



**HAL**  
open science

## Digital phenotype of mood disorders: A conceptual and critical review

Redwan Maatoug, Antoine Oudin, Vladimir Adrien, Bertrand Saudreau,  
Olivier Bonnot, Bruno Millet, Florian Ferreri, Stephane Mouchabac, Alexis  
Bourla

► **To cite this version:**

Redwan Maatoug, Antoine Oudin, Vladimir Adrien, Bertrand Saudreau, Olivier Bonnot, et al.. Digital phenotype of mood disorders: A conceptual and critical review. *Frontiers in Psychiatry*, 2022, 13, 10.3389/fpsyt.2022.895860 . hal-04590124

**HAL Id: hal-04590124**

**<https://nantes-universite.hal.science/hal-04590124v1>**

Submitted on 28 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



## OPEN ACCESS

## EDITED BY

Nicholas C. Jacobson,  
Dartmouth College, United States

## REVIEWED BY

Olusola Ajilore,  
University of Illinois at Chicago,  
United States  
Nathan Allen,  
The University of Auckland,  
New Zealand  
Frederick Sundram,  
The University of Auckland,  
New Zealand, in collaboration with  
reviewer NA

## \*CORRESPONDENCE

Alexis Bourla  
alexis.bourla@aphp.fr

†These authors have contributed  
equally to this work

## SPECIALTY SECTION

This article was submitted to  
Digital Mental Health,  
a section of the journal  
Frontiers in Psychiatry

RECEIVED 14 March 2022

ACCEPTED 07 July 2022

PUBLISHED 26 July 2022

## CITATION

Maatoug R, Oudin A, Adrien V,  
Saudreau B, Bonnot O, Millet B,  
Ferreri F, Mouchabac S and Bourla A  
(2022) Digital phenotype of mood  
disorders: A conceptual and critical  
review.

*Front. Psychiatry* 13:895860.  
doi: 10.3389/fpsy.2022.895860

## COPYRIGHT

© 2022 Maatoug, Oudin, Adrien,  
Saudreau, Bonnot, Millet, Ferreri,  
Mouchabac and Bourla. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which  
does not comply with these terms.

# Digital phenotype of mood disorders: A conceptual and critical review

Redwan Maatoug<sup>1,2†</sup>, Antoine Oudin<sup>1,2†</sup>, Vladimir Adrien<sup>2,3</sup>,  
Bertrand Saudreau<sup>2,4</sup>, Olivier Bonnot<sup>5,6</sup>, Bruno Millet<sup>1,2</sup>,  
Florian Ferreri<sup>2,3</sup>, Stephane Mouchabac<sup>2,3</sup> and  
Alexis Bourla<sup>2,3,7\*</sup>

<sup>1</sup>Service de Psychiatrie Adulte de la Pitié-Salpêtrière, Institut du Cerveau (ICM), Sorbonne Université, Assistance Publique des Hôpitaux de Paris (AP-HP), Paris, France, <sup>2</sup>iCRIN (Infrastructure for Clinical Research in Neurosciences), Paris Brain Institute (ICM), Sorbonne Université, INSERM, CNRS, Paris, France, <sup>3</sup>Department of Psychiatry, Sorbonne Université, Hôpital Saint Antoine-Assistance Publique des Hôpitaux de Paris (AP-HP), Paris, France, <sup>4</sup>Département de Psychiatrie de l'Enfant et de l'Adolescent, Assistance Publique des Hôpitaux de Paris (AP-HP), Sorbonne Université, Paris, France, <sup>5</sup>CHU de Nantes, Department of Child and Adolescent Psychiatry, Nantes, France, <sup>6</sup>Pays de la Loire Psychology Laboratory, Nantes, France, <sup>7</sup>INICEA Korian, Paris, France

**Background:** Mood disorders are commonly diagnosed and staged using clinical features that rely merely on subjective data. The concept of digital phenotyping is based on the idea that collecting real-time markers of human behavior allows us to determine the digital signature of a pathology. This strategy assumes that behaviors are quantifiable from data extracted and analyzed through digital sensors, wearable devices, or smartphones. That concept could bring a shift in the diagnosis of mood disorders, introducing for the first time additional examinations on psychiatric routine care.

**Objective:** The main objective of this review was to propose a conceptual and critical review of the literature regarding the theoretical and technical principles of the digital phenotypes applied to mood disorders.

**Methods:** We conducted a review of the literature by updating a previous article and querying the PubMed database between February 2017 and November 2021 on titles with relevant keywords regarding digital phenotyping, mood disorders and artificial intelligence.

**Results:** Out of 884 articles included for evaluation, 45 articles were taken into account and classified by data source (multimodal, actigraphy, ECG, smartphone use, voice analysis, or body temperature). For depressive episodes, the main finding is a decrease in terms of functional and biological parameters [decrease in activities and walking, decrease in the number of calls and SMS messages, decrease in temperature and heart rate variability (HRV)], while the manic phase produces the reverse phenomenon (increase in activities, number of calls and HRV).

**Conclusion:** The various studies presented support the potential interest in digital phenotyping to computerize the clinical characteristics of mood disorders.

#### KEYWORDS

mood disorders, digital phenotyping, machine learning, artificial intelligence, depressive disorder, bipolar disorder

## Introduction

The diagnosis of mood disorders currently relies purely on clinical interviews based on the identification of symptoms that can be either subjective (sadness, anhedonia, exaltation, etc.), or that could potentially be objectified (attention disorders, psychomotor retardation, sleep disorders, etc.). The search for biomarkers is one of the major challenges in this field, and the concept of the “Digital Phenotype” (DP) of a pathology can be understood as a kind of digital biomarker, sustaining the extended phenotype, a concept introduced by Richard Dawkins that implies that phenotypes should not be limited just to biological processes (1).

Defined in 2015 by S. H. Jain (2) and shortly after by J. Torous (3) for psychiatry, the digital phenotype refers to the real-time capture by computerized measurement tools of certain characteristics specific to a psychiatric disorder. Some behaviors or symptoms could be quantifiable, which would bring a shift in the assessment of psychiatric semiology, offering a new branch of investigation constituted by an “e-semiology.” An accelerometer or an actigraphic device can detect changes in a motor symptomatology (e.g., acceleration during a manic episode, decreased activation during a depressive episode, or replacement of graphorrhoea by an increase in the number of SMS messages sent). Models based on these new signs are starting to emerge since the miniaturization of sensors and the ubiquitous use of smartphones allow extensive data collection to which psychiatrists did not have access before.

To highlight objective symptoms of mood disorders, the most often studied criteria are those relating to motor aspects (slowing down and restlessness), speech characteristics (speed, prosody, tenor), or sleep disorders (insomnia or hypersomnia) as well as biometric data detectable by sensors [heart rate (HR), temperature, etc.]. This collection method is called “passive data gathering”; no intervention is necessary, reducing the weight of the observers and mitigating cognitive bias of the clinicians. This term is opposed to “active data gathering,” which requires the involvement of the patient in the collection of the data (e.g., Ecological Momentary Assessment EMA) (Figure 1).

Passive data are collected automatically in real time, without requiring any input from the user and relying on tools such as the accelerometer (number of steps, motor behavior), GPS, mobile phone-based software sensing (e.g., sleep analysis), or wearables that measure HR, heart rate variability (HRV), galvanic skin conductance, temperature, blood pressure or others indicators that could be considered as potential biomarkers of certain psychiatric disorders. Smartphone use (e.g., number of SMS messages, call log, voice analysis, social media posts, internet use, online shopping, music, pictures, calendar) provides access to psychosocial functioning and to passive assessment of content.

All that information can be considered “big data” since it comes from multiple sources (e.g., multimodal passive data) and aggregates different features, and many research teams aim to determine the digital phenotypes of several mood disorders using machine learning, regression analysis or natural language processing approaches.

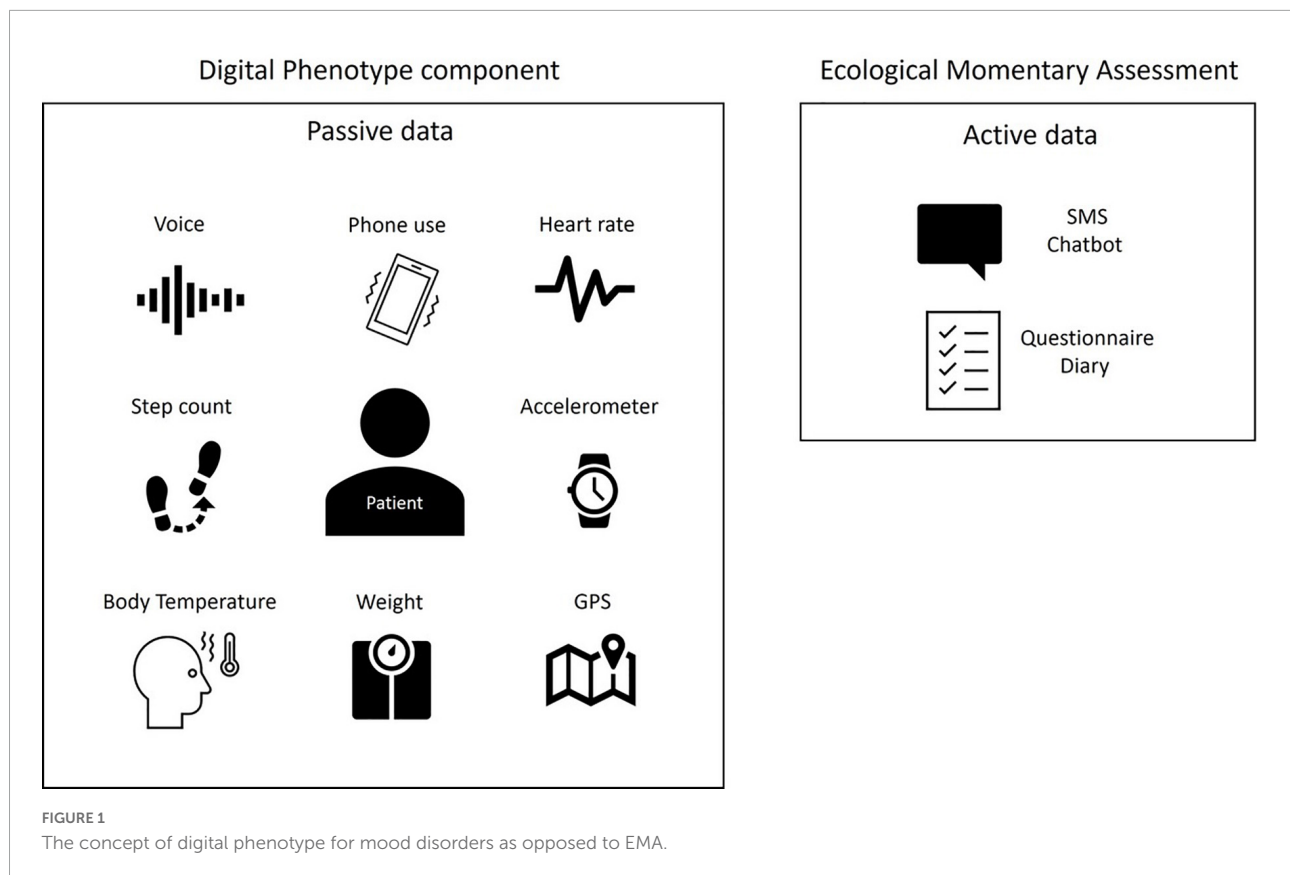
In this review we propose to investigate the digital phenotype of mood disorders (depressive disorder and bipolar disorder).

## Methods

We conducted a review of the literature by updating a previous article (4) covering the literature from 2010 to February 2017. The present review was conducted by querying the PubMed database between February 2017 and November 2021 for titles with the terms [computer] OR [computerized] OR [mobile] OR [automatic] OR [automated] OR [machine learning] OR [sensor] OR [heart rate variability] OR [HRV] OR [actigraphy] OR [actimetry] OR [digital] OR [motion] OR [temperature] with each of the terms AND each of the following: [mood], [bipolar], [depression], [depressive], [manic], [mania]. For studies published before 2017 see Bourla et al. (4). Article selection can be seen in the PRISMA diagram (Figure 2).

Exclusion criteria were:

- Reviews and meta-analyses
- The use of digital phenotype for evaluating treatment response



- The use of digital phenotype for evaluating interventions
- The use of medical equipment for HRV/HR assessment (only smartphone sensors)
- Therapeutic interventions
- Postpartum depression or pregnancy studies
- Child or adolescent studies
- Age above 65 years
- Comorbidities (e.g., post-stroke depression, chronic obstructive pulmonary disease, Human Immunodeficiency Virus, cardiosurgical patients)
- Multimodal evaluation with blood or imagery biomarkers
- Study protocols
- Case reports
- Ecological Momentary Assessment (EMA) or self-rating questionnaires or self-reports not associated with digital phenotype.

## Results

Some studies use multiple sensors at the same time (multimodal data source), while other studies focus specifically on one type of sensor (unimodal data source). Forty-five articles were included and classified by data source: Smartphone and

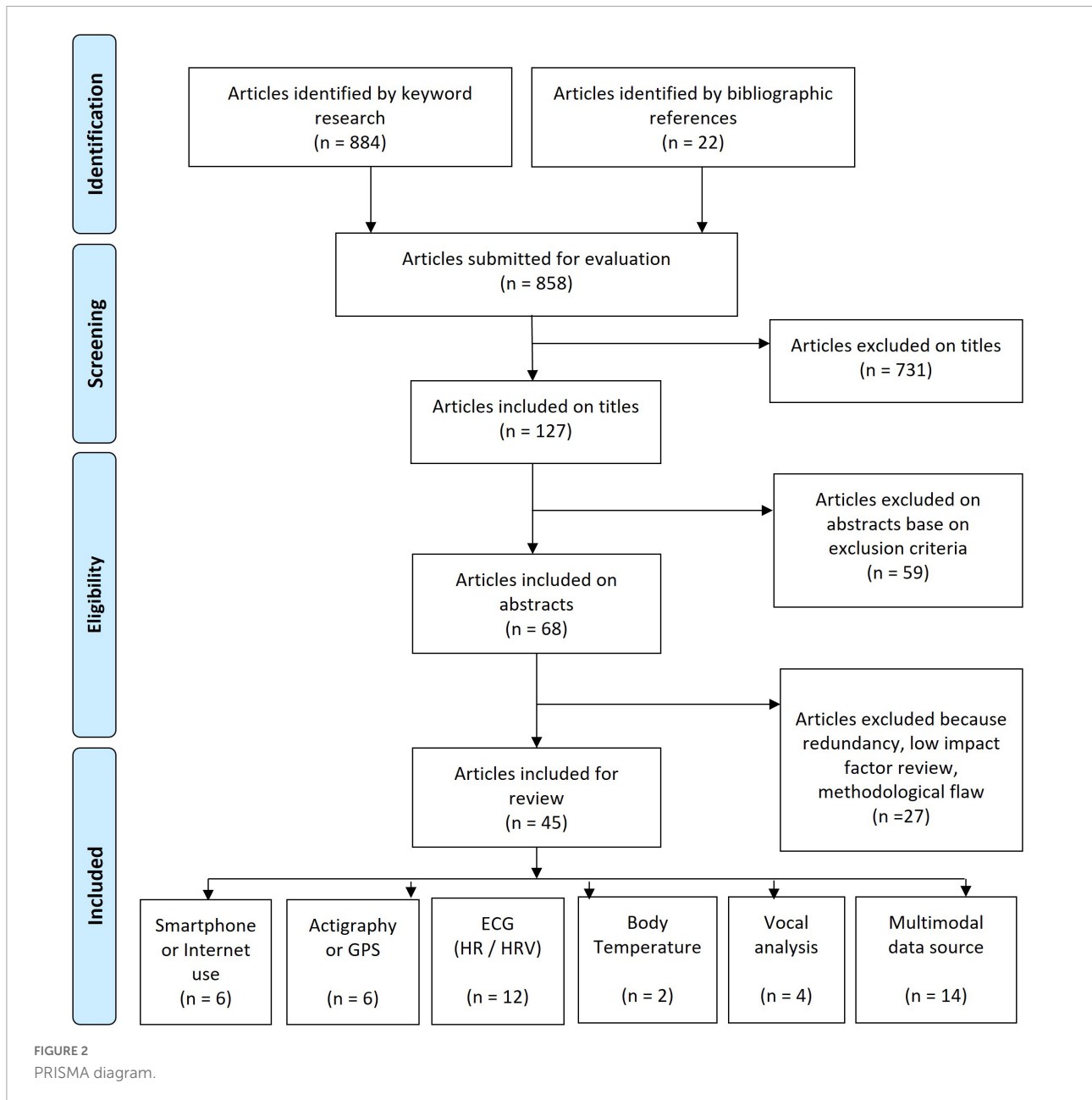
Internet use, actigraphy and GPS, electrocardiogram (ECG), voice analysis, body temperature and multimodal data source.

## Unimodal data source

Results using a unimodal data source are summarized in [Table 1](#).

### Internet use

Safa et al. (5) and Islam et al. (6) explored the potential of linguistic approaches on Twitter and Facebook. They focus on tweets or comments, and use n-gram language model, or Linguistic Inquiry and Word Count (LIWC), based on the relationships between patterns of language, first person pronouns, anger words, various negative emotions and mental disorders. They show 91 and 83% accuracy in predicting depressive symptoms, respectively. Linguistic data from Facebook comments provides the highest accuracy. Yue et al. (7) explore metadata of internet traffic on smartphones for depression screening. They develop techniques to identify internet usage sessions and create their own algorithm to predict depression correlating with the Patient Health Questionnaire 9 (PHQ9). The internet traffic data was divided into application categories (e.g., mail, social, video, audio, game, shopping,



study) with a focus on usage (total duration, number of sessions, duration during morning afternoon, and night, and screen on-off events) and they achieve a specificity up to 77% for depression prediction.

**Smartphone use  
Depressive disorders**

Opoku Asare et al. (8) studied 629 individuals assessing multiple features: battery consumption, time zone, time stamped data, foreground app usage, internet connectivity, screen lock and unlock logs with demographic information and self-reports (PHQ8). They find a positive correlation between

screen status-normalized entropy (defined as the degree of variability, complexity and randomness in behavior states, e.g., disconnection and connection states, frequency of use, etc.) and depression. But they find no exploitable association between other screen, app, and internet connectivity features. Using their best supervised machine learning, they achieved accuracy of up to 92.51%.

**Bipolar disorder**

Gillett et al. (9) showed a significant interaction between a bipolar disorder population and smartphone use. A negative correlation is highlighted between total

outgoing call frequency, total duration, total outgoing SMS messaging frequency and bipolar depressive episodes. Razavi et al. (10) confirmed that call durations and number of text messages had a negative correlation with depressive symptoms. Furthermore, using machine

learning (random forest classifier), they found an accuracy of up to 81.1% for diagnosing bipolar depression. Zulueta et al. (11) demonstrate a significant positive correlation between accelerometer displacement, average interKEY delay, session count and autocorrect rate and

TABLE 1 Unimodal data source.

|                 | Study (year)                | Disorder        | n       | Questionnaire              | Features                                       | Accuracy                     | Sensitivity | Specificity |
|-----------------|-----------------------------|-----------------|---------|----------------------------|--|------------------------------|-------------|-------------|
| Smartphone      | Safa et al. (5)             | MDD             | 1023    | LIWC                       | Twitter posts and bio-text                     | 83–91%                       | n.c.        | n.c.        |
|                 | Yue et al. (7)              | MDD             | 79      | PHQ9                       | Metadata of internet traffic                   | 80%                          | n.c.        | n.c.        |
|                 | Gillett et al. (9)          | BD/BPD/HC       | 55      | QIDS                       | Phone calls and SMS messaging                  | n.c.                         | n.c.        | n.c.        |
|                 | Razavi et al. (10)          | BD              | 412     | BDI – II                   | Phone use (calls and text messages)            | 76–81%                       | n.c.        | n.c.        |
|                 | Islam et al. (6)            | MDD             | 7145 uc | LIWC                       | Comments on Facebook                           | n.c.                         | n.c.        | n.c.        |
|                 | Zulueta et al. (11)         | BD              | 16      | HDRS/YMRS                  | Metadata of keystroke entry with accelerometer | n.c.                         | n.c.        | n.c.        |
| Actigraphy      | Difrancesco et al. (14)     | MDD/GAD         | 359     | Deutsch 30-IDS             | Sleep parameters                               | n.c.                         | n.c.        | n.c.        |
|                 | Minaeva et al. (16)         | MDD             | 179     | IDSR/CIDI/BDI-II           | Global activity                                | n.c.                         | n.c.        | n.c.        |
|                 | Jakobsen et al. (17)        | BD              | 55      | MADRS                      | Global activity                                | n.c.                         | 82%         | 84%         |
|                 | Kaufmann et al. (20)        | BD              | 131     | YMRS                       | Sleep parameters                               | n.c.                         | n.c.        | n.c.        |
|                 | Merikangas et al. (13)      | BDI/BDII<br>MDD | 242     | PHQ9                       | Global activity                                | n.c.                         | n.c.        | n.c.        |
|                 | Tonon et al. (15)           | BD              | 15      | YMRS                       | Global activity                                | n.c.                         | 71%         | 100%        |
|                 | Zhang et al. (12)           | MDD             | 308     | PHQ8                       | Bluetooth features                             | n.c.                         | n.c.        | n.c.        |
| HR or HRV       | Gregório et al. (33)        | HP/BD           | 36      | MINI/BrMaS                 | Heart parameters                               | n.c.                         | n.c.        | n.c.        |
|                 | Ortiz et al. (30)           | BDI/BDII        | 53      | IBI/SADSL/MADRS/<br>YMRS   | Heart parameters                               | n.c.                         | n.c.        | n.c.        |
|                 | Brugnera et al. (28)        | HP              | 65      | BDI II                     | Heart parameters during stress protocol        | n.c.                         | n.c.        | n.c.        |
|                 | Byun et al. (25)            | HP/MDD          | 78      | HAMD                       | Heart parameters during stress protocol        | 74%                          | 73%         | 75.6%       |
|                 | Byun et al. (26)            | MDD             | 66      | HAMD                       | Heart parameters during stress protocol        | 70%                          | 64%         | 76%         |
|                 | Hartmann et al. (24)        | HP/MDD          | 127     | HDRS 17                    | Heart parameters                               | n.c.                         | n.c.        | n.c.        |
|                 | Lesnewich et al. (23)       | HP              | 152     | BDI II                     | Heart parameters                               | n.c.                         | n.c.        | n.c.        |
|                 | Faurholt-Jepsen et al. (17) | BD              | 16      | HDRS17/YMRS                | Heart parameters                               | n.c.                         | n.c.        | n.c.        |
|                 | Wazen et al. (31)           | BDI             | 19      | MINI/BRMS                  | Heart parameters during hospitalization        | n.c.                         | n.c.        | n.c.        |
|                 | Carnevali et al. (21)       | HP              | 42      | RRS                        | Heart parameters                               | n.c.                         | n.c.        | n.c.        |
|                 | Chen et al. (17)            | HP/MDD          | 80      | No                         | Heart parameters during stress protocol        | n.c.                         | n.c.        | n.c.        |
|                 | Kuang et al. (27)           | MDD             | 76      | PID                        | Heart parameters during stress protocol        | 86.4%                        | 89.5%       | 84.2%       |
|                 | Temperature                 | Ma et al. (35)  | MDD/SR  | 62                         | HAMD17/PHQ9/HAMA                               | Temperature during treatment | n.c.        | n.c.        |
| Kim et al. (36) |                             | MDD             | 67      | HAMD                       | Electrodermal activity during stress protocol  | 74%                          | 74%         | 71%         |
| Voice           | Shin et al. (37)            | MDD             | 93      | MINI/BAI/HDRS/<br>BIS/PHQ9 | Voice characteristics                          | n.c.                         | 65.6%       | 66.2%       |
|                 | Weiner et al. (39)          | BP              | 56      | YMRS/QIDSC16               | Voice characteristics                          | 83–86%                       | n.c.        | n.c.        |
|                 | Weintraub et al. (40)       | BP              | 123     | LIWC                       | Emotional expression                           | 75.2–81.8%                   | 70%         | 80%         |
|                 | Zhang et al. (38)           | MDD             | n.c.    | PHQ9                       | Voice characteristics in audio files           | n.c.                         | n.c.        | n.c.        |

BAI, Beck Anxiety Inventory; BDI II, Beck Depression Inventory; BP, Bipolar Disorder; BPI, Bipolar Disorder Type I; BPII, Bipolar Disorder Type II; BrMaS, Bech-Rafaelsen Mania Scale; GAD, Generalized Anxiety Disorder; HAMD17 or HDRS, Hamilton Depression Rating Scale; HAMA, Hamilton for Anxiety; HP, Healthy Patient; IDSR, Inventory of Depressive Symptomatology (self-report); LIWC, Linguistic Inquiry and Word Count; MADRS, Montgomery Asberg Depression Rating Scale; MDD, Major Depressive Disorder; MINI, Mini-International Neuropsychiatric Interview; PHQ9, Patient Health Questionnaire 9; QIDS, Quick Inventory Depression Scale; uc, user comments; n.c., not communicated; SR, Suicidal Risk; SA, History of Suicidal Attempt; w, week; YMRS, Young Mania Rating Scale.



depressive symptomatology using the Hamilton Depression Rating Scale (HDRS).

### Actigraphy or GPS use Depressive disorders

Zhang Y. et al. (12) extracted 49 Bluetooth features including periodicity and regularity of individuals' life rhythms using Nearby Bluetooth Device Count (NBDC; detected by Bluetooth sensors in mobile phones). During a two-year follow up study with 308 patients, they tried to correlate these features with a bi-weekly PHQ8 questionnaire assessing for depression. They show that before a depressive episode, several changes were found in the preceding 2 weeks of Bluetooth data (the amount, the variance and the periodicity decreased and NBDC sequence became more irregular). Merikangas et al. (13) highlight a direct and positive correlation of sleep and energy (assessed by minute-to-minute activity counts from an actigraphy device worn on the non-dominant wrist for 2 weeks) with mood in a large nested-case control study. They found a unidirectional association between motor activity and subjective mood level and a bidirectional association between motor activity and subjective energy level and sleep duration. Difrancesco et al. (14) obtained similar results, with a negative correlation of gross motor activity (GMA) and sleep parameters with depressive symptoms. Tonon et al. (15) demonstrate the reliability of using an actigraphic strategy for evaluating the intensity of depression. They show a sensitivity up to 71% and a specificity up to 100% for evaluating melancholia in depressed patients, since nocturnal activity was significantly higher in non-melancholic patients. Minaeva et al. (16) explore a predictive model based on actigraphy and Experience Sampling Methodology (ESM). The actigraphy model was provided by the GMA and time and maximal activity level across the 24-h period. The ESM model has a fixed design, with questionnaires including items on current mood states, social interactions, daily experiences and behaviors. They found reasonable discriminative ability for the actigraphy model alone and excellent discriminative ability for both the ESM and the combined-domains model (actigraphy + ESM). These results were based on a strong correlation between depression and lower levels of physical activity.

### Bipolar disorder

Jakobsen et al. (17) used several machine-learning techniques using actigraphy and observed an 84% accuracy, sensitivity and specificity when differentiating between depressed bipolar patients and healthy individuals. Jacobson et al. (18) re-analyzed Jakobsen's data using novel methods and their machine-learning algorithm correctly predicted the diagnostic status 89% of the time. These results remain inconsistent, as observed by Freyberg et al. (19) since they found no difference between healthy controls and younger bipolar patients in activity energy expenditure, suggesting that

these outcomes could progress along with disease duration. Kaufman et al. (20) explored several sleep features using night actigraphy: total sleep time, waking after sleep onset, percent of sleep and number of awakenings. Using a ML algorithm (LASSO regression) they found that none of those features differ between bipolar patients and healthy controls. However, there is considerable variability among those with bipolar disorder. Zulueta et al. (11) demonstrate a significant correlation between accelerometer displacement and HDRS or YMRS during depressive or manic episodes.

### Heart rate/heart rate variability Depressive disorders

Carnevali et al. (21) have published a long-term follow-up study of 3 years, assessing HRV at several points (T0, 13th and 34th month). They find that resting HRV is negatively correlated with both rumination and depressive symptoms. They suggest a link between HRV at T0 and the evolution from rumination and depressive symptoms at month 13. They also conclude that a low vagal tone is a characteristic of depressive symptoms, which is consistent with Chen et al. (22), who have suggested the implication of an over-activated parasympathetic nervous system under long-term depression. Thus, a negative correlation is usually found between HRV and diagnosis of depression. But use of current antidepressants can turn this correlation into a positive one as Lesnewich et al. demonstrated in (23). Hartmann et al. (24) demonstrate that HRV could be used as a unique biomarker which could vary with the depressive symptomatology and be correlated with symptom severity. Byun et al. (25) used machine learning on 20 HRV indices and achieved an accuracy of 74.4%, a sensitivity of 73%, and a specificity of 75.6% for depression detection. In another study (26), they found a similar result, proving that entropy features of HR are lower in depressive patients in stress conditions. All those results are in line with Kuang et al. (27) who obtained an accuracy of 86.4%, sensitivity of 89.5%, and specificity of 84.2% for depression. Brugnera et al. (28) showed a significant and positive relationship with a higher resting-state HRV, but confirmed a blunted reactivity to the stress protocol. They suggest that healthy individuals with higher depressive symptoms have atypical cardiovascular responses to stressful events. Despite these results, Sarlon et al. (29) have not found any interaction between HRV and depression severity.

### Bipolar disorder

Ortiz et al. (30) showed a positive correlation between reduction of HRV and illness duration, number of depressive episodes, duration of the most severe manic/hypomanic episodes, co-morbid anxiety disorder and family history of suicide. Wazen et al. (31) observed an increase in HR and decrease in HRV in mania relative to euthymia. Faurholt-Jepsen et al. (32) found an increase of 18% of HRV in mania

state compared with a depressed state, but no significant HRV difference has been found between euthymia and the depressed stage in bipolar disorder. Gregorio et al. (33) showed non-linear HRV dynamics consistent with high sympathetic heat modulation with less vagal modulation compared to healthy controls. They show a possible reduction of sympathetic modulation after treatment, with an increased vagal or parasympathetic modulation.

## Body temperature

Lorenz et al. (34) first showed no significant temperature difference between mood disorders and healthy control groups, but after excluding antidepressant-medicated participants, they found that the depression group has lower skin temperature amplitude and a less stable skin temperature rhythm. Ma et al. (35) find a peripheral body temperature rhythm higher in major depressive disorder patients than in healthy controls. They highlighted the existence of phase delay of temperature that was greater in mood disorder patients with suicidal risk than those without suicidal risk and healthy controls. The same results can be found with patients before treatment introduction. They suggest that the body temperature anomalies can diminish with the improvement of depressive symptoms after the treatment with antidepressants. Kim et al. (36) used a machine learning approach to analyze an electrodermal activity data set and predict MD diagnoses during a stress protocol. They observed 74% accuracy and sensitivity and 71% specificity for depression.

## Vocal recording Depressive disorders

Shin et al. (37) used machine learning on data extracted from semistructured interview recordings. Previous research showed that depressed patients had simpler, lifeless voices with lower volume. Their movements of the vocal tract were slow and they spoke in low voices. Thus, the features extracted are from four aspects: namely glottal, tempo-spectral, formant, and other physical features. Glottal features included information about sound articulation with the vocal cords, obtained by parameterizing each numeric after drawing a waveform. They used it for calculating three parameters: the opening phase, closing phase, and closed phase. They extracted tempo-spectral features with the help of “librosa,” an audio processing toolkit. Temporal features refer to the time or length of the interval when participants continue an utterance. For spectral features, they used averaged spectral centroid, spectral bandwidth, roll-off frequency, and root mean square energy. Formant features are phonetic information obtained through linear prediction. The formant was considered as the resonance of the vocal tract and as the local maximum of the spectrum. The first to third formants were exploited with their bandwidths. Other physical attributes were obtained as the mean and variance of pitch and magnitude, zero crossing rate (which indicated how intensely

the voice was uttered), and the voice portions (which indicated how frequently they appeared). Silence was represented by frames with zero crossing rates below the average. This study was conducted on 93 patients in three different groups: non-depressed, major depression, and minor depression. The minor depression group had the lowest voices and more pitch changes during speech. The major depression group was between the non-depressed and the minor depression group. A multilayer processing method achieved 65.9% accuracy, 65.6% sensitivity, and 66.2% specificity for distinguishing depressed people from healthy controls. Zhang L. et al. (38) were able to predict the PHQ9 score with an AUC of 0.825 with features extracted from recorded audio of depressed subjects.

## Bipolar disorder

Weiner et al. (39) applied machine learning on verbal fluency tasks (letter, semantic, free word generation, and associational fluency data) to classify two distinct acute episodes in each of 56 patients as manic, mixed manic, depressive, or mixed depressive. They used a two-step procedure with a first one beginning by selection of single words using a voice activity detection algorithm. Then, speech features were calculated for each word. The word detection algorithm used the energy of the audio signal to analyze the temporal and spectral features. Then, they used the Camacho SWIPE algorithm according to a spectral matching procedure, which analyses signal intensity, zero crossing rate, spectral strength, and finally permitted detection of single words. Speech feature estimation was based on the estimation of specific features related to prosody and voice quality. Pauses calculated between words, word length and estimating F0 (pitch) dynamics (temporal windows of 10 milliseconds) were obtained with the SWIPE algorithm and supplied the prosodic features. For each word, they supplied estimates of F0, median, median absolute deviation. Then they extended the use of these features to all the voiced segments within each word and finally they reported the resulting features as amplitude, duration and tilt (mean of amplitude and duration). Voice quality was obtained by estimating the long-term muscular setting of the larynx and vocal tract that deviated from the neutral point which was calculated with the help of a DYPSA algorithm. They highlighted that in the mixed manic and manic groups, voice quality patterns were elevated in subjects with a high score in the Quick Inventory of Depressive Symptomatology (QIDS-c16) questionnaire. Higher score on the Young Mania Rating Scale (YMRS) questionnaire was correlated with higher median pitch, with a higher variability as expressed by the dispersion of voiced sound fundamental frequency, and with higher tilt. Given these results, they selected features for a mixed versus non-mixed and depression algorithm detection. They achieved an accuracy of 84% for discriminating depression from mixed depression, 86% for discriminating hypomania from mixed hypomania. They suggest that vocal



features *via* verbal fluency tasks could be reliable biomarkers and help to improve diagnostic accuracy. Weintraub J. et al. (40) also used machine learning to achieve an algorithm with an accuracy of at least 75.2% for detecting high or low expressed mood.

Several studies tried to demonstrate a new correlation between DP and mood disorders, while others tried to prove the reliability of predictive models using different machine learning algorithms based on the previous correlations. Only a few focused on specific symptoms of mood disorders.

## Multimodal data source

Results from multimodal data sources are summarized in (Table 2).

Meyerhoff et al. (41) studied a cohort of 282 healthy individuals who used wearable devices measuring steps, energy level, HR, and sleep change and trained a supervised machine learning algorithm to study the interaction between those passive data and a PHQ8 questionnaire. They observed changes in GPS features, exercise duration, and use of active apps before the rise of depressive symptoms, suggesting a directional correlation between changes in behaviors and subsequent changes in depressive symptoms. Bai R. et al. (42) recruited 334 patients with a major depressive disorder to a study using an app called “Mood Mirror” that allows active data and passive data collection (phone and wearable wristband) in order to classify patients between several mood states: steady remission, mood Swing (drastic or moderate), and steady depressed. They tested several combinations of data in order to achieve the best classification. The best features were passive data (1 feature from phone usage and 3 from the wearable) and achieved over 75% accuracy.

Rykov et al. (43) recruited 267 healthy people and used wearables to record physical activity, sleep patterns, circadian rhythms (CRs) for step usage, HR, energy expenditure and sleep data. They used supervised machine learning and found an accuracy of 80%, a sensitivity of 82%, and a specificity of 78% for detection of depression, but only in subsamples of depressed and healthy participants with no risk of depression (no or minimal depressive symptoms). Apart from this, the ability of the digital biomarkers to detect depression in the whole sample was limited. Nickels et al. (44) recruited 415 individuals (around 80% with MDD) and created a large specific operating system that recorded accelerometer, ambient audio, phone information, barometric air pressure, battery charge, Bluetooth, light level, network, gyroscope, physical activity level, phone calls, ping, proximity, screen state, step count, text messages, volume, and Wi-fi network. Using a subset of 34 DP features, they found that 11 features showed a significant correlation with PHQ9. They found that a more negative sentiment of the voice diary, obtained from a derived measure from a sentiment classification algorithm, is associated with a higher PHQ9 score.

Moreover, bad self-reported sleep quality, higher ambient audio level, letting the phone ring for longer periods until the call was missed, fewer different locations visited in a given week, fewer words spoken per minute, longer duration of voice diary, lower weekly mean battery percentage, the number of emojis in outgoing text messages, receiving or making more phone calls per week, and less variability in where participants spent time were correlated with higher PHQ9 score. They achieved a logistic regression model resulting in a 10 fold cross-validated mean AUC of 0.656 (SD 0.079).

Pedrelli et al. (45) observed that it was not possible to determine if one modality (smartphone, wearable or both) could outperform the others. They highlighted that the most predictive features were related to phone engagement, activity level, skin conductance and HRV, but stated that further studies are needed to increase strategy accuracy. Jacobson et al. (46) conserved a rate of accuracy to predict depression with only a 1 week recording study with 15 participants with MDD. They combined only two assessing tools with actigraphy, which records continuous movement, to estimate global activity and with ambient light exposure to estimate social activity. A deep neural network combined with SMOTE class balancing technique achieved an accuracy of 84%, a sensitivity of 82%, and a specificity of 84%. A prospective observational cohort study was performed by Cho et al. (47) on 55 patients with MDD and bipolar disorder type 1 and type 2 during 2 years using a smartphone-based EMA and a wearable activity tracker (Fitbit). They processed the digital phenotypes into 130 features based on circadian rhythms (e.g., steps before bedtime, light exposure during daytime, and HR amplitude) and performed mood classification using a random forest algorithm.

Di Matteo et al. (48) designed an Android app to collect periodic measurements including samples of ambient audio, GPS location, screen state, and light sensor data during a 2-week observational study. They found good accuracy with an AUC of 0.64.

Jacobson et al. (49) used passive sensor data, including GPS, location type based on the Google Places location type, local weather information (temperature, humidity, precipitation), light level, HR information (average HR and HRV), and outgoing phone calls, and used machine learning algorithms to correlate these data with a dynamic mood assessment using EMA. They found good accuracy, with predicted depressed mood scores that were highly correlated with the observed depressed mood scores from the models.

Narziev et al. (50) designed a Short-Term Depression Detector using EMA and various passive sensors available on a smartphone (phone calls, app usage, unlocked state, stationary state, light sensor, accelerometer, step detection) and on a smartwatch (HRM, accelerometer). They used support vector machine and random forest models to achieve group classification with an accuracy of 96.00%.

TABLE 2 Multimodal data source.

| Study (year)             | Disorder     | n   | Smart-phone | GPS or actigraphy | HR – HRV | Body temperature | Voice | Light exposure | Scale          | Accuracy |
|--------------------------|--------------|-----|-------------|-------------------|----------|------------------|-------|----------------|----------------|----------|
| Bai et al. (42)          | MDD          | 334 | Yes         | Yes               | Yes      | Yes              | Yes   | No             | PHQ9           | 76.67%   |
| Meyerhoff et al. (41)    | MDD/GAD/SAD  | 282 | Yes         | Yes               | Yes      | No               | No    | No             | PHQ8/DAS7      | n.c.     |
| Nickels et al. (44)      | MDD          | 415 | Yes         | Yes               | No       | No               | No    | Yes            | PHQ9           | n.c.     |
| Opoku Asare et al. (8)   | MDD          | 629 | Yes         | No                | No       | No               | No    | No             | PHQ8           | 98%      |
| Rykov et al. (43)        | MDD          | 267 | Yes         | Yes               | Yes      | No               | No    | No             | PHQ9           | 80%      |
| Sarlon et al. (29)       | MDD          | 89  | No          | No                | Yes      | Yes              | No    | No             | BDI II         | No       |
| Di Matteo et al. (48)    | Gen pop      | 112 | Yes         | Yes               | No       | No               | Yes   | Yes            | PHQ8           | n.c.     |
| Jacobson and Chung, (49) | MDD          | 31  | Yes         | Yes               | Yes      | No               | No    | Yes            | PANAS-X        | n.c.     |
| Narziev et al. (50)      | MDD          | 20  | Yes         | Yes               | Yes      | No               | No    | Yes            | PHQ9           | 96%      |
| Freyberg et al. (19)     | BD           | 60  | No          | Yes               | Yes      | No               | No    | No             | HDRS17         | n.c.     |
| Pedrelli et al. (45)     | MDD          | 31  | Yes         | Yes               | Yes      | No               | No    | No             | HDRS17         | n.c.     |
| Cho et al. (47)          | MDD/BDI/BDII | 55  | Yes         | Yes               | Yes      | No               | No    | Yes            | Mood chart app | 85/94%   |
| Jacobson et al. (46)     | BDII         | 15  | No          | Yes               | No       | No               | No    | Yes            | MADRS          | 84%      |
| Lorenz et al. (34)       | MDD          | 242 | No          | Yes               | Yes      | Yes              | No    | No             | CES D score    | n.c.     |

BDI, Bipolar Disorder Type 1; BDII, Bipolar Disorder Type 2; BDI II, Beck Depression Inventory; CES D, Centre for Epidemiological Studies-Depression; DAS7, Dyadic Adjustment Scale; GAD, General Anxiety Disorder; Gen pop, General population; HDRS, Hamilton Depression Rating Scale; HP, Healthy People; MADRS, Montgomery Asberg Depression Rating Scale; MDD, Major Depressive Disorder; n.c., not communicated; PANAS-X, Positive and Negative Affect Schedule Expanded; PHQ9, Patient Health Questionnaire 9; SAD, Social Anxiety Disorder; w, week.

## Discussion

### Principal results

Digital phenotypes compute the clinical characteristics specific to various mental states, sometimes with better precision than a clinician can achieve, and with the possibility of doing it remotely. **Table 3** summarizes the most relevant features according to this review and our previous one (4).

The unimodal data source type is the first and the most common type of study. As we can observe in **Table 1**, we have as many studies focused on the new correlation demonstration as new predictive models with different levels of precision. It can be explained by a faster and easier protocol, with more innovative possibilities. Results in unimodal data sources have an accuracy around 75% with a sensitivity range between 64 and 91%, and a specificity between 66.2 and 100%. Studies focusing on smartphone use are the most dynamic unimodal approach, providing disruptive ways to a better understanding of mood disorders. The tools which are used for that purpose are multiple: call frequency, SMS message frequency, SMS message length, keystroke entry date, accelerometer displacement, activity tracker *via* Bluetooth, usage sessions from internet traffic, and linguistic analyses on Twitter and Facebook posts with the help of EMA provided by smartphone apps. In most cases, applications are developed specially for each study. ECG recording is the most common type of wearable found in this review. Features were Resting HR, HRV, Respiration Rate, high-frequency HRV, low-frequency HRV, and root mean square of the successive differences, with experimental exposition using a protocol for assessing the

autonomic responses to stress and recovery. Body temperature recording is also a promising source. This extracts global body temperatures using a Holter monitor that detects 24-h peripheral body temperature. Several studies combined temperature with actigraphy, making it possible to categorize body temperature throughout the day's phases, and adjust the amplitude with activity. While promising, voice analysis still remains one of the poorest study domains. The data provided come from speech text provided by smartphone and use of acoustic data automatically recorded. Other metrics can also be calculated: total speech activity, the proportion of speech during a day, analysis of voice and linguistic patterns, GPS location provided by environmental sounds, and verbal fluency tasks (letter, semantic, free word generation, and associational fluency). Activity of humans in daily life is often represented through GPS location or actigraphy on wearables that can be used to study physical activity, recording GMA, and sleep quality with CR. This technique helps us to understand the link between state of health, activity and sleep conditions. They try to implement reliable values such as total sleep time, wake after sleep onset, percent sleep and number of awakenings, sleep latency, sleep efficiency, and relative amplitude between daytime and night-time activity.

The multimodal data source approach uses combinations of the research presented from the unimodal data. We can observe that in most cases, it is focused on the development of new predictive models for diagnosing mood disorders. Therefore, in most cases it uses machine learning for training new helpful algorithms for increased diagnostic accuracy. Different models are proposed to achieve this goal. However, multimodal data sources, despite the promising possibilities of

this technique, have hardly succeeded in demonstrating new disruptive comprehension of the DP of mood disorders.

Beyond the aspects mentioned in this article, the digital phenotype offers interesting perspectives for treatment and clinical research, in particular for the dimensional approach to mental disorders. Because mood disorders manifest themselves

in a heterogeneous way on the different components of the mind, psychiatric classifications have often been criticized, particularly for their validity, and more contemporary approaches attempt to increase their reliability by using more integrative approaches, the most successful of which is the RDoC matrix (51). This model proposed by the NIMH

TABLE 3 Digital phenotype of features relevant to mood disorders.

|             | MDD or bipolar depression   | Mania   |
|-------------|---|---|
| Actigraphy  | Decreased daytime activities<br>Decreased walking<br>Nearby Bluetooth Device Count: Number, variance and periodicity are decreased and NBDC sequence becomes more irregular<br>Nocturnal activity is significantly higher in non-melancholic patients<br>Unidirectional association between motor activity and subjective mood level and a bidirectional association between motor activity and subjective energy level or sleep duration.<br>Negative correlation between gross motor activity and sleep parameters with depressive symptoms<br>Lower level of activity<br>Late start of activities<br>Peak activity at noon<br>Poor evening activities  | Increase in activities<br>Increase in number of locations   |
| HR and HRV  | Severity-dependent decrease in HRV<br>Increase in low-frequency HRV and low-frequency/high-frequency ratio<br>Decreased high-frequency HRV<br>Higher heart rate and lower HRV in bipolar than in unipolar depression<br>Increase in the low-frequency/high-frequency ratio and decrease in high-frequency HRV in BPII<br>Resting HRV is negatively correlated with both rumination and depressive symptoms<br>Reduction of HRV is correlated with illness severity  | Decreased RR interval (increased HR), variance, low-frequency HRV, and high-frequency HRV;<br>Increase in the low-frequency/high-frequency ratio<br>Increased and decreased HRV were both found |
| Temperature | Decrease in temperature<br>Night-time temperature increase<br>Decreased amplitude<br>Phase advance<br>Depression group has a lower skin temperature amplitude and a less-stable skin temperature rhythm<br>Body temperature rhythm higher in major depressive disorder patients than healthy<br>Body temperature anomalies can diminish with the improvement of depressive symptoms   | n.c.  |
| Smartphone  | Decrease in smartphone use (number of SMS messages, number of calls)<br>Change in the duration of calls<br>Linguistic data: patterns of language, first person pronouns, anger words, various negative emotions expressed on Twitter or Facebook<br>Internet traffic data (mail, social, video, audio, game, shopping, study): total duration, number of sessions, duration in morning, afternoon, and midnight, and screen on-off events<br>Smartphone use: outgoing call frequency, total duration, total outgoing SMS messaging frequency, call durations, number of text messages, average interKEY delay, session count and autocorrect rate<br>Positive correlation between screen status-normalized entropy and depression | Number of calls increased;<br>Increased call duration;<br>Number of messages increased  |
| Voice       | Increased response latency;<br>Number of breaks and length of breaks increased  | Reduced number of breaks<br>Increased verbal fluency  |
| Multimodal  | The most predictive features were related to phone engagement, activity level, skin conductance, and heart rate variability   |   |

MDD, Major Depressive Disorder; HRV, Heart Rate Variability; HR, Heart Rate; SMS, Short Message Service; NBDC, Nearby Bluetooth Device Count.

offers a relevant framework to better exploit the links existing between clinical and biological markers from a dimensional perspective. The numerous passive and active markers of DP make it possible to collect information specifically related to each clinical dimension (emotional and affective, cognitive, conative, and physical) and therefore to propose a more detailed evaluation of semiology. In this perspective, Torous et al. (3) state that it will thus be possible to collect preclinical information at the community level, thus establishing a reference level of the psychological dimensions of the general population and to compare them with the pathological variations observed in clinical populations. We suppose that it will be possible to determine predictive patterns for the occurrence of a clinical episode; in short, to make predictions.

From a therapeutic point of view, use by clinical dimension therefore makes it possible to offer, in real time, more specific and therefore personalized interventions. For example, in a recent article we have listed the interventions more specific to the conative dimension (which refers to goal-directed behavior intentionally based on motivational factors), which can be provided in support of active and passive DP data (52).

Finally, as mentioned above, DP can be an effective tool for bias mitigation. Indeed, the processes related to decision-making can benefit at several levels from the data collected, both active and passive. For illustration, the passive collection of data, less subject to the subjectivity of the patient and the clinician, makes it possible to have a more efficient control of certain cognitive human constraints (recall bias for example).

## Limitations

Some limitations have been found in the review. The majority of apps are available on Android, but only a few on iOS. That means that data collected could introduce a new type of bias based on social grounds. A significant number of studies show an over-representation of the female population. Large cohort studies remain the exception to the rule, and the ability of digital phenotype to be widely and easily deployed has remained untapped. That could be explained by the legal procedures for protection of privacy, which differ significantly between countries. Therefore, there is an overrepresentation of the American or Asian population in digital phenotype studies.

The majority of the recent studies focused on smartphone use, HRV, GPS and multiple wearable strategies. Multimodal strategies can be observed to present better reliability than the unimodal approach. Questionnaires used by studies are different, making comparisons more difficult, although use of the PHQ9 (or PHQ8) questionnaire is common. Further exploration could be achieved through other technologies, bringing more accuracy with less recording, which is logically more efficient.

## Comparison with prior work

Most authors recommend the use of passive data preferentially to active data in the context of bipolar disorder because this type of automatically generated data makes it possible to limit bias and limit the feeling of intrusion that self-questionnaires can cause (especially if they must be filled in regularly or if they appear in a “pop-up”). Some authors emphasize the limitations of the actual DSM-5 approach, which is based on clinical statistical observation. As reported above, the digital phenotype provides new insight into the classification of disorders from a behavioral perspective.

The concept of a digital phenotype, which is materially supported by a technological tool that is theoretically accessible at any time, thus makes it possible to track an individual's behavioral change processes over a more sustained period of time compared to periodic visits to the practitioner (53). However, it is not recommended for all clinical approaches and as reported by Patoz et al. (54), many studies suggest that these applications would be more appropriate for mild and moderate stages of depression. Indeed, their use in severe mood disorders is potentially limited.

We have also emphasized the appeal of passive data, which encourages the involvement and commitment of subjects more easily; for example, it requires little effort to make an inventory of one's symptoms. O'Brien and Toms (55) point out that engagement is not a static process, but a multi-stage process: one starts at the point of engagement, then comes a period of engagement, and finally it is possible to encounter a point of disengagement and a period of re-engagement. Finally, in the absence of symptoms, the feeling of need for care may diminish and thus cause the subject to disengage. These tools allow for a more detailed and personalized behavioral follow-up, and therefore to propose effective corrective actions, such as feedback screen for promoting the reward dimension (56).

## Conclusion

Ultimately, it appears that the digital phenotype of a pathology is the computer translation of objectifiable signs of mental illness, and it should therefore be understood as a means of strengthening the observation capacities of psychiatrists. Regarding depressive disorders, the main elements are the decrease in functional and biological parameters (decrease in activity and walking, decrease in the number of calls and SMS messages, decrease in temperature and HRV) while the manic phase results in the reverse phenomenon (increase in activity, number of calls, and HRV) as one would expect. As of now, most of the studies have focused on one tool, with significant accuracy. But there still remains lack of evidence of the usability of these technologies for long-term follow-up.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

AB and SM: conceptualization. AB and RM: methodology. FF and BM: validation. AO, AB, and RM: writing – original draft preparation. AB, AO, and VA: writing – review and editing. BS and OB: visualization. SM and AB: supervision. FF, OB, and BM: project administration. All authors have read and agreed to the published version of the manuscript.

## References

- Pasztor E, Dawkins, R. 1989. (original edition 1982; reprinted 1983, 1987, 1988, 1989): the extended phenotype. Oxford Paperbacks, Oxford University Press, Oxford, UK. 307 Pp. F6.95. *J Evol Biol.* (1991) 4:161–161. doi: 10.1046/j.1420-9101.1991.4010161.x
- Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. *Nat Biotechnol.* (2015) 33:462–3. doi: 10.1038/nbt.3223
- Torous J, Onnela J-P, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. *Transl Psychiatry.* (2017) 7:e1053–1053. doi: 10.1038/tp.2017.25
- Bourla A, Ferreri F, Ogorzelec L, Guinchard C, Mouchabac S. Évaluation des troubles thymiques par l'étude des données passives : le concept de phénotype digital à l'épreuve de la culture de métier de psychiatre. *L'Encéphale.* (2018) 44:168–75. doi: 10.1016/j.encep.2017.07.007
- Safa R, Bayat P, Moghtader L. Automatic detection of depression symptoms in twitter using multimodal analysis. *J Supercomput.* (2021):1–36. doi: 10.1007/s11227-021-04040-8
- Islam MR, Kabir MA, Ahmed A, Kamal ARM, Wang H, Ulhaq A. Depression detection from social network data using machine learning techniques. *Health Inf Sci Syst.* (2018) 6:8. doi: 10.1007/s13755-018-0046-0
- Yue C, Ware S, Morillo R, Lu J, Shang C, Bi J, et al. Automatic depression prediction using internet traffic characteristics on smartphones. *Smart Health.* (2020) 18:100137. doi: 10.1016/j.smhl.2020.100137
- Opoku Asare K, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D. Predicting depression from smartphone behavioural markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. *JMIR MHealth UHealth* (2021) 9:e26540. doi: 10.2196/26540
- Gillett G, McGowan NM, Palmius N, Bilderbeck AC, Goodwin GM, Saunders KEA. Digital communication biomarkers of mood and diagnosis in borderline personality disorder, bipolar disorder, and healthy control populations. *Front Psychiatry.* (2021) 12:610457. doi: 10.3389/fpsy.2021.610457
- Razavi R, Gharipour A, Gharipour M. Depression screening using mobile phone usage metadata: a machine learning approach. *J Am Med Inform Assoc.* (2020) 27:522–30. doi: 10.1093/jamia/oc221
- Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: a biaffected digital phenotyping study. *J Med Internet Res.* (2018) 20:e241. doi: 10.2196/jmir.9775
- Zhang Y, Folarin AA, Sun S, Cummins N, Ranjan Y, Rashid Z, et al. Predicting depressive symptom severity through individuals' nearby bluetooth device count data collected by mobile phones: preliminary longitudinal study. *JMIR MHealth UHealth.* (2021) 9:e29840. doi: 10.2196/29840
- 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,

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- energy, mood, and sleep in adults with bipolar disorder. *JAMA Psychiatry.* (2019) 76:190. doi: 10.1001/jamapsychiatry.2018.3546
- Difrancesco S, Penninx BWJH, Riese H, Giltay EJ, Lamers F. The role of depressive symptoms and symptom dimensions in actigraphy-assessed sleep, circadian rhythm, and physical activity. *Psychol Med.* (2021):1–7. doi: 10.1017/S0033291720004870
  - Tonon AC, Fuchs DFP, Barbosa Gomes W, Levandovski R, Pio de Almeida Fleck M, Hidalgo MPL, et al. Nocturnal motor activity and light exposure: objective actigraphy-based marks of melancholic and non-melancholic depressive disorder. Brief report. *Psychiatry Res.* (2017) 258:587–90. doi: 10.1016/j.psychres.2017.08.025
  - Minaeva O, Riese H, Lamers F, Antypa N, Wichers M, Booij SH. Screening for depression in daily life: development and external validation of a prediction model based on actigraphy and experience sampling method. *J Med Internet Res.* (2020) 22:e22634. doi: 10.2196/22634
  - 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) 15:e0231995. doi: 10.1371/journal.pone.0231995
  - Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. *NPJ Digit Med.* (2019) 2:3. doi: 10.1038/s41746-019-0078-0
  - Freyberg J, Brage S, Kessing LV, Faurholt-Jepsen M. Differences in psychomotor activity and heart rate variability in patients with newly diagnosed bipolar disorder, unaffected relatives, and healthy individuals. *J Affect Disord.* (2020) 266:30–6. doi: 10.1016/j.jad.2020.01.110
  - Kaufmann CN, Lee EE, Wing D, Sutherland AN, Christensen C, Ancoli-Israel S, et al. Correlates of poor sleep based upon wrist actigraphy data in bipolar disorder. *J Psychiatr Res.* (2021) 141:385–9. doi: 10.1016/j.jpsychires.2021.06.038
  - Carnevali L, Thayer JF, Brosschot JF, Ottaviani C. Heart rate variability mediates the link between rumination and depressive symptoms: a longitudinal study. *Int J Psychophysiol.* (2018) 131:131–8. doi: 10.1016/j.ijpsycho.2017.11.002
  - Chen X, Yang R, Kuang D, Zhang L, Lv R, Huang X, et al. Heart rate variability in patients with major depression disorder during a clinical autonomic test. *Psychiatry Res.* (2017) 256:207–11. doi: 10.1016/j.psychres.2017.06.041
  - Lesnewich LM, Conway FN, Buckman JF, Brush CJ, Ehmann PJ, Eddie D, et al. Associations of depression severity with heart rate and heart rate variability in young adults across normative and clinical populations. *Int J Psychophysiol.* (2019) 142:57–65. doi: 10.1016/j.ijpsycho.2019.06.005
  - Hartmann R, Schmidt FM, Sander C, Hegerl U. Heart rate variability as indicator of clinical state in depression. *Front Psychiatry.* (2019) 9:735. doi: 10.3389/fpsy.2018.00735
  - Byun S, Kim AY, Jang EH, Kim S, Choi KW, Yu HY, et al. Detection of major depressive disorder from linear and nonlinear heart rate variability features



- during mental task protocol. *Comput Biol Med.* (2019) 112:103381. doi: 10.1016/j.combiomed.2019.103381
26. Byun S, Kim AY, Jang EH, Kim S, Choi KW, Yu HY, et al. Entropy analysis of heart rate variability and its application to recognize major depressive disorder: a pilot study. *Technol Health Care.* (2019) 27:407–24. doi: 10.3233/THC-199037
27. Kuang D, Yang R, Chen X, Lao G, Wu F, Huang X, et al. Depression recognition according to heart rate variability using bayesian networks. *J Psychiatr Res.* (2017) 95:282–7. doi: 10.1016/j.jpsychires.2017.09.012
28. Brugnera A, Zarbo C, Tarvainen MP, Carlucci S, Tasca GA, Adorni R, et al. Higher levels of depressive symptoms are associated with increased resting-state heart rate variability and blunted reactivity to a laboratory stress task among healthy adults. *Appl Psychophysiol Biofeedback.* (2019) 44:221–34. doi: 10.1007/s10484-019-09437-z
29. Sarlon J, Staniloiu A, Kordon A. Heart rate variability changes in patients with major depressive disorder: related to confounding factors, not to symptom severity? *Front Neurosci.* (2021) 15:675624. doi: 10.3389/fnins.2021.675624
30. Ortiz A, Bradler K, Moorti P, MacLean S, Husain MI, Sanches M, et al. Reduced heart rate variability is associated with higher illness burden in bipolar disorder. *J Psychosom Res.* (2021) 145:110478. doi: 10.1016/j.jpsychores.2021.110478
31. Wazen GLL, Gregório ML, Kemp AH, de Godoy MF. Heart rate variability in patients with bipolar disorder: from mania to euthymia. *J Psychiatr Res.* (2018) 99:33–8. doi: 10.1016/j.jpsychires.2018.01.008
32. Faurholt-Jepsen M, Brage S, Kessing LV, Munkholm K. State-related differences in heart rate variability in bipolar disorder. *J Psychiatr Res.* (2017) 84:169–73. doi: 10.1016/j.jpsychires.2016.10.005
33. Gregório ML, Wazen GLL, Kemp AH, Milan-Mattos JC, Porta A, Catai AM, et al. Non-linear analysis of the heart rate variability in characterization of manic and euthymic phases of bipolar disorder. *J Affect Disord.* (2020) 275:136–44. doi: 10.1016/j.jad.2020.07.012
34. Lorenz N, Spada J, Sander C, Riedel-Heller SG, Hegerl U. Circadian skin temperature rhythms, circadian activity rhythms and sleep in individuals with self-reported depressive symptoms. *J Psychiatr Res.* (2019) 117:38–44. doi: 10.1016/j.jpsychires.2019.06.022
35. Ma X, Cao J, Zheng H, Mei X, Wang M, Wang H, et al. Peripheral body temperature rhythm is associated with suicide risk in major depressive disorder: a case-control study. *Gen Psychiatry.* (2021) 34:e100219. doi: 10.1136/gpsych-2020-100219
36. 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) 8:17030. doi: 10.1038/s41598-018-35147-3
37. Shin D, Cho WI, Park CHK, Rhee SJ, Kim MJ, Lee H, et al. Detection of minor and major depression through voice as a biomarker using machine learning. *J Clin Med.* (2021) 10:3046. doi: 10.3390/jcm10143046
38. Zhang L, Duvvuri R, Chandra KKL, Nguyen T, Ghomi RH. Automated voice biomarkers for depression symptoms using an online cross-sectional data collection initiative. *Depress Anxiety.* (2020) 37:657–69. doi: 10.1002/da.23020
39. Weiner L, Guidi A, Doignon-Camus N, Giersch A, Bertschy G, Vanello N. Vocal features obtained through automated methods in verbal fluency tasks can aid the identification of mixed episodes in bipolar disorder. *Transl Psychiatry.* (2021) 11:415. doi: 10.1038/s41398-021-01535-z
40. Weintraub MJ, Posta F, Arevian AC, Miklowitz DJ. Using machine learning analyses of speech to classify levels of expressed emotion in parents of youth with mood disorders. *J Psychiatr Res.* (2021) 136:39–46. doi: 10.1016/j.jpsychires.2021.01.019
41. Meyerhoff J, Liu T, Kording KP, Ungar LH, Kaiser SM, Karr CJ, et al. Evaluation of changes in depression, anxiety, and social anxiety using smartphone sensor features: longitudinal cohort study. *J Med Internet Res.* (2021) 23:e22844. doi: 10.2196/22844
42. 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 multicentre study. *JMIR MHealth UHealth.* (2021) 9:e24365. doi: 10.2196/24365
43. Rykov Y, Thach T-Q, Bojic I, Christopoulos G, Car J. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modelling. *JMIR MHealth UHealth.* (2021) 9:e24872. doi: 10.2196/24872
44. Nickels S, Edwards MD, Poole SF, Winter D, Gronsbell J, Rozenkrants B, et al. Toward a mobile platform for real-world digital measurement of depression: user-centred design, data quality, and behavioral and clinical modelling. *JMIR Ment Health.* (2021) 8:e27589. doi: 10.2196/27589
45. 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) 11:584711. doi: 10.3389/fpsy.2020.584711
46. Jacobson NC, Weingarden H, Wilhelm S. Using digital phenotyping to accurately detect depression severity. *J Nerv Ment Dis.* (2019) 207:893–6. doi: 10.1097/NMD.0000000000001042
47. Cho C-H, Lee T, Kim M-G, In HP, Kim L, Lee H-J. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. *J Med Internet Res.* (2019) 21:e11029. doi: 10.2196/11029
48. Di Matteo D, Fotinos K, Lokuge S, Mason G, Sternat T, Katzman MA, et al. Automated screening for social anxiety, generalized anxiety, and depression from objective smartphone-collected data: cross-sectional study. *J Med Internet Res.* (2021) 23:e28918. doi: 10.2196/28918
49. Jacobson NC, Chung YJ. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. *Sensors (Basel).* (2020) 20:3572. doi: 10.3390/s20123572
50. Narziev N, Goh H, Toshnazarov K, Lee SA, Chung KM, Noh Y. STDD: Short-term depression detection with passive sensing. *Sensors (Basel).* (2020) 20:1396. doi: 10.3390/s20051396
51. Cuthbert BN. Research domain criteria (RDoC): progress and potential. *Curr Dir Psychol Sci.* (2022) 31:107–14. doi: 10.1177/09637214211051363
52. Mouchabac S, Maatoug R, Conejero I, Adrien V, Bonnot O, Millet B, et al. In search of digital dopamine: how apps can motivate depressed patients, a review and conceptual analysis. *Brain Sci.* (2021) 11:1454. doi: 10.3390/brainsci11111454
53. Patoz M-C, Hidalgo-Mazzei D, Blanc O, Verdolini N, Pacchiarotti I, Murru A, et al. Patient and physician perspectives of a smartphone application for depression: a qualitative study. *BMC Psychiatry.* (2021) 21:65. doi: 10.1186/s12888-021-03064-x
54. Agarwal A, Patel M. Prescribing behaviour change: opportunities and challenges for clinicians to embrace digital and mobile health. *JMIR MHealth UHealth.* (2020) 8:e17281. doi: 10.2196/17281
55. Holdener M, Gut A, Angerer A. Applicability of the user engagement scale to mobile health: a survey-based quantitative study. *JMIR MHealth UHealth.* (2020) 8:e13244. doi: 10.2196/13244
56. O'Brien H, Cairns P editors. *Why Engagement Matters.* Cham: Springer International Publishing (2016).