Sleep Medicine:X 7 (2024) 100114

Contents lists available at ScienceDirect

Sleep Medicine: X

journal homepage: www.sciencedirect.com/journal/sleep-medicine-x

The predictive value of supervised machine learning models for insomnia symptoms through smartphone usage behavior

Laura Simon [a,*], Yannik Terhorst [a], Caroline Cohrdes [b], Rüdiger Pryss [c], Lisa Steinmetz [d], Jon D. Elhai [e,f], Harald Baumeister [a]

[a] Institute of Psychology and Education, Department of Clinical Psychology and Psychotherapy, Ulm University, Lise-Meitner-Str. 16, Ulm, Germany
[b] Mental Health Research Unit, Department of Epidemiology and Health Monitoring, Robert Koch Institute Berlin, Nordufer 20, Berlin, Germany
[c] Institute of Clinical Epidemiology and Biometry, Julius-Maximilians-University of Würzburg, Sanderring 2, Würzburg, Germany
[d] Department of Psychiatry and Psychotherapy, Faculty of Medicine, Medical Center, University of Freiburg, Hauptstraße 5, Freiburg, Germany
[e] Department of Psychology, University of Toledo, 2801 West Bancroft Street, Toledo, OH, USA
[f] Department of Psychiatry, University of Toledo, 3000 Arlington Avenue, Toledo, OH, USA

ARTICLE INFO

Keywords:
Digital phenotyping
Mobile sensing
Insomnia
Smartphone usage
Machine learning

ABSTRACT

Introduction: Digital phenotyping can be an innovative and unobtrusive way to improve the detection of insomnia. This study explores the correlations between smartphone usage features (SUF) and insomnia symptoms and their predictive value for detecting insomnia symptoms.
Methods: In an observational study of a German convenience sample, the Insomnia Severity Index (ISI) and smartphone usage data (e.g., time the screen was active, longest time the screen was inactive in the night) for the previous 7 days were obtained. SUF (e.g., min, mean) were calculated from the smartphone usage data. Correlation analyses between the ISI and SUF were conducted. For the specification of the machine learning models (ML), 80 % of the data was allocated to training, 20 % to testing, and five-fold cross-validation was used. Six algorithms (support vector machine, XGBoost, Random Forest, k-Nearest-Neighbor, Naive Bayes, and Logistic Regressions) were specified to predict ISI scores $\geq$15.
Results: 752 participants (51.1 % female, mean ISI = 10.23, mean age = 41.92) were included in the analyses. Small correlations between some of the SUF and insomnia symptoms were found. In the ML models, sensitivity was low, ranging from 0.05 to 0.27 in the testing subsample. Random Forest and Naive Bayes were the best-performing algorithms. Yet, their AUCs (0.57, 0.58 respectively) in the testing subsample indicated a low discrimination capacity.
Conclusions: Given the small magnitude of the correlations and low discrimination capacity of the ML models, SUFs, as measured in this study, do not appear to be sufficient for detecting insomnia symptoms. Further research is necessary to explore whether examining intra-individual variations and subpopulations or employing alternative smartphone sensors yields more promising outcomes.

Abbreviations

Abbreviation	Definition
AUC	area under the receiver operating characteristic curve
autoR	autocorrelation
CASMIN	Comparative Analyses of Social Mobility in Industrial Nations
CI	confidence interval
GAD-7	Generalized Anxiety Disorder questionnaire-7
ISI	Insomnia Severity Index
LR	logistic regression

(*continued on next column*)

(*continued*)

Abbreviation	Definition
ML	machine learning
kNN	k-Nearest-Neighbor
NB	naive bayes
kurt	kurtosis
PHQ-9	Patient Health Questionnaire-9
RF	random forest
rmssd	square root of the average squared successive difference
skew	skewnes

(*continued on next page*)

* Corresponding author.
E-mail address: laura.simon@uni-ulm.de (L. Simon).

https://doi.org/10.1016/j.sleepx.2024.100114
Received 23 November 2023; Received in revised form 7 December 2023; Accepted 2 May 2024
Available online 4 May 2024

(*continued*)

Abbreviation	Definition
SVM	radial basis function support vector machines via kernlab
TRIPOD	Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis

1. Introduction

Up to 30 % of the adult population is affected by insomnia symptoms, and approximately 10 % fulfill the criteria for insomnia disorder [1]. Insomnia disorder is associated with a substantial personal [2,3] and societal burden [4,5]. Moreover, insomnia disorder increases the risk of developing or worsening other somatic and mental health conditions [6,7]. Hence, prompt diagnoses followed by guideline-compliant treatment are needed to reduce the consequences and costs associated with insomnia disorder. Nevertheless, insomnia disorder often remains undiagnosed and untreated [8–10].

This matter calls for innovative approaches to detect insomnia disorder and educate affected persons about available treatment possibilities. Given that insomnia disorder is not routinely screened for by primary care physicians [11] and the burden of filling out questionnaires, passive data collection methods could be of particular interest. Actigraphy, a passive data collection method, has been extensively studied in sleep research in the last decades. Yet, as evidence for the utility of actigraphy in the diagnostic process for insomnia disorder is limited [12], and wearing actigraphy can be inconvenient, the question arises whether our digital footprint may provide opportunities to enhance the detection of insomnia disorder. The analysis of individuals' interactions with digital technologies to measure behavior, cognition, and mood and using this information for monitoring and diagnosing potential disorders is called digital phenotyping [13–15], and the potential of digital phenotyping is increasingly attracting attention in various mental health domains [16] as depression [17,18] or schizophrenia and bipolar disorder [19] as digital phenotyping might provide innovative means for the prevention and diagnosis of mental disorders and the continuous monitoring of symptoms.

Given the ubiquity of smartphones [20] and the range of smartphone sensors that may be used to infer the (mental health) status of a person [21], data collected from smartphone sensors may offer an innovative and unobtrusive approach for detecting insomnia disorder [22]. Smartphone usage behavior (e.g., the duration a screen or application is active, duration in the time in which a smartphone is not used) appears to be particularly promising for detecting insomnia disorder given the potential associations between smartphone usage and insomnia disorder: Difficulties initiating or maintaining sleep or early awakenings may lead to irregular or excessive smartphone usage or vice versa. Indeed, research indicates an association between self-reported excessive smartphone usage and sleep disturbances in adolescents and young adults [23,24]. Thus far, the focus of digital phenotyping studies has not been directly on insomnia disorder but only on monitoring sleep duration and predicting sleep quality [25–27]. A systematic review [28] points to encouraging findings for the utilization of various smartphone sensors for digital phenotyping in other mental health domains, with the accuracy ranging between 59 and 86 % for depressive symptoms or precision rates up to 97 % for state changes in bipolar disorder.

Given the promising results of other mental health domains and to fill this research gap in the field of insomnia, this study investigates the predictive value of objectively measured smartphone usage behavior and self-reported insomnia symptoms. Given the limited research in this field and to overcome some limitations of traditional statistical analyses (e.g., overfitting, predictor collinearity, and linearity; [29]), we choose an exploratory approach utilizing supervised machine learning. In particular, a data set featuring a validated self-report questionnaire for assessing insomnia symptoms (ISI; [30])) and objectively measured smartphone usage data of the previous seven days was analyzed to investigate the following research questions:

1. What are the bivariate correlations between objectively measured smartphone usage behavior and insomnia symptoms?
2. Can supervised ML algorithms using features from objectively measured smartphone usage behavior be developed to classify groups of persons with or without insomnia symptoms?

2. Material and methods

2.1. Source of data

We present a sub-study of the CORONA HEALTH project, a large-scale observational smartphone application-based exploratory study evaluating the effects of the COVID-19 pandemic on mental health. This sub-study focuses on insomnia symptoms, and the research intentions have been preregistered on the Open Science Framework before the data collection (https://osf.io/wz4af). This work aims to answer research questions 4 and 5 of the pre-registration. Deviating from the registration, this work features correlation analyses between objectively measured smartphone usage data and insomnia symptoms, and the predictive value was analyzed using ML algorithms instead of multilevel regression models, given the exploratory nature of this study. Moreover, as neural networks require very large sample sizes, we abstained from specifying neural networks. The CORONA HEALTH study was conducted following the Declaration of Helsinki, the German medical products law, and approved by the ethics committee and data protection officer of the University of Würzburg (No. 130/20-me). The data were collected from July 17, 2020, to June 13, 2022. The assessments were conducted using a smartphone application developed explicitly for this project. Beierle et al. [31] provide a detailed overview describing the application and how the objectively measured smartphone usage data were measured.

2.2. Participants

Participants had to be aged eighteen or older to participate in the study. The sample consists of a German convenience sample recruited by a nationwide open recruitment strategy (e.g., institutional homepages of the research consortium, social media channels, mailing lists, and media reports). Study participation was voluntary, and each participant provided informed consent. Due to Apple's data protection regulations, only participants using Android smartphones, and consenting to share their objective smartphone usage data were included in this sub-study. Participants did not receive financial compensation. However, they were provided automated feedback on their well-being (based on standardized cut-off scores for symptoms of loneliness and depression). Additionally, a news ticker was implemented in the application to provide information about the COVID-19 pandemic. The study size was not a priori-defined. Thus, all individuals who downloaded the CORONA HEALTH application from the Google Play Store could participate in the study, given that they fulfilled the inclusion criteria.

2.3. Outcome and predictors

2.3.1. Self-reported data

Participants were asked to fill out a self-report questionnaire after downloading the application. Smartphone usage data for the day the self-report questionnaire was answered and the previous seven days were retrospectively collected.

Insomnia symptoms were the outcome with which correlations were explored, and that was to be predicted by the ML algorithms. We assessed insomnia symptoms using the German version of the Insomnia Severity Index (ISI; [30,32]). The ISI comprises seven 5-point Likert scale items (0–4 points; total score range: 0–28 points), which assess the perceived severity of insomnia symptoms. Deviating from the original

scale, participants were asked to rate insomnia symptoms for the past week for this study. The ISI has been shown to have good internal consistency and validity [30]. Moreover, for the descriptive data on the participants, symptoms of depression were assessed using the German version of the Patient Health Questionnaire-9 (PHQ-9; [33,34]) and anxiety using the German version of the Generalized Anxiety Disorder questionnaire-7 (GAD-7; [35,36]). Participants were asked to provide their highest educational degree. These answers were classified according to the Comparative Analyses of Social Mobility in Industrial Nations (CASMIN) into low (i.e., elementary to middle school education), moderate (i.e., high school and specialized vocational training), and high (i.e., at least short circle tertiary education or bachelor or equivalent university degree) education [37]. No particular actions have been taken to blind the assessment of the outcome and other self-report data. However, the automated and anonymous data collection may have contributed to reducing biases.

2.3.2. Smartphone usage data

The following smartphone usage data were collected and processed for the analyses: *ActiveTime*, describing the total time any application was visible in the screen's foreground; *UseTime*, describing the total time in which the screen was active (including no application being used or being woken up from a notification). Moreover, the time stamps of the start and end times, in which the smartphone was inactive for at least 1 h, were collected. As the smartphone usage data were passively and unobtrusively collected via the smartphone application, no measures to blind the assessments were employed.

2.4. Data processing

Data were investigated for the plausibility of answers (i.e., correspondence between similar items), careless responding (i.e., straightlining and intraindividual response variability), and extreme outliers (i.e., Mahalanobis distance). Following the scoring procedures of the ISI, the sum score was calculated [30]. Using the guidelines for scoring the ISI [30], we divided participants into two groups: no or subclinical insomnia symptoms ISI (sum scores 0–14) and insomnia symptoms (sum scores 15 and above).

Following Schoedel et al. [38], the longest nightly inactive smartphone period *(NightlyInactivty*; i.e., no active screen) was derived as follows: First, a time window between 6 p.m. (first day) and 2 p.m. (of the following day) was defined in which the nightly inactive smartphone period had to occur. Then, the longest period of inactivity within this time window and the respective time stamp marking the beginning and end of this period were extracted. The time stamp marking the beginning of the inactive smartphone period is likely to be the last time the participant used the smartphone before going to sleep. The time stamp marking the end of the period is likely to be the first time the person used the smartphone after awakening. Hence, we labeled these variables as *LastUsage* and *FirstUsage*.

2.4.1. Missing data

Participants who did not fill out the ISI or for which no objective smartphone usage data at all was available were excluded. Missingness in the smartphone usage data was allowed and handled in the preprocessing of the ML models through imputation via the k-Nearest-Neighbor approach.

2.4.2. Calculation of the smartphone usage features

In digital phenotyping, statistical indices are commonly calculated from the sensing data to be used as features in the predictive models. Therefore, for each processed smartphone usage variable (i.e., ActiveTime, UseTime, NightlyInactivity, LastUsage, FirstUsage), the mean, median, standard deviation (SD), minimum (min), and maximum (max) were calculated to be analyzed as features in the correlation analyses and ML models. Furthermore, for ActiveTime and UseTime, the following additional measures of variability were calculated to be used as features: the autocorrelation with the lag of 1 (autoR), the square root of the average squared successive difference (rmssd), and the entropy reflecting the variability between ActiveTime and non-ActiveTime. Moreover, the skewness (skew) and kurtosis (kurt) were calculated for distribution measures.

2.5. Statistical analyses

2.5.1. Correlation analyses

Correlations between the ISI sum score and the smartphone usage features (i.e., mean, median, SD, min, and max for ActiveTime, UseTime, NighltyInactivity, LastUsage, and FirstUsage) were calculated using Pearson's correlation. Additionally, for ActiveTime and UseTime, correlations between the ISI sum score and the following features were calculated: autoR, rmssd, entropy, skew, and kurt. Given the exploratory nature of this study, we chose to only report the 95 % confidence intervals (CI) and abstained from significance tests.

2.5.2. Predicting insomnia symptoms with supervised machine learning

Binary predictive classification ML algorithms were specified to categorize participants as experiencing insomnia symptoms (i.e., ISI sum scores of 15 and above) or no insomnia symptoms (i.e., ISI sum scores below 15) using the smartphone usage features as predictors. The following supervised ML algorithms were specified: Radial basis function support vector machines via kernlab (SVM), XGBoost (XGB), Random Forest (RF), k-Nearest-Neighbor (kNN), Naive Bayes (NB), and Logistic Regression (LR). We selected these algorithms because they are among the most accurate algorithms across research domains and in digital phenotyping [39]. An automatic grid search was used to optimize the respective hyperparameters.

To preserve the class distribution, the dichotomous outcome variable, thus the presence of insomnia symptoms, was used as the stratification variable; we randomly allocated 80 % ($n = 600$) of the data to the training and 20 % ($n = 152$) to the testing subsample. Five-fold repeated cross-validations were employed to reduce overfitting. Hence, the training data was split into five folds, and four folds were used to simulate the training data and the fifth fold to simulate test data. This process was repeated nine times, yielding a total of 50 replications. The models fitted in the training subsample were subsequently applied to the testing subsample.

We followed the model-specific preprocessing recommendations by Kuhn & Silge [40]. For all algorithms, missing data were imputed using the kNN approach. Near zero variance features and highly correlated features ($\geq$0.90) were removed. Data were transformed using a simple Yeo-Johnson transformation in SVM, kNN, and LR. Z-score normalization was used for SVM and kNN. The ML models' performance was evaluated using the following metrics: sensitivity, specificity, accuracy, and the area under the receiver operating characteristic curve (AUC).

A Bayesian random intercept model, which considers resampling, was fitted to derive the posterior distributions of the AUC, and the credible intervals from the model posterior distributions for the AUCs were plotted to compare the ML algorithms formally. For the best-performing algorithms, variable importance plots were generated, provided the algorithm allowed for the computation of variable importance scores.

2.6. Software/packages

The data preprocessing from the application was performed in Python [41]. The statistical software R was used for all analyses [42]. Correlations were calculated using the "psych" package [43]. The ML analyses were conducted using the "tidymodel" collection [44].

This study has been reported following the Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) recommendations [45]; the TRIPOD checklist is reported in

the supplemental material.

3. Results

3.1. Participants

2507 participants enrolled in the Corona Health Study. Of these, data from 752 participants were eligible for the present sub-study. See Fig. 1 for reasons for exclusion.

The participants of this sub-study covered a broad age range from 18 to 84 years (M = 41.92, SD = 13.61). Of the participants, 53.1 % identified as female, 46 % as male, and 0.9 % as diverse. Table 1 provides a summary of the sample characteristics. The sample was highly educated; 63 % of the respondents indicated a high education level, whereas only 7.7 % indicated a low education level.

The mean ISI value (M = 10.28, SD = 6.54) indicated that the sample had overall a subclinical insomnia severity [30]. Table 2 shows a categorization of the participants based on their insomnia severity. The study's distribution of insomnia symptoms aligns closely with the prevalence observed in other studies conducted in Germany [46].

3.2. Correlation of smartphone usage features and insomnia symptoms

Table 3 presents Pearson's correlations between smartphone usage features derived from the smartphone usage data (ActiveTime, UseTime, NightlyInactivy, LastUsage, FirstUsage) and insomnia symptoms. For the respective smartphone usage variables, ActiveTime SD (r = 0.13), UseTime kurt (r = −0.08), NightlyInactivity min (r = −0.13), LastUsage median (r = 0.13), and FirstUsage SD (r = 0.13) yielded the largest correlations.

Table 1
Sample characteristics.

	M (SD)/N(%)
Age	41.92 (13.61)
Gender	
female	399 (53.06 %)
male	346 (46.01 %)
diverse	7 (0.93 %)
CASMIN educational classification	
low	58 (7.71 %)
moderate	204 (27.13 %)
high	474 (63.02 %)
NA	16 (2.12 %)
Depression symptoms (PHQ-9)	9.61 (6.65)
Anxiety symptoms (GAD-7)	8.36 (5.19)
Insomnia symptoms (ISI)	10.28 (6.54)

Note. Data are *M* (*SD*) or %. CASMIN = Comparative Analyses of Social Mobility in Industrial Nations [37]; NA = not available; PHQ-9 = Patient Health Questionnaire-9 [33]; GAD-7 = Generalized Anxiety Disorder 7-item Scale [36]; ISI = insomnia severity index [30].

Table 2
Distribution of participants based on their insomnia severity.

Category	ISI sum score	n(%)
Classified as experiencing no insomnia symptoms		
No insomnia	0–7	294 (39.10 %)
Subthreshold insomnia	8–14	257 (34.18 %)
Classified as experiencing symptoms of insomnia		
Moderate insomnia	15–21	159 (21.14 %)
Clinically significant insomnia	22–28	42 (5.59 %)

Note. ISI = insomnia severity index [30].

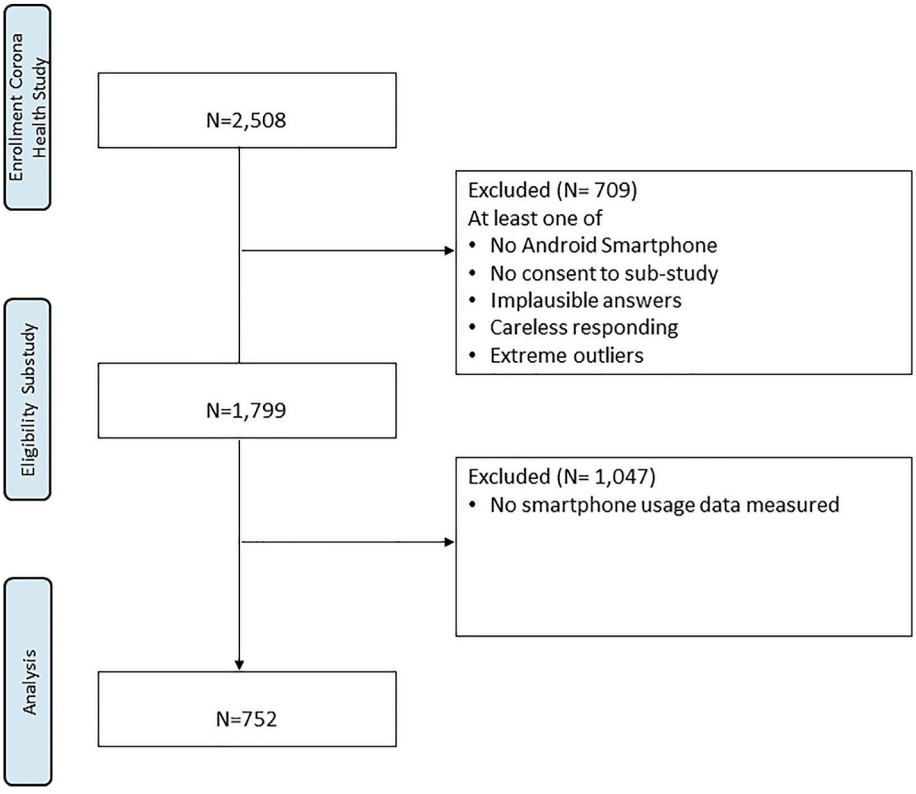

Fig. 1. Flowchart.
Note. Participants were excluded if no objective smartphone usage data at all was measured.

Table 3
Correlation analyses of the sensing features derived from the smartphone usage data.

	Pearson's r	Lower bound 95%- CI	Upper bound 95%- CI
ActiveTime			
Mean	**0.10**	**0.03**	**0.17**
Median	**0.09**	**0.01**	**0.16**
SD	**0.13**	**0.06**	**0.20**
Min	0.02	−0.05	0.09
Max	**0.12**	**0.05**	**0.19**
autoR	0.03	−0.05	0.10
rmssd	**0.12**	**0.05**	**0.19**
Skewness	−0.01	−0.09	0.07
Kurtosis	−0.01	−0.09	0.07
Entropy	**0.11**	**0.04**	**0.19**
Usetime			
Mean	0.03	−0.04	0.10
Median	0.03	−0.04	0.10
SD	0.03	−0.04	0.11
Min	0.00	−0.07	0.07
Max	0.03	−0.04	0.10
autoR	0.03	−0.04	0.11
rmssd	0.02	−0.05	0.09
Skewness	−0.07	−0.15	0.00
Kurtosis	**−0.08**	**−0.15**	**−0.01**
Entropy	−0.01	−0.08	0.06
NightlyInactivity			
Mean	**−0.11**	**−0.19**	**−0.03**
Median	**−0.11**	**−0.19**	**−0.03**
SD	**0.12**	**0.04**	**0.20**
Min	**−0.13**	**−0.21**	**−0.06**
Max	−0.04	−0.12	0.03
LastUsage			
Mean	**0.12**	**0.04**	**0.19**
Median	**0.13**	**0.05**	**0.20**
SD	**0.12**	**0.04**	**0.20**
Min	0	−0.07	0.08
Max	**0.11**	**0.03**	**0.19**
FirstUsage			
Mean	−0.01	−0.09	0.07
Median	0.00	−0.08	0.07
SD	**0.13**	**0.05**	**0.20**
Min	−0.08	−0.16	0.00
Max	0.05	−0.02	0.13

Note. Pearson's correlations. Given the explorative nature of this study p-values are not presented but non-zero including confidence intervals are highlighted in bold; CI = confidence interval; SD = standard deviation; autoR = autocorrelation; rmssd = square root of the average squared successive difference.

3.3. Predicting insomnia symptoms with supervised machine learning

The predictive performance of the supervised ML models trained with the smartphone usage features is presented in Table 4. The performance of the algorithms varied across the metrics. However, both the sensitivity and AUC were low across all investigated algorithms.

According to the credible intervals derived from the model posterior distributions of the AUCs (Fig. 2), RF and NB achieved the highest performance. Yet, even for RF and NB, the sensitivity was low (0.17 and 0.27, respectively) in the testing subsample, and the AUC was only at 0.57 for RF and 0.58 for NB in the testing subsample. Fig. 3 depicts the variable importance plot for RF. The variable importance plot indicates that seven of the ten most important features were associated with the

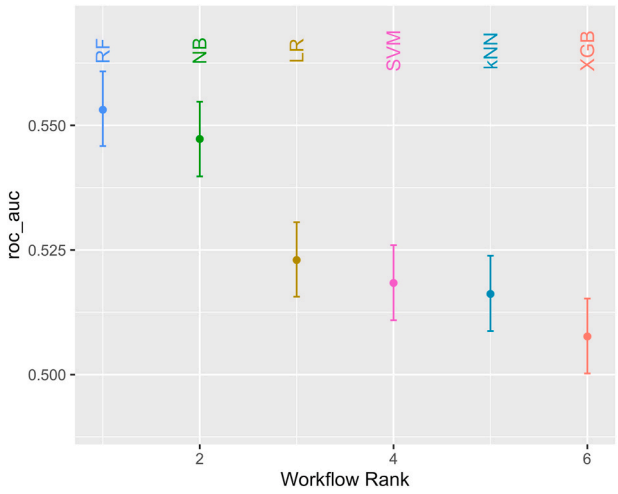

Fig. 2. Comparison of the machine learning algorithms.
Note. Credible intervals derived from the Bayesian random intercept model posterior distributions.
roc_auc = the area under the receiver operating characteristic curve; RF = Random Forest; NB = Naive Bayes; LR = Logistic Regression; SVM = Radial basis function support vector machines via kernlab; kNN = k-Nearest-Neighbor; XGB = XGBoost.

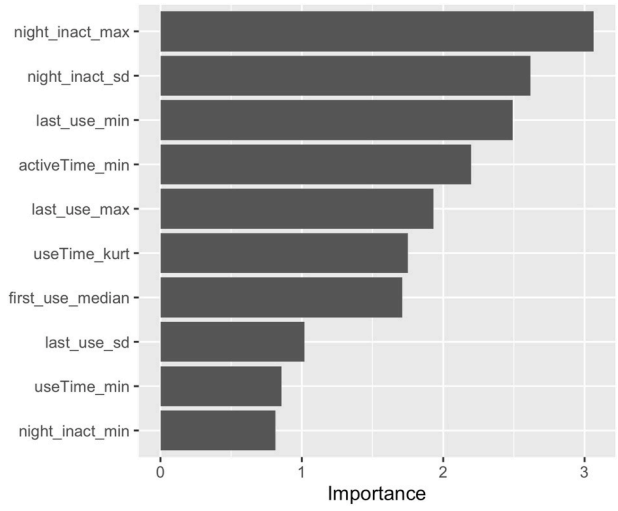

Fig. 3. Variable importance plot Random Forest.

Table 4
Machine learning algorithms predicting insomnia symptoms.

	Performance metrics for classification over repeated cross-validations in the training sample				Performance metrics for classification in the testing sample			
	Sensitivity	Specificity	Accuracy	AUC	Sensitivity	Specificity	Accuracy	AUC
Support vector machine	0.094	0.944	0.717	0.523	0.049	0.910	0.678	0.488
XGBoost	0.144	0.847	0.660	0.508	0.195	0.856	0.678	0.547
Random Forest	0.129	0.883	0.682	0.551	0.171	0.937	0.730	0.566
k-Nearest Neighbor	0.267	0.737	0.612	0.516	0.195	0.766	0.612	0.499
Naive Bayes	0.251	0.801	0.654	0.546	0.268	0.856	0.697	0.584
Logistic Regression	0.061	0.950	0.713	0.520	0.073	0.973	0.730	0.529

Note. AUC = Area under the curve. The machine learning algorithms were specified to categorize participants as experiencing insomnia symptoms (insomnia severity index [30] ≥15) or no insomnia symptoms (insomnia severity index [30]<15) using smartphone usage features as predictors.

features derived from the longest nightly inactivity.

4. Discussion

This study explored the correlations and predictive value of smartphone usage features for self-reported insomnia symptoms in a sample covering a large age range and a distribution of insomnia symptoms representative of the general population in Germany. Given the small magnitude of the correlations and low discrimination capacity of the ML models, smartphone usage features, as measured in this study, do not appear sufficient for detecting insomnia symptoms.

While the direction of the found correlations between the smartphone usage features is consistent with studies relying on self-report data [23,24], the magnitude of the found correlation was rather small ($r = 0.00$ to 0.13). The supervised ML models, which do not assume linear relationships, yielded suboptimal performances. The model comparison revealed that RF and NB exhibited the highest performance, albeit with AUCs of 0.57 (RF) and 0.58 (NB) and sensitivities of 0.17 (RF) and 0.27 (NB) in the testing subsample. Thus, the performance metrics indicate a limited ability to distinguish individuals with insomnia symptoms from those without and appear to miss the majority of individuals experiencing insomnia severity symptoms.

In light of these results, the question arises as to whether there is indeed no or only a minimal relationship between smartphone usage features and insomnia symptoms or if there is an underlying relationship that could not be identified in this study. One possibility could be that the relationship between smartphone usage features and insomnia symptoms does not hold across the population but that this relationship exists only in particular subpopulations. This possibility appears particularly plausible as insomnia is a heterogeneous disorder, and different subtypes of insomnia may exist [47]. For example, relatively robust evidence exists for a differentiation based on the objective sleep duration [47]. Given the possibility that distinct insomnia subtypes vary in their smartphone usage behaviors, it would be of interest to conduct digital phenotyping studies across the proposed insomnia subtypes, particularly as digital phenotyping might be an additional marker to differentiate between subtypes. Moreover, smartphone usage features may offer only very limited information for people who only sporadically interact with their smartphones. This may particularly apply to older adults, as younger populations tend to show more excessive smartphone usage behavior [48]. Hence, it may be plausible that the association between smartphone usage behavior and insomnia symptoms is more robust in younger populations. Indeed, most digital phenotyping studies target students or younger adults [49]. Instead, we investigated an extensive age range from 18 to 84 years in our study. However, we could not specify subgroup analyses investigating different age groups, as even larger sample sizes would be necessary.

Moreover, it could be the case that the relationship between insomnia symptoms and smartphone usage features is more robust if one considers intraindividual and not interindividual variations. In our analyses, only interindividual variations were studied. However, it could be that not particular smartphone usage patterns but individual changes in these patterns provide relevant information on insomnia symptoms. Therefore, exploring intraindividual variations in longitudinal data sets appears relevant, including multiple self-reports and objective measurements.

In this study, we have only explored smartphone usage features. However, smartphones have a range of sensors that may be leveraged. It could be that larger associations and a greater discrimination capacity may be found using the features of other smartphone sensors. For example, in a study examining the relationship between smartphone sensors and the psychological states depression, anxiety, and stress, no correlations were found for smartphone usage behavior, whereas correlations were found for GPS features [50]. This observation is consistent with a recent meta-analysis, which found between-person correlations between GPS features and depression [49]. While actigraphy alone may not be sufficient for identifying insomnia disorder [12], it may be fruitful to collate all available smartphone sensors (e.g., actigraphy, GPS, light sensor) and explore whether a combination of these sensors improves the discrimination capacity.

Besides the fact that only smartphone usage features were analyzed in this study, other limitations should be considered when interpreting the results. The sample consisted of highly educated participants and featured only Android users. Differences between Android and iOS users may exist [51], for example, because the monthly budget may influence the choice of smartphone [52]. Moreover, as this study was conducted during the COVID-19 pandemic, these exceptional circumstances may have confounded the data. In addition, as in all sensing studies relying on consumer devices, the variations of devices and software updates may have led to noise in the data collection [53]. Lastly, given the number of decisions that had to be taken for the preprocessing and specification of the algorithms, other decisions might have led to different results.

It is essential to consider that the acceptance towards digital phenotyping is currently low to moderate [54]. The limitations of this study and the low acceptance of digital phenotyping underscore the challenges of digital phenotyping. Exploring strategies to mitigate these challenges is crucial as the broad scope of potential applications of digital phenotyping could majorly improve the care of insomnia disorder. While this study focused on symptom detection, other areas of application have already been proposed: For example, digital phenotyping may also be utilized to tailor interventions to individual needs (i.e., just-in-time-adaptive interventions) or for relapse prevention strategies [22]. Therefore, in future studies, it is imperative to delve not only into the utility of specific features or models but also to consider societal conditions and, in particular, the acceptance of such new approaches.

5. Conclusion

Given that the investigation of digital phenotyping for insomnia disorder is currently in its infancy, this study is an essential first step in exploring this new field, notably as this work features a large sample size and a broad age range. However, the findings indicate that smartphone usage features, at least as analyzed in the current study, are insufficient for detecting insomnia symptoms. Future studies should investigate whether the performance of smartphone usage features increases for specific populations or in longitudinal designs and whether additional smartphone sensors can improve the discrimination capacity.

Statement

During the preparation of this work the authors used ChatGPT in order to improve the readability and language of some paragraphs. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data availability statement

Further data supporting this study's findings are available from the corresponding author upon reasonable request.

Funding statement

The authors received no financial support for the research, authorship, and/or publication of this article.

Ethics approval statement

The Corona Health Study was approved by the ethics committee and the data protection officer of the University of Würzburg, Germany (No. 130/20-me).

Patient consent statement

The participants provided informed consent prior to participating in the study.

CRediT authorship contribution statement

Laura Simon: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Yannik Terhorst:** Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization. **Caroline Cohrdes:** Writing – review & editing, Project administration, Investigation, Conceptualization. **Rüdiger Pryss:** Writing – review & editing, Software, Methodology, Investigation, Data curation, Conceptualization. **Lisa Steinmetz:** Writing – review & editing. **Jon D. Elhai:** Writing – review & editing, Methodology, Formal analysis. **Harald Baumeister:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jon D. Elhai receives royalties for several books published on post-traumatic stress disorder (PTSD); is a paid, full-time faculty member at the University of Toledo; occasionally serves as a paid expert witness on PTSD legal cases. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sleepx.2024.100114.

References

[1] Riemann D, Benz F, Dressle RJ, Espie CA, Johann AF, Blanken TF, Leerssen J, Wassing R, Henry AL, Kyle SD, Spiegelhalder K, Van Someren EJW. Insomnia disorder: state of the science and challenges for the future. J Sleep Res 2022;(March):1–18. https://doi.org/10.1111/jsr.13604.

[2] Benz F, Knoop T, Ballesio A, Bacaro V, Johann AF, Rücker G, Feige B, Riemann D, Baglioni C. The efficacy of cognitive and behavior therapies for insomnia on daytime symptoms: a systematic review and network meta-analysis. Clin Psychol Rev 2020;80(May):101873. https://doi.org/10.1016/j.cpr.2020.101873.

[3] Kyle SD, Morgan K, Espie CA. Insomnia and health-related quality of life. Sleep Med Rev 2010;14(1):69–82. https://doi.org/10.1016/j.smrv.2009.07.004.

[4] Daley M, Morin CM, LeBlanc M, Grégoire JP, Savard J, Baillargeon L. Insomnia and its relationship to health-care utilization, work absenteeism, productivity and accidents. Sleep Med 2009;10(4):427–38. https://doi.org/10.1016/j.sleep.2008.04.005.

[5] Thiart H, Ebert DD, Lehr D, Nobis S, Buntrock C, Berking M, Smit F, Riper H. Internet-based cognitive behavioral Therapy for insomnia: a health economic evaluation. Sleep 2016;39(10). https://doi.org/10.5665/sleep.6152.

[6] Benz F, Meneo D, Baglioni C, Hertenstein E. Insomnia symptoms as risk factor for somatic disorders : an umbrella review of systematic reviews and meta-analyses. J Sleep Res 2023;32(6). https://doi.org/10.1111/jsr.13984.

[7] Hertenstein E, Feige B, Gmeiner T, Kienzler C, Spiegelhalder K, Johann A, Jansson-Fröjmark M, Palagini L, Rücker G, Riemann D, Baglioni C. Insomnia as a predictor of mental disorders: a systematic review and meta-analysis. Sleep Med Rev 2019;43:96–105. https://doi.org/10.1016/j.smrv.2018.10.006.

[8] Benca RM. Diagnosis and treatment of chronic insomnia: a review. Psychiatr Serv 2005;56(3):332–43. https://doi.org/10.1176/appi.ps.56.3.332.

[9] Hassed C, Antoniades J, Jones K, Rajaratnam S, Kiropolous L, Naughton M, Piterman L. An examination of australian general practitioners' knowledge, attitudes and practices in relation to sleep disorders. Malays Fam Physician 2012;7(1):16–23.

[10] Roach M, Juday T, Tuly R, Chou JW, Jena AB, Doghramji PP. Challenges and opportunities in insomnia disorder. Int J Neurosci 2021;131(11):1058–65. https://doi.org/10.1080/00207454.2020.1773460.

[11] Yamamoto M, Lim CT, Huang H, Spottswood M, Huang H. Insomnia in primary care: considerations for screening, assessment, and management. J Med Access 2023;7. https://doi.org/10.1177/27550834231156727.

[12] Rösler L, van der Lande G, Leerssen J, Cox R, Ramautar JR, van Someren EJW. Actigraphy in studies on insomnia: worth the effort? J Sleep Res 2023;32(1):1–9. https://doi.org/10.1111/jsr.13750.

[13] Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA 2017;318(13):1215–6. https://doi.org/10.1001/jama.2017.11295.

[14] Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015;33(5):462–3. https://doi.org/10.1038/nbt.3223.

[15] Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016;41(7):1691–6. https://doi.org/10.1038/npp.2016.7.

[16] Dlima SD, Shevade S, Menezes SR, Ganju A. Digital phenotyping in health using machine learning approaches: scoping review. JMIR Bioinform Biotech 2022;3(1). https://doi.org/10.2196/39618.

[17] Abd-Alrazaq A, AlSaad R, Shuweihdi F, Ahmed A, Aziz S, Sheikh J. Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. Npj Digital Med 2023;6(1):1–16. https://doi.org/10.1038/s41746-023-00828-5.

[18] De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, Lavelle G, Matcham F, Pace A, Mohr DC, Dobson R, Hotopf M. Digital health tools for the passive monitoring of depression: a systematic review of methods. Npj Digital Med 2022;5(1):3. https://doi.org/10.1038/s41746-021-00548-8.

[19] Benoit J, Onyeaka H, Keshavan M, Torous J. Systematic review of digital phenotyping and machine learning in psychosis spectrum illnesses. Harv Rev Psychiatr 2020;28(5):296–304. https://doi.org/10.1097/hrp.0000000000000268.

[20] Statista. Smartphone-nutzung in deutschland. 2022. did-71707-1.

[21] Garatva P, Terhorst Y, Messner E-M, Karlen W, Pryss R, Baumeister H. Smart sensors for health research and improvement. 2023. p. 395–411. https://doi.org/10.1007/978-3-030-98546-2_23.

[22] Terhorst Y, Knauer J, Baumeister H. Smart sensing enhanced diagnostic expert systems. In: Montag C, Baumeister H, editors. Digital phenotyping and mobile sensing. Studies in neuroscience, psychology and behavioral economics. Springer International Publishing; 2023. p. 413–25. https://doi.org/10.1007/978-3-030-98546-2_24.

[23] Hale L, Guan S. Screen time and sleep among school-aged children and adolescents: a systematic literature review. Sleep Med Rev 2015;21:50–8. https://doi.org/10.1007/978-3-319-57196-6_33.

[24] Yang J, Fu X, Liao X, Li Y. Association of problematic smartphone use with poor sleep quality, depression, and anxiety: a systematic review and meta-analysis. Psychiatr Res 2020;284. https://doi.org/10.1016/j.psychres.2019.112686.

[25] Bai Y, Xu B, Ma Y, Sun G, Zhao Y. Will you have a good sleep tonight? sleep quality prediction with mobile phone. In: 7th international conference on body area networks; 2012.

[26] Langholm C, Jin A, Byun S, Mullington J, Torous J. Monitoring sleep using smartphone data in a population of college students. Npj Mental Health Res 2023;2(3):1–6. https://doi.org/10.1038/s44184-023-00023-0.

[27] Min JK, Doryab A, Wiese J, Amini S, Zimmerman J, Hong JI. Toss "N" turn: smartphone as sleep and sleep quality detector. Conf Human Factor Comput Syst - Proc 2014:477–86. https://doi.org/10.1145/2556288.2557220.

[28] Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inf 2018;77:120–32. https://doi.org/10.1016/j.jbi.2017.12.008.

[29] Elhai JD, Montag C. The compatibility of theoretical frameworks with machine learning analyses in psychological research. Curr Opinion Psychol 2020;36:83–8. https://doi.org/10.1016/j.copsyc.2020.05.002.

[30] Bastien CH, Vallières A, Morin CM. Validation of the insomnia severity index as an outcome measure for insomnia research. Sleep Med 2001;2(4):297–307. https://doi.org/10.1016/S1389-9457(00)00065-4.

[31] Beierle F, Schobel J, Vogel C, Allgaier J, Mulansky L, Haug F, Haug J, Schlee W, Holfelder M, Stach M, Schickler M, Baumeister H, Cohrdes C, Deckert J, Deserno L, Edler JS, Eichner FA, Greger H, Hein G, Pryss R. Corona health—a study-and sensor-based mobile app platform exploring aspects of the covid-19 pandemic. Int J Environ Res Publ Health 2021;18(14). https://doi.org/10.3390/ijerph18147395.

[32] Gerber M, Lang C, Lemola S, Colledge F, Kalak N, Holsboer-Trachsler E, Pühse U, Brand S. Validation of the German version of the insomnia severity index in adolescents, young adults and adult workers: results from three cross-sectional studies. BMC Psychiatr 2016;16(1):1–14. https://doi.org/10.1186/s12888-016-0876-8.

[33] Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

[34] Löwe B, Zipfel S, Herzog W. Deutsche Übersetzung und Validierung des "Brief Patient Health Questionnaire (Brief PHQ).". Pfizer; 2002.

[35] Löwe B, Müller S, Brähler E, Kroenke K, Albani C, Decker O. Validierung und Normierung eines kurzen Selbstratinginstrumentes zur Generalisierten Angst (GAD-7) in einer repräsentativen Stichprobe der deutschen Allgemeinbevölkerung. PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie 2007;57(2):A050.

[36] Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092–7. https://doi.org/10.1001/archinte.166.10.1092.

[37] Schneider SL. Die Konzeptualisierung, Erhebung und Kodierung von Bildung in nationalen und internationalen Umfragen. GESIS—Leibniz-Institut Für Sozialwissenschaften; 2015. *Retrieved from*, https://Www.Gesis.Org/Filea.Dmin/Upload/SDMwi.Ki/Archiv/Messu.Ng_Von_Bildu.Ng_Schne.Ider_1. 0_, 8102, 15.

[38] Schoedel R, Pargent F, Au Q, Völkel ST, Schuwerk T, Bühner M, Stachl C. To challenge the morning lark and the night owl: using smartphone sensing data to investigate day–night behaviour patterns. Eur J Pers 2020;34(5):733–52. https://doi.org/10.1002/per.2258.

[39] Opuku Asare K, Moshe I, Terhorst Y, Vega J, Hosio S, Baumeister H, Pulkki-Råback L, Ferreira D. 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. https://doi.org/10.1016/j.pmcj.2022.101621.
[40] Kuhn M, Silge J. Tidy modeling with R. O'Reilly Media, Inc.; 2022.
[41] Van Rossum G, Drake FL. Python 3 reference manual. CreateSpace 2009.
[42] R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2023. https://www.r-project.org/.
[43] Revelle W. Psych: procedures for psychological, psychometric, and personality research. Northwestern University 2023;2(3.9). 2.3.9, https://cran.r-project.org/package=psych.
[44] Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. 2020. Boston, MA, USA.
[45] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. Circulation 2015;131(2):211–9. https://doi.org/10.1186/s12916-014-0241-z.
[46] Baglioni C, Altena E, Bjorvatn B, Blom K, Bothelius K, Devoto A, Espie CA, Frase L, Gavriloff D, Tuuliki H, Hoflehner A, Högl B, Holzinger B, Järnefelt H, Jernelöv S, Johann AF, Lombardo C, Nissen C, Palagini L, Riemann D. The European academy for cognitive behavioural therapy for insomnia: an initiative of the European insomnia network to promote implementation and dissemination of treatment. J Sleep Res 2020;29(2):1–29. https://doi.org/10.1111/jsr.12967.
[47] Nyhuis CC, Fernandez-Mendoza J. Insomnia nosology: a systematic review and critical appraisal of historical diagnostic categories and current phenotypes. J Sleep Res 2023;(April):1–21. https://doi.org/10.1111/jsr.13910.
[48] Csibi S, Griffiths MD, Demetrovics Z, Szabo A. Analysis of problematic smartphone use across different age groups within the 'components model of addiction.'. Int J Ment Health Addiction 2019:616–31. https://doi.org/10.1007/s11469-019-00095-0.
[49] Terhorst Y, Knauer J, Philippi P, Baumeister H. The Relation between passively collected GPS features and depressive symptoms : a systematic review and meta-analysis . Table of Contents. JMIR Preprints 2023:51875. 11/09/2023.
[50] Moshe I, Terhorst Y, Opoku Asare K, Sander LB, Ferreira D, Baumeister H, Mohr DC, Pulkki-Råback L. Predicting symptoms of depression and anxiety using smartphone and wearable data. Front Psychiatr 2021;12(January):1–12. https://doi.org/10.3389/fpsyt.2021.625247.
[51] Pryss R, Schlee W, Reichert M, Kurthen I, Giroud N, Jagoda L, Neuschwander P, Meyer M, Neff P, Schobel J, Hoppenstedt B, Spiliopoulou M, Langguth B, Probst T. Ecological momentary assessment based differences between android and iOS users of the TrackYourHearing mHealth crowdsensing platform. In: Proceedings of the annual international conference of the IEEE engineering in medicine and biology society. EMBS; 2019. p. 3951–5. https://doi.org/10.1109/EMBC.2019.8857854.
[52] Gotz FM, Stieger S, Reips UD. Users of the main smartphone operating systems (iOS, Android) differ only little in personality. PLoS One 2017;12(5):1–18. https://doi.org/10.1371/journal.pone.0176921.
[53] Stisen A, Blunck H, Bhattacharya S, Prentow TS, Kjærgaard MB, Dey A, Sonne T, Jensen MM. Smart devices are different: assessing and mitigating mobile sensing heterogeneities for activity recognition. In: SenSys 2015 - proceedings of the 13th ACM conference on embedded networked sensor systems; 2015. p. 127–40. https://doi.org/10.1145/2809695.2809718.
[54] Terhorst Y, Weilbacher N, Suda C, Simon L, Messner EM, Sander LB, Baumeister H. Acceptance of smart sensing: a barrier to implementation—results from a randomized controlled trial. Frontiers in Digital Health 2023;5(July):1–10. https://doi.org/10.3389/fdgth.2023.1075266.