Investigating Rhythmicity in App Usage to Predict Depressive Symptoms: Protocol For a Personalized Framework Development and Validation Through a Countrywide Study

Md. Sabbir Ahmed, Tanvir Hasan, Salekul Islam, Nova Ahmed

Submitted to: JMIR Research Protocols
on: August 06, 2023

Table of Contents

Investigating Rhythmicity in App Usage to Predict Depressive Symptoms: Protocol For a Personalized Framework Development and Validation Through a Countrywide Study

Md. Sabbir Ahmed[1] BSc; Tanvir Hasan[1] BSc; Salekul Islam[2] PhD; Nova Ahmed[1] PhD

[1]Design Inclusion and Access Lab North South University Dhaka BD
[2]Dept. of Computer Science and Engineering United International University Dhaka BD

Corresponding Author:
Md. Sabbir Ahmed BSc
Design Inclusion and Access Lab
North South University
Bashundhara
Dhaka
BD

Abstract

Background: Knowing a student's depressive symptoms can facilitate significantly more precise diagnosis and treatment. But very few studies focused on depressive symptom predictions through unobtrusive systems and these studies are limited by smaller sample sizes, lower performance, and the requirement for higher resources. Besides, it is unexplored whether there exist statistically significant rhythms based on different app usage behavioral markers (e.g., app usage sessions) which could be potential in finding subtle differences to predict with higher accuracy like the models based on rhythms of physiological data.

Objective: The main objective of this study is to explore whether there exist statistically significant rhythms in resource-insensitive app usage behavioral markers and predict depressive symptoms through these markers-based rhythmic features. Another objective of this study is to understand whether there is a potential link between rhythmic features and depressive symptoms.

Methods: Through a countrywide study, we collected 2952 students' raw app usage data and responses to the 9 depressive symptoms of the Patient Health Questionnaire (PHQ-9). The students' app usage behavioral data were retrieved through our developed tool which was previously used in our pilot studies in Bangladesh on different research problems. To explore whether there is a rhythm based on app usage data, we will do a zero-amplitude test. Additionally, we will develop a cosinor model for each participant to extract the rhythmic parameters (e.g., acrophase). Besides, to get a comprehensive picture of the rhythms, we will explore the non-parametric rhythmic features (e.g., interdaily stability) as well. Apart from these, we will do a regression analysis to understand the association of rhythmic features with depressive symptoms. Finally, we will develop a personalized multi-task learning (MTL) framework to predict symptoms through rhythmic features.

Results: After excluding participants to satisfy different requirements (e.g., having app usage data of at least 2 days to explore rhythmicity), 2902 students' are kept for analysis where there are 24.48 million app usage events data and 7 days' app usage data of 98.17% (n=2849) students. The participants were from all divisions of Bangladesh, both public and private universities, 52 different departments, and 19 different universities of Bangladesh. We are analyzing the data and will be publishing the findings in a peer-reviewed venue.

Conclusions: Having an in-depth understanding of the app usage rhythms and their connection with depressive symptoms through a countrywide study can significantly help healthcare professionals and researchers to better understand depressed students and may create possibilities for using app usage-based rhythms for intervention. Additionally, the MTL framework based on app usage rhythmic features may more accurately predict the symptoms due to the rhythms' capability to find subtle differences.

(JMIR Preprints 06/08/2023:51540)
DOI: https://doi.org/10.2196/preprints.51540

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
Please make my preprint PDF available to anyone at any time (recommended).
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
✓ **No, I do not wish to publish my submitted manuscript as a preprint.**
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Investigating Rhythmicity in App Usage to Predict Depressive Symptoms: Protocol For a Personalized Framework Development and Validation Through a Countrywide Study

Md Sabbir Ahmed[1], Tanvir Hasan[1], Salekul Islam[2], and Nova Ahmed[1]

[1]Design Inclusion and Access Lab, North South University, Bangladesh

[2]Dept. of CSE, United International University, Bangladesh

msg2sabbir@gmail.com, tanvirfuad00@gmail.com, salekul@cse.uiu.ac.bd, nova.ahmed@northsouth.edu

Abstract

Background: Knowing a student's depressive symptoms could facilitate significantly more precise diagnosis and treatment. However, very few studies focused on depressive symptom predictions through unobtrusive systems, and these studies are limited by smaller sample sizes, lower performance, and the requirement for higher resources. Besides, it is unexplored whether there exist statistically significant rhythms based on different app usage behavioral markers (e.g., app usage sessions) which could be potential in finding subtle differences to predict with higher accuracy like the models based on rhythms of physiological data.
Objective: The main objective of this study is to explore whether there exist statistically significant rhythms in resource-insensitive app usage behavioral markers and predict depressive symptoms through these markers-based rhythmic features. Another objective of this study is to understand whether there is a potential link between rhythmic features and depressive symptoms.
Methods: Through a countrywide study, we collected 2952 students' raw app usage data and responses to the 9 depressive symptoms of the Patient Health Questionnaire (PHQ-9). The students' app usage behavioral data were retrieved through our developed tool which was previously used in our pilot studies in Bangladesh on different research problems. To explore whether there is a rhythm based on app usage data, we will do a zero-amplitude test. Additionally, we will develop a cosinor model for each participant to extract the rhythmic parameters (e.g., acrophase). Besides, to get a comprehensive picture of the rhythms, we will explore the non-parametric rhythmic features (e.g., interdaily stability) as well. Apart from these, we will do a regression analysis to understand the association of rhythmic features with depressive symptoms. Finally, we will develop a personalized multi-task learning (MTL) framework to predict symptoms through rhythmic features.
Results: After excluding participants to satisfy different requirements (e.g., having app usage data of at least 2 days to explore rhythmicity), 2902 students' data are kept for analysis where there are 24.48 million app usage events data and 7 days' app usage data of 98.17% (n=2849) students. The participants were from all divisions of Bangladesh, both public and private universities, 52 different departments, and 19 different universities. We are analyzing the data and will be publishing the findings in a peer-reviewed venue.
Conclusions: Having an in-depth understanding of the app usage rhythms and their connection with depressive symptoms through a countrywide study can significantly help healthcare professionals and researchers to better understand depressed students and may create possibilities for using app usage-based rhythms for intervention. Additionally, the MTL framework based on app usage rhythmic features may more accurately predict the symptoms due to the rhythms' capability to find subtle differences.

Keywords: Depressive symptoms, app-usage rhythm, behavioral markers, personalization, multi-task learning framework.

Introduction

Background

The need for the identification of depressive symptoms

On every 40 seconds, a person commits suicide and there are more than 20 attempts globally [1]. Among suicide attempters, major depressive disorder (MDD) is common [3] and people having MDD are at greater risk of suicidality [2]. Despite these facts, there is a significant lack of taking steps to mitigate depression due to which the depression rate is increasing. In fact, it is estimated that depression will rank first as a global burden of disease by 2030 [4]. Additionally, due to the COVID-19 pandemic, a significantly higher number of people faced MDD [5], and such a negative impact may persist for a prolonged period. In Bangladesh, the depression prevalence is higher than the overall rate in South Asia [10]. Compared to Bangladeshi people of any other professions, the depression rate (e.g., 82.4% of students have at least mild depression [42]) is higher among university students [6] which is alarming.

To significantly facilitate early intervention to mitigate depression, there is an urgent need for its early identification [9]. A person is identified as depressed if symptoms (e.g., hopelessness) appear for a time period such as most days for a minimum of 14 days [39, 40]. However, it is difficult to precisely assess depressive symptoms [96]. In fact, primary care providers fail to identify depression in more than 50% of cases [13, 14]. Knowing the depressive symptoms (e.g., 9 symptoms of the PHQ-9 scale [8]) of a person in real-time can facilitate mental healthcare professionals significantly to understand the illness more precisely, to early identify depression, and take steps accordingly for intervention.

Pervasive health research in low and middle-income countries

Despite the fact that over 80% of depression's burden is found in low and middle-income countries (LMICs) [7], there remains a severe scarcity of mental healthcare professionals in LMICs. In Bangladesh, there are only 565 psychologists [10] though it has a population of over 165 million [12] and 1.234 million university students [11]. In these cases, a pervasive technology such as a smartphone-based monitoring system can play a significant role which is available to a large number of people of LMICs [43] Additionally, to minimize the barriers to healthcare facilities in low-resource settings, AI-based mobile applications can be potential [41]. However, almost all of the studies that demonstrated the pervasive technology-based automated system to identify depression have been conducted in the context of high-income countries. For instance, all of the studies which have been included in a recent systematic review are from countries other than the LMICs [20] which indicates how less pervasive health research has been conducted in the context of LMICs. As a result, the developed models based on participants from high-income countries may not be applicable in LMICs since the behavior (e.g., app usage behavior [21]) varies among countries, and various factors such as socioeconomic status [22] and culture [21] impact behavior.

App usage rhythm that may resemble biobehavioral rhythm

Zeitgebers, the social and environmental cues, help a person's rhythms to synchronize well [80] which can impact their daily activities. Rhythms based on pervasive device-sensed physiological data change depending on external cues such as light exposure, eating time, and physical activity [80]. Similarly, smartphone usage behavior is linked with factors such as eating behavior [81]. Also, there is a relation of app usage with alertness, chronotype [76], and physiological data such as sleep [75, 76, 84]. Like the physiological data, app usage behavioral markers vary depending on the hour of the day [59, 82, 83] and exercise [84]. These facts show app usage behavior may also have a rhythmic pattern with reproducible waveforms similar to the rhythms based on physiological data.

Though a recent study extracted parametric rhythmic feature dominant periods from smartphone usage data [24], the study is limited by not exploring rhythmic features such as the acrophase, interdaily stability, intradaily variability, and relative amplitude. Besides, that study explored mere screen usage, without any exploration of more informative features [25] such as entropy data-based features. In the case of the other previous app usage data-based studies, researchers used descriptive statistical methods [59, 82, 83] to find whether there is any variation over the day and inferential statistical methods [16, 34] to find the difference in terms of aggregated data of the 4 time periods, namely, morning, afternoon, evening, and night. These approaches have some limitations. First, it lacks statistical evidence to show whether there exists any rhythm that could be resolved by the zero-amplitude test [67, 97] which is used to detect rhythm in the field of chronobiology. Additionally, the mere analysis in presenting the difference between the aggregated data of morning and evening can not present whether there is a cycle that can repeat over the days. The average data may also miss the microscopic view [98] of data. On the other hand, an analysis fitting the mathematical models (e.g., cosinor model) to time series data can present the microscopic view of the data and inferential statistical estimates of the rhythmic properties [98]. Besides, many informative behavioral markers such as the dominant period, stability in the behavior, and the peak time of the oscillation in the rhythm can not be extracted from just finding the differences between periods (e.g., morning vs. night period).

The potential of app usage-based rhythms in identifying depressive symptoms in LMICs

In human life, physiological changes reappear in a cyclable waveform [98]. The rhythm features based on the physiological data have been explored both in the chronobiology [85] and pervasive health [24] areas. Researchers found a relation between the physiological data-based rhythmic markers and health status [24]. These markers can find out the subtle differences that enable these markers-based features to predict hospital readmission [86], to identify loneliness and depression class [24]. This shows the possibility of improving the performance of the models upon the incorporation of app usage rhythmic features to predict the symptoms of depression. However, previous studies [24, 86] mostly relied on wearables to extract the rhythmic features which are costly and may not be affordable [87] by the low-income people. On the other hand, smartphones are economically attainable [88] and app usage data are resource-insensitive [25]. As a result, app usage data-based systems may be feasible in LMICs.

Predicting depression and depressive symptoms through an unobtrusive method

Classification of the depressed and non-depressed: Most existing research based on AI for mental health has worked on classification problems [52]. Researchers classified the depressed and non-depressed participants by developing personalized models [18] and using the contextually filtered features where the rule mining technique was incorporated [19]. Some other studies (e.g., [23, 27] relied on sensing data (e.g., GPS data) along with smartphone data call history to predict depression. Researchers also leveraged internet usage data [32], location data retrieved through the campus WiFi infrastructure [29, 30], and GPS [28, 30, 31]. Recently, some researchers focused on exploring rhythmic features to assess depression. For instance, [24] leveraged the rhythm-based features to classify the depressed and non-depressed participants. However, the classification does not provide precise information about a participant’s depressive status since scores of all the symptoms are aggregated together to keep a participant into a particular group (e.g., depressed or moderate depression) which loses the complexity of the psychological problem depression.

Predicting depression scores: Compared to the classification research problems, there is relatively less research on predicting the depression score (e.g., a person's PHQ-9 score 11). These studies in the pervasive health area can be broadly categorized into the studies which developed models leveraging data based on both smartphone and wearable [17, 51, 54, 55], only smartphones [17, 54], various sensing devices along with Social Media [101], subjective and smartphone data [57]. There remain mixed findings on whether the smartphone has superior performance than the wearable. Smartphone-sensed data showed higher performance in a previous study [54] when models were evaluated after splitting the training and testing data based on time. However, in the same study [54], researchers found inferior performance of smartphones in another evaluation criterion. Regardless of their superiority, both wearables and smartphones show promising performance in automated assessment of depression [17] which can play a role in real-time remote monitoring of the sufferers [58]. The wearable and smartphone-sensed behavioral markers such as call duration [57] and heart rate [56] as well as inferred markers such as circadian rhythm [56] have a significant correlation with the depression score which may explain the fact of enabling the sensed data to predict depression. However, like the classification, in the prediction of depression scores also, there remains the inability to understand the depressive symptoms precisely since a person's depression score (e.g., 11) can be found in combination with different frequencies of different symptoms' appearance.

Predicting symptoms and multi-task learning: With great interest, researchers did a network analysis of the depressive symptoms and presented the possible viable target for intervention [46] and central symptoms for possible focused treatments [47]. The relation of ecological momentary assessment of depressive symptoms with PHQ-9 score [53] and also the relation of depressive symptoms with behavioral data [45] was also investigated in previous studies. Exploring pervasive device-sensed data, researchers [45] found a link of higher spending duration in dorms with higher fatigue which is a depressive symptom. Though there are studies predicting symptoms of other psychological problems such as Schizophrenia [48] and ADHD [49], there are very few studies [33] that predicted the appearance of depression symptoms. In a previous study [33], authors predicted depressive symptoms through the data of iPhone users and Android users. However, their study has several limitations. First, their models' performance is low in the case of the maximum depressive symptoms, particularly, the specificity score is around 60% or below 60% in many models developed by leveraging smartphone data. Second, they developed a separate model to predict each of the symptoms which makes the model resource sensitive. Third, they considered each symptom-predicting task as a separate one which could be improved by leveraging the shareable information among symptoms through the multi-task learning (MTL) framework as MTL has superior performance than the single-task learning technique [17, 48]. However, researchers grouped all the predicting tasks into a single model in previous studies [17, 48] which may lower performance since all tasks do not help each other to improve the performance [50].

The Objective of This Study

To overcome the limitations mentioned above, the main objective of this study is to explore app usage data-based rhythmicity detection and develop and validate an MTL framework leveraging app usage data-based parametric and non-parametric rhythmic parameters through a countrywide study in Bangladesh. Another objective of this study is to explore whether the app usage rhythmic features relates to depressive symptoms.

Methods

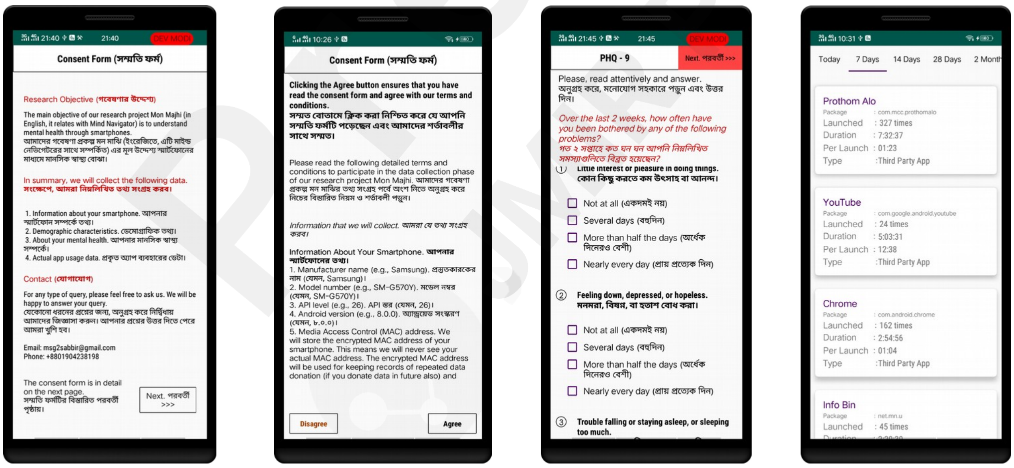
Ethical Considerations

Our study was approved (reference: 2022/OR-NSU/IRB/0704) by the North South University Institutional Review Board/Ethics Review Committee. While collecting the data, we presented the consent form in Bangla (native language) and English languages in short (Figure 1(a)) where we provided a summary of the data that we were going to collect. In addition, we provided a consent form in detail (Figure 1(b)) about different matters including data safety, privacy, and each data that we are going to collect. Apart from these, before data collection, we briefly described the research to the participants. All participants provided their consent. After collecting data, we provided an incentive of around USD 0.3 equivalent food tokens to each participant.

All participants' data are stored in the Google Firebase database which only one of the researchers can access with 2-factor authentication. From the participants, we did not collect any personal information such as name, email, and phone number. Additionally, from the participants, we did not collect the names of the departments and universities so that participants could feel more comfortable in donating the data. It can be mentioned in this paper, we reported the number of departments and universities based on our aggregated notes.

Retrieval of App Usage Behavioral Markers

We focus on developing the system in a way so that the users of our system and also the healthcare professionals can be informed about the depressive symptoms in real-time. For instance, a healthcare professional, having the consent of a user to access information, can be notified if our systems' automatic prediction in a day shows that the depressive symptoms can get worse. For this research, we used our previously developed tool [15] which can retrieve the participants' past 7 days' foreground and background app usage events' data within a second, once the participant provides the consent. The average required time to retrieve the events is 307.94 milliseconds (SD=1.1 seconds) [25]. For each app usage event, there is data on the app name, package name, and timestamp of the event which we will use to extract the behavioral markers. The tool (Figure 1) was used in our previous studies to explore different research problems including research on students' academic results [89, 92, 93], depression [16, 25, 34, 91], and loneliness [26, 90] showing the reliability and validity of the tool.



(a) Consent form in short (b)Consent form in detail (c) PHQ-9 scale to collect depressive symptoms' data (d) App usage data

Figure 1: Screenshot of the data collection tool showing the (a) consent form in short in Bangla and English languages, (b) consent form in detail, (c) the PHQ-9 scale, (d) app usage data.

Sample Size Determination and Data Collection

In Bangladesh, there are around 1.234 million university students [11] and to have a sample that represents the behavior of this population with a confidence level of 95% and a margin of error of 5%, there is a need for 385 university students as we found by using the SurveyMonkey [61] where the formula for finite sample size [105] is used. Using the same formula, we found the required sample size is 384 for each to represent the 0.448 million female and 0.786 million male students

[11]. Since the depression rate differs between the students of public and private universities in Bangladesh [44], we collected data from both types of universities. In the cases of 0.329 million and 0.902 million students of the public and private universities respectively [11], we found the requirement of 384 students as a sample in each case. However, there is no fixed sample size that can ensure the generalizable performance of the machine learning (ML) models. Therefore, to develop an impactful model, we tried to maximize the number of participants by conducting a countrywide study. Additionally, we tried to maximize the number of students from public universities as the largest number of students study there [11].

We collected data using the multi-stage convenient sampling method in all 8 divisions of Bangladesh. From each division, we collected data from at least 1 university and multiple departments. By collecting data from different divisions, universities, and departments, we tried to maximize the diversity among the participants because the socioeconomic status and many other demographic characteristics vary by region of a country which can have an impact on mental health [44] as well as smartphone usage behavior [102, 103].

We collected data from 2952 participants from September 2022 to March 2023. While collecting data, first, through our app [15], the participants responded to the questions about demographic characteristics (e.g., gender) after giving consent. After that, they responded to the items of the PHQ-9 scale [8] using the same app where they responded based on their experiences of the past 14 days. After saving responses to all questions of the psychological scales, the app automatically retrieved the app usage data which may take less than 1 second in almost all cases as shown in our previous estimation [25]. After retrieval, the app saved the app usage data instantly.

Data Pre-processing and Dataset Description

Our tool could not retrieve any app usage data from 14 participants' phones. One of the plausible reasons for having such a problem could be due to having system problems in the phones as shared by the participants. Several of these participants shared that they do not use the original version of the phones. Besides, in the case of the 2 participants, there were missing age values and we did not impute these values as age information was not required for the primary purpose of the research. Apart from these, there was missing data in the case of the 82 participants' professions. We imputed those missing professions based on 2 pieces of information. First, in our study, there were only 2 participants who were not students, and to whom, we did not reach out for data collection. These 2 persons donated data through installing the app from the Play Store which can be based on their own interests. This presents that there is a very low probability of having a profession other than the student. Second, all of these 82 participants donated data when we reached the universities for data collection as we found by matching the study dates and timestamps of these participants' data donations. Thus, we imputed these 82 participants' professions as students.

While the participants were required to donate data at least once, we encouraged the participants to donate as many times as possible. After excluding the 14 participants for whom our tool could not collect any app usage events, there were 2936 participants remaining. Among them, 71 participants donated at least twice. However, since there were very few participants who donated data twice, in this study, we will keep only the first-time donated data for the next steps of the analysis.

Since in this study, we will estimate the rhythm, we are following previous studies to find out the minimum number of days required to estimate rhythm. It is suggested to have data of at least 1 day for estimating circadian rhythm [62]. However, sensed data from 2 days can estimate the rhythm sufficiently [63]. Researchers [48] also found rhythmic features extracted based on the data from 2 days can predict more accurately the physiological and mental changes. Therefore, inspired by these previous studies, we excluded the participants (n=34) who had app usage data of less than 2 days.

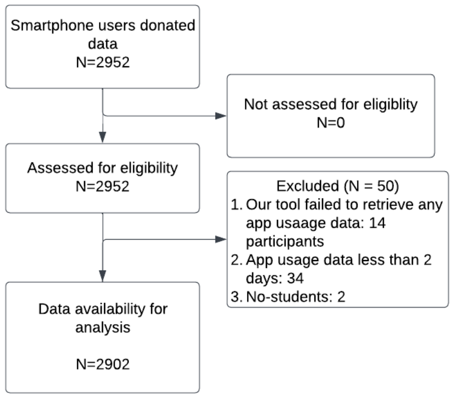


Figure 2: STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) flowchart.

Finally, after excluding the 2 non-students, 14 students without any app usage data, and 34 students having app usage data of less than 2 days, there are 2902 students' app usage data (Figure 2) which will be analyzed in the next steps. In total, there are 24.48 million app usage foreground and background events. These data are of 24.84 million minutes. Among the 2902 students, 98.17% (n=2849) participants have app usage data of 7 days (Figure 3).

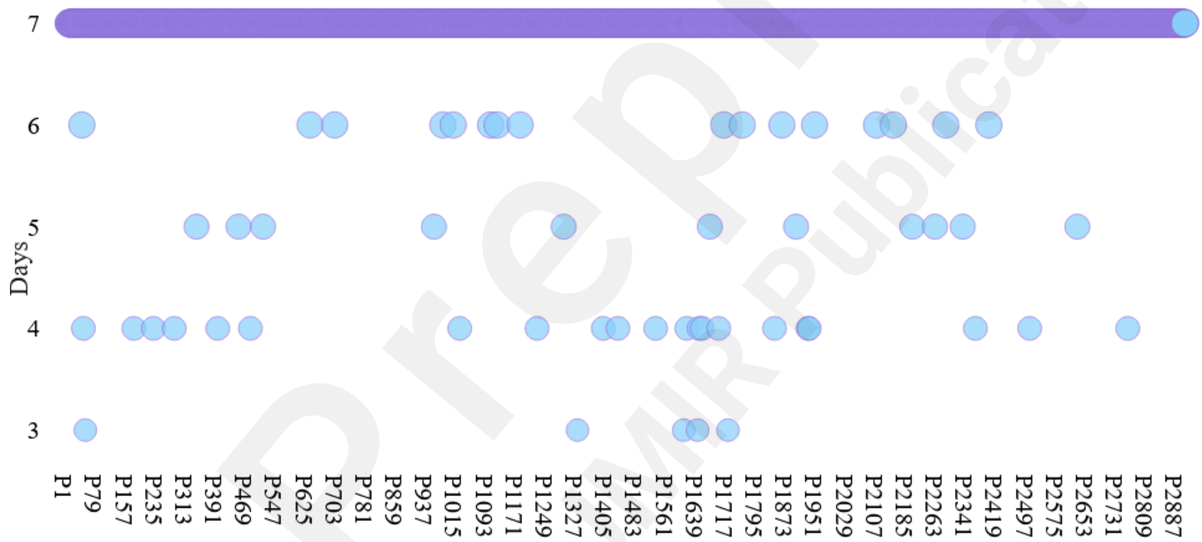


Figure 3: Number of days' data that each participant has for the final analysis.

Ground Truth Data to Measure Depressive Symptoms

We used the Patient Health Questionnaire-9 (PHQ-9) scale [8] which is one of the most commonly used scales to assess depression [20]. In our study, we used the translated PHQ-9 scale in Bengali [64] which has been validated and has also been largely used in Bangladesh (e.g., [42, 44]. The PHQ-9 score of 10 has a sensitivity and specificity of 88% in identifying major depressive disorder [8]. Based on this cut-off score, we will categorize the participants as depressed, if the PHQ-9 score is at least 10.

Table 1: Depressive symptom's name for each item of the PHQ-9 scale.

Symptoms	Symptom in the PHQ-9 Scale [8]
Symptom 1	Little interest or pleasure in doing things.
Symptom 2	Feeling down, depressed, or hopeless.
Symptom 3	Trouble falling or staying asleep, or sleeping too much.

Symptoms	Symptom in the PHQ-9 Scale [8]
Symptom 4	Feeling tired or having little energy.
Symptom 5	Poor appetite or overeating.
Symptom 6	Feeling bad about yourself - or that you are a failure or have let yourself or your family down.
Symptom 7	Trouble concentrating on things, such as reading the newspaper or watching television.
Symptom 8	Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual.
Symptom 9	Thoughts that you would be better off dead or of hurting yourself in some way.

In a previous study [33], a person was categorized as having a depressive symptom, if the respective symptom appeared several days or more. However, we will consider the threshold more than half of the days since NIMH and WHO define a person as depressed if symptoms appear for a time frame such as most days for a minimum of 14 days [39, 40]. That being said, if a participant reported that he/she/they were bothered by a depressive symptom (e.g., Hopelessness) available in the PHQ-9 scale for more than half of the days or nearly every day of the past 14 days, we will categorize that participant as having the depressive symptom.

Extraction of Behavioral Markers

As we will test the rhythmicity where the time series data are used [67] and we will explore the rhythmicity of the app usage behavioral markers instead of exploring the raw foreground and background app usage events, we will set a time frame of 15 minutes based on which each app usage behavioral marker will be extracted. But to check the robustness of the findings, we will follow the same process to explore rhythmicity using two other time frames as well: 10 minutes and 20 minutes.

Apart from calculating the usage duration, frequency of launching apps, entropy-based on duration and entropy-based on frequency of launching apps, and app usage sessions (please, see details about these behavioral markers in our recent publications [25]), we will calculate the following behavioral markers based on which we will extract the parametric and non-parametric rhythmic features.

Relative importance of the app categories: In pervasive health research, aggregated usage data (e.g., in [18, 19, 23]) are widely used where the individual app categories remain unexplored. In our previous studies [25, 26], we found that the app category-based features are more important than the aggregated smartphone usage regardless of category. However, in those studies, features of an app category were extracted independently regardless of the usage behavior of other categories. As a result, the individual category itself may not give higher information for the app categories which are less used but contain distinguishable markers. For instance, if a participant launches Social Media apps 100 times and Health & Fitness apps only 10 times, data from the Health & Fitness category may get lower importance. However, data from the Health & Fitness category can be more important since depressed students use apps that contain features for smoking prevention and reducing body weight which may not be used more frequently, but contain enough information to be significantly different from the non-depressed students [16]. Hence, we calculate Term Frequency - Inverse Document Frequency (TF-IDF) which is a widely used technique in natural language processing [94] where the less frequent terms across the documents can get more importance. To adapt TF-IDF in the context of app usage, we will use data from all time frames over all days. In each time frame f, the app usage sessions s will work as the set of documents, and in each session (i.e., a document) j, the list of categories of the used apps will act as the words.

Let the set of documents of a participant i in f be $\{D_{ij}, D_{ij+1}, ..., D_{is}\}$; where $i \epsilon n$, $j \epsilon s$; n presents the number of participants. We will calculate the $TF-IDF_{icj}$ for each app category c based on the participant's data of sessions s. $TF-IDF_{icj} = TF_{icj} * IDF_{icj}$; $TF_{icj} = \log ¿$); $IDF_{icj} = \log(\frac{1+s}{1+df})$; $freq(c)$ = number of times apps of c was launched in session j; df = number of sessions where c was used; s presents the total number of sessions in f. After calculating the $TF-IDF_{icj}$ for each session, the mean $TF-IDF_{\square}$ value will be calculated for each category over the sessions of a time frame. Finally, using that mean value, we will extract the rhythmic features to understand how the relative importance of the participant in using a particular category c varied or remained constant over the days, and over the periods of a day, whether there is a rhythm in behavioral marker based on TF-IDF.

It can be mentioned that to categorize the apps, we will follow relevant previous studies (e.g., [82]) and the process of our previous studies [16, 25, 26, 34, 89] where we categorized the apps more than 20 categories after exploring developers' referred categories in Google Play Store and other app stores, and discussing with graduate students of the Computer Science and Engineering (CSE) department.

Personalized top-n apps: We will calculate the entropy based on usage duration and frequency of launch of top-n apps which can vary by student. To find the value of n, we will use the probability distribution and will plot the probability of using apps in the y-axis and the number of apps in the x axis. The point where the curve will fall will be considered as the threshold (i.e., the value of n) to find the top-n apps. To find the cut-off value, we are inspired by a previous study [23] where cut-off values to exclude the participants having missing values were found in this approach.

Co-usage of apps: A significant portion of app usage consists of switching from one app to another app [59]. One plausible reason for having such behavior can be changing moods while using the apps [60] and the users may do it to seek support, overcome negative emotions, etc. To quantify how the participants switch from one app to another, we will calculate the co-usage of apps where two apps will be considered as co-app usage if they were used in a single app usage session and also were used consecutively. However, in a smartphone, there remain many system apps that get opened automatically to support the function of another app. The users do not need to use those apps intentionally and as a result, the inclusion of those apps may misrepresent the behavior. Besides, to switch from one app to another, a user returns to the home screen of a phone where the launcher app will be opened automatically. Considering the aforementioned issues, we will exclude those system apps and launcher apps before quantifying the co-app usage.

To find the subtle difference in variation of app usage patterns, we will build a network based on the co-usage where each edge presents a co-usage of 2 different apps. The weight of the edge will be calculated by point mutual information (PMI): $PMI(A_1, A_2) = \log \frac{p(A_1, A_2)}{p(A_1)\, p(A_2)}$ where $p(A_1, A_2)$ presents the probability to appear A_1 and A_2 apps consecutively in the same app usage session; $p(A¿¿1)¿$ and $p(A¿¿2)¿$ presents the probability to appear A_1 and A_2 apps respectively in that session regardless of the consecutiveness. After that, we will calculate the centrality and graph edit distance. The calculation of centrality will help to understand the most influential app in the network which is connected with the maximum number of nodes. The graph edit distance between two different sessions of the same time frame will inform whether the behavior differed. Finally, we will use the average data of each session of a time frame to explore the rhythmicity of the centrality and graph edit distance.

Rhythm Analysis and Extraction of Rhythmic Features

The parametric cosinor method is one of the most widely used approaches to finding out the rhythmic parameters. However, the cosinor analysis can not find out the fragmentation in the rest-activity rhythm [68] which can be detected by extracting the nonparametric rhythmic parameters such as interdaily stability (IS). Therefore, to get a comprehensive picture of the rhythms, we will do both parametric and non-parametric tests and will extract the respective rhythmic features as described below. In our study, instead of focusing on only all students' data-based global models, to extract the rhythms, we will also develop individual models and extract the rhythmic parameters for each participant. The main reason is the physiological data-based nonparametric rhythmic [70, 71] as well as parametric rhythmic [72] parameters vary by the characteristics of people and thus, there may also be a variation in parameters by the individual participant's rhythm which is solely based on app usage data.

Dominant period: In cosinor analysis, a cosine curve is fitted on the given period (e.g., 24 hours) by the least squares method where the model reduces the difference between the observed and estimated values. However, the cosinor analysis itself can not estimate the best-fitting period [69]. Thus, we will empirically investigate through the periodogram analysis and by setting the periods from 1 to 24 to find out. Later, the cosinor model will be developed for each given period (e.g., 13 hours). The best fitting period will be counted as the dominant period where the proportion of explained variance is maximum.

Rhythm detection: Cosinor analysis is a parametric method and hence, we will test the normality [65] first. Later, we will process the non-normally distributed data through log transformation. After that, to find out whether there exists a statistically significant rhythm of a participant, we will do a zero-amplitude test [67] by setting the significance level to 0.05. To find out whether the individual participant's rhythm differs from the rhythm based on all participants, we will do population mean cosinor analysis which will be based on all of the participants where a cosinor model will be developed for each participant and the average of the cosinor rhythmic parameters is calculated afterwards [69].

MESOR, amplitude, and acrophase: After developing the cosinor model, we will extract the parametric rhythmic parameters midline estimating statistic of rhythm (MESOR), amplitude, and acrophase for each participant. The acrophase presents the timing of the high values recurring in each cycle of the rhythm [67], the mesor is the rhythm-adjusted mean [67], and the amplitude presents the difference between the equilibrium position and the peak point of the rhythm oscillation [66]. Since amplitude can present the strength of the rhythm [74], while comparing the diurnal (12 hours period) and circadian (24 hours period) rhythms to find out the rhythm that has more strength, we will compare the amplitude. Moreover, we will compare the coefficient of determination as it presents how well the model is fitted in a given period.

Interdaily stability and intradaily variability: Mental state has a relation with interdaily stability (IS) and intradaily variability (IV). For example, patients with bipolar disorder have less IS and higher IV [73]. Similar patterns may be found in the case of IS and IV based on app usage data which motivated us to calculate following a previous study [68] on actimetry:

$IV = \frac{(N)\sum_{i=2}^{N}(x_i - x_{i-1})^2}{(N-1)\sum_{i=1}^{N}(x_i - x_m)^2}$; Here, x_i is the behavioral marker's value at a time frame and x_m is the mean value of the same behavioral marker in all time frames; N presents the number of time frames.

$IS = \frac{(N)\sum_{t=1}^{p}(x_t - x_m)^2}{(p)\sum_{i=1}^{N}(x_i - x_m)^2}$; Here, x_t presents the behavioral marker's average value in that time frame over the days; p presents the number of time frames per day. The range of the value of the IV is 0 to 2 whereas the range of the value of IS is 0 to 1 [77]. The higher the IS, the greater the stability as the name implies. On the other hand, the higher the IV, the higher the fragmentation in the rhythm. If there remains a difference, for example, if someone sleeps daytime and keeps waking at night time, the IV will be higher [68].

M10, L5, and relative amplitude: M10 presents the mean value of the most active consecutive 10 hours and provides information about the diurnal activity. The active persons have a higher M10 [68] and the lower M10 can be associated with exercise reduction [68] and also with a negative mental state [68, 78]. On the other hand, the L5 presents the mean value of least active 5 consecutive hours of the day. It is a measure of nocturnal activity and the higher L5 may present the activity during the rest cycle [68]. After calculating the M10 and L5, we calculate the relative amplitude (RA), $RA = \frac{M10 - L5}{M10 + L5}$. As can be understood from the formula of RA, the RA is actually the normalized difference between M10 and L5. The larger the difference between L5 and M10, the larger the RA. People having psychological problems can have a lower RA [48, 68].

Statistical Analysis

To explore whether there is any relation between depressive symptoms and app usage rhythmic features, we will do binomial logistic regression. As having a variance inflation factor (VIF) of more than 5 can create a biased regression model [79], we will eliminate the variable if the VIF of any variable becomes more than 5.

ML Models Development

Participant-similarity and development of personalized models

Most of the existing models (e.g., [19, 23, 25, 26]) to predict depression used training data to predict the outcome regardless of the characteristics of the participant whose class is going to be predicted. These models may have issues regarding generalizability since all individuals have unique characteristics that may not be captured in one-size-fits-all models (i.e., one set of training data to predict the depression of all test participants). Indeed, through empirical investigation, it has been found that the personalized model performs better than the one-size-fits-all model [35, 37]. The one-size-fits-all or global model is likely to capture the general or the "average" characteristics of the participants for the prediction due to which the model may perform well only on the "average" participants and may not work in the case of the participants whose behavior deviates from the "average" participants [36]. On the other hand, in the personalized models, models are trained dynamically for each participant which can facilitate finding the most relevant set of features to predict the outcome of the test participant [35]. Therefore, a personalized model may perform better in predicting depressive symptoms since each of the depressed students has a unique app usage signature [16] and the depressed students have statistically significantly different smartphone usage behavior than the non-depressed students [34] as we found in our previous studies [16, 34].

In our system to predict the depressive symptoms, we will use the participant i's rhythmic parameters based on each of the app usage behavioral markers of m_i to calculate the cosine similarity with all other participants $n-1$. $s_{ij}=\frac{B_i B_j}{¿B_i \vee ¿B_j \vee ¿¿}$; B_i, B_j present the vectors ($\forall x(x \in m_i \rightarrow x \in B_i), \forall x(x \in m_j \rightarrow x \in B_j)$) based on the rhythmic parameters set m_i and m_j of i and j participants respectively. Cosine similarity s_{ij} will be close to 1 when there is a higher similarity between i and j participants. After finding the similarities of participant i with all participants, we will be using the most similar N participants to train the multi-task learning framework for i. The value of N will be decided empirically through a search on values 200, 250, 300, and 350.

MTL framework development and validation

While there remain some similar tasks, the development of the model through the MTL strategy can facilitate improving the performance of the models through information sharing. Besides, if different models are developed for different tasks, then, the system can be resource inefficient. In a previous study [48], researchers used the MTL technique to develop systems for predicting the symptoms of Schizophrenia. However, in that system, all symptom prediction tasks were considered the same. In reality, all the tasks included in a model do not help each other to improve performance [50]. Therefore, we will be finding the similarities among the symptoms and will be using similar symptoms prediction tasks in a model while other similar symptoms' prediction tasks will be in another model. To group the tasks, we will calculate the correlation coefficients among the symptoms. The symptoms which are highly (coefficients > 0.7), moderately (0.4 <= coefficients <= 0.7) [109], and less than moderately correlated, or not correlated will be kept in 3 different groups.

While developing the MTL framework, we will use the hard parameter-sharing technique since in this approach, the model can find a common representation to capture all of the tasks which can reduce the potential risk of overfitting [108]. Combining multiple loss functions brings promising performances [110]. In our study, we will use a weighted loss of the hinge and cross-entropy. However, we will not use a fixed weight. Instead, we will tune the weight using the Bayesian Search optimization technique which selects the next parameter based on the performance of the previously selected one.

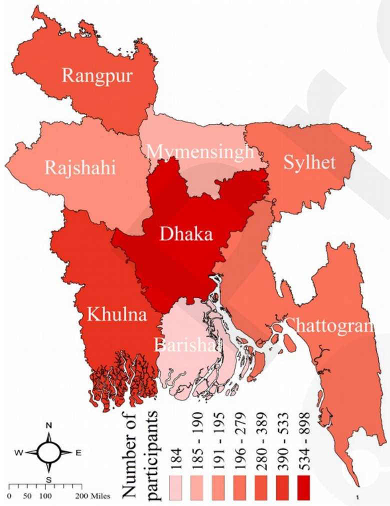
To validate the framework, we will use a nested cross-validation method since this approach was found to have a generalizable performance compared to K-fold cross-validation (CV) [99]. In the outer loop, the Leave One Out Cross Validation (LOOCV) method will be used and in the inner loop, there will be a 10-fold CV where in each interaction, 9 folds will be used for tuning the hyper-parameters and the remaining 1 fold will be used for validation. It can be mentioned that we are aware of the fact using LOOCV will increase the time complexity since there are 2902 participants and for each of them, a personalized model will be developed. However, we choose to use LOOCV because it will work like the real-world scenario where the model will predict a single participant's depressive symptoms at a time. Besides, this process will help in personalizing the model where only the participants who are similar (in terms of app usage behavior without using any information about the depressive symptoms) to the student for whom the model will predict the symptoms, will be used in model training as we discussed in detail in the section "Participant-similarity and development of personalized models". During the development of the models, we will maximize the balanced accuracy as it is based on sensitivity and specificity, and having a higher balanced accuracy can lead to higher precision and F1 score.

To understand the robustness of the proposed framework, we will compare it with the performance in the following approaches.

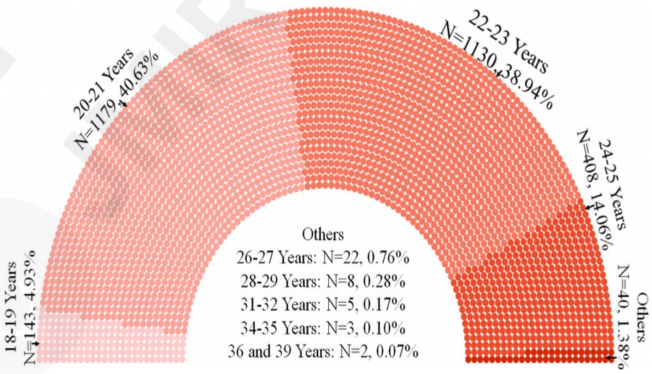
1. Comparison with single-task learning-based models: We will compare the performance of the proposed personalized MTL framework with the single-task learning (STL) models. This will resemble the approach presented in the previous study [33] where each symptom was considered as a single task. To develop the STL model, we will use ML algorithms such as RF, support vector machine (SVM), decision tree (DT) [106], and logistic regression [107] which are widely used in medical informatics as shown in systematic reviews [106, 107].
2. Comparison with non-personalized MTL framework: Since we expect that the personalization may provide better performance as discussed in the last subsection, we will compare the performance of the personalized MTL framework with a non-personalized MTL framework where we will use $n-1$ participants' data for training, instead of using a personalized subset of data.
3. Comparison with MTL framework without grouping the tasks: To understand how grouping tasks based on similarity impact on performance, we will compare performance of the MTL framework with and without grouping tasks.

Results

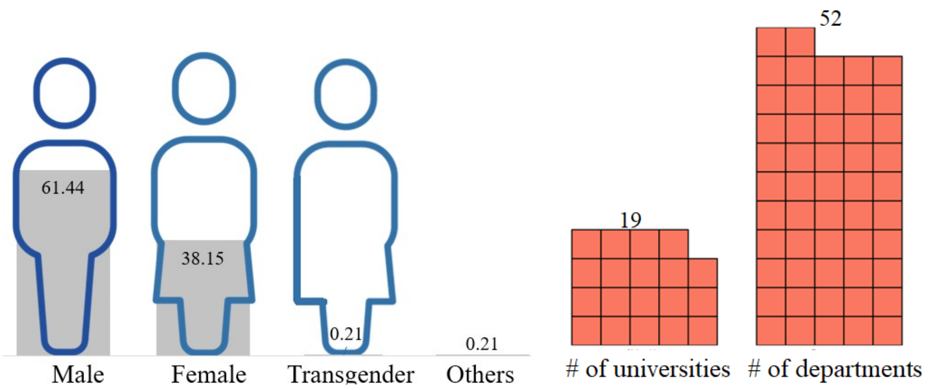
We are working on rhythm detection, rhythmic feature extraction, and the MTL framework development. We are expecting to publish our findings by June 2024. We present the findings regarding participants' demographic characteristics below.



(a) Map of Bangladesh presenting living division



(b) Age



(c) Gender (in percentage) (d) Institutions in number

Figure 4: Participants' demographic characteristics on (a) living division, (b) age, (c) gender, and (d) institutions

The participants were from each of the 8 divisions of Bangladesh (Figure 4 (a)). Most participants (N=887, 30.56%) were from the Dhaka division which also reflects the fact that the majority of the university students of Bangladesh reside in this division. The participants' age varied from 18 to 39 years and 79.57% of participants were aged 20-23 years (Figure 4(b)). Among 2902 participants, 38.15% and 61.44% were female and male participants respectively (Figure 4(c)). The students from public and private universities were 83.74% (N=2430) and 16.26% (N=472) respectively. The participants were from 19 universities (Figure 4(d)) including specialized universities in particular subjects namely, agriculture, engineering, and textile. They were from 52 different departments including students from the Arts (e.g., Dept. of Sculpture), Business (e.g., Dept. of Management Studies), Engineering (e.g., Dept. of Petroleum & Mining Engineering), Science (e.g., Dept. of Botany), Textile (e.g., Dept. of Apparel Engineering), Public health, and Law faculties.

Discussion

Significance

By using the dataset constructed through a countrywide study on 2902 students having over 24 million app usage events, we will explore whether there is a statistically significant rhythm based on the different app usage behavioral markers. We hypothesize the app usage behavioral markers such as the relative importance of an app category have rhythmic patterns with reproducible waveforms because the app usage markers vary depending on factors such as time of the day [59, 82, 83] like the physiological data. Additionally, since rhythmic features based on physiological [85, 95] and activity [100] data showed the potential application in problems such as knowing the participants having a higher risk of disease [95], knowing sedentary behavior [85], finding the subtle changes in detecting COVID-19 [95], an in-depth exploration of the app usage marker-based rhythms may show the alternative source of data to know the rhythms in human life. App usage marker-based rhythm has the possibility to be used for different purposes. For example, having a statistically significant relation between the rhythmicity of app usage and depressive symptoms can create the possibility of using these rhythmic features for the intervention to mitigate depression.

Apart from these, by predicting symptoms, our study will extend the previous studies since most studies (details in a recent systematic review [52]) in the pervasive health area developed classification models (e.g., to classify the depressed and non-depressed) where the complexity of the psychological problem depression may be lost. For instance, a participant with a PHQ-9 score of 10 has moderate depression [8] and a score of 10 can be found in a different combination of the scores of the PHQ-9 scale's 9 symptoms. As a result, by classifying the participants into some groups based on the overall score of a scale, it is not possible to precisely know the depressive symptoms that bothered the student. However, it is very important to know since each depressive symptom (e.g., symptoms in the PHQ-9 scale [8]) presents a unique phenomenon (e.g., Anhedonia, sleep disturbance, suicidal ideation) [38]. Therefore, depending on our proposed personalized MTL framework's performance based on real-time data, the proposed system can contribute to early diagnosis of depressive symptoms and precisely understanding a depressed student which in turn may contribute to mitigating depression prevalence.

Our previous pilot studies in Bangladesh regarding the relation of app usage with depression [34, 91] and loneliness [90], classifying the depressed and non-depressed [25], lonely and non-lonely [26] students showed the promise of models solely based on resource insensitive [25] app usage behavioral markers. Incorporation of the app usage rhythmic features and also the MTL framework by leveraging the similarities among the symptoms' prediction tasks so that tasks do not hurt each other's performance, may help the researchers and developers in developing more robust models to predict the symptoms of psychological problems through solely app usage data. Besides, our tool's reliance on retrieved data from a smartphone within a second [25] may also make it feasible in low and middle-income countries since smartphones are more affordable [88] compared to unaffordable [87] wearables which are usually used to get the physiological data and extract the rhythmic features.

Strengths

The median sample size of the previous studies that classified or predicted depression is 58 and none of the studies that developed computational models for prediction tasks had a sample size over 500 as presented by a recent systematic review [20]. On the other hand, we constructed a large dataset having 2902 students. In addition, the participants of our study are from each division of Bangladesh, both from public and private universities and 51 different departments. As far as we have seen, this is the largest dataset having data on both app usage and depressive symptoms. Considering these facts, our findings based on the proposed methods may be generalizable, robust enough to be impactful in the real world, and may contribute significantly to advancing the knowledge in mobile and pervasive health research areas.

To our best knowledge, we are the first to explore in-depth the rhythms based on different app usage behavioral markers which can open the opportunity to find an alternative source of knowing rhythms of daily life without depending on the physiological data-based rhythms which are usually retrieved by costly wearables.

In our recent work based on app usage [25], our developed system had a higher performance in predicting depression than the existing systems which were based on app usage as well. Through feature analysis (for details, please see [25]), we found higher importance of our newly explored behavioral markers (e.g., ratio of hamming distance [25]) compared to the features used in the previous studies. That being said, depending on the behavioral markers used in a model, the performance varies. Hence, our presented some novel behavioral markers (e.g., relative importance of the app categories, co-usage of apps) in this protocol which were not explored in previous pervasive health research have significance. Besides, to predict depressive symptoms, we will develop a personalized MTL framework. Though an MTL framework was developed in some previous studies (e.g., in [48]) to predict mental state, our study will add knowledge by showing the performance of personalized MTL.

Limitations

Following previous studies [48, 62, 63], we kept the 2902 participants to analyze rhythms who had app usage data of at least 2 days and 98.17% (n=2849) of them had app usage data of 7 days.

However, having data of more than 7 days could help us to better understand the stability of the rhythms, the rhythm disruption over weeks, and its potential effect on the depressive symptoms' appearance. However, we believe our study can work as a primary cursor for future studies to explore more the app usage rhythms.

Though our study included over 2900 students from different regions of Bangladesh, our proposed system may not be generalizable for every student since behavior varies depending on many factors such as season, region, [21], and socioeconomic status [22]. We recommend future studies to include participants based on more factors that can impact the behavior. Moreover, in our study, though we included participants from all 8 divisions of Bangladesh, we could not include participants from all 64 districts of the divisions. Additionally, having more participants from very rural areas could have the potential to get a more reliable picture of the students' behavior which in turn can be useful to develop a better system.

Conclusions

Predicting depressive symptoms accurately could help in better diagnosis of depression and taking appropriate steps accordingly. However, the existing models regarding the symptoms' prediction are limited by various issues including the lower performance (e.g., specificity is around or below 60% for most symptoms). Our proposed approach to explore rhythmic features from the app usage behavioral markers and development and validation of the MTL framework through our constructed large-scale dataset may bring new insights regarding rhythms and higher performance for predicting depressive symptoms.

Acknowledgements

This research has been funded by North South University (ID: NonCTRG-22-26) and the Institute of Advanced Research (ID: UIU-IAR-01-2022-SE-21) of the United International University. We are grateful to the faculties of universities located in different regions of Bangladesh without whose cordial support, it would not be possible to conduct this countrywide study. We also thank the participants for their time and willingness to donate data.

Data Availability

Since the app usage data is sensitive, making the data publicly available can raise different data privacy and safety concerns (e.g., reidentification of the participants [16]). Thus, we do not have a plan to upload the data to any public data repositories. But to give access to the data, we will consider a reasonable request that can be sent to the corresponding author.

Conflicts of Interest

None declared.

Abbreviations

IS: Interdaily Stability
IV: Intradaily Variability
LOOCV: Leave One Out Cross Validation
MTL: Multi-Task Learning
PHQ-9: Patient Health Questionnaire-9
RA: Relative Amplitude

References

1. WHO. Suicide prevention. Who.int. Available from: https://www.who.int/health-topics/suicide [accessed Mar 18, 2023]
2. Cai H, Xie X-M, Zhang Q, Cui X, Lin J-X, Sim K, Ungvari GS, Zhang L, Xiang Y-T. Prevalence of suicidality in major depressive disorder: A systematic review and meta-analysis of comparative studies. Front Psychiatry 2021;12:690130. PMID:34603096
3. Dong M, Zeng L-N, Lu L, Li X-H, Ungvari GS, Ng CH, Chow IHI, Zhang L, Zhou Y, Xiang Y-T. Prevalence of suicide attempt in individuals with major depressive disorder: a meta-analysis of observational surveys. Psychol Med 2019;49(10):1691–1704. PMID:30178722
4. WHO. The global burden of disease. Who.int. 2008. Available from: https://apps.who.int/iris/bitstream/handle/10665/43942/9789241563710_eng.pdf [accessed Mar 18, 2023]
5. 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. PMID:34634250
6. Hosen I, Al-Mamun F, Mamun MA. Prevalence and risk factors of the symptoms of depression, anxiety, and stress during the COVID-19 pandemic in Bangladesh: a systematic review and meta-analysis. Glob Ment Health (Camb) 2021;8(e47):e47. PMID:35145709
7. WHO. Depression and Other Common Mental Disorders: Global Health Estimates. Who.int. 2017. Available from: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed Mar 18, 2023]
8. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606–613. PMID:11556941
9. Fusar-Poli P, McGorry PD, Kane JM. Improving outcomes of first-episode psychosis: an overview. World Psychiatry 2017;16(3):251–265. PMID:28941089
10. WHO. Bangladesh WHO special initiative for mental health situational assessment. Who.int. 2019. Available from: https://www.who.int/docs/default-source/mental-health/special-initiative/who-special-initiative-country-report---bangladesh---2020.pdf?sfvrsn=c2122a0e_2 [accessed Mar 18, 2023]
11. BANBEIS. Gov.bd. 2021. Available from: http://www.banbeis.gov.bd/sites/default/files/files/banbeis.portal.gov.bd/files/142326c5_ea2f_43fc_ae28_6f3bedc6059e/Bangladesh%20Education%20Statistics%202021_compressed-236-404.pdf [accessed Mar 18, 2023]
12. BBS. POPULATION & HOUSING CENSUS 2022. 2022. Available from: http://www.bbs.gov.bd/site/page/47856ad0-7e1c-4aab-bd78-892733bc06eb/Population-&-Housing
13. Coyne JC, Schwenk TL, Fechner-Bates S. Nondetection of depression by primary care physicians reconsidered. Gen Hosp Psychiatry 1995;17(1):3–12. PMID:7737492
14. Cepoiu M, McCusker J, Cole MG, Sewitch M, Ciampi A. Recognition of depression in older medical inpatients. J Gen Intern Med 2007;22(5):559–564. PMID:17443361
15. Ahmed MS. Mon Majhi. Google.com. 2022. Available from: https://play.google.com/store/apps/details/?id=net.mn.u [accessed Mar 18, 2023]
16. Ahmed MS, Ahmed N. Exploring unique app signature of the depressed and non-depressed through their fingerprints on apps. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering Cham: Springer International Publishing; 2022. p. 218–239.
17. Lu J, Shang C, Yue C, Morillo R, Ware S, Kamath J, Bamis A, Russell A, Wang B, Bi J. Joint modeling of heterogeneous sensing data for depression assessment via multi-task learning. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018;2(1):1–21. doi: 10.1145/3191753
18. Xu X, Chikersal P, Dutcher JM, Sefidgar YS, Seo W, Tumminia MJ, Villalba DK, Cohen S, Creswell KG, Creswell JD, Doryab A, Nurius PS, Riskin E, Dey AK, Mankoff J. Leveraging collaborative-filtering for personalized behavior modeling: A case study of depression detection among college students. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021;5(1):1–27. doi: 10.1145/3448107
19. Xu X, Chikersal P, Doryab A, Villalba DK, Dutcher JM, Tumminia MJ, Althoff T, Cohen S, Creswell KG, Creswell JD, Mankoff J, Dey AK. Leveraging routine behavior and contextually-filtered features for depression detection among college students. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019;3(3):1–33. doi: 10.1145/3351274
20. 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 Digit Med 2022;5(1):3. PMID:35017634
21. Sekara V, Alessandretti L, Mones E, Jonsson H. Temporal and cultural limits of privacy in smartphone app usage. Sci Rep 2021;11(1):3861. PMID:33594096
22. Tu Z, Cao H, Lagerspetz E, Fan Y, Flores H, Tarkoma S, Nurmi P, Li Y. Demographics of mobile app usage: long-term analysis of mobile app usage. CCF Trans Pervasive Comp Interact 2021;3(3):235–252. doi: 10.1007/s42486-020-00041-3

23. Chikersal P, Doryab A, Tumminia M, Villalba DK, Dutcher JM, Liu X, Cohen S, Creswell KG, Mankoff J, Creswell JD, Goel M, Dey AK. Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing: A machine learning approach with robust feature selection. ACM Trans Comput Hum Interact 2021;28(1):1–41. doi: 10.1145/3422821
24. Yan R, Liu X, Dutcher J, Tumminia M, Villalba D, Cohen S, Creswell D, Creswell K, Mankoff J, Dey A, Doryab A. A computational framework for modeling biobehavioral rhythms from mobile and wearable data streams. ACM Trans Intell Syst Technol 2022;13(3):1–27. doi: 10.1145/3510029
25. Ahmed MS, Ahmed N. A Fast and Minimal System to Identify Depression Using Smartphones: Explainable Machine Learning–Based Approach. JMIR Forma 2023; doi: 10.2196/28848
26. Ahmed MS, Ahmed N. Less is more: Leveraging digital behavioral markers for real-time identification of loneliness in resource-limited settings. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering Cham: Springer Nature Switzerland; 2023. p. 460–476.
27. Dogrucu A, Perucic A, Isaro A, Ball D, Toto E, Rundensteiner EA, Agu E, Davis-Martin R, Boudreaux E. Moodable: On feasibility of instantaneous depression assessment using machine learning on voice samples with retrospectively harvested smartphone and social media data. Smart Health 2020;17(100118):100118. doi: 10.1016/j.smhl.2020.100118
28. Canzian L, Musolesi M. Trajectories of depression: Unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing New York, NY, USA: ACM; 2015.
29. Ware S, Yue C, Morillo R, Lu J, Shang C, Kamath J, Bamis A, Bi J, Russell A, Wang B. Large-scale automatic depression screening using meta-data from WiFi infrastructure. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018;2(4):1–27. doi: 10.1145/3287073
30. Yue C, Ware S, Morillo R, Lu J, Shang C, Bi J, Kamath J, Russell A, Bamis A, Wang B. Fusing location data for depression prediction. IEEE Trans Big Data 2021;7(2):355–370. PMID:35498556
31. Farhan AA, Yue C, Morillo R, Ware S, Lu J, Bi J, Kamath J, Russell A, Bamis A, Wang B. Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data. 2016 IEEE Wireless Health (WH) IEEE; 2016. p. 1–8.
32. Yue C, Ware S, Morillo R, Lu J, Shang C, Bi J, Kamath J, Russell A, Bamis A, Wang B. Automatic depression prediction using Internet traffic characteristics on smartphones. Smart Health 2020;18:100137. PMID:33043105
33. Ware S, Yue C, Morillo R, Lu J, Shang C, Bi J, Kamath J, Russell A, Bamis A, Wang B. Predicting depressive symptoms using smartphone data. Smart Health 2020;15(100093):100093. doi: 10.1016/j.smhl.2019.100093
34. Ahmed MS, Rony RJ, Hasan T, Ahmed N. Smartphone usage behavior between depressed and non-depressed students: An exploratory study in the context of Bangladesh. Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers New York, NY, USA: ACM; 2020b. doi: 10.1145/3410530.3414441
35. Ng K, Sun J, Hu J, Wang F. Personalized predictive modeling and risk factor identification using patient similarity. AMIA Summits Transl Sci Proc 2015;2015:132–136. PMID:26306255
36. Sharafoddini A, Dubin JA, Lee J. Patient similarity in prediction models based on health data: A scoping review. JMIR Med Inform 2017;5(1):e7. PMID:28258046
37. Lee J, Maslove DM, Dubin JA. Personalized mortality prediction driven by electronic medical data and a patient similarity metric. PLoS One 2015;10(5):e0127428. PMID:25978419
38. Yun J-Y, Yun YH. Health-promoting behavior to enhance perceived meaning and control of life in chronic disease patients with role limitations and depressive symptoms: a network approach. Sci Rep 2023;13(1):4848. PMID:36964273
39. NIMH. Depression. National Institute of Mental Health (NIMH). 2023. Available from: https://www.nimh.nih.gov/health/topics/depression [accessed Apr 14, 2023]
40. WHO. Depressive disorder (depression). Who.int. 2023. Available from: https://www.who.int/news-room/fact-sheets/detail/depression [accessed Apr 14, 2023]
41. Owoyemi A, Owoyemi J, Osiyemi A, Boyd A. Artificial intelligence for healthcare in Africa. Front Digit Health 2020;2:6. PMID:34713019
42. Islam MA, Barna SD, Raihan H, Khan MNA, Hossain MT. Depression and anxiety among university students during the COVID-19 pandemic in Bangladesh: A web-based cross-sectional survey. PLoS One 2020;15(8):e0238162. PMID:32845928
43. Delaporte A, Bahia K. The State of Mobile Internet Connectivity report 2022 - mobile for development. Mobile for Development. 2022. Available from: https://www.gsma.com/r/somic/ [accessed Apr 15, 2023]
44. Hossain MJ, Ahmmed F, Khandokar L, Rahman SMA, Hridoy A, Ripa FA, Emran TB, Islam MR, Mitra S, Alam M. Status of psychological health of students following the extended university closure in Bangladesh: Results from a web-based cross-sectional study. PLOS Glob Public Health 2022;2(3):e0000315. PMID:36962237
45. Wang R, Wang W, daSilva A, Huckins JF, Kelley WM, Heatherton TF, Campbell AT. Tracking depression dynamics

in college students using mobile phone and wearable sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018;2(1):1–26. doi: 10.1145/3191775
46. Briganti G, Scutari M, Linkowski P. Network structures of symptoms from the Zung Depression Scale. Psychol Rep 2021;124(4):1897–1911. PMID:32686585
47. Cheung T, Jin Y, Lam S, Su Z, Hall BJ, Xiang Y-T, International Research Collaboration on COVID-19. Network analysis of depressive symptoms in Hong Kong residents during the COVID-19 pandemic. Transl Psychiatry 2021;11(1):460. PMID:34489416
48. Tseng VW-S, Sano A, Ben-Zeev D, Brian R, Campbell AT, Hauser M, Kane JM, Scherer EA, Wang R, Wang W, Wen H, Choudhury T. Using behavioral rhythms and multi-task learning to predict fine-grained symptoms of schizophrenia. Sci Rep 2020;10(1):15100. PMID:32934246
49. Ware S, Knouse LE, Draz I, Enikeeva A. Predicting ADHD symptoms using smartphone sensing data. Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp/ISWC ’22 Adjunct), September 11â•ﬁ15, 2022, Cambridge, United Kingdom 2022; Available from: https://ubicomp-mental-health.github.io/papers/2022/Ware-ADHD.pdf
50. Standley T, Zamir AR, Chen D, Guibas L, Malik J, Savarese S. Which tasks should be learned together in multi-task learning? Proceedings of the 37th International Conference on Machine Learning (ICML) 2020. Available from: http://proceedings.mlr.press/v119/standley20a/standley20a.pdf [accessed Apr 16, 2023]
51. Mullick T, Radovic A, Shaaban S, Doryab A. Predicting depression in adolescents using mobile and wearable sensors: Multimodal machine learning-based exploratory study. JMIR Form Res 2022;6(6):e35807. PMID:35749157
52. Abd-Alrazaq A, AlSaad R, Aziz S, Ahmed A, Denecke K, Househ M, Farooq F, Sheikh J. Wearable artificial intelligence for anxiety and depression: Scoping review. J Med Internet Res 2023;25(1):e42672. PMID:36656625
53. Baryshnikov I, Aledavood T, Rosenström T, Heikkilä R, Darst R, Riihimäki K, Saleva O, Ekelund J, Isometsä E. Relationship between daily rated depression symptom severity and the retrospective self-report on PHQ-9: A prospective ecological momentary assessment study on 80 psychiatric outpatients. J Affect Disord 2023;324:170–174. PMID:36586594
54. Pedrelli P, Fedor S, Ghandeharioun A, Howe E, Ionescu DF, Bhathena D, Fisher LB, Cusin C, Nyer M, Yeung A, Sangermano L, Mischoulon D, Alpert JE, Picard RW. Monitoring changes in depression severity using wearable and mobile sensors. Front Psychiatry 2020;11:584711. PMID:33391050
55. Ghandeharioun A, Fedor S, Sangermano L, Ionescu D, Alpert J, Dale C, Sontag D, Picard R. Objective assessment of depressive symptoms with machine learning and wearable sensors data. 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII) IEEE; 2017. p. 325–332.
56. Rykov Y, Thach T-Q, Bojic I, Christopoulos G, Car J. Digital biomarkers for depression screening with wearable devices: Cross-sectional study with machine learning modeling. JMIR MHealth UHealth 2021;9(10):e24872. PMID:34694233
57. Cao J, Truong AL, Banu S, Shah AA, Sabharwal A, Moukaddam N. Tracking and predicting depressive symptoms of adolescents using smartphone-based self-reports, parental evaluations, and passive phone sensor data: Development and usability study. JMIR Ment Health 2020;7(1):e14045. PMID:32012072
58. Doryab A. Critical thinking of mobile sensing for health. Crossroads 2021;28(1):34–37. doi: 10.1145/3481831
59. Gordon ML, Gatys L, Guestrin C, Bigham JP, Trister A, Patel K. App usage predicts cognitive ability in older adults. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems New York, NY, USA: ACM; 2019.
60. Sarsenbayeva Z, Marini G, van Berkel N, Luo C, Jiang W, Yang K, Wadley G, Dingler T, Kostakos V, Goncalves J. Does Smartphone Use Drive our Emotions or vice versa? A Causal Analysis. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems New York, NY, USA: ACM; 2020.
61. SurveyMonkey. Sample size calculator: Understanding sample sizes. SurveyMonkey. Available from: https://www.surveymonkey.com/mp/sample-size-calculator/ [accessed Jul 6, 2023]
62. Dijk D-J, Duffy JF. Novel approaches for assessing circadian rhythmicity in humans: A review. J Biol Rhythms 2020;35(5):421–438. PMID:32700634
63. Thomas KA, Burr RL. Circadian research in mothers and infants: how many days of actigraphy data are needed to fit cosinor parameters? J Nurs Meas 2008;16(3):201–206. PMID:19886472
64. Chowdhury A, Ghosh S, Sanyal D. Bengali adaptation of Brief Patient Health Questionnaire for screening depression at primary care. Journal of the Indian Medical Association 2004;
65. Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Ann Card Anaesth 2019;22(1):67–72. PMID:30648682
66. Khan AH, Tian X-L. Circadian Amplitude. Encyclopedia of Gerontology and Population Aging Cham: Springer International Publishing; 2021. p. 1003–1012.
67. Cornelissen G. Cosinor-based rhythmometry. Theor Biol Med Model 2014;11(1):16. PMID:24725531
68. Gonçalves BSB, Adamowicz T, Louzada FM, Moreno CR, Araujo JF. A fresh look at the use of nonparametric analysis in actimetry. Sleep Med Rev 2015;20(C):84–91. PMID:25065908

69. Mutak A. cosinor2 vignette. R-project.org. 2018. Available from: https://cran.r-project.org/web/packages/cosinor2/vignettes/cosinor2.html [accessed Jul 21, 2023]
70. Anderson ST, FitzGerald GA. Sexual dimorphism in body clocks. Science 2020;369(6508):1164–1165. doi: 10.1126/science.abd4964
71. Duarte LL, Menna-Barreto L. Chronotypes and circadian rhythms in university students. Biol Rhythm Res 2022;53(7):1058–1072. doi: 10.1080/09291016.2021.1903791
72. Mitchell JA, Quante M, Godbole S, James P, Hipp JA, Marinac CR, Mariani S, Cespedes Feliciano EM, Glanz K, Laden F, Wang R, Weng J, Redline S, Kerr J. Variation in actigraphy-estimated rest-activity patterns by demographic factors. Chronobiol Int 2017;34(8):1042–1056. PMID:28650674
73. Jones SH, Hare DJ, Evershed K. Actigraphic assessment of circadian activity and sleep patterns in bipolar disorder. Bipolar Disord 2005;7(2):176–186. PMID:15762859
74. Masubuchi S, Hashimoto S, Endo T, Honma S, Honma K-I. Amplitude reduction of plasma melatonin rhythm in association with an internal desynchronization in a subject with non-24-hour sleep–wake syndrome. Psychiatry Clin Neurosci 1999;53(2):249–251. doi: 10.1046/j.1440-1819.1999.00535.x
75. Abdullah S, Matthews M, Murnane EL, Gay G, Choudhury T. Towards circadian computing: “Early to bed and early to rise” makes some of us unhealthy and sleep deprived. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing New York, NY, USA: ACM; 2014.
76. Murnane EL, Abdullah S, Matthews M, Kay M, Kientz JA, Choudhury T, Gay G, Cosley D. Mobile manifestations of alertness: Connecting biological rhythms with patterns of smartphone app use. Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services New York, NY, USA: ACM; 2016.
77. Esaki Y, Obayashi K, Saeki K, Fujita K, Iwata N, Kitajima T. Association between circadian activity rhythms and mood episode relapse in bipolar disorder: a 12-month prospective cohort study. Transl Psychiatry 2021;11(1):525. PMID:34645802
78. Mayeli A, LaGoy AD, Smagula SF, Wilson JD, Zarbo C, Rocchetti M, Starace F, Zamparini M, Casiraghi L, Calza S, Rota M, D’Agostino A, de Girolamo G, DiAPAson Consortium, Ferrarelli F. Shared and distinct abnormalities in sleep-wake patterns and their relationship with the negative symptoms of Schizophrenia Spectrum Disorder patients. Mol Psychiatry 2023;1–9. PMID:37055512
79. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: With applications in R. New York, NY: Springer US; 2021. ISBN:9781071614174
80. Quante M, Mariani S, Weng J, Marinac CR, Kaplan ER, Rueschman M, Mitchell JA, James P, Hipp JA, Cespedes Feliciano EM, Wang R, Redline S. Zeitgebers and their association with rest-activity patterns. Chronobiol Int 2019;36(2):203–213. PMID:30365354
81. Ryu S, Oh H. Duration and content type of smartphone use in relation to diet and adiposity in 53,133 adolescents. Curr Dev Nutr 2021;5:1088. doi: 10.1093/cdn/nzab053_081
82. Böhmer M, Hecht B, Schöning J, Krüger A, Bauer G. Falling asleep with Angry Birds, Facebook and Kindle: A large scale study on mobile application usage. Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services New York, NY, USA: ACM; 2011.
83. Morrison A, Xiong X, Higgs M, Bell M, Chalmers M. A large-scale study of iPhone app launch behaviour. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems New York, NY, USA: ACM; 2018.
84. Mehrotra A, Müller SR, Harari GM, Gosling SD, Mascolo C, Musolesi M, Rentfrow PJ. Understanding the role of places and activities on mobile phone interaction and usage patterns. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017;1(3):1–22. doi: 10.1145/3131901
85. Miyagi R, Sasawaki Y, Shiotani H. The influence of short-term sedentary behavior on circadian rhythm of heart rate and heart rate variability. Chronobiol Int 2019;36(3):374–380. doi: 10.1080/07420528.2018.1550422
86. Doryab A, Dey AK, Kao G, Low C. Modeling biobehavioral rhythms with passive sensing in the wild: A case study to predict readmission risk after pancreatic surgery. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019;3(1):1–21. doi: 10.1145/3314395
87. Elias J. Why wearables are out of reach for people who need them most. Forbes. 2015. Available from: https://www.forbes.com/sites/jenniferelias/2015/10/27/the-leftovers-part-i-the-cost-of-activity/?sh=a0e5ffc63c9b [accessed Jul 28, 2023]
88. Lohchab H. Affordable smartphone sales shoot up as demand grows due to e-learning needs. Economic Times. 2020. Available from: https://economictimes.indiatimes.com/tech/hardware/affordable-smartphone-sales-shoot-up-as-demand-grows/articleshow/78241563.cms [accessed Jul 28, 2023]
89. Ahmed MS, Rony RJ, Hadi MA, Hossain E, Ahmed N. A Minimalistic Approach to Predict and Understand the Relation of App Usage with Students’ Academic Performances. PACM on Human-Computer Interaction 2023;7(MHCI). doi: 10.1145/3604240
90. Ahmed MS, Khan SS, Ahmed N. Is There Any Relation Between Smartphone Usage and Loneliness During the

COVID-19 Pandemic?: A Study by Exploring Two Objective App Usage Datasets. EAI Endorsed Transactions on Pervasive Health and Technology 2023. DOI: 10.4108/eetpht.9.3663

91. Ahmed MS, Hasan T, Rahman MM, Ahmed N. A rule mining and Bayesian network analysis to explore the link between depression and digital behavioral markers of games app usage. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering Cham: Springer Nature Switzerland; 2023. p. 557–569.
92. Ahmed MS, Jahangir Rony R, Ahmed N. Identifying high and low academic result holders through smartphone usage data. Asian CHI Symposium 2021 New York, NY, USA: ACM; 2021.
93. Ahmed MS, Rony RJ, Ashhab MM, Ahmed N. An empirical study to analyze the impact of Instagram on students' academic results. 2020 IEEE Region 10 Symposium (TENSYMP) IEEE; 2020a. p. 666–669.
94. Chiny M, Chihab M, Bencharef O, Chihab Y. Netflix recommendation system based on TF-IDF and cosine similarity algorithms. Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning SCITEPRESS - Science and Technology Publications; 2021.
95. Sarwar A, Agu EO, Almadani A. CovidRhythm: A deep learning model for passive prediction of covid-19 using biobehavioral rhythms derived from wearable physiological data. IEEE Open J Eng Med Biol 2023;4:21–30. doi: 10.1109/ojemb.2023.3261223
96. Davidsen AS, Fosgerau CF. What is depression? Psychiatrists' and GPs' experiences of diagnosis and the diagnostic process. Int J Qual Stud Health Well-being 2014;9(1):24866. doi: 10.3402/qhw.v9.24866
97. Moškon M. CosinorPy: a python package for cosinor-based rhythmometry. BMC Bioinformatics 2020;21(1). doi: 10.1186/s12859-020-03830-w
98. Bingham C, Arbogast B, Guillaume GC, Lee JK, Halberg F. Inferential statistical methods for estimating and comparing cosinor parameters. Chronobiologia 1982;9(4):397–439. PMID:7168995
99. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One 2019;14(11):e0224365. doi: 10.1371/journal.pone.0224365
100. Xu Y, Su S, Li X, Mansuri A, McCall WV, Wang X. Blunted rest-activity circadian rhythm increases the risk of all-cause, cardiovascular disease and cancer mortality in US adults. Sci Rep 2022;12(1):20665. PMID:36450759
101. Saha K, Grover T, Mattingly SM, Swain VD, Gupta P, Martinez GJ, Robles-Granda P, Mark G, Striegel A, De Choudhury M. Person-centered predictions of psychological constructs with social media contextualized by multimodal sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021;5(1):1–32. doi: 10.1145/3448117
102. Tu Z, Li R, Li Y, Wang G, Wu D, Hui P, Su L, Jin D. Your apps give you away: Distinguishing mobile users by their app usage fingerprints. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018;2(3):1–23. doi: 10.1145/3264948
103. Malmi E, Weber I. You are what apps you use: Demographic prediction based on user's apps. Proceedings of the International AAAI Conference on Web and Social Media 2021;10(1):635–638. doi: 10.1609/icwsm.v10i1.14776
104. UsageStatsManager. Android Developers. Available from: https://developer.android.com/reference/android/app/usage/UsageStatsManager [accessed Dec 22, 2023]
105. Sample Size Calculation when the Sampling Method is Simple Random Sampling (SRS). Uidaho.edu. Available from: https://www.webpages.uidaho.edu/ed571/OFFLINE-571-Modules/Offline_Modules/M3-S/Sample_Size_Calculations-Formulas.pdf [accessed Dec 22, 2023]
106. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, Collins GS, Bajpai R, Riley RD, Moons KGM, Hooft L. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ 2021;n2281. doi: 10.1136/bmj.n2281
107. Bernert RA, Hilberg AM, Melia R, Kim JP, Shah NH, Abnousi F. Artificial intelligence and suicide prevention: A systematic review of machine learning investigations. Int J Environ Res Public Health 2020;17(16):5929. doi: 10.3390/ijerph17165929
108. Ruder S. An overview of multi-task learning in deep neural networks. arXiv [csLG]. 2017. Available from: http://arxiv.org/abs/1706.05098 [accessed Dec 25, 2023]
109. Durrheim K, Tredoux C. Numbers, hypotheses & conclusions: A course in statistics for the social sciences. 6th ed. Juta and Company; 2004. ISBN:9781919713311
110. Mao Y, Wang Z, Liu W, Lin X, Xie P. MetaWeighting: Learning to weight tasks in multi-task learning. In: Muresan S, Nakov P, Villavicencio A, editors. Findings of the Association for Computational Linguistics: ACL 2022 Stroudsburg, PA, USA: Association for Computational Linguistics; 2022. p. 3436–3448.