

Recognizing Academic Performance, Sleep Quality, Stress Level, and Mental Health using Personality Traits, Wearable Sensors and Mobile Phones

Akane Sano^{*1}, Andrew J. Phillips^{*2}, Amy Z. Yu^{*1}, Andrew W. McHill^{*2}, Sara Taylor^{*1}, Natasha Jaques^{*1}, Charles A. Czeisler^{*2}, Elizabeth B. Klerman^{*2}, Rosalind W. Picard^{*1}

^{*1} Media Lab

Massachusetts Institute of Technology
Cambridge, USA

{akanes, ayu, sataylor, jaquesn, picard}@media.mit.edu

^{*2} Division of Sleep and Circadian Disorders,

Departments of Medicine and Neurology,
Brigham and Women's Hospital

Division of Sleep Medicine, Harvard Medical School,
Boston, USA

{ajphillips, amchill}@partners.org,
ebklerman@hms.harvard.edu,

caczeisler@research.bwh.harvard.edu

Abstract— What can wearable sensors and usage of smart phones tell us about academic performance, self-reported sleep quality, stress and mental health condition? To answer this question, we collected extensive subjective and objective data using mobile phones, surveys, and wearable sensors worn day and night from 66 participants, for 30 days each, totaling 1,980 days of data. We analyzed daily and monthly behavioral and physiological patterns and identified factors that affect academic performance (GPA), Pittsburg Sleep Quality Index (PSQI) score, perceived stress scale (PSS), and mental health composite score (MCS) from SF-12, using these month-long data. We also examined how accurately the collected data classified the participants into groups of high/low GPA, good/poor sleep quality, high/low self-reported stress, high/low MCS using feature selection and machine learning techniques. We found associations among PSQI, PSS, MCS, and GPA and personality types. Classification accuracies using the objective data from wearable sensors and mobile phones ranged from 67-92%.

Keywords— *mobile phone; smart phone; wearable sensor; sleep; stress; mental health; personality; big five personality; phone call; SMS; skin conductance; accelerometer; machine learning*

I. INTRODUCTION

Many mobile and wearable devices have been recently commercialized to capture 24/7 behavior. Wearable sensors now provide estimates of the number of steps taken, physical activity levels, rest/activity patterns and physiological outcomes such as heart rate and skin conductance. Mobile phones can measure location, distance traveled, social interactions (phone call and short message service: SMS), application usage, and acceleration and light levels. Researchers have used wearable sensors and/or mobile phone data to understand factors such as personality type [1], mood [2, 3], sleep [3, 4, 5] and self-reported stress [6, 7]. In previous work, we collected 5 days of data (wearable sensor, mobile phone and surveys) from 18 participants and were able to

classify them into high and low perceived stress groups using machine learning [6]. We also found features in mobile phone usage, wearable sensor and survey data that were significantly related to perceived stress level using correlation analysis. In the current study, we have increased our sampling period to ~30-days per person and our population to 66 participants to collect more intensive multi-modal data including perceived stress, sleep, personality, physiological, behavioral and social interaction data that are important factors in academic performance, sleep, stress, and mental health in addition to what were monitored on the phone in previous studies [4, 5, 8, 9, 10].

Academic performance, sleep quality, stress, and mental health can depend on internal and external factors such as personality traits, physiology, behavior and social interaction. While previous studies have examined multiple days and nights (from one week to a couple months) of daily behaviors including sleep, activity and social interactions, [4, 5, 8, 9, 10], few studies have examined capturing both internal and external factors to understand which factors are related and how, and fewer have used objective sensor data. Understanding these associations can be used to design tools to reduce stress, improve academic performance, sleep quality, and mental health. For example, using data accumulated on the phone and the model developed from the data from both the entire set of users as well as for each individual, personalized feedback can be provided to users reporting high stress, together with insights into what behaviors are likely to be related to their high stress, and which behaviors they might change to reduce it.

The interactions between academic performance, sleep quality, self-reported stress, self-reported mental health and personality categories have been previously characterized using self-reported data. Specifically, academic performance

has been correlated with personal traits (conscientiousness, openness and agreeableness) [11, 12] and sleep parameters have been reported to be influenced by personality traits of neuroticism [13], extraversion [14], and agreeableness [15]. Vollrath summarized the relationship between stress and the Big Five Inventory Personality Test categories and identified that neuroticism was a predictor of stress [16].

We are particularly interested in examining the association between behavioral factors and health, since behaviors may be more modifiable than personality. Daily technology use may influence sleep and perceived stress; for example, increased mobile phone use is related to poor sleep quality and perceived stress [17, 18, 19]. Social interactions also play important roles in health: Christakis et al. investigated how health-related behaviors such as obesity and happiness spread over social networks and found that happy and unhappy people are clustered in social networks and happiness spread up to three degrees of separation over social networks [20, 21]. Moturu et al. investigated the association among sociability, sleep quality and mood with self-reported surveys and mobile phone proximity data from 54 participants for one month [3]: They found lower sociability was related to poorer mood and higher median daily sociability occurred when people slept for 7-8 hours.

In this paper, we collected 30-days of multi-modal data from undergraduate students who are socially connected, and aim to identify (1) which factors characterizing individuals in daily life separate high/low GPA groups, self-reported good/poor sleepers (from PSQI score), self-reported high/low stress groups (from PSS score) and self-reported mentally healthy/unhealthy groups (from MCS score); (2) how accurately we can classify the groups from the data; and (3) how the data collected objectively with mobile phones and wearable sensors compare to the data collected subjectively by questionnaires or surveys.

II. DATA COLLECTION

Sixty-six undergraduate students participated in a 30-day experiment (47 males, 19 females, average age=20.1 \pm 1.5, mean \pm SD) providing 1,980 days of data. Participants were recruited through email. Prior to the experiment, participants filled out the Pittsburgh Sleep Quality Index (PSQI) [22], the Big Five Inventory Personality Test [23], the Horne-Ostberg Morningness-Eveningness Questionnaire (MEQ) [24], the Perceived Stress Scale (PSS) [25] and the SF-12 Physical and Mental Health Composite Scale (MCS for mental health) [26]. During the 30-day experiment, participants wore a wrist sensor on their dominant hand (Q-sensor, Affectiva, USA) to measure three-axis accelerometer data (ACC), skin temperature (ST) and skin conductance (SC, a measure of sympathetic nervous system activity) at 8 Hz and a wrist actigraphy monitor on their non-dominant hand (MotionLogger, AMI, USA) to measure activity and light exposure levels. Participants also installed an Android phone application adapted by the first author from the funf open

source framework [27] to measure call, SMS, location, internet usage and “screen on” timing. During the study period, they filled out surveys every morning and evening about academic, extracurricular, and exercise activities, sleep, caffeinated drink intake, social interaction, and self-reported general health, mood, alertness, tiredness and stress level. At the end of the study, they filled out the PSS, State-Trait Anxiety Index [28] and SF-12. Grade point average (GPA) was reported by the participants at the end of the semester in which the experiment occurred. Email usage during the experiment (to, from, cc and timestamps) was collected through the MIT website *Immersion* at the end of the study [29]. In addition, based on their call, SMS and email usage during the experiment, participants were asked to characterize the interactions with their frequent contacts. The Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects pre-approved this study and all participants gave informed consent.

Sleep/wake onsets were determined by a combination of wrist actigraphy and sleep diaries. We computed sleep regularity as a value of 0 - 1 based on the likelihood of sleep/wake state being the same time-points 24 hours apart, because sleep researchers have pointed out the importance of this sleep regularity measure [30] in addition to sleep duration.

We collected phone and email usage for two main reasons: First, 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 [31, 32]. Second, phone and email usage and location data give clues to sociability. The timing of calls, SMS, emails and “screen on” provide an estimate of how often participants interact with their phone during the day and the night, while the number of calls, SMS and emails and the number of people they interact with helps quantify their social interaction. For phone and SMS features, we computed the entropy p as

$$p = - \sum_n \left(\frac{n}{\sum n} * \log_2 \left(\frac{n}{\sum n} \right) \right), \quad (1)$$

where n is the frequency of phone calls or messages for each telephone number

Skin conductance (SC) was measured because it represents autonomic arousal during the day and provides a stress index; its responses during sleep are highly likely to occur in either non-REM Stage 2 sleep or Slow Wave Sleep, and help to characterize sleep better than using only acceleration data from actigraphy [33]. Skin temperature also helps to understand sleep/wake patterns [34], while acceleration helps show activity and sleep patterns. We hypothesize that physiology combined with daily behavior data can be used to predict aspects of sleep behaviors, academic performance, and self-reported stress and mental health better than any of these measures alone.

III. ANALYSIS

A. Feature Extraction

Table 1 shows 700 features extracted from the collected data. For the classification of high/low GPA, PSQI, PSS, and mental health condition (MCS), we used features computed from month-long data. SC was processed first by low-pass filtering (cutoff frequency 0.4 Hz, 32nd order FIR filter). Since there are individual differences in SC amplitude, we 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 exceeds a value of 0.02 μ S per second. We detected SC “peaks” based on those that exceeded this threshold and counted the number of peaks per each 30-second epoch. The SC peaks during sleep provide an index of deeper sleep stages (SWS and NREM2) [33]. For ACC, we computed the mean activity level based on the root square values of the 3-axis accelerometer. For ACC data in wakefulness, we separated the data into sit, walk and run episodes based on thresholds we computed with another set of ACC data from 48 people who did sitting, walking and running with the same sensor on their non-dominant wrist. We used these thresholds to compute SC features for sit, walk, and run episodes in order to separate SC responses into psychological and activity-related ones. In our previous paper [6], we only extracted mean, median and SD of the whole day’s ACC and SC data and lost rich information behind the data. In this paper, to capture subtle changes in ACC, SC and TEMP patterns, we added more detailed features from histograms and power spectrum density data of each signal.

B. Classification of long-term profile using one month data

From GPA, PSQI, PSS and MCS from SF-12, we defined the top 20% and the bottom 20% of the participants as high/low GPA groups, good/poor sleepers, high/low stress groups and mentally healthy/unhealthy groups. The distribution of the scores is illustrated in Figure 1. GPA is the weighted average of grades by credits of classes in the semester of the study and ranges from 2.0(D) to 5.0(A). PSQI is a standardized survey to evaluate sleep patterns and quality, and differentiates good and poor sleep quality by measuring subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction over the last month with 18 questions. A PSQI score above 5 was considered as poor sleep quality [22]. PSS is a score from 0 to 40 and higher scores mean high perceived stress. MCS is computed using the scores of twelve questions (SF-12) and ranges from 0 to 100, where a zero score indicates the lowest level and 100 indicates the highest level of mental health measured by the scales [26].

TABLE I. COMPUTED FEATURES

Pre Experiment Survey (15 features)	PSS score (about the past month), regular bedtime, wakeup time, duration, PSQI score, MEQ, State and Trait Anxiety Score, Big Five Test (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism), physical and mental health composite scores (PCS and MCS) from SF-12
Morning survey (17 features)	Sleep time, wake time, sleep latency, sleep regularity, social interactions before sleep (with person in person or through electronic devices), how they wake up (alarm or spontaneously), the number of awakenings, duration of awakenings, the number of naps, duration of naps, alertness, happiness, sluggishness, healthiness and calmness when wake up
Evening survey (12 features)	Total hours of academic, exercise, and extracurricular activities, the number of cups of caffeinated drinks, existence of memorable positive and negative and very negative social interactions, alertness, happiness, sluggishness healthiness and calmness before sleep
Post-Experiment Survey (23 features)	PSS score, PCS and MCS from SF-12, social interactions about the past one month (# of the top 20 people to interact through face to face, email, SMS and phone, total # of people with positive, neutral and negative interactions, # of family members, friends, work-related colleagues each participant interacted frequently in the past one month)
Wearable sensor (SC) (297 features)	Mean, median, SD, frequency (11 bins) and power spectral density (5 bins) of normalized EDA for day time, sit, walk run, entire sleep and 1-4 quarters of sleep Frequency (11 bins), percentage of SC peaks, and entropies of # of SC peaks and SC storms, for day time, sit, walk, run entire sleep, and 1-4 quarters of sleep
Wearable sensor (ACC) (174 features)	Mean % of sit, walk and run activities per day, mean, median and SD of RMS values, frequency (11 bins) and power spectral density (5 bins) of RMS values for day time, sit, walk, run, entire sleep, and 1-4 quarters of sleep, mean objective sleep quality from actigraphy
Wearable sensor (ST) (114 features)	Mean, median, SD, frequency (11 bins) and power spectral density (5 bins) of normalized skin temperature for day time, entire sleep and 1-4 quarters of sleep
Phone (CALL) (12 features)	Time of each call, duration for each call, total # of people and entropy, for each of: all calls, incoming and outgoing calls
Phone (SMS) (12 features)	Time of each SMS message, total # of SMS messages, total # of people and entropy for each of: all, received and sent SMS
Phone (Screen on/off) (10 features)	Time of each screen on/off, total # of screen on/off, total duration of screen on between 0-3am, 3-6am, 6-9am, 9am-12pm, 12-3pm, 3-6pm, 6-9pm and 9pm-0am
Phone (MOB: mobility) (2 features)	Total distance per day and standard deviation of the distance
Phone (Internet) (2 features)	Time, total duration
Email (10 features)	Total # of sent emails, mean and SD of # of daily received/sent emails, # of people to send emails, mean and SD of timestamps of received and sent emails

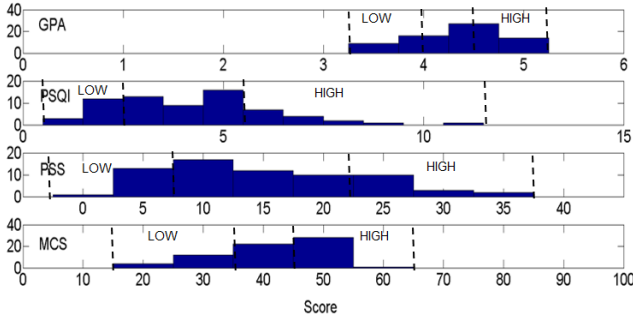


Fig. 1. Distribution of GPA, PSQI, PSS, and MCS from 66 participants

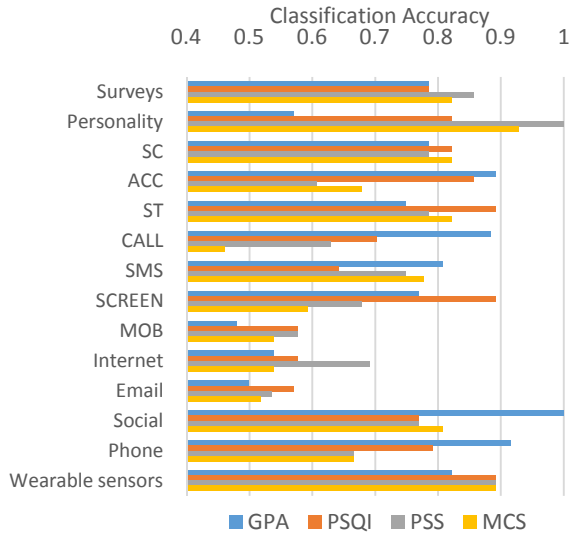


Fig.2. Classification accuracies for GPA, PSQI, PSS and MCS

We first applied sequential forward feature selection to find the best combinations of 1- 3 features for each modality: 1) Surveys (all the pre-, morning and evening surveys that weren't also part of what we were predicting, were used except for personality and social interaction questions, which are used separately below; for PSS classification, we also excluded pre and post study MCS since some of the questions were similar about stress; for MCS classification, we also excluded pre- and post-PSS because of the similar questions) 2) Personality (scores from Big Five Test) 3) SC 4) ACC 5) ST 6) CALL 7) SMS 8) SCREEN 9) MOB 10) Internet 11) Email 12) Social (features from social interaction questions in post-study surveys). We then compared how these features performed using two different classifiers: Support vector machine (SVM) with a linear kernel and SVM with a radial basis function kernel.

For each classification, we examined the accuracy using leave-one-participant-out approach. We selected features and trained models from all except one participant data and tested the model against the left-out participant's data. This procedure was repeated for 26-28 participants (40 % of the entire participants).

IV. RESULTS AND DISCUSSIONS

Figure 2 shows classification accuracies from each modality, taken across the runs of the SVM-L or SVM-RBF, whichever was best. Table 2 shows the top performing 1-3 selected features, used to obtain the results in Fig 2. Of the items measured by surveys, higher GPA was related to longer total hours of academic activities and lower PSQI or earlier wake time. PSQI and agreeableness we found related to academic performance in our data are the factors previously reported [11, 12]. Good/poor PSQI groups were best classified using PSS, daily subjective measures of healthiness and sleep regularity [30] which is a relatively new measure that is not captured by PSQI's metrics. Poor sleepers showed high PSS, low healthiness and low sleep regularity. The relationship between poor sleep quality and high stress level has been previously reported so this finding is a replication in a new group of college students [35, 36]. High PSS groups showed subjective low happiness, high sluggishness and high sickness. The mentally healthy group showed high happiness and low stress level.

Surveys and personality types contributed more to accurately classifying groups of high/low PSS and high/low MCS than they did to high/low GPA and high/low PSQI. Higher neuroticism is a common factor associated with worse scores on PSQI, PSS and MCS. Using personality information alone was more than 80% accurate at classifying high/low PSQI, PSS, and MCS.

Wearable features yielded higher classification accuracies for high/low PSS, PSQI and MCS groups than phone features. For predicting high/low GPA, ACC solo features from the wearable showed higher accuracies than all features from the wearable; however, the combination of features from the phone boosted the overall high/low GPA classification accuracy. The high GPA group showed earlier mean call timestamps and low entropy.

We found some interesting results related to late night phone usage. The poor sleep quality group showed longer 3-6am screen on duration and shorter 9am-12pm duration. The high stress group and the mentally unhealthy group used SMS less frequently but showed later mean timestamp.

Social features during the daytime and right before sleep played also important roles. For example, it is surprising that the high GPA group showed higher number of negative email contacts and negative interactions. The high stress and mentally unhealthy group showed higher frequency of interaction with person before sleep.

Future work will further analyze the data in order to interpret the physiological or behavioral meaning as to why specific features from wearable sensors and mobile phones were selected by the algorithms and what they imply in terms of recommending healthy behavioral choices. Wearable sensors capture mainly how we move, sleep and exercise, while phones capture mainly how we communicate with others. We found that detailed SC, ST and ACC features contributed more accuracy than mobile phone features toward classifying the high/low PSS, PSQI and MCS groups.

Complex physiological and behavioral patterns are embedded in daily behaviors and feature selection and machine learning techniques unveiled subtle differences in the groups we tried to classify. Future work will delve more deeply into what these features suggest in terms of making healthier daily decisions. This work provides a significant first step showing that objective wearable and mobile data carry information that might be used by individuals in order to make better predictions about the impact of behavioral choices on GPA, sleep, stress, and mental health.

ACKNOWLEDGMENT

We are grateful to Mr. Conor O'Brien for his tremendous support helping run the College Sleep Project and collecting the data, and to Dr. Cesar Hidalgo for helping design social network surveys. We also appreciate all the participants and support from the MIT Media Lab Consortium especially a generous donation by Samsung and funding from NIH grants R01GM105018, K99HL119618 (AJP), K24HL105664 (EBK), and T32HL007901(AWM).

REFERENCES

- [1] Y. A. de Montjoye, J. Quoidbach, F. Robic, and A. S. Pentland, "Predicting personality using novel mobile phone-based metrics," *Proc. 6th Int. Conf. Soc. Comput. Behav. Model. Predict.*, 2013.
- [2] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, "MoodScope," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services - MobiSys '13*, 2013, p. 465.
- [3] S. T. Moturu, I. Khayal, N. Aharony, W. Pan, and A. (Sandy) Pentland, "Using Social Sensing to Understand the Links between Sleep, Mood, and Sociability," in *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, 2011, pp. 208–214.
- [4] Z. Chen, M. Lin, F. Chen, N. D. Lane, G. Cardone, R. Wang, T. Li, Y. Chen, T. Choudhury, and A. T. Campbell, "Unobtrusive Sleep Monitoring using Smartphones," *2013 7th Int. Conf. Pervasive Comput. Technol. Healthc. Work.*, pp. 145–152, 2013.
- [5] J.-K. Min, A. Doryab, J. Wiese, S. Amini, J. Zimmerman, and J. I. Hong, "Toss 'n' turn," in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 2014, pp. 477–486.
- [6] A. Sano, P. Picard, "Stress Recognition using Wearable Sensors and Mobile Phones," *2013 Hum. Assoc. Conf. Affect. Comput. Intell. Interact.*, 2013.
- [7] A. Muaremi, B. Arnrich, and G. Tröster, "Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep," *Bionanoscience*, vol. 3, pp. 172–183, 2013.
- [8] J. Bauer, S. Consolvo, B. Greenstein, J. Schooler, E. Wu, N. F. Watson, and J. Kientz, "ShutEye," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, 2012, p. 1401.
- [9] S. Lawson, S. Jamison-Powell, A. Garbett, C. Linehan, E. Kucharczyk, S. Verbaan, D. A. Rowland, and K. Morgan, "Validating a mobile phone application for the everyday, unobtrusive, objective measurement of sleep," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 2013, p. 2497.
- [10] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "StudentLife," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct*, 2014, pp. 3–14.
- [11] A. E. Poropat, "A meta-analysis of the five-factor model of personality and academic performance," *Psychol. Bull.*, vol. 135, pp. 322–338, 2009.
- [12] E. E. Nofhle and R. W. Robins, "Personality predictors of academic outcomes: big five correlates of GPA and SAT scores," *J. Pers. Soc. Psychol.*, vol. 93, pp. 116–130, 2007.
- [13] A. M. Soehner, K. S. Kennedy, and T. H. Monk, "Personality correlates with sleep-wake variables," *Chronobiol. Int.*, vol. 24, no. 5, pp. 889–903, Jan. 2007.
- [14] W. D. S. Killgore, J. M. Richards, D. B. Killgore, G. H. Kamimori, and T. J. Balkin, "The trait of Introversion-Extraversion predicts vulnerability to sleep deprivation," *J. Sleep Res.*, vol. 16, no. 4, pp. 354–63, Dec. 2007.
- [15] J. K. Clark, "Relationship of personality and sleep to academic success in the United States Military Academy: a perspective utilizing the five factor model of personality," 2007.
- [16] M. Vollrath, "Personality and stress," *Scand. J. Psychol.*, vol. 42, no. 4, pp. 335–47, Sep. 2001.
- [17] 2011 Sleep in American Poll, National Sleep Foundation
- [18] G. S. Brunborg, R. A. Mentzoni, H. Molde, H. Myrseth, K. J. M. Skouevørøe, B. Bjorvatn, and S. Pallesen, "The relationship between media use in the bedroom, sleep habits and symptoms of insomnia," *J. Sleep Res.*, vol. 20, no. 4, pp. 569–75, Dec. 2011.
- [19] S. Thomée, A. Härenstam, and M. Hagberg, "Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults—a prospective cohort study," *BMC Public Health*, vol. 11, p. 66, Jan. 2011.66.
- [20] N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years," *N. Engl. J. Med.*, vol. 357, no. 4, pp. 370–9, Jul. 2007.
- [21] J. H. Fowler and N. A. Christakis, "Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study," *BMJ*, vol. 337, p. a2338, 2008.
- [22] D. J. Buysse, C. F. Reynolds, T. H. Monk, S. R. Berman, and D. J. Kupfer, "The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research," *Psychiatry Res.*, vol. 28, no. 2, pp. 193–213, May 1989.
- [23] O. P. John and S. Srivastava, "The Big Five trait taxonomy: History, measurement, and theoretical perspectives," In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: Guilford, 1999.
- [24] J. Horne and O. Östberg, "A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms," *International Journal of Chronobiology*, 4, 97–100, 1976.
- [25] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *J. Health Soc. Behav.*, vol. 24, no. 4, pp. 385–96, Dec. 1983.
- [26] J. E. Ware, M. Kosinski, and S.D. Keller, "A 12-item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity," *Medical Care* 1996, 34:220-233
- [27] N. Aharony, W. Pan, C. Ip, I. Khayal, and A. Pentland, "Social fMRI: Investigating and shaping social mechanisms in the real world," *Pervasive Mob. Comput.*, vol. 7, no. 6, pp. 643–659, Dec. 2011.
- [28] C. D. Spielberger., R. L. Gorsuch, P. R. Lushene, P. R. Vagg, G. and A. Jacobs, "Manual for the State-Trait Anxiety Inventory," Consulting Psychologists Press, Inc, 1983.
- [29] <https://immersion.media.mit.edu/>
- [30] W. Clerx, A. Phillips, S. Lockley, C. O'Brien, E. Klerman, and C. Czeisler. "Impact of irregularity of sleep-wake schedules on circadian phase and amplitude in college undergraduates," The 2014 Meeting of the Society for Research on Biological Rhythms, 2014
- [31] C. Cajochen, S. Frey, D. Anders, J. Späti, M. Bues, A. Pross, R. Mager, A. Wirz-Justice, and O. Stefani, "Evening exposure to a light-emitting

- diodes (LED)-backlit computer screen affects circadian physiology and cognitive performance.,” *J. Appl. Physiol.*, vol. 110, pp. 1432–1438, 2011.
- [32]A.-M. Chang, D. Aeschbach, J. F. Duffy, and C. A. Czeisler, “Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness,” *Proc. Natl. Acad. Sci.*, 2014.
- [33]A. Sano, R. W. Picard, and R. Stickgold, “Quantitative analysis of wrist electrodermal activity during sleep,” *Int. J. Psychophysiol.*, vol. 94, pp. 382–9, 2014.
- [34]A. Sano and R. W. Picard, “Comparison of sleep-wake classification using electroencephalogram and wrist-worn multi-modal sensor data,” in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 930–933.
- [35]F. C. Brown, W. C. Buboltz, and B. Soper, “Relationship of sleep hygiene awareness, sleep hygiene practices, and sleep quality in university students,” *Behav. Med.*, vol. 28, pp. 33–38, 2002.
- [36]D. Jensen, “Understanding Sleep Disorders in a College Student Population,” *J. Coll. Couns.*, vol. 6, pp. 25–34, 2003.

TABLE II. SELECTED FEATURES*

Outcome Features	High vs Low GPA	High vs Low PSQI score	High vs Low PSS	High vs Low MCS
Surveys**	PSQI + Total hours of academic activity (n=15), wake time + Total hours of academic activity (n=13)	PSS score (pre-study), Sleep regularity, Sickness (evening)	Happiness (morning) + Sluggishness (evening) (n=15), Sluggishness (morning) + Sickness (evening) (n=13)	Happiness (evening), Stress level (evening)
Personality	Conscientiousness, Agreeableness	Neuroticism	Neuroticism, Openness + Agreeableness (n=15), Conscientiousness + Extraversion (n=13)	Openness, Conscientiousness, Neuroticism
SC	2Q sleep amp frequency (0.7-0.8)	Day mean + sit amp frequency (0.2-0.3) + 4Q sleep amp frequency (0.5-0.6) (n=15), run amp median + 1Q sleep amp frequency (0.3-0.4) + Entropy of walk storm (n=13)	Sit amp frequency (0.9-1.0), storm entropy(n=15), 3Q sleep amplitude frequency (0.6-0.7) (n=13)	SD of sit amplitude (n=15), amp frequency (0.8-0.9) (n=13)
ACC	Day SD, Day RMS frequency (2-3) + 4Q sleep (2-4) (n=13), Day RMS frequency (3-4) + whole night sleep RMS frequency (14-16) (n=15)	Day RMS frequency(2-3) + Walk RMS median (n=15), Walk RMS frequency(10) + 3Q sleep frequency (2-3) (n=13)	1Q sleep frequency (18-20) + 2Q sleep frequency (0-2) (n=15), 2Q sleep mean + 3Q sleep median (n=13)	Day RMS frequency (2) + Day PSD (1-2 Hz) (n=15), 3Q sleep frequency (5) + 4Q sleep frequency (10) (n=13)
ST	Day frequency (0.9-1) + 1Q sleep mean (n=13), 4Q sleep frequency (0.3-0.4) + 4Q sleep frequency(0.9-1) (n=15)	Day PSD (0.2-0.3Hz), 1Q sleep frequency (0.6-0.7) (n=13), 3Q Sleep frequency (0.9-1.0) (n=15), 4Q sleep SD	1Q sleep frequency(0.8-0.9), 2Q sleep frequency (0.7-0.8) + 4Q sleep frequency (0.3-0.4) (n=15), 3Q sleep PSD(4) + 1Q sleep frequency (0.3-0.4) (n=13)	Day frequency (0.1-0.2) + sleep frequency (0.1) + 1Q sleep mean (n=13), 1Q sleep frequency (0.3-0.4) + 1Q sleep frequency (0.9-1.0) + 3Q frequency(0-0.1) (n=15)
CALL	Mean timestamp (n=14), entropy (n=12)	Mean timestamp, total duration	Entropy and # of people	Entropy + total duration (n=14), Entropy + mean timestamp (n=12)
SMS	Mean timestamp + entropy	# of SMS (n=15), mean timestamp (n=13)	# of SMS (n=15), mean timestamp (n=13)	# of SMS (n=13), timestamp (n=14)
SCREEN	# of screen on, 6-9am duration, 3am-6am (n=12), 9pm-12pm duration (n=14)	3-6am duration (n=13), 9am-12pm duration (n=15)	6pm-9pm duration, 3-6am duration + 9am-12pm duration (n=15), 6-9am duration + mean timestamp (n=13)	0am-3am duration + 6am-9am duration (n=13), total duration + 6am-9am duration (n=14)
MOB		SD of daily distance traveled	Daily distance traveled, SD of daily distance traveled	Daily distance traveled
INTERNET	Timestamp	Timestamp and total duration	Timestamp and total duration	Timestamp and total duration
Email	# of contacts over sent emails, median timestamp for received emails	SD of timestamp for sent emails	Median of # of received emails, SD of timestamp for sent emails	Median of # of received emails (n=14), SD of sent emails (n=13), median of timestamp for received emails
Social	Total # of negative email contacts (n=12), Frequency of very negative interactions (n=14)	Frequency of memorable positive (n=12) and # of family contacts (n=14)	Frequency of interaction with person before sleep	Frequency of interaction with person before sleep, Face to face neutral interaction (n=12), total # of neutral contacts(n=14)

* Numbers in brackets means the number of participants who showed the combination of the features as the best features.

** Outcome variables were never included as inputs: For high/low PSQI classification, we excluded all PSQI questions, etc. Also, for high/low PSS classification, we excluded both PSS and pre and post study MCS questions since some of the MCS questions were about stress; similarly for MCS classification, we also excluded pre- and post-PSS questions because of similarity to outputs being classified.