

Original Paper

Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones: Observational Study

Akane Sano¹, PhD; Sara Taylor¹, MS; Andrew W McHill^{2,3}, PhD; Andrew JK Phillips^{2,3}, PhD; Laura K Barger^{2,3}, PhD; Elizabeth Klerman^{2,3}, PhD, MD; Rosalind Picard¹, ScD

¹Affective Computing Group, Media Lab, Massachusetts Institute of Technology, Cambridge, MA, United States

²Brigham and Women's Hospital, Boston, MA, United States

³Harvard Medical School, Boston, MA, United States

Corresponding Author:

Akane Sano, PhD

Affective Computing Group

Media Lab

Massachusetts Institute of Technology

75 Amherst Street

Cambridge, MA, 02139

United States

Phone: 1 6178999468

Email: akanes@media.mit.edu

Abstract

Background: Wearable and mobile devices that capture multimodal data have the potential to identify risk factors for high stress and poor mental health and to provide information to improve health and well-being.

Objective: We developed new tools that provide objective physiological and behavioral measures using wearable sensors and mobile phones, together with methods that improve their data integrity. The aim of this study was to examine, using machine learning, how accurately these measures could identify conditions of self-reported high stress and poor mental health and which of the underlying modalities and measures were most accurate in identifying those conditions.

Methods: We designed and conducted the 1-month SNAPSHOT study that investigated how daily behaviors and social networks influence self-reported stress, mood, and other health or well-being-related factors. We collected over 145,000 hours of data from 201 college students (age: 18-25 years, male:female=1.8:1) at one university, all recruited within self-identified social groups. Each student filled out standardized pre- and postquestionnaires on stress and mental health; during the month, each student completed twice-daily electronic diaries (e-diaries), wore two wrist-based sensors that recorded continuous physical activity and autonomic physiology, and installed an app on their mobile phone that recorded phone usage and geolocation patterns. We developed tools to make data collection more efficient, including data-check systems for sensor and mobile phone data and an e-diary administrative module for study investigators to locate possible errors in the e-diaries and communicate with participants to correct their entries promptly, which reduced the time taken to clean e-diary data by 69%. We constructed features and applied machine learning to the multimodal data to identify factors associated with self-reported poststudy stress and mental health, including behaviors that can be possibly modified by the individual to improve these measures.

Results: We identified the physiological sensor, phone, mobility, and modifiable behavior features that were best predictors for stress and mental health classification. In general, wearable sensor features showed better classification performance than mobile phone or modifiable behavior features. Wearable sensor features, including skin conductance and temperature, reached 78.3% (148/189) accuracy for classifying students into high or low stress groups and 87% (41/47) accuracy for classifying high or low mental health groups. Modifiable behavior features, including number of naps, studying duration, calls, mobility patterns, and phone-screen-on time, reached 73.5% (139/189) accuracy for stress classification and 79% (37/47) accuracy for mental health classification.

Conclusions: New semiautomated tools improved the efficiency of long-term ambulatory data collection from wearable and mobile devices. Applying machine learning to the resulting data revealed a set of both objective features and modifiable behavioral

features that could classify self-reported high or low stress and mental health groups in a college student population better than previous studies and showed new insights into digital phenotyping.

(*J Med Internet Res* 2018;20(6):e210) doi: [10.2196/jmir.9410](https://doi.org/10.2196/jmir.9410)

KEYWORDS

mobile health; mood; machine learning; wearable electronic devices; smartphone; mobile phone; mental health; psychological stress

Introduction

Background

Recent advances in wearable and mobile technologies have enabled individuals to monitor their daily lives and enabled scientific investigators to passively collect real-time data without disrupting people's habitual routines. Two examples of such devices are wrist-wearable devices that collect activity and other physiological data (eg, activity or sleep; heart rate; skin conductance, SC; blood pressure; and blood sugar level) and mobile phones (eg, smartphones) that monitor location, activity, social interaction over calls and texts (short message service, SMS), app use, screen on or off, and environmental data such as ambient light exposure and humidity.

Leveraging data from wearable and mobile devices to gain meaningful information about human health has been called digital phenotyping [1-3]. Digital phenotyping is defined as the moment-by-moment quantification of the individual-level human phenotype *in situ* using data from personal digital devices. Data from personal digital devices may be used to understand health and behaviors with a goal of preventing or minimizing disorders and diseases. For example, current health status, behavior history, and potential future health trajectories information might help (1) individuals become more aware of their risk profiles and enable them to make better informed decisions and take actions to change their behaviors to reduce potential negative physical and mental outcomes and (2) clinicians monitor changes in their client's or patient's status.

Mobile phones have been used to monitor stress and mental health [4-10]. The pioneering Student Life study that monitored 48 college students across a 10-week term using objective Android mobile phone sensors and usage investigated the relationship between well-being measures such as self-reported stress, depression, flourishing and loneliness, and academic performance [4]. Lower Perceived Stress Scale (PSS) score was correlated with higher conversation frequency during the day (9 AM-6 PM: the time frame participants might be in classes) and the evening (6 PM-0 AM), longer conversation duration during the day, and longer sleep duration. One study that evaluated self-reported depression using mobile phones for 2 weeks (N=28) [8] showed that mobility patterns (ie, regularity in 24-hour mobility patterns, as well as location variance) from Global Positioning System and phone usage features including usage duration and frequency were correlated with depressive symptom severity on a self-reported depression survey, the Patient Health Questionnaire-9 (PHQ-9) [8]. Another mobile phone-based study that lasted 12 weeks (N=73) identified mobile phone features that predicted clinically diagnosed depressed mood with 0.74 area under the curve; these features

including the total count of outgoing calls, the count of unique numbers texted, absolute distance traveled, dynamic variation of the voice, speaking rate, and voice quality [10].

The combination of wearable sensor and mobile phone data has also been used to study self-reported stress in daily life [11-14]. Muaremi et al, using both wearable sensors and mobile phones, developed a way to automate the recognition of self-reported daily stress levels using sleep parameters and 37 physiological responses (including heart rate, heart rate variability (HRV) and SC) from wearable sensors (N=10, 19 days), or mobile phone usage and sleep HRV from wearable sensors (N=35, 4 months). They showed 61% 3-class stress level classification accuracy with a combination of phone usage and sleep HRV features and 73% accuracy using sleep duration, upper body posture, and sleep HRV features [11,12]. Sano et al also investigated 5-day self-reported high or low stress recognition (N=18) and 1-month high or low stress recognition (N=66) using wearable sensor and mobile phone data; they showed 75% and 90% accuracy using leave-one participant-out or 10-fold cross-validation, respectively [13,14].

Objectives

These previous studies focused on only mobile phone usage or on phone usage plus wearable sensor data only during sleep and have not taken advantage of 24/7 multimodal phone + wearable data during wake and sleep to understand behaviors and physiology for long-term study of self-reported stress and mental health. We chose to approach this goal beginning with college students, most of whom report high stress, and some of whom are at risk of low or declining mental health [15,16]. According to the 2017 National College Health Assessment that examined data from 47,821 college students at 92 schools in the United States, more than half of the respondents said that their stress levels were higher than average, more than one-third had difficulty functioning because of depression, and two-thirds said they felt overwhelming anxiety in the last year [15]. Students' high stress and low mental health could negatively impact their academic performance [17]. Moreover, one-tenth of the students had a plan for suicide. Suicide rate is increasing, and suicide is the second leading cause of death for college students [18]. More students are seeking help, and 34% of counseling centers have a treatment waitlist [19]. Under these conditions, development of improved tools for screening, monitoring, and intervening for self-reported stress and poor mental health through wearable sensors and mobile phones in daily life settings will be beneficial. We aim to ultimately detect stress and mental health changes before clinical interventions are required and provide personalized early warnings together with data-driven suggestions of individualized behaviors that might promote better mental health outcomes.

Our SNAPSHOT study was designed to collect and examine rich multimodal information in participants' everyday life using wearable sensors and mobile phones for phenotyping sleep, stress, and mental health, all of which are major health issues in modern society. This paper has three main elements. First, we introduce a methodology and tools to capture long-term, large-scale ambulatory data on physiological and behavioral characteristics using sensors installed in wearable devices and mobile phones. The dataset from the SNAPSHOT study is one of the first large multimodal datasets that contains continuous physiology from a healthy college student population. The dataset currently includes approximately 145,000 hours of data from 201 participants at one university. Second, as real-world ambulatory data are messy, we describe tools we developed and deployed to improve the integrity and quality of the collected data and to reduce the time experimenters spend checking for and fixing errors. Third, we identify objective physiological markers and modifiable behaviors that successfully classify self-reported high or low stress and mental health and examine the separate contributions of wearable sensors and mobile phone data.

Methods

The 1-month SNAPSHOT study is a long-term and large-scale study developed to measure Sleep, Networks, Affect, Performance, Stress, and Health using Objective Techniques. Our aim was to investigate how daily behaviors and social networks influence sleep, self-reported stress, mood, performance, and other well-being-related factors. For each of five Fall and Spring semesters starting in Fall 2013, we collected approximately 1 month of data per person from college students who were socially connected and at a single New England university. Students were only allowed to participate in the study once. There was a total of 201 participants; Fall 2013: 20, Spring 2014: 48, Fall 2014: 46, Spring 2015: 47, Fall 2015: 40; ages 18 to 25 years; 129 male, 72 female; 70 freshman, 49 sophomore, 44 junior, 36 senior, and 2 unreported. The approximately 1 month of data collection was between the start of semester and midterms.

Recruitment

We intentionally recruited college students from a single academic institution who were socially connected because of our interest in how social networks affect sleep and health behaviors. Our definition of socially connected was making a call or SMS at least once a week with each other. Each semester, we recruited groups of at least 5 people who knew each other and interacted socially. We posted our study advertisement to undergraduate students' mailing lists. Potential participants filled out screening questionnaires to determine eligibility. Our exclusion criteria were as follows: (1) non-Android phone users, (2) inability to wear wrist sensors (eg, irritated skin on wrist), (3) pregnant women, (4) travel across more than one time zone 1 week before the study or have plans to travel more than one time zone away during the study, and (5) age <18 years or >60 years. In our study, we targeted only Android phone users because other mobile phones (eg, iPhone) did not allow us to monitor phone usage as needed for this study.

Eligible participants attended information and consent sessions. For each session, we invited approximately 15 participants and explained in detail the study and tasks that participants would perform during the study. After participants gave written informed consent, they completed prestudy questionnaires, started wearing devices, and installed an Android app (described below) on their phone. The study obtained a National Institutes of Health Certificate of Confidentiality so that potentially sensitive information such as drug or alcohol use provided by the participants could not be revealed for legal purposes; this was important protection for the students as the daily diary included requests for such information.

The participants received financial compensation at the end of the study; the amount depended on the number of days they completed diaries, wore the sensors, and completed other protocol tasks.

Study protocols were approved by the Massachusetts Institute of Technology and Partners HealthCare Institutional Review Boards. The study was registered on clinicaltrials.gov (NCT02846077).

Data Collection

All data were deidentified before analysis, although location information could potentially be used to reidentify people. Phone numbers, email addresses, and actual names from the social network surveys were hashed.

Start of the Study Questionnaires

At the start of the study, participants completed the Morningness-Eveningness Questionnaire [20], the Pittsburgh Sleep Quality Index [21], the Myers Brigg Personality test, the Big Five Inventory Personality Test [22], the PSS [23], the 12-Item Short Form Health Survey (SF-12) for physical and mental component summary (MCS) scores [24], and a set of social network surveys assessing with whom participants spent their time to help map their social networks. We also collected age, sex, academic major, and living situation (eg, dorm name and whether single or multiple occupancy room) information.

Ambulatory Monitoring

Wearable Sensors

Participants wore two sensors on their wrists: a Q-sensor (Affectiva, Boston, MA, United States) to measure SC, skin temperature (ST), three-axis *acceleration* (ACC) on their dominant wrist and a Motion Logger (AMI, Ardsley, NY, United States) on their nondominant wrist to measure acceleration and ambient light data. ACC can be used to estimate activity levels and sleep or wake patterns. SC reflects autonomic arousal during the day, providing a stress index during wakefulness; SC increases during sleep are highly likely to occur in either non-rapid eye movement (non-REM) stage 2 sleep or slow-wave sleep (SWS) [25]. The sensor data were logged into the flash memory of the sensors. Participants were instructed to remove sensors only in instances when the sensor could become wet or risked being broken.

Mobile Phone App

We wrote a custom Android phone app based on funf [26] that monitored location, receivers, senders, and timings of calls and SMS text messages, screen on or off timings, and phone app usage. No content of emails, calls, or SMS text messages was recorded. Phone usage was measured for two main reasons. First, phone usage and location data give clues to social interactions. The timing of calls, SMS, and screen on provide an estimate of how often participants interact with their phone during the day and the night, whereas the number of calls, SMS, and the number of people they interact with helps quantify their social interaction. Second, lighting from the interaction with mobile phones or emailing late at night could disturb the biological circadian clock and increase alertness, both of which

can influence sleep patterns [27,28]. We asked our participants not to use third-party messaging apps, if possible, during the study for the last two cohorts.

Twice-Daily Electronic Diaries

Participants completed electronic diaries (e-diaries): upon awakening and at bedtime each day. These diaries contained questions about sleep and wake times; naps; exercise; academic and extracurricular activity times; social interactions; caffeine, alcohol, and drug intake; overall health condition; sleep; mood; and self-reported stress (Figure 1). Participants received emails that included a URL to the morning and evening diaries. They could complete the diaries using computers, tablets, or mobile phones.

Figure 1. An example evening e-diary. For some questions, if yes is chosen, additional questions are presented.

Evening daily diary for 2016-10-06

Did you attend any academic activities today (including classes, e-classes, sections, seminars, labs, study groups)? Yes No

How many?

List each: when the activity started and for how long? [put extra events under 'Comments']

Activity 1 start 09:00 AM [Clear entry](#)

For how long?

How many hours did you study by yourself today, not including any of the academic activities in the previous question?

Did you engage in any exercise-based activities today (including sports, gym, cycling, etc.)? Yes No

Did you attend any other extracurricular activities today, besides academic activities and exercise-based activities? Yes No

For each of the following, indicate how you feel right now by clicking on each line and adjusting the sliders.

Sleepy	<input type="range"/>		Alert
Sad	<input type="range"/>		Happy
Sluggish	<input type="range"/>		Energetic
Sick	<input type="range"/>		Healthy
Stressed Out	<input type="range"/>		Calm Relaxed

In addition, did you have an emotionally charged interaction with someone today? No

Yes, a memorable positive interaction

Yes, a somewhat negative interaction

Yes, and it was very negative

Please enter any comments for today

Save »

Poststudy Questionnaires and Other Measurements

At the end of the month of intensive data collection:

1. Academic performance as measured by grade point average was self-reported by each participant for the semester previous to the study and the current study semester.
2. Email usage during the experiment (ie, to, from, cc, and time stamps) was collected through the Massachusetts Institute of Technology (MIT) website Immersion [29].
3. On the basis of their phone call, SMS, and email usage objectively measured during the experiment, participants were asked to self-report whether they had positive or neutral or negative interactions with each frequent contact as a whole over the month. Participants also indicated to which category each frequent contact belonged to (ie, family, social, work, others).
4. The PSS, the SF-12, the set of social network surveys, and the State-Trait Anxiety Index [30] were completed.

Data Preprocessing

Ambulatory data measured with wearable sensors, mobile phones, and surveys tend to be noisy. Examples include (1) AM vs PM errors when participants complete survey items about their sleep and activity times; (2) participants forgetting to charge or wear sensors; (3) sensors breaking or the signals becoming noisy; and (4) mobile phone connectivity, hardware sensor functionality, and mobile software updates, which can break and interfere with data integrity. To address these issues, various techniques have been applied, such as data cleaning before data analysis [31]; data quality evaluation [32], detecting faulty data, noise reduction [33], and interpolating faulty or missing values [34,35]. To reduce the occurrence or impact of these issues, additional approaches can be used during ambulatory data collection. For example, during the study, an e-diary system can notify participants about potential inaccurate answers before they submit their answers, and a study investigator can check data quality of incoming data and provide feedback to the participants. For this study, we developed tools for improving the quality of the collected data and for supporting more efficient human checking and correcting of the phone, sensor, and e-diary data.

Preprocessing Twice-Daily Electronic Diaries

We collected a total of 6077 days of e-diary data. In the first year of the SNAPSHOT study, we set up an e-diary system that automatically sent surveys to our participants every morning and evening and then sent reminders if the participants did not complete the surveys within 12 hours. We implemented logic check functions on the system that prompted users to revise their answers if certain types of errors or missing answers were detected (eg, if two activity events overlapped, or if their reported wake time was earlier than their reported bedtime). During this first year, study investigators manually checked participants' answers every 1 to 2 days and emailed them to revise their answers when errors were found.

In year 2 of the study, we installed raster plots that visualize participants' activities over time (Figure 2). These raster plots were displayed to participants after they submitted their answers, allowing users to visually confirm their responses and return to

their survey to correct any errors. These raster plots reduced about half of the daily diary errors. The raster plots also reduced the total average time taken to preprocess 1 month of a participant's e-diary data by 53%: from 145 min (year 1) to 68 min (year 2).

Finally, in year 3, we created and installed an administrative module that includes three components to further improve data validity: a calendar view, interactive checking system, and a summary view. Every day, a study investigator logged into the diary system and saw the calendar view (Multimedia Appendix 1) that showed the number of participants in the study, the number of participants whose morning and evening diaries were checked, the number of unchecked diaries, the number of diaries that needed to be rechecked, and participants' comments. The interactive checking system automatically flagged missing answers in the e-diary and allowed the study investigator to check daily diaries just by flagging sections of the e-diary as error (Figure 3). Emails were automatically sent to participants if there were errors or missing answers. The summary view (Multimedia Appendix 2) showed the daily diary status for each participant in different colors (eg, green-acceptable, red-missing, and pink-error). These plots enabled the study investigator to understand which participants had filled out the daily surveys and which participants they needed to contact (eg, if there were repeated errors or missing entries in the diaries). This module further reduced the total average time taken to preprocess 1 month of a participant's e-diary data from 68 min (year 2) to 45 min (year 3). The combined changes in raster plots and the administrative modules reduced the total average time taken to clean 1 month of one participant's e-diary data by 69%: from 145 min (year 1) to 45 min (year 3). Overall, participants' daily diary completion rates ranged between 92% and 97% with no significant differences across semesters.

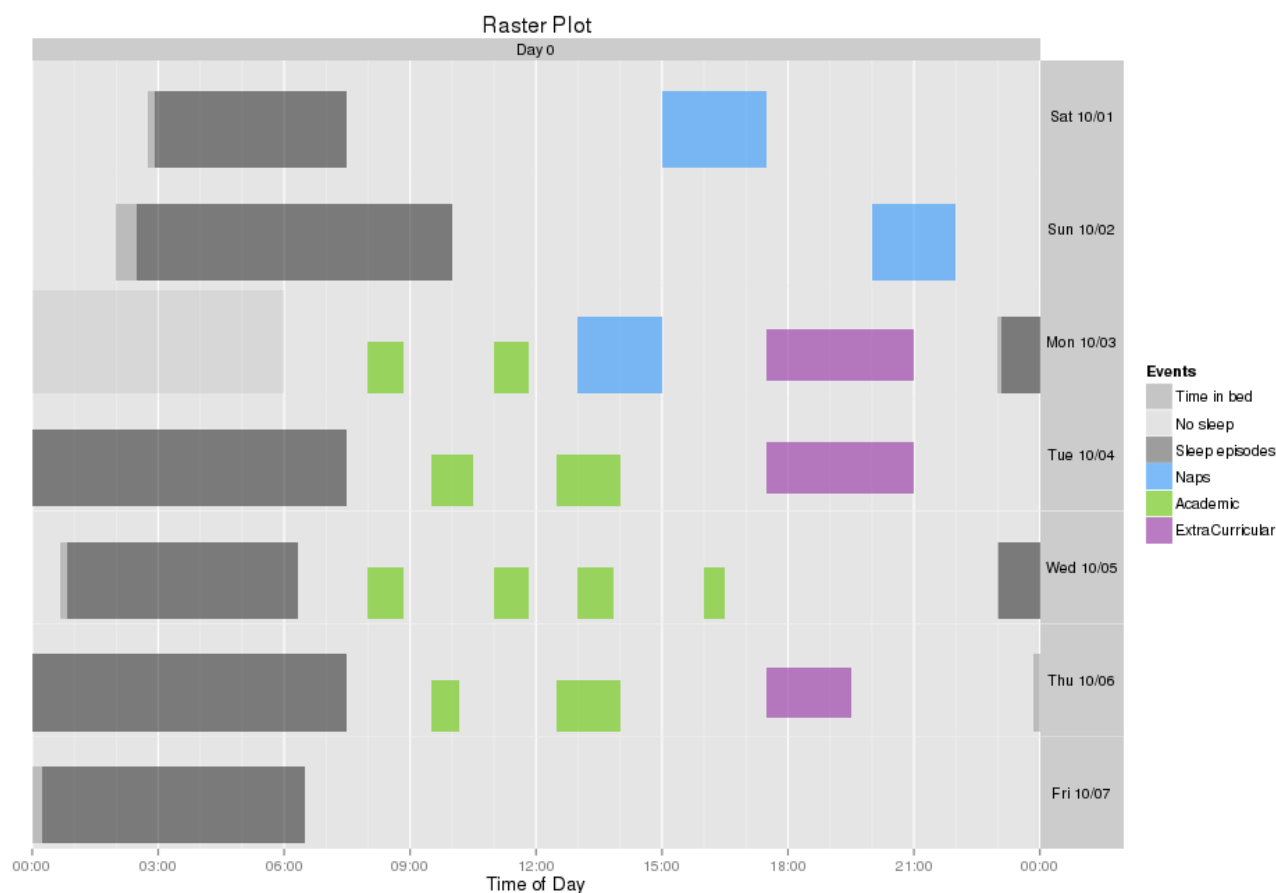
Preprocessing Sensor or Mobile Phone Data

Every week, a study investigator had a face-to-face meeting with each participant to download sensor data and to check if sensors were working correctly, if the participants were wearing them properly, and if sensor electrodes needed replacement. We developed scripts to download the data from sensors and check sensor readings automatically for quality using a previously developed and tested automated classifier [36]. This classifier separated clean epochs and noisy epochs of SC data for further analysis.

We collected 6309 days of Q-sensor data for a total of 125,413 hours. We computed how much data were within a typical range per published guidelines: for SC, 83% were within the range of 0.01 to 30 microS [37-39], and for ST, 99.7% were within the range of 20 to 42 degrees Celsius [40]. In addition, 92% of the collected SC data were classified as clean data using an artifact detection algorithm [36]. Thus, among the collected SC data, 80% of the data were used for further analysis.

Mobile phone data were sent automatically to a server by the custom funf-based app. On the server, another set of scripts that we wrote checked the data quality every day and sent notification to a participant if a problem was found in their data (eg, not receiving phone data for a day). Phone data were collected on 85% of the days.

Figure 2. Plot of daily activity timing (raster plot) with time of day (midnight to midnight) on the y-axis and each day plotted on a separate line. Participants saw this plot after filling out their surveys and before they submitted their answers. Different activities were marked with different colors.



Identifying Risk Factors, Objective Biomarkers, and Modifiable Behavioral Features Related to Stress and Mental Health

We defined high stress and low stress groups based on their poststudy PSS scores (Figure 4). PSS scores range from 0 to 40: higher scores indicate higher perceived stress. A PSS score of 14.2 is the average for the age group of 18 to 29 years, and a score over 16 is considered as high stress and of high health concern [23]. Our participants' average PSS score was 17.1. We used the value of $PSS \geq 16$ to construct the high stress group ($N=109$, top 57.7% [109/189]) and $PSS < 16$ for the low stress group ($N=80$, bottom 42.3% [80/189]). Because we originally had an unbalanced set of data for high stress and for low stress, we first reduced the size of the high stress group by the method of random sampling of its data to equalize the size of the high and low stress classes at $N=80$. Thus, the prior probabilities on both classes were made to be 0.5, so that a random classifier would be expected to attain accuracy of 50%.

We defined high mental health and low mental health groups based on their poststudy MCS from the SF-12 (Figure 4). For the MCS, a value ≥ 50 is considered good mental health [41,42], and 11.8% (23/195) of our population scored ≥ 50 . We therefore extracted the top and bottom 12% to form the two groups: high mental health group ($MCS \geq 50$, top 11.8% [23/195], $N=23$) and low mental health group ($MCS \leq 29.4$, bottom 12.3% [24/195], $N=24$). Thus, the data in the high and low mental health groups

were balanced so that the prior probability of either group would be 0.5, with a random classifier expected to have an accuracy of 50%.

Feature Extraction

To quantify the relative importance of the many measures, we compared the classification performance using the following separate categories of features: (1) Big Five personality + gender, (2) wearable sensors (eg, ST, SC, and ACC), (3) mobile phone (eg, call, SMS, screen on, and location), and (4) objective features (combining wearable sensors and mobile phone metrics). We also separately defined (5) modifiable behaviors as features that can potentially be controlled by participants, such as sleep and activity timing and phone usage; these are important features to measure for future behavioral interventions (Table 1). Note that some features such as phone features and ACC feature are found in more than one of the five categories.

SC was processed first using low-pass filtering (cutoff frequency 0.4 Hz, 32nd order finite impulse response filter). Because there are individual differences in SC amplitude, we extracted features from both unnormalized and normalized SC data based on the maximum and minimum amplitude of each day within each individual. To detect SC peaks, we obtained the first derivative of the low-pass-filtered non-normalized SC data and then determined where the slope exceeded a value of 0.02 μS per second [43]. We detected SC peaks based on those that exceeded this threshold and counted the number of peaks in each 30-second epoch.

Figure 3. Interactive diary check system. The left panel shows a participant’s answers. The right panel shows if there are any detected errors or missing entries and enables adding comments. After the study investigator clicked the Save button, the system sent an email to a participant about any missing or erroneous entries if appropriate.

Did you attend any academic activities today (including classes, e-classes, sections, seminars, labs, study groups)? Yes No

How many?

List each: when the activity started and for how long? [put extra events under 'Comments']

Activity 1 start Clear entry

For how long? hh mm

How many class deadlines did you have today? (Class deadlines include homework deadlines, projects, quiz, mid-terms and finals)

How many hours did you study by yourself today, not including any of the academic activities in the previous question?

Did you engage in any exercise-based activities today (including sports, gym, cycling, etc.)? Yes No

How many times?

List each: when the activity started and for how long [put extra under 'Comments']

Activity 1 start Clear entry

For how long?

Did you attend any other extracurricular activities today, besides academic activities and exercise-based activities? Yes No

How many times?

List each: when the activity started and for how long [put extra under 'Comments']

Activity 1 start Clear entry

For how long?

Did you miss or were you late for any scheduled events (e.g., academic activities, exercise-based activities, other extracurricular activities) because you overslept today? Yes No

When did the event start? Clear entry

How many total servings of caffeine did you have today?

When is the latest hour you consumed caffeine? Clear entry

Please refer to this guide to help calculate servings

Cola 12 oz (1 can)	1/2 serving
Tea (1 cup)	1/2 serving
Home brew coffee 8 oz (1 cup)	1 serving
Energy drink 16 oz (1 large can)	2 servings
5 hour energy	2 servings
Large coffee 16 oz (large or grande)	2 servings
Caffeine pill - 100mg	1 serving
Caffeine pill - 200mg	2 servings

Besides caffeine, did you use any other medications, drugs, or alcohol today? Yes No

For each of the following, indicate how you feel right now by clicking on each line and adjusting the sliders.

Sleepy Alert

Sad Happy

Sluggish Energetic

Sick Healthy

Stressed Out Calm/Relaxed

In addition, did you have an emotionally charged interaction with someone today? No Yes, a memorable positive interaction Yes, a somewhat negative interaction Yes, and it was very negative

Please enter any comments for today

Save »

Academic activity

OK Missing Error

AMPM Wrong Date Overlap

Email content

Exercise

OK Missing Error

AMPM Wrong Date Overlap

Email content

Extracurricular

OK Missing Error

Overslept

OK Missing Error

Caffeine or Drugs

OK Missing Error

Scales

OK Missing Error

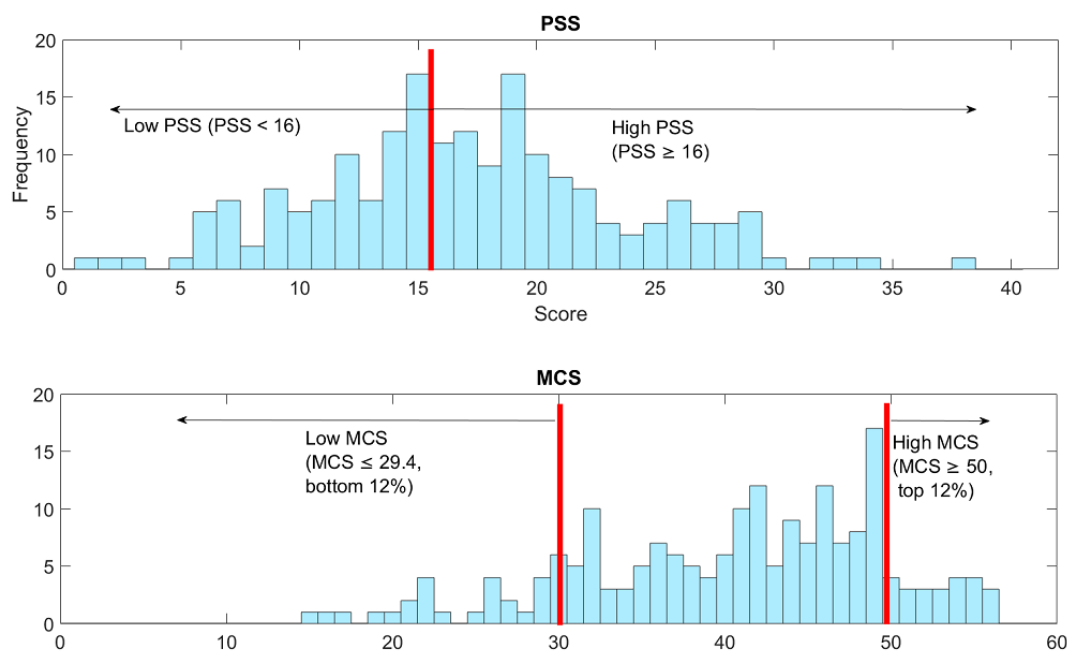
Emotional Interaction

OK Missing Error

Comments

Admin comments (not sent to participant)

« Back
Save »
Skip

Figure 4. (1) Distribution of poststudy Perceived Stress Scale (PSS) and (2) Distribution of poststudy mental component summary (MCS) scores.

We used four different times of interest for analyses: day (9 AM-6 PM), night (6 PM-0 AM), late night (0 AM-3 AM), and sleep time (estimated for each individual from actigraphy and daily sleep diaries) as physiological responses, such as SC and ACC during daytime and sleep time have different meanings [44] and late night phone and exercise activities could relate to self-reported stress and mental health [45].

Bedtime and sleep regularity were calculated from the daily sleep diaries, and sleep duration and sleep efficiency were estimated from actigraphy with help of the daily sleep diaries. Sleep regularity was computed because a relationship between irregular sleep and low mental health was found in a previous study using this index [46]. The Sleep Regularity Index (SRI; Figure 5) captures the probability of an individual being in the same state (asleep vs awake) at any two time points 24 hours apart with 1 minute resolution, averaged across the entire study [47], where $s(t)=1$ during wake and $s(t)=-1$ during sleep for each minute. Assume data are collected for $[0, T]$ with T =total number of hours of data and $\tau=24$ hours.

In practice, individuals will only display sleep patterns that range between an SRI of 0 (random) and 100 (periodic: an individual who sleeps and wakes at exactly the same times each day). Values less than 0 are theoretically possible (eg, alternating 24 h of sleep and 24 h of wake) but very unlikely to be observed.

Phone usage and location data can provide information on sociability. We computed the timing and the number of calls,

SMS, and screen on, which provide an estimate of how often participants interact with their phone during the day and the night. Previous studies showed the relationships between long phone usage duration and high stress [45] and long and frequent phone usage and severe depressive symptoms [8]. We also computed the number of people each participant interacted with over calls and SMS to help quantify their social interaction. For mobility features, we computed the distance and radius based on locations to which our participants travelled as these features were shown to be important in previous studies [8,48]. Additionally, because our population spent most of their time on campus or at their residence, we computed whether the day's mobility pattern varied from the typical routine based on a Gaussian mixture model trained for each participant's 1-month mobility patterns [49].

Classification

For classifying high or low stress groups and high or low mental health groups, we compared the methods of least absolute shrinkage and selection operator (LASSO), support vector machine (SVM) with linear kernel classifier, and SVM with radial basis function (RBF) kernel classifier; these algorithms were used in previous related work [8,10]. LASSO is a logistic regression that performs regularization and feature selection by minimizing the least squares objective function with an L1 penalty [50].

Table 1. List of features.

Modality	Features
All	Personality types, gender, diary, sensor, and phone features
Big Five personality types, gender (6 features)	Openness, conscientiousness, extraversion, agreeableness, neuroticism, gender
Sensors (17 features x 4 time frames x 3=204 features)	Mean, median, SD of 0 AM-3 AM, sleep, 9 AM-6 PM, 6 PM-0 AM for SC ^a , ACC ^b , and ST ^c Skin conductance: Area under the curve for 30 s epochs, max, mean, median, and SD of amplitude; mean, median and SD of peaks for 30 s epochs; mean, median, and SD of normalized amplitude Acceleration: total # of zero crossing for 30 s epochs Skin temperature: max, min, mean, median, and SD of temperature
Phone (25 features (call, SMS ^d , screen) x 3 time frames x 3 + 4 features (mobility) x 3 features=237 features)	Mean, median, SD of 0 AM-24 AM, 0 AM-3 AM, 6 PM-0 AM for call, SMS, and screen (not mobility) Call: Mean, median, and SD of duration and time stamp of calls per day; total duration per day, total number per day, and number of unique people per day SMS: Mean, median, and SD of duration and time stamp of SMS per day; total number per day and number of unique people per day Screen: Mean, median, and SD of screen-on duration and screen-on time stamp per day; total duration per day and total number of on or off per day Mobility: Total distance per day, 5-min distance, radius per day, and log likelihood of each day
Objective (441 features)	Phone and sensor features (see above)
Modifiable behaviors (296 features)	Sleep Regularity Index Mean, median, and SD of bedtime and sleep duration Diary features (see below) ACC total # of zero crossing for 30 s epochs Phone features (see above)
Diary (17 x 3=51 features)	Mean, median, SD of sleep or no sleep (pulled an all-nighter; binary valued), pre sleep electronic media interaction (emails, calls, SMS, Skype, chat, and online games; binary valued), pre sleep personal interaction(binary valued), # of naps, nap duration, # of academic activities per day, total academic duration, study duration, # of extracurricular activities, total extracurricular activities, # of exercise, exercise duration, # of caffeinated drink intake, memorable positive interaction(binary valued), somewhat negative interaction (binary valued), very negative interaction(binary valued), last caffeine intake time
Sleep (1 + 3 x 8=25 features)	Sleep Regularity Index Mean, median, and SD of bedtime, sleep duration, sleep efficiency, sleep or no sleep (pulled an all-nighter; binary valued), pre sleep electronic media interaction (emails, calls, SMS, Skype, chat, and online games; binary valued), pre sleep personal interaction (binary valued), # of naps and nap duration

^aSC: skin conductance.

^bACC: acceleration.

^cST: skin temperature.

^dSMS: short message service.

Figure 5. Equation of Sleep Regularity Index.

$$Sleep\ Regularity\ Index = 200 \times \left(\frac{\int_0^{T-\tau} s(t)s(t+\tau) dt}{T-\tau} - 0.5 \right)$$

For training and testing models, we used nested-cross validation. To evaluate model performance, we applied leave-one-cohort-out: training a model with all except one semester cohort's data and testing the model against the left-out

cohort's data, repeating this process for the total number of cohorts (ie, 5 times). First we (1) split the data into two datasets: a training set made up of four cohorts and a test set made up of

one cohort. We then left the test set out until step (5) or (8) below.

For training the SVM models, we applied sequential forward feature selection to the training data to reduce overfitting and find the best combinations. (2) We applied a *t* test to each feature of the training datasets and selected 100 features with the lowest *P* values for finding features to separate two groups effectively then (3) applied sequential forward feature selection [51]: applied an SVM RBF classifier with 10-fold cross validation to find the best up to five combinations from these 100 features and optimized hyperparameters (*C* for SVM linear and *C* and *gamma* for SVM RBF). Then, (4) we trained the SVM linear or RBF models with the selected features of the training data, (5) tested the models against the test data, and (6) repeated this process (1-6) five times.

For LASSO, (7) the penalization parameter was determined with the training data by 10-fold cross validation and (8) the trained model was tested using the test data. This process (1, 7, and 8) was repeated five times.

We computed overall accuracy and F1 scores by concatenating the five-cohort predicted output to compare the performance of the models to reduce the bias from splitting [52]. We computed 95% confidence levels using adjusted Wald test [53]. The F1 score is a measure of performance computed using precision (also known as positive predictive value) and recall (also known as sensitivity) as described in Equation 1, where precision is the number of correct positive results divided by the number of all positive results, and recall is the number of correct positive results divided by the number of positive results that should have been returned.

$$(1) F1 = 2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$$

We also compared the performance of the models using features based on data from the entire 1-month study period with that using features based only on using the data from the week before the PSS and MCS surveys were completed.

We applied *t* tests or Mann-Whitney *U* tests (for non-Gaussian distributions) to examine if the means of the features were statistically different between the high or low PSS groups or the high or low MCS groups. We adjusted for the multiple comparisons using false discovery rate (FDR).

Results

Relationships Among Prestudy and Poststudy Perceived Stress Scores and Mental Component Summary

There were no differences in the poststudy PSS or MCS among the five cohorts; one-way analysis of variance ($P=.20$, $F=1.50$). Students' poststudy scores (both PSS and MCS) were highly correlated with prestudy scores ($r=.59$, $.60$, Pearson correlation). Poststudy PSS scores statistically increased (mean prestudy PSS: 15.0, poststudy PSS: 17.1, paired *t* test, $P<.001$) and MCS scores decreased compared with the prestudy scores (mean prestudy MCS: 44.4, poststudy MCS: 40.4, Wilcoxon signed rank test, $P<.001$). Thus, the students reported worsening stress and mental health over the 1 month of measurement.

The poststudy PSS was inversely correlated with the poststudy MCS ($r=-.71$, Pearson correlation; [Multimedia Appendix 3](#)): (1) 83% (19/23) of the students in the high MCS group belonged to the low PSS group and (2) 88% (21/24) of the students in the low MCS group belonged to the high PSS group. The low MCS group had higher PSS scores than the rest of the students in the high PSS group: low MCS group's average PSS score was 25.2, whereas the rest in the high PSS group's average PSS score was 20.7 ($P<.001$).

Stress and Mental Health Classification

Overall, we found SVM models with the RBF kernel worked better than LASSO and linear SVM models using RBF kernels for all of the metrics ([Figures 6 and 7](#); see [Multimedia Appendix 4](#) for accuracy and F1 scores and [Multimedia Appendices 5 and 6](#) for F1 scores for all results). SVM with the RBF kernel can model more complex decision boundaries. Sensor features showed higher performance than phone features both for PSS and MCS.

We also compared the performance of the SVM RBF models using features from only the last week of the 1-month period to using the features from the entire month. Overall, the performances with the 1 month of features were better (classification accuracy improved by 1-16%) than those using just the last week of features, except in the case of the SVM models using all features.

The accuracy for PSS classification was highest when using all features (82%), followed by when using features from only sensors (78%), only behaviors (74%), only the Big Five (71%), or only objective data (70%). The same rank ordering also held when comparing F1 scores. For MCS, sensor features and objective features showed the highest accuracy (87%), followed by Big Five (85%), behaviors (79%), and all (77%). The ranking of the F1 scores was similar except for all features had a slightly higher F1 than behaviors. The means and SD of the accuracy and F1 scores from leave-one-cohort-out cross validation are presented in [Multimedia Appendix 7](#).

We also tested different cutoffs: (1) instead of PSS cutoffs ≥ 16 for high and <16 for low stress, we used PSS ≥ 14 for high stress group and PSS <14 for low stress, as (as noted above) 14.2 is the reported average for people aged 18 to 29 years [23] and (2) instead of extreme MCS cutoffs (top and bottom 12%), we used MCS \geq median (42.05) for high mental health group and MCS $<$ median for low mental health group). This was done to test if the rankings of performances were sensitive to the exact cutoff values. Sensor and modifiable behavior features worked best with both cutoff values ([Multimedia Appendices 8 and 9](#)). Compared with the extreme MCS cutoffs, the median cutoff showed much lower classification performance (the accuracy decreased by 21 to 6 %).

We summarize the features most commonly selected by the algorithms as useful for high or low PSS detection ([Figure 8](#)) and high or low MCS detection ([Figure 9](#)) using the full 1 month of data. Percentages indicate the percent time these features are selected across 10-fold cross validation over five cohorts and five feature modalities (all, Big Five + gender, sensor, phone, objective and modifiable behavior features).

For PSS classification for self-reported stress, neuroticism and conscientiousness were the most often selected features (90% and 70% of the models, respectively). The high PSS group had higher neuroticism (q [which is the FDR-adjusted P value]=.0004). The high stress group had a larger extracurricular activity duration SD ($q=.04$).

In the MCS classification for self-reported mental health, the low MCS group showed higher neuroticism ($q<.001$) and lower conscientiousness ($q=.04$) than the high MCS group. The low MCS group had naps more frequently (40%; $q=.04$). In the MCS classification models using only the last week of data, the low MCS group showed a lower probability of interacting with electronic media (eg, emails, calls, SMS, Skype, chat, and online games) before sleep (30%; $q=.004$) and lower SD of the number of SC peaks during the time frame of 0 AM to 3 AM (20%; $q=.03$), as well as higher neuroticism ($q<.001$).

The percentages of time each feature was selected for each fold of leave-one-cohort cross validation are presented in [Multimedia Appendices 10 and 11](#).

We also tried building models only with sleep features (eg, features in the sleep category and some sleep related features in the survey category). We obtained 72% and 65% accuracy for classifying high or low PSS and high or low MCS. Mean nap duration was the most common feature used for the PSS models (80% of the models), followed by median bed time and the frequency of pulling all-nighters (60%). The frequency of pulling all-nighters (100% of the models), mean number of naps, sleep duration, and sleep efficiency (60%) were commonly selected features by the MCS classification models. Average sleep duration was not significantly different statistically in the high vs low PSS groups or in the high vs low MCS groups (high PSS: 6 hours 42 min vs low PSS: 6 hours 51 min [$P=.09$], high MCS: 6 hours 40 min, low MCS: 6 hours 34 min [$P=.72$]). Instead, the low MCS group's more frequent napping was one of the most discriminating features.

Figure 6. High or low Perceived Stress Scale (PSS) classification results. Top: comparison of performance using 1 month of data with three machine learning algorithms. Bottom: comparison of performance using 1 month of data vs only the last week of data with support vector machine radial basis function (SVM RBF). Accuracy scores for Big Five + Gender data are not shown in the bottom graph because these data are collected only once. Error bars indicate the 95% CIs based on adjusted Wald test.

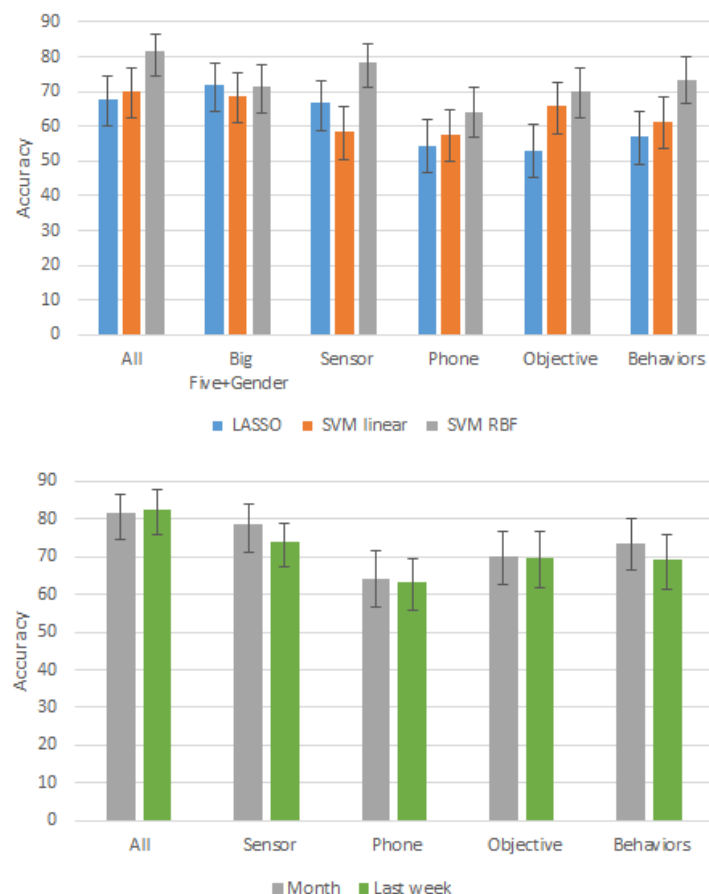


Figure 7. As in Figure 6 with high or low mental component summary score classification results, accuracy scores for Big Five + Gender data are not shown in the bottom graph because these data are collected only once. Error bars indicate the 95% CIs based on adjusted Wald test.

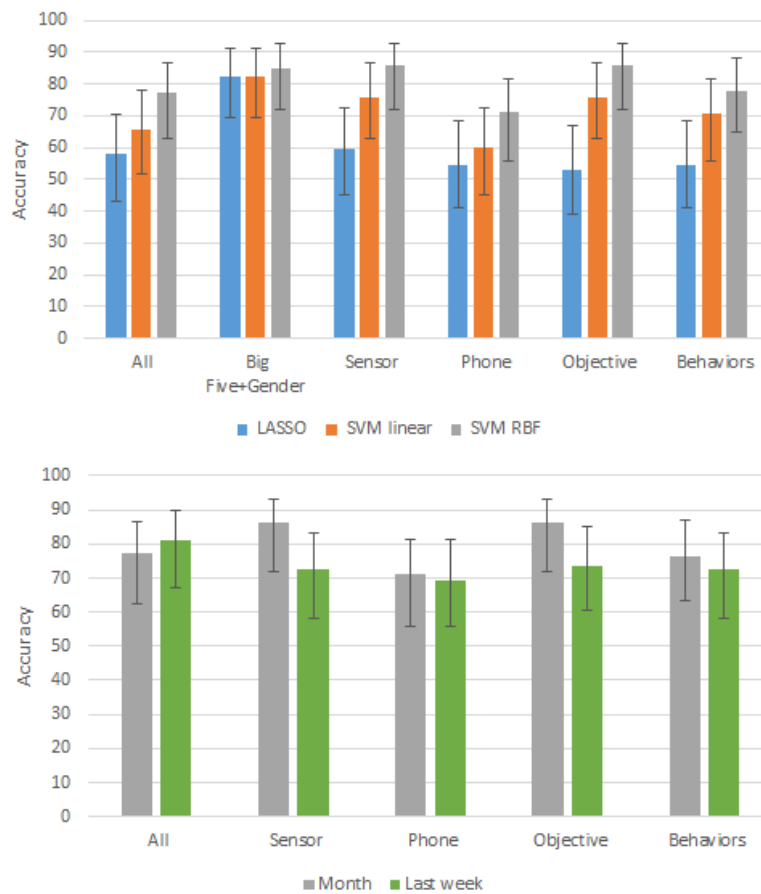


Figure 8. Percentage of time each feature was selected across 10-cross-validation for high or low Perceived Stress Scale (PSS) classification models with 1 month of data.

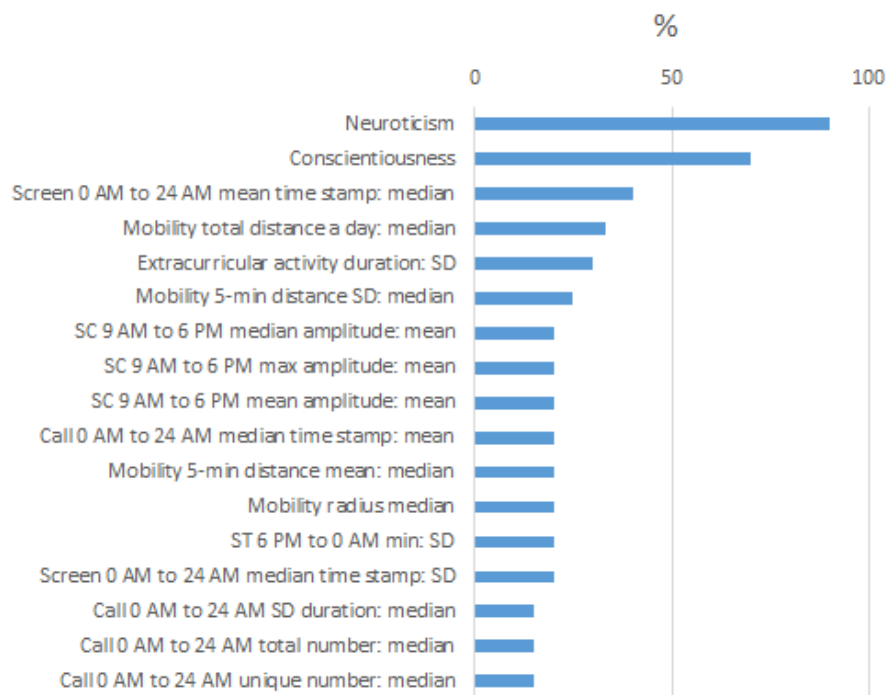
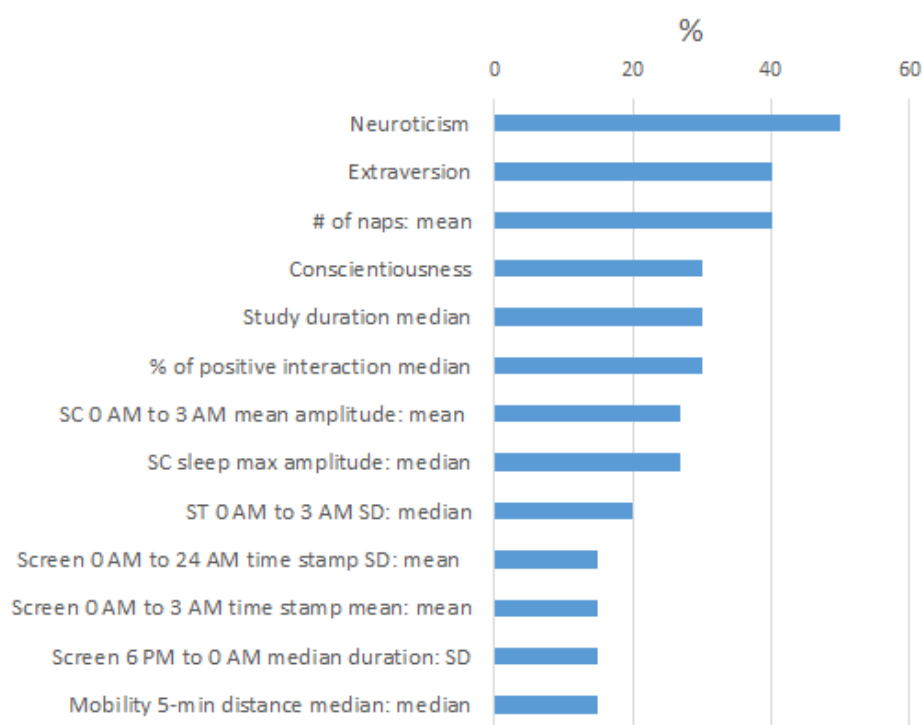


Figure 9. Percentage of times each feature was selected across 10-cross-validation for high or low mental component summary (MCS) classification models with 1 month of data.



Discussion

Principal Findings

In this paper, we developed novel tools to collect and process objective physiological and behavioral measures using online diaries, wearable sensors, and mobile phones. We aimed to investigate how accurately these measures could identify conditions of self-reported high stress and poor mental health and features most accurate in identifying these conditions.

Physiological sensor, phone, and mobility features were the best predictors for distinguishing self-reported high or low stress and mental health. Wearable sensor features, including SC and ST, reached 79% accuracy for classifying high or low stress groups and 87% accuracy for classifying high or low mental health groups. Modifiable behaviors, including number of naps, studying duration, phone calls (number, time stamp and duration of calls), mobility patterns, and phone-screen-on time, reached 74% accuracy for high or low stress group classification and 78% accuracy for high or low mental health group classification.

Comparison With Prior Work and Interpretations of Our Results

Our analysis showed that relatively high accuracy and F1 scores can be achieved using the leave-one-semester-cohort-out testing of the machine learning classifier for high or low stress measured by PSS and high or low mental health measured by MCS. Of all the features tested, the sensor features resulted in approximately 14% higher classification accuracies in both PSS and MCS than the phone features. In particular, SC responses during the time frame of 9 AM to 6 PM were one of the best predictors for PSS. SC has been considered as a biomarker for stress [44] because SC quantifies eccrine sweat activity that is

controlled by only sympathetic nervous activity. These findings (1) are among the first to show the potential contribution of SC in stress detection using a wrist wearable sensor in a 24/7 daily life setting and (2) agree with previous findings that use a conventional finger SC sensor or a wearable SC sensor in settings where a person is seated, eg, driving a car. For example, Healey et al measured SC, heart rate, HRV, respiration, and electromyogram in Boston drivers and reported that SC was the most associated with stress [54]. Additionally, Hernandez et al discriminated stressful and nonstressful calls at a call center environment using SC features with 78% accuracy [55], and Setz et al automatically classified SC responses from cognitive load and stress with accuracy higher than 80% [56].

As we examined more closely which sensor features were most discriminative, we found that SC responses during the time frame of 0 AM to 3 AM and during sleep were predictors for separating high and low self-reported mental health. Some studies have shown that finger-based SC are reduced for patients with depression measured in a short-term lab study [57-59]. One possible explanation of how low SC responses during sleep could be related to MCS scores is that there is a decrease in SWS in depression [60] and other psychiatric disorders [61], and the largest SC responses during sleep are likely to occur during non-REM stage 2 and SWS [25]. Note that in our data, (1) 0 AM to 3 AM could include both awake and asleep conditions, and if it included sleep, we would expect it to include more SWS being at the start of the night for this cohort; (2) our low mental score groups are based on self-report; and (3) we do not know if any of our participants had clinically defined depression or other psychiatric disorders as that information was not gathered as part of this study.

We found that ST features were also predictors for PSS and MCS. A previous study has shown that acute stress does reduce distal finger ST but does not statistically significantly reduce wrist ST in a laboratory stress test setting [62]. Furthermore, another study showed that ST is one of the strongest discriminants to distinguish sleep and wake states [63]. Another study showed that patients with depression have less rhythmicity in ST [64], which would also be consistent with less regular sleep in depression. To our knowledge, this paper is the first to report that ambulatory wrist ST features are related to self-reported stress.

For phone features, our results showed phone usage time stamp and duration can be predictors for PSS and MCS. These results are consistent with several previous studies. People with a PHQ-9 score higher than 5 showed longer phone usage and higher phone usage frequency than those with a PHQ-9 score lower than 5 in a 2-week study with mobile phones [8]. A questionnaire-based study also showed a relationship between high mobile phone usage, stress, and symptoms of depression [45].

Mobility, specifically travel distance per day and SD of the distance traveled as measured by phone geolocation data, is a predictor both for self-reported stress and mental health. This result agrees with one study [8] reporting that normalized mobility entropy (distribution of frequency of visiting different places) and location variance were negatively correlated with depression symptoms and another study reporting that mobility patterns were highly related to stress level [13]. The relationship between reduced activity levels and mobility patterns and high stress and low mental health has been studied [48,65]. These behavioral markers could be an objective index for monitoring self-reported low mental health. It is possible that encouraging people to move more could be an effective intervention to reduce stress and improve mental health.

Consistent with previous studies [7,13,14], personality types were one of the most influential and statistically significant factors for self-reported stress and mental health in this college population. In the Big Five Inventory Personality Test categories, neuroticism was a predictor of stress [66]. The combination of low extraversion and low conscientiousness or low agreeableness contributed to the high stress group; these directions of the associations in our analysis were consistent with prior work [67]. High neuroticism and low extraversion have previously been associated with low MCS [68].

There is a known association between sleep deficiency and mental health status (eg, [61]). Our results, however, did not show that sleep duration was a strong discriminant feature for self-reported stress and mental health.

Limitations

There are multiple limitations of this study:

1. Selection of a feature as discriminating between two categories does not mean it is an important feature or causative of that behavior.
2. These results do not tell us the causality (eg, does a student sleep later and less regularly because of higher stress or have higher stress because of later or more irregular sleep?).

3. Our participants were limited to Android phone users because we wanted to log detailed phone usage, which is not allowed by other phone systems such as iPhone. As about half of the undergraduate students were Android users on the campus, a selection bias might exist. A previous study showed slight differences in personality types and economic status between iPhone users and Android users [69].
4. A total of 64% of our study population were male participants. It has been reported that females report higher perceived stress levels and more depressive symptoms [70-73], and there are gender differences in psychological and biological stress responses [74]. In our dataset, the ratios of female participants in the high or low PSS and MCS groups were 45% and 20% (high and low PSS) and 22% and 54% (high and low MCS). Modeling stress and mental health differently in males and females might help understand the mechanism. Gender was included as a potential feature in our models and was not selected frequently.
5. Our data come from college students at one New England university over 4 years. The work needs to be applied to other populations to determine generalizability.
6. Our data come from socially connected student groups. We might observe some statistically coherent behaviors in our dataset because of these connections.

Future Work

These new tools and methods can allow multimodal data in daily life to be captured more continuously, with greater accuracy and integrity of the data, and for long-term and at great scale. We are planning to collect a larger amount of data for an even longer time to study long-term behaviors and physiological responses and build predictive models. To do this, we need to build a new system for consenting people in remote locations, fully automate checking their measurement status and data accuracy automatically, and let the participants know about errors so they can fix them to keep study compliance and data accuracy high.

We will continue our data analysis for understanding behaviors, physiological responses, and traits that impact health and well-being. One of our hypotheses is that health-related behaviors will be contagious within social networks and that social network data we obtained from call, SMS, and email data could capture the social contagion quantitatively instead of requiring self-report to capture it. We are also interested in studying how phone usage influences sleep and health and how we can predict stress and mental health using previous behaviors and physiology.

These machine learning models are not limited to modalities and features we measured and computed in this study but can also be used for other modalities such as heart rate and heart rate variability that are controlled by autonomous activities, and other features such as app usage, ambient light, and audio or sentiment-based patterns extracted from text or speech could be added to improve the models. The features and models presented in this paper can be tested in similar multimodal ambulatory datasets collected in other future studies. Tracking

stress and mental health conditions would help students better understand their stress and mental health conditions over multiple semesters, as well as help clinicians see how treatment affects students' conditions if they receive treatment.

Conclusions

In this paper, we introduced a methodology and tools we developed to measure ambulatory multimodal data and improve the integrity of collected data to study self-reported stress and mental health in the daily lives of college students. We showed that objective and modifiable behavioral features collected over

1 month can classify these college students as high or low stress based on the PSS and as having high or low mental health based on MCS from SF-12 collected at the end of that month with over 70% accuracy, whereas sensor features alone could classify high or low mental health and achieve over 88% on an F1 score. For classifying high or low stress groups, we found that combining phone and sensor features typically gave the best results over using either modality alone, whereas for classifying high or low mental health groups, the use of wearable sensor features performed comparable to wearable + phone features.

Acknowledgments

The authors are grateful to Mr Conor O'Brien, Mr Justin Buie, Mr Salim Qadri, Ms Natalie Virayan, Mr Michael Shreeve, Mr Omer Zaidi, Ms Natasha Jaques, Mr Weixuan Chen, Ms Asma Ghandeharioun, Mr Daniel Lopez Martinez, Ms Ehimwenma Nosakhare, Ms Amy Zhao Yu, Mr Daniel Smilkov, Ms Jade Philipoom, Ms Yuna Hahn, Ms Sienna Ramos, Ms Jihyun Gia Min, Ms Tania Yu, Ms Shirley Chen, Ms Laura Breiman, and Dr Catherine Ricciardi for their tremendous support helping run the SNAPSHOT study and collecting the data, and to Dr Cesar Hidalgo for helping design social network surveys. They also appreciate all the participants and support from the MIT Media Lab Consortium, especially a generous donation by Samsung Electronics, NEC, and funding from NIH grants R01GM105018, R00HL119618 (AJKP), K24HL105664 (EBK), and KL2TR002370, F32DK107146, T32HL007901 (AWM) and Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health Award UL1 TR001102), and financial contributions from Harvard University and its affiliated academic health care centers. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic health care centers, or the National Institutes of Health.

Conflicts of Interest

RP is a cofounder of and shareholder in Affectiva, who commercialized the original sensors used in this study. RP is also a cofounder and shareholder in Empatica, a company that makes wearable sensors that can collect ambulatory data similar to the data collected in this study. EK has consulted for legal firms and for Pfizer Pharmaceuticals.

Multimedia Appendix 1

Calendar view. For each day, the investigator can see how many participants are in the study, how many surveys have been verified, how many need to be re-examined, and participant comments.

[\[PNG File, 102KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Summary view. The investigator can see the daily diary status for each participant in different colors: green: acceptable; red: missing; pink: error; red/white slash: missing but participants still have time to complete the diary.

[\[PNG File, 100KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

The relationship between poststudy Perceived Stress Scale (PSS) and poststudy Mental Component Score (MCS) for each participant (blue circle). Chosen cutoffs for high and low PSS and high and low MCS are indicated.

[\[PNG File, 57KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Performance of PSS and MCS classification models with 1 month or last week of data.

[\[PDF File \(Adobe PDF File\), 28KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

High or low Perceived Stress Scale (PSS) classification results. Top: Comparison of F1 scores for PSS classification with three machine learning algorithms (LASSO, SVM linear, and SVM RBF) on 1 month of different types of data (All, Big Five + Gender,

Sensor, Phone, Objective, Behaviors). Bottom: Comparison of F1 scores with SVM RBF machine learning algorithm on 1 month of data versus on only the last week of the same data types. Accuracy scores for Big Five + Gender data are not shown in the bottom graph because these data are collected only once.

[\[PNG File, 21KB-Multimedia Appendix 5\]](#)

Multimedia Appendix 6

High or low Mental Component Score (MCS) classification results. Top: Comparison of F1 scores for MCS classification with three machine learning algorithms (LASSO, SVM linear, and SVM RBF) on 1 month of different types of data (All, Big Five + Gender, Sensor, Phone, Objective, Behaviors). Bottom: Comparison of F1 scores with SVM RBF machine learning algorithm on 1 month of data versus on only the last week of the same data types. Accuracy scores for Big Five + Gender data are not shown in the bottom graph because these data are collected only once.

[\[PNG File, 22KB-Multimedia Appendix 6\]](#)

Multimedia Appendix 7

Mean and SD of accuracy and F1 scores from leave-one-cohort-out PSS and MCS classification models with 1 month of data and SVM RBF.

[\[PDF File \(Adobe PDF File\), 26KB-Multimedia Appendix 7\]](#)

Multimedia Appendix 8

High or low Perceived Stress Scale (PSS) and Mental Component Score (MCS) classification results. Comparison of F1 for PSS and MCS classification scores with SVM RBF and 1 month of different types of data (All, Big Five + Gender, Sensor, Phone, Objective, Behaviors). Cutoff: 14 (the average in the 18-29 years age group) for PSS and 42.05 (median) for MCS.

[\[PNG File, 14KB-Multimedia Appendix 8\]](#)

Multimedia Appendix 9

Performance of PSS and MCS classification models with 1 month of data and SVM RBF. PSS cutoff: 14 (the average in the 18-29 years age group) and MCS cutoff: 42.05 (median).

[\[PDF File \(Adobe PDF File\), 16KB-Multimedia Appendix 9\]](#)

Multimedia Appendix 10

Percentages of the number of times each feature was selected for each fold of leave-one-cohort-out cross validation for 1-month PSS models.

[\[PDF File \(Adobe PDF File\), 20KB-Multimedia Appendix 10\]](#)

Multimedia Appendix 11

Percentages of the number of times each feature was selected for each fold of leave-one-cohort-out cross validation for 1-month MCS models.

[\[PDF File \(Adobe PDF File\), 19KB-Multimedia Appendix 11\]](#)

References

1. Torous J, Kiang MV, Lorme J, Onnela JP. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Ment Health* 2016 May 5;3(2):e16 [[FREE Full text](#)] [doi: [10.2196/mental.5165](https://doi.org/10.2196/mental.5165)] [Medline: [27150677](https://pubmed.ncbi.nlm.nih.gov/27150677/)]
2. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. *Nat Biotechnol* 2015 May;33(5):462-463. [doi: [10.1038/nbt.3223](https://doi.org/10.1038/nbt.3223)] [Medline: [25965751](https://pubmed.ncbi.nlm.nih.gov/25965751/)]
3. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schüssler-Fiorenza Rose SM, et al. Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. *PLoS Biol* 2017 Jan;15(1):e2001402 [[FREE Full text](#)] [doi: [10.1371/journal.pbio.2001402](https://doi.org/10.1371/journal.pbio.2001402)] [Medline: [28081144](https://pubmed.ncbi.nlm.nih.gov/28081144/)]
4. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. 2014 Presented at: UbiComp '14 Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 13-17, 2014; Seattle, Washington. [doi: [10.1145/2632048.2632054](https://doi.org/10.1145/2632048.2632054)]

5. Bogomolov A, Lepri B, Pianesi F. Happiness recognition from mobile phone data. 2013 Presented at: International Conference on Social Computing; September 8-14, 2013; Alexandria, VA, USA. [doi: [10.1109/SocialCom.2013.118](https://doi.org/10.1109/SocialCom.2013.118)]
6. Bauer G, Lukowicz P. Can smartphones detect stress-related changes in the behaviour of individuals? 2012 Presented at: IEEE International Conference on Pervasive Computing and Communications Workshops; March 19-23, 2012; Lugano, Switzerland. [doi: [10.1109/PerComW.2012.6197525](https://doi.org/10.1109/PerComW.2012.6197525)]
7. Bogomolov A, Lepri B, Ferron M, Pianesi F, Pentland A. Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits. 2014 Presented at: Proceedings of the 22nd ACM international conference on Multimedia; November 3-7, 2014; Orlando, Florida, USA. [doi: [10.1145/2647868.2654933](https://doi.org/10.1145/2647868.2654933)]
8. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *J Med Internet Res* 2015 Jul 15;17(7):e175 [FREE Full text] [doi: [10.2196/jmir.4273](https://doi.org/10.2196/jmir.4273)] [Medline: [26180009](https://pubmed.ncbi.nlm.nih.gov/26180009/)]
9. Chow PI, Fua K, Huang Y, Bonelli W, Xiong H, Barnes LE, et al. Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. *J Med Internet Res* 2017 Mar 3;19(3):e62 [FREE Full text] [doi: [10.2196/jmir.6820](https://doi.org/10.2196/jmir.6820)] [Medline: [28258049](https://pubmed.ncbi.nlm.nih.gov/28258049/)]
10. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. *J Med Internet Res* 2017 Mar 16;19(3):e75 [FREE Full text] [doi: [10.2196/jmir.6678](https://doi.org/10.2196/jmir.6678)] [Medline: [28302595](https://pubmed.ncbi.nlm.nih.gov/28302595/)]
11. Muaremi A, Arnrich B, Tröster G. Towards measuring stress with smartphones and wearable devices during workday and sleep. *Bionanoscience* 2013;3:172-183 [FREE Full text] [doi: [10.1007/s12668-013-0089-2](https://doi.org/10.1007/s12668-013-0089-2)] [Medline: [25530929](https://pubmed.ncbi.nlm.nih.gov/25530929/)]
12. Muaremi A, Bexheti A, Gravenhorst F, Arnrich B, Tröster G. Monitoring the Impact of Stress on the Sleep Patterns of Pilgrims using Wearable Sensors. 2014 Presented at: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); June 1-4, 2014; Valencia, Spain. [doi: [10.1109/BHI.2014.6864335](https://doi.org/10.1109/BHI.2014.6864335)]
13. Sano A, Picard RW. Stress Recognition Using Wearable Sensors and Mobile Phones. 2013 Presented at: Humaine Association Conference on Affective Computing and Intelligent Interaction; September 2-5, 2013; Geneva, Switzerland. [doi: [10.1109/ACII.2013.117](https://doi.org/10.1109/ACII.2013.117)]
14. Sano A, Phillips AJ, Yu AZ, McHill AW, Taylor S, Jaques N, et al. Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. 2015 Presented at: IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN); June 9-12, 2015; Cambridge, MA, USA. [doi: [10.1109/BSN.2015.7299420](https://doi.org/10.1109/BSN.2015.7299420)]
15. CACUSS. American College Health Association-National College Health Assessment II: Canadian Reference Group Executive Summary Spring 2017 URL: http://www.cacuss.ca/health_data.htm [accessed 2018-05-16] [WebCite Cache ID [6zT9bD7yo](https://www.webcitation.org/6zT9bD7yo)]
16. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization world mental health surveys. *Psychol Med* 2016 Oct;46(14):2955-2970 [FREE Full text] [doi: [10.1017/S00332971716001665](https://doi.org/10.1017/S00332971716001665)] [Medline: [27484622](https://pubmed.ncbi.nlm.nih.gov/27484622/)]
17. Eisenberg D, Golberstein E, Hunt J. Mental health and academic success in college. *B E J Econom Anal Policy* 2009;9(1):- [doi: [10.2202/1935-1682.2191](https://doi.org/10.2202/1935-1682.2191)]
18. The American Association of Suicidology. Suicidology. 2016. College Students & Suicide Fact Sheet: 2016 Fact Sheet URL: <http://www.suicidology.org/Portals/14/Re-Formatted%20College%20Students%20Fact%20Sheet.pdf?ver=2016-11-16-110354-547> [accessed 2018-05-16] [WebCite Cache ID [6zTBIkcVH](https://www.webcitation.org/6zTBIkcVH)]
19. The Association for University and College Counseling Center Directors. The association for university and college counseling center directors annual survey URL: http://files.cmcglobal.com/AUCCCD_2013_Monograph_Public.pdf [accessed 2018-05-16] [WebCite Cache ID [6zT9rC5tt](https://www.webcitation.org/6zT9rC5tt)]
20. Horne JA, Ostberg O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Int J Chronobiol* 1976;4(2):97-110. [Medline: [1027738](https://pubmed.ncbi.nlm.nih.gov/1027738/)]
21. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989 May;28(2):193-213. [Medline: [2748771](https://pubmed.ncbi.nlm.nih.gov/2748771/)]
22. John O, Srivastava S. Moityca.com.br. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives URL: http://moityca.com.br/pdfs/bigfive_John.pdf [accessed 2018-05-18] [WebCite Cache ID [6zVEjcmQv](https://www.webcitation.org/6zVEjcmQv)]
23. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983 Dec;24(4):385-396. [Medline: [6668417](https://pubmed.ncbi.nlm.nih.gov/6668417/)]
24. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996 Mar;34(3):220-233. [Medline: [8628042](https://pubmed.ncbi.nlm.nih.gov/8628042/)]
25. Sano A, Picard RW, Stickgold R. Quantitative analysis of wrist electrodermal activity during sleep. *Int J Psychophysiol* 2014 Dec;94(3):382-389 [FREE Full text] [doi: [10.1016/j.ijpsycho.2014.09.011](https://doi.org/10.1016/j.ijpsycho.2014.09.011)] [Medline: [25286449](https://pubmed.ncbi.nlm.nih.gov/25286449/)]
26. Aharony N, Pan W, Ip C, Khayal I, Pentland A. Social fMRI: investigating and shaping social mechanisms in the real world. *Pervasive Mob Comput* 2011 Dec;7(6):643-659. [doi: [10.1016/j.pmcj.2011.09.004](https://doi.org/10.1016/j.pmcj.2011.09.004)]

27. Cajochen C, Frey S, Anders D, Späti J, Bues M, Pross A, et al. Evening exposure to a light-emitting diodes (LED)-backlit computer screen affects circadian physiology and cognitive performance. *J Appl Physiol* (1985) 2011 May;110(5):1432-1438 [FREE Full text] [doi: [10.1152/jappphysiol.00165.2011](https://doi.org/10.1152/jappphysiol.00165.2011)] [Medline: [21415172](https://pubmed.ncbi.nlm.nih.gov/21415172/)]
28. Chang AM, Aeschbach D, Duffy JF, Czeisler CA. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. *Proc Natl Acad Sci U S A* 2015 Jan 27;112(4):1232-1237 [FREE Full text] [doi: [10.1073/pnas.1418490112](https://doi.org/10.1073/pnas.1418490112)] [Medline: [25535358](https://pubmed.ncbi.nlm.nih.gov/25535358/)]
29. Immersion. A people-centric view of your email life URL: <https://immersion.media.mit.edu/> [accessed 2018-05-18] [WebCite Cache ID 6zVK7Rb2d]
30. Spielberger C. Manual for the State-Trait Anxiety Inventory (Form Y). Palo Alto, CA: Consulting Psychologists Press; 1983.
31. Van den Broeck J, Cunningham SA, Eeckels R, Herbst K. Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS Med* 2005 Oct;2(10):e267 [FREE Full text] [doi: [10.1371/journal.pmed.0020267](https://doi.org/10.1371/journal.pmed.0020267)] [Medline: [16138788](https://pubmed.ncbi.nlm.nih.gov/16138788/)]
32. Clifford GD, Lopez D, Li Q, Rezek I. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. 2011 Presented at: Computing in Cardiology; September 18-21, 2011; Hangzhou, China.
33. Wood LB, Asada HH. Noise cancellation model validation for reduced motion artifact wearable PPG sensors using MEMS accelerometers. 2006 Presented at: International Conference of the IEEE Engineering in Medicine and Biology Society; August 30-September 03, 2006; New York, NY, USA. [doi: [10.1109/IEMBS.2006.260359](https://doi.org/10.1109/IEMBS.2006.260359)]
34. Engels J. Imputation of missing longitudinal data: a comparison of methods. *J Clin Epidemiol* 2003 Oct;56(10):968-976. [doi: [10.1016/S0895-4356\(03\)00170-7](https://doi.org/10.1016/S0895-4356(03)00170-7)]
35. Schmitt P, Mandel J, Guedj M. A comparison of six methods for missing data imputation. *J Biom Biostat* 2015;6:224. [doi: [10.4172/2155-6180.1000224](https://doi.org/10.4172/2155-6180.1000224)]
36. Taylor S, Jaques N, Chen W, Fedor S, Sano A, Picard R. Automatic identification of artifacts in electrodermal activity data. 2015 Presented at: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 25-29, 2015; Milan, Italy URL: <http://europepmc.org/abstract/MED/26736662> [doi: [10.1109/EMBC.2015.7318762](https://doi.org/10.1109/EMBC.2015.7318762)]
37. Doberenz S, Roth WT, Wollburg E, Maslowski NI, Kim S. Methodological considerations in ambulatory skin conductance monitoring. *Int J Psychophysiol* 2011;80(2):87-95. [doi: [10.1016/j.ijpsycho.2011.02.002](https://doi.org/10.1016/j.ijpsycho.2011.02.002)]
38. Dawson ME, Schell AM, Filion DL. The electrodermal system. In: *Handbook of Psychophysiology*. New York, NY, US: Cambridge University Press; 2007:159-181.
39. Wenzel A. *The SAGE Encyclopedia of Abnormal and Clinical Psychology*. Thousand Oaks, California: SAGE Publications, Inc; 2017.
40. Texas Instruments. TIDA-00824 Human Skin Temperature Sensing for Wearable Applications Reference Design URL: <http://www.ti.com/lit/ug/tiduay7/tiduay7.pdf> [accessed 2018-05-16] [WebCite Cache ID 6zTANV7XJ]
41. Bolge SC, Flores NM, Phan JH. The burden of poor mental well-being among patients with type 2 diabetes mellitus: examining health care resource use and work productivity loss. *J Occup Environ Med* 2016 Nov;58(11):1121-1126 [FREE Full text] [Medline: [27820762](https://pubmed.ncbi.nlm.nih.gov/27820762/)]
42. Wolff JL, Roter DL. Older adults' mental health function and patient-centered care: does the presence of a family companion help or hinder communication? *J Gen Intern Med* 2012 Jun;27(6):661-668 [FREE Full text] [doi: [10.1007/s11606-011-1957-5](https://doi.org/10.1007/s11606-011-1957-5)] [Medline: [22180197](https://pubmed.ncbi.nlm.nih.gov/22180197/)]
43. Boucsein W, Fowles DC, Grimnes S, Ben-Shakhar G, Roth WT, Dawson ME, Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures. Publication recommendations for electrodermal measurements. *Psychophysiology* 2012 Aug;49(8):1017-1034. [doi: [10.1111/j.1469-8986.2012.01384.x](https://doi.org/10.1111/j.1469-8986.2012.01384.x)] [Medline: [22680988](https://pubmed.ncbi.nlm.nih.gov/22680988/)]
44. Boucsein W. *Electrodermal Activity*. United States: Springer US; 2012.
45. Thomée S, Härenstam A, Hagberg M. Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults--a prospective cohort study. *BMC Public Health* 2011 Jan;11:66 [FREE Full text] [doi: [10.1186/1471-2458-11-66](https://doi.org/10.1186/1471-2458-11-66)] [Medline: [21281471](https://pubmed.ncbi.nlm.nih.gov/21281471/)]
46. Sano A, Phillips A, McHill A, Taylor S, Barger L, Czeisler C, et al. Influence of weekly sleep regularity on self-reported wellbeing. *Sleep* 2017;40(suppl_1):A67-A68. [doi: [10.1093/sleepj/zsx050.181](https://doi.org/10.1093/sleepj/zsx050.181)]
47. Phillips AJK, Clerx WM, O'Brien CS, Sano A, Barger LK, Picard RW, et al. Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. *Sci Rep* 2017 Jun 12;7(1):3216 [FREE Full text] [doi: [10.1038/s41598-017-03171-4](https://doi.org/10.1038/s41598-017-03171-4)] [Medline: [28607474](https://pubmed.ncbi.nlm.nih.gov/28607474/)]
48. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. 2015 Presented at: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7-11, 2015; Osaka, Japan. [doi: [10.1145/2750858.2805845](https://doi.org/10.1145/2750858.2805845)]
49. Jaques N, Taylor S, Azaria A, Ghandeharioun A, Sano A, Picard R. Predicting students' happiness from physiology, phone, mobility, and behavioral data. 2015 Presented at: International Conference on Affective Computing and Intelligent Interaction (ACII); September 21-24, 2015; Xi'an, China URL: <http://europepmc.org/abstract/MED/28515966> [doi: [10.1109/ACII.2015.7344575](https://doi.org/10.1109/ACII.2015.7344575)]
50. Tibshirani R. Regression selection shrinkage via the Lasso. *J R Stat Soc Series B Stat Methodol* 2011;73(3):273-282. [doi: [10.1111/j.1467-9868.2011.00771.x](https://doi.org/10.1111/j.1467-9868.2011.00771.x)]

51. Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell* 1997 Dec;97(1-2):273-324. [doi: [10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)] [Medline: [356583](https://pubmed.ncbi.nlm.nih.gov/356583/)]
52. Forman G, Scholz M. Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. *SIGKDD Explor* 2010;12:49-57. [doi: [10.1145/1882471.1882479](https://doi.org/10.1145/1882471.1882479)]
53. Agresti A, Coull BA. Approximate is better than “Exact” for interval estimation of binomial proportions. *Am Stat* 1998 May;52(2):119-126. [doi: [10.1080/00031305.1998.10480550](https://doi.org/10.1080/00031305.1998.10480550)]
54. Healey JA, Picard RW. Affect.media.mit.edu. 2005. Detecting Stress During Real-World Driving Tasks Using Physiological Sensors URL: <https://affect.media.mit.edu/pdfs/05.healey-picard.pdf> [accessed 2018-05-18] [WebCite Cache ID 6zVI8MWT1]
55. Hernandez J, Morris RR, Picard RW. Call Center Stress Recognition with Person-Specific Models. 2011 Presented at: International Conference on Affective Computing and Intelligent Interaction; October 09-12, 2011; Memphis, USA. [doi: [10.1007/978-3-642-24600-5_16](https://doi.org/10.1007/978-3-642-24600-5_16)]
56. Setz C, Arnrich B, Schumm J, La Marca R, Tröster G, Ehlert U. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Trans Inf Technol Biomed* 2010 Mar;14(2):410-417. [doi: [10.1109/TITB.2009.2036164](https://doi.org/10.1109/TITB.2009.2036164)] [Medline: [19906598](https://pubmed.ncbi.nlm.nih.gov/19906598/)]
57. Ward NG, Doerr HO, Storrie MC. Skin conductance: a potentially sensitive test for depression. *Psychiatry Res* 1983 Dec;10(4):295-302. [Medline: [6583718](https://pubmed.ncbi.nlm.nih.gov/6583718/)]
58. Argyle N. Skin conductance levels in panic disorder and depression. *J Nerv Ment Dis* 1991 Sep;179(9):563-566. [Medline: [1919559](https://pubmed.ncbi.nlm.nih.gov/1919559/)]
59. Rottenberg J. Mood and emotion in major depression. *Curr Dir Psychol Sci* 2005;14(3):167-170. [doi: [10.1111/j.0963-7214.2005.00354.x](https://doi.org/10.1111/j.0963-7214.2005.00354.x)]
60. Nutt D, Wilson S, Paterson L. Sleep disorders as core symptoms of depression. *Dialogues Clin Neurosci* 2008;10(3):329-336 [FREE Full text] [Medline: [18979946](https://pubmed.ncbi.nlm.nih.gov/18979946/)]
61. Abad VC, Guilleminault C. Sleep and psychiatry. *Dialogues Clin Neurosci* 2005;7(4):291-303 [FREE Full text] [Medline: [16416705](https://pubmed.ncbi.nlm.nih.gov/16416705/)]
62. Vinkers CH, Penning R, Hellhammer J, Verster JC, Klaessens JH, Olivier B, et al. The effect of stress on core and peripheral body temperature in humans. *Stress* 2013 Sep;16(5):520-530. [doi: [10.3109/10253890.2013.807243](https://doi.org/10.3109/10253890.2013.807243)] [Medline: [23790072](https://pubmed.ncbi.nlm.nih.gov/23790072/)]
63. Sano A, Picard RW. Comparison of sleep-wake classification using electroencephalogram and wrist-worn multi-modal sensor data. *Conf Proc IEEE Eng Med Biol Soc* 2014;2014:930-933 [FREE Full text] [doi: [10.1109/EMBC.2014.6943744](https://doi.org/10.1109/EMBC.2014.6943744)]
64. Barbini B, Benedetti F, Colombo C, Guglielmo E, Campori E, Smeraldi E. Perceived mood and skin body temperature rhythm in depression. *Eur Arch Psychiatry Clin Neurosci* 1998;248(3):157-160. [Medline: [9728735](https://pubmed.ncbi.nlm.nih.gov/9728735/)]
65. Zschucke E, Gaudlitz K, Ströhle A. Exercise and physical activity in mental disorders: clinical and experimental evidence. *J Prev Med Public Health* 2013 Jan;46 Suppl 1:S12-S21 [FREE Full text] [doi: [10.3961/jpmph.2013.46.S.S12](https://doi.org/10.3961/jpmph.2013.46.S.S12)] [Medline: [23412549](https://pubmed.ncbi.nlm.nih.gov/23412549/)]
66. Vollrath M. Personality and stress. *Scand J Psychol* 2001;42(4):335-347. [Medline: [11547909](https://pubmed.ncbi.nlm.nih.gov/11547909/)]
67. Ebstrup JF, Eplöv LF, Pisinger C, Jørgensen T. Association between the Five Factor personality traits and perceived stress: is the effect mediated by general self-efficacy? *Anxiety Stress Coping* 2011 Jul;24(4):407-419. [doi: [10.1080/10615806.2010.540012](https://doi.org/10.1080/10615806.2010.540012)] [Medline: [21213153](https://pubmed.ncbi.nlm.nih.gov/21213153/)]
68. van Straten A, Cuijpers P, van Zuuren FJ, Smits N, Donker M. Personality traits and health-related quality of life in patients with mood and anxiety disorders. *Qual Life Res* 2007 Feb;16(1):1-8 [FREE Full text] [doi: [10.1007/s11136-006-9124-x](https://doi.org/10.1007/s11136-006-9124-x)] [Medline: [17033892](https://pubmed.ncbi.nlm.nih.gov/17033892/)]
69. Götz FM, Stieger S, Reips UD. Users of the main smartphone operating systems (iOS, Android) differ only little in personality. *PLoS One* 2017 May;12(5):e0176921 [FREE Full text] [doi: [10.1371/journal.pone.0176921](https://doi.org/10.1371/journal.pone.0176921)] [Medline: [28467473](https://pubmed.ncbi.nlm.nih.gov/28467473/)]
70. Mayor E. Gender roles and traits in stress and health. *Front Psychol* 2015 Jun;6:779 [FREE Full text] [doi: [10.3389/fpsyg.2015.00779](https://doi.org/10.3389/fpsyg.2015.00779)] [Medline: [26106354](https://pubmed.ncbi.nlm.nih.gov/26106354/)]
71. Zuckerman DM. Stress, self-esteem, and mental health: How does gender make a difference? *Sex Roles* 1989 Apr;20(7-8):429-444. [doi: [10.1007/BF00288001](https://doi.org/10.1007/BF00288001)]
72. Sigmon ST, Pells JJ, Boulard NE, Whitcomb-Smith S, Edenfield TM, Hermann BA, et al. Gender differences in self-reports of depression: the response bias hypothesis revisited. *Sex Roles* 2005;53(5-6):401-411. [doi: [10.1007/s11199-005-6762-3](https://doi.org/10.1007/s11199-005-6762-3)]
73. Sandanger I, Nygård JF, Sørensen T, Moum T. Is women's mental health more susceptible than men's to the influence of surrounding stress? *Soc Psychiatry Psychiatr Epidemiol* 2004 Mar;39(3):177-184. [doi: [10.1007/s00127-004-0728-6](https://doi.org/10.1007/s00127-004-0728-6)] [Medline: [14999449](https://pubmed.ncbi.nlm.nih.gov/14999449/)]
74. Verma R, Balhara YP, Gupta CS. Gender differences in stress response: role of developmental and biological determinants. *Ind Psychiatry J* 2011 Jan;20(1):4-10 [FREE Full text] [doi: [10.4103/0972-6748.98407](https://doi.org/10.4103/0972-6748.98407)] [Medline: [22969173](https://pubmed.ncbi.nlm.nih.gov/22969173/)]

Abbreviations

ACC: acceleration

E-diary: electronic diary

FDR: false discovery rate
HRV: heart rate variability
LASSO: least absolute shrinkage and selection operator
MCS: mental component summary
MIT: Massachusetts Institute of Technology
PHQ-9: patient health questionnaire-9
PSS: perceived stress scores
RBF: radial basis function
REM: rapid eye movement
SC: skin conductance
SF-12: 12-Item Short Form Health Survey
SMS: short message service
SRI: Sleep Regularity Index
ST: skin temperature
SVM: support vector machine
SWS: slow-wave sleep

Edited by G Eysenbach; submitted 13.11.17; peer-reviewed by S Saeb, L Barnes, C Gorrostieta, J Apolinário-Hagen; comments to author 16.12.17; revised version received 24.02.18; accepted 22.04.18; published 08.06.18

Please cite as:

Sano A, Taylor S, McHill AW, Phillips AJK, Barger LK, Klerman E, Picard R
Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones: Observational Study
J Med Internet Res 2018;20(6):e210
URL: <http://www.jmir.org/2018/6/e210/>
doi: [10.2196/jmir.9410](https://doi.org/10.2196/jmir.9410)
PMID: [29884610](https://pubmed.ncbi.nlm.nih.gov/29884610/)

©Akane Sano, Sara Taylor, Andrew W McHill, Andrew JK Phillips, Laura K Barger, Elizabeth Klerman, Rosalind Picard. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 08.06.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.