

Original Paper

Investigating Smartphone-Based Sensing Features for Depression Severity Prediction: Observation Study

Yannik Terhorst^{1,2,3}, MSc; Eva-Maria Messner¹, PhD; Kennedy Opoku Asare⁴, PhD; Christian Montag⁵, Prof Dr; Christopher Kannen⁵, MSc; Harald Baumeister¹, Prof Dr

¹Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education, Ulm University, Ulm, Germany

²Department of Psychology, LMU Munich, Munich, Germany

³German Center for Mental Health (DZPG), Partner Site Munich-Augsburg, Munich, Germany

⁴Center for Ubiquitous Computing, University of Oulu, Oulu, Finland

⁵Department of Molecular Psychology, Institute of Psychology and Education, Ulm University, Ulm, Germany

Corresponding Author:

Yannik Terhorst, MSc

Department of Clinical Psychology and Psychotherapy

Institute of Psychology and Education

Ulm University

Lise-Meitner-Str. 16

Ulm, 89081

Germany

Phone: 49 8921805057

Email: yannik.terhorst@psy.lmu.de

Abstract

Background: Unobtrusively collected objective sensor data from everyday devices like smartphones provide a novel paradigm to infer mental health symptoms. This process, called smart sensing, allows a fine-grained assessment of various features (eg, time spent at home based on the GPS sensor). Based on its prevalence and impact, depression is a promising target for smart sensing. However, currently, it is unclear which sensor-based features should be used in depression severity prediction and if they hold an incremental benefit over established fine-grained assessments like the ecological momentary assessment (EMA).

Objective: The aim of this study was to investigate various features based on the smartphone screen, app usage, and call sensor alongside EMA to infer depression severity. Bivariate, cluster-wise, and cluster-combined analyses were conducted to determine the incremental benefit of smart sensing features compared to each other and EMA in parsimonious regression models for depression severity.

Methods: In this exploratory observational study, participants were recruited from the general population. Participants needed to be 18 years of age, provide written informed consent, and own an Android-based smartphone. Sensor data and EMA were collected via the INSIGHTS app. Depression severity was assessed using the 8-item Patient Health Questionnaire. Missing data were handled by multiple imputations. Correlation analyses were conducted for bivariate associations; stepwise linear regression analyses were used to find the best prediction models for depression severity. Models were compared by adjusted R^2 . All analyses were pooled across the imputed datasets according to Rubin's rule.

Results: A total of 107 participants were included in the study. Ages ranged from 18 to 56 (mean 22.81, SD 7.32) years, and 78% of the participants identified as female. Depression severity was subclinical on average (mean 5.82, SD 4.44; Patient Health Questionnaire score ≥ 10 : 18.7%). Small to medium correlations were found for depression severity and EMA (eg, valence: $r = -0.55$, 95% CI -0.67 to -0.41), and there were small correlations with sensing features (eg, screen duration: $r = 0.37$, 95% CI 0.20 to 0.53). EMA features could explain 35.28% (95% CI 20.73% to 49.64%) of variance and sensing features (adjusted $R^2 = 20.45\%$, 95% CI 7.81% to 35.59%). The best regression model contained EMA and sensing features ($R^2 = 45.15\%$, 95% CI 30.39% to 58.53%).

Conclusions: Our findings underline the potential of smart sensing and EMA to infer depression severity as isolated paradigms and when combined. Although these could become important parts of clinical decision support systems for depression diagnostics and treatment in the future, confirmatory studies are needed before they can be applied to routine care. Furthermore, privacy, ethical, and acceptance issues need to be addressed.

KEYWORDS

smart sensing; digital phenotyping; depression; observation study; smartphone; mHealth; mobile health; app; mental health; symptoms; assessments

Introduction

Depression is associated with high personal burden, impaired social participation and functioning, increased mortality, and high economic burden [1-3]. In 2020, depression was one of the leading causes of disability-adjusted life years worldwide (49.4 million; 95% CI 33.6-68.7) and is expected to be the leading cause by 2030 [1,3,4]. Despite its severity and the existence of effective treatments for depression [5-7], only 41.8% of people with major depressive disorder (MDD) receive any mental health services, and less than 30% of people with MDD receive adequate treatment [1,8]. Although several barriers contribute to this issue (eg, availability, accessibility, and acceptability of treatment) [1,8,9], a fundamental prerequisite to any health services is a timely and accurate diagnosis of MDD, or more generally, the assessment of depression severity to initiate an informed treatment process [10-14].

Well-established assessments like structured clinical interviews are often not feasible to be conducted in primary care or for preventive screening purposes (eg, due to time pressure or limited availability of qualified personnel) [10-13,15]. Furthermore, if they are implemented, they typically take place at a fixed time point and assess symptoms retrospectively, which makes them subject to several biases (eg, recall bias) and unable to assess the dynamic and fluctuating nature of mental health [16-18]. Hence, novel diagnostic approaches, which can be easily integrated into daily living to monitor depression severity with high ecological validity, could make an important contribution to improve and augment current diagnostic procedures for depression—particularly if they provide fine-grained insights into mental health symptomology (eg, on daily level). Given the omnipresence of smartphones in everyday life, the unobtrusive collection of objective sensor data (eg, total haversine distance between tracked GPS coordinates, social contacts, screen, and app usage) might be a promising step toward improved diagnoses [19-21]. This process is also referred to as smart sensing (also known as mobile sensing or digital phenotyping) in the context of depression [19-22]. Applications of smart sensing range from supporting initial diagnosis to integration during treatment (eg, just-in-time-adaptive interventions) or after treatment (eg, just-in-time interventions in relapse prevention) [19,23].

In the field of depression so far, many studies followed a classificatory approach classifying persons as depressed or not depressed (eg, based on the 8-item Patient Health Questionnaire [PHQ-8] cutoff ≥ 10) [24-28]. For instance, a first meta-analysis on supervised machine learning models to predict depression status based on wearable data reported an average accuracy of 0.89, 95% CI 0.83-0.93 (sensitivity: 0.87, 95% CI 0.79-0.92; specificity: 0.93, 95% CI 0.87-0.97) [24]. However, such a classificatory understanding of depression seems questionable when looking at (1) the poor agreement of domain experts in

the diagnosis of depression [29,30], (2) the heterogeneity of symptom networks in depressed patients [31,32], and (3) the profound evidence for depression and general psychopathology being a continuous spectrum [29,33]. Hence, studies operationalizing depression as a continuous spectrum and understanding the prediction as a regression instead of a classification problem are highly needed in the field.

For example, for GPS features (eg, total distance, number of significant places), a meta-analysis shows that robust correlations between sensing features and depression severity as continuous dimension exist (eg, distance: $r=-0.25$, 95% CI -0.29 to -0.21), time spent home: $r=0.10$ (95% CI $0-0.19$), or normalized entropy: $r=-0.17$, 95% CI -0.29 to -0.04) [34]. Besides, initial studies highlight the potential of features obtained from the screen (eg, smartphone usage duration), app (app usage), and call (eg, number of incoming calls) sensors [35,36]. However, so far, analyses are often limited to bivariate correlations and do not extend to (1) the variance in depression severity, which can be explained by the features, and (2) the combination of multiple features and which incremental benefit they provide (eg, in explained variance) [35].

Against this background, this study aimed to extend the evidence for various sensor modalities collected via the smartphone (ie, screen features, app usage features, location or GPS features, and call features) by (1) investigating bivariate correlations, (2) exploring the explained variance and incremental benefit of features in cluster-wise regression models (eg, limited to location features), and (3) cluster-combined regression models. Accordingly, the following research questions will be answered:

1. Which bivariate correlations are present between depression severity and (a) screen, (b) app, (c) location, and (d) call features?
2. How much variance in depression severity can be explained in parsimonious cluster-wise regression analyses (eg, limited to location features)?
3. How much variance in depression severity can be explained by the best cluster-combined regression model?

Besides, we wanted to compare the unobtrusively and objectively collected sensor features against features based on ecological momentary assessments (EMA; eg, average valence, average arousal), which, similarly to sensor data, provide a continuous assessment over time but require active input [18,35,37]. Therefore, we investigated the following questions:

4. Which bivariate correlations are present between depression severity and EMA features?
5. How much variance in depression severity can be explained in regression analyses using EMA features?
6. How much variance in depression severity can be explained in regression analyses using EMA and sensing features?

7. What is the difference in explained variance in depression severity between regression models limited to EMA features or sensor features compared to their combination?

Methods

Study Design

This study is an exploratory observation study investigating the associations between smart sensing features and depression severity. Accordingly, we followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [38] (see [Multimedia Appendix 1](#) for the STROBE checklist).

Ethical Considerations

All procedures were assessed and approved by the local ethics committee of Ulm University, Germany (259/16-CL/bal). Informed consent was given by all participants, and participants were informed about their rights according to the European General Data Protection Regulation. Privacy and data security were assessed by Ulm University. Students of Ulm University participating in the study were eligible for study credits as an expense allowance. No other compensation was provided to participants.

Study Population and Procedure

Aiming for a general population sample, participants (see eligibility criteria below) were recruited using an open recruitment strategy involving digital channels (eg, email lists and social media posts) and offline channels (eg, flyers at public institutions). Participants were informed about the purposes of and procedures in the study in an online survey and asked for their written informed consent. If given, they were instructed to install the smart sensing application of the INSIGHTS framework [39,40] after providing basic personal characteristics (ie, gender and age) in the online survey; afterwards, all data were collected via the app. See the assessment and features below for a description of all features and the technical framework paper for further details on the software [39].

Eligibility Criteria for Participants and Episodes

Participants were included in the study if (1) informed consent was given and (2) participants were 18 years or older. (3) Due to the technical requirements of the sensing framework, participants were required to have a smartphone running on Android. Neither a diagnosis nor a minimum level of depression was required to be included in the study. Furthermore, we only included participants' data in the analysis if (4) participants completed the depression severity questionnaire (PHQ-8; see details below) at least once. We structured the data of the participants in 14-day periods consisting of the depression questionnaire assessing the average depression severity in the last 14 days and the corresponding sensing data per day. The number of episodes varied across participants (mean 4.09, SD 3.55; range: 1-42). To avoid the biasing influence of participants being represented more often than others in the dataset and to maximize the data quality, we (5) included only the episode with the lowest amount of missingness per person. Missingness was determined across all days in the episode and features. In

addition, (6) we excluded all participants with more than 50% missing data in EMA and sensing features during an episode to ensure missing data handling procedures (see below) were reliable [41,42]. Excluded participants, according to (6), did not significantly differ in age, gender, or depression symptomology from included participants, underlying the assumption that technical issues leading to the exclusion (eg, app not working) did not occur systematically ([Multimedia Appendix 2](#)).

Assessment and Features

Overview

The assessment consisted of (1) self-report severity measures, (2) EMA, (3) smartphone screen features, (4) app usage features, (5) location features, and (6) call features. Following the validated procedures from previous studies, we used Python, Snakemake, and the Reproducible Analysis Pipeline for Data Streams (RAPIDS) framework in the data preprocessing and extraction pipeline of all smartphone features [28,43-45]. Smartphone features were calculated for each day and aggregated across the 14-day window (eg, average daily smartphone usage duration across 14 days). A summary of all included smartphone features can be found below.

Clinical Questions

We used the PHQ-8 for the assessment of depression severity. The PHQ-8 consists of 8 self-report items asking how often a symptom was present in the last 14 days (0=not at all to 3=nearly every day). Higher PHQ-8 sum scores indicate higher depression severity. The PHQ-8 is a reliable instrument (Cronbach $\alpha=.87$, $\omega=.94$) [46,47] (see [Multimedia Appendix 3](#) for an overview of the PHQ-8 items).

EMA Questions

All items were rated from 0 (lowest imaginable) to 100 (highest imaginable). We assessed valence (higher values indicate positive affect), arousal (higher values indicate higher energy levels), and stress (higher values indicate higher stress levels) 3 times per day (morning, midday, and evening). Additionally, sleep quality (high values indicate good sleep) was assessed in the morning, and satisfaction (higher values indicate stronger satisfaction) with the quality of social interaction, the number of social interactions, activity, and nutrition was rated in the evening once per day. See [40] and [Multimedia Appendix 4](#) for an overview of the EMA questions.

Screen Features

The screen sensor tracked the lock and unlock events of the smartphone. Based on this, we determined the number of usage sessions, duration of usage (sum, average, maximum), regularity index of all unlock episodes, entropy, and normalized entropy of unlock events [28,44].

App Features

For each app, we tracked the start and end of each usage session to calculate the count of all used apps, the mean duration of app usage per day, the regularity index of app usage, and the frequency entropy of app usage [44].

Location Features

Using the GPS sensor of the smartphone, we determined the total distance, logarithmic location variance, number of significant places, stay duration (average, maximum, SD, at top 1 location, at top 2 location, and at top 3 location), the ratio between the time spent at nonsignificant places to all clusters (percent of outlier time), location entropy, normalized location entropy, circadian movement, location routine index, number of location transitions, and moving to static ratio [44,48-54].

Call Features

Incoming, outgoing, and missed calls were tracked separately. For each of them, we tracked the *count* and distinct contacts. Furthermore, for incoming and outgoing calls, we calculated the duration (average, sum, and maximum) and entropy [28,44,54].

See [Table 1](#) for an overview of the feature definition, interpretation, and references, providing an in-depth introduction to the background and calculations of the features. For further details on the Reproducible Analysis Pipeline for Data Streams framework, refer to its paper [44].

Table 1. Feature definition and interpretation.

Feature	Definition	Units and interpretation
Ecological momentary assessment		
Valence	Affect ratings from users	0 (lowest) to 100 (highest); higher values indicate positive valence
Arousal	Arousal ratings from users	0 (lowest) to 100 (highest); higher values indicate higher energy levels
Stress	Stress ratings from users	0 (lowest) to 100 (highest); higher values indicate higher stress
Sleep	Sleep quality ratings from users	0 (lowest) to 100 (highest); higher values indicate higher sleep quality
Social quantity	Satisfaction with quality of social contacts reported by users	0 (lowest) to 100 (highest); higher values indicate higher satisfaction
Social quality	Satisfaction with quality of social contacts reported by users	0 (lowest) to 100 (highest); higher values indicate higher satisfaction
Nutrition	Nutrition ratings from users	0 (lowest) to 100 (highest); higher values indicate healthy nutrition
Physical activity	Physical activity ratings from users	0 (lowest) to 100 (highest); higher values indicate higher intensity activity
App usage		
App count	Number of app usage episodes in foreground	Count
App duration	Duration of app usage episodes in foreground	Time in hours
App frequency entropy	Entropy is a measurement of the degree of variability between the users' behavior states. The frequency of use per app over a 24-hour period was used to calculate app frequency entropy. For calculation details, refer to [44,54]	Higher app frequency entropy reflects a more distributed usage of apps (ie, high variability). Lower values indicate that users used one app more often (ie, low variability)
App regularity index	Regularity index captures the similarity of behavior between the same hours across different days. App regularity index refers to the similarity of the most frequently used app at the same hour across days. Calculations were based on [44,54,55]	Higher regularity index indicates a user used the same app at the same hours across days
Screen usage		
Screen episode count	Number of screen unlock episodes	Count
Screen duration	Duration of unlock episodes	Time in hours
Screen regularity index	Regularity index captures the similarity of behavior between the same hours across different days. Screen regularity index refers to the similarity of the most frequent screen status (on or off) at the same hour across days. Refer to [44,55]	Higher regularity index indicates a user used their smartphone at the same hours across days
Screen entropy	Entropy is a measurement of the degree of variability between the users' behavior states. The frequency of screen states (on or off) over a 24-hour period was used to calculate screen entropy. For calculation details, refer to [44,54]	Higher screen entropy reflects a more distributed unlock of the screen (ie, high variability). Lower values indicate that users screen is more often in one state (ie, low variability).
Screen normalized entropy	Normalized entropy is the entropy divided by the logarithm of the number of states (N)	Entropy divided by $\log(N)$
Call		
Missed calls count	Number of missed calls	Count
Missed calls distinct contacts	Number of distinct contacts whose calls were missed	Count
Incoming calls count	Number of incoming calls	Count
Incoming calls distinct contacts	Number of distinct contacts whose calls were answered	Count
Incoming calls duration	Duration of incoming calls	Time in hours

Feature	Definition	Units and interpretation
Incoming calls entropy	Entropy is a measurement of the degree of variability between the users' behavior states. The duration of incoming calls over a 24-hour period was used to calculate incoming calls entropy. For calculation details, refer to [44,54]	Higher incoming calls entropy reflects a more distributed call duration across incoming calls (ie, high variability)
Outgoing calls count	Number of outgoing calls	Count
Outgoing calls distinct contacts	Number of distinct contacts who were called.	Count
Outgoing calls duration	Duration of outgoing calls	Time in hours
Outgoing calls entropy	Entropy is a measurement of the degree of variability between the users' behavior states. The duration of outgoing calls over a 24-hour period was used to calculate outgoing calls entropy. For calculation details, refer to [44,54]	Higher outgoing calls entropy reflects a more distributed call duration across outgoing calls (ie, high variability)
Location		
Total distance	Total distance (haversine) between tracked location coordinates	Distance in km
Location variance	Logarithm of the combined variance in latitude and longitude [44,48-50]	
Moving to static ratio	Ratio of moving states (speed >1 km/h) to static states (speed <1 km/h) [50]	Ratio; higher values reflect more moving states compared to static states
Number of significant clusters	Clusters are determined by k-means clustering of stationary location coordinates (speed <1 km/h). Clusters needed to be 400 meters from each other. Pauses within 200 meters of cluster were counted as cluster visits after the initial clustering. Only clusters with a time duration of 10 minutes were counted as significant. For further details, see [44,48-51]	Count
Staying time at clusters	Time spent at significant clusters	Time in hours
Time at top 1 location	Total time spent at the most significant cluster	Time in hours
Time at top 2 location	Total time spent at the second most significant cluster	Time in hours
Time at top 3 location	Total time spent at the third most significant cluster	Time in hours
Location entropy	Entropy is a measurement of the degree of variability between the users' behavior states. The duration at significant clusters over a 24-hour period was used to calculate location entropy. For calculation details, refer to [44,48-50,54]	Higher location entropy reflects more distributed time spent at significant clusters. Lower values indicate that users spent more time at some significant clusters.
Location normalized entropy	Entropy divided by the logarithm of the number of significant clusters (N)	Entropy divided by the log(N)
Location circadian movement	The extent to which a person's visits at significant clusters follow a 24-hour circadian rhythm. For further details, see [44,50,52,53,56]	Low values indicate a break from routine, whereas high values indicate that a person followed a daily routine.
Time spent at nonsignificant clusters (at outliers)	Time spent at nonsignificant clusters divided by the time spent at all significant cluster	Ratio; higher values indicate more time spent at nonsignificant clusters

Preprocessing and Missing Data Handling

To account for missing data in the dataset, we performed multiple imputations by chained equations [42]. For a missingness overview before imputation, see [Multimedia Appendix 5](#). As outlined before, we constructed 14-day periods consisting of the PHQ-8 and sensing features per day. The following imputations were conducted on the day level (ie, 14 days for each participant and episode). Given the nested data structure (ie, multiple days of the same participant), we applied 2-level predictive mean matching with random intercepts for all variables. A total of 20 complete datasets were obtained. Convergence was achieved after 10 iterations. In a second step, we aggregated the data from the daily level in each imputed dataset to the episode level to match the data structure of the

PHQ-8 and sensing features. Aggregation consisted of the mean and SD across the 14 days. Therefore, each imputed dataset contained a single data entry for each person consisting of the average depression severity in the last 14 days (PHQ-8) and the average and SD of the sensing features in the last 14 days.

Statistical Analysis

Given the exploratory character of this study, we decided not to report any *P* values. Instead, all analyses report on the point estimates and corresponding 95% CIs. Throughout the analyses, we conducted analyses on each imputed dataset separately and pooled the data using Rubin's rule with Barnhard-Rubin adjustment for degrees of freedom [42,57-59]. The analysis strategy was structured in 3 steps. First, we calculated bivariate between-person Pearson correlations (*r*) and their 95% CIs

between EMA and sensing features with depression. The correlation is a statistical measure of the magnitude of a linear relationship between 2 variables. It ranges from -1 (perfect negative linear relationship) to $+1$ (perfect positive linear relationship). Since higher values of the PHQ-8 indicate higher levels of depression, positive correlations between a feature and the PHQ-8 imply that higher values of this feature are associated with higher depression. However, correlations do not allow any causal inference. See [Table 1](#) for interpretation guidance on all features.

Second, variables with non-zero correlation CIs were included as candidates in cluster-wise regression analyses (ie, limited to EMA, screen, app, location, and call features only). The best models per cluster were determined by stepwise backward exclusion of predictors with zero-including CIs. Predictors, starting with the least influential (standardized β) and broadest zero-including CI, were removed one at a time, and regression models were refitted and compared against each other based on adjusted R^2 after each step. This stepwise backward elimination process was continued until all predictors with zero-including CIs were excluded and regression models no longer improved regarding adjusted R^2 . For each final cluster-wise regression, we determined the adjusted R^2 and its 95% CI to quantify the explained variance in depression severity.

In the third step, we constructed regression models (1) combining the different sensing feature clusters in one model to evaluate the potential of smart sensing as a stand-alone paradigm and (2) including EMA and sensing feature clusters to evaluate their combined potential. As before, predictors were eliminated following stepwise backward exclusion. Overall model performance was evaluated based on adjusted R^2 . Differences in adjusted R^2 and the information criteria Akaike information criterion and Bayesian information criterion were used for model comparisons. Model parameters were standardized and adjusted for age and gender in sensitivity analysis.

Software

All analyses and data preparation were conducted in R. The mice and miceadds packages were used for the imputation and

pooling of analysis results [42,60]. For a full list of all packages, see [Multimedia Appendix 6](#).

Results

Overview

A total of 201 participants answered at least 1 PHQ-8. In total, 94 of the responders were excluded due to poor data quality ($>50\%$ average missingness in EMA and sensing features). Hence, a total of 107 participants were included in the analysis.

The mean age of the participants was 22.81 (SD 7.32), with the oldest participant being 56 years and the youngest 18 years. In total, 83 (77.6%) of the participants identified themselves as female (male: $n=24$, 22.4%). On average, participants showed subclinical depression levels (mean 5.82, SD 4.44), with 20 (18.7%) above the PHQ-8 ≥ 10 cutoff indicating clinically relevant depression severity. For a summary of the PHQ-8 item level, see [Multimedia Appendix 7](#).

EMA features revealed an average valence of 65.98 (SD 12.82), arousal of 50.17 (SD 12.16), and stress of 38.85 (SD 15.66). Daily screen usage of the smartphone was 2.72 hours (SD 2.82) on average. A complete summary of all $k=102$ features and their respective means and SDs can be found in [Multimedia Appendix 8](#).

Correlations

Correlation analyses revealed small to medium correlations of EMA and sensing features with depression severity. Average valence ratings showed the highest correlation to depression severity of the EMA features ($r=-0.55$, 95% CI -0.67 to -0.41), while the SD of app frequency entropy in app usage ($r=-0.19$, 95% CI -0.37 to -0.00), mean duration of outgoing calls ($r=0.25$, 95% CI 0.04 to 0.43), location routine index ($r=0.23$, 95% CI 0.02 to 0.42), and the average screen duration ($r=0.37$, 95% CI 0.20 to 0.53) were the strongest features in their respective clusters. See [Table 2](#) for a summary of all correlations with zero, excluding CIs. The complete correlation summary of all investigated features can be found in [Multimedia Appendix 9](#).

Table 2. Pooled bivariate correlation results between depression and features.^a

	<i>r</i> (95% CI)
Ecological momentary assessment features	
Average valence	−0.55 (−0.67 to −0.41)
Average satisfaction with social quality	−0.51 (−0.64 to −0.35)
Average sleep quality	−0.50 (−0.63 to −0.34)
Average arousal	−0.42 (−0.57 to −0.25)
Average social quantity	−0.39 (−0.54 to −0.22)
Average nutrition	−0.25 (−0.45 to −0.03)
SD of valence	0.23 (0.04 to 0.41)
Average stress	0.42 (0.25 to 0.56)
App features	
SD of frequency entropy	−0.19 (−0.37 to −0.00)
Call features	
Average incoming call duration	0.21 (0.00 to 0.39)
SD of incoming call duration	0.21 (0.01 to 0.40)
Average outgoing call duration	0.25 (0.04 to 0.43)
SD of outgoing call duration	0.25 (0.04 to 0.44)
Features	
Average location routine index	0.23 (0.02 to 0.42)
Screen features	
Average total screen duration	0.23 (0.04 to 0.40)
Average max screen duration	0.24 (0.05 to 0.41)
Average SD of screen episode count	0.27 (0.09 to 0.44)
Average duration of screen episodes	0.37 (0.20 to 0.53)

^aCorrelations are pooled correlations based on multiple imputations. Only correlations with their 95% CI excluding zero are displayed. A full correlation summary can be found in [Multimedia Appendix 9](#). For feature definition and interpretation, refer to [Table 1](#).

Regression

We included the features identified in the correlation analysis in step-wise regression analyses to investigate their incremental contribution to the explained variance in depression severity. The final regression model using EMA features included average valence ($\hat{\beta} = -0.39$, 95% CI −0.58 to −0.21) and social quality as predictors ($\hat{\beta} = -0.29$, 95% CI −0.48 to −0.10). Combined, they explained 35.28% of the variance (95% CI 20.73% to 49.64%).

In the app cluster, SD of app frequency entropy ($\hat{\beta} = -0.19$, 95% CI −0.38 to −0.00) explained adjusted $R^2 = 2.81\%$ of the variance (95% CI 0.00%–12.02%), while location routine index ($\hat{\beta} = 0.23$, 95% CI 0.02 to 0.44) explained adjusted $R^2 = 4.39\%$ (95% CI 0.00% to 16.71%) in the location cluster. From the call cluster, the SD of incoming call duration ($\hat{\beta} = 0.20$, 95% CI 0.03 to 0.41) and average outgoing call duration ($\hat{\beta} = 0.24$, 95% CI 0.04 to 0.44) were the final included predictors explaining adjusted $R^2 = 8.68\%$ of the variance (95% CI 0.88% to 22.32%). Of all 4

candidates in the screen cluster, the average screen duration was the only included predictor in the final model ($\hat{\beta} = 0.37$, 95% CI 0.19 to 0.55; adjusted $R^2 = 13.09\%$, 95% CI 3.40% to 26.65%).

Combining all sensing features in a parsimonious model yielded a model explaining adjusted $R^2 = 20.45\%$ of the variance (95% CI 7.81% to 35.59%) with average screen duration ($\hat{\beta} = 0.39$, 95% CI 0.21 to 0.56), SD of app frequency entropy ($\hat{\beta} = -0.19$, 95% CI −0.36 to −0.02), and SD of incoming call duration ($\hat{\beta} = 0.21$, 95% CI 0.02 to 0.41) as predictors.

The highest variance was explained when combining EMA and sensing features: adj. $R^2 = 45.15\%$ of the variance (95% CI 30.39% to 58.53%). Features included in this prediction model were average valence ($\hat{\beta} = -0.36$, 95% CI −0.53 to −0.19), social quality ($\hat{\beta} = -0.24$, 95% CI −0.41 to −0.06), average screen duration ($\hat{\beta} = 0.22$, 95% CI 0.07 to 0.37), SD of app frequency entropy ($\hat{\beta} = -0.17$, 95% CI −0.31 to −0.02), and average

duration of outgoing calls ($\hat{\beta}=0.17$, 95% CI 0.01 to 0.33). See [Table 3](#) for a summary of the parsimonious EMA, sensing, and combined regression models.

Table 3. Regression results for depression predicted by ecological momentary assessment and smartphone features in stand-alone and combined models.^a

	$\hat{\beta}$ (95% CI)	Adjusted R^2 (95% CI), %	AIC ^b	BIC ^c	Δ adjusted R^2 , %
EMA^d cluster		35.28 (20.73 to 49.64)	262.45	273.14	N/A ^e
Average valence	-0.39 (-0.58 to -0.21)				
Average social quality	-0.29 (-0.48 to -0.10)				
Sensing cluster		20.45 (7.81 to 35.59)	284.53	297.90	EMA: -14.83
Average screen duration	0.39 (0.21 to 0.56)				
Average app frequency entropy	-0.19 (-0.36 to -0.02)				
Incoming call SD	0.21 (0.02 to 0.41)				
Combined		45.15 (30.39 to 58.53)	249.42	268.13	EMA: 9.87; sensing: 24.70
Average valence	-0.36 (-0.53 to -0.19)				
Average social quality	-0.24 (-0.41 to -0.06)				
Average screen duration	0.22 (0.07 to 0.37)				
Average app frequency entropy	-0.17 (-0.31 to -0.02)				
Average duration of outgoing calls	0.17 (0.01 to 0.33)				

^aAll results were obtained by pooling results from multiple imputations according to Rubin's rule. All estimates were fully standardized. For feature definition and interpretation, refer to [Table 1](#).

^bAIC: Akaike information criterion.

^cBIC: Bayesian information criterion.

^dEMA: ecological momentary assessment.

^eNot applicable.

Results were robust when adjusting for age and gender, yielding nonsignificant main effects in the parsimonious EMA (age: -0.08 to 0.23; gender: -0.06 to 0.26), sensing (age: -0.16 to 0.21; gender: -0.26 to 0.10), and combined EMA and sensing model (age: -0.09 to 0.22; gender: -0.09 to 0.21).

Discussion

Principal Findings

To explore the potential of smart sensing for depression, this study investigated the bivariate correlations and explainable variance in regression models built on smartphone sensors and EMA features. Across sensor modalities, we found small correlations between smart sensing features and depression severity. Combined smart sensing features could explain 20.45% (95% CI 7.81%-35.59%) of the depression severity variance in a parsimonious model. Conversely, we found small to medium correlations for EMA features, which could explain 35.28% (95% CI 20.73%-49.64%) of the variance. The best model was the combination of smart sensing and EMA features, which explained 45.15% (95% CI 30.39%-58.53%; Δ adjusted R^2 to smart sensing only: 24.70%; Δ adjusted R^2 to EMA only: 9.87%).

Comparison to Prior Work

The EMA findings are in line with previous studies and reviews highlighting the potential of EMA as a continuous assessment to infer depression severity [35,37,40,61,62]. However, while the bivariate correlations, as well as the explained variance, are higher for EMA compared to features obtained from smartphone sensors, it is important to note that the sensing cluster alone could explain about 20% of the variance. Given the unobtrusive nature of the collection of sensor data, this approach has a crucial advantage over EMA, which requires active user or patient involvement over a long period (eg, multiple daily responses over 14 days). In particular, for clinical application, it should be evaluated whether the additional burden for patients to answer EMA is proportional to the gain in explained variance. Furthermore, various issues in EMA (eg, interpretation of momentary questions, usage of comparison standards) could be avoided by the collection of objective sensor data [19,21,62,63]. That said, to maximize the explained variance, the combination of sensor features and EMA seems to be best. This result extends the findings by Moshe et al [35], who previously evaluated GPS features in conjunction with physiological wearable data and EMA similarly, showing that the combination of sensing features with EMA yields the best regression model for depression severity.

Although these findings seem promising, it is also important to note that this study and others so far are exploratory [34-36,64]. Although we have more than 2 decades of research on EMA and its application for mental health [18,37], the field of smart sensing is still in its infancy. Facing heterogeneous methodology and study quality, as well as potential publication bias in the field, confirmatory studies are highly needed in the field of smart sensing before clinical application [34,36,64,65]. Besides, it is important to note that this study followed a rather data-driven approach to investigating features, which were collectible by the here-used framework. In the context of smart sensing, a central question is which features are needed and provide an incremental benefit. Although this study can give first insights into this topic, predictors like screen duration, app usage entropy, and call features (eg, SD in incoming calls) should only be incorporated in clinical systems if replicated in future studies. Also, an extension to other sensors (eg, language analysis based on LIWC or sentiment analysis) [66-70], app content (eg, usage of social media apps) [71], network usage [28], and the combination with other wearable devices (eg, biophysiological data from smart watches) or different data sources (eg, journaling data) would be a promising addition [23,24,35,67].

However, a closer inspection is needed not only for smart sensing features but also for EMA. For example, our study highlighted that many of the investigated features (eg, arousal or physical activity) did not hold any incremental benefit besides valence and social quality. Since EMA is associated with additional burden for patients, it is especially important to identify the core set of items that maximize predictive power while reducing the item load. Despite a long research history, surprisingly little systematic and meta-analytical evidence is available on which questions should be asked in the context of depression and more precisely when (eg, morning or evening) and on which schedule (eg, multiple fixed time points or microrandomized assessments) [37,63,72-74].

Besides the empirical evidence for the applicability of smart sensing, questions of acceptance, data security and privacy, and ethical challenges surrounding smart sensing need to be addressed [19,75-77]. For instance, recent studies found only moderate acceptance of smart sensing in the context of mental health and highlighted the impact of data types and recipients in smart sensing [77-79]. Only when these barriers and challenges are overcome, can smart sensing unfold its potential fully.

Lastly, we want to point out that the current developments in smart sensing for mental health mainly focus on the prediction of psychopathology [24,26,27,36,64]. Future studies could further advance the field by not only focusing on pathology but also investigating applications to assess risk factors (eg, loneliness, stress) or mediators and mechanisms of change (eg, rumination, therapeutic alliance). Understanding the underlying process of how treatment works is a crucial step to optimizing treatments and pathways in mental health care [80-83]. Since smart sensing allows for a fine-grained and unobtrusive assessment, it might become a promising and feasible paradigm to unveil the mechanism of change in mental health care and better understand the dynamics of therapeutic processes.

Limitations

In addition to the already highlighted exploratory nature of this study, we would like to emphasize a few limitations of this study. First, this study followed a cross-sectional design investigating which predictors explain the variance between persons. Hence, any causal interpretation of the results is not eligible and would require different study designs. For instance, the increased smartphone usage (eg, screen duration) could be caused by depression, but also vice versa or even explained by a third variable. Also, the analysis of trajectories and dynamics over time was not in the scope of this study. Based on the multiple episodes of participants in this study, a longitudinal perspective would be a valuable addition to this study and a field with so far underrepresented research on longitudinal models [34,84,85].

Second, the present sample was a convenient sample recruited in the general population, which resulted in a rather young (mean 22.81, SD 7.32, range 18-56) and unequal gender distribution (female: 77.6%). Furthermore, only 18.7% of the participants showed clinically relevant symptomology. Therefore, the generalizability of the findings to other samples, especially clinical samples, remains open. However, by showing the feasibility of smart sensing predictions and their incremental benefit in explaining variance in depression, this study lays a strong foundation to move to clinical populations and studies. To further increase the quality of clinical studies, also methodological points need to be addressed; for instance, self-report instruments, as applied in this study (ie, PHQ-8), are prone to several sources of bias (eg, social desirability or recall biases). Hence, the application of more reliable and valid assessments like clinician ratings or medical diagnosis should be considered in future studies alongside measurements, which maximize reliability in a specific depression severity range of interest (eg, high reliability in subclinical or severe depression levels) [86-88].

Third, alongside the sample characteristics, the sample size needs to be considered. Adequate sample size is key to designing confirmatory studies aiming to test an assumed clinically relevant effect with sufficient power. In depression research, a standardized mean difference of 0.24, which transfers to a correlation of $r=0.12$ [89], is argued to be a clinically relevant effect [90]. However, to be able to test such a correlation with sufficient power (eg, 80%), a sample size of $N=542$ would be required (assuming a 2-sided test with $\alpha=5\%$, following a bivariate normal model). Hence, our study was highly underpowered to test for a minimally clinically relevant correlation of $r=0.12$. Accordingly, we did not report any P values and solely reported on the 95% CIs to provide a range of reasonable assumable magnitudes of estimates, which could guide future studies in their sample size planning.

Lastly, we would like to emphasize that we opted for linear regression analysis to investigate the incremental contribution of various predictors and allow for direct comparison of various models. This method provides a straightforward approach to answering the present research question. However, sensor data are also very complex, and previous studies have highlighted the potential of nonlinear machine learning models for

depression severity to fully exploit the smart sensing data [25,28]. Although previously applied to classification, machine learning models like extreme gradient-boosted regression trees may also be promising in regression. On the flip side, these models go hand in hand with challenges such as overfitting and difficulties in the interpretation and explanations of the models [25,28,91]. Therefore, even if proven to improve the predictive accuracy, it should be carefully considered whether the complexity and downsides (eg, explainability) of potential machine learning models justify their usage over simpler but easy-to-interpret statistical regression models [92].

Conclusions

Smart sensing and EMA provide potent paradigms to infer and predict depression severity. Our results show that EMA and sensing features alone can substantially explain variance in depression severity. In isolation, EMA was superior to sensing features in terms of explained variance. However, sensing features alone could explain about 20% of the variance, emphasizing the potential of this unobtrusive and objective assessment of depression. To maximize explainable variance, EMA and sensing features should be combined. However, while these findings are promising, confirmatory studies, particularly in clinical settings and samples, are highly needed before robust conclusions can be drawn.

Acknowledgments

We would like to thank all the student assistants involved in the project. No generative AI (eg, ChatGPT) was used in this project and manuscript. This study was self-funded by the authors. YT is supported by the initial phase of the German Center for Mental Health (Deutsches Zentrum für Psychische Gesundheit; grant 01EE2303A).

Data Availability

The primary data obtained in this observation study can be provided by HB and YT on reasonable request. Data-sharing agreements may have to be signed depending on the request. Requests should be directed at the corresponding author (YT). Support is depending on available resources.

Authors' Contributions

YT, EMM, CM, and HB were involved in the conceptualization. Data curation was done by YT, EMM, and KOA. Formal analysis was done by YT. CM and HB were involved in funding acquisition. EMM, YT, HB, CK, and CM were involved in the investigation. YT and KOA were involved in the methodology. Project administration was done by YT and EMM. YT, EMM, CM, CK, and HB contributed to resources; software was contributed by YT, KOA, and CK. YT, EMM, and HB were involved in supervision. Validation was done by YT, EMM, KOA, and CK. Visualization was done by YT. YT also contributed to writing the original draft. YT, EMM, KOA, CM, CK, and HB contributed to reviewing and editing the manuscript.

Conflicts of Interest

All authors declare no conflicts of interest. However, for reasons of transparency, CM notes that he has received (to Ulm University and earlier University of Bonn) grants from agencies such as the German Research Foundation. CM has performed grant reviews for several agencies, edited journal sections and articles, given academic lectures at clinical or scientific venues or for companies, and generated books or book chapters for publishers of mental health texts. For some of these activities, he received royalties, but never from gaming or social media companies. CM was part of a discussion circle (Digitalität und Verantwortung: <https://about.fb.com/de/news/h/gespraechskreis-digitalitaet-und-verantwortung/>) debating ethical questions linked to social media, digitalization, and society/democracy at Facebook. In this context, he received no salary for his activities. Finally, he currently functions as an independent scientist on the scientific advisory board of the Nymphenburg Group. This activity is financially compensated. Moreover, he is on the scientific advisory board of Applied Cognition, an activity which is also compensated.

Multimedia Appendix 1

STROBE checklist.

[\[DOCX File , 22 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Sensitivity analysis on study exclusion.

[\[DOCX File , 16 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Patient Health Questionnaire items.

[\[DOCX File , 16 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Ecological momentary assessment items.

[\[DOCX File , 15 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Overview of missingness per feature before multiple imputation.

[\[DOCX File , 17 KB-Multimedia Appendix 5\]](#)

Multimedia Appendix 6

Software information.

[\[DOCX File , 15 KB-Multimedia Appendix 6\]](#)

Multimedia Appendix 7

Additional sample characteristics.

[\[DOCX File , 15 KB-Multimedia Appendix 7\]](#)

Multimedia Appendix 8

Feature means and SDs.

[\[DOCX File , 21 KB-Multimedia Appendix 8\]](#)

Multimedia Appendix 9

Full correlations between depression and features.

[\[DOCX File , 35 KB-Multimedia Appendix 9\]](#)

References

1. Herrman H, Patel V, Kieling C, Berk M, Buchweitz C, Cuijpers P, et al. Time for united action on depression: a Lancet-World Psychiatric Association Commission. *Lancet*. 2022;399(10328):957-1022. [doi: [10.1016/S0140-6736\(21\)02141-3](https://doi.org/10.1016/S0140-6736(21)02141-3)] [Medline: [35180424](https://pubmed.ncbi.nlm.nih.gov/35180424/)]
2. 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;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/)]
3. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. Nov 10, 2018;392(10159):1789-1858. [FREE Full text] [doi: [10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7)] [Medline: [30496104](https://pubmed.ncbi.nlm.nih.gov/30496104/)]
4. World mental health report: transforming mental health for all. Geneva. World Health Organization; 2022. URL: <https://www.who.int/publications/i/item/9789240049338> [accessed 2024-12-20]
5. Cuijpers P, Noma H, Karyotaki E, Vinkers CH, Cipriani A, Furukawa TA. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry*. 2020;19(1):92-107. [FREE Full text] [doi: [10.1002/wps.20701](https://doi.org/10.1002/wps.20701)] [Medline: [31922679](https://pubmed.ncbi.nlm.nih.gov/31922679/)]
6. Cuijpers P, Karyotaki E, Reijnders M, Ebert DD. Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression. *Epidemiol Psychiatr Sci*. 2019;28(1):21-30. [FREE Full text] [doi: [10.1017/S2045796018000057](https://doi.org/10.1017/S2045796018000057)] [Medline: [29486804](https://pubmed.ncbi.nlm.nih.gov/29486804/)]
7. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: a meta-analytic review. *Psychol Bull*. 2021;147(8):749-786. [doi: [10.1037/bul0000334](https://doi.org/10.1037/bul0000334)] [Medline: [34898233](https://pubmed.ncbi.nlm.nih.gov/34898233/)]
8. Vigo D, Haro JM, Hwang I, Aguilar-Gaxiola S, Alonso J, Borges G, et al. Toward measuring effective treatment coverage: critical bottlenecks in quality- and user-adjusted coverage for major depressive disorder. *Psychol Med*. Oct 20, 2020;1-11. [FREE Full text] [doi: [10.1017/S0033291720003797](https://doi.org/10.1017/S0033291720003797)] [Medline: [33077023](https://pubmed.ncbi.nlm.nih.gov/33077023/)]
9. Andrade LH, Alonso J, Mneimneh Z, Wells J, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. *Psychol Med*. 2014;44(6):1303-1317. [FREE Full text] [doi: [10.1017/S0033291713001943](https://doi.org/10.1017/S0033291713001943)] [Medline: [23931656](https://pubmed.ncbi.nlm.nih.gov/23931656/)]
10. Trautmann S, Beesdo-Baum K. The treatment of depression in primary care. *Dtsch Arztebl Int*. 2017;114(43):721-728. [FREE Full text] [doi: [10.3238/arztebl.2017.0721](https://doi.org/10.3238/arztebl.2017.0721)] [Medline: [29143731](https://pubmed.ncbi.nlm.nih.gov/29143731/)]
11. Kroenke K, Unutzer J. Closing the false divide: sustainable approaches to integrating mental health services into primary care. *J Gen Intern Med*. 2017;32(4):404-410. [FREE Full text] [doi: [10.1007/s11606-016-3967-9](https://doi.org/10.1007/s11606-016-3967-9)] [Medline: [28243873](https://pubmed.ncbi.nlm.nih.gov/28243873/)]

12. Thombs BD, Coyne JC, Cuijpers P, de Jonge P, Gilbody S, Ioannidis J, et al. Rethinking recommendations for screening for depression in primary care. *CMAJ*. 2012;184(4):413-418. [FREE Full text] [doi: [10.1503/cmaj.111035](https://doi.org/10.1503/cmaj.111035)] [Medline: [21930744](https://pubmed.ncbi.nlm.nih.gov/21930744/)]
13. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet*. 2009;374(9690):609-619. [doi: [10.1016/S0140-6736\(09\)60879-5](https://doi.org/10.1016/S0140-6736(09)60879-5)] [Medline: [19640579](https://pubmed.ncbi.nlm.nih.gov/19640579/)]
14. Terhorst Y, Sander LB, Ebert DD, Baumeister H. Optimizing the predictive power of depression screenings using machine learning. *Digit Health*. 2023;9:20552076231194939. [FREE Full text] [doi: [10.1177/20552076231194939](https://doi.org/10.1177/20552076231194939)] [Medline: [37654715](https://pubmed.ncbi.nlm.nih.gov/37654715/)]
15. Lotfi L, Flyckt L, Krakau I, Mårtensson B, Nilsson GH. Undetected depression in primary healthcare: occurrence, severity and co-morbidity in a two-stage procedure of opportunistic screening. *Nord J Psychiatry*. 2010;64(6):421-427. [doi: [10.3109/08039481003786378](https://doi.org/10.3109/08039481003786378)] [Medline: [20429746](https://pubmed.ncbi.nlm.nih.gov/20429746/)]
16. First MB, Rebello TJ, Keeley JW, Bhargava R, Dai Y, Kulygina M, et al. Do mental health professionals use diagnostic classifications the way we think they do? A global survey. *World Psychiatry*. 2018;17(2):187-195. [FREE Full text] [doi: [10.1002/wps.20525](https://doi.org/10.1002/wps.20525)] [Medline: [29856559](https://pubmed.ncbi.nlm.nih.gov/29856559/)]
17. Wichers M. The dynamic nature of depression: a new micro-level perspective of mental disorder that meets current challenges. *Psychol Med*. 2014;44(7):1349-1360. [doi: [10.1017/S0033291713001979](https://doi.org/10.1017/S0033291713001979)] [Medline: [23942140](https://pubmed.ncbi.nlm.nih.gov/23942140/)]
18. Trull TJ, Ebner-Priemer U. Ambulatory assessment. *Annu Rev Clin Psychol*. 2013;9:151-176. [FREE Full text] [doi: [10.1146/annurev-clinpsy-050212-185510](https://doi.org/10.1146/annurev-clinpsy-050212-185510)] [Medline: [23157450](https://pubmed.ncbi.nlm.nih.gov/23157450/)]
19. Terhorst Y, Knauer J, Baumeister H. Smart sensing enhanced diagnostic expert systems. In: Montag C, Baumeister H, editors. *Digital Phenotyping and Mobile Sensors*. Berlin. Springer; 2023:413-425.
20. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annu Rev Clin Psychol*. 2017;13:23-47. [FREE Full text] [doi: [10.1146/annurev-clinpsy-032816-044949](https://doi.org/10.1146/annurev-clinpsy-032816-044949)] [Medline: [28375728](https://pubmed.ncbi.nlm.nih.gov/28375728/)]
21. Garatva P, Terhorst Y, Messner EM, Karlen W, Pryss R, Baumeister H. Smart sensors for health research and improvement. In: Montag C, Baumeiste H, editors. *Digital and Phenotyping Mobile Sensors*. Berlin. Springer; 2023:395-411.
22. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*. 2016;41(7):1691-1696. [FREE Full text] [doi: [10.1038/npp.2016.7](https://doi.org/10.1038/npp.2016.7)] [Medline: [26818126](https://pubmed.ncbi.nlm.nih.gov/26818126/)]
23. Steele R, Hillsgrove T, Khoshavi N, Jaimes LG. A survey of cyber-physical system implementations of real-time personalized interventions. *J Ambient Intell Human Comput*. 2021;13(5):2325-2342. [FREE Full text] [doi: [10.1007/s12652-021-03263-0](https://doi.org/10.1007/s12652-021-03263-0)]
24. Abd-Alrazaq A, AlSaad R, Shuweihi F, Ahmed A, Aziz S, Sheikh J. Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. *NPJ Digit Med*. 2023;6(1):84. [FREE Full text] [doi: [10.1038/s41746-023-00828-5](https://doi.org/10.1038/s41746-023-00828-5)] [Medline: [37147384](https://pubmed.ncbi.nlm.nih.gov/37147384/)]
25. Dlima SD, Shevade S, Menezes SR, Ganju A. Digital phenotyping in health using machine learning approaches: scoping review. *JMIR Bioinform Biotechnol*. 2022;3(1):e39618. [FREE Full text] [doi: [10.2196/39618](https://doi.org/10.2196/39618)] [Medline: [38935947](https://pubmed.ncbi.nlm.nih.gov/38935947/)]
26. Zarate D, Stavropoulos V, Ball M, de Sena Collier G, Jacobson NC. Correction: exploring the digital footprint of depression: a PRISMA systematic literature review of the empirical evidence. *BMC Psychiatry*. 2022;22(1):530. [FREE Full text] [doi: [10.1186/s12888-022-04153-1](https://doi.org/10.1186/s12888-022-04153-1)] [Medline: [35932004](https://pubmed.ncbi.nlm.nih.gov/35932004/)]
27. Nouman M, Khoo SY, Mahmud MAP, Kouzani AZ. Recent advances in contactless sensing technologies for mental health monitoring. *IEEE Internet Things J*. 2022;9(1):274-297. [FREE Full text] [doi: [10.1109/jiot.2021.3097801](https://doi.org/10.1109/jiot.2021.3097801)]
28. Opoku Asare K, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. *JMIR mHealth uHealth*. 2021;9(7):e26540. [FREE Full text] [doi: [10.2196/26540](https://doi.org/10.2196/26540)] [Medline: [34255713](https://pubmed.ncbi.nlm.nih.gov/34255713/)]
29. Kotov R, Krueger R, Watson D, Achenbach T, Althoff R, Bagby R, et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. *J Abnorm Psychol*. 2017;126(4):454-477. [doi: [10.1037/abn0000258](https://doi.org/10.1037/abn0000258)] [Medline: [28333488](https://pubmed.ncbi.nlm.nih.gov/28333488/)]
30. Regier DA, Narrow W, Clarke D, Kraemer H, Kuramoto S, Kuhl E, et al. DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses. *Am J Psychiatry*. 2013;170(1):59-70. [doi: [10.1176/appi.ajp.2012.12070999](https://doi.org/10.1176/appi.ajp.2012.12070999)] [Medline: [23111466](https://pubmed.ncbi.nlm.nih.gov/23111466/)]
31. Fried EI, Nesse RM. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. *J Affect Disord*. 2015;172:96-102. [FREE Full text] [doi: [10.1016/j.jad.2014.10.010](https://doi.org/10.1016/j.jad.2014.10.010)] [Medline: [25451401](https://pubmed.ncbi.nlm.nih.gov/25451401/)]
32. Fried EI. The 52 symptoms of major depression: lack of content overlap among seven common depression scales. *J Affect Disord*. 2017;208:191-197. [FREE Full text] [doi: [10.31234/osf.io/vbk59](https://doi.org/10.31234/osf.io/vbk59)]
33. Chevance A, Ravaud P, Tomlinson A, Le Berre C, Teufer B, Touboul S, et al. Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey. *Lancet Psychiatry*. 2020;7(8):692-702. [doi: [10.1016/S2215-0366\(20\)30191-7](https://doi.org/10.1016/S2215-0366(20)30191-7)] [Medline: [32711710](https://pubmed.ncbi.nlm.nih.gov/32711710/)]
34. Terhorst Y, Knauer J, Philippi P, Baumeister H. The relation between passively collected GPS mobility metrics and depressive symptoms: systematic review and meta-analysis. *J Med Internet Res*. 2024;26:e51875. [FREE Full text] [doi: [10.2196/51875](https://doi.org/10.2196/51875)] [Medline: [39486026](https://pubmed.ncbi.nlm.nih.gov/39486026/)]

35. Moshe I, Terhorst Y, Opoku Asare K, Sander LB, Ferreira D, Baumeister H, et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. *Front Psychiatry*. 2021;12:625247. [FREE Full text] [doi: [10.3389/fpsy.2021.625247](https://doi.org/10.3389/fpsy.2021.625247)] [Medline: [33584388](https://pubmed.ncbi.nlm.nih.gov/33584388/)]
36. Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review. *JMIR mHealth uHealth*. 2018;6(8):e165. [FREE Full text] [doi: [10.2196/mhealth.9691](https://doi.org/10.2196/mhealth.9691)] [Medline: [30104184](https://pubmed.ncbi.nlm.nih.gov/30104184/)]
37. Trull TJ, Ebner-Priemer UW. Ambulatory assessment in psychopathology research: a review of recommended reporting guidelines and current practices. *J Abnorm Psychol*. 2020;129(1):56-63. [doi: [10.1037/abn0000473](https://doi.org/10.1037/abn0000473)] [Medline: [31868388](https://pubmed.ncbi.nlm.nih.gov/31868388/)]
38. von Elm E, Altman D, Egger M, Pocock S, Gøtzsche PC, Vandenbroucke J, et al. STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573-577. [FREE Full text] [doi: [10.7326/0003-4819-147-8-200710160-00010](https://doi.org/10.7326/0003-4819-147-8-200710160-00010)] [Medline: [17938396](https://pubmed.ncbi.nlm.nih.gov/17938396/)]
39. Montag C, Baumeister H, Kannen C, Sariyska R, Meßner E, Brand M. Concept, possibilities and pilot-testing of a new smartphone application for the social and life sciences to study human behavior including validation data from personality psychology. *J*. 2019;2(2):102-115. [FREE Full text] [doi: [10.3390/j2020008](https://doi.org/10.3390/j2020008)]
40. Messner EM, Sariyska R, Mayer B, Montag C, Kannen C, Schwerdtfeger A, et al. Insights – future implications of passive smartphone sensing in the therapeutic context. *Verhaltenstherapie*. 2019;32(Suppl. 1):86-95. [FREE Full text] [doi: [10.1159/000501951](https://doi.org/10.1159/000501951)]
41. Enders CK. *Applied Missing Data Analysis*. New York. The Guilford Press; 2010.
42. van Buuren S, Groothuis-Oudshoorn CG. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011;45(3). [FREE Full text] [doi: [10.32614/cran.package.mice](https://doi.org/10.32614/cran.package.mice)]
43. Köster J, Rahmann S. Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics*. 2012;28(19):2520-2522. [doi: [10.1093/bioinformatics/bts480](https://doi.org/10.1093/bioinformatics/bts480)] [Medline: [22908215](https://pubmed.ncbi.nlm.nih.gov/22908215/)]
44. Vega J, Li M, Aguilera K, Goel N, Joshi E, Khandekar K, et al. Reproducible analysis pipeline for data streams: open-source software to process data collected with mobile Devices. *Front Digit Health*. 2021;3:769823. [FREE Full text] [doi: [10.3389/fdgh.2021.769823](https://doi.org/10.3389/fdgh.2021.769823)] [Medline: [34870271](https://pubmed.ncbi.nlm.nih.gov/34870271/)]
45. Opoku Asare K, Moshe I, Terhorst Y, Vega J, Hosio S, Baumeister H, et al. Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: a longitudinal data analysis. *Pervasive Mob Comput*. 2022;83:101621. [FREE Full text] [doi: [10.1016/j.pmcj.2022.101621](https://doi.org/10.1016/j.pmcj.2022.101621)]
46. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. *J Affect Disord*. 2009;114(1-3):163-173. [doi: [10.1016/j.jad.2008.06.026](https://doi.org/10.1016/j.jad.2008.06.026)] [Medline: [18752852](https://pubmed.ncbi.nlm.nih.gov/18752852/)]
47. Arias de la Torre J, Vilagut G, Ronaldson A, Valderas J, Bakolis I, Dregan A, et al. Reliability and cross-country equivalence of the 8-item version of the patient health questionnaire (PHQ-8) for the assessment of depression: results from 27 countries in Europe. *Lancet Reg Health Eur*. 2023;31:100659. [FREE Full text] [doi: [10.1016/j.lanepe.2023.100659](https://doi.org/10.1016/j.lanepe.2023.100659)] [Medline: [37332385](https://pubmed.ncbi.nlm.nih.gov/37332385/)]
48. Saeb S, Zhang M, Kwasny MM, Karr CJ, Kording K, Mohr D. The relationship between clinical, momentary, and sensor-based assessment of depression. *Int Conf Pervasive Comput Technol Healthc*. 2015;2015:103. [FREE Full text] [doi: [10.4108/icst.pervasivehealth.2015.259034](https://doi.org/10.4108/icst.pervasivehealth.2015.259034)] [Medline: [26640739](https://pubmed.ncbi.nlm.nih.gov/26640739/)]
49. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *J Med Internet Res*. 2015;17(7):e175. [FREE Full text] [doi: [10.2196/jmir.4273](https://doi.org/10.2196/jmir.4273)] [Medline: [26180009](https://pubmed.ncbi.nlm.nih.gov/26180009/)]
50. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ*. 2016;4:e2537. [FREE Full text] [doi: [10.7717/peerj.2537](https://doi.org/10.7717/peerj.2537)] [Medline: [28344895](https://pubmed.ncbi.nlm.nih.gov/28344895/)]
51. Arthur D, Vassilvitskii S. K-Means++: the advantages of careful seeding. 2007. Presented at: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms; January 7-9, 2007; New Orleans, Louisiana, USA.
52. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. 2015. Presented at: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp; September 7-11, 2015:1293-1304; Osaka, Japan. URL: <https://doi.org/10.1145/2750858.2805845> [doi: [10.1145/2750858.2805845](https://doi.org/10.1145/2750858.2805845)]
53. Barnett I, Onnela JP. Inferring mobility measures from GPS traces with missing data. *Biostatistics*. 2020;21(2):e98-e112. [FREE Full text] [doi: [10.1093/biostatistics/kxy059](https://doi.org/10.1093/biostatistics/kxy059)] [Medline: [30371736](https://pubmed.ncbi.nlm.nih.gov/30371736/)]
54. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948;27(3):379-423. [doi: [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x)]
55. Wang W, Harari GM, Wang R, Müller SR, Mirjafari S, Masaba K, et al. Sensing behavioral change over time. *Proc ACM Interact Mob Wearable Ubiquitous Technol*. 2018;2(3):1-21. [FREE Full text] [doi: [10.1145/3264951](https://doi.org/10.1145/3264951)]
56. Wu C, McMahon M, Fritz H, Schnyer DM. Circadian rhythms are not captured equal: exploring circadian metrics extracted by different computational methods from smartphone accelerometer and GPS sensors in daily life tracking. *Digit Health*. 2022;8:20552076221114201. [FREE Full text] [doi: [10.1177/20552076221114201](https://doi.org/10.1177/20552076221114201)] [Medline: [35874860](https://pubmed.ncbi.nlm.nih.gov/35874860/)]
57. Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc*. 1996;91(434):473-489. [doi: [10.2307/2291635](https://doi.org/10.2307/2291635)]

58. Barnard J, Rubin DB. Miscellanea. Small-sample degrees of freedom with multiple imputation. *Biometrika*. 1999;86(4):948-955. [[FREE Full text](#)] [doi: [10.1093/biomet/86.4.948](https://doi.org/10.1093/biomet/86.4.948)]
59. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, New Jersey. John Wiley & Sons; 1987.
60. Robitzsch A, Grund S, Henke T. miceadds: some additional multiple imputation functions, especially for mice. 2018. URL: <https://alexanderrobitzsch.r-universe.dev/miceadds> [accessed 2024-11-23]
61. Hall M, Scherner PV, Kreidel Y, Rubel JA. A systematic review of momentary assessment designs for mood and anxiety symptoms. *Front Psychol*. 2021;12:642044. [[FREE Full text](#)] [doi: [10.3389/fpsyg.2021.642044](https://doi.org/10.3389/fpsyg.2021.642044)] [Medline: [34079492](https://pubmed.ncbi.nlm.nih.gov/34079492/)]
62. Colombo D, Fernández-Álvarez J, Patané A, Semonella M, Kwiatkowska M, García-Palacios A, et al. Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: a systematic review. *J Clin Med*. 2019;8(4):465. [[FREE Full text](#)] [doi: [10.3390/jcm8040465](https://doi.org/10.3390/jcm8040465)] [Medline: [30959828](https://pubmed.ncbi.nlm.nih.gov/30959828/)]
63. Stone AA, Schneider S, Smyth JM. Evaluation of pressing issues in ecological momentary assessment. *Annu Rev Clin Psychol*. 2023;19:107-131. [[FREE Full text](#)] [doi: [10.1146/annurev-clinpsy-080921-083128](https://doi.org/10.1146/annurev-clinpsy-080921-083128)] [Medline: [36475718](https://pubmed.ncbi.nlm.nih.gov/36475718/)]
64. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. *J Biomed Inform*. 2018;77:120-132. [[FREE Full text](#)] [doi: [10.1016/j.jbi.2017.12.008](https://doi.org/10.1016/j.jbi.2017.12.008)] [Medline: [29248628](https://pubmed.ncbi.nlm.nih.gov/29248628/)]
65. De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. *NPJ Digit Med*. 2022;5(1):3. [[FREE Full text](#)] [doi: [10.1038/s41746-021-00548-8](https://doi.org/10.1038/s41746-021-00548-8)] [Medline: [35017634](https://pubmed.ncbi.nlm.nih.gov/35017634/)]
66. Rathner EM, Djamali J, Terhorst Y, Schuller B, Cummins N, Salamon G, et al. How did you like 2017? Detection of language markers of depression and narcissism in personal narratives. 2018. Presented at: Interspeech 2018; September 2-6, 2018:3388-3392; Hyderabad, India. URL: <https://doi.org/10.21437/Interspeech.2018-2040> [doi: [10.21437/interspeech.2018-2040](https://doi.org/10.21437/interspeech.2018-2040)]
67. Kathan A, Triantafyllopoulos A, He X, Milling M, Yan T, Rajamani ST, et al. Journaling data for daily PHQ-2 depression prediction and forecasting. *Annu Int Conf IEEE Eng Med Biol Soc*. 2022;2022:2627-2630. [doi: [10.1109/EMBC48229.2022.9871015](https://doi.org/10.1109/EMBC48229.2022.9871015)] [Medline: [36086268](https://pubmed.ncbi.nlm.nih.gov/36086268/)]
68. Zantvoort K, Scharfenberger J, Boß L, Lehr D, Funk B. Finding the best match - a case study on the (text-)feature and model choice in digital mental health interventions. *J Healthc Inform Res*. 2023;7(4):447-479. [[FREE Full text](#)] [doi: [10.1007/s41666-023-00148-z](https://doi.org/10.1007/s41666-023-00148-z)] [Medline: [37927375](https://pubmed.ncbi.nlm.nih.gov/37927375/)]
69. 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-centered design, data quality, and behavioral and clinical modeling. *JMIR Ment Health*. 2021;8(8):e27589. [[FREE Full text](#)] [doi: [10.2196/27589](https://doi.org/10.2196/27589)] [Medline: [34383685](https://pubmed.ncbi.nlm.nih.gov/34383685/)]
70. Hussain F, Stange JP, Langenecker SA, McInnis M, Zulueta J, Piscitello A, et al. Passive sensing of affective and cognitive functioning in mood disorders by analyzing keystroke kinematics and speech dynamics. In: Baumeister H, Montag C, editors. *Mobile Sensing and Digital Phenotyping in Psychoinformatics*. Berlin. Springer; 2019.
71. Lin LY, Sidani JE, Shensa A, Radovic A, Miller E, Colditz JB, et al. Association between social media use and depression among U.S. young adults. *Depress Anxiety*. 2016;33(4):323-331. [[FREE Full text](#)] [doi: [10.1002/da.22466](https://doi.org/10.1002/da.22466)] [Medline: [26783723](https://pubmed.ncbi.nlm.nih.gov/26783723/)]
72. Eisele G, Vachon H, Lafit G, Kuppens P, Houben M, Myin-Germeys I, et al. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *PsyArXiv*. Preprint posted online on February 20, 2020. [[FREE Full text](#)] [doi: [10.31234/osf.io/zf4nm](https://doi.org/10.31234/osf.io/zf4nm)]
73. Reiter T, Schoedel R. Never miss a beep: Using mobile sensing to investigate (non-)compliance in experience sampling studies. *Behav Res Methods*. 2024;56(4):4038-4060. [[FREE Full text](#)] [doi: [10.3758/s13428-023-02252-9](https://doi.org/10.3758/s13428-023-02252-9)] [Medline: [37932624](https://pubmed.ncbi.nlm.nih.gov/37932624/)]
74. Cloos L, Ceulemans E, Kuppens P. Development, validation, and comparison of self-report measures for positive and negative affect in intensive longitudinal research. *Psychol Assess*. 2023;35(3):189-204. [doi: [10.1037/pas0001200](https://doi.org/10.1037/pas0001200)] [Medline: [36480406](https://pubmed.ncbi.nlm.nih.gov/36480406/)]
75. Torous J, Nebeker C. Navigating ethics in the digital age: introducing Connected and Open Research Ethics (CORE), a tool for researchers and institutional review boards. *J Med Internet Res*. 2017;19(2):e38. [[FREE Full text](#)] [doi: [10.2196/jmir.6793](https://doi.org/10.2196/jmir.6793)] [Medline: [28179216](https://pubmed.ncbi.nlm.nih.gov/28179216/)]
76. McCradden MD, Joshi S, Mazwi M, Anderson JA. Ethical limitations of algorithmic fairness solutions in health care machine learning. *Lancet Digit Health*. 2020;2(5):e221-e223. [[FREE Full text](#)] [doi: [10.1016/S2589-7500\(20\)30065-0](https://doi.org/10.1016/S2589-7500(20)30065-0)] [Medline: [33328054](https://pubmed.ncbi.nlm.nih.gov/33328054/)]
77. Terhorst Y, Weilbacher N, Suda C, Simon L, Messner EM, Sander LB, et al. Acceptance of smart sensing: a barrier to implementation-results from a randomized controlled trial. *Front Digit Health*. 2023;5:1075266. [[FREE Full text](#)] [doi: [10.3389/fdgh.2023.1075266](https://doi.org/10.3389/fdgh.2023.1075266)] [Medline: [37519894](https://pubmed.ncbi.nlm.nih.gov/37519894/)]
78. Nicholas J, Shilton K, Schueller SM, Gray EL, Kwasny MJ, Mohr DC. The role of data type and recipient in individuals' perspectives on sharing passively collected smartphone data for mental health: cross-sectional questionnaire study. *JMIR mHealth uHealth*. 2019;7(4):e12578. [[FREE Full text](#)] [doi: [10.2196/12578](https://doi.org/10.2196/12578)] [Medline: [30950799](https://pubmed.ncbi.nlm.nih.gov/30950799/)]

79. Rottstädt F, Becker E, Wilz G, Croy I, Baumeister H, Terhorst Y. Enhancing the acceptance of smart sensing in psychotherapy patients: findings from a randomized controlled trial. *Front Digit Health*. 2024;6:1335776. [FREE Full text] [doi: [10.3389/fdgh.2024.1335776](https://doi.org/10.3389/fdgh.2024.1335776)] [Medline: [38698889](https://pubmed.ncbi.nlm.nih.gov/38698889/)]
80. Pfammatter M, Tschacher W. Klassen allgemeiner wirkfaktoren der psychotherapie und ihr zusammenhang mit therapietechniken [Article in German]. *Z Psychosom Med Psychother*. 2016;45(1):1-13. [FREE Full text] [doi: [10.1026/1616-3443/a000331](https://doi.org/10.1026/1616-3443/a000331)]
81. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. *Br Med J*. 2015;350:h1258. [FREE Full text] [doi: [10.1136/bmj.h1258](https://doi.org/10.1136/bmj.h1258)] [Medline: [25791983](https://pubmed.ncbi.nlm.nih.gov/25791983/)]
82. Domhardt M, Grund S, Mayer A, Büscher R, Ebert DD, Sander LB, et al. Unveiling mechanisms of change in digital interventions for depression: study protocol for a systematic review and individual participant data meta-analysis. *Front Psychiatry*. 2022;13:899115. [FREE Full text] [doi: [10.3389/fpsy.2022.899115](https://doi.org/10.3389/fpsy.2022.899115)] [Medline: [36262633](https://pubmed.ncbi.nlm.nih.gov/36262633/)]
83. Kazdin AE. Mediators and mechanisms of change in psychotherapy research. *Annu Rev Clin Psychol*. 2007;3:1-27. [doi: [10.1146/annurev.clinpsy.3.022806.091432](https://doi.org/10.1146/annurev.clinpsy.3.022806.091432)] [Medline: [17716046](https://pubmed.ncbi.nlm.nih.gov/17716046/)]
84. Zhang Y, Folarin A, Sun S, Cummins N, Vairavan S, Bendayan R, et al. RADAR-CNS consortium. Longitudinal relationships between depressive symptom severity and phone-measured mobility: dynamic structural equation modeling study. *JMIR Ment Health*. 2022;9(3):e34898. [FREE Full text] [doi: [10.2196/34898](https://doi.org/10.2196/34898)] [Medline: [35275087](https://pubmed.ncbi.nlm.nih.gov/35275087/)]
85. Müller SR, Peters H, Matz SC, Wang W, Harari GM. Investigating the relationships between mobility behaviours and indicators of subjective well-being using smartphone-based experience sampling and GPS tracking. *Eur J Pers*. 2020;34(5):714-732. [FREE Full text] [doi: [10.1002/per.2262](https://doi.org/10.1002/per.2262)]
86. Wahl I, Löwe B, Bjorner JB, Fischer F, Langs G, Voderholzer U, et al. Standardization of depression measurement: a common metric was developed for 11 self-report depression measures. *J Clin Epidemiol*. 2014;67(1):73-86. [doi: [10.1016/j.jclinepi.2013.04.019](https://doi.org/10.1016/j.jclinepi.2013.04.019)] [Medline: [24262771](https://pubmed.ncbi.nlm.nih.gov/24262771/)]
87. Brehaut E, Neupane D, Levis B, Wu Y, Sun Y, Krishnan A, et al. Depression prevalence using the HADS-D compared to SCID major depression classification: an individual participant data meta-analysis. *J Psychosom Res*. 2020;139:110256. [doi: [10.1016/j.jpsychores.2020.110256](https://doi.org/10.1016/j.jpsychores.2020.110256)] [Medline: [33069051](https://pubmed.ncbi.nlm.nih.gov/33069051/)]
88. Levis B, Yan XW, He C, Sun Y, Benedetti A, Thombs BD. Comparison of depression prevalence estimates in meta-analyses based on screening tools and rating scales versus diagnostic interviews: a meta-research review. *BMC Med*. 2019;17(1):65. [FREE Full text] [doi: [10.1186/s12916-019-1297-6](https://doi.org/10.1186/s12916-019-1297-6)] [Medline: [30894161](https://pubmed.ncbi.nlm.nih.gov/30894161/)]
89. Nordahl-Hansen A, Øien RA, Volkmar F, Shic F, Cicchetti DV. Enhancing the understanding of clinically meaningful results: a clinical research perspective. *Psychiatry Res*. 2018;270:801-806. [FREE Full text] [doi: [10.1016/j.psychres.2018.10.069](https://doi.org/10.1016/j.psychres.2018.10.069)] [Medline: [30551328](https://pubmed.ncbi.nlm.nih.gov/30551328/)]
90. Cuijpers P, Turner EH, Koole SL, van Dijke A, Smit F. What is the threshold for a clinically relevant effect? The case of major depressive disorders. *Depress Anxiety*. 2014;31(5):374-378. [doi: [10.1002/da.22249](https://doi.org/10.1002/da.22249)] [Medline: [24677535](https://pubmed.ncbi.nlm.nih.gov/24677535/)]
91. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol*. 2018;14:91-118. [doi: [10.1146/annurev-clinpsy-032816-045037](https://doi.org/10.1146/annurev-clinpsy-032816-045037)] [Medline: [29401044](https://pubmed.ncbi.nlm.nih.gov/29401044/)]
92. Jacobucci R, Grimm KJ. Machine learning and psychological research: the unexplored effect of measurement. *Perspect Psychol Sci*. 2020;15(3):809-816. [doi: [10.1177/1745691620902467](https://doi.org/10.1177/1745691620902467)] [Medline: [32348703](https://pubmed.ncbi.nlm.nih.gov/32348703/)]

Abbreviations

EMA: ecological momentary assessment

MDD: major depressive disorder

PHQ-8: 8-item Patient Health Questionnaire

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Coristine; submitted 08.12.23; peer-reviewed by R Antunes, H Hsin; comments to author 18.04.24; revised version received 30.06.24; accepted 18.10.24; published 30.01.25

Please cite as:

Terhorst Y, Messner E-M, Opoku Asare K, Montag C, Kannen C, Baumeister H

Investigating Smartphone-Based Sensing Features for Depression Severity Prediction: Observation Study

J Med Internet Res 2025;27:e55308

URL: <https://www.jmir.org/2025/1/e55308>

doi: [10.2196/55308](https://doi.org/10.2196/55308)

PMID: [39883512](https://pubmed.ncbi.nlm.nih.gov/39883512/)

©Yannik Terhorst, Eva-Maria Messner, Kennedy Opoku Asare, Christian Montag, Christopher Kannen, Harald Baumeister. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 30.01.2025. 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 the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.