

## Research and Applications

# Depression screening using mobile phone usage metadata: a machine learning approach

Rouzbeh Razavi,<sup>1</sup> Amin Gharipour,<sup>2</sup> and Mojgan Gharipour<sup>3</sup>

<sup>1</sup>Department of Management and Information Systems, Kent State University, Kent, OH, USA, <sup>2</sup>School of Information Technology, Griffith University, Gold Coast, Australia, and <sup>3</sup>Isfahan Cardiovascular Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

Corresponding Author: Rouzbeh Razavi, Department of Management and Information Systems, College of Business Administration, Kent State University, 475 Terrace Dr, Kent, OH 44240, USA (rrazavi@kent.edu)

Received 22 August 2019; Revised 2 December 2019; Accepted 15 December 2019

### ABSTRACT

**Objective:** Depression is currently the second most significant contributor to non-fatal disease burdens globally. While it is treatable, depression remains undiagnosed in many cases. As mobile phones have now become an integral part of daily life, this study examines the possibility of screening for depressive symptoms continuously based on patients' mobile usage patterns.

**Materials and Methods:** 412 research participants reported a range of their mobile usage statistics. Beck Depression Inventory—2nd ed (BDI-II) was used to measure the severity of depression among participants. A wide array of machine learning classification algorithms was trained to detect participants with depression symptoms (ie, BDI-II score  $\geq 14$ ). The relative importance of individual variables was additionally quantified.

**Results:** Participants with depression were found to have fewer saved contacts on their devices, spend more time on their mobile devices to make and receive fewer and shorter calls, and send more text messages than participants without depression. The best model was a random forest classifier, which had an out-of-sample balanced accuracy of 0.768. The balanced accuracy increased to 0.811 when participants' age and gender were included.

**Discussions/Conclusion:** The significant predictive power of mobile usage attributes implies that, by collecting mobile usage statistics, mental health mobile applications can continuously screen for depressive symptoms for initial diagnosis or for monitoring the progress of ongoing treatments. Moreover, the input variables used in this study were aggregated mobile usage metadata attributes, which has low privacy sensitivity making it more likely for patients to grant required application permissions.

**Key words:** depression, mobile usage, mobile health, machine learning

## INTRODUCTION

Depression is known as 1 of the most prevalent psychiatric disorders and affects nearly 15% of individuals globally.<sup>1</sup> An estimated 10%–20% of primary care visits are related to depression, making it the second most prevalent chronic disorder observed by primary care physicians.<sup>2</sup> Depression is also known to impose more societal burdens than any other mental and neurological disorder.<sup>3</sup> In the US

alone, the loss of productivity as a result of depression is estimated to cost \$44 billion annually.<sup>4</sup>

Depression is treatable using different methods, such as antidepressants and psychotherapy. However, many individuals who need such treatments may not receive it, since depression remains undiagnosed in many cases. For example, it is reported that primary care physicians, who deliver the majority of care for depression,<sup>5</sup> identify

only about 50% of depression cases.<sup>2</sup> Part of the challenge in diagnosing depression is the fact that the correct diagnosis requires patients to recall how they felt over long periods of time, which makes the diagnosis prone to errors and judgment biases.<sup>6,7</sup> Moreover, early diagnosis of depression may significantly benefit patients since early treatment and intervention is shown to be associated with a better prognosis.<sup>8</sup> Therefore, the possibility of screening for depressive symptoms passively and continuously would be highly useful.

Moreover, mobile devices have become an integral part of the modern lifestyle. It is estimated that more than 5 billion people have a mobile device, with more than 60% of those (77% in developed countries) being smart devices.<sup>9</sup> In addition to their widespread penetration, mobile devices are taking center stage in users' lives. A study by the research firm, Dscout, shows that a typical user initiates 76 separate sessions a day and touches his or her mobile device 2617 times every day.<sup>10</sup> In light of how closely users are attached to their mobile devices (physically, cognitively, and emotionally), their interactions with their devices leave behind unique footprints that can reveal a wide array of their characteristics. Most smart mobile devices are now equipped with built-in sensors and standard Application Programming Interfaces (APIs) to measure and collect users' data, including their mobility patterns (eg, using GPS traces),<sup>11</sup> physical activity (eg, using accelerometers), crowd density (eg, using Bluetooth detection), and the time spent indoor versus outdoor (eg, Wi-Fi signal measurements).<sup>11</sup>

With regard to depression, some previous studies have shown correlations between different mobile phone usage attributes and depressive behaviors.<sup>3,11–19</sup> For example, depressed users have been found to make fewer calls and browse the internet on their mobile phones less frequently.<sup>2</sup> By deploying a wide range of machine learning classification algorithms (which are capable of capturing complex and nonlinear relationships) and analyzing the mobile usage metadata of more than 400 demographically diverse individuals, this study seeks to examine the possibility of detecting depressive symptoms from individuals' mobile usage patterns. More specifically, this study tries to answer the following questions:

1. How do mobile usage attributes differ for depressed and not-depressed users?
2. How accurately can mobile usage attributes and mobile device characteristics detect depressive symptoms?
3. How does the prediction accuracy increase if basic demographic attributes, such as age and gender, are additionally included?

This research is different from the existing literature from the following perspectives. First, while the relationship between depression and some mobile usage attributes was examined in some past studies, with the exception of the study by Saeb et al,<sup>3</sup> none of the previous studies quantified the predictive power of such attributes for depression diagnosis classifications.<sup>13,15,16,18,19</sup> Understanding the predictive power of mobile usage attributes is required to assess the feasibility, reliability, and accuracy of using these attributes for depression screening. Second, unlike previous studies that used privacy-sensitive information (such as user's time–location information)<sup>3</sup> or content attributes information (such as the content of social media posts),<sup>20</sup> this study relies only on aggregated metadata (ie, context) mobile usage attributes for depression diagnosis. Considering the increasing importance of privacy for adopting mobile health applications by users,<sup>21</sup> the sensitivity of the information collected by such applications is likely to influence their adoption rate by

users. Third, most past studies have considered either small sample sizes (eg,  $N=8$  in<sup>14</sup> and  $N=28$  in<sup>3</sup>) or less representative samples (eg, university students in<sup>15,16,18,19</sup>) in their studies, which makes it difficult to generalize the study findings. Finally, previous studies used only linear models to explore the relationship between depression and mobile usage attributes, which, unlike machine learning models, cannot capture complex and nonlinear relationships that may exist between the dependent and independent variables. This may result in underestimating the predictive power of mobile usage attributes for depression screening.

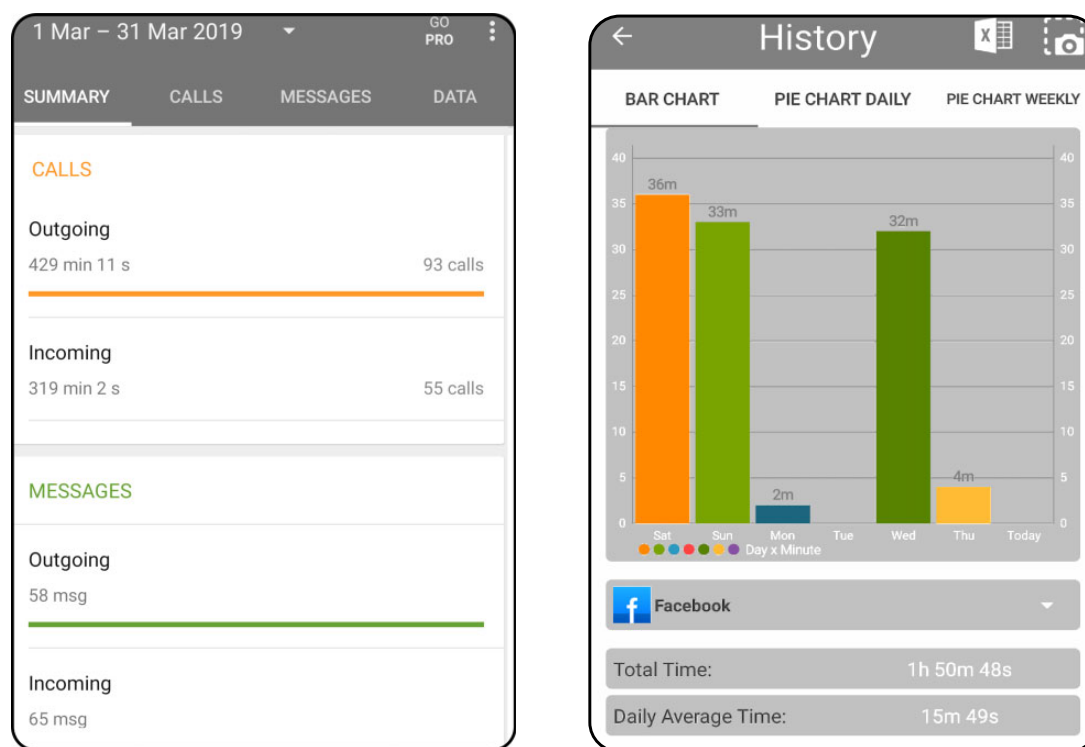
## MATERIALS AND METHODS

### Data source

Data collection was conducted using a survey through the Amazon Mechanical Turk (MTurk) portal. MTurk provides an efficient way to collect data from geographically distributed survey participants. Past research indicates that MTurk participants are more demographically diverse than other standard internet samples and that the collected data is as reliable as that of traditional methods.<sup>22</sup> Survey eligibility criteria were presented at the beginning, which required individuals to be at least 18 years old, be a resident of the United States, own an Android smartphone, and be willing to install 2 specific mobile applications to report their mobile usage statistics. Respondents were also required to answer all survey questions, hence there was no missing data. Upon successful completion of the survey, participants were paid a reward of \$2.50.

The first part of the survey captured respondents' demographic information. The second part of the survey captured information about participants' mobile devices and their usage. For usage attributes, the average number of daily calls and text messages (inbound and outbound) and the average time spent on talking, social media, web browsing, and entertainment applications were recorded. To ensure that the mobile usage attributes were accurately captured, participants were asked to download 2 mobile applications: "Callistics" (ver. 2.6.8) and "StayFree" (ver. 3.0.9) and to report their usage statistics over the past 14 days prior to completion of the survey. Callistics is a manager app that provides historical statistics about the number (and duration) of calls and text messages, while StayFree provides information about the time spent on each application. Figure 1 shows the snapshots of these 2 mobile applications (The requirement to install the 2 mobile applications was stated at the beginning of the survey. Moreover, to check if participants had installed these mobile apps, they were asked to provide the name of the third menu option on each of the 2 apps. Instructions to locate this information on the apps were provided). In addition, age, make, and model of participants' mobile devices were captured. The estimated dollar value of devices was subsequently determined using the Kimovil website (<https://www.kimovil.com/en/>).

Moreover, this study used Beck Depression Inventory—2nd ed (BDI-II),<sup>23</sup> which is 1 of the most reliable and widely used psychometric tests to measure the severity of depression. The validity and reliability of BDI-II has been examined and confirmed in several studies including studies considering individuals in different age groups,<sup>24,25</sup> in different cultures,<sup>26,27</sup> and among individuals with various disorders.<sup>28,29</sup> The test consists of 21 multiple choice questions each assessing a different depressive symptom that participants may have experienced in the last 14 days, including feelings of frustration and hopelessness, cognitions such as guilt or feelings of being punished, as well as physical symptoms such as weight loss, fatigue,



**Figure 1.** A snapshot of a) *Callistics* and b) *StayFree* applications.

and lack of interest in sex. Each symptom is on a 4-point rating scale (0–3). The BDI-II is scored by summing the ratings, which ranges from 0 to 63. According to the manual,<sup>23</sup> scores between 0 and 13 indicate minimal, between 14 and 19 mild, between 20 and 28 moderate, and between 29 and 63 severe depression. In this study, depression is defined based on the BDI-II scores. More specifically, considering the importance of detecting depressive symptoms as early as possible in this study, classification models are trained to separate participants with minimal depression from those with mild to severe depression (ie, BDI-II score  $\leq 13$  versus BDI-II score  $\geq 14$ ).

Responses from a total of 456 participants were gathered. However, 35 respondents failed to provide valid answers to the questions about the applications' menu options, which were asked to ensure that participants installed those apps; 9 respondents answered at least 1 of the 2 attention check questions of the survey incorrectly. Moreover, an outlier detection algorithm based on the indegree number algorithm,<sup>30</sup> suitable for detecting outliers in questionnaire data,<sup>31</sup> was applied to responses from the remaining participants. However, no outlier was found by the algorithm. Out of the remaining 412 participants, 210 (51%) were women, and 202 (49%) were men, and the average age was 40.27 (SD = 18.87).

### Machine learning classification algorithms and performance measures

To capture complex and nonlinear relationships between the individuals' mobile usage attributes and their depressive symptoms, an array of machine learning classification algorithms was used. Each classification algorithm has its own strengths and shortcomings.<sup>32</sup> Thus, to ensure that the reported predictive power of input attributes was not limited by the capabilities of a given algorithm, various algorithms were examined in this study. Table 1 shows the classification algorithms used in this study along with a list of their tuning parameters.

To avoid overfitting, data was split into 2 mutually exclusive subsets: training (80% of records) and test (20% of records) sets.<sup>32</sup> Including 80% of the data in the training set allows for better training of the model and tuning of the hyperparameters. Subsequently, using the adaptive search algorithm proposed in,<sup>32</sup> the optimal values of tuning parameters were found through 10-fold cross-validations, using only the training set. The performance of the model was then benchmarked on the test subset. This process ensured that test samples had not been exposed to the model during the training or parameter tuning steps, which minimized the risk of overfitting.<sup>32</sup> To ensure that the performance of the models was assessed on a sufficient number of samples and that the reported prediction performance was not specific to a single test subset, the above process was repeated 10 times, each time with new random training and test partitions and performance results were subsequently averaged.<sup>32</sup> A balanced random splitting method proposed in<sup>32</sup> was used to ensure the similarity of data distribution in the training and test sets.

With regard to performance metrics, this study used area under the curve (AUC), balanced accuracy, sensitivity, and specificity. Balanced accuracy is calculated by averaging the proportion of data points that are correctly classified for each class, and it is appropriate when there are class imbalance (eg, in this study where there are more nondepressed samples). Sensitivity and specificity are the fraction of positive and negative cases, respectively, that are correctly identified by the model. Formally, these measures can be defined as follows:

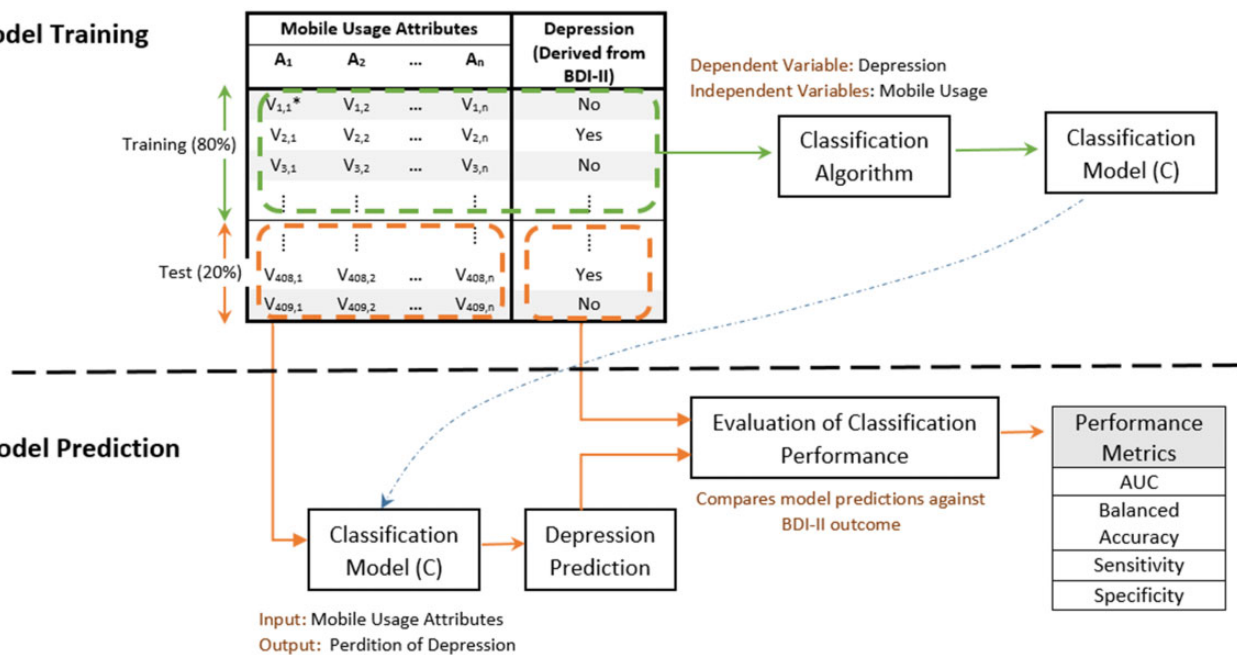
$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad \text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Balanced Accuracy} = \frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right)$$

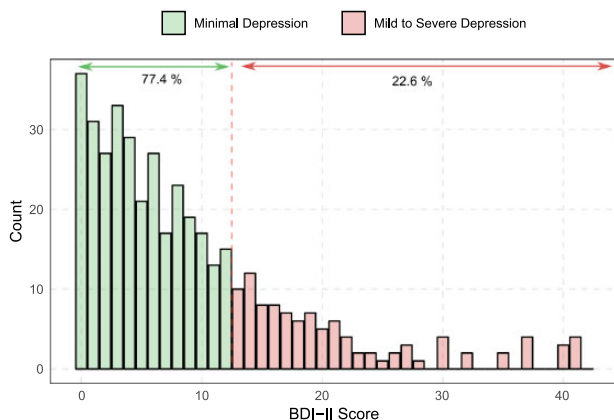
where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. Once a classification

**Table 1.** List of machine learning algorithms used in this study and their tuning parameters

Algorithm	Acronym	Tuning parameters
Classification and Regression Trees	CART	Complexity Parameter, <i>cp</i>
Gradient Boosting Machines	GBM	No. of trees, <i>n.trees</i> Depth, <i>interaction.depth</i> Shrinkage parameter, <i>shrinkage</i>
K-Nearest Neighbors	KNN	No. of Neighbors, <i>k</i>
Logistic Regression	LR	—
Neural Networks	NN	Weight decay, <i>decay</i> No. of hidden layer units, <i>size</i>
Random Forests	RF	No. of variables available for splitting, <i>mtry</i>
Support Vector Machines-Linear Kernel	SVM-L	Cost parameter, <i>c</i>
SVM-Radial Basis Function Kernel	SVM-RBF	Cost parameter, <i>c</i>

**Model Training**

\*  $V_{m,n}$  represents the value of the  $n^{\text{th}}$  mobile attributes of the  $m^{\text{th}}$  participant

**Figure 2.** An illustrative overview of procedures for model training and model training.**Figure 3.** Distribution of BDI-II scores. Classification models are trained to separate individuals with minimal depression from those with mild to severe depression.

model is constructed, the importance of the independent variables can be quantified by removing each of the variables from the model and by measuring the level of performance drop as a result. Following the previously proposed methodology,<sup>32</sup> the variable importance values were scaled to have a maximum value of 100 in this study. Figure 2 illustrates an overview of the steps involved in the model training and prediction phases, where training data is used for model construction, and then the model is used for making predictions on test data.

**RESULTS****Depression and mobile usage attributes**

Figure 3 shows the frequency distribution of BDI-II scores, where 93 (22.6%) participants were shown to have mild to severe depression. The BDI-II scores were internally consistent (Cronbach's  $\alpha = 0.92$ ). Table 2 shows the variations of demographics and mobile usage attributes for depressed (ie, BDI-II score  $\geq 14$ ) and not-depressed

**Table 2.** The mean and standard deviations (SD) of participants' mobile usage and device characteristics attributes. The \*\*\*, \*\*, and \* refer to the 1%, 5%, and 10% levels of statistical significance, respectively

Mobile usage attribute	No/Minimal Depress/Mild to Severe Depression				P value
	Mean	SD	Mean	SD	
Avg. Total Daily Mobile Usage (min)	127.13	111.21	164.48	135.08	0.012 **
Avg. No. of Calls Received Daily	5.15	4.08	4.23	3.82	0.044 **
Avg. Duration of Received Calls (min)	6.12	4.51	5.29	4.50	0.082 *
Avg. No. of Calls Initiated Daily	4.36	2.95	3.35	2.54	0.001 ***
Avg. Duration of Initiated Calls (min)	7.89	5.30	6.84	4.36	0.036 **
Avg. No. of Text Messages Received Daily	17.89	22.73	17.05	21.36	0.74
Avg. No. of Text Messages Sent Daily	13.55	13.63	18.01	17.66	0.021 **
Avg. Time Spent Web Browsing Daily (min)	39.1	40.53	44.85	43.64	0.2581
Avg. Time Spent on Social Media Apps Daily (min)	39.55	38.88	54.24	49.72	0.009 ***
Avg. Time Spent on Entertainment Apps (min)	41.34	48.55	47.6	54.88	0.32
No. of Saved Contacts on the Device	72.57	78.88	56.43	58.05	0.003 ***
Estimated Device Value (\$)	491.94	260.85	414.78	223.07	0.005 ***
Device Age (months)	20.19	13.04	22.49	13.53	0.14

(ie, BDI-II score  $\leq 13$ ) participants, where 2-sample *t*-tests on mean/proportion differences were additionally performed. According to the table, compared to participants with no or minimal depression, the overall mobile usage was notably higher for participants with depression (164.48 minutes/day versus 127.13 minutes/day). Participants with depressive symptoms were found to make and receive fewer calls, and their calls (incoming and outgoing) were shorter. For example, participants with depression made 30% fewer calls than those with minimal depressive symptoms. Instead, participants with depressive symptoms were found to send 33% more text messages, and spend 37% more time on social media applications.

The results shown in Table 2 suggest that participants with more extensive social ties (eg, those with a higher number of contacts and calls<sup>33</sup>) are less likely to show depressive symptoms, which is in line with past literature that shows a positive correlation between social ties and psychological well-being.<sup>34</sup> In addition, the results show a positive correlation between depression and the intensive use of mobile devices, where the overall mobile device usage was 29.3% higher for participants with depression. The positive association between depression and increased use of mobile devices was also reported in the past literature.<sup>35</sup> Finally, previous research shows that people with depressive symptoms tend to choose less direct communication channels,<sup>35</sup> which may explain the increased number of text messages sent and reduced number of outgoing calls among study participants with depression.

Moreover, participants with depression were shown to own cheaper mobile devices, on average, which is likely due to the income effect. Literature suggests a negative correlation between income and the prevalence of depression.<sup>36,37</sup> A positive correlation (0.48) was observed between participants' income and the estimated value of their mobile devices. The results show no statistically significant pattern between depression and the use of entertainment applications, web browsing, and the number of text messages received. In addition to mobile usage attributes, demographic variables such as gender and age have been reported to be associated with the prevalence of depression.<sup>36–38</sup> In this study, the average age for participants with no or minimal depression was 35.41 (SD = 16.94), and it was 41.33 (SD = 19.19) for those with mild to severe depression (*P* value = .004). Moreover, 27.14% (ie, 57 out of 210) of female participants and 17.82% (ie, 36 out of 202) of male participants had mild to severe depression (*P* value = .031).

Moreover, to examine the potential multi-collinearity effects as a result of correlations between independent variables, a variance inflation factor (VIF) analysis, which is widely used for detecting multi-collinearity,<sup>39</sup> was performed. For a given variable, the VIF measures how much the variance of the estimated coefficient is increased because of collinearity. VIF has a minimum value of 1 (ie, the complete absence of collinearity), and values above 5 indicate a level of multi-collinearity that can be problematic for modeling (ie, inaccurate estimation of model's parameters or unreliable evaluation of the predictive power of variables).<sup>39</sup> In this study, the VIF values have been shown to be smaller than 2 for all independent variables.

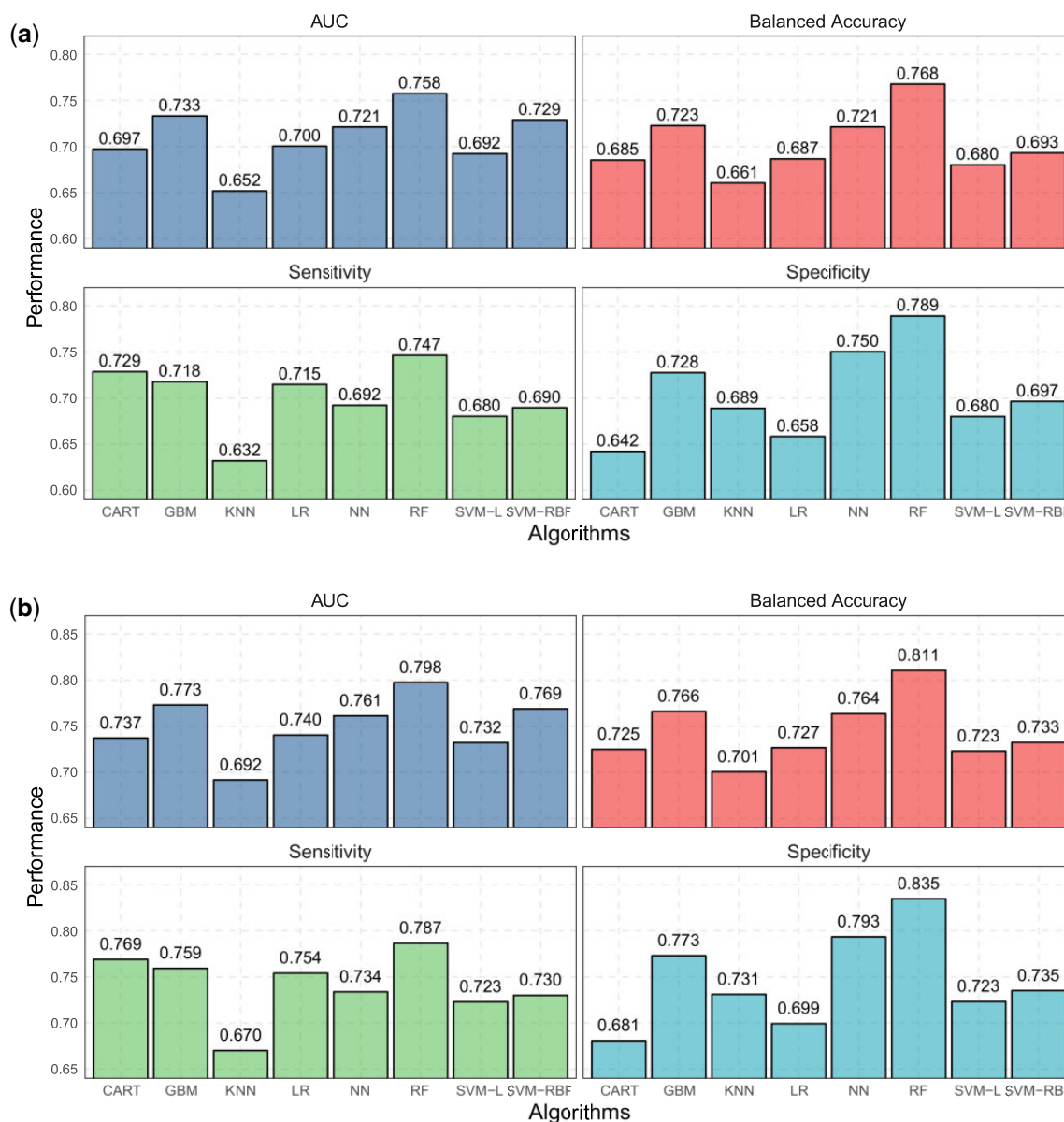
### Predicting depression from mobile usage attributes

Figure 4a shows the performance of various classification algorithms for detecting depressed users from mobile usage attributes listed in Table 2. Random forest (RF) classifier showed the best performance, where the balanced accuracy was 0.768. The examination of sensitivity and specificity metrics suggests that most classification models have comparable capabilities in detecting positive and negative cases. Moreover, the analysis of misclassified samples showed that 81.2% of misclassified samples had a BDI-II score between 7 and 20 (ie, mostly borderline individuals with minimal or mild depression). This observation also suggests that the small representation of participants with severe depression (ie, BDI-II score  $> 29$ ) in the training samples, as shown in Figure 3, did not impact the performance of the classification model significantly. No statistically significant differences were observed between the age (*P* = .72) and gender (*P* = .45) of misclassified and correctly classified participants. By additionally including users' age and gender as 2 independent variables, the results in Figure 4 show that the prediction performance improved, where balanced accuracy reached 0.811. Finally, Figure 5 shows the relative importance of the top 10 attributes in the RF model, where the average number of outgoing calls is the most significant variable, followed by the total average daily usage (RVI = 94.17), and the number of contacts saved on the device (RVI = 84.68).

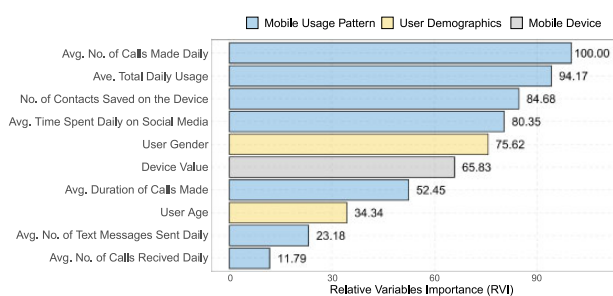
## DISCUSSION

The results showed the significant predictive power of mobile usage attributes in inferring users' depressive symptoms. The mobile usage





**Figure 4.** Predictive performance of various classification algorithms to detect users with depression from a) mobile usage attributes and b) mobile usage attributes and users' age and gender.



**Figure 5.** Significance of the top 10 variables in the RF model.

attributes and device characteristics considered in this study can be obtained through different mobile application permissions ("Callistics" and "StayFree" apps). For example, in Android systems, many of the predictive mobile usage attributes (eg, frequency

and duration of incoming and outgoing calls) can be collected when users grant an application the "Read Call Log" permission. The analysis of more than 1 million Android applications in<sup>40</sup> revealed that this permission is the 20th most common permission (out of 235 permissions) requested by Android applications. Mental health mobile applications can ask users to provide their age and gender, or alternatively, this information can be estimated from other mobile attributes. For example, it is shown that the list of installed mobile applications could reveal users' gender with an accuracy of 82.3%.<sup>41</sup>

Once these attributes are collected, a mobile health application can use a pretrained predictive classification model to diagnose depression. Integrating a trained classification model into mobile applications makes it possible to continuously screen depression in a passive way and over long periods either for diagnosis or for monitoring the progress of ongoing treatments, which is especially valuable considering the high prevalence and low diagnosis rate of

depression.<sup>2,42</sup> Upon diagnosis of depression, various Internet-based interventions can be delivered by the mobile health application, where many of these interventions have been found to be promising. Such interventions include computerized cognitive behavioral therapy (CBT) that is delivered over the Internet (iCBT),<sup>43</sup> guided iCBT, where iCBT is provided with the support of a therapist<sup>44</sup> or a non-clinical support staff,<sup>45</sup> cognitive bias modification therapy (CBMT),<sup>46</sup> behavioral activation (BA) intervention,<sup>47,48</sup> acceptance and commitment therapy (ACT),<sup>48</sup> and mindfulness-based (MB) intervention.<sup>47</sup> The study<sup>49</sup> provides a systematic review and comparison of mobile-based interventions for depression, where many of these interventions have been found to be highly effective. Moreover, the study in<sup>11</sup> provides an overview of different implementations of mobile interfaces to support therapy for mental health disorders, including depression.

Despite the benefits, privacy concerns have been cited as a significant obstacle toward the adoption and use of mobile health applications.<sup>42</sup> For example, in a recent study, nearly 60% of respondents expressed concerns about the privacy of personal information that is collected by mobile health applications.<sup>42</sup> This study showed that depression could be detected with relatively good accuracy (ie, balanced accuracy = 0.768) even with aggregated metadata usage attributes, which are less privacy-sensitive compared to other attributes such as the exact users' location information<sup>3</sup> or the content of their social media posts and activities.<sup>20</sup> According to the literature, the degree of a mobile application's information transparency, as perceived by users, plays a central role in users' decision-making process when they are faced with a personalization–privacy paradox.<sup>30</sup> Therefore, to increase adoption, it is recommended that mobile health applications clearly state what information is collected from users and how it is processed and used. This is especially important considering that the descriptions for many of the mobile application permissions are generic and unclear.<sup>50</sup> In addition to the ethical and legal aspects, past research suggests that users are more willing to disclose their mobile information when the purpose of mobile data collection is specific rather than when it is vague and ambiguous.<sup>51</sup> Moreover, the results show that the prediction accuracy can be further increased to 0.811 if users provide their age and gender, which again are less sensitive demographic information compared to other attributes such as income or education.

While this study shows some promising results in terms of the feasibility and accuracy of depression screening based on mobile usage attributes, the study framework can be extended by considering additional attributes and information sources. For example, as shown in,<sup>3</sup> users' location information is likely to improve depression screening models. While the combination of time and location information is considered to be highly privacy-sensitive<sup>52</sup> and may meet resistance from mobile users, aggregated mobility and location information attributes, such as the average distance traveled daily or the time spent indoor versus outdoor, are still likely to be predictive of depression. Considering location attributes would also make it possible to include other publicly available information, such as zip-code level demographic information from census data. Moreover, measurements of Bluetooth and Wi-Fi signals are likely to be useful in characterizing users' environment (eg, indoor versus outdoor<sup>11</sup>) and the extent of their interactions and exposures to other people (eg, crowd density),<sup>53</sup> which can be also predictive of depression. Furthermore, the overall accuracy of depression classification is likely to improve when combining predictions over successive intervals (eg, multiple 2-week windows), using methods such as collaborative filtering or weighted probability averaging.<sup>54</sup> Finally, it

should be noted that in this research, the study participants were all residents of the US. However, past research suggests that cross-cultural factors can influence both depression prevalence<sup>55</sup> and the patterns of mobile device usage,<sup>56</sup> and therefore, a follow-up study in other geographies is recommended.

## CONCLUSION

By deploying a range of machine learning classification algorithms and analyzing the BDI-II survey data from more than 400 participants, this study showed the possibility of inferring depression from a range of aggregated mobile usage and device characteristics attributes. The study highlighted the differences between the mobile usage patterns of respondents with minimal depression and those with mild to severe depression. The RF classification model showed a balanced accuracy of 0.768 and the AUC of 0.733 for classifying the 2 groups. Including respondents' age and gender increased the balanced accuracy to 0.811. The average number of outgoing calls was the most predictive variable, followed by the average total daily usage (RVI = 94.17), the number of saved contacts on the device (RVI = 84.68), and then the average time spent daily on social media (RVI = 80.35). The implications of these results are significant for mental health mobile applications, since the possibility of screening for depressive symptoms passively and continuously over time is essential both for diagnosis as well as monitoring the progress of ongoing treatments.

## FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

## AUTHOR CONTRIBUTIONS

RR, AG, and MG conceptualized the scope of the manuscript, defined the objectives, and wrote the manuscript. MG provided a literature review. RR and AG conducted all major analyses and developed the Methods and Results sections. MG provided feedback and revisions on both quantitative analyses and drafts of the manuscript.

## CONFLICT OF INTEREST STATEMENT

None declared.

## REFERENCES

- Huang S, LePendu P, Iyer S, *et al.* Toward personalizing treatment for depression: predicting diagnosis and severity. *J Am Med Inform Assoc* 2014; 21 (6): 1069–75.
- Mitchell A, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009; 374 (9690): 609–19.
- Saeb S, Zhang M, Karr CJ, *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.
- Cain RA. Navigating the sequenced treatment alternatives to relieve depression (STAR\*D) study: practical outcomes and implications for depression treatment in primary care. *Prim Care* 2007; 34 (3): 505–19.
- Weel CV, Weel-Baumgarten EV, Rijswijk EV. Treatment of depression in primary care. *BMJ* 2009; 338: b934.
- Wells J, Horwood L. How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports. *Psychol Med* 2004; 34 (6): 1001–11.

7. Mineka S, Sutton S. Cognitive biases and the emotional disorders. *Psychol Sci* 1992; 3 (1): 65–9.
8. Kamphuis MH, Stegenga BT, Zuitho NP, *et al.* Does recognition of depression in primary care affect outcome? *Fam Pract* 2011; 21: 16–23.
9. Taylor K, Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Technical report, Pew Research, 2019. <https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/> Accessed January 8, 2020.
10. Winnick M. Putting a finger on our phone obsession. Technical report, Dscout Research, 2016. <https://blog.dscout.com/mobile-touches> Accessed January 8, 2020
11. Gravenhorst F, Muaremi A, Bardram J, *et al.* Mobile phones as medical devices in mental disorder treatment: an overview. *Pers Ubiquit Comput* 2015; 19 (2): 335–53.
12. BinDhim NF, Shaman AM, Trevena L, *et al.* Depression screening via a smartphone app: cross-country user characteristics and feasibility. *J Am Med Inform Assoc* 2015; 22 (1): 29–34.
13. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. *JMIR Mhealth Uhealth* 2014; 21: e2.
14. Burns MN, Begale M, Duffeey J, *et al.* Harnessing context sensing to develop a mobile intervention for depression. *J Med Internet Res* 2011; 13 (3): e55.
15. Snjic A, Velicovic V, Sokolovic M, *et al.* Relationship between the manner of mobile phone use and depression, anxiety, and stress in university students. *Int J Environ Res Public Health* 2018; 15(4): e697.
16. Harwood J, Dooley JJ, Scott AJ, *et al.* Constantly connected – the effects of smart-devices on mental health. *Comput Human Behav* 2014; 34: 267–72.
17. Kumar S, Abowd GD, Abraham WT, *et al.* Center of excellence for mobile sensor data-to-knowledge (MD2K). *J Am Med Inform Assoc* 2015; 22 (6): 1137–42.
18. Thomee S, Harenstam A. Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults: a prospective cohort study. *BMC Public Health* 2011; 11: 66.
19. Thomee S, Dellve L, Harenstam A, *et al.* Perceived connections between information and communication technology use and mental symptoms among young adults: a qualitative study. *BMC Public Health* 2010; 10: 66.
20. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. In: *7th International AAAI Conference on Weblogs and Social Media*; Cambridge, MA; July 8–11, 2013.
21. Dehling T, Gao F, Schneider S, *et al.* Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. *JMIR Mhealth Uhealth* 2015; 3 (1): e8.
22. Kwang T, Buhrmester M. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality data? *Perspect Psychol Sci* 2011; 6(1): 3–5.
23. Beck AT, Steer RA, Brown GK, *et al.* *Manual for the Beck Depression Inventory-II*. Vol. 1. San Antonio, TX: Psychological Corporation; 1996: 82.
24. Osman A, Kopper BA, Barrios F, *et al.* Reliability and validity of the Beck depression inventory–II with adolescent psychiatric inpatients. *Psychol Assess* 2004; 16 (2): 120–32.
25. Segal DL, Coolidge FL, Cahill BS, *et al.* Psychometric properties of the Beck Depression Inventory–II (BDI-II) among community-dwelling older adults. *Behav Modif* 2008; 32 (1): 3–20.
26. Kojima M, Furukawa TA, Takahashi H, *et al.* Cross-cultural validation of the Beck Depression Inventory-II in Japan. *Psychiatry Res* 2002; 110(3): 291–9.
27. Ghassemzadeh H, Mojtabei R, Karamghadiri N, *et al.* Psychometric properties of a Persian-language version of the Beck Depression Inventory–Second edition: BDI-II-Persian. *Depress Anxiety* 2005; 21 (4): 185–92.
28. Sacco R, Santangelo G, Stamenova S, *et al.* Psychometric properties and validity of Beck Depression Inventory II in multiple sclerosis. *Eur J Neurol* 2016; 23 (4): 744–50.
29. Mystakidou K, Tsilika E, Parpa E, *et al.* Beck Depression Inventory: exploring its psychometric properties in a palliative care population of advanced cancer patients. *Eur J Cancer Care* 2007; 16 (3): 244–50.
30. Hautamaki V, Karkkainen I, Franti P. Outlier detection using k-nearest neighbor graph. In: *proceedings of the IEEE 17th International Conference on Pattern Recognition (ICPR)*; Cambridge, UK., 26–26 August 2004.
31. Zijlstra WP, van der Ark LA, Sijsma K. Outliers in questionnaire data: can they be detected and should they be removed? *J Educ Behav Stat* 2011; 36 (2): 186–212.
32. Kuhn M, Johnson K. *Applied Predictive Modeling*. New York: Springer; 2013: 26.
33. Chae J. Re-examining the relationship between social media and happiness: the effects of various social media platforms on re-conceptualized happiness. *Telematics Inform* 2018; 35 (6): 1656–64.
34. Kawachi I, Berkman LF. Social ties and mental health. *J Urban Health* 2001; 78 (3): 458–67.
35. Kim JH, Seo M, David P. Alleviating depression only to become problematic mobile phone users: can face-to-face communication be the antidote? *Comput Hum Behav* 2015; 51: 440–7.
36. Domenech-Abella Mundo J, Leonardi M. The association between socioeconomic status and depression among older adults in Finland, Poland and Spain: a comparative cross-sectional study of distinct measures and pathways. *J Affect Disord* 2018; 241: 311–8.
37. Freeman A, Tyrovolas S, Koyanagi A, *et al.* The role of socio-economic status in depression: results from the courage. *BMC Public Health* 2016; 16 (1): 1098.
38. Albert PA. Why is depression more prevalent in women? *J Psychiatry Neurosci* 2015; 40 (4): 219–21.
39. James G, Witten D, Hastie T, *et al.* *An Introduction to Statistical Learning*. New York: Springer; 2013.
40. Atkinson M. Apps Permissions in the Google Play Store. Pew Research Technical Report; 2015. <https://www.pewresearch.org/internet/2015/11/10/apps-permissions-in-the-google-play-store/>. Accessed January 8, 2020.
41. Malmi E, Weber I. You are what apps you use: Demographic prediction based on user's apps. In: *10th International AAAI Conference on Web and Social Media*. Cologne, Germany, 17–20 May 2016.
42. Lipschitz J, Miller CJ, Hogan TP, *et al.* Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement. *JMIR Ment Health* 2019; 6 (1): e11334.
43. Williams AD, Andrews G. The effectiveness of internet cognitive behavioural therapy (iCBT) for depression in primary care: a quality assurance study. *PloS One* 2013; 8 (2): e57447.
44. Berger T, Hämmerli K, Gubser N, *et al.* Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. *Cogn Behav Ther* 2011; 40 (4): 251–66.
45. Titov N, Andrews G, Davies M, *et al.* Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. *PloS One* 2010; 5 (6): e10939.
46. Blackwell SE, Browning M, Mathews A, *et al.* Positive imagery-based cognitive bias modification as a web-based treatment tool for depressed adults: a randomized controlled trial. *Clin Psychol Sci* 2015; 3 (1): 91–111.
47. Ly KH, Trüschel A, Jarl L, *et al.* Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. *BMJ Open* 2014; 4 (1): e003440.
48. Carlbring P, Hägglund M, Luthström A, *et al.* Internet-based behavioral activation and acceptance-based treatment for depression: a randomized controlled trial. *J Affect Disord* 2013; 148 (2–3): 331–7.
49. Josephine K, Josefine L, Philipp D, *et al.* Internet-and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. *J Affect Disord* 2017; 223: 28–40.
50. Kelley PG, Cranor LF, Sadeh N. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; Paris, France, 1–4 April 2013.



51. Awad NF, Krishnan MS. The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS Quarterly* 2006; 30(1): 13–28.
52. Bettini C, Wang XS, Jajodia S. Protecting privacy against location-based personal identification. In: proceedings of the 2<sup>nd</sup> international conference on Secure Data Management; Trondheim, Norway, 2-3 September 2005.
53. Hong H, Luo C, Chan M. Socialprobe: Understanding social interaction through passive Wi-Fi monitoring. In: proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services; Hiroshima, Japan Nov 28 -Dec 1 2016.
54. Armstrong JS. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. New York: Springer Science & Business Media; 2001.
55. Scorza P, Masyn K, Salomon P, *et al*. The impact of measurement differences on cross-country depression prevalence estimates: a latent transition analysis. *Plos One* 2018; 23: 1–14.
56. Qin L, Kim Y, Tan X. Understanding the intention of using mobile social networking apps across cultures. *Int J Human Comput Interact* 2018; 34 (12): 1183–93.

# Depression screening using mobile phone usage metadata: a machine learning approach

Razavi, Rouzbeh; Gharipour, Amin; Gharipour, Mojgan

01	Titus Pellegrom	Page 1
4/4/2021 15:18		
02	Titus Pellegrom	Page 3
4/4/2021 15:21		
03	Titus Pellegrom	Page 3
4/4/2021 15:21		
04	Titus Pellegrom	Page 3
4/4/2021 15:21		
05	Titus Pellegrom	Page 4
4/4/2021 15:20		
06	Titus Pellegrom	Page 5
4/4/2021 15:23		