

Original Paper

Exploring Digital Biomarkers of Illness Activity in Mood Episodes: Hypotheses Generating and Model Development Study

Gerard Anmella^{1,2,3,4,5*}, MD; Filippo Corponi^{6*}, MD; Bryan M Li^{6*}, MSc; Ariadna Mas^{1,2,3,4,5}, PsyD; Miriam Sanabra^{1,2,3,4,5}, PsyD, PhD; Isabella Pacchiarotti^{1,2,3,4,5}, MD, PhD; Marc Valentí^{1,2,3,4,5}, MD, PhD; Iria Grande^{1,2,3,4,5}, MD, PhD; Antoni Benabarre^{1,2,3,4,5}, MD, PhD; Anna Giménez-Palomo^{1,2,3,4,5}, MD; Marina Garriga^{1,2,3,4,5}, MD, PhD; Isabel Agasi¹, RNC; Anna Bastidas^{1,4,5}, RNC; Myriam Caveró^{1,2,3,4,5}, MD, PhD; Tabatha Fernández-Plaza¹, MD; Néstor Arbelo^{1,3,4,5,7}, MD; Miquel Bioque^{1,3,4,5,7}, MD, PhD; Clemente García-Rizo^{1,3,4,5,7}, MD, PhD; Norma Verdolini^{1,2,3,4,5}, MD, PhD; Santiago Madero^{1,3,4,5,7}, MD; Andrea Murru^{1,2,3,4,5}, MD, PhD; Silvia Amoretti^{1,2,3,4,5}, PsyD, PhD; Anabel Martínez-Aran^{1,2,3,4,5}, PsyD, PhD; Victoria Ruiz¹, RNC; Giovanna Fico^{1,2,3,4,5}, MD; Michele De Prisco^{1,2,3,4,5}, MD; Vincenzo Oliva^{1,2,3,4,5}, MD; Aleix Solanes^{2,4,5,8}, MSc, PhD; Joaquim Radua^{2,4,5,8,9,10}, MD, PhD; Ludovic Samalin^{11,12}, MD, PhD; Allan H Young¹³, MD, PhD; Eduard Vieta^{1,2,3,4,5}, MD, PhD; Antonio Vergari⁶, MSc, PhD; Diego Hidalgo-Mazzei^{1,2,3,4,5}, MD, PhD

¹Department of Psychiatry and Psychology, Institute of Neuroscience, Hospital Clínic de Barcelona, Barcelona, Catalonia, Spain

²Bipolar and Depressive Disorders Unit, Digital Innovation Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Catalonia, Spain

³Biomedical Research Networking Centre Consortium on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain

⁴Department of Medicine, School of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Catalonia, Spain

⁵Institute of Neurosciences (UBNeuro), University of Barcelona, Barcelona, Catalonia, Spain

⁶School of Informatics, University of Edinburgh, Edinburgh, United Kingdom

⁷Barcelona Clinic Schizophrenia Unit, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Catalonia, Spain

⁸Imaging of Mood- and Anxiety-Related Disorders (IMARD) Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Catalonia, Spain

⁹Early Psychosis: Interventions & Clinical-detection (EPIC) Lab, Department of Psychosis Studies, Institute of Psychiatry Psychology and Neuroscience, King's College London, London, United Kingdom

¹⁰Center for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

¹¹Department of Psychiatry, Centre Hospitalier Universitaire (CHU) Clermont-Ferrand, University of Clermont Auvergne, Centre National de la Recherche Scientifique (CNRS), Clermont Auvergne INP, Institut Pascal (UMR 6602), Clermont-Ferrand, France

¹²Association Française de Psychiatrie Biologique et Neuropsychopharmacologie (AFPBN), Paris, France

¹³Centre for Affective Disorders, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

* these authors contributed equally

Corresponding Author:

Diego Hidalgo-Mazzei, MD, PhD
Department of Psychiatry and Psychology
Institute of Neuroscience
Hospital Clínic de Barcelona
Villarroel St, 170
Barcelona, Catalonia, 08036
Spain
Phone: 34 932275400 ext 4189
Email: dahidalg@clinic.cat

Abstract

Background: Depressive and manic episodes within bipolar disorder (BD) and major depressive disorder (MDD) involve altered mood, sleep, and activity, alongside physiological alterations wearables can capture.

Objective: Firstly, we explored whether physiological wearable data could predict (aim 1) the severity of an acute affective episode at the intra-individual level and (aim 2) the polarity of an acute affective episode and euthymia among different individuals. Secondly, we explored which physiological data were related to prior predictions, generalization across patients, and associations between affective symptoms and physiological data.

Methods: We conducted a prospective exploratory observational study including patients with BD and MDD on acute affective episodes (manic, depressed, and mixed) whose physiological data were recorded using a research-grade wearable (Empatica E4) across 3 consecutive time points (acute, response, and remission of episode). Euthymic patients and healthy controls were recorded during a single session (approximately 48 h). Manic and depressive symptoms were assessed using standardized psychometric scales. Physiological wearable data included the following channels: acceleration (ACC), skin temperature, blood volume pulse, heart rate (HR), and electrodermal activity (EDA). Invalid physiological data were removed using a rule-based filter, and channels were time aligned at 1-second time units and segmented at window lengths of 32 seconds, as best-performing parameters. We developed deep learning predictive models, assessed the channels' individual contribution using permutation feature importance analysis, and computed physiological data to psychometric scales' items normalized mutual information (NMI). We present a novel, fully automated method for the preprocessing and analysis of physiological data from a research-grade wearable device, including a viable supervised learning pipeline for time-series analyses.

Results: Overall, 35 sessions (1512 hours) from 12 patients (manic, depressed, mixed, and euthymic) and 7 healthy controls (mean age 39.7, SD 12.6 years; 6/19, 32% female) were analyzed. The severity of mood episodes was predicted with moderate (62%-85%) accuracies (aim 1), and their polarity with moderate (70%) accuracy (aim 2). The most relevant features for the former tasks were ACC, EDA, and HR. There was a fair agreement in feature importance across classification tasks (Kendall W=0.383). Generalization of the former models on unseen patients was of overall low accuracy, except for the intra-individual models. ACC was associated with "increased motor activity" (NMI>0.55), "insomnia" (NMI=0.6), and "motor inhibition" (NMI=0.75). EDA was associated with "aggressive behavior" (NMI=1.0) and "psychic anxiety" (NMI=0.52).

Conclusions: Physiological data from wearables show potential to identify mood episodes and specific symptoms of mania and depression quantitatively, both in BD and MDD. Motor activity and stress-related physiological data (EDA and HR) stand out as potential digital biomarkers for predicting mania and depression, respectively. These findings represent a promising pathway toward personalized psychiatry, in which physiological wearable data could allow the early identification and intervention of mood episodes.

(*JMIR Mhealth Uhealth* 2023;11:e45405) doi: [10.2196/45405](https://doi.org/10.2196/45405)

KEYWORDS

depression; mania; bipolar disorder; major depressive disorder; machine learning; deep learning; physiological data; digital biomarker; wearable; Empatica E4

Introduction

Mood disorders, including bipolar disorder (BD) and major depressive disorder (MDD), are ranked among the top 25 leading causes of disease burden worldwide [1] and are associated with recurrent depressive and manic episodes. Manic episodes are characterized by increased activity and self-esteem, reduced need for sleep, and expansive mood and behavior, whereas during depressive episodes, patients experience decreased energy and activity, sadness, low self-esteem, and social withdrawal [2-4]. These changes in mood, sleep, and activity during mood episodes translate to changes in physiological data that novel research-grade wearables can capture with high precision in real time [5,6]. Linking these digital signals with illness activity could potentially identify digital biomarkers [7].

Biomarkers are characteristics that are measured as an indicator of pathogenic processes (disease-associated biomarkers) or responses to an exposure or intervention (drug-related biomarkers) [8]. These can include molecular, histological, radiographic, or physiological characteristics. Digital biomarkers are objective, quantifiable, and physiological, and behavioral measures are collected using digital devices that are portable, wearable, implantable, or ingestible [9]. Traditional biomarkers can be invasive and expensive to measure and are difficult to

collect over time, thus giving an incomplete view of the complexity and dynamism of the disease. Alternatively, digital biomarkers are usually noninvasive, modular, and cheaper to measure, and they provide access to continuous and longitudinal measurements, both qualitative and quantitative. Moreover, they offer novel ways of measuring health status by providing perspectives into diseases that were unavailable before, which can supplement and enhance conclusions from traditional biomarkers [10]. Digital biomarkers have the potential to redefine diagnosis, improve the accuracy of diagnostic methods, enhance monitoring, and personalize interventions [11], leading to precision medicine, especially in psychiatric diseases [12].

In the last decade, there has been an exponential growth in the number of digital biomarker studies in the health domain, especially in cardiovascular and respiratory diseases [9]. Wearables are the most common type of digital devices used in digital biomarker studies, especially those incorporating accelerometer sensors that measure physical activity [13]. Wearable devices include wristbands, smartwatches, smart shirts, smart rings, smart electrodes, smart headsets, smart glasses, and so on. Wrist-worn devices are the most common type of wearable device in mental health studies and have shown to be effective in diagnosing anxiety and depression. However, none of the studies used it for treatment. The most commonly

used category of data for model development was physical activity data, followed by sleep and heart rate (HR) data [14]. There are several areas in health care in which wearable devices have shown potential, including monitoring, diagnosis, treatment, and rehabilitation of diseases. Even though wearables have shown accurate activity-tracking measurements and are acceptable for users [15], including feasibility studies in people with mental health problems [16], their implementation in usual clinical practice is still challenging [17].

Wearables collecting actigraphy, the noninvasive method of monitoring human rest and activity [18], can capture altered sleep rhythms in remitted BD [19] and also depressive symptoms [20]. In addition, actigraphy data from wearables have shown to accurately predict mood disorder diagnoses and symptom change [21]. Moreover, wearables collecting blood pulse have shown differences in HR variability (HRV) between BD and healthy controls (HCs) [22], as well as between affective states in BD [23]. In addition, people with bipolar and unipolar depression and suicidal behavior have long shown autonomic alterations that can be captured as hyporeactive electrodermal activity (EDA) [24,25], and in recent years, research-grade wearables have incorporated sensors allowing continuous EDA collection [26]. With these upgrades, in the latest years, it is now feasible to monitor mood changes in patients with MDD [27] and also predict the presence and severity of depressive states in BD and MDD with promising accuracy using wearable physiological data [28]. Despite these promising results, the specific roles of these digital signals and their longitudinal potential to measure illness activity and treatment response in mood disorders are still unknown.

The conjuncture of advances in machine learning [29] and the improved precision of wearable devices [30] may help identify physiological patterns of illness activity in mood disorders. Firstly, considering this promising background, we explored whether physiological wearable data could predict the severity of an acute affective episode at the intra-individual level (aim 1) and the polarity of an acute affective episode and euthymia among different individuals (aim 2). Secondly, we explored which physiological data were related to prior predictions, generalization across patients, and associations between affective symptoms and physiological data.

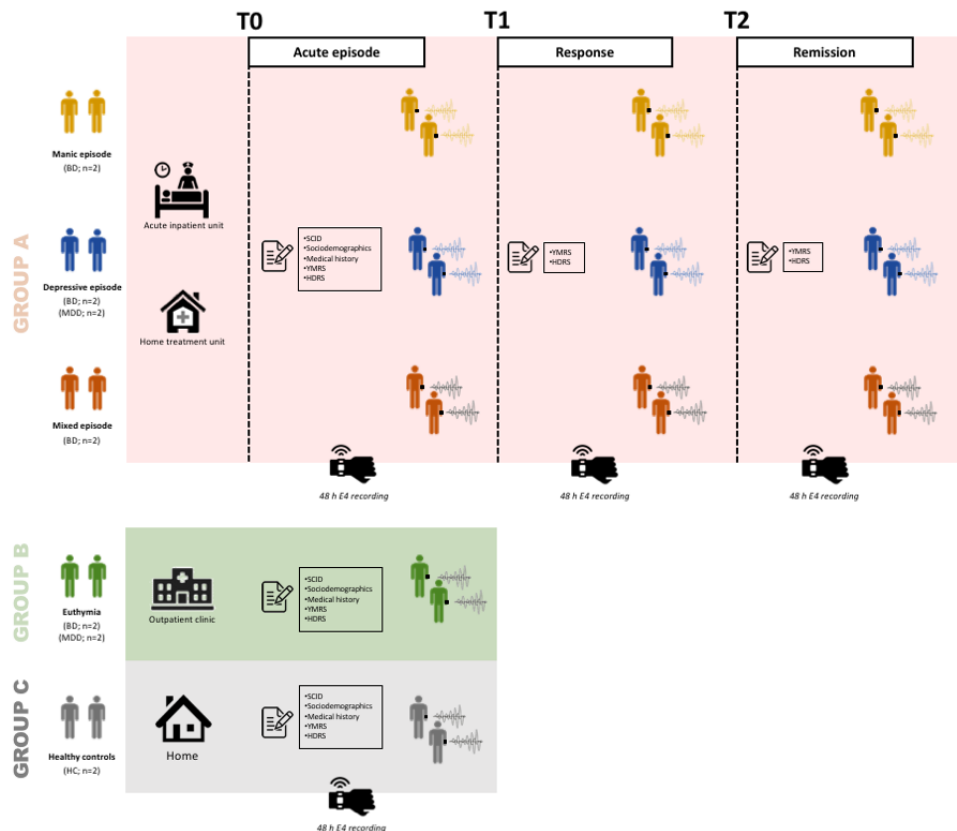
Methods

Study Design

A prospective exploratory observational study with 3 independent groups (Figure 1): group A, patients on acute affective episodes, manic episodes in BD (n=2), major depressive episodes in BD (n=2) and MDD (n=2), and mixed features manic episodes in BD (n=2); group B, euthymic patients with BD (n=2) and MDD (n=2); and group C, HC (n=7). Potential participants were identified at the outpatient and the acute inpatient or hospitalization at home units by their clinicians (ie, psychiatrists). Physiological data were recorded across 3 consecutive time points for group A: T0-acute (T0): current acute affective episodes according to the Diagnostic and Statistical Manual of Mental Disorders–5 (DSM-5); T1-response (T1): symptom response, as more than 30% improvement in the Young Mania Rating Scale (YMRS) score or the 17-item Hamilton Depression Rating Scale (HDRS) score; and T2-remission (T2): symptomatic remission, with YMRS and HDRS score ≤ 7 [31]. Euthymic patients (group B) and HCs (group C) were recorded during a single session.

The inclusion criteria were as follows: (1) aged above 18 years; (2) having a diagnosis according to the DSM-5 [32] criteria confirmed with the Structured Clinical Interview for DSM-5 Disorders [33]; and (3) willingness and ability to give consent (reconfirmed upon clinical remission). In addition, euthymic patients (group B) should also (4) score ≤ 7 on the YMRS and HDRS for at least 8 weeks [31]. HC (group C) should present no current or previous psychiatric disorder according to the DSM-5 criteria and confirmed using the Structured Clinical Interview for DSM-5 Disorders, excluding nicotine substance use disorder. Exclusion criteria for all groups were as follows: (1) concomitant severe cardiovascular or neurological medical conditions with a potential autonomic dysfunction, ongoing cardiovascular arrhythmia, or pacemaker; (2) comorbid current substance use disorder according to the DSM-5 criteria, excluding nicotine substance use disorder; (3) comorbid current psychiatric disorder with great interference of symptoms (eg, obsessive compulsive disorder with ritualized behaviors); (4) current pharmacological treatment with β -blockers or other pharmacological treatments affecting the autonomic nervous system; and (5) ongoing pregnancy.

Figure 1. Study design and recordings. BD: bipolar disorder; HC: healthy controls; HDRS: Hamilton Depression Rating Scale; MDD: major depressive disorder; SCID: Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders; T0: current acute Diagnostic and Statistical Manual of Mental Disorders–5 affective episodes; T1: symptoms' response; T2: symptomatic remission; YMRS: Young Mania Rating Scale.



Assessments

The following sociodemographic variables were collected: age, sex, DSM-5 psychiatric diagnoses [32], medical and psychiatric comorbidities, years of illness duration, first-degree relative with mental illness, and drug misuse habits. Psychopathological assessments were conducted using the YMRS [34,35] for manic symptoms and the 17-item HDRS [36,37] for depressive symptoms. Clinical assessments were performed during a single session for euthymic patients (group B) and HCs (group C) and at 3 consecutive time points (T0-acute, T1-response, and T2-remission) for patients on acute affective episodes (group A), as described in Figure 1.

Research-Grade Wearable Device for Recording

When choosing a wearable device for a research project, there are several factors that should be considered, including (1) the signals of interest to be captured (eg, stress-related and actigraphy); (2) the users who will be studied (eg, inpatients, outpatients, and HCs); (3) the pragmatic needs of the study (eg, budget, battery life, placement of the devices, and confidentiality of participants); (4) establishing assessment procedures (eg, stress elicitation task, resting, and sleep); and (5) performing qualitative and quantitative analyses on resulting data (eg, visually inspecting the data registered, quantifying data loss, assessing the quality of data, and comparing the data of different wearable devices) [38]. Considering the previous points, the E4 wristband from Empatica [39] was the preferred wearable device for the purpose of our study for several reasons. First, the E4 has shown accuracy in measuring HR, HRV [40], and EDA

compared with laboratory conditions [41], as well as for sleep staging [42]. As previously mentioned, these physiological parameters have been shown to be altered in mood disorders and mood episodes [19-23,25-28]. Second, the E4 has been validated in scientific research for detecting emotional arousal, stress [43,44], and mental effort [45] using the aforementioned physiological signals. Furthermore, the E4 has proven to be useful in predicting depressive symptoms in MDD with low relative errors [46,47], predicting self-reported depressive states [48], and identifying and quantifying the severity of anxiety states [49]. In patients with BD, the E4 has shown to be useful in distinguishing manic from euthymic mood states [50,51]. Third, the inpatients included in the study were in a highly restricted setting, which would not allow the use of user-dependent wearables or devices providing external communication (eg, an internet connection). This requirement was fulfilled by the E4 device. Finally, the data recorded by the E4 are of high precision and quality [40,41], with minimal data loss when performing the analyses (see the *Results* section).

Recording Procedure of Physiological Data

For each recording, patients and HCs were provided with an E4 wristband [39] (Multimedia Appendix 1) for approximately 48 hours (limited by battery life). The research team collected the wearables after each session. Individuals' behavior was not externally influenced in any manner, further to the requirement of wearing the wristband. Patients with acute affective episodes (group A), during their psychiatric admission in the inpatient unit, were not allowed to leave the hospital at any point until discharge, as it is the standard practice with inpatients. T0-acute,

T1-response, and T2-remission recordings were usually carried out in this setting. This was not the case with patients at the hospitalization at home or outpatient units (a minority of all cases), in which patients were not subject to mobility restrictions. In all cases, both for patients and HCs, participants were asked to wear the wristband during their daily life, with little to no interference in their behavior. They were also asked to put the wristband themselves at the beginning of the recording while researchers checked for adequate contact between the sensors and the skin wrist. Participants received instructions to remove the device when taking a shower to preserve the integrity of the device.

The E4 wristband has sensors that collect physiological data at different sampling rates. The physiological data signals from each recording session were collected from the following channels and sampling rates as raw data: 3D acceleration (ACC) in space over time on an x-, y-, and z-axis (ACC, 32 Hz); EDA (4 Hz); skin temperature (TEMP, 4 Hz); and blood volume pulse (BVP, 64 Hz); or in a processed format: interbeat intervals (IBIs, the time between 2 consecutive heart ventricular contractions) and HR (1 Hz). The BVP signal is obtained using a photoplethysmography sensor that measures volume changes in the blood. Empatica uses 2 algorithms on the BVP signal to construct an IBI with which HR (and HRV) can be calculated. The 2 algorithms are optimized to detect heartbeats and discard beats that contain artifacts [39,40].

Preprocessing of Physiological Data

Owing to the naturalistic setting of the recording sessions, the data obtained from the E4 wristband are inherently noisy. For instance, some patients show low levels of compliance during an affective episode (eg, mania), which can lead to poor skin contact from the device, hence inaccurate readings for certain channels, or complete removal of the wearable device, resulting in unusable data. To that end, we removed invalid physiological data enforcing the rules-based filter by Kleckner et al [52] and an additional rule to remove HR values that exceed the physiologically plausible range (25-250 bpm) to quality control the raw data and remove physiologically impossible recordings (Table 1). Quality controlling physiological data from wearable devices is common practice, as this type of data is particularly

noisy, and failing to quality control the data favors spurious correlations, and previous works have advised against imputing data in this scenario [53].

We did not use IBI data because of the disproportionately high number of missing values (approximately 70%) relative to data from different channels [54], especially because it is only a derivation of BVP. Therefore, we did not calculate HRV features. In sum, a total of 7 channels from the E4 device (ACC_X, ACC_Y, ACC_Z, BVP, EDA, HR, and TEMP) were used as physiological data to build the prediction models. Different time units (μ) and window lengths (w) were explored during tuning, and the best combination was selected. Because the sampling rate varied across different channels, the recordings were time aligned. If a channel's sampling rate was higher than 1 Hz, that channel was downsampled by taking the average value across samples within μ . We compared different time units ($\mu=1, 2, 4, 32, \text{ and } 64 \text{ Hz}$), and we used 1 Hz because it showed the best performance; therefore, a time unit $\mu=1$ second was set across all channels. Upon time alignment, each recording was then segmented into a predefined number of segments using a tunable window length (w), taking values in real-time seconds (s) (only powers of 2, specifically from 2^0 [1 s] to 2^{11} [2048 s], were explored for computational convenience). Of note, by tuning the hyperparameter w , an interesting pattern appeared across tasks, whereby a value of 2^5 (ie, 32 s) emerged as an optimal point, whereas smaller or higher values were associated with a deterioration in validation performance (U-shaped performance); therefore, $\mu=1 \text{ Hz}$ and $w=2^5$ (32) seconds were used for analyses as the best-performing algorithm (Multimedia Appendix 2).

To obtain an equal number of segments from each class for model evaluation, we randomly selected 20 segments from each session and stored them as a held-out test set, which was never observed by the model during either training or validation. We then randomly assigned the remaining segments to the train and validation sets with ratios of 80% and 20%, respectively. Each segment was normalized (scaled to [0, 1]) using the per-channel global (across all segments) minimum and maximum values derived from the train set.

Table 1. Rules-based filter for invalid physiological data.

Rules	Filter for invalid data	Range
1	To prevent “floor” artifacts (eg, electrode loses contact with skin) and “ceiling” artifacts (circuit is overloaded)—EDA ^a not in a valid range	0.05 to 60 μS^b
2	EDA changes too quickly—EDA slope not in a valid range	-10 to +10 $\mu\text{S}/\text{second}$
3	Skin temperature suggests the EDA sensor is not being worn—skin temperature not in a valid range	30 to 40 $^{\circ}\text{C}$
4 ^c	HR ^d not in a valid range	25 to 250 bpm ^e
5	Transitional data surrounding segments identified as invalid via the preceding rules—account for transition effects	Within 5 seconds

^aEDA: electrodermal activity.

^b μS : microsiemens.

^cAddition to the algorithm used by Kleckner et al [52].

^dHR: heart rate.

^ebpm: beats per minute.

Data Analyses

Tasks

The recording segments produced with the preprocessing steps described earlier were used in supervised learning experiments as input to the supervised models. For aim 1, models were trained on 3-class classification tasks (T0-acute, T1-response, and T2-remission) for each individual on an acute affective episode (manic BD, depressed BD, depressed MDD, and mixed BD). For aim 2, one model was trained on a 7-class classification task (manic BD, depressed BD, mixed BD, depressed MDD, euthymic BD, euthymic MDD, and HCs).

Segments from each class under a given task were extracted in the same number to obtain perfectly balanced classes. As sets were designed to be perfectly balanced, we adopted accuracy as our primary metric but also reported the F_1 -score, precision, and recall and computed the area under the receiver operating characteristic (AUROC) curves. It should be noted that ours is a multiclass setting, but as we had perfectly balanced sets, micro-, macro-, and weighted averages coincided. For the AUROC curves, the one-vs-rest multiclass strategy was adopted, also known as one-vs-all, which amounts to computing a receiver operating characteristic (ROC) curve for each class, so that at a given step, a given class is regarded as positive and the remaining classes are lumped together as a single negative class.

As part of our exploratory data analysis, to quantify the association between physiological data and affective symptoms measured by the YMRS and HDRS scale items, their normalized mutual information (NMI) was computed.

For each task, with the exception of the one about distinguishing members of a group of only HCs, as we were interested in testing the degree to which a model can generalize to different individuals, unseen during training, and sharing the same psychiatric label (diagnosis and psychopathological status), we prepared a test set of segments from recordings collected from an independent group of individuals. Therefore, the model was tested on this extra, independent holdout set to obtain an estimate of the out-of-sample generalization performance.

Model

We elected a Bidirectional Long Short-Term Memory (BiLSTM) model [55] as our model architecture. BiLSTM is a type of recurrent neural network (RNN), a class of deep learning model specifically designed to handle sequence data such as time series. RNNs process streams of data one time step at a time, and they store information regarding previous time steps in a hidden unit, such that the model output at each time step is informed by the current time step as well as by previous ones. Long short-term memory (LSTM) units represent an improvement over vanilla RNNs, as they address gradient instability by modeling the hidden state with cells that decide what to keep in memory and what to discard. This feature makes LSTM more efficient in capturing long-range dependencies. In contrast to a simple LSTM, BiLSTM reads the input sequence in 2 directions, from start to end and from end to start, thereby

allowing for a richer representation. Although other deep learning architectures suitable for time series have been developed (more recently, the transformer [56]), as the aim of this work was exploratory rather than benchmarking different models, we contented ourselves with a single popular architectural choice for time series. By the same token, we used a simple shallow BiLSTM with 128 hidden units and tanh activation, followed by a single dense layer with softmax activation, to output the possible classes. The BiLSTM model was trained using the Adam optimizer [57] for 120 epochs with a learning rate of 0.001 and a batch size of 32 to minimize the cross-entropy between the ground-truth distribution over classes and the probability distribution of belonging to such classes outputted by the last network layer. To reduce overfitting, dropout [58] and early stopping were used. The choice of hyperparameters was based on a random search that yielded the best performance in the validation set.

Permutation Feature Importance

To assess the channels' individual impact on the test set performance in the aforementioned tasks, we adopted a perturbation-based approach. For each channel at a time, we randomly permuted its values in the test set segments and computed the difference in performance relative to the baseline model. We chose this approach because it has a straightforward interpretation and provides a highly compressed, global insight into the importance of the channels. Agreement on channels' relevance across different tasks was measured using the Kendall W .

Code and Data Availability

The codebase was written in Python (version 3.8; Python Software Foundation), where the deep learning models were implemented in TensorFlow and developed on a single NVIDIA RTX 2080Ti. The repository for this study can be found on the internet [59].

Ethics Approval and Confidentiality

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki and Good Clinical Practice and the Hospital Clinic Ethics and Research Board (HCB/2021/104). All participants provided written informed consent before their inclusion in the study. All data were collected anonymously and stored encrypted in servers complying with all General Data Protection Regulation and Health Insurance Portability and Accountability Act regulations.

Results

Overview

A total of 35 sessions from 12 patients (manic, depressed, mixed, and euthymic) and 7 HCs (mean age 39.7, SD 12.6 years; 6/19, 32% female) were analyzed, totaling 1512 hours recorded. The median percentage of data per recording session dropped from further analysis of quality control was 11.05 (range 2.50-34.21). A clinical demographic overview of the study sample is presented in Table 2.

Table 2. Clinical demographic overview of the study sample.

Diagnosis	Age (years)	Sex	HDRS ^a score			YMRS ^b score		
			T0 ^c	T1 ^d	T2 ^e	T0	T1	T2
Manic BD ^f	40	Male	5	4	4	24	8	2
Manic BD ^g	21	Male	3	5	4	23	15	1
Depressed BD ^h	33	Male	23	6	4	0	0	0
Depressed BD ^{g,h}	36	Male	17	12	3	2	4	2
Mixed BD	30	Female	8	4	4	30	20	5
Mixed BD ^g	40	Male	11	2	1	29	10	3
Depressed MDD ⁱ	57	Male	33	13	7	7	2	0
Depressed MDD ^g	45	Male	27	11	7	4	1	1
Euthymic BD	54	Male	3	— ^j	—	0	—	—
Euthymic BD ^g	61	Male	1	—	—	3	—	—
Euthymic MDD	60	Female	4	—	—	0	—	—
Euthymic MDD ^g	60	Male	3	—	—	0	—	—
HC ^k	32	Female	0	—	—	0	—	—
HC ^g	34	Male	0	—	—	0	—	—
HC	28	Female	0	—	—	1	—	—
HC	29	Male	0	—	—	2	—	—
HC	31	Male	2	—	—	1	—	—
HC	32	Female	1	—	—	3	—	—
HC	31	Female	0	—	—	1	—	—

^aHDRS: Hamilton Depression Rating Scale.

^bYMRS: Young Mania Rating Scale.

^cT0: current acute Diagnostic and Statistical Manual of Mental Disorders–5 affective episodes or only register for euthymic patients and healthy controls.

^dT1: symptoms' response.

^eT2: symptomatic remission.

^fBD: bipolar disorder.

^gThe recording segments extracted from the marked subjects were used to check the models' ability to generalize to clinically similar subjects, unseen during training.

^hAll registers performed at the hospitalization at home or outpatient units.

ⁱMDD: major depressive disorder.

^jEuthymic patients and healthy controls were recorded during a single session (T0).

^kHC: healthy control.

Aim 1: Prediction of the Severity of an Acute Affective Episode at the Intra-individual Level

The 3-class classification tasks (T0-acute, T1-response, T2-remission; accuracy expected by chance: $1/3=33\%$) to predict the severity of an acute affective episode showed accuracies ranging from 62% (depressed BD) to 85% (depressed MDD). The generalization models on unseen patients showed accuracies ranging from 28% (depressed MDD) to 57% (manic BD; [Table 3](#)). The confusion matrix is shown in [Multimedia Appendix 3](#). This means that the model showed moderate to high accuracies for classifying the severity of each acute affective episode, with the best prediction models classifying individuals with depressed

MDD and manic BD. However, generalization of the models was of very low accuracy for depressed MDD and mixed BD (by chance; approximately 30%), of low accuracy (slightly above chance; >40%) for mixed BD, and of moderate accuracy (>55%) for manic BD.

The permutation importance analysis for the classification tasks for aims 1 and 2 is shown in [Figure 2](#). Kendall W was 0.383, indicating fair agreement in feature importance across both intra- and inter-individual classification tasks. ACC was the most relevant channel for predicting mania, whereas EDA and HR, followed by TEMP, were the most relevant channels for predicting both BD and unipolar depression (aim 1). The BVP

channel did not change performance for either better or worse (Figure 2).

Table 3. Prediction of the severity of an acute affective episode: model and generalization on unseen patients.

Individuals with affective episodes and performance metric	Model	Generalization
Manic BD^a		
Accuracy ^b (%)	70	56.67
F_1 -score	0.6978	0.5279
Precision	0.6979	0.5381
Recall	0.7000	0.5667
AUROC ^c	0.6980	0.5432
Depressed BD		
Accuracy ^b (%)	61.67	41.67
F_1 -score	0.6171	0.3968
Precision	0.6273	0.4085
Recall	0.6167	0.4167
AUROC	0.6115	0.4067
Mixed BD		
Accuracy ^b (%)	63.33	30
F_1 -score	0.6333	0.2576
Precision	0.6333	0.3004
Recall	0.6333	0.3068
AUROC	0.6333	0.3012
Depressed MDD^d		
Accuracy ^b (%)	85	28.33
F_1 -score	0.8492	0.2451
Precision	0.8774	0.2581
Recall	0.8500	0.2833
AUROC	0.8672	0.2856

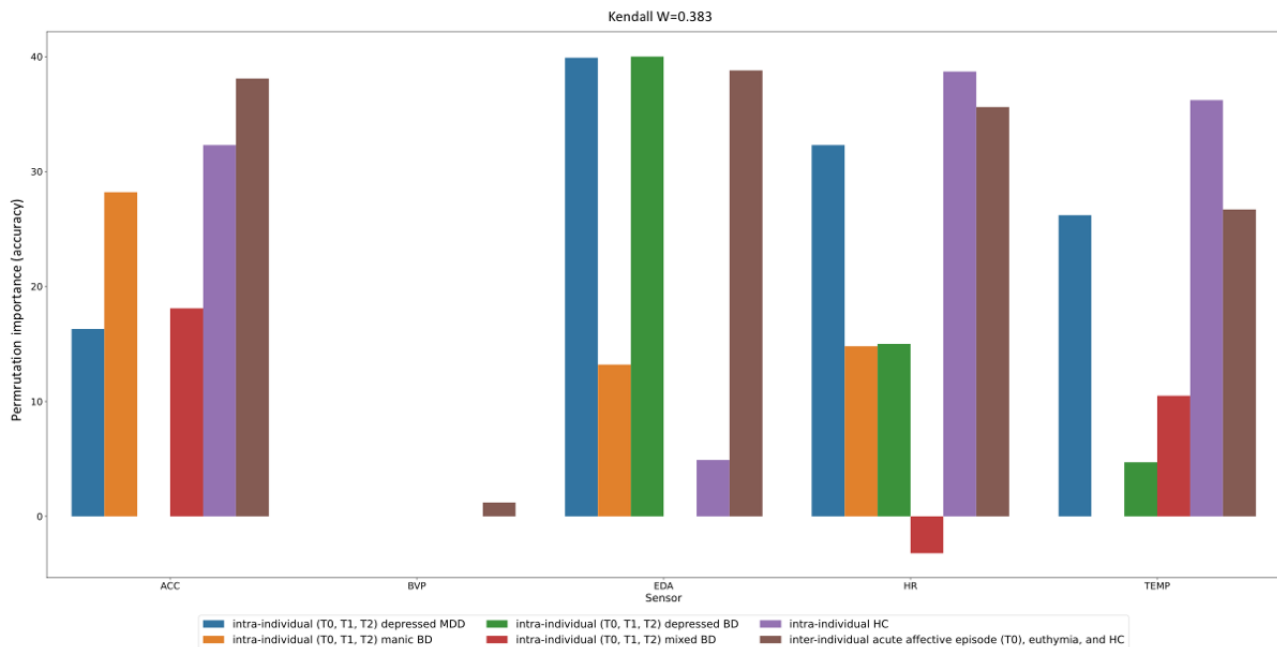
^aBD: bipolar disorder.

^bAccuracy expected by chance for a 3-class classification task is $1/3=33\%$. Thus, accuracies above 33% suggest that the model can predict outcomes better than random guessing, and higher values for accuracy indicate better predictive capacity of the model. Note that the test set was designed to have the same number of samples in each class. This is reflected in the values of F_1 -score, precision, and recall being very close to each other and to that of accuracy.

^cAUROC: area under the receiver operating characteristic.

^dMDD: major depressive disorder.

Figure 2. Permutation importance analysis. The height of the bars shows the change in accuracy at test time upon scrambling a channel through a random permutation of its values. A positive (negative) permutation importance value means that scrambling that channel results in a drop (increase) in accuracy relatively to the baseline where original (nonpermuted) values were used across all channels, that is, the channel's permutation deteriorates (improves) the performance. A "0" permutation importance value indicates that a random permutation of the channel's values does not affect accuracy in either direction. For instance, electrodermal activity (EDA) shows a positive change in accuracy of 40% for the intra-individual depressed BD severity prediction model; this means that removing this channel from the model would result in a decrease of prediction accuracy of 40%—from 62% to 22%—thus EDA is highly relevant for that model. Different colors correspond to the different tasks being investigated. ACC: acceleration; BD: bipolar disorder; BVP: blood volume pulse; HC: healthy controls; HR: heart rate; MDD: major depressive disorder; TEMP: temperature; T0: current acute Diagnostic and Statistical Manual of Mental Disorders–5 affective episodes; T1: symptoms' response; T2: symptomatic remission.



Aim 2: Prediction of the Polarity of an Acute Affective Episode and Euthymia Among Different Individuals

The 7-class classification task (accuracy expected by chance: $1/7=14\%$) to predict the polarity of affective episodes and euthymia showed an accuracy of 70%. The best classifications were depressed and euthymic MDD, followed by depressed BD, and the worst was manic BD, followed by HCs. The generalization model showed an accuracy of 15.7% (slightly above chance). The classification task for 7 HCs showed an accuracy of 50% (Table 4). The confusion matrix is shown in Multimedia Appendix 4. Thus, both models showed predictions above chance, but their generalization was poor. Moreover, the model including patients with acute affective episodes obtained

higher accuracy (70%) than the model including 7 HCs (50%). This increased prediction capacity suggests that psychopathological symptoms during acute affective episodes may translate into physiological alterations that are not present in HCs.

The most relevant channels for predicting the polarity of affective episodes, euthymia, and HCs among different individuals (aim 2) were EDA, followed by ACC, HR, and TEMP (all channels showed >30% permutation importance). The BVP channel permutation importance was approximately 0%. These results were highly similar for the classification task of 7 HCs, but EDA showed only 4.9% permutation importance (Figure 2).

Table 4. Prediction of the polarity of an acute affective episode and euthymia among different individuals: model and generalization on unseen patients.

Individuals with affective episodes and performance metric	Model	Generalization
6 patients (acute affective episodes and euthymia) and 1 HC^a		
Accuracy ^b (%)	70	15.7
F ₁ -score	0.6927	0.1516
Precision	0.6889	0.1513
Recall	0.6934	0.1517
AUROC ^c	0.6900	0.1510
7 HCs		
Accuracy ^b (%)	50	— ^d
F ₁ -score	0.4923	—
Precision	0.4911	—
Recall	0.4988	—
AUROC	0.4998	—

^aHC: healthy control.

^bAccuracy expected by chance for a 3-class classification task is 1/3=33%. Thus, accuracies above 33% suggest that the model can predict outcomes better than random guessing, and higher values for accuracy indicate better predictive capacity of the model. Note that the test set was designed to have the same number of samples in each class. This is reflected in the values of F₁-score, precision, and recall being very close to each other and to that of accuracy.

^cAUROC: area under the receiver operating characteristic.

^dAs we were interested in predicting affective psychopathology, we tested the degree to which a model can generalize to different individuals for each task except for the one about distinguishing members of a group of only HCs.

Symptom Association With Physiological Data

The tile plots for the NMI between physiological data and the YMRS and HDRS scale items for the former intra-individual (aim 1) and between-individuals (aim 2) classification tasks are

shown in Figures 3 and 4, respectively. TEMP had the highest association with psychometric scales (NMI approximately 1.0), and BVP had the lowest consistency (NMI scores oscillating from 0 to 1).

Figure 3. Tile plots for the normalized mutual information analysis between physiological data and psychometric scales' items: intra-individual level. For each scales' item the mutual information (MI) with respect to each of the channels was measured and scaled to 0 to 1 dividing by the maximum MI value for that item. Values of zero indicate no associations, values of 1 indicate the maximum recorded MI across all channels for an individual item. ACC_X: x-axis acceleration; ACC_Y: y-axis acceleration; ACC_Z: z-axis acceleration; BD: bipolar disorder; BVP: blood volume pulse; EDA: electrodermal activity; HDRS: Hamilton Depression Rating Scale; HR: heart rate; MDD: major depressive disorder; TEMP: temperature; YMRS: Young Mania Rating Scale.

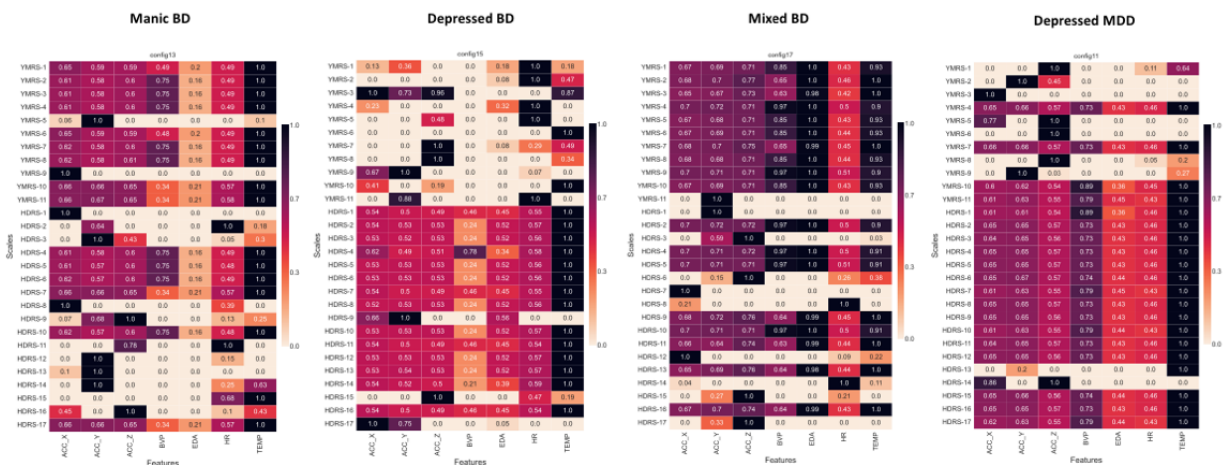
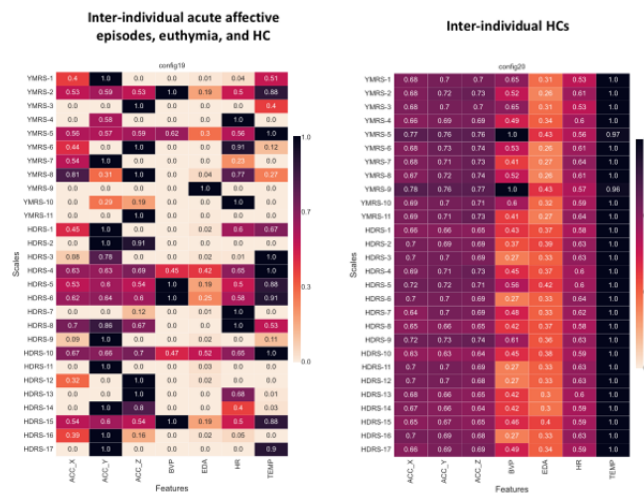


Figure 4. Tile plot for the normalized mutual information analysis between physiological data and psychometric scales' items: between-individual level. For each scales' item, the mutual information (MI) with respect to each of the channels was measured and scaled to 0 to 1 dividing by the maximum MI value for that item. Values of "0" indicate no associations; values of 1 indicate the maximum recorded MI across all channels for an individual item. ACC_X: x-axis acceleration; ACC_Y: y-axis acceleration; ACC_Z: z-axis acceleration; BVP: blood volume pulse; EDA: electrodermal activity; HC: healthy controls; HDRS: Hamilton Depression Rating Scale; HR: heart rate; TEMP: temperature; YMRS: Young Mania Rating Scale.



Intra-individual NMI Analysis

Motor activity (ACC) channels were highly associated with manic symptoms (NMI>0.6), and stress-related channels (EDA and HR) with depressive symptoms (NMI from 0.4 to 1.0), as shown in Figure 3.

Between-Individuals NMI Analysis

“Increased motor activity” (YMRS item 2 [YMRS2]) was associated with ACC (NMI>0.55), “aggressive behavior” (YMRS9) with EDA (NMI=1.0), “insomnia” (HDRS4-6) with ACC (NMI~0.6), “motor inhibition” (HDRS8) with ACC (NMI~0.75), and “psychic anxiety” (HDRS10) with EDA (NMI=0.52), as shown in Figure 4.

Discussion

Principal Findings

Although other studies have used raw physiological data to predict mental health status, this is the first study to present a novel fully automated method for the analysis of raw physiological data from a research-grade wearable device, including a rules-based filter for invalid physiological data, whereas all other studies presented methods that required manual interventions at some point in the pipeline [46,47,51,60], thus hindering the replicability and scalability of results. Moreover, our preprocessing pipeline is strictly based on the best-performing algorithm for analysis (ie, not arbitrarily decided), whereas other studies decided arbitrary cutoff points for analyzing raw physiological data (eg, ACC data recorded at 32 Hz sampling rates analyzed arbitrarily in 1-min epochs [50]). Our method may allow other research teams to use a viable supervised learning pipeline for time-series analyses for a popular research-grade wristband [39]. In addition, our work integrates physiological digital data from all sensors captured by a research-grade wearable, and we assessed the relevance of each channel (ACC, TEMP, BVP, HR, and EDA) in the prediction models. In contrast, other studies have focused on specific digital signals, such as actigraphy [50], or used

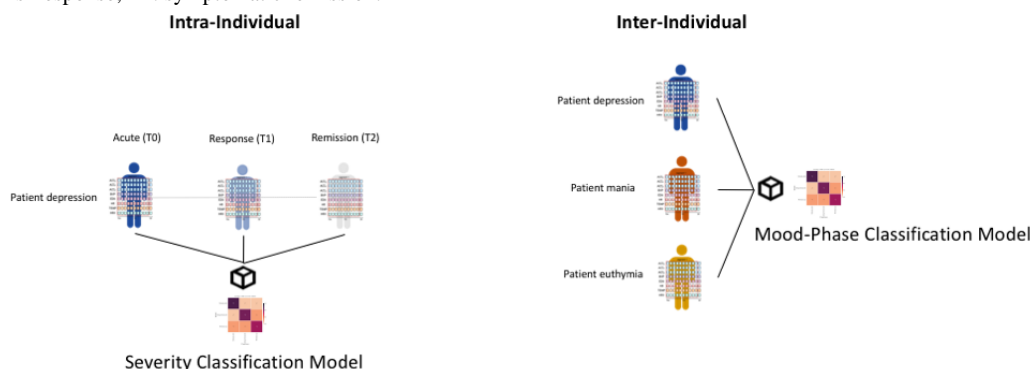
combinations of digital signals (such as actigraphy and EDA) and predesigned features (eg, amplitude of skin conductance response peaks) [51] but arbitrarily disregarded other digital signals, such as TEMP, or derived features, such as HRV. Furthermore, we aimed to distinguish the severity of mania and depression in a progressive and longitudinal manner according to the usual clinical resolution of mood episodes. We believe that the potential quantification of affective episodes is harder but a clinically more relevant task that may allow a more accurate and precise understanding of the disease rather than a mere dichotomous (acute vs remission) classification, as done in previous studies [50,51]. In addition, we included in the same work analyses at the intra-individual level and between different individuals, analyses targeting specific mood symptoms and generalization of the models on unseen patients. We believe that the use of different analysis methods allows us to examine the data from complementary perspectives to answer specific research questions. In addition, these different approaches may reveal random associations or artifacts that would stay hidden without replication. On the basis of these exploratory results, we propose hypotheses for future testing [61] in current and other similar projects.

Note that both (1) intra- and (2) inter-individual analyses approach different research questions: the (1) intra-individual analytical approach looks at the course of an index episode within a single patient and examines whether different states (from the acute phase to response and remission) can be distinguished from each other; on the other hand, the (2) inter-individual analytical approach takes a cross-sectional view and studies the degree to which different mood disorder states (comprising the full spectrum from depression to mixed state, mania, and euthymia) can be separated. Both analyses try to identify digital biomarkers of illness activity using physiological data collected with a wristband. However, intra-individual analyses look for a fine-grained quantification of illness activity that may allow the identification of low-severity mood states (or prodromal phases) in comparison with moderate to severe ones. Conversely, inter-individual analyses could potentially

distinguish between mood phases (mania vs depression) or cases from HCs but may not be suitable for assessing the severity of mood episodes, as represented in Figure 5. Studies in similar areas, such as brain computer interfaces for the rehabilitation of motor impairments [62] or seizure forecasting [63],

emphasized the importance of the subject-wise approach (modeling each subject separately). In many instances, despite work on domain adaptation [64] to learn subject-invariant representations, a model has to be fine-tuned to the level of the single patient.

Figure 5. Severity versus Mood-Phase Classification Models: visual grounds for both intra- and inter-individual analyses. On the left, a severity classification model for a patient with depression (acute-response-remission phases). On the right, a mood-phase classification model (depression, mania, and euthymia). Note that on the left model, the same individual is compared at 3 different states (corresponding to a reduction in depressive psychopathology). Thus, individual-level characteristics (age, sex, and gait) should go through little to no variation across; should remain the same on the 3 longitudinal registers; and therefore, the shift in the covariate distribution should be relatively contained and not influence the classification of the model (capturing mood-relevant signals). In contrast, on the right, 3 different individuals at 3 different mood states are compared. In this case, the model would potentially distinguish between mood phases (mania vs depression), or cases from healthy controls, but may not be able to distinguish longitudinal changes in disease severity over the course of an index episode. In addition, in the latter model, subject-specific characteristics may be overlapped with mood-relevant signals, thus acting as confounders for the model. T0: current acute Diagnostic and Statistical Manual of Mental Disorders–5 affective episodes; T1: symptoms' response; T2: symptomatic remission.



Studies comparing intra- and inter-individual models show that although intra-individual (cross-subject or patient-specific) models are trained on the data of a single subject, they perform better than intersubject (within-subject or generalized) models [65]. However, some studies have shown that hybrid models trained on multiple subjects and then fine-tuned on subject-specific data led to the best performance, without requiring as much data from a specific subject [66]. In intersubject studies, models generally see more data, as multiple subjects are included, but must contend with greater data variability, which introduces different challenges. In fact, there is both intra- and intersubject variability owing to time-variant factors related to the experimental setting and underlying psychological parameters. This impedes direct transferability or generalization among sessions and subjects [62]. To illustrate this, in a study aimed at evaluating a seizure detection model using physiological data and determining its application in a real-world setting, 2 procedures were applied: intra- and intersubject evaluation. Intrasubject evaluation focuses on the performance of the methodology when applied to data from a single patient, whereas intersubject evaluation assesses the performance of multiple patients with potentially different types of epilepsy and seizure manifestations [63].

Notably, the out-of-sample generalizations of both models differ vastly. Whereas the intra-individual model requires multiple seizures recorded per subject and will produce individualized models tailored to a single patient, the inter-individual model requires seizures recorded from multiple participants and will provide intersubject models to be used over wider populations. For this purpose, intersubject variability plays a key role: focal seizures have a multitude of possible clinical manifestations that can occur in sequence or in parallel and can be repeated or

not occur at all, in a single seizure. For instance, preictal tachycardia appears to be a phenomenon that is not generalizable to patient cohorts. Furthermore, although there may oftentimes be little change in the semiology of seizures for a single patient, they can be very heterogeneous across populations. Intra-individual models optimized for each patient can robustly detect seizures in some patients with epilepsy, but they may fail, especially when the seizures have differing semiologies that are not represented in the training data for the model. Intersubject models perform worse than if trained in an individualized manner, at least in terms of either sensitivity or false-alarm rates [63]. This is equivalent to a study aimed at evaluating a model for mood episode detection and determining its application in a real-world setting. During acute affective episodes, a huge combination of symptoms can be present in 2 different patients [67,68], and recurrent longitudinal affective episodes in a single patient can present with a similar combination of symptoms, but this is not always the case [69-72]. At the intrasubject level, out-of-sample generalization would require multiple episodes of disease occurrence longitudinally in a single patient. In fact, similar studies with intra-individual models have achieved high detection accuracies with low sample sizes and better performance than intersubject classification [63,73]. In contrast, at the intersubject level, out-of-sample generalization does not require longitudinal episodes but only cross-sectional episodes in different patients. Therefore, both models serve different but complementary purposes to build a real-world model for the detection of prodromal affective symptoms. Future studies combining intra- and inter-individual analyses should determine which of these approaches may work best to identify affective episodes, giving guidance for the design of future studies in the field.

Clinically, the end goal is to have a model inferring mood states at the individual level, regardless of whether such a model is shared across subjects or if each subject has a tailored model. Although most digital biomarker research has focused on diagnosis classification, few studies have aimed to detect longitudinal symptom change. Developing methods to detect changes in mood symptoms has the potential to prompt just-in-time interventions to prevent full-blown affective relapses and clinical deterioration and evaluate the response to pharmacological treatments with objective measures [21].

In our sample, both intra- and inter-individual models for respectively assessing differences in severity of acute affective episodes over time (Table 3) and differences in the polarity of acute affective episodes, euthymia, and HCs (Table 4) showed accuracies considerably above chance. Although preliminary, these results indicate that there may be objective differences in digital signals (ie, digital biomarkers) according to the psychopathological severity of patients (intra-individual models) and that patients with BD or MDD may present particular patterns of digital signals for mood episodes of mania and depression (inter-individual models). However, with few patients and measurements per model, these digital biomarkers may be challenging to identify and even harder to generalize.

Motor activity (from ACC) was the most relevant digital signal for predicting the severity of mania and mixed mania (but not for unipolar or bipolar depression) and also for predicting the polarity of acute affective episodes between individuals (Figure 2). In line with our results, other research groups have found that wearable motor activity data can distinguish mania from remission in patients with BD at the intra-individual level [50]. Moreover, other studies have shown that motor activity data could identify mood episodes and euthymia among different individuals, including mania versus euthymia [51], depression versus HCs [60], and mania versus depression versus HCs [74]. In fact, “activation,” which comprises having objective (motor activity) and related subjective (energy) levels emerging from underlying physiological changes, has been widely recognized as a key feature from mania [75]. Previous literature proposes that mood and activation represent distinct dimensions of BD [76] with distinct intervention approaches [77]. In addition, dysregulation of patterns of activity has been observed in BD both in acute phases and euthymia and has been proposed as a potential biomarker for BD [78]. However, it should be noted that mania may be better characterized by differences in robustness, variability, predictability, or complexity of activation rather than mean levels of activity [75], so future analyses should explore which characteristics of motor activity are key for the former predictions.

In contrast, “stress-related” digital signals (EDA and HR) were the most relevant for predicting the severity of both unipolar and bipolar depression (but not mania or mixed mania) and were also prominent for predicting the polarity of acute affective episodes between individuals (Figure 2). In fact, when looking at psychic anxiety as a symptom (item 10 from HDRS), EDA and HR showed strong associations (Figure 4). Moreover, EDA showed relevance for predicting the polarity of affective episodes between individuals but did not differentiate between HCs (38% vs 4.9%), as shown in Figure 2. This suggests that

EDA may be a specific marker for psychopathological alterations that are not present in HCs. Furthermore, skin TEMP (a proposed marker of stress) was also a relevant physiological signal for predicting the severity of unipolar and bipolar depression (Figure 2). These findings are in line with previous literature [26,79-82] and reinforce the hypothesis that stress plays a key role in people with depression. Whereas patients with manic episodes usually lack insight into their symptoms, patients with depression are usually aware of their altered state and bear much distress and anxiety [83], which may be translated into physiological alterations, as suggested in our findings.

Generalizations of the former models on unseen patients were of overall low accuracy, which may be due to high psychopathological and individual heterogeneity, as well as external factors. Although mood episodes share many psychopathological aspects, they can present with multiple combinations of symptoms [68,76,84]. Each digital signal may provide information on a specific symptom dimension (altered motor activity, sleep disturbances, and stress-related symptoms) rather than the entire affective episode (manic, depressive, or mixed). We hypothesized that training the models with a larger sample, including patients with different symptom combinations for each affective episode, will result in more precise generalizations. Thus, exploring how patients cluster according to physiological data might help toward a dimensional (rather than categorical) disease classification. Deep learning is a promising approach for clustering high-dimensional, unstructured data [85], and new methods have been proposed specifically for data from wearable devices (multivariate time series) [86,87]. Apart from polymorphic psychopathological presentations in mood episodes, there is high between-subject heterogeneity in physiological data. For instance, skin TEMP, HR, and EDA vary within a physiological range in the same individual according to external (ie, atmospheric humidity or ambient TEMP) or internal factors (ie, hydration, diet, caffeine intake, and drugs) [52], and there are also individual-level patterns (eg, specific gaits, circadian rhythms, basal skin TEMP, or HR). This calls for ad-hoc techniques to disentangle between-patient heterogeneity from mood-related signals [88] and consider the role of potential confounders in the models (eg, drugs, medical comorbidities, physical activity, atmospheric conditions, and diet). Notwithstanding, generalizations of the intra-individual models for manic BD and depressed BD were above chance, in contrast to the generalization of the inter-individual model (almost by chance). This may suggest that individual heterogeneity is partially controlled for when comparing the same individual at different time points. This way, physiological changes may be more related to psychopathology rather than simply to individual characteristics (eg, gait, sex, and age). However, intra-individual comparisons do not control for external factors (eg, humidity, atmospheric TEMP, exercise, or hydration), which should be considered and controlled for.

When exploring the association between affective symptoms and physiological data, skin TEMP showed the highest association with psychometric scales (NMI approximately 1.0; Figures 3 and 4). Skin TEMP has been proposed as an objective

physiological marker of stress [89,90], and it has been shown that people with mood disorders present objective reductions in peripheral skin TEMP (due to vasoconstriction) after stress-oriented interventions [91]. Moreover, skin TEMP from wearable data has been used to study circadian rhythms in patients with mood disorders, showing alterations in their chronobiology [92]. Even so, thermoregulatory dysfunction has been proposed in a subgroup of patients with BD [93]. However, the skin TEMP continuously recorded with wearables has been relatively understudied in mood disorders, and further efforts should be made in this direction.

Regarding the most relevant inputs for the previous models, physiological data related to specific symptom dimensions (eg, ACC with motor activity and EDA and HR variation with stress response or anxiety) seemed to be more relevant signals for predicting mood episode severity and polarity rather than more raw data, such as BVP with nearly 0% permutation importance in all models (Figures 2-4), which do not seem to have a direct clinical translation to physiological alterations related to mental health symptoms. We hypothesized that complex features with potential clinical translation (ie, indicating stress response or autonomic dysfunction), such as HRV [22,23,94], which is calculated from BVP, and EDA reactivity, calculated from EDA [26], may be of greater value than second-to-second changes in motor activity (ACC), EDA, pulse (BVP), and TEMP. We hypothesized that adding derived features as input to the models will probably result in better predictions, as shown by other research groups when identifying mood states in BD using the same wristband device [51]. Therefore, we are currently exploring derived features from raw data (ie, statistical, time-domain, and frequency-domain features) [53], assessing EDA reactivity by extracting information on the tonic and phasic components of skin conductance using novel automated methods [18,53,95], and performing stress elicitation to assess potential alterations (hyporeactivity) in the phasic component of EDA during mood episodes [26]. Finally, considering the sleep and circadian rhythm disturbances in mood disorders in both euthymia [19,96] and acute phases [97-99], we are exploring automated methods to separate sleep from wake times [87,100,101]. Our goal is to evaluate sleep disturbances and differences in physiological signals during sleep and wake periods during mood episodes [77].

Limitations

We acknowledge several limitations in this study. First, the limited sample size for model development does not allow us to make strong claims about generalization performance [102]. However, most recordings were longer than 40 hours and each patient on an acute mood episode was recorded longitudinally at 3 time points (acute, response, and remission). In fact, our data set in terms of recording hours is well above other data sets modeled with deep learning in health care settings: the deep convolutional approach proposed by Musallam et al [103] was

applied to 60 hours of electroencephalogram recordings [104]. In addition, the wearable device used (E4), allows fine-grained collection of digital physiological data (from 1 Hz to 64 Hz) for precision longitudinal time-series analyses. Regarding sample size in terms of the number of subjects, previous endeavors used as few as 12 subjects [46]. Unfortunately, this type of data, that is, recorded with a research-grade wearable device on a population with a psychiatric condition (arguably interfering with compliance to instructions), is expensive and time-consuming to collect. Second, potential confounding variables such as sex, age, pharmacological treatments, exercise, or BMI were not controlled for, and some of the study sample was not matched by age and sex. This may have biased the results, as those variables have been found to affect motor activity data, especially in between-group comparisons [60]. The within-subject design allows partial mitigation of both the weakness of a small sample size and the influence of confounders, so the models can capture mood-related signals. Therefore, we performed intra-individual comparisons across consecutive time points. In fact, the generalization of intra-individual models obtained substantially better accuracies, showing glimpses of capturing the severity of manic and depressive psychopathology.

Future works will further explore the capabilities of advanced automated machine learning models for identifying affective illness activity and the role of confounders in this association. Of particular interest are the application of clustering algorithms [87], exploring derived features (HRV [94] and EDA reactivity [26]), the role of wake and sleep periods [77,105], and the potential of physiological data to predict treatment responses and detect prodromal signs of mood episodes [106]. Future projects will include (1) studying the role of psychotic symptoms in patients with affective disorders, as well as in patients with schizophrenia; (2) assessing the role of smartphone-based derived data, including ecologic momentary assessments and passive data [107-109], in patients with BD using the SIMPLE smartphone app [110,111]; and (3) investigating the potential of combining physiological wearable data with peripheral biomarkers [112,113] and speech features [114-118].

Conclusions

Physiological wearable data may have the potential to identify and predict the severity of mania and depression in mood disorders as well as specific symptoms quantitatively. Motor activity appears to be the most relevant digital biomarker for predicting mania, whereas stress-related digital biomarkers (EDA and HR) appear to be the most relevant for predicting both bipolar and unipolar depression. In the context of biomarkers in mood disorders, these findings represent a promising pathway toward personalized psychiatry, in which clinical decisions and treatments could be supported by passive continuous and objective digital data.

Acknowledgments

The authors acknowledge the contribution of all the participants of the study.

GA is supported by a Rio Hortega 2021 grant (CM21/00017) from the Spanish Ministry of Health financed by the Instituto de Salud Carlos III (ISCIII) and cofinanced by Fondo Social Europeo Plus (FSE+). FC and BML are supported by the United Kingdom Research and Innovation (grant EP/S02431X/1), UK Research and Innovation (UKRI) Centre for Doctoral Training in Biomedical AI at the University of Edinburgh, School of Informatics. A Mas is supported by an Agència de Gestió d'Ajudes Universitàries i de Investigació (AGAUR)—PANDÈMIES 2020 grant (PI047003) from the Generalitat de Catalunya. MS is supported by a grant from the Baszucki Brain Research Fund. IG thanks the support of the Spanish Ministry of Science and Innovation (PI19/00954) integrated into the Plan Nacional de I+D+I and cofinanced by the ISCIII-Subdirección General de Evaluación y el Fondos Europeos de la Unión Europea (FEDER, FSE, Next Generation EU/Plan de Recuperación Transformación y Resiliencia_PRTR); the ISCIII; the CIBER of Mental Health (CIBERSAM); and the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (2017 SGR 1365), Centres de Recerca de Catalunya (CERCA) Programme or Generalitat de Catalunya as well as the Fundació Clínic per la Recerca Biomèdica (Pons Bartran 2022-FRCB_PB1_2022). AG-P is supported by a Rio Hortega 2021 grant (CM21/00094) from the Spanish Ministry of Health financed by ISCIII and cofinanced by Fondo Social Europeo Plus (FSE+). MB thanks the Spanish Ministry of Health and ISCIII (PI20/01066). NV thanks the Biomedicine International Training Research Programme for Excellent Clinician-Scientists (BITRECS) project, which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754550 and from "La Caixa" Foundation (ID 100010434), under the agreement LCF/PR/GN18/50310006. SM is supported by the grant "Contracte de Recerca Emili Letang-Josep Font" provided by Hospital Clínic de Barcelona. A Murru thanks the support of the Spanish Ministry of Science and Innovation (PI19/00672) integrated into the Plan Nacional de I+D+I and cofinanced by the ISCIII-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER). SA has been supported by a Sara Borrell contract (CD20/00177), funded by ISCIII and cofunded by the European Social Fund "Investing in your future." AM-A thanks the support of the Spanish Ministry of Science and Innovation (PI18/00789, PI21/00787) integrated into the Plan Nacional de I+D+I and cofinanced by ISCIII-Subdirección General de Evaluación and the FEDER; the ISCIII; the CIBERSAM; the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (2017 SGR 1365), the CERCA Programme, and the Departament de Salut de la Generalitat de Catalunya for the PERIS grant SLT006/17/00177. GF is supported by a fellowship from "La Caixa" Foundation (ID 100010434)—fellowship code—LCF/BQ/DR21/11880019. JR is supported by a Miguel Servet II contract (CPII19/00009), funded by ISCIII and cofunded by the European Social Fund "Investing in your future." EV thanks the support of the Spanish Ministry of Science, Innovation and Universities (PI15/00283, PI18/00805, PI19/00394, CPII19/00009) integrated into the Plan Nacional de I+D+I and cofinanced by the ISCIII-Subdirección General de Evaluación and the FEDER; the ISCIII; the CIBERSAM; the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (2017 SGR 1365), and the CERCA Programme or Generalitat de Catalunya. The authors would like to thank the Departament de Salut de la Generalitat de Catalunya for the PERIS grant SLT006/17/00357. DH-M is supported by a Juan Rodés JR18/00021 granted by the ISCIII.

This project was funded by the ISCIII (FIS PI21/00340, TIMEBASE Study), cofunded by the European Union, as well as a Baszucki Brain Research Fund grant (PI046998) from the Milken Foundation. The ISCIII or the Milken Foundation had no further role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Data Availability

The data supporting the findings of this study are available upon request from the corresponding author.

Authors' Contributions

GA and DH-M were responsible for study planning, project conception, and coordination. A Mas, MS, IP, MV, IG, A Benabarre, AG-P, MG, IA, A Bastidas, MC, TF-P, NA, MB, CG-R, NV, SM, SA, AM-A, and VR were responsible for recruitment. FC, BML, AV, MDP, VO, AS, and JR were responsible for data analysis. GA, FC, BML, and DH-M were responsible for manuscript preparation. All authors revised the final manuscript.

Conflicts of Interest

GA has received continuing medical education (CME)-related honoraria or consulting fees from Janssen-Cilag, Lundbeck, Lundbeck and Otsuka, and Angelini. IP has received CME-related honoraria, or consulting fees from ADAMED, Janssen-Cilag, and Lundbeck. IG has received grants and served as consultant, advisor or CME speaker for the following identities: Angelini, Casen Recordati, Ferrer, Janssen Cilag, and Lundbeck, Lundbeck-Otsuka, Luye, SEI Healthcare. AG-P has received CME-related honoraria, or consulting fees from Janssen-Cilag, Lundbeck, Casen Recordati and Angelini. MC has received grants and served as consultant, advisor or CME speaker for the following entities: Lundbeck, Esteve, Pfizer. NA has received CME-related financing from Janssen-Cilag, Lundbeck, Adamed, Pfizer, Angelini and Boston Scientific. MB has been a consultant for, received grant/research support and honoraria from, and been on the speakers/advisory board of has received honoraria from talks and/or consultancy of Adamed, Angelini, Casen-Recordati, Exeltis, Ferrer, Janssen, Lundbeck, Neuraxpharm, Otsuka, Pfizer and Sanofi. NV has received financial support for CME activities and travel funds from the following entities: Angelini, Janssen-Cilag,

Lundbeck, Otsuka. SM has received CME-related honoraria, or consulting fees from Janssen-Cilag, Lundbeck, Lundbeck/Otsuka, and Angelini. A Murru has received grants and served as consultant, advisor or CME speaker for the following entities: Angelini, Idorsia, Lundbeck, Pfizer, Takeda. LS has received CME-related honoraria, or consulting fees from Boehringer -Ingelheim, Janssen, Lundbeck/Otsuka, Sanofi-Aventis. AHY has received honoraria for lectures and advisory boards for all major pharmaceutical companies with drugs used in affective and related disorders. EV has received research support from or served as consultant, adviser or speaker for AB-Biotics, Abbott, Abbvie, Adamed, Angelini, Biogen, Celon, Dainippon Sumitomo Pharma, Ferrer, Gedeon Richter, GH Research, Glaxo SmithKline, Janssen, Lundbeck, Organon, Otsuka, Rovi, Sage pharmaceuticals, Sanofi-Aventis, Shire, Sunovion, Takeda, and Viatri. DH-M has received CME-related honoraria and served as consultant for Abbott, Angelini, Ethypharm Digital Therapy and Janssen-Cilag. All authors report no financial or other relationship relevant to the subject of this article.

Multimedia Appendix 1

Empatica E4.

[\[PNG File , 667 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Validation set performance (accuracy) as a function of time alignment (Hz) and window length (w).

[\[PNG File , 150 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Confusion matrix for the prediction of the severity of an acute affective episode: models and generalization. BD: bipolar disorder; MDD: major depressive disorder.

[\[PNG File , 296 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Confusion matrix for the prediction of the polarity of affective episodes, euthymia, and healthy controls: models and generalization. BD: bipolar disorder; HC: healthy controls; MDD: major depressive disorder; T0: current acute Diagnostic and Statistical Manual of Mental Disorders–5 affective episodes.

[\[PNG File , 341 KB-Multimedia Appendix 4\]](#)

References

1. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet* 2021 Nov 06;398(10312):1700-1712 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(21\)02143-7](https://doi.org/10.1016/S0140-6736(21)02143-7)] [Medline: [34634250](https://pubmed.ncbi.nlm.nih.gov/34634250/)]
2. Vieta E, Berk M, Schulze TG, Carvalho AF, Suppes T, Calabrese JR, et al. Bipolar disorders. *Nat Rev Dis Primers* 2018 Mar 08;4:18008. [doi: [10.1038/nrdp.2018.8](https://doi.org/10.1038/nrdp.2018.8)] [Medline: [29516993](https://pubmed.ncbi.nlm.nih.gov/29516993/)]
3. Carvalho AF, Firth J, Vieta E. Bipolar disorder. *N Engl J Med* 2020 Jul 02;383(1):58-66. [doi: [10.1056/NEJMr1906193](https://doi.org/10.1056/NEJMr1906193)] [Medline: [32609982](https://pubmed.ncbi.nlm.nih.gov/32609982/)]
4. Otte C, Gold SM, Penninx BW, Pariante CM, Etkin A, Fava M, et al. Major depressive disorder. *Nat Rev Dis Primers* 2016 Sep 15;2:16065. [doi: [10.1038/nrdp.2016.65](https://doi.org/10.1038/nrdp.2016.65)] [Medline: [27629598](https://pubmed.ncbi.nlm.nih.gov/27629598/)]
5. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. *Nat Biotechnol* 2015 May;33(5):462-463. [doi: [10.1038/nbt.3223](https://doi.org/10.1038/nbt.3223)] [Medline: [25965751](https://pubmed.ncbi.nlm.nih.gov/25965751/)]
6. Hidalgo-Mazzei D, Young AH, Vieta E, Colom F. Behavioural biomarkers and mobile mental health: a new paradigm. *Int J Bipolar Disord* 2018 May 06;6(1):9 [[FREE Full text](#)] [doi: [10.1186/s40345-018-0119-7](https://doi.org/10.1186/s40345-018-0119-7)] [Medline: [29730832](https://pubmed.ncbi.nlm.nih.gov/29730832/)]
7. Sheikh M, Qassem M, Kyriacou PA. Wearable, environmental, and smartphone-based passive sensing for mental health monitoring. *Front Digit Health* 2021 Apr 07;3:662811 [[FREE Full text](#)] [doi: [10.3389/fdgth.2021.662811](https://doi.org/10.3389/fdgth.2021.662811)] [Medline: [34713137](https://pubmed.ncbi.nlm.nih.gov/34713137/)]
8. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) resource. Food and Drug Administration, National Institutes of Health. Silver Spring, MD, USA: Food and Drug Administration; 2016. URL: <https://www.fdanews.com/ext/resources/files/2020/11-24-20-BEST.pdf?1606261388> [accessed 2023-02-15]
9. Motahari-Nezhad H, Fgaier M, Mahdi Abid M, Péntek M, Gulácsi L, Zrubka Z. Digital biomarker-based studies: scoping review of systematic reviews. *JMIR Mhealth Uhealth* 2022 Oct 24;10(10):e35722 [[FREE Full text](#)] [doi: [10.2196/35722](https://doi.org/10.2196/35722)] [Medline: [36279171](https://pubmed.ncbi.nlm.nih.gov/36279171/)]
10. Babrak LM, Menetski J, Rebhan M, Nisato G, Zinggeler M, Brasier N, et al. Traditional and digital biomarkers: two worlds apart? *Digit Biomark* 2019 Aug 16;3(2):92-102 [[FREE Full text](#)] [doi: [10.1159/000502000](https://doi.org/10.1159/000502000)] [Medline: [32095769](https://pubmed.ncbi.nlm.nih.gov/32095769/)]

11. Insel TR. Digital phenotyping: technology for a new science of behavior. *JAMA* 2017 Oct 03;318(13):1215-1216. [doi: [10.1001/jama.2017.11295](https://doi.org/10.1001/jama.2017.11295)] [Medline: [28973224](https://pubmed.ncbi.nlm.nih.gov/28973224/)]
12. Salagre E, Vieta E. Precision psychiatry: complex problems require complex solutions. *Eur Neuropsychopharmacol* 2021 Nov;52:94-95. [doi: [10.1016/j.euroneuro.2021.07.003](https://doi.org/10.1016/j.euroneuro.2021.07.003)] [Medline: [34325190](https://pubmed.ncbi.nlm.nih.gov/34325190/)]
13. Motahari-Nezhad H, Al-Abdulkarim H, Fgaier M, Abid MM, Péntek M, Gulácsi L, et al. Digital biomarker-based interventions: systematic review of systematic reviews. *J Med Internet Res* 2022 Dec 21;24(12):e41042 [FREE Full text] [doi: [10.2196/41042](https://doi.org/10.2196/41042)] [Medline: [36542427](https://pubmed.ncbi.nlm.nih.gov/36542427/)]
14. Abd-Alrazaq A, AlSaad R, Aziz S, Ahmed A, Denecke K, Househ M, et al. Wearable artificial intelligence for anxiety and depression: scoping review. *J Med Internet Res* 2023 Jan 19;25:e42672 [FREE Full text] [doi: [10.2196/42672](https://doi.org/10.2196/42672)] [Medline: [36656625](https://pubmed.ncbi.nlm.nih.gov/36656625/)]
15. Germini F, Noronha N, Borg Debono V, Abraham Philip B, Pete D, Navarro T, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. *J Med Internet Res* 2022 Jan 21;24(1):e30791 [FREE Full text] [doi: [10.2196/30791](https://doi.org/10.2196/30791)] [Medline: [35060915](https://pubmed.ncbi.nlm.nih.gov/35060915/)]
16. de Angel V, Adeleye F, Zhang Y, Cummins N, Munir S, Lewis S, et al. The feasibility of implementing remote measurement technologies in psychological treatment for depression: mixed methods study on engagement. *JMIR Ment Health* 2023 Jan 24;10:e42866 [FREE Full text] [doi: [10.2196/42866](https://doi.org/10.2196/42866)] [Medline: [36692937](https://pubmed.ncbi.nlm.nih.gov/36692937/)]
17. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. *JMIR Mhealth Uhealth* 2020 Nov 09;8(11):e18907 [FREE Full text] [doi: [10.2196/18907](https://doi.org/10.2196/18907)] [Medline: [33164904](https://pubmed.ncbi.nlm.nih.gov/33164904/)]
18. de Looft P, Duursma R, Noordzij M, Taylor S, Jaques N, Scheepers F, et al. Wearables: an R package with accompanying shiny application for signal analysis of a wearable device targeted at clinicians and researchers. *Front Behav Neurosci* 2022 Jun 23;16:856544 [FREE Full text] [doi: [10.3389/fnbeh.2022.856544](https://doi.org/10.3389/fnbeh.2022.856544)] [Medline: [35813597](https://pubmed.ncbi.nlm.nih.gov/35813597/)]
19. Geoffroy PA, Scott J, Boudebessé C, Lajnef M, Henry C, Leboyer M, et al. Sleep in patients with remitted bipolar disorders: a meta-analysis of actigraphy studies. *Acta Psychiatr Scand* 2015 Feb;131(2):89-99. [doi: [10.1111/acps.12367](https://doi.org/10.1111/acps.12367)] [Medline: [25430914](https://pubmed.ncbi.nlm.nih.gov/25430914/)]
20. Rykov Y, Thach TQ, Bojic I, Christopoulos G, Car J. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. *JMIR Mhealth Uhealth* 2021 Oct 25;9(10):e24872 [FREE Full text] [doi: [10.2196/24872](https://doi.org/10.2196/24872)] [Medline: [34694233](https://pubmed.ncbi.nlm.nih.gov/34694233/)]
21. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. *NPJ Digit Med* 2019 Feb 01;2:3 [FREE Full text] [doi: [10.1038/s41746-019-0078-0](https://doi.org/10.1038/s41746-019-0078-0)] [Medline: [31304353](https://pubmed.ncbi.nlm.nih.gov/31304353/)]
22. Faurholt-Jepsen M, Kessing LV, Munkholm K. Heart rate variability in bipolar disorder: a systematic review and meta-analysis. *Neurosci Biobehav Rev* 2017 Feb;73:68-80. [doi: [10.1016/j.neubiorev.2016.12.007](https://doi.org/10.1016/j.neubiorev.2016.12.007)] [Medline: [27986468](https://pubmed.ncbi.nlm.nih.gov/27986468/)]
23. Faurholt-Jepsen M, Brage S, Kessing LV, Munkholm K. State-related differences in heart rate variability in bipolar disorder. *J Psychiatr Res* 2017 Jan;84:169-173 [FREE Full text] [doi: [10.1016/j.jpsychires.2016.10.005](https://doi.org/10.1016/j.jpsychires.2016.10.005)] [Medline: [27743529](https://pubmed.ncbi.nlm.nih.gov/27743529/)]
24. Iacono WG, Lykken DT, Peloquin LJ, Lumry AE, Valentine RH, Tuason VB. Electrodermal activity in euthymic unipolar and bipolar affective disorders. A possible marker for depression. *Arch Gen Psychiatry* 1983 May;40(5):557-565. [doi: [10.1001/archpsyc.1983.01790050083010](https://doi.org/10.1001/archpsyc.1983.01790050083010)] [Medline: [6838333](https://pubmed.ncbi.nlm.nih.gov/6838333/)]
25. Sarchiapone M, Gramaglia C, Iosue M, Carli V, Mandelli L, Serretti A, et al. The association between electrodermal activity (EDA), depression and suicidal behaviour: a systematic review and narrative synthesis. *BMC Psychiatry* 2018 Jan 25;18(1):22 [FREE Full text] [doi: [10.1186/s12888-017-1551-4](https://doi.org/10.1186/s12888-017-1551-4)] [Medline: [29370787](https://pubmed.ncbi.nlm.nih.gov/29370787/)]
26. Greco A, Valenza G, Lanata A, Rota G, Scilingo EP. Electrodermal activity in bipolar patients during affective elicitation. *IEEE J Biomed Health Inform* 2014 Nov;18(6):1865-1873. [doi: [10.1109/JBHI.2014.2300940](https://doi.org/10.1109/JBHI.2014.2300940)] [Medline: [25375684](https://pubmed.ncbi.nlm.nih.gov/25375684/)]
27. Bai R, Xiao L, Guo Y, Zhu X, Li N, Wang Y, et al. Tracking and monitoring mood stability of patients with major depressive disorder by machine learning models using passive digital data: prospective naturalistic multicenter study. *JMIR Mhealth Uhealth* 2021 Mar 08;9(3):e24365 [FREE Full text] [doi: [10.2196/24365](https://doi.org/10.2196/24365)] [Medline: [33683207](https://pubmed.ncbi.nlm.nih.gov/33683207/)]
28. Tazawa Y, Liang KC, Yoshimura M, Kitazawa M, Kaise Y, Takamiya A, et al. Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. *Heliyon* 2020 Feb;6(2):e03274 [FREE Full text] [doi: [10.1016/j.heliyon.2020.e03274](https://doi.org/10.1016/j.heliyon.2020.e03274)] [Medline: [32055728](https://pubmed.ncbi.nlm.nih.gov/32055728/)]
29. Bhatt P, Liu J, Gong Y, Wang J, Guo Y. Emerging artificial intelligence-empowered mHealth: scoping review. *JMIR Mhealth Uhealth* 2022 Jun 09;10(6):e35053 [FREE Full text] [doi: [10.2196/35053](https://doi.org/10.2196/35053)] [Medline: [35679107](https://pubmed.ncbi.nlm.nih.gov/35679107/)]
30. Huhn S, Axt M, Gunga HC, Maggioni MA, Munga S, Obor D, et al. The impact of wearable technologies in Health research: scoping review. *JMIR Mhealth Uhealth* 2022 Jan 25;10(1):e34384 [FREE Full text] [doi: [10.2196/34384](https://doi.org/10.2196/34384)] [Medline: [35076409](https://pubmed.ncbi.nlm.nih.gov/35076409/)]
31. Tohen M, Frank E, Bowden CL, Colom F, Ghaemi SN, Yatham LN, et al. The International Society for Bipolar Disorders (ISBD) task force report on the nomenclature of course and outcome in bipolar disorders. *Bipolar Disord* 2009 Aug;11(5):453-473. [doi: [10.1111/j.1399-5618.2009.00726.x](https://doi.org/10.1111/j.1399-5618.2009.00726.x)] [Medline: [19624385](https://pubmed.ncbi.nlm.nih.gov/19624385/)]
32. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. 5th edition. Washington, DC, USA: American Psychiatric Association; May 18, 2013.

33. First MB, Williams JB, Karg RS, Spitzer RL. Structured clinical interview for DSM-5, research version. American Psychiatric Association. 2015 Nov 05. URL: <https://www.appi.org/products/structured-clinical-interview-for-dsm-5-scid-5> [accessed 2023-02-15]
34. Colom F, Vieta E, Martínez-Arán A, Garcia-Garcia M, Reinares M, Torrent C, et al. [Spanish version of a scale for the assessment of mania: validity and reliability of the Young Mania Rating Scale]. *Med Clin (Barc)* 2002 Sep 28;119(10):366-371. [doi: [10.1016/s0025-7753\(02\)73419-2](https://doi.org/10.1016/s0025-7753(02)73419-2)] [Medline: [12372167](https://pubmed.ncbi.nlm.nih.gov/12372167/)]
35. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry* 1978 Nov;133:429-435. [doi: [10.1192/bjp.133.5.429](https://doi.org/10.1192/bjp.133.5.429)] [Medline: [728692](https://pubmed.ncbi.nlm.nih.gov/728692/)]
36. Ramos-Brieva JA, Cordero-Villafafila A. A new validation of the Hamilton Rating Scale for Depression. *J Psychiatr Res* 1988;22(1):21-28. [doi: [10.1016/0022-3956\(88\)90024-6](https://doi.org/10.1016/0022-3956(88)90024-6)] [Medline: [3397906](https://pubmed.ncbi.nlm.nih.gov/3397906/)]
37. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960 Feb;23(1):56-62 [FREE Full text] [doi: [10.1136/jnnp.23.1.56](https://doi.org/10.1136/jnnp.23.1.56)] [Medline: [14399272](https://pubmed.ncbi.nlm.nih.gov/14399272/)]
38. Kleckner IR, Feldman MJ, Goodwin MS, Quigley KS. Framework for selecting and benchmarking mobile devices in psychophysiological research. *Behav Res Methods* 2021 Apr;53(2):518-535 [FREE Full text] [doi: [10.3758/s13428-020-01438-9](https://doi.org/10.3758/s13428-020-01438-9)] [Medline: [32748241](https://pubmed.ncbi.nlm.nih.gov/32748241/)]
39. Empatica E4. Empatica. URL: <https://www.empatica.com/en-gb/research/e4/> [accessed 2023-02-15]
40. Schuurmans AA, de Looft P, Nijhof KS, Rosada C, Scholte RH, Popma A, et al. Validity of the Empatica E4 wristband to measure Heart Rate Variability (HRV) parameters: a comparison to Electrocardiography (ECG). *J Med Syst* 2020 Sep 23;44(11):190 [FREE Full text] [doi: [10.1007/s10916-020-01648-w](https://doi.org/10.1007/s10916-020-01648-w)] [Medline: [32965570](https://pubmed.ncbi.nlm.nih.gov/32965570/)]
41. Menghini L, Gianfranchi E, Cellini N, Patron E, Tagliabue M, Sarlo M. Stressing the accuracy: wrist-worn wearable sensor validation over different conditions. *Psychophysiology* 2019 Nov;56(11):e13441. [doi: [10.1111/psyp.13441](https://doi.org/10.1111/psyp.13441)] [Medline: [31332802](https://pubmed.ncbi.nlm.nih.gov/31332802/)]
42. Li Q, Li Q, Cakmak AS, Da Poian G, Bliwise DL, Vaccarino V, et al. Transfer learning from ECG to PPG for improved sleep staging from wrist-worn wearables. *Physiol Meas* 2021 May 13;42(4):1088/1361-6579/abf1b0 [FREE Full text] [doi: [10.1088/1361-6579/abf1b0](https://doi.org/10.1088/1361-6579/abf1b0)] [Medline: [33761477](https://pubmed.ncbi.nlm.nih.gov/33761477/)]
43. Alinia P, Sah RK, McDonnell M, Pendry P, Parent S, Ghasemzadeh H, et al. Associations between physiological signals captured using wearable sensors and self-reported outcomes among adults in alcohol use disorder recovery: development and usability study. *JMIR Form Res* 2021 Jul 21;5(7):e27891 [FREE Full text] [doi: [10.2196/27891](https://doi.org/10.2196/27891)] [Medline: [34287205](https://pubmed.ncbi.nlm.nih.gov/34287205/)]
44. Ollander S, Godin C, Campagne A, Charbonnier S. A comparison of wearable and stationary sensors for stress detection. In: Proceedings of the 2016 International Conference on Systems, Man, and Cybernetics. 2016 Presented at: SMC '16; October 9-12, 2017; Budapest, Hungary p. 4362-4366 URL: <https://ieeexplore.ieee.org/abstract/document/7844917> [doi: [10.1109/smc.2016.7844917](https://doi.org/10.1109/smc.2016.7844917)]
45. Romine W, Schroeder N, Banerjee T, Graft J. Toward mental effort measurement using electrodermal activity features. *Sensors (Basel)* 2022 Sep 28;22(19):7363 [FREE Full text] [doi: [10.3390/s22197363](https://doi.org/10.3390/s22197363)] [Medline: [36236461](https://pubmed.ncbi.nlm.nih.gov/36236461/)]
46. Ghandeharioun A, Fedor S, Sangermano L, Ionescu D, Alpert J, Dale C, et al. Objective assessment of depressive symptoms with machine learning and wearable sensors data. In: proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction. 2017 Presented at: ACII '17; October 23-26, 2017; San Antonio, TX, USA p. 325-332 URL: <https://www.computer.org/csdl/proceedings/acii/2017/12OmNCcbEdf> [doi: [10.1109/acii.2017.8273620](https://doi.org/10.1109/acii.2017.8273620)]
47. Pedrelli P, Fedor S, Ghandeharioun A, Howe E, Ionescu DF, Bhatena D, et al. Monitoring changes in depression severity using wearable and mobile sensors. *Front Psychiatry* 2020 Dec 18;11:584711 [FREE Full text] [doi: [10.3389/fpsy.2020.584711](https://doi.org/10.3389/fpsy.2020.584711)] [Medline: [33391050](https://pubmed.ncbi.nlm.nih.gov/33391050/)]
48. Choi J, Lee S, Kim S, Kim D, Kim H. Depressed mood prediction of elderly people with a wearable band. *Sensors (Basel)* 2022 May 31;22(11):4174 [FREE Full text] [doi: [10.3390/s22114174](https://doi.org/10.3390/s22114174)] [Medline: [35684797](https://pubmed.ncbi.nlm.nih.gov/35684797/)]
49. Shaukat-Jali R, van Zalk N, Boyle DE. Detecting subclinical social anxiety using physiological data from a wrist-worn wearable: small-scale feasibility study. *JMIR Form Res* 2021 Oct 07;5(10):e32656 [FREE Full text] [doi: [10.2196/32656](https://doi.org/10.2196/32656)] [Medline: [34617905](https://pubmed.ncbi.nlm.nih.gov/34617905/)]
50. Jakobsen P, Stautland A, Riegler MA, Côté-Allard U, Sepasdar Z, Nordgreen T, et al. Complexity and variability analyses of motor activity distinguish mood states in bipolar disorder. *PLoS One* 2022 Jan 21;17(1):e0262232 [FREE Full text] [doi: [10.1371/journal.pone.0262232](https://doi.org/10.1371/journal.pone.0262232)] [Medline: [35061801](https://pubmed.ncbi.nlm.nih.gov/35061801/)]
51. Côté-Allard U, Jakobsen P, Stautland A, Nordgreen T, Fasmer OB, Oedegaard KJ, et al. Long-short ensemble network for bipolar manic-euthymic state recognition based on wrist-worn sensors. *IEEE Pervasive Comput* 2022 Apr 1;21(2):20-31 [FREE Full text] [doi: [10.1109/mprv.2022.3155728](https://doi.org/10.1109/mprv.2022.3155728)]
52. Kleckner IR, Jones RM, Wilder-Smith O, Wormwood JB, Akcakaya M, Quigley KS, et al. Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data. *IEEE Trans Biomed Eng* 2018 Jul;65(7):1460-1467 [FREE Full text] [doi: [10.1109/TBME.2017.2758643](https://doi.org/10.1109/TBME.2017.2758643)] [Medline: [28976309](https://pubmed.ncbi.nlm.nih.gov/28976309/)]
53. Föll S, Maritsch M, Spinola F, Mishra V, Barata F, Kowatsch T, et al. FLIRT: a feature generation toolkit for wearable data. *Comput Methods Programs Biomed* 2021 Nov;212:106461 [FREE Full text] [doi: [10.1016/j.cmpb.2021.106461](https://doi.org/10.1016/j.cmpb.2021.106461)] [Medline: [34736174](https://pubmed.ncbi.nlm.nih.gov/34736174/)]

54. How is IBI.csv obtained? Empatica E4. 2021 Jun 17. URL: <https://support.empatica.com/hc/en-us/articles/201912319-How-is-IBI-csv-obtained> [accessed 2022-07-05]
55. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. *IEEE Trans Signal Process* 1997 Nov;45(11):2673-2681 [FREE Full text] [doi: [10.1109/78.650093](https://doi.org/10.1109/78.650093)]
56. Vaswani A, Shazeer N, Parmar N, Erban R, Uszkoreit J, Jones L, et al. Attention is all you need. In: *Proceedings of the 31st Annual Conference on Neural Information Processing Systems*. 2017 Presented at: NeurIPS '17; December 4-9, 2017; Long Beach, CA, USA.
57. Kingma DP, Ba LJ. Adam: a method for stochastic optimization. In: *Proceedings of the International Conference on Learning Representations*. 2015 May Presented at: ICLR' 15; May 7-9, 2015; San Diego, CA, USA URL: <https://dare.uva.nl/search?identifier=a20791d3-1aff-464a-8544-268383c33a75> [doi: [10.5860/choice.189890](https://doi.org/10.5860/choice.189890)]
58. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov RR. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014 Jan 01;15(1):1929-1958 [FREE Full text]
59. Anmella G, Corponi F, Li BM, Mas A, Sanabra M, Pacchiarotti I, et al. INTREPIBD/JMIR2023: Code for JMIR mHealth and uHealth paper "Exploring digital biomarkers of illness activity in mood episodes: hypotheses generating and model development study". GitHub. 2023. URL: <https://github.com/INTREPIBD/JMIR2023> [accessed 2023-04-19]
60. Jakobsen P, Garcia-Ceja E, Riegler M, Stabell LA, Nordgreen T, Torresen J, et al. Applying machine learning in motor activity time series of depressed bipolar and unipolar patients compared to healthy controls. *PLoS One* 2020 Aug 24;15(8):e0231995 [FREE Full text] [doi: [10.1371/journal.pone.0231995](https://doi.org/10.1371/journal.pone.0231995)] [Medline: [32833958](https://pubmed.ncbi.nlm.nih.gov/32833958/)]
61. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. *Proc Natl Acad Sci U S A* 2018 Mar 13;115(11):2600-2606 [FREE Full text] [doi: [10.1073/pnas.1708274114](https://doi.org/10.1073/pnas.1708274114)] [Medline: [29531091](https://pubmed.ncbi.nlm.nih.gov/29531091/)]
62. Saha S, Baumert M. Intra- and inter-subject variability in EEG-based sensorimotor brain computer interface: a review. *Front Comput Neurosci* 2019 Jan 21;13:87 [FREE Full text] [doi: [10.3389/fncom.2019.00087](https://doi.org/10.3389/fncom.2019.00087)] [Medline: [32038208](https://pubmed.ncbi.nlm.nih.gov/32038208/)]
63. Böttcher S, Bruno E, Epitashvili N, Dümpelmann M, Zabler N, Glasstetter M, et al. Intra- and inter-subject perspectives on the detection of focal onset motor seizures in epilepsy patients. *Sensors (Basel)* 2022 Apr 26;22(9):3318 [FREE Full text] [doi: [10.3390/s22093318](https://doi.org/10.3390/s22093318)] [Medline: [35591007](https://pubmed.ncbi.nlm.nih.gov/35591007/)]
64. Özdenizci O, Wang YE, Koike-Akino T, Erdoğan D. Learning invariant representations from EEG via adversarial inference. *IEEE Access* 2020;8:27074-27085 [FREE Full text] [doi: [10.1109/access.2020.2971600](https://doi.org/10.1109/access.2020.2971600)] [Medline: [33747669](https://pubmed.ncbi.nlm.nih.gov/33747669/)]
65. Roy Y, Banville H, Albuquerque I, Gramfort A, Falk TH, Faubert J. Deep learning-based electroencephalography analysis: a systematic review. *J Neural Eng* 2019 Aug 14;16(5):051001. [doi: [10.1088/1741-2552/ab260c](https://doi.org/10.1088/1741-2552/ab260c)] [Medline: [31151119](https://pubmed.ncbi.nlm.nih.gov/31151119/)]
66. Page A, Shea C, Mohsenin T. Wearable seizure detection using convolutional neural networks with transfer learning. In: *Proceedings of the 2016 International Symposium on Circuits and Systems*. 2016 Presented at: ISCAS '16; May 22-25, 2016; Montreal, Canada p. 1086-1089 URL: <https://ieeexplore.ieee.org/abstract/document/7527433> [doi: [10.1109/iscas.2016.7527433](https://doi.org/10.1109/iscas.2016.7527433)]
67. Corponi F, Anmella G, Verdolini N, Pacchiarotti I, Samalin L, Popovic D, et al. Symptom networks in acute depression across bipolar and major depressive disorders: a network analysis on a large, international, observational study. *Eur Neuropsychopharmacol* 2020 Jun;35:49-60. [doi: [10.1016/j.euroneuro.2020.03.017](https://doi.org/10.1016/j.euroneuro.2020.03.017)] [Medline: [32409261](https://pubmed.ncbi.nlm.nih.gov/32409261/)]
68. Ostergaard SD, Jensen SO, Bech P. The heterogeneity of the depressive syndrome: when numbers get serious. *Acta Psychiatr Scand* 2011 Dec;124(6):495-496. [doi: [10.1111/j.1600-0447.2011.01744.x](https://doi.org/10.1111/j.1600-0447.2011.01744.x)] [Medline: [21838736](https://pubmed.ncbi.nlm.nih.gov/21838736/)]
69. Andrade-González N, Álvarez-Cadenas L, Saiz-Ruiz J, Lahera G. Initial and relapse prodromes in adult patients with episodes of bipolar disorder: a systematic review. *Eur Psychiatry* 2020 Feb 12;63(1):e12 [FREE Full text] [doi: [10.1192/j.eurpsy.2019.18](https://doi.org/10.1192/j.eurpsy.2019.18)] [Medline: [32093795](https://pubmed.ncbi.nlm.nih.gov/32093795/)]
70. Solomon DA, Leon AC, Coryell WH, Endicott J, Li C, Fiedorowicz JG, et al. Longitudinal course of bipolar I disorder: duration of mood episodes. *Arch Gen Psychiatry* 2010 Apr;67(4):339-347 [FREE Full text] [doi: [10.1001/archgenpsychiatry.2010.15](https://doi.org/10.1001/archgenpsychiatry.2010.15)] [Medline: [20368510](https://pubmed.ncbi.nlm.nih.gov/20368510/)]
71. Mignogna KM, Goes FS. Characterizing the longitudinal course of symptoms and functioning in bipolar disorder. *Psychol Med* 2022 Jun 14:1-11. [doi: [10.1017/S0033291722001489](https://doi.org/10.1017/S0033291722001489)] [Medline: [35699120](https://pubmed.ncbi.nlm.nih.gov/35699120/)]
72. Nandi A, Beard JR, Galea S. Epidemiologic heterogeneity of common mood and anxiety disorders over the lifecourse in the general population: a systematic review. *BMC Psychiatry* 2009 Jun 01;9:31 [FREE Full text] [doi: [10.1186/1471-244X-9-31](https://doi.org/10.1186/1471-244X-9-31)] [Medline: [19486530](https://pubmed.ncbi.nlm.nih.gov/19486530/)]
73. Bin Heyat MB, Akhtar F, Abbas SJ, Al-Sarem M, Alqarafi A, Stalin A, et al. Wearable flexible electronics based cardiac electrode for researcher mental stress detection system using machine learning models on single lead electrocardiogram signal. *Biosensors (Basel)* 2022 Jun 17;12(6):427 [FREE Full text] [doi: [10.3390/bios12060427](https://doi.org/10.3390/bios12060427)] [Medline: [35735574](https://pubmed.ncbi.nlm.nih.gov/35735574/)]
74. Krane-Gartiser K, Henriksen TE, Morken G, Vaaler A, Fasmer OB. Actigraphic assessment of motor activity in acutely admitted inpatients with bipolar disorder. *PLoS One* 2014 Feb 20;9(2):e89574 [FREE Full text] [doi: [10.1371/journal.pone.0089574](https://doi.org/10.1371/journal.pone.0089574)] [Medline: [24586883](https://pubmed.ncbi.nlm.nih.gov/24586883/)]
75. Scott J, Murray G, Henry C, Morken G, Scott E, Angst J, et al. Activation in bipolar disorders: a systematic review. *JAMA Psychiatry* 2017 Feb 01;74(2):189-196. [doi: [10.1001/jamapsychiatry.2016.3459](https://doi.org/10.1001/jamapsychiatry.2016.3459)] [Medline: [28002572](https://pubmed.ncbi.nlm.nih.gov/28002572/)]
76. Martino DJ, Valerio MP, Parker G. The structure of mania: an overview of factorial analysis studies. *Eur Psychiatry* 2020 Feb 10;63(1):e10 [FREE Full text] [doi: [10.1192/j.eurpsy.2020.18](https://doi.org/10.1192/j.eurpsy.2020.18)] [Medline: [32093802](https://pubmed.ncbi.nlm.nih.gov/32093802/)]

77. Merikangas KR, Swendsen J, Hickie IB, Cui L, Shou H, Merikangas AK, et al. Real-time mobile monitoring of the dynamic associations among motor activity, energy, mood, and sleep in adults with Bipolar Disorder. *JAMA Psychiatry* 2019 Feb 01;76(2):190-198 [FREE Full text] [doi: [10.1001/jamapsychiatry.2018.3546](https://doi.org/10.1001/jamapsychiatry.2018.3546)] [Medline: [30540352](https://pubmed.ncbi.nlm.nih.gov/30540352/)]
78. Shou H, Cui L, Hickie I, Lameira D, Lamers F, Zhang J, et al. Dysregulation of objectively assessed 24-hour motor activity patterns as a potential marker for bipolar I disorder: results of a community-based family study. *Transl Psychiatry* 2017 Aug 22;7(8):e1211 [FREE Full text] [doi: [10.1038/tp.2017.136](https://doi.org/10.1038/tp.2017.136)] [Medline: [28892068](https://pubmed.ncbi.nlm.nih.gov/28892068/)]
79. Kircanski K, Williams LM, Gotlib IH. Heart rate variability as a biomarker of anxious depression response to antidepressant medication. *Depress Anxiety* 2019 Jan;36(1):63-71 [FREE Full text] [doi: [10.1002/da.22843](https://doi.org/10.1002/da.22843)] [Medline: [30311742](https://pubmed.ncbi.nlm.nih.gov/30311742/)]
80. Schiweck C, Piette D, Berckmans D, Claes S, Vrieze E. Heart rate and high frequency heart rate variability during stress as biomarker for clinical depression. A systematic review. *Psychol Med* 2019 Jan;49(2):200-211. [doi: [10.1017/S0033291718001988](https://doi.org/10.1017/S0033291718001988)] [Medline: [30134999](https://pubmed.ncbi.nlm.nih.gov/30134999/)]
81. Kim AY, Jang EH, Kim S, Choi KW, Jeon HJ, Yu HY, et al. Automatic detection of major depressive disorder using electrodermal activity. *Sci Rep* 2018 Nov 19;8(1):17030 [FREE Full text] [doi: [10.1038/s41598-018-35147-3](https://doi.org/10.1038/s41598-018-35147-3)] [Medline: [30451895](https://pubmed.ncbi.nlm.nih.gov/30451895/)]
82. Vos G, Trinh K, Sarnyai Z, Rahimi Azghadi M. Generalizable machine learning for stress monitoring from wearable devices: a systematic literature review. *Int J Med Inform* 2023 May;173:105026 [FREE Full text] [doi: [10.1016/j.ijmedinf.2023.105026](https://doi.org/10.1016/j.ijmedinf.2023.105026)] [Medline: [36893657](https://pubmed.ncbi.nlm.nih.gov/36893657/)]
83. Stone LB, McCormack CC, Bylsma LM. Cross system autonomic balance and regulation: associations with depression and anxiety symptoms. *Psychophysiology* 2020 Oct;57(10):e13636 [FREE Full text] [doi: [10.1111/psyp.13636](https://doi.org/10.1111/psyp.13636)] [Medline: [33460174](https://pubmed.ncbi.nlm.nih.gov/33460174/)]
84. Corponi F, Anmella G, Pacchiarotti I, Samalin L, Verdolini N, Popovic D, et al. Deconstructing major depressive episodes across unipolar and bipolar depression by severity and duration: a cross-diagnostic cluster analysis on a large, international, observational study. *Transl Psychiatry* 2020 Jul 19;10(1):241 [FREE Full text] [doi: [10.1038/s41398-020-00922-2](https://doi.org/10.1038/s41398-020-00922-2)] [Medline: [32684621](https://pubmed.ncbi.nlm.nih.gov/32684621/)]
85. Aljalbout E, Golkov V, Siddiqui Y, Strobel M, Cremers D. Clustering with deep learning: taxonomy and new methods. *arXiv* 2018 Jan 23 [FREE Full text] [doi: [10.48550/arxiv.1801.07648](https://doi.org/10.48550/arxiv.1801.07648)]
86. Ienco D, Interdonato R. Deep multivariate time series embedding clustering via attentive-gated autoencoder. In: *Proceedings of the 24th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*. 2020 May Presented at: PAKDD' 2020; May 11–14, 2020; Singapore p. 318-329 URL: https://link.springer.com/chapter/10.1007/978-3-030-47426-3_25#citeas [doi: [10.1007/978-3-030-47426-3_25](https://doi.org/10.1007/978-3-030-47426-3_25)]
87. Park S, Lee SW, Han S, Cha M. Clustering insomnia patterns by data from wearable devices: algorithm development and validation study. *JMIR Mhealth Uhealth* 2019 Dec 05;7(12):e14473 [FREE Full text] [doi: [10.2196/14473](https://doi.org/10.2196/14473)] [Medline: [31804187](https://pubmed.ncbi.nlm.nih.gov/31804187/)]
88. Aglinskas A, Hartshorne JK, Anzellotti S. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. *Science* 2022 Jun 03;376(6597):1070-1074. [doi: [10.1126/science.abm2461](https://doi.org/10.1126/science.abm2461)] [Medline: [35653486](https://pubmed.ncbi.nlm.nih.gov/35653486/)]
89. Vinkers CH, Penning R, Hellhammer J, Verster JC, Klaessens JH, Olivier B, et al. The effect of stress on core and peripheral body temperature in humans. *Stress* 2013 Sep;16(5):520-530. [doi: [10.3109/10253890.2013.807243](https://doi.org/10.3109/10253890.2013.807243)] [Medline: [23790072](https://pubmed.ncbi.nlm.nih.gov/23790072/)]
90. Herborn KA, Graves JL, Jerem P, Evans NP, Nager R, McCafferty DJ, et al. Skin temperature reveals the intensity of acute stress. *Physiol Behav* 2015 Dec 01;152(Pt A):225-230 [FREE Full text] [doi: [10.1016/j.physbeh.2015.09.032](https://doi.org/10.1016/j.physbeh.2015.09.032)] [Medline: [26434785](https://pubmed.ncbi.nlm.nih.gov/26434785/)]
91. Klainin-Yobas P, Ignacio J, He HG, Lau Y, Ngooi BX, Koh SQ. Effects of a stress-management program for inpatients with mental disorders: a feasibility study. *Biol Res Nurs* 2016 Mar;18(2):213-220. [doi: [10.1177/1099800415595877](https://doi.org/10.1177/1099800415595877)] [Medline: [26183182](https://pubmed.ncbi.nlm.nih.gov/26183182/)]
92. Serrano-Serrano AB, Marquez-Arrico JE, Navarro JF, Martinez-Nicolas A, Adan A. Circadian characteristics in patients under treatment for substance use disorders and severe mental illness (schizophrenia, major depression and bipolar disorder). *J Clin Med* 2021 Sep 25;10(19):4388 [FREE Full text] [doi: [10.3390/jcm10194388](https://doi.org/10.3390/jcm10194388)] [Medline: [34640406](https://pubmed.ncbi.nlm.nih.gov/34640406/)]
93. Murphy PJ, Frei MG, Papolos D. Alterations in skin temperature and sleep in the fear of harm phenotype of pediatric bipolar disorder. *J Clin Med* 2014;3(3):959-971 [FREE Full text] [doi: [10.3390/jcm3030959](https://doi.org/10.3390/jcm3030959)] [Medline: [25530872](https://pubmed.ncbi.nlm.nih.gov/25530872/)]
94. Stautland A, Jakobsen P, Fasmer OB, Osnes B, Torresen J, Nordgreen T, et al. Heart rate variability as biomarker for bipolar disorder. *medRxiv* 2022 Feb 15 (forthcoming). [doi: [10.1101/2022.02.14.22269413](https://doi.org/10.1101/2022.02.14.22269413)]
95. Hernando-Gallego F, Luengo D, Artes-Rodríguez A. Feature extraction of galvanic skin responses by nonnegative sparse deconvolution. *IEEE J Biomed Health Inform* 2018 Sep;22(5):1385-1394. [doi: [10.1109/JBHI.2017.2780252](https://doi.org/10.1109/JBHI.2017.2780252)] [Medline: [29990244](https://pubmed.ncbi.nlm.nih.gov/29990244/)]
96. Meyer N, Faulkner SM, McCutcheon RA, Pillinger T, Dijk DJ, MacCabe JH. Sleep and circadian rhythm disturbance in remitted schizophrenia and bipolar disorder: a systematic review and meta-analysis. *Schizophr Bull* 2020 Mar 10;46(5):1126-1143 [FREE Full text] [doi: [10.1093/schbul/sbaa024](https://doi.org/10.1093/schbul/sbaa024)] [Medline: [32154882](https://pubmed.ncbi.nlm.nih.gov/32154882/)]
97. Lewis KS, Gordon-Smith K, Forty L, Di Florio A, Craddock N, Jones L, et al. Sleep loss as a trigger of mood episodes in bipolar disorder: individual differences based on diagnostic subtype and gender. *Br J Psychiatry* 2017 Sep;211(3):169-174 [FREE Full text] [doi: [10.1192/bjp.bp.117.202259](https://doi.org/10.1192/bjp.bp.117.202259)] [Medline: [28684405](https://pubmed.ncbi.nlm.nih.gov/28684405/)]

98. Murru A, Guiso G, Barbuti M, Anmella G, Verdolini N, Samalin L, BRIDGE-II-Mix Study Group. The implications of hypersomnia in the context of major depression: results from a large, international, observational study. *Eur Neuropsychopharmacol* 2019 Apr;29(4):471-481 [FREE Full text] [doi: [10.1016/j.euroneuro.2019.02.011](https://doi.org/10.1016/j.euroneuro.2019.02.011)] [Medline: [30846287](https://pubmed.ncbi.nlm.nih.gov/30846287/)]
99. Steinan MK, Scott J, Lagerberg TV, Melle I, Andreassen OA, Vaaler AE, et al. Sleep problems in bipolar disorders: more than just insomnia. *Acta Psychiatr Scand* 2016 May;133(5):368-377 [FREE Full text] [doi: [10.1111/acps.12523](https://doi.org/10.1111/acps.12523)] [Medline: [26590799](https://pubmed.ncbi.nlm.nih.gov/26590799/)]
100. Liu J, Zhao Y, Lai B, Wang H, Tsui KL. Wearable device heart rate and activity data in an unsupervised approach to personalized sleep monitoring: algorithm validation. *JMIR Mhealth Uhealth* 2020 Aug 05;8(8):e18370 [FREE Full text] [doi: [10.2196/18370](https://doi.org/10.2196/18370)] [Medline: [32755887](https://pubmed.ncbi.nlm.nih.gov/32755887/)]
101. Wei J, Boger J. Sleep detection for younger adults, healthy older adults, and older adults living with dementia using wrist temperature and actigraphy: prototype testing and case study analysis. *JMIR Mhealth Uhealth* 2021 Jun 01;9(6):e26462 [FREE Full text] [doi: [10.2196/26462](https://doi.org/10.2196/26462)] [Medline: [34061038](https://pubmed.ncbi.nlm.nih.gov/34061038/)]
102. Vos G, Trinh K, Sarnyai Z, Azghadi M. Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices. *arXiv* 2022 Sep 30. [doi: [10.48550/arxiv.2209.15146](https://doi.org/10.48550/arxiv.2209.15146)]
103. Musallam YK, AlFassam NI, Muhammad G, Amin SU, Alsulaiman M, Abdul W, et al. Electroencephalography-based motor imagery classification using temporal convolutional network fusion. *Biomed Signal Process Control* 2021 Aug;69:102826 [FREE Full text] [doi: [10.1016/j.bspc.2021.102826](https://doi.org/10.1016/j.bspc.2021.102826)]
104. Kaya M, Binli MK, Ozbay E, Yanar H, Mishchenko Y. A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. *Sci Data* 2018 Oct 16;5:180211 [FREE Full text] [doi: [10.1038/sdata.2018.211](https://doi.org/10.1038/sdata.2018.211)] [Medline: [30325349](https://pubmed.ncbi.nlm.nih.gov/30325349/)]
105. Zhang Y, Folarin AA, Sun S, Cummins N, Bendayan R, Ranjan Y, RADAR-CNS Consortium. Relationship between major depression symptom severity and sleep collected using a wristband wearable device: multicenter longitudinal observational study. *JMIR Mhealth Uhealth* 2021 Apr 12;9(4):e24604 [FREE Full text] [doi: [10.2196/24604](https://doi.org/10.2196/24604)] [Medline: [33843591](https://pubmed.ncbi.nlm.nih.gov/33843591/)]
106. Van Assche E, Antoni Ramos-Quiroga J, Pariante CM, Sforzini L, Young AH, Flossbach Y, et al. Digital tools for the assessment of pharmacological treatment for depressive disorder: state of the art. *Eur Neuropsychopharmacol* 2022 Jul;60:100-116. [doi: [10.1016/j.euroneuro.2022.05.007](https://doi.org/10.1016/j.euroneuro.2022.05.007)] [Medline: [35671641](https://pubmed.ncbi.nlm.nih.gov/35671641/)]
107. Dunster GP, Swendsen J, Merikangas KR. Real-time mobile monitoring of bipolar disorder: a review of evidence and future directions. *Neuropsychopharmacology* 2021 Jan;46(1):197-208 [FREE Full text] [doi: [10.1038/s41386-020-00830-5](https://doi.org/10.1038/s41386-020-00830-5)] [Medline: [32919408](https://pubmed.ncbi.nlm.nih.gov/32919408/)]
108. Gillett G, Saunders KE. Remote monitoring for understanding mechanisms and prediction in psychiatry. *Curr Behav Neurosci Rep* 2019 May 2;6(2):51-56 [FREE Full text] [doi: [10.1007/s40473-019-00176-3](https://doi.org/10.1007/s40473-019-00176-3)]
109. Kessing LV, Faurholt-Jepsen M. Mood instability - a new outcome measure in randomised trials of bipolar disorder? *Eur Neuropsychopharmacol* 2022 May;58:39-41. [doi: [10.1016/j.euroneuro.2022.02.005](https://doi.org/10.1016/j.euroneuro.2022.02.005)] [Medline: [35219178](https://pubmed.ncbi.nlm.nih.gov/35219178/)]
110. García-Estela A, Cantillo J, Angarita-Osorio N, Mur-Milà E, Anmella G, Pérez V, et al. Real-world implementation of a smartphone-based psychoeducation program for bipolar disorder: observational ecological study. *J Med Internet Res* 2022 Feb 02;24(2):e31565 [FREE Full text] [doi: [10.2196/31565](https://doi.org/10.2196/31565)] [Medline: [35107440](https://pubmed.ncbi.nlm.nih.gov/35107440/)]
111. Hidalgo-Mazzei D, Mateu A, Reinares M, Murru A, Del Mar Bonnín C, Varo C, et al. Psychoeducation in bipolar disorder with a SIMPLE smartphone application: feasibility, acceptability and satisfaction. *J Affect Disord* 2016 Aug;200:58-66. [doi: [10.1016/j.jad.2016.04.042](https://doi.org/10.1016/j.jad.2016.04.042)] [Medline: [27128358](https://pubmed.ncbi.nlm.nih.gov/27128358/)]
112. Stanislaus S, Faurholt-Jepsen M, Vinberg M, Poulsen HE, Kessing LV, Coello K. Associations between oxidative stress markers and patient-reported smartphone-based symptoms in patients newly diagnosed with bipolar disorder: an exploratory study. *Eur Neuropsychopharmacol* 2022 Sep;62:36-45. [doi: [10.1016/j.euroneuro.2022.07.002](https://doi.org/10.1016/j.euroneuro.2022.07.002)] [Medline: [35896055](https://pubmed.ncbi.nlm.nih.gov/35896055/)]
113. Anmella G, Sanabra M, Mas-Musons A, Hidalgo-Mazzei D. Combining digital with peripheral biomarkers in bipolar disorder. *Eur Neuropsychopharmacol* 2022 Oct;63:71-72. [doi: [10.1016/j.euroneuro.2022.07.187](https://doi.org/10.1016/j.euroneuro.2022.07.187)] [Medline: [36081269](https://pubmed.ncbi.nlm.nih.gov/36081269/)]
114. Zhang J, Pan Z, Gui C, Xue T, Lin Y, Zhu J, et al. Analysis on speech signal features of manic patients. *J Psychiatr Res* 2018 Mar;98:59-63. [doi: [10.1016/j.jpsychires.2017.12.012](https://doi.org/10.1016/j.jpsychires.2017.12.012)] [Medline: [29291581](https://pubmed.ncbi.nlm.nih.gov/29291581/)]
115. Guidi A, Schoentgen J, Bertschy G, Gentili C, Landini L, Scilingo EP, et al. Voice quality in patients suffering from bipolar disease. *Annu Int Conf IEEE Eng Med Biol Soc* 2015;2015:6106-6109. [doi: [10.1109/EMBC.2015.7319785](https://doi.org/10.1109/EMBC.2015.7319785)] [Medline: [26737685](https://pubmed.ncbi.nlm.nih.gov/26737685/)]
116. Weiner L, Doignon-Camus N, Bertschy G, Giersch A. Thought and language disturbance in bipolar disorder quantified via process-oriented verbal fluency measures. *Sci Rep* 2019 Oct 03;9(1):14282 [FREE Full text] [doi: [10.1038/s41598-019-50818-5](https://doi.org/10.1038/s41598-019-50818-5)] [Medline: [31582814](https://pubmed.ncbi.nlm.nih.gov/31582814/)]
117. Carrillo F, Mota N, Copelli M, Ribeiro S, Sigman M, Cecchi G, et al. Emotional intensity analysis in bipolar subjects. *arXiv* 2016 Jun 07 [FREE Full text]
118. Carrillo F, Sigman M, Fernández Slezak DF, Ashton P, Fitzgerald L, Stroud J, et al. Natural speech algorithm applied to baseline interview data can predict which patients will respond to psilocybin for treatment-resistant depression. *J Affect Disord* 2018 Apr 01;230:84-86. [doi: [10.1016/j.jad.2018.01.006](https://doi.org/10.1016/j.jad.2018.01.006)] [Medline: [29407543](https://pubmed.ncbi.nlm.nih.gov/29407543/)]

Abbreviations

ACC: acceleration
AUROC: area under the receiver operating characteristic
BD: bipolar disorder
BiLSTM: Bidirectional Long Short-Term Memory
BVP: blood volume pulse
DSM-5: Diagnostic and Statistical Manual of Mental Disorders–5
EDA: electrodermal activity
HC: healthy control
HDRS: Hamilton Depression Rating Scale
HR: heart rate
HRV: heart rate variability
IBI: interbeat interval
LSTM: long short-term memory
MDD: major depressive disorder
NMI: normalized mutual information
RNN: recurrent neural network
ROC: receiver operating characteristic
T0: current acute Diagnostic and Statistical Manual of Mental Disorders–5 affective episodes
T1: symptoms' response
T2: symptomatic remission
TEMP: temperature
YMRS: Young Mania Rating Scale

Edited by L Buis; submitted 29.12.22; peer-reviewed by OR Patil, Y Zhang, V Gupta, S Tedesco; comments to author 25.01.23; revised version received 20.02.23; accepted 07.03.23; published 04.05.23

Please cite as:

Anmella G, Corponi F, Li BM, Mas A, Sanabra M, Pacchiarotti I, Valentí M, Grande I, Benabarre A, Giménez-Palomo A, Garriga M, Agasi I, Bastidas A, Cavero M, Fernández-Plaza T, Arbelo N, Bioque M, García-Rizo C, Verdolini N, Madero S, Murru A, Amoretti S, Martínez-Aran A, Ruiz V, Fico G, De Prisco M, Oliva V, Solanes A, Radua J, Samalin L, Young AH, Vieta E, Vergari A, Hidalgo-Mazzei D

Exploring Digital Biomarkers of Illness Activity in Mood Episodes: Hypotheses Generating and Model Development Study

JMIR Mhealth Uhealth 2023;11:e45405

URL: <https://mhealth.jmir.org/2023/1/e45405>

doi: [10.2196/45405](https://doi.org/10.2196/45405)

PMID: [36939345](https://pubmed.ncbi.nlm.nih.gov/36939345/)

©Gerard Anmella, Filippo Corponi, Bryan M Li, Ariadna Mas, Miriam Sanabra, Isabella Pacchiarotti, Marc Valentí, Iria Grande, Antoni Benabarre, Anna Giménez-Palomo, Marina Garriga, Isabel Agasi, Anna Bastidas, Myriam Cavero, Tabatha Fernández-Plaza, Néstor Arbelo, Miquel Bioque, Clemente García-Rizo, Norma Verdolini, Santiago Madero, Andrea Murru, Silvia Amoretti, Anabel Martínez-Aran, Victoria Ruiz, Giovanna Fico, Michele De Prisco, Vincenzo Oliva, Aleix Solanes, Joaquim Radua, Ludovic Samalin, Allan H Young, Eduard Vieta, Antonio Vergari, Diego Hidalgo-Mazzei. Originally published in JMIR mHealth and uHealth (<https://mhealth.jmir.org>), 04.05.2023. 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 mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on <https://mhealth.jmir.org/>, as well as this copyright and license information must be included.