.88 0.298	-0.05 0.95 1.0	0.01 1.01 0.943	0.01 1.01 0.773	0.23 1.26 1.0	0.07 1.07 1.0	0.14 1.15 0.213	**0.21** **1.24** **0.007**	0.00 1.00 1.0	0.01 1.01 0.521
	Kilometers slow	-0.22 0.80 1.0	0.33 1.39 1.0	0.14 1.15 1.0	**0.35** **1.41** **0.002**	**0.07** **1.07** **<0.001**	**0.12** **1.13** **<0.001**	0.86 2.37 0.843	0.12 1.13 1.0	**0.39** **1.48** **<0.001**	0.32 1.37 0.057	**0.15** **1.16** **<0.001**	**0.13** **1.14** **<0.001**
Communication	Outgoing Calls	0.20 1.22 1.0	-0.46 0.63 1.0	0.37 1.45 0.642	**0.81** **2.26** **<0.001**	0.16 1.17 1.0	0.23 1.26 1.0	0.21 1.23 1.0	0.71 2.03 1.0	0.18 1.19 1.0	0.09 1.09 1.0	0.06 1.06 1.0	0.13 1.14 1.0
	Incoming missed calls	0.49 1.63 1.0	0.21 1.23 1.0	0.39 1.48 0.429	0.17 1.18 1.0	-0.05 0.95 1.0	-0.08 0.92 1.0	-0.85 0.43 1.0	-0.84 0.43 1.0	-0.02 0.98 1.0	**0.70** **2.01** **0.022**	0.07 1.08 1.0	0.36 1.37 1.0
	Outgoing not reached calls	**1.01** **3.0** **<0.001**	-0.02 0.98 0.928	-0.29 0.75 1.0	-0.54 0.58 0.116	-0.37 0.69 0.788	-0.60 0.55 0.053	-0.98 0.38 0.157	0.40 1.50 1.0	0.21 1.23 1.0	**0.53** **1.70** **0.032**	0.05 1.05 1.0	0.31 0.98 1.0
	Minutes Call Duration	0.22 1.24 1.0	-0.61 0.55 1.0	0.11 1.12 1.0	0.21 1.23 0.073	-0.01 0.99 1.0	0.01 1.01 1.0	-0.79 0.45 0.983	0.55 1.74 1.0	0.13 1.14 1.0	-0.03 0.97 1.0	-0.03 0.97 1.0	-0.02 0.98 1.0
	Unique conv. partners	**1.22** **3.39** **0.003**	0.29 1.34 1.0	0.12 1.13 1.0	0.38 1.46 0.466	-0.03 0.97 1.0	-0.05 0.95 1.0	-0.30 0.74 1.0	-0.17 0.84 1.0	0.06 1.07 1.0	-0.32 0.73 1.0	0.19 1.22 1.0	0.16 1.18 1.0
Sleep	Hours asleep	-0.13 0.88 1.0	-0.85 0.43 0.255	**0.89** **2.43** **<0.001**	**0.69** **2.00** **0.009**	**0.49** **1.64** **<0.001**	0.35 1.43 0.066	-0.41 0.66 1.0	0.45 1.56 1.0	**0.78** **2.18** **0.009**	-0.31 0.73 1.0	0.03 1.03 1.0	-0.07 0.93 1.0
	Time got up	-0.53 0.59 1.0	**-1.39** **0.25** **0.001**	0.31 1.37 1.0	**0.87** **2.39** **<0.001**	-0.02 0.98 0.816	0.09 1.09 1.0	**1.36** **3.89** **0.044**	**1.87** **6.50** **0.001**	-0.69 0.50 0.193	**-1.32** **0.27** **<0.001**	0.24 1.28 1.0	0.12 1.13 1.0
Phone Use	Count display on	0.18 1.20 1.0	-0.35 0.70 1.0	-0.03 0.97 1.0	-0.09 0.91 1.0	-0.02 0.98 0.091	0.00 1.00 1.0	-1.35 0.26 0.069	-1.49 0.23 0.058	-0.15 0.86 1.0	-0.06 0.94 1.0	0.00 1.00 1.0	0.03 1.03 0.118
	Minutes Display on	**1.11** **3.02** **0.002**	0.33 1.39 1.0	0.03 1.03 1.0	0.10 1.10 0.394	0.00 1.00 1.0	**0.01** **1.01** **0.013**	0.29 1.34 1.0	-0.76 0.47 1.0	0.10 1.11 1.0	-0.05 0.95 1.0	0.00 1.00 1.0	0.00 1.00 1.0
	Minutes Phone Inactive	-0.79 0.46 0.200	-0.43 0.65 1.0	0.16 1.18 0.074	**0.23** **1.26** **0.007**	0.00 1.00 0.085	**0.00** **1.00** **0.002**	0.86 2.35 0.559	0.61 1.84 1.0	-0.09 0.91 1.0	0.08 1.08 1.0	0.00 1.00 1.0	0.00 1.00 1.0
Latent Variables	Communication	**1.05** **2.87** **0.001**	-0.42 0.66 1.0	**0.39** **1.47** **0.004**	**0.52** **1.68** **<0.001**	1.36 3.90 0.206	1.68 5.36 0.065	-0.72 0.49 1.0	-0.62 0.54 1.0	0.03 1.03 1.0	0.00 1.0 0.986	**2.20** **9.00** **<0.001**	**2.18** **8.85** **0.003**
	Active	-0.16 0.85 1.0	0.18 1.20 1.0	-0.12 0.89 1.0	**0.25** **1.29** **0.041**	-0.27 0.76 1.0	0.12 1.13 1.0	-0.61 0.54 1.0	-0.15 0.86 1.0	**-0.31** **0.73** **0.035**	-0.06 0.94 1.0	0.02 1.02 0.951	-1.13 0.32 0.305
	Sleep	0.17 1.18 1.0	-0.61 0.54 0.074	0.13 1.14 0.066	**0.26** **1.29** **<0.001**	0.44 1.56 1.0	0.43 1.54 1.0	0.49 1.64 1.0	0.35 0.41 1.0	0.05 1.05 1.0	-0.16 0.85 0.413	0.53 1.70 1.0	0.18 1.19 1.0

b: Estimated unstandardized regression coefficients. OR: Odds ratio. P-values are adjusted for multiple testing within each affective polarity (depression/mania) and for autocorrelation, variance and mean values separately via Bonferroni-Holm correction, this means there were six calculations for 40 variables each (e.g. autocorrelation in depressive pre-episode weeks, variance in depressive pre-episode weeks). P-values < 0.05 are highlighted in bold and shaded grey. Latent variables Communication, Active, and Sleep were derived via structural equation modeling. Sleep was assessed via e-diary.

promising than those based on AR for identifying pre-episode depressive states. However, only the variance of 'hours asleep' consistently differentiated both pre-episode weeks from euthymia.

3.1.1.2. Predicting Early and Late Pre-Episode weeks of Depression with Moving Averages. Differentiating between the content-based information of specific variables and their temporal dynamics in the sense of dynamical systems theory, we also examined the effects of the moving averages of the parameters. Results for the prediction of both pre-episode weeks are found in the two columns next to AR and variance in Table 2. Moving averages of five variables, 'minutes on foot', 'kilometers slow', 'hours asleep', 'minutes display on', 'minutes phone inactive' significantly differentiated being in a pre-episode week vs. being euthymic. Interestingly, all of the effects of these five predictors were positive. Clinically, this translates to patients exhibiting increased walking, more sleep and altered phone use behavior in pre-episode weeks when compared to euthymia. The mean value of kilometers slow predicted early and late pre-episode weeks, while all others only significantly differentiated one of the pre-episode weeks.

3.1.1.3. Performance of latent factor parameters (AR, variance and average) predicting Early and Late Pre-Episode Weeks of Depression. Results for the three latent variables can be found on the left-hand side in the last three rows of Table 2. In detail, AR of communication predicted early pre-episode weeks, while variance of communication predicted early as well as late pre-episode weeks. Moreover, variance of active and sleep significantly predicted late pre-episode weeks. All observed significant effects were positive, thus complying with dynamical systems theory.

These results indicate that latent parameters, especially the *variance* of the latent parameters, show promise in identifying pre-episode depressive weeks.

3.1.2. Mania

3.1.2.1. Predicting Early and Late Pre-Episode Weeks of Mania with EWS (AR/variance). Results for predicting manic pre-episode weeks using EWS are presented in four columns on the right-hand side of Table 2. The *AR* of the parameters 'steps', 'minutes still' and 'time got up' significantly predicted pre-episode weeks compared to euthymic days. Among these, the AR of 'time got up' emerged as the most consistent predictor, distinguishing both early and late pre-episode weeks from euthymia. All significant effects were associated with positive coefficients (b), indicating that higher AR values for these parameters correspond to an increased likelihood of being in a manic pre-episode week. The *variance* of 'steps', 'minutes in vehicle', 'kilometers total', 'kilometers fast', 'kilometers slow', 'incoming missed calls', 'outgoing not reached calls', 'hours asleep' and 'time got up' significantly predicted being in a pre-episode week vs. euthymia. Kilometers total predicted early as well as late pre-episode weeks. The direction of effects differed across predictors. The effects of variance of 'steps', 'minutes in vehicle' and 'time got up' were negative, indicating lower variance corresponds to a higher likelihood of being in a manic pre-episode week. This means patients had more consistent walking, driving and time got up in the weeks prior to emerging manic episodes. This contradicts dynamical systems theory, which posits increased variance before critical transitions.

In sum, significant results for pre-episode weeks of mania clustered around activity parameters. In contrast to depression, several predictors for mania showed reversed effects (i.e., lower variance before episodes).

3.1.2.2. Predicting Early and Late Pre-Episode Weeks of Mania with Moving Averages. The final two columns show the results for moving averages of all variables to differentiate pre-episode manic weeks (Table 2). The moving average of four variables, 'steps', 'minutes on foot', 'minutes still', and 'kilometers slow' distinguished between pre-episode weeks of mania and euthymia. Notably, all of those are activity parameters and all observed effects were positive. This translates to increased walking and increased time spent still during pre-episode manic weeks compared to euthymia. 'Steps' and 'kilometers slow' significantly predicted early as well as late pre-episode weeks.

Compared to EWS, moving averages for mania were more selective, identifying fewer predictors but with greater consistency across both pre-episode weeks, particularly for physical activity features.

3.1.2.3. Performance of latent factor parameters (AR, variance and average) predicting Early and Late Pre-Episode Weeks of Mania. Results for the latent variables predicting pre-episode manic weeks are shown in the last three rows in the right-hand side of Table 2. Of the three latent parameters, only the variance of 'active' and the moving averages of 'communication' significantly differentiated pre-episode weeks from euthymia. However, the effect of the variance of active was negative, contradicting the concept of critical slowing down. The moving averages of communication predicted early as well as late prodromal weeks and was positive.

Taken together, latent predictors for mania yielded fewer significant effects and more inconsistencies with theoretical expectations, limiting their potential compared to latent variables before depressive episode prediction.

3.2. Predictive performance

ROC curves (Fig. 2) assessed the clinical utility of latent variable predictors. The x-axis represents the false positive rate (1-specificity), while the y-axis represents the true positive rate (sensitivity). A diagonal 45-degree line would indicate chance-level prediction (AUC = 0.5), where sensitivity equals the false positive rate, meaning that any predictor falling along this line performs no better than random guessing.

Following previous work (Ludwig et al., 2025), we define a target area for clinical applicability: a sensitivity of at least 50 %, ensuring detection of at least every second emerging episode, and a specificity of at least 95 %, minimizing false alarms. In our sample, this specificity roughly translated to one false alarm per month, though this estimate depends on episode prevalence and overall predictive accuracy and may not generalize to other clinical settings. We illustrate this target area as a shaded region in Fig. 2. Results for nine different models – using AR, Variance and moving average for each of the three latent parameters to predict pre-episode week vs. euthymia - are presented in Fig. 2 stratified by pre-episode week. The AUCs of the models range from 0.852 to 0.865 for early pre-episode depression, from 0.853 to 0.87 for late pre-episode depression, from 0.859 to 0.878 for early pre-episode mania, and from 0.867 to 0.881 for late pre-episode mania. Notably, the models include random intercepts to mirror the results of the multilevel analyses. This improves fit but also means the models are not truly predictive.

While all predictors perform above chance (AUC > 0.5), none reach the clinical utility threshold (i.e., the target area). However, several parameters come close in the late pre-episode depressive weeks and early pre-episode manic weeks, making this setting the most promising. Additionally, predictions for late pre-episode weeks of depression appear more promising than those for early pre-episode weeks of depression or any pre-episode weeks of mania.

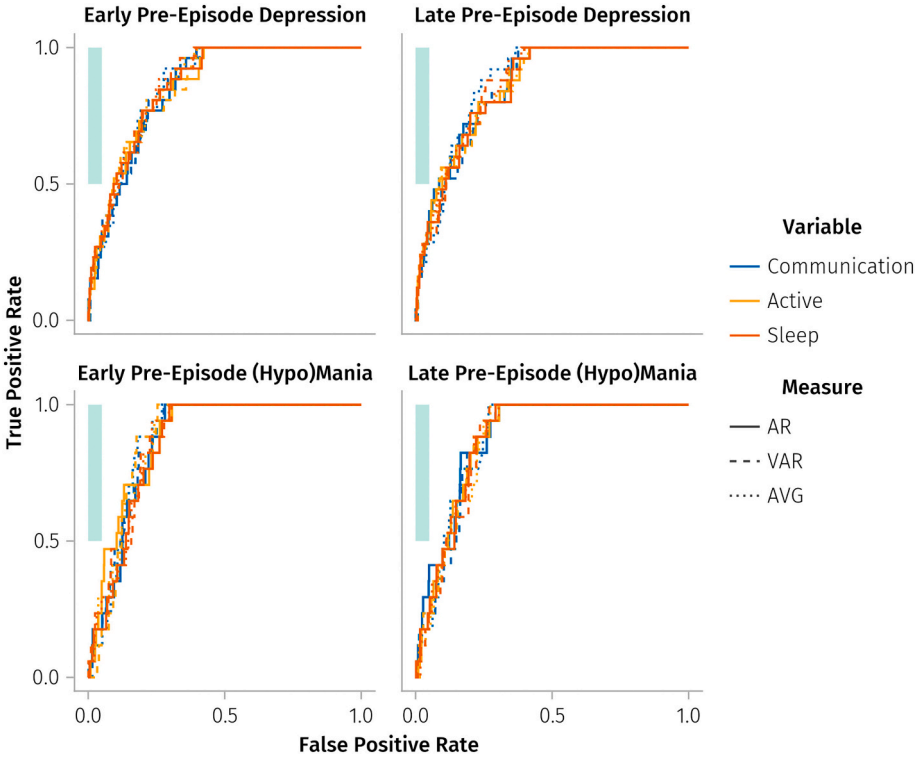


Fig. 2. Summarized results of individual models plotting true and false positive rates of autocorrelation (AR), variance (VAR) and moving average (AVG) of the three latent parameters (active, communication, sleep) for pre-episode prediction of depression and mania. Shaded in green is our pre-defined target area (Sensitivity >50 %, Specificity >95 %) required for clinical application. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

4.1. Main results

To our knowledge, this is the first study to investigate whether indicators of critical slowing down, specifically increased AR or variance, occur in smartphone-based digital phenotyping prior to affective episodes in individuals diagnosed with BD, with the goal of enhancing early detection of prodromal symptoms. Several parameters significantly differentiated pre-episode weeks from euthymic days. However, rather than a uniform effect, significant predictors were scattered across different parameters (e.g. steps, time got up), time frames (early vs. late pre-episode weeks), and models (AR, variance, mean values). In detail, most predictors identified either early or late pre-episode weeks, not both. Although false positives were statistically adjusted for, the lack of convergence across predictors, time frames and models suggest that isolated effects may not be robust. We therefor propose that focusing on predictive patterns is more informative than discussing single predictors. In particular, three promising patterns emerged: activity-related parameters for mania prediction, EWS in sleep, and variance in latent parameters.

Activity-related parameters were strong predictors of manic pre-episode weeks. Increased AR of steps and minutes still indicated slower dynamics and more similar walking or resting time between successive days before the onset of mania. The significantly increased variance of kilometers travelled before manic episodes represents more extreme (lower and higher) distances travelled compared to euthymic periods, which complies with critical slowing down. However, the variance of steps and minutes in vehicle is significantly *reduced* before manic episodes, translating to lower fluctuations of daily steps or minutes spent in a vehicle on successive days, thereby contradicting dynamical systems theory. This reduced variance before manic episodes may indicate a different pattern of behavioural fluctuations in the pre-episode weeks, which has been reported in other studies on EWS in BD, where changes in AR and variance can occur in both directions before manic episodes (Kunkels et al., 2021; Ludwig et al., 2024). Increased moving averages of activity were significantly associated with being in a manic pre-episode week compared to euthymic days. Notably, the moving averages of "kilometers slow" (i.e. distance travelled at less than 20 km/h typically reflecting walking, running or moderate cycling) emerged as the most consistent predictor, significantly associated with *both manic and depressive pre-episode weeks*, always in the direction of increased activity (positive effects only). While this pattern may seem contradictory given the well-established link between *reduced* walking and depressive symptoms (Bizzozero-Peroni et al., 2024; Ludwig et al., 2018), we emphasize that our focus is on the pre-episode phase where atypical changes in behavior may reflect increased external demands or disruptions (e.g. work travel, moving house or social obligations). These disruptions, which may involve more physical activity than usual, could function as stressors that contribute to recurrences, regardless of episode polarity. However, as we lack contextual information about the nature of these movements, this interpretation remains speculative and should be considered with caution.

The second pattern were EWS in sleep-variables for the prediction of mood shifts in any direction. The AR and variance of the two variables hours asleep and time got up significantly differentiated several early and late pre-episode weeks in depression as well as mania. This indicates that sleeping patterns become more rigid (increased AR), but at the same time more extreme in their expression (increased variance), before mood shifts, consistent with critical slowing down. However, moving averages did not show these effects. Importantly, these variables were derived from e-diaries rather than passive sensing, highlighting the need for future studies to extract sleep data from passive sensing parameters to determine whether subjective sleep reports or actual sleep patterns are more predictive.

Lastly, the three latent variables (communication, active, sleep) were

included with the expectation of reducing error variance compared to individual variables and showed promise in differentiating depressive pre-episode weeks. The variance of all three parameters predicted late (active, sleep) or both (communication) depressive pre-episode weeks. Of the three variables, communication not only predicted both depressive pre-episode weeks (variance) and early depressive pre-episode weeks (AR), but also both manic pre-episode weeks (mean). Clinically this manifests in more rigid (increased AR) and extreme communication behavior (increased variance), as well as more extreme manifestation of activity and sleep parameters (increased variances) before depressive episodes. Interestingly, none of the mean values of the individual markers of communication were related to manic pre-episode weeks, whereas the mean of the latent communication variable was very strongly related to the occurrence of manic pre-episode weeks. Possibly, rather than the increase in a specific communication behavior (e.g., more outgoing call *or* longer call durations), the concomitant increase in different communicative behaviors (e.g., more outgoing call *and* longer call durations) predict the occurrence of a manic episode. This pattern emphasizes the complexities in the association between digital phenotyping and clinical outcomes and it emphasizes the need to target this issue using multivariate approaches such as the latent variable approach pursued in the present work.

Smartphone-based digital phenotyping shows definite potential in differentiating pre-episode weeks from euthymic days in BD. If replicated, these findings could transform BD patient care by facilitating earlier interventions. However, ROC curve analyses showed that while predictions were above-chance, even the most promising predictors, the latent variables, did not meet our proposed clinical thresholds.

4.2. Strengths and limitations

While the number of participants in this study was relatively small, the dataset itself represents a key strength and compares favorably to existing studies in this field. To date, this study included the most emerging affective episodes in BD patients investigating indicators of critical slowing down. Moreover, the 'ground truth' assessment is outstanding, with bi-weekly clinical expert interviews categorically rating the psychopathological status on a day-level over one year. Additionally, effects were found in early and late pre-episode weeks. Taking into account that every day represents aggregated data of the previous 14 days, our analyses cover 28 days preceding episode onset. This is the period of time typically reported for prodromal sub-threshold symptoms before relapse (Jackson et al., 2003; Mantere et al., 2008; Van Meter et al., 2016) and corresponds well to previous findings. However, it is important to acknowledge that the results are based on only 30 depressive and 20 manic episodes in 22 patients, which limits statistical power, particularly for mania, and constrains the generalizability of our findings.

Our approach deliberately limited the number of derived smartphone-parameters to minimize alpha-error inflation by leveraging established variables from previous research on this dataset (Ebner-Priemer et al., 2020) and the parameters of the follow-up study BipoLife A3 (Mühlbauer et al., 2018). This conservative strategy limited false positives, but may have missed more suitable indicators. Also, we did not include potentially relevant clinical variables as covariates, such as lifetime number of episodes or subthreshold symptoms. Future studies could use pure data driven machine-learning approaches to identify optimal variables, unlocking the full potential of smartphone-based data (Sameh et al., 2025), or further investigate these data including clinical predictors. Additionally, we did not pre-register our data analyses, as this study was an exploratory investigation using an existing data set. Accordingly, replication of our findings in an independent and larger dataset is warranted. We acknowledge that clinical factors such as medication use or specific comorbidities may influence smartphone-based measures. However, due to limited quality and availability of continuous time-series data on these variables, controlling for them was not feasible in this study and remains a limitation.

Applying dynamical systems theory to psychopathology is challenging. While EWS are often assumed to be generic and universal, recent studies challenge this assumption (Dablander et al., 2022; Scheffer et al., 2024a, 2024b). Their detection relies on identifying a genuine critical transition- and the correct type – based on underlying processes governing the change (Helmich et al., 2024). This is particularly difficult in psychiatry, where exact relapse mechanisms remain unclear, increasing the risk of misinterpreting or misapplying EWS (Dakos and Carpenter, 2015; Helmich et al., 2024). Proper identification also depends on aligning the variables and the temporal scale with the systems dynamics, as misalignment can produce false negatives or positives. Furthermore, the performance of EWS indicators such as autocorrelation and variance can be affected by stochastic fluctuations (O'Regan and Burton, 2018). While our multilevel framework accounts for some sources of variation, we acknowledge that stochasticity may still influence these patterns. To support the assumption of stationarity, we detrended our time series using local polynomial regression, following established procedures in the EWS literature.

Despite these challenges, dismissing dynamical systems theory for relapse prediction in affective disorders seems premature. Our high-quality dataset provides the strongest foundation to detect signals of critical slowing down, if present. In the face of highly complex, poorly understood systems (humans) and the potentially transformative impact of episode prediction, a data-driven, inductive approach with strict clinical quality criteria may help determine whether this framework can be a valuable tool for managing BD. However, our findings also underscore that without clearly defined boundary conditions, the application of DST to psychological data remains theoretically ambiguous. Observing reduced autocorrelation or variance before transitions, as in our data, questions whether the system undergoes a genuine critical transition at all — or whether different mechanisms underlie relapse. This aligns with recent calls for caution in interpreting EWS patterns as universal signals of instability. Future studies should not only replicate EWS empirically, but also refine the theoretical models guiding when, how, and why DST applies in psychopathology.

Lastly, latent variables were derived based on an iterative, data-driven strategy (Ebner-Priemer et al., 2020). While the procedure was also informed by conceptual and theoretical considerations, an a-priori theoretical definition of latent factors would be preferable. This approach was necessary due to the vast number of possible smartphone-parameters (Ebner-Priemer and Santangelo, 2020) and the lack of clear theoretical models for synthesizing them. Future research should refine our theoretical understanding of mobile sensing parameters and develop more robust latent measurement models for smartphone-based phenotyping.

4.3. Future directions

EWS in smartphone-based parameters show promise, but fully realizing their potential requires refining parameters derived from raw data and optimizing analytical frameworks for psychological data. Parameter improvement could be driven by data-based machine learning algorithms (Sameh et al., 2025) or by deriving smarter digital phenotypes that better reflect psychopathology (Wadle and Ebner-Priemer, 2023). E.g. the DSM-criterion increased 'goal-directed behaviour or energy' for the diagnosis of mania is probably not optimally captured by the parameters along the lines of 'daily kilometers travelled', and more fitting smartphone sensing parameters might be developed and used as proxies. Also, focusing on sleep and investigating EWS in passively sensed parameters of sleep warrant further investigation. While dynamical systems theory remains a valuable approach for BD, alternative, truly prospective frameworks such as statistical process control (Ludwig et al., 2025) should also be explored. The ultimate goal is to identify well-aligned smartphone-based passive sensing parameters and the most effective analytical frameworks to enhance early relapse prediction in

BD.

5. Conclusions

Our analysis of smartphone-based digital phenotyping identified several predictors that differentiated pre-episode weeks from euthymic days in BD. However, emphasizing promising patterns over isolated predictors appears most informative. While activity and sleep parameters as well as latent variables show some promise for early detection, none of the predictors have yet met our clinical thresholds. Further research should refine these measures, establish the optimal analytic framework, and replicate findings to improve relapse prediction in BD.

CRediT authorship contribution statement

Vera M. Ludwig: Writing – original draft, Data curation, Conceptualization. **Carl A. Bittendorf:** Writing – review & editing, Visualization, Methodology, Data curation. **Iris Reinhard:** Writing – review & editing, Methodology, Formal analysis. **Andreas B. Neubauer:** Writing – review & editing, Methodology, Formal analysis. **Eva Mennigen:** Writing – review & editing, Supervision. **Esther Mühlbauer:** Writing – review & editing, Project administration, Investigation, Data curation. **Wolfram E. Severus:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. **Michael Bauer:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Ulrich W. Ebner-Priemer:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Michael Bauer reports financial support was provided by German Research Foundation. Michael Bauer reports a relationship with Alfred E Tiefenbacher GmbH und Co KG that includes: consulting or advisory. Michael Bauer reports a relationship with Compass Pathfinder Limited that includes: consulting or advisory. Michael Bauer reports a relationship with GH Research that includes: consulting or advisory. Michael Bauer reports a relationship with MedLink International Inc. that includes: consulting or advisory. Michael Bauer reports a relationship with Janssen Global Services LLC that includes: consulting or advisory. Michael Bauer reports a relationship with LivaNova USA, Inc. that includes: consulting or advisory. Michael Bauer reports a relationship with Novartis Pharma Switzerland that includes: consulting or advisory. Michael Bauer reports a relationship with Sunovion Respiratory Development Inc. that includes: consulting or advisory. Michael Bauer reports a relationship with MedTriX GmbH that includes: speaking and lecture fees. Michael Bauer reports a relationship with Streamedup GmbH that includes: speaking and lecture fees. Ulrich Ebner-Priemer reports a relationship with Boehringer Ingelheim GmbH that includes: consulting or advisory. Ulrich Ebner-Priemer reports a relationship with Angelini Pharma Germany GmbH that includes: speaking and lecture fees. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

MB has received competitive grant support by Deutsche Forschungsgemeinschaft (DFG), Bundesministerium für Bildung und Forschung (BMBF), European Commission, Sächsische Aufbaubank (SAB), and served as an advisor to Alfred E. Tiefenbacher GmbH Co. KG, COMPASS Pathfinder Ltd., GH Research, MedEd-Link Inc. Janssen Global Services, LLC, Livanova, Novartis Switzerland, Sunovion, and has received lecture fees from MedTrix GmbH, Streamedup GmbH. UE-P reports consultancy for Boehringer-Ingelheim and speaker honorarium from Angelini Pharma. The author asserts that the mentioned companies had no influence over the content of this article. VML, CAB, IR, ABN, EM, EM, and WES report no potential conflict of interest.

Acknowledgement

Research reported in this publication was supported in part by two grants from the German Research Foundation/Deutsche Forschungsgemeinschaft (DFG Project number 521379614 [CRC 393 “Trajectories of affective disorders: cognitive-emotional mechanisms of symptom change”] and project number GRK 2773/1 [IRTG 2773 “Risk Factors and Pathomechanisms of Affective Disorders”]).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2025.119972.

References

Bauer, M., Grof, P., Gyulai, L., Rasgon, N., Glenn, T., Whybrow, P.C., 2004. Using technology to improve longitudinal studies: self-reporting with Chrono Record in bipolar disorder. Bipolar Disord. 6 (1), 67–74. https://doi.org/10.1046/j.1399-5618.2003.00085.x.

Bauer, M., Wilson, T., Neuhaus, K., Sasse, J., Pfennig, A., Lewitzka, U., Grof, P., Glenn, T., Rasgon, N., Bschor, T., Whybrow, P.C., 2008. Self-reporting software for bipolar disorder: validation of ChronoRecord by patients with mania. Psychiatry Res. 159 (3), 359–366. https://doi.org/10.1016/j.psychres.2007.04.013.

Bayani, A., Hadaeghi, F., Jafari, S., Murray, G., 2017. Critical slowing down as an early warning of transitions in episodes of bipolar disorder: a simulation study based on a computational model of circadian activity rhythms. Chronobiol. Int. 34 (2), 235–245. https://doi.org/10.1080/07420528.2016.1272608.

Bech, P., Bolwig, T.G., Kramp, P., Rafaelsen, O.J., 1979. The Bech-Rafaelsen Mania Scale and the Hamilton Depression Scale. Acta Psychiatr. Scand. 59 (4), 420–430. https://doi.org/10.1111/j.1600-0447.1979.tb04484.x.

Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B., 2017. Julia: a fresh approach to numerical computing. SIAM Rev. 59 (1), 65–98. https://doi.org/10.1137/141000671.

Bizzozero-Peroni, B., Díaz-Goñi, V., Jiménez-López, E., Rodríguez-Gutiérrez, E., Sequí-Domínguez, I., Núñez De Arenas-Arroyo, S., López-Gil, J.F., Martínez-Vizcaíno, V., Mesas, A.E., 2024. Daily step count and depression in adults: a systematic review and meta-analysis. JAMA Netw. Open 7 (12). https://doi.org/10.1001/jamanetworkopen.2024.51208.

Bos, F.M., Schreuder, M.J., George, S.V., Doornbos, B., Bruggeman, R., van der Krieke, L., Haarman, B.C.M., Wichers, M., Snippe, E., 2022. Anticipating manic and depressive transitions in patients with bipolar disorder using early warning signals. Int. J. Bipolar Disord. 10 (1), 12. https://doi.org/10.1186/s40345-022-00258-4.

Brietzke, E., Hawken, E.R., Idzikowski, M., Pong, J., Kennedy, S.H., Soares, C.N., 2019. Integrating digital phenotyping in clinical characterization of individuals with mood disorders. Neurosci. Biobehav. Rev. 104, 223–230. https://doi.org/10.1016/j.neubiorev.2019.07.009.

Csikszentmihalyi, M., Larson, R., 1987. Validity and reliability of the experience-sampling method. J. Nerv. Ment. Dis. 175 (9), 526–536. https://doi.org/10.1097/00005053-198709000-00004.

Curtiss, J., Fulford, D., Hofmann, S.G., Gershon, A., 2019. Network dynamics of positive and negative affect in bipolar disorder. J. Affect. Disord. 249, 270–277. https://doi.org/10.1016/j.jad.2019.02.017.

Dablander, F., Pichler, A., Cika, A., Bacilieri, A., 2022. Anticipating critical transitions in psychological systems using early warning signals: theoretical and practical considerations. Psychol. Methods. https://doi.org/10.1037/met0000450.

Dakos, V., Carpenter, S.R., 2015. Resilience indicators:prospects and limitations for early warnings ofregime shifts. Phil. Trans. R. Soc. B 370, 20130263. https://doi.org/10.1098/rstb.2013.0263.

Dakos, V., Carpenter, S.R., Brock, W.A., Ellison, A.M., Guttal, V., Ives, A.R., Kéfi, S., Livina, V., Seekell, D.A., van Nes, E.H., Scheffer, M., 2012. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. PLoS ONE 7 (7). https://doi.org/10.1371/journal.pone.0041010.

Dakos, V., Boulton, C.A., Buxton, J.E., Abrams, J.F., Armstrong McKay, D.I., Bathiany, S., Blaschke, L., Boers, N., Dylewsky, D., López-Martínez, C., Parry, I., Ritchie, P., van der Bolt, B., van der Laan, L., Weinans, E., Kéfi, S., 2023. Tipping Point Detection and Early-warnings in Climate, Ecological, and Human Systems. https://doi.org/10.5194/egusphere-2023-1773.

de Assis da Silva, R., Mograbi, D.C., Silveira, L.A.S., Nunes, A.L.S., Novis, F.D., Landeira-Fernandez, J., Cheniaux, E., 2015. Insight across the different mood states of bipolar disorder. Psychiatry Q. 86 (3), 395–405. https://doi.org/10.1007/s11126-015-9340-z.

de Azevedo Cardoso, T., Kochhar, S., Torous, J., Morton, E., 2024. Digital tools to facilitate the detection and treatment of bipolar disorder: key developments and future directions. JMIR Ment. Health 11 (1), e58631. https://doi.org/10.2196/58631.

Ebner-Priemer, U., Santangelo, P., 2020. Digital phenotyping: hype or hope? Lancet Psychiatry 7 (4), 297–299. https://doi.org/10.1016/S2215-0366(19)30380-3.

Ebner-Priemer, U.W., Kubiak, T., Pawlik, K., 2009. Ambulatory assessment. Eur. Psychol. 14 (2), 95–97. https://doi.org/10.1027/1016-9040.14.2.95.

Ebner-Priemer, U., Mühlbauer, E., Neubauer, A.B., Hill, H., Beier, F., Santangelo, P.S., Ritter, P., Kleindienst, N., Bauer, M., Schmiedek, F., Severus, E., 2020. Digital phenotyping: towards replicable findings with comprehensive assessments and integrative models in bipolar disorders. Int. J. Bipolar Disord. 8 (1). https://doi.org/10.1186/s40345-020-00210-4.

First, M.B., Williams, J.B., Karg, R.S., Spitzer, R.L., 2015. User's Guide for the Structured Clinical Interview for DSM-5 Disorders, Research Version (SCID-5-RV). American Psychiatric Association.

Goodwin, G.M., Haddad, P.M., Ferrier, I.N., Aronson, J.K., Barnes, T.R.H., Cipriani, A., Coghill, D.R., Fazel, S., Geddes, J.R., Grunze, H., Holmes, E.A., Howes, O., Hudson, S., Hunt, N., Jones, I., MacMillan, I.C., McAllister-Williams, H., Miklowitz, D.R., Morriss, R., Young, A.H., 2016. Evidence-based guidelines for treating bipolar disorder: revised third edition recommendations from the British Association for Psychopharmacology. J. Psychopharmacol. 30 (6), 495–553. https://doi.org/10.1177/0269881116636545.

Goossens, P.J.J., Kupka, R.W., Beentjes, T.A.A., van Achterberg, T., 2010. Recognising prodromes of manic or depressive recurrence in outpatients with bipolar disorder: a cross-sectional study. Int. J. Nurs. Stud. 47 (10), 1201–1207. https://doi.org/10.1016/j.ijnurstu.2010.01.010.

Helmich, M.A., Schreuder, M.J., Bringmann, L.F., Riese, H., Snippe, E., Smit, A.C., 2024. Slow down and be critical before using early warning signals in psychopathology. Nat. Rev. Psychol. https://doi.org/10.1038/s44159-024-00369-y.

Ikäheimonen, A., Luong, N., Baryshnikov, I., Darst, R., Heikkilä, R., Holmen, J., Martikkala, A., Riihimäki, K., Saleva, O., Isometsä, E., Aledavood, T., 2024. Predicting and monitoring symptoms in patients diagnosed with depression using smartphone data: observational study. J. Med. Internet Res. 26, e56874. https://doi.org/10.2196/56874.

Insel, T.R., 2018. Digital phenotyping: a global tool for psychiatry. In: World Psychiatry, vol. 17 (3). Blackwell Publishing Ltd., pp. 276–277. https://doi.org/10.1002/wps.20550

Jackson, A., Cavanagh, J., Scott, J., 2003. A systematic review of manic and depressive prodromes. J. Affect. Disord. 74 (3), 209–217. https://doi.org/10.1016/S0165-0327(02)00266-5.

Jakobsen, P., Côté-Allard, U., Riegler, M.A., Stabell, L.A., Stautland, A., Nordgreen, T., Torresen, J., Fasmer, O.B., Oedegaard, K.J., 2024. Early warning signals observed in motor activity preceding mood state change in bipolar disorder. Bipolar Disord. 26 (5), 468–478. https://doi.org/10.1111/bdi.13430.

Kunkels, Y.K., Riese, H., Knapen, S.E., Riemersma - van der Lek, R.F., George, S.V., van Roon, A.M., Schoevers, R.A., Wichers, M., 2021. Efficacy of early warning signals and spectral periodicity for predicting transitions in bipolar patients: an actigraphy study. Transl. Psychiatry 11 (1). https://doi.org/10.1038/s41398-021-01465-w.

Ludwig, V.M., Bayley, A., Cook, D.G., Stahl, D., Treasure, J.L., Asthworth, M., Greenough, A., Winkley, K., Bornstein, S.R., Ismail, K., 2018. Association between depressive symptoms and objectively measured daily step count in individuals at high risk of cardiovascular disease in South London, UK: a cross-sectional study. BMJ Open 8 (4). https://doi.org/10.1136/bmjopen-2017-020942.

Ludwig, V.M., Reinhard, I., Mühlbauer, E., Hill, H., Severus, W.E., Bauer, M., Ritter, P., Ebner-Priemer, U.W., 2024. Limited evidence of autocorrelation signaling upcoming affective episodes: a 12-month e-diary study in patients with bipolar disorder. Psychol. Med. 1–9. https://doi.org/10.1017/S0033291723003811.

Ludwig, V.M., Bittendorf, C.A., Reinhard, I., Guth, M., Mühlbauer, E., Hartnagel, L.-M., Severus, W.E., Bauer, M., Ritter, P., Ebner-Priemer, U.W., 2025. Predicting depressive and manic episodes in patients with bipolar disorder using statistical process control methods on passive sensing data. Journal of Psychopathology and Clinical Science. https://doi.org/10.1037/abn0001002.

Mantere, O., Suominen, K., Valtonen, H.M., Arvilommi, P., Isometsä, E., 2008. Only half of bipolar I and II patients report prodromal symptoms. J. Affect. Disord. 111 (2–3), 366–371. https://doi.org/10.1016/j.jad.2008.03.011.

Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. Br. J. Psychiatry 134 (4), 382–389. https://doi.org/10.1192/bjp.134.4.382.

Mühlbauer, E., 2020. Langzeitmonitoring symptomnaher Verhaltensparameter bei Patienten mit bipolaren Erkrankungen. TU Dresden.

Mühlbauer, E., Bauer, M., Ebner-Priemer, U., Ritter, P., Hill, H., Beier, F., Kleindienst, N., Severus, E., 2018. Effectiveness of smartphone-based ambulatory assessment (SBAA-BD) including a predicting system for upcoming episodes in the long-term treatment of patients with bipolar disorders: study protocol for a randomized controlled single-blind trial. BMC Psychiatry 18 (1). https://doi.org/10.1186/s12888-018-1929-y.

Nakamura, T., Kim, J., Sasaki, T., Yamamoto, Y., Takei, K., Taneichi, S., 2013. Intermittent locomotor dynamics and its transitions in bipolar disorder. In: 2013 22nd International Conference on Noise and Fluctuations (ICNF), pp. 1–4. https://doi.org/10.1109/ICNF.2013.6578924.

Nunes, A., Singh, S., Allman, J., Becker, S., Ortiz, A., Trappenberg, T., Alda, M., 2022. A critical evaluation of dynamical systems models of bipolar disorder. Transl. Psychiatry (1). https://doi.org/10.1038/s41398-022-02194-4.

O'Regan, S.M., Burton, D.L., 2018. How stochasticity influences leading indicators of critical transitions. Bull. Math. Biol. 80 (6), 1630–1654. https://doi.org/10.1007/s11538-018-0429-z.

Ortiz, A., Maslej, M.M., Husain, M.I., Daskalakis, Z.J., Mulsant, B.H., 2021. Apps and gaps in bipolar disorder: a systematic review on electronic monitoring for episode prediction. J. Affect. Disord. 295, 1190–1200. https://doi.org/10.1016/j.jad.2021.08.140.

Perich, T., Mitchell, P.B., Loo, C., Hadzi-Pavlovic, D., Roberts, G., Frankland, A., Lau, P., Wright, A., 2013. Clinical and demographic features associated with the detection of early warning signs in bipolar disorder. J. Affect. Disord. 145 (3), 336–340. https://doi.org/10.1016/j.jad.2012.08.014.

Salvi, J.D., Rauch, S.L., Baker, J.T., 2021. Behavior as physiology: how dynamical-systems theory could advance psychiatry. Am. J. Psychiatry 178 (9), 791–792. https://doi.org/10.1176/appi.ajp.2020.20081151.

Sameh, A., Rostami, M., Oussalah, M., Korpelainen, R., Farrahi, V., 2025. Digital phenotypes and digital biomarkers for health and diseases: a systematic review of machine learning approaches utilizing passive non-invasive signals collected via wearable devices and smartphones. Artif. Intell. Rev. 58 (2). https://doi.org/10.1007/s10462-024-11009-5.

Scheffer, M., Carpenter, S.R., Lenton, T.M., Bascompte, J., Brock, W., Dakos, V., van de Koppel, J., van de Leemput, I.A., Levin, S.A., van Nes, E.H., Pascual, M., Vandermeer, J., 2012. Anticipating critical transitions. Science 338 (6105), 344–348. https://doi.org/10.1126/science.1225244.

Scheffer, M., Bockting, C.L., Borsboom, D., Cools, R., Delecroix, C., Hartmann, J.A., Kendler, K.S., Van De Leemput, I., Van Der Maas, H.L.J., Van Nes, E., Mattson, M., McGorry, P.D., Nelson, B., 2024a. A dynamical systems view of psychiatric disorders - practical implications: a review. JAMA Psychiatry 81 (6), 624–630. https://doi.org/10.1001/jamapsychiatry.2024.0228.

Scheffer, M., Bockting, C.L., Borsboom, D., Cools, R., Delecroix, C., Hartmann, J.A., Kendler, K.S., Van De Leemput, I., Van Der Maas, H.L.J., Van Nes, E., Mattson, M., McGorry, P.D., Nelson, B., 2024b. A dynamical systems view of psychiatric disorders - theory: a review. JAMA Psychiatry 81 (6), 618–623. https://doi.org/10.1001/jamapsychiatry.2024.0215.

Stone, A.A., Shiffman, S., 1994. Ecological Momentary Assessment (Ema) in behavioral medicine. Ann. Behav. Med. 16 (3), 199–202. https://doi.org/10.1093/abm/16.3.199.

Trefois, C., Antony, P.M.A., Goncalves, J., Skupin, A., Balling, R., 2015. Critical transitions in chronic disease: transferring concepts from ecology to systems medicine. Curr. Opin. Biotechnol. 48–55. https://doi.org/10.1016/j.copbio.2014.11.020.

Trull, T.J., Ebner-Priemer, U.W., 2009. Using Experience Sampling Methods/Ecological Momentary Assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. Psychol. Assess. 21 (4), 457–462. https://doi.org/10.1037/a0017653.

Van De Leemput, I.A., Wichers, M., Cramer, A.O.J., Borsboom, D., Tuerlinckx, F., Kuppens, P., Van Nes, E.H., Viechtbauer, W., Giltay, E.J., Aggen, S.H., Derom, C., Jacobs, N., Kendler, K.S., Van Der Maas, H.L.J., Neale, M.C., Peeters, F., Thiery, E., Zachar, P., Scheffer, M., 2014. Critical slowing down as early warning for the onset and termination of depression. Proc. Natl. Acad. Sci. USA 111 (1), 87–92. https://doi.org/10.1073/pnas.1312114110.

Van Meter, A.R., Burke, C., Youngstrom, E.A., Faedda, G.L., Correll, C.U., 2016. The bipolar prodrome: meta-analysis of symptom prevalence prior to initial or recurrent mood episodes. J. Am. Acad. Child Adolesc. Psychiatry 55 (7), 543–555. https://doi.org/10.1016/j.jaac.2016.04.017.

Wadle, L.-M., Ebner-Priemer, U.W., 2023. Smart digital phenotyping. Eur. Neuropsychopharmacol. 76 (11), 1–2. https://doi.org/10.1016/j.euroneuro.2023.07.002.

Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., 1978. A rating scale for mania: reliability, validity and sensitivity. Br. J. Psychiatry 133 (11), 429–435. https://doi.org/10.1192/bjp.133.5.429.