Computers in Biology and Medicine 179 (2024) 108749

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Predicting long-term sleep deprivation using wearable sensors and health surveys

Rafael Trujillo [a,*], Enshi Zhang [a], John Michael Templeton [b], Christian Poellabauer [a]

[a] Florida International University - Knight Foundation School of Computing and Information Sciences, 11200 SW 8th St, Miami, FL, 33199, USA
[b] University of South Florida - Department of Computer Science and Engineering, 4202 E Fowler Ave, Tampa, FL, 33620, USA

ARTICLE INFO

Keywords:
Sleep prediction
Wearable devices
Machine learning
Recursive feature elimination
Sleep deprivation

ABSTRACT

Sufficient sleep is essential for individual well-being. Inadequate sleep has been shown to have significant negative impacts on our attention, cognition, and mood. The measurement of sleep from in-bed physiological signals has progressed to where commercial devices already incorporate this functionality. However, the prediction of sleep duration from previous awake activity is less studied. Previous studies have used daily exercise summaries, actigraph data, and pedometer data to predict sleep during individual nights. Building upon these, this article demonstrates how to predict a person's long-term average sleep length over the course of 30 days from Fitbit-recorded physical activity data alongside self-report surveys. Recursive Feature Elimination with Random Forest (RFE-RF) is used to extract the feature sets used by the machine learning models, and sex differences in the feature sets and performances of different machine learning models are then examined. The feature selection process demonstrates that previous sleep patterns and physical exercise are the most relevant kind of features for predicting sleep. Personality and depression metrics were also found to be relevant. When attempting to classify individuals as being long-term sleep-deprived, good performance was achieved across both the male, female, and combined data sets, with the highest-performing model achieving an AUC of 0.9762. The best-performing regression model for predicting the average nightly sleep time achieved an R-squared of 0.6861, with other models achieving similar results. When attempting to predict if a person who previously was obtaining sufficient sleep would become sleep-deprived, the best-performing model obtained an AUC of 0.9448.

1. Introduction

Lack of sleep is a widespread public health problem in modern industrialized societies. Estimates have placed economic losses due to lack of sleep in the hundreds of billions of dollars in the United States alone, and tens of billions in other countries such as Germany, Japan, and the United Kingdom [1]. Increasingly sedentary lifestyles, excessive electronic device usage, and psychosocial stressors have been identified as widespread causes of insufficient sleep [1]. Widespread lack of sleep causes this economic damage through reduced worker productivity and increased medical costs [1]. Insufficient sleep has been associated with several chronic conditions such as obesity, diabetes, and hypertension, as well as life-threatening diseases such as heart attack and stroke [2]. Up to 20% of road accidents have been identified as resulting from drowsiness from lack of sufficient sleep. These types of drowsiness-induced accidents are also twice as likely to be fatal than accidents caused by other factors [3].

Among university students specifically, sufficient sleep has been shown to be necessary for academic success and the maintenance of both physical and mental health [4]. Poor sleep among university students has been correlated with poor academic performance and mental health [5], increased alcohol consumption [6], and increased stress [7]. However, many college students suffer from poor sleep quality, with a recent meta-analysis finding one-third of students suffering from sleep disturbances [8].

In order to measure sleep characteristics, researchers and clinicians regularly use polysomnography (PSG) machines. PSG is considered the most reliable method of measuring sleep. However, traditional polysomnography machines are expensive, intrusive, and may even interfere with sleep [9]. Commercial wearable devices have opened the door for researchers to conduct sleep studies without the use of PSG machines. Since 2017 Fitbit has manufactured models capable of measuring user sleep stages. Various studies have analyzed the accuracy of these models in measuring different sleep parameters. These studies find that recent Fitbit models show no statistical difference from PSG when measuring total sleep time, sleep efficiency, and wake after sleep

* Corresponding author.
E-mail addresses: rtruj023@fiu.edu (R. Trujillo), ezhan004@fiu.edu (E. Zhang), jtemplet@usf.edu (J.M. Templeton), cpoellab@fiu.edu (C. Poellabauer).

https://doi.org/10.1016/j.compbiomed.2024.108749
Received 10 October 2023; Received in revised form 21 May 2024; Accepted 8 June 2024
Available online 2 July 2024
0010-4825/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

onset [9]. Previous studies have found the Fitbit Charge HR, the model that collected the sleep data in this study, to estimate total sleep time with sufficient accuracy for sleep research [10–12].

Previous studies have found a wide range of daytime influences on sleep quality. For instance, researchers have found that phone usage before sleep time will reduce sleep quality [13]. Excessive exercise has been found to negatively impact sleep [14]. Certain foods have also been found to correlate with sleep quality [15].

With the mass adoption of smartphones, research into the usage of mobile phones to deliver healthcare interventions has gained popularity. Mobile health (mHealth) interventions aimed at behavioral change have been investigated in various contexts such as weight management, exercise promotion, medication adherence, and psychological well-being. These interventions have been used with populations suffering from various ailments such as HIV, hypertension, diabetes, and asthma [16]. mHealth interventions have also begun to utilize machine learning techniques to both monitor and predict the course of ailments, as well as identify their causes [17].

This study aims to showcase the efficacy of diverse health-related features in predicting sleep and evaluate the feasibility of forecasting long-term sleep patterns. The overarching goal is to contribute to the development of machine learning-assisted mHealth interventions, specifically targeting the identification of factors contributing to poor sleep and its enhancement. This research holds considerable significance for the future, as it lays the foundation for personalized interventions delivered through smartphones or other devices. These interventions, equipped with sensor data from commercial wearables and Internet of Things (IoT) devices, could offer tailored recommendations for behavior changes. Moreover, they can track user compliance and assess the effectiveness of these recommendations on sleep. As an example, sleep prediction functionality could be integrated into a future mHealth intervention aimed at controlling high blood pressure. This mHealth application could track food consumption and exercise habits by integrating with IOT-enabled kitchen appliances and wearable exercise trackers. In addition to predicting the effects of diet and exercise on blood pressure directly, it could also anticipate their future effects on sleep. Given that poor sleep contributes to hypertension [2], this would allow the application to recognize further the long-term detrimental effects of substances and habits that do not directly increase blood pressure but do so through degrading the user's sleep. In addition to cardiovascular health, numerous psychiatric illnesses are correlated with sleep disorders. These sleep disorders are associated with worse quality of life and prognosis [18]. An mHealth application for managing these disorders will have to consider sleep habits and aid sleep-disordered users in obtaining healthy sleep habits. Given the numerous correlations between poor sleep and negative health outcomes, many future mHealth interventions would benefit from incorporating knowledge about and predicting the user's sleep habits.

In contrast to previous studies investigating sleep prediction from daytime activity, this study investigates the prediction of long-term objectively measured sleep duration, as opposed to focusing solely on individual nights or sleep quality. By expanding the focus to long-term sleep patterns, this study allows for the examination of how chronic lifestyle factors, physical conditions, and psychological factors influence sleep.

Predicting individual nights' sleep may reveal short-term effects, such as the impact of alcohol consumption on sleep quality. However, it fails to capture the effects of chronic behaviors, such as excessive alcohol consumption over time. Additionally, focusing on long-term sleep patterns enables future behavioral interventions that may not be feasible when targeting individual nights. For instance, an mHealth system designed to improve nightly sleep may identify a potential sleep disturbance in advance, but by the time the prediction is made, the user may be unable to address the underlying causes, such as excessive caffeine consumption or nighttime phone use.

Furthermore, individual nights' sleep quality is inherently more variable and influenced by transient factors, such as occasional noisy neighbors or temporary pain, compared to long-term sleep patterns. Therefore, studying long-term sleep trends provides a more stable and comprehensive understanding of sleep behavior.

This study presents four distinct long-term sleep prediction schemes based on wearable data and survey responses. These schemes employ Recursive Feature Elimination (RFE) on sex-segregated datasets. The results underscore the relevance of specific factors — physical activity, exercise motivation, levels of depression, indicators of obesity, and, when included, the previous month's sleep behavior — in accurately predicting long-term sleep. The inclusion of exercise motivation and Big Five openness to experience additionally open up future avenues of investigation for sleep researchers, as these have not previously been found as relevant factors influencing sleep.

This paper will be structured as follows. An overview of similar work focused on predicting sleep from daytime features will be presented in Section 2. Following this, the NetHealth data set will be presented and the features used throughout this study will be explored in Section 3.1. The data filtering criteria, feature selection methodology, RFE-RF trials, and the four sleep prediction schemes will then be described in Sections 3.2 and 3.3 respectively. The results from the RFE-RF trials and sleep prediction schemes will then be presented in Section 4 and their relevance to future sleep-related work will be discussed in Section 5. The paper concludes by examining the limitations of its methodology and data and discussing how future work can overcome these in Section 6.

2. Related work

Previous efforts have been made to predict sleep from self-reported lifestyle factors. In [19], university students' health status, caffeine and drug intake, demographic information, and electronics usage patterns were used to predict if they were sleep disturbed according to their Pittsburgh Sleep Quality Index (PSQI) scores. The authors of [20] utilized self-reported mood, physical self-reported activity levels, work status, and smartphone usage patterns. Alongside the self-reported data, an activity-tracking app (Google Fit) was utilized to measure objective exercise metrics. These features were used to classify the students as good or bad sleepers according to the PSQI.

Wearable sensors have previously been used to predict sleep characteristics from daytime activity. Sathyanarayan et al. [21] had high school students wear commercial actigraphs for a week. The resulting accelerometer time-series data were then used to predict if the wearer achieved at least 85% sleep efficiency the following night.

The socially contagious nature of sleep behaviors have also been used in prediction of sleep. Sano et al. [22] leveraged social network information gathered from call and SMS data in combination with physiological data gathered from a worn Q-sensor to create SleepNet. This system was used to predict participants' next day sleep duration, utilizing both the physiological data and a social graph network constructed from the cell phone data.

Fitbit devices have been used in the past for both sleep measurement and prediction. Fellger et al. [23] obtained data from patients in rehabilitation wearing one of three separate wearable devices, either one of two commercial actigraphs or a Fitbit Charge pedometer. The step count and activity time series from the devices were used to predict both the following night's total sleep time and the next day's physical activity.

One other study, conducted by Kilic et al. [24], utilizes the same dataset as this study (the NetHealth Project) for sleep prediction. The authors utilized a Convolutional Neural Network to predict sleep efficiency utilizing daytime data and health surveys. However, the low range of the predicted sleep efficiency only being between 93% and 95% limits its applicability in predicting medically relevant outcomes. This study, in contrast, focuses on long term average sleep duration with multiple of the NetHealth participants sleeping less than the minimum six hours of sleep recommended by the National Sleep Foundation [25].

Previous research that exclusively used Fitbit devices has attempted to use daily activity summaries instead of time-series data. Hidayat et al. used Fitbit Alta's to predict changes in an individual's sleep quality from daily physical activity summaries [26]. Phan et al. used Fitbit Charge HRs to collect sensor data from freshmen for 106 days [27]. The participants' daily calories, distance and steps walked, and total time at different activity levels were used to predict whether individuals would achieve a minimum of 6 h and 42 min of sleep.

The unique contributions of this study in contrast to previous efforts are as follows. This study examines the prediction of long-term objectively measured sleep duration, in contrast to individual nights or sleep quality. A wide variety of long-term survey and wearable data gathered over the course of months are incorporated into the machine learning models; previous efforts focusing on sleep duration have mainly focused on the usage of sensor data. The relative importance of different features in predicting long-term sleep is examined through the usage of RFE; special focus is given to the relative importance of past sleep data as previous attempts that have incorporated it have not attempted to investigate its relative importance when compared to daytime sensor data. Finally, sex differences in long-term sleep predictors, RFE feature selection, and the performance of machine learning models are examined in detail.

3. Methods

This section provides an overview of the methods employed in this study. It begins by detailing the NetHealth dataset and describing the features utilized. Subsequently, it outlines the data filtering processes implemented to ensure data quality. This is followed by describing the RFE-RF trials conducted to determine the relative importance of different features in four future sleep prediction schemes. The integration of the RFE-RF technique into the overall prediction process is also explained.

3.1. NetHealth data set

The data used in this study comes from a public pre-existing data set, the *NetHealth* project,[1] a multi-year data collection project studying the effects of social media on student health [28,29]. 698 students wore Fitbit Charge HR devices over the course of several months up to multiple years. Recruiting occurred from the Fall 2015 semester to the Spring 2016 semester. Students participated for up to four years after initial recruitment. While wearing these devices, they completed multiple surveys in "waves" every few months. For this study, answers given during the second wave "W2" in the winter of 2016 were used as features in the machine learning models. Demographic information asked during the first survey collection wave was also included. Alongside these responses, Fitbit data acquired in the months before and after these surveys was also used. This wave was selected as it still retained a large cohort of participants, had a range of potentially relevant surveys, and had multiple prior months of Fitbit data available.

3.1.1. Daytime measurements

The W2 wave contained a variety of surveys relating to participants' daytime behaviors. Surveys asked about mental and physical health, personality, physical exercise habits, technology use, and drug and alcohol consumption. Additional features used were from Fitbit recorded physical activity over the course of 120 days, three 30-day periods before the surveys were administered, and the 30 days after. These measurements include daily steps and time spent at different heart rate zones. Alongside these, we included time spent at different Fitbit-defined levels of physical activity, which Fitbit calculates through expended metabolic equivalents (METs) [30].

3.1.2. Sleep measurements

Alongside daytime behavior, measurements of previous sleep patterns were also used for prediction. For each of the three previous months, a participant's Fitbit calculated average sleep time, sleep efficiency, minutes in bed before falling asleep, and minutes in bed after waking were all included. From the W2 surveys, answers to the PSQI were excluded as they are derived from self-reported sleep problems, including insufficient sleep [31].

3.1.3. Surveys

The NetHealth data set includes a variety of standardized psychological surveys:

- **Center for Epidemiological Studies-Depression** or **CES-D:** a 20-question survey originally designed to measure levels of depression in adolescents and young adults [32].
- **Self Regulation Questionnaire Exercise** or **SRQ-E:** a questionnaire aimed at identifying the perceived sources of motivation for why a person engages in physical exercise [33].
- **Beck Depression Inventory** or **BDI:** a standardized self-report inventory measuring levels of depression [34].
- **Beck Anxiety Inventory** or **BAI:** a self-report questionnaire designed to measure a person's levels of anxiety [35].
- **Big Five Inventory:** a self-report inventory designed to measure a person's Big Five personality traits; these are neuroticism, agreeableness, openness to experience, extroversion, and conscientiousness [36].
- **The Social & Emotional Loneliness for Adults** or **SELSA:** this scale measures the degree of an adult's social isolation from family, romantic partners, and broader society [37].
- **Morningness–Eveningness Questionnaire** or **MEQ:** this is designed to measure a person's preference for performing activities during the morning or evening [38].
- **Spielberger State-Trait Anxiety Inventory** or **STAI:** this measures both the taker's immediate and habitual levels of anxiety [39].

In addition to standardized questionnaires, participants were also asked questions regarding:

- Usage of alcohol, prescription and recreational drugs, caffeine, and tobacco.
- Political and societal views.
- Perceived degree of physical and mental health, alongside the presence of hereditary diseases.
- The presence of major life changes.
- Computer and smartphone usage habits.
- Physical exercise habits, including solitary and team sports.
- Perceived self-efficacy in exercise and dietary habits.
- Race and ethnic background.
- Sexual orientation and activity.

These survey questions were all made available starting on January 31, 2016. Participants submitted their responses to all the questions in the wave simultaneously via Qualtrics.

3.1.4. Data filtering

In order to ensure high-quality data, information from days when students wore the Fitbit for less than 90% of their time awake was excluded. If a 30-day period did not contain at least 10 days of sleep and physical activity data, then it was excluded. Any participant who had a

[1] The data set is available at the *NetHealth Project* website: https://sites.nd.edu/nethealth/

Table 1
Sleep characteristics by sex.

	Data set		
	Combined	Male	Female
Size	257	144	113
% Poor Sleepers	31.9	36.1	26.5
Mean Sleep in Minutes	379.8	376.9	383.4
Sleep in Minutes Standard Deviation	55.6	53	58.7

30-day period with less than 10 days of sleep or 10 days of physical activity was excluded from this analysis. These cutoffs were chosen to allow for sufficient data quantity (reducing how many samples) while maintaining quality (having samples from individuals with high compliance) for Machine Learning outcomes. Participants were then classified as "poor" sleepers if their average sleep time for the 30 days after submitting their responses to the survey wave was less than or equal to six hours, otherwise, they were labeled as "normal" sleepers. This data was then split according to sex, with a male-only, female-only, and combined data set. The six-hour threshold was chosen since a normal young adult requires 7–8 h of sleep, and the National Sleep Foundation recommends no less than six hours per night for adults aged 18–64 [25]. Sleeping less than six hours per night has been associated with increased cardiovascular disease, risk of diabetes, obesity, and cancer [40–43]. The sleep characteristics of the three data sets can be seen in Table 1.

It must also be noted that Fitbit recorded daytime napping is considered a daytime feature and not part of a person's average sleep. This is because previous sleep research has shown that chronic insufficient nighttime sleep has harmful effects that cannot be negated with subsequent napping nor short-term "makeup" sleep [44]. As a result, even if a person accumulates at least 6 h of mean sleep per day through a combination of napping and primary nighttime sleep, they are still classified as a poor sleeper if they do not achieve at least six hours of sleep during the night.

To exclude napping, a participant's sleep records were labeled as either a nap or the day's main sleep. These labels were derived as follows, if a sleep record ended (the participant woke up) on a certain date, and there are no other sleep records that end on that date, then it is classified as main sleep. If multiple records end on the same date, then the record with the earliest end time is considered the main sleep, and the others are classified as napping. Neither the time of day during which sleep occurs nor its duration was used to determine if a sleep episode was the day's primary sleep. This definition was used as it allows for very irregular and poor sleep schedules to be accommodated. For example, an individual starting to sleep after sunrise due to insomnia and then waking up later in the day is correctly handled. If the criteria were based on the time of day then this might be misclassified. Another extreme example would be an individual only managing to obtain two hours of sleep during the night, napping three hours in the afternoon, and then sleeping a normal eight hours. If sleep duration was used as a criterion, then napping might be classified as the main sleep.

3.2. Feature selection

In order to reduce the dimensionality of the data, which totals 157 features, in a manner that allows us to determine the relative importance of different features, Recursive Feature Elimination was used alongside a Random Forest model (RFE-RF) [45]. RFE is an iterative process for selecting features alongside a machine learning algorithm. The process attempts to make predictions with an initial set of variables and then discards the lowest-ranked features. The relative importance of features is extracted from the Random Forest model training on the data [46]. This process repeats until the target number of features is achieved. Given that this process must be repeated over different training sets during cross-validation to avoid over-fitting, the selected features vary with each iteration [45]. The data sets were split according to 5-fold cross-validation, with each training set being used to perform feature selection. The resulting subset of features was then used to train a Random Forest model, which was validated against the test set. This process was then repeated 3 times over the different desired amount of features for each of the three data sets for both regression and classification. For regression models, the feature number that yielded the maximum R-squared with the test data was selected as optimal, while for classification models the largest ROC-AUC was used. In order to determine some of the most relevant features, RFE-RF was again performed within a 5-fold cross-validation loop iterated 20 times with the previously determined optimal number of features.

After the determination of the most important features, different machine-learning models were used in conjunction with RFE-RF. K-Nearest Neighbor (with "brute-force" nearest neighbor selection) and Support Vector Machine models were used for both regression and classification, while linear regression and logistic regression were used for regression and classification respectively. To prevent over-fitting, the process of determining the optimal set of features was not performed as part of hyper-parameter selection, as this would have used information from the entire data set. Instead, within each cross-validation loop, RFE was performed over multiple possible desired numbers of features and only on the training set. The complete model selection is as follows: a set of hyper-parameters was searched over for the different machine learning models. For each set of hyper-parameters, the data set was shuffled and split for 5-fold stratified cross-validation. For each of these cross-validation splits, the data was scaled and then RFE-RF was performed 20 times with different desired feature numbers. Within each RFE-RF loop, the machine learning model was trained and validated on the training set using the selected features using 5-fold cross-validation. For classification, the features that achieved the highest ROC-AUC score were selected to be used to train the model on the entire training set and validated against the test set. The same process was applied for regression, except the features that yielded the highest R-squared were chosen. The entire process is visualized in Fig. 1.

3.3. Mixed and daytime activity only models

This work examines four different schemes for predicting sleep. The first is an attempt at predicting the average amount of time a person sleeps per night over the course of one month. The second is an attempt to classify a participant's sleep status as a healthy or unhealthy sleeper. These two initial schemes utilize all the above-described features.

Previous sleep prediction work which has utilized both exercise and sleep Fitbit data have not yet attempted to determine the relative contributions of past sleep patterns in predicting future sleep. In an effort to determine this, further attempts were made to predict sleep entirely from daytime activity and survey data, excluding previous nighttime sleep records entirely. The first of these is again attempting to classify someone as a healthy or unhealthy sleeper. The second is an attempt at predicting whether a participant previously classified as being a healthy sleeper will transition into an unhealthy sleeper two months later. These later two schemes also included additional changes in the features being used. Various lifestyle and drug, alcohol, and caffeine usage features were removed after being identified as irrelevant to the previous mixed schemes. Additional Fitbit features were also added, which indicate changes in average daily exercise between the first month and the current month.

4. Results

This section presents the results obtained from the RFE-RF feature selection trials and shows the relative importance of different features in the four prediction schemes. The performance of various machine learning algorithms using these four schemes is then presented.

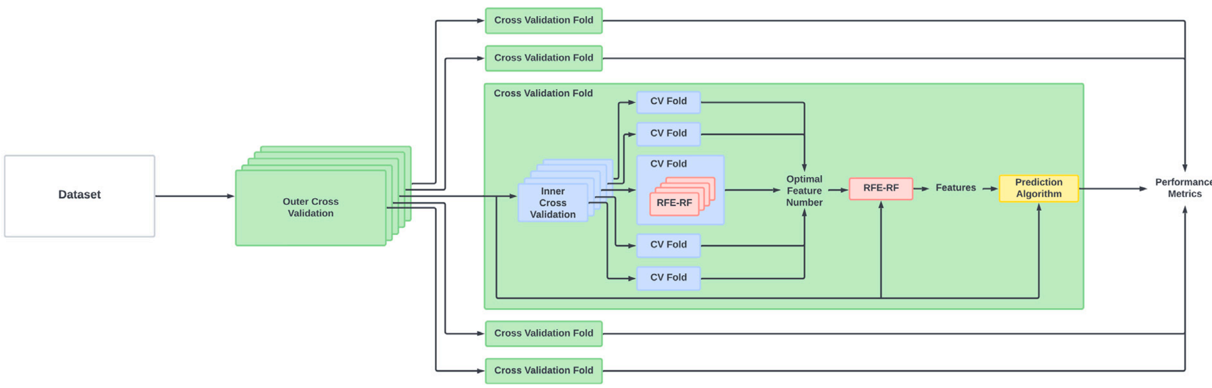

Fig. 1. Diagram illustrating the feature selection and prediction process. This process was iteratively performed for each of the presented machine learning algorithms, with varying sets of hyper-parameter values. Note that every cross-validation fold (CV Fold) is processed the same way, but the process is only illustrated for a single outer and inner fold for clarity.

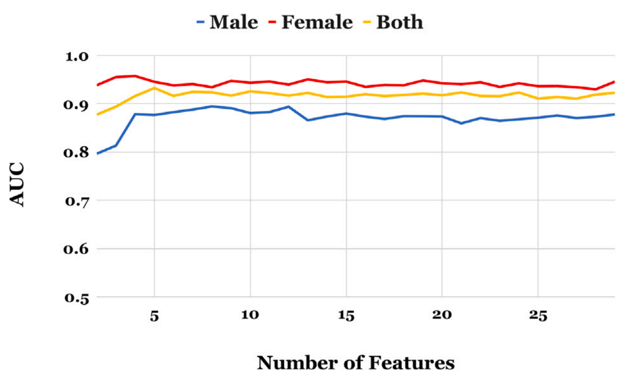

Fig. 2. Mixed daytime and nighttime Random Forest Classifier, predicting sleeper status using the n features selected from RFE-RF.

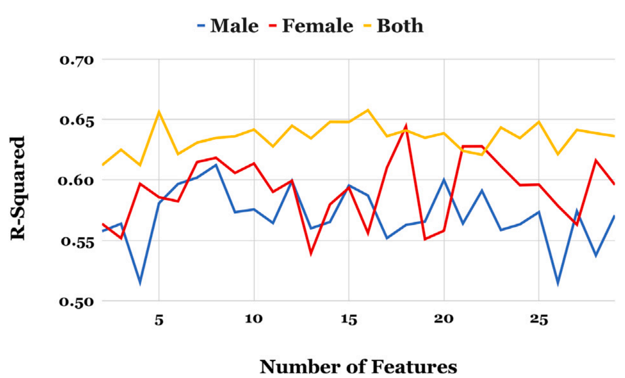

Fig. 3. Mixed daytime and nighttime Random Forest Regressor R-Squared Scores using the n features selected from RFE-RF.

Table 2
Optimal number of features.

	Daytime and nighttime		Daytime only	
	Classification	Regression	Classification	Degradation
Combined	5	16	24	17
Male	8	8	28	9
Female	4	18	4	18

4.1. RFE-RF

Figs. 2–5 show the performance changes of the random forest model as the target number of features is changed for RFE-RF. The optimal

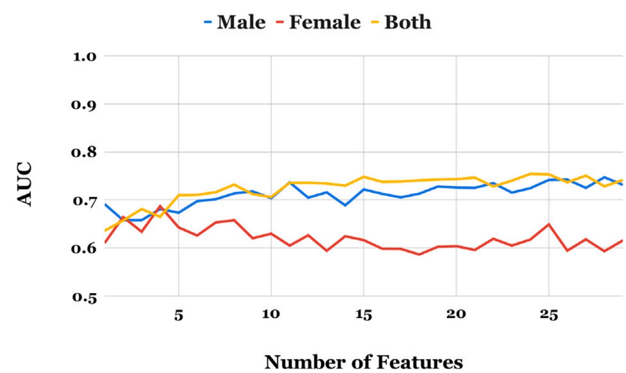

Fig. 4. Daytime activity only RFE-RF AUC scores predicting sleeper status.

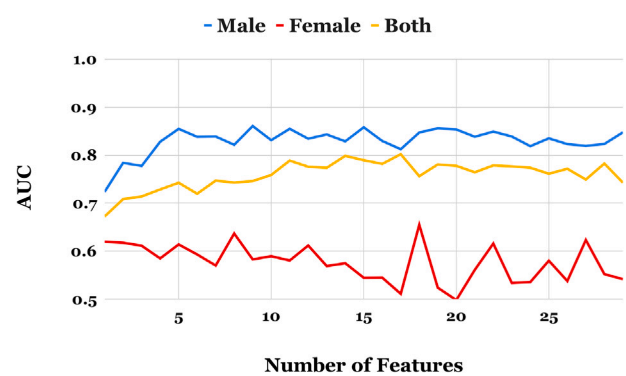

Fig. 5. Performance of a Random Forest Classifier predicting sleeper status degradation using the n features selected from RFE-RF.

feature numbers for the data sets are shown in Table 2. Tables 3–5 show the types of features that were selected when RFE-RF is performed with the optimal feature number, as well as the percentage of times that feature type was selected.

4.1.1. Mixed daytime and nighttime models

Figs. 2 and 3 show that good performances were achieved with relatively few features for the mixed daytime and nighttime models across all three data sets. The models' performances then suffer a gradual decline as more features are introduced passed the optimal number. Classification models generally required fewer features and more stable performances as the number of features varied. For classification, only four features were required to reach optimal performance

Table 3
Percentage of times a feature category was selected by RFE-RF for sleeper status classification using both daytime and sleep features. In parenthesis is the number of features being selected for.

Daytime and nighttime classification			
Category	Both (5)	Male (8)	Female (4)
Fitbit daytime activity	100%	100%	64%
Fitbit previous months' daytime activity	3%	48%	31%
Fitbit previous months' sleep	100%	100%	100%
Napping	0%	0%	2%
Depression	0%	3%	2%
Big Five openness	0%	3%	0%
SRQ-E exercise motivation	0%	80%	0%
Exercise self-efficacy	0%	7%	0%
Obesity	2%	34%	0%
Self reported exercise frequency	0%	2%	0%

Table 4
Percentage of times a feature category was selected by RFE-RF for average sleep duration regression using both daytime and sleep features. In parentheses is the number of features being selected for.

Daytime and nighttime regression			
Category	Both (16)	Male (8)	Female (18)
Fitbit daytime activity	100%	100%	100%
Fitbit previous months' daytime activity	100%	98%	100%
Fitbit previous months' sleep	100%	100%	100%
Napping	48%	9%	46%
Depression	11%	1%	29%
Big Five openness	10%	0%	4%
Big Five extraversion	0%	0%	53%
Big Five conscientiousness	3%	2%	1%
Big Five agreeableness	0%	0%	16%
Big Five neuroticism	1%	0%	2%
SRQ-E exercise motivation	20%	11%	37%
Exercise self-efficacy	5%	0%	0%
Obesity	11%	4%	65%
Self reported exercise frequency	21%	4%	14%
Physical health	0%	3%	0%
Daily activities	0%	0%	12%
MEQ preference	0%	0%	5%
Caffeine usage	0%	0%	10%

Table 5
Percentage of times a feature category was selected by RFE-RF for sleeper status classification using only daytime features. In parentheses is the number of features being selected for.

Daytime only classification			
Category	Both (24)	Male (28)	Female (4)
Fitbit daytime activity	100%	100%	100%
Diff. in Fitbit activity from first month	100%	100%	42%
Fitbit previous months' daytime activity	100%	100%	97%
Depression	93%	91%	1%
Big Five openness	24%	20%	0%
Big Five extraversion	1%	1%	0%
Big Five conscientiousness	1%	5%	0%
Big Five agreeableness	0%	1%	0%
SRQ-E exercise motivation	81%	100%	0%
Exercise self efficacy	81%	82%	6%
Obesity	100%	96%	19%
Self-esteem	20%	9%	0%
Physical health	1%	3%	0%

for the female data set, while for regression, the minimum number of features selected was for the male data set with only eight. For both classification and regression, the models were better able to predict female sleep than male sleep.

Tables 3 and 4 show that when incorporating previous sleep features, the average sleep lengths of the previous 30-day periods are extremely important features, appearing in practically all the feature selections. Indicators of physical exercise, such as average steps per day and average minutes with an elevated heart rate, also appear to be very relevant. Not only is current physical activity relevant, but indicators of physical activity from previous 30-day periods are also consistently being selected as well. Alongside actual physical activity, scores for the Exercise Self Regulation Questionnaire (SRQ-E) also appear relevant. The SRQ-E is a questionnaire aimed at identifying the perceived sources of motivation for why a person engages in physical exercise [33]. Big Five personality traits also appeared in the results. The Big Five personality model comprises five statistically independent traits [47]. The first is agreeableness, marked by a tendency towards friendliness, a large degree of concern for others, and a preference for cooperation over competition. Extraversion is characterized by a preference for large amounts of sensory stimulation and social interaction. Openness is an individual's desire for novel experiences and ideas, and an appreciation of art. Individuals high in conscientiousness tend to be responsible, organized, and methodical. The final trait is neuroticism, or a person's propensity towards emotional instability, negative thoughts and feelings, depression, and anxiety [47]. Depressive symptoms also appeared relevant, as shown by the inclusion of CES-D and BDI scores. A person's weight and Body Mass Index (BMI) also appeared relevant. Finally, an individual's self-reported average nap duration was also relevant.

Table 6
Percentage of times a feature category was selected by RFE-RF for sleeper status degradation using only daytime features. In parentheses is the number of features being selected for.

Daytime only sleep status degradation			
Category	Both (17)	Male (9)	Female (18)
Fitbit daytime activity	100%	100%	100%
Diff. in Fitbit activity from first month	100%	100%	100%
Fitbit previous months' daytime activity	100%	85%	100%
Depression	96%	62%	81%
Big Five openness	100%	65%	88%
Big Five extraversion	8%	0%	3%
Big Five agreeableness	6%	0%	0%
SRQ-E exercise motivation	99%	99%	85%
Exercise self efficacy	39%	7%	0%
Obesity	97%	54%	86%
Has mental health problems	0%	0%	27%
Self reported happiness	0%	0%	4%
Self-esteem	0%	0%	8%
Physical health	1%	0%	4%
Anxiety	0%	0%	1%

4.1.2. Daytime only models

The daytime-only models, in contrast to the mixed ones, showed a general increase in performance as the number of features increased for the male and combined sets while showing a general decline for the female data set.

Tables 5 and 6 show that when only considering daytime features, Fitbit recorded exercise, motivation for exercise, levels of depression, and BMI are the most predictive features. These are essentially the same results as those obtained by the models that include previous sleep records, with the obvious exception of previous sleep records.

4.2. Model performance

Given the large number of training and validation cycles required to perform hyper-parameter tuning and RFE-RF, the trials were run using Florida International University Instructional and Research Computing Center's High-Performance Computing system.[2] In total, the trials took 98,883 compute minutes to complete.

Tables 7 and 8 show the performance achieved by the classification and regression models that utilize previous months' sleep on

[2] The Instructional and Research Computing Center's website is available at https://ircc.fiu.edu

Table 7
Daytime and nighttime classification models.

Algorithm	Male			Female			Combined		
	AUC	Features		AUC	Features		AUC	Features	
		Mean	SD		Mean	SD		Mean	SD
Logistic Regression	0.8879	11	6.099	0.9224	11	4.775	0.911	10.2	5.268
KNN	0.9267	7.2	1.72	0.9762	4	2.607	0.9394	6.8	4.956
SVM	0.8974	9.8	8.611	0.9475	5.8	6.675	0.9223	6.6	3.878
Random Forest	0.8648	8.6	4.317	0.9409	11.4	6.829	0.915	16	2.449

Table 8
Daytime and nighttime regression models.

Algorithm	Male			Female			Combined		
	R2	Features		R2	Features		R2	Features	
		Mean	SD		Mean	SD		Mean	SD
Linear Regression	0.6703	13.6	4.445	0.6756	15.4	2.154	0.6532	8.2	2.417
KNN	0.4705	6.4	2.871	0.5773	6.2	6.997	0.61	5.6	3.262
SVR	0.65	16	2.191	0.6651	4.8	5.154	0.6861	15.4	2.653
Random Forest	0.5858	6.2	4.707	0.5241	9.8	4.956	0.6638	7.4	3.007

Table 9
Daytime only sleep status degradation models.

Algorithm	Combined			Male			Female		
	AUC	Features		AUC	Features		AUC	Features	
		Mean	SD		Mean	SD		Mean	SD
Logistic Regression	0.7784	17	3.347	0.874	7.6	5.643	0.7086	8	5.797
KNN	0.7709	6.6	4.454	0.8946	11.8	7.359	0.8107	11	5.177
SVM	0.8545	8.8	1.72	0.9448	11	4.335	0.7633	11.4	5.678
Random Forest	0.6725	8.6	4.079	0.8046	15	6.033	0.6945	10.2	3.187

Table 10
Daytime only sleep status classification models.

Algorithm	Combined			Male			Female		
	AUC	Features		AUC	Features		AUC	Features	
		Mean	SD		Mean	SD		Mean	SD
Logistic Regression	0.7534	6.4	4.08	0.7584	9	6.419	0.6347	8.8	3.06
KNN	0.7473	13.6	6.151	0.8212	13.6	4.587	0.7335	5.2	2.034
SVM	0.8156	13.6	3.911	0.7806	12.4	6.344	0.8159	7	6.066
Random Forest	0.6936	16.6	3.2	0.707	10.8	5.036	0.6348	15.2	5.706

the different data sets, alongside the mean number of features and their standard deviations used in the training cycles. Tables 9 and 10 show the performance achieved by the classification models only using daytime features.

A common result across all the prediction schemes is the high standard deviation in the number of features used. This indicates that the optimal feature number is not a stable property across cross-validation folds.

For the mixed daytime and nighttime classification models shown in Table 7, The KNN model was the best-performing model consistently across the three data sets, with an AUC of 0.9267 for the male set, and 0.9762 and 0.9394 for the female and combined sets respectively. It also achieved these results with fewer features than other models. It utilized the fewest mean number of features for both the male and female sets, and the second-fewest for the combined set.

In the mixed daytime and nighttime regression models shown in Table 8, in contrast to the mixed classification models, there was no single best-performing algorithm across the different data sets. Instead, linear regression performed best for both the male and female data sets, with an R-squared of 0.6703 and 0.6756, respectively.

The removal of previous sleep records led to a significant decline in AUC scores, with most models achieving less than 0.8 AUC. In contrast to the mixed daytime and nighttime models, the performance among different algorithms showed much larger performance differences. For example, when predicting the degradation of sleep quality among males the SVM models achieved an AUC of 0.9448, while the random forest models only achieved a score of 0.8046, a difference of 0.1402, for the mixed models this difference was only 0.0326.

5. Discussion

From the results obtained by the RFE-RF selection process, and the performance of the models, various conclusions can be drawn. The most important of these are as follows. The majority of the predictive power in mixed daytime and nighttime models derives from past sleep features. Past sleep is more predictive of future female sleep than male sleep. Predicting long term sleep is still possible without past sleep records. Daytime activity features and survey data is more predictive of sleep for males than for females. The most relevant health survey related features were exercise motivation, indicators of obesity, and signs of depression. And finally, various features appeared as relevant despite previous sleep research not suggesting their importance, such as exercise motivation, and Big Five openness to experience.

5.1. RFE-RF feature selection

Special care should be given when interpreting the feature selection results. For a single prediction scheme, such as the daytime only sleeper status classification, the number of features being used can have large variance, in this case 4 for the female set and 28 for the male set. Differences in the selection rate for different features must be analyzed with this in mind, large differences in the optimal number of features makes interpretation difficult. For example, in Table 5 features indicating depression are selected 91% for the male set for daytime only classification, but are selected only 1% for the female set. This does not indicate that depression is almost irrelevant for female sleep prediction, as there are only 4 features being chosen. Its actual relevance should be obtained by looking at its presence in the combined dataset, alongside its very high selection rate in Table 6 for all of the data sets.

With the above considerations in mind, there appear to be sex differences in feature selection. For the combined and female data sets, regression achieved optimal results with more features than classification. This was not present for the male data set. Both classification and regression performed better with the female set than the combined or male one. This indicates that these features, in combination, are better able to predict long-term sleep duration for females than males. The male data set performed worse than the other two for both classification and regression. The random forest model was more able to predict sleep quantity for the female data set than the combined data set. However, this reverses for classification, where it was able to classify the combined data set better. For classification tasks, RFE-RF obtained optimal results when using fewer features when compared with regression.

5.1.1. Daytime fitbit features

Fitbit measurements of physical activity from the predicted month were repeatedly selected, with the number of minutes a person is sedentary in a day being consistently selected as a feature. The number of steps taken and the amount of time in different Fitbit-defined heart rate zones are also likely proxies for physical activity. Interestingly, both physical activity measures from previous months and motivations for exercising are also consistently selected. These likely indicate the extent to which a person regularly exercises. Previous research has indicated that insufficient physical exercise is a leading cause of lack of sleep [1]. The inclusion of physical activity metrics may also be related to the inclusion of BMI scores. Obesity has been linked with poor sleep [48] and is also a result of insufficient physical exercise.

5.1.2. Previous sleep features

When including historic sleep quantity data from the previous three months, said data is clearly the most important set of features when predicting sleep quantity, with these features appearing in 100% of the iterations across the three data sets for both classification and regression. The high importance of previous months' sleep measurements is to be expected, an individual who does not obtain sufficient sleep over months due to various lifestyle factors is likely to keep suffering from lack of sleep if they do not significantly change their behavior. Oddly, previous months' sleep efficiency also consistently appears as a selected feature. Sleep efficiency is the total time sleeping divided by the actual total time spent in bed and is used as a measure of sleep quality. Traditionally, sleep quality is viewed as orthogonal to sleep quantity, with sleep efficiency being viewed as an objective measure of sleep quantity [49]. Hence, the inclusion of sleep efficiency is unusual.

5.1.3. Survey features

Among the survey features, the most important factors are a person's weight, levels of depression, and motivation for exercise. Academic sleep literature may reveal the reasons for including these features. Mental health problems have been associated with poor sleep [1]. Depression has been shown to both cause sleep disturbances [50] as well be worsened by sleep deprivation [51]. These findings are likely reflected through the inclusion of an individual's CESD score. This also likely explains the inclusion of a person's Big Five trait neuroticism, as this trait is directly linked with a person's tendency towards emotional instability and anxiety [47]. The inclusion of the Big Five trait extraversion could potentially be related to the excessive usage of social media, especially among women, where it was chosen 59% of the time for regression. Both being a woman [52] and having higher levels of extraversion [53,54] have been associated with increased social media usage and addiction. Social media usage has been directly linked with worse sleep among university students [55].

An outlier among these survey features is a person's level of Big Five openness. This trait appears not very relevant in the mixed daytime and nighttime models, nor in the daytime-only models that predict sleeper status. However, in the models that predict the degradation of a previously healthy sleeper into an unhealthy one, it is one of the most chosen features. Previous research analyzing the relationship between sleep and personality has not shown openness to experience to be relevant [56,57].

Another unusual result is the consistent inclusion of exercise motivation in the form of SRQ-E results. To the authors' best knowledge, no previous research has been done examining the motivation for exercising with sleep. A person's motivation for performing physical exercise is likely correlated to the intensity and frequency of said exercise, leading to its inclusion as a form of proxy for these. However, its persistent inclusion would indicate that it contains more information than found in the Fitbit recorded features. It could also potentially be correlated with other facets of exercise that are not captured by Fitbit, such as the degree to which the physical activity reduces stress, is perceived as fun, or leads to feelings of tiredness or physical soreness, all of which might relate to sleep outcomes.

The lack of inclusion of demographic information is also an interesting result in the feature selection process. Features such as self-identified race and whether one was a foreign or domestic student were included in the models. Racial differences in sleep outcomes have been well established [58,59], so their exclusion is unexpected. This may result from other features correlating with race sufficiently that the actual racial differences may not provide additional information, though further investigation is warranted.

5.2. Mixed daytime nighttime classification models

When including previous nighttime data, all the models achieved relatively similar performance. The worst-performing model and data combination was with the random forest model on the male data set, with an AUC of 0.8648, a decrease of only 0.11 from the best-performing one. Interestingly, there is a large amount of variability in the optimal number of features used by the models, even when they achieve similar performance. For example, the best-performing model, the KNN model on the female data set, achieved an AUC of 0.9762 with an average of only four features, while the Random Forest model on the combined data set utilized on average four times as many features and achieved an AUC of 0.915. This demonstrates that a majority of the model's predictive power comes from a few features, as the addition of more features did not generally result in increased performance even across different algorithms. The RFE-RF feature selection results indicates that these highly predictive features are participant's previous Fitbit measured sleep measurements and physical activity scores, given that these were used by all models for all the data sets. This indicates that the features used were more potent when predicting female sleep, as the highest performance was achieved with the fewest features. The opposite is true for the male data set, which achieved the worst results while using more features on average. When combined with the previously stated observation that the majority of the predictive power of the features derives from past sleep, it can be concluded that past sleep is more predictive for the female data set than for male one.

5.3. Mixed daytime nighttime regression models

In contrast to the classification tasks, the best-performing models for a data set always utilized a relatively large number of features among the different models when contrasted with the mixed classification models, with an average of 13.6, 15.4, and 15.4 features being used for the male, female, and combined data sets, respectively. When combined with the This would indicate that predicting the specific mean time sleeping requires more features, than classifying individuals as good or bad sleepers.

5.4. Daytime activity only models

The decrease in performance shown by the daytime-only models reinforces that previous sleep data held considerable predictive power. Even without these features though, sleep was still being effectively predicted, indicating it is still possible to classify participants as good or bad sleepers without knowing a person's previous sleep patterns.

In a reversal from the mixed models, daytime activity-only predictions on the male data set generally performed better than on the female data set. The mean male AUC across different algorithms for sleep quantity prediction is 0.7668, while for females, it is 0.704725. When considering the degradation of sleep, the mean male AUC is 0.8795 and 0.744275 for females. The poor results obtained with the female data set, in contrast with the other two, indicate that the included features are much more predictive for the male data set than the female.

6. Limitations and future work

It must also be noted that further work is necessary to assess the generalizability of the results presented in this study. Future research should focus on using an independent test set that is completely separate from the model training and feature selection process to provide a more accurate evaluation of how the proposed models would perform in real-world scenarios.

Additionally, there are inherent limitations in the dataset used in this study. The presented results are based on data from a single cohort of freshman students from the same university over a few months. Future research should aim to examine non-student populations or, at the very least, students from other geographic locations. It should also explore individuals' sleep patterns at different time periods, as all the presented results pertain to the same period of time.

Furthermore, the current study did not investigate methods for improving prediction performance and generalizability, such as employing more advanced feature selection or model selection techniques.

Future work should also focus on identifying valuable features that can indicate a person's sleep. This could be done by incorporating insights from sleep researchers on the factors that influence a person's sleep. More specifically, special attention should be paid to any factors that lead to chronic sleeplessness. Diet-related features derived from IOT-enabled devices are a promising direction, allowing further investigation of the effects of different foods on sleep. Non-wearable sensors measuring environmental factors are also a promising direction for future research, with a recent study from Fritz et al. [60] having shown promising results by examining the effects of air quality on sleep quality. More features can also be obtained from more fine-grained sensor data. Fitbit provides access to minute-by-minute records of a person's physiological and physical activity state. These were not available for The NetHealth Project, so future research should attempt to utilize other data sets that include this information. In addition, participants' ages were not recorded in the NetHealth data set. Although all participants were incoming university students, and thus the majority were likely young adults, including age would likely yield additional insights and increased prediction accuracy.

Even without minute-by-minute Fitbit data, there are still further features that can be derived that were not included in this study. For example, the average time and variability of when a person goes to bed and when a person wakes up are all potentially useful features that should be investigated. Daily estimates of caloric consumption could be calculated from available features, the percentage of days where the participant is physically active, and breakdowns of exercise by day of the week, among other daytime features that could prove useful.

Alongside obtaining more features, the effects of modifying the data-cleaning process should be looked at. The current study removed all samples that had less than ten days of sleep or physical activity, and days with less than 90% Fitbit wear time were excluded. This cutoff was chosen semi-arbitrarily as a good middle point that permitted sufficient samples to perform machine learning while still retaining quality. This could be loosened, either reducing the day and sleep requirements or reducing the wear time requirements. This would lead to an increased data set size, potentially improving model performance. Conversely, stricter filtering might lead to better results as the resultant data would be of higher quality. The authors of this study believe a more empirical approach to data filtering examining both possibilities would likely lead to improved results.

7. Conclusions

The above results demonstrate that long-term sleep can be predicted from various daytime behaviors alongside physical and psychological traits, including exercise habits, exercise motivation, obesity, Big Five openness, daytime napping, and indicators of depression among others. The most important of these factors were measurements of physical exercise. A novel finding was that both the actual amount of exercise as well as the motivations for engaging in exercise were significant predictors. To the authors' best knowledge, this correlation between exercise motivation and sleep has not been previously examined in sleep literature. After physical exercise, the most powerful features were indicators of obesity and depression. When previous months' sleep measurements were included as features, these in isolation provided powerful predictors of future sleep without needing daytime activity. Large sex differences exist in feature selection and model performance. Daytime habits were more strongly correlated with male sleep than female sleep, as were mental health and obesity indicators. As a result, daytime-only models performed better for males than females. Mixed models that incorporated previous sleep performed better for females than males, likely resulting from previous sleep patterns correlating more strongly with female sleep than for male sleep.

CRediT authorship contribution statement

Rafael Trujillo: Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Enshi Zhang:** Writing – review & editing, Writing – original draft. **John Michael Templeton:** Writing – review & editing. **Christian Poellabauer:** Writing – review & editing, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge the Instructional & Research Computing Center (IRCC) at Florida International University for providing computing resources that have contributed to the research results reported within this paper. Web: https://ircc.fiu.edu

References

[1] M. Hafner, M. Stepanek, J. Taylor, W.M. Troxel, C. Van Stolk, Why sleep matters—the economic costs of insufficient sleep: a cross-country comparative analysis, Rand Health Q. 6 (2017) http://dx.doi.org/10.7249/rr1791.

[2] N.G. Altman, B. Izci-Balserak, E. Schopfer, N. Jackson, P. Rattanaumpawan, P.R. Gehrman, N.P. Patel, M.A. Grandner, Sleep duration versus sleep insufficiency as predictors of cardiometabolic health outcomes, Sleep Med. 13 (2012) 1261–1270, http://dx.doi.org/10.1016/j.sleep.2012.08.005.

[3] S. Garbarino, P. Durando, O. Guglielmi, G. Dini, F. Bersi, S. Fornarino, A. Toletone, C. Chiorri, N. Magnavita, Sleep apnea, sleep debt and daytime sleepiness are independently associated with road accidents. a cross-sectional study on truck drivers, PLoS One 11 (2016) e0166262, http://dx.doi.org/10.1371/journal.pone.0166262.

[4] H. Noland, J.H. Price, J. Dake, S.K. Telljohann, Adolescents' sleep behaviors and perceptions of sleep, J. Sch. Health 79 (2009) 224–230, http://dx.doi.org/10.1111/j.1746-1561.2009.00402.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1746-1561.2009.00402.x.

[5] K.M. Orzech, D.B. Salafsky, L.A. Hamilton, The state of sleep among college students at a large public university, J. Am. Coll. Health 59 (2011) 612–619, http://dx.doi.org/10.1080/07448481.2010.520051.

[6] N. Digdon, K. Landry, University students' motives for drinking alcohol are related to evening preference, poor sleep, and ways of coping with stress, Biol. Rhythm Res. 44 (2013) 1–11, http://dx.doi.org/10.1080/09291016.2011.632235.

[7] M. Gardani, D.R. Bradford, K. Russell, S. Allan, L. Beattie, J.G. Ellis, U. Akram, A systematic review and meta-analysis of poor sleep, insomnia symptoms and stress in undergraduate students, Sleep Med. Rev. 61 (2022) 101565, http://dx.doi.org/10.1016/j.smrv.2021.101565.

[8] J. Deng, F. Zhou, W. Hou, Z. Silver, C.Y. Wong, O. Chang, A. Drakos, Q.K. Zuo, E. Huang, The prevalence of depressive symptoms, anxiety symptoms and sleep disturbance in higher education students during the covid-19 pandemic: A systematic review and meta-analysis, Psychiatry Res. 301 (2021) 113863, http://dx.doi.org/10.1016/j.psychres.2021.113863, URL: https://www.sciencedirect.com/science/article/pii/S0165178121001608.

[9] S. Haghayegh, S. Khoshnevis, M.H. Smolensky, K.R. Diller, R.J. Castriotta, Accuracy of wristband fitbit models in assessing sleep: Systematic review and meta-analysis, J. Med. Internet Res. 21 (2019) e16273, http://dx.doi.org/10.2196/16273, URL: http://www.jmir.org/2019/11/e16273/.

[10] H.A. Lee, H.J. Lee, J.H. Moon, T. Lee, M.G. Kim, H. In, C.H. Cho, L. Kim, Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults, Psychiatry Investig. 14 (2017) 179, http://dx.doi.org/10.4306/pi.2017.14.2.179.

[11] M. de Zambotti, F.C. Baker, A.R. Willoughby, J.G. Godino, D. Wing, K. Patrick, I.M. Colrain, Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents, Physiol. Behav. 158 (2016) 143–149, http://dx.doi.org/10.1016/j.physbeh.2016.03.006.

[12] J.G. Godino, D. Wing, M. de Zambotti, F.C. Baker, K. Bagot, S. Inkelis, C. Pautz, M. Higgins, J. Nichols, T. Brumback, et al., Performance of a commercial multi-sensor wearable (fitbit charge hr) in measuring physical activity and sleep in healthy children, PLoS One 15 (2020) e0237719, http://dx.doi.org/10.1371/journal.pone.0237719.

[13] S. Vhaduri, C. Poellabauer, Impact of different pre-sleep phone use patterns on sleep quality, in: 2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks, BSN, 2018, pp. 94–97, http://dx.doi.org/10.1109/BSN.2018.8329667.

[14] J. Stutz, R. Eiholzer, C.M. Spengler, Effects of evening exercise on sleep in healthy participants: a systematic review and meta-analysis, Sports Med. 49 (2019) 269–287.

[15] M.P. St-Onge, A. Mikic, C.E. Pietrolungo, Effects of diet on sleep quality, Adv. Nutr. 7 (2016) 938–949, http://dx.doi.org/10.3945/an.116.012336.

[16] W. Cao, M.W. Milks, X. Liu, M.E. Gregory, D. Addison, P. Zhang, L. Li, Mhealth interventions for self-management of hypertension: Framework and systematic review on engagement, interactivity, and tailoring, JMIR Mhealth Uhealth 10 (2022) e29415, http://dx.doi.org/10.2196/29415, URL: https://mhealth.jmir.org/2022/3/e29415.

[17] Z. Wang, H. Xiong, J. Zhang, S. Yang, M. Boukhechba, D. Zhang, L.E. Barnes, D. Dou, From personalized medicine to population health: A survey of mhealth sensing techniques, IEEE Internet Things J. 9 (2022) 15413–15434, http://dx.doi.org/10.1109/JIOT.2022.3161046.

[18] A. Hombali, E. Seow, Q. Yuan, S.H.S. Chang, P. Satghare, S. Kumar, S.K. Verma, Y.M. Mok, S.A. Chong, M. Subramaniam, Prevalence and correlates of sleep disorder symptoms in psychiatric disorders, Psychiatry Res. 279 (2019) 116–122, http://dx.doi.org/10.1016/j.psychres.2018.07.009, URL: https://www.sciencedirect.com/science/article/pii/S0165178118302683.

[19] A.A. Alghwiri, F. Almomani, A.A. Alghwiri, S.L. Whitney, Predictors of sleep quality among university students: the use of advanced machine learning techniques, Sleep Breath. 25 (2021) 1119–1126, http://dx.doi.org/10.1007/s11325-020-02150-w.

[20] H. Sanchez-Trigo, E. Molina-Martínez, M. Grimaldi-Puyana, B. Sañudo, Effects of lifestyle behaviours and depressed mood on sleep quality in young adults. a machine learning approach, Psychol. Health (2022) 1–16, http://dx.doi.org/10.1080/08870446.2022.2067331, pMID: 35475409.

[21] A. Sathyanarayana, S. Joty, L. Fernandez-Luque, F. Ofli, J. Srivastava, A. Elmagarmid, T. Arora, S. Taheri, Sleep quality prediction from wearable data using deep learning, JMIR Mhealth Uhealth 4 (2016) e125, http://dx.doi.org/10.2196/mhealth.6562, URL: http://mhealth.jmir.org/2016/4/e125/.

[22] M. Khalid, E.B. Klerman, A.W. McHill, A.J.K. Phillips, A. Sano, Sleepnet: Attention-enhanced robust sleep prediction using dynamic social networks, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 8 (2024) http://dx.doi.org/10.1145/3643508.

[23] A. Fellger, G. Sprint, D. Weeks, E. Crooks, D.J. Cook, Wearable device-independent next day activity and next night sleep prediction for rehabilitation populations, IEEE J. Transl. Eng. Health Med. 8 (2020) 1–9, http://dx.doi.org/10.1109/JTEHM.2020.3014564.

[24] O. Kilic, B. Saylam, O. Durmaz Incel, Sleep quality prediction from wearables using convolution neural networks and ensemble learning, in: Proceedings of the 2023 8th International Conference on Machine Learning Technologies, Association for Computing Machinery, New York, NY, USA, 2023, pp. 116–120, http://dx.doi.org/10.1145/3589883.3589900.

[25] M. Hirshkowitz, K. Whiton, S.M. Albert, C. Alessi, O. Bruni, L. DonCarlos, N. Hazen, J. Herman, E.S. Katz, L. Kheirandish-Gozal, et al., National sleep foundation's sleep time duration recommendations: Methodology and results summary, Sleep Health 1 (2015) 40–43, http://dx.doi.org/10.1016/j.sleh.2014.12.010.

[26] W. Hidayat, T.D. Tambunan, R. Budiawan, Empowering wearable sensor generated data to predict changes in individual's sleep quality, in: 2018 6th International Conference on Information and Communication Technology, ICoICT, 2018, pp. 447–452.

[27] D.V. Phan, C.L. Chan, D.K. Nguyen, Applying deep learning for prediction sleep quality from wearable data, in: Proceedings of the 4th International Conference on Medical and Health Informatics, Association for Computing Machinery, New York, NY, USA, 2020, pp. 51–55, http://dx.doi.org/10.1145/3418094.3418114.

[28] Y. Kim, S. Vhaduri, C. Poellabauer, Comprensión de los comportamientos de los estudiantes universitarios en materia de llamadas telefónicas para una solución sostenible de salud y bienestar móvil, Actas Congr. Int. Ingen. Sistemas 1 (2021) 7–24, http://dx.doi.org/10.26439/ciis2020.5517, URL: https://revistas.ulima.edu.pe/index.php/CIIS/article/view/5517.

[29] R. Purta, S. Mattingly, L. Song, O. Lizardo, D. Hachen, C. Poellabauer, A. Striegel, Experiences measuring sleep and physical activity patterns across a large college cohort with fitbits, in: Proceedings of the 2016 ACM International Symposium on Wearable Computers, 2016, pp. 28–35, http://dx.doi.org/10.1145/2971763.2971767.

[30] M. Jetté, K. Sidney, G. Blümchen, Metabolic equivalents (mets) in exercise testing, exercise prescription, and evaluation of functional capacity, Clin. Cardiol. 13 (1990) 555–565, http://dx.doi.org/10.1002/clc.4960130809.

[31] D.J. Buysse, C.F. Reynolds III, T.H. Monk, S.R. Berman, D.J. Kupfer, The pittsburgh sleep quality index: a new instrument for psychiatric practice and research, Psychiatry Res. 28 (1989) 193–213, http://dx.doi.org/10.1016/0165-1781(89)90047-4.

[32] L.S. Radloff, The use of the center for epidemiologic studies depression scale in adolescents and young adults, J. Youth Adolesc. 20 (1991) 149–166, http://dx.doi.org/10.1007/BF01537606.

[33] R.M. Ryan, J.P. Connell, Perceived locus of causality and internalization: examining reasons for acting in two domains, J. Pers. Soc. Psychol. 57 (1989) 749, http://dx.doi.org/10.1037//0022-3514.57.5.749.

[34] A.T. Beck, R.A. Steer, M.G. Carbin, Psychometric properties of the beck depression inventory: Twenty-five years of evaluation, Clin. Psychol. Rev. 8 (1988) 77–100, http://dx.doi.org/10.1016/0272-7358(88)90050-5.

[35] A.T. Beck, N. Epstein, G. Brown, R. Steer, Beck anxiety inventory, 1988, PsycTESTS Dataset.

[36] D.A. Cobb-Clark, S. Schurer, The stability of big-five personality traits, Econom. Lett. 115 (2012) 11–15, http://dx.doi.org/10.1016/j.econlet.2011.11.015.

[37] E. DiTommaso, B. Spinner, The development and initial validation of the social and emotional loneliness scale for adults (selsa), Pers. Individ. Differ. 14 (1993) 127–134, http://dx.doi.org/10.1016/0191-8869(93)90182-3.

[38] J. Horne, O. Ostberg, A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms, Int. J. Chronobiology 4 (1976) 97–110, URL: http://europepmc.org/abstract/MED/1027738.

[39] C.D. Spielberger, State-trait anxiety inventory for adults, 1983, PsycTESTS Dataset.

[40] F. Gu, Q. Xiao, L.W. Chu, K. Yu, C.E. Matthews, A.W. Hsing, N.E. Caporaso, Sleep duration and cancer in the nih-aarp diet and health study cohort, PLoS One 11 (2016) http://dx.doi.org/10.1371/journal.pone.0161561.

[41] F.P. Cappuccio, L. D'Elia, P. Strazzullo, M.A. Miller, Quantity and quality of sleep and incidence of type 2 diabetes, Diabetes Care 33 (2009) 414–420, http://dx.doi.org/10.2337/dc09-1124.

[42] K.L. Knutson, Sleep duration and cardiometabolic risk: A review of the epidemiologic evidence, Best Pract. Res. Clin. Endocrinol. Metab. 24 (2010) 731–743, http://dx.doi.org/10.1016/j.beem.2010.07.001.

[43] O.M. Buxton, E. Marcelli, Short and long sleep are positively associated with obesity, diabetes, hypertension, and cardiovascular disease among adults in the United States, Soc. Sci. Med. 71 (2010) 1027–1036, http://dx.doi.org/10.1016/j.socscimed.2010.05.041.

[44] D. Leger, J.B. Richard, O. Collin, F. Sauvet, B. Faraut, Napping and weekend catchup sleep do not fully compensate for high rates of sleep debt and short sleep at a population level (in a representative nationwide sample of 12 637adults), Sleep Med. 74 (2020) 278–288, http://dx.doi.org/10.1016/j.sleep.2020.05.030.

[45] P.M. Granitto, C. Furlanello, F. Biasioli, F. Gasperi, Recursive feature elimination with random forest for ptr-ms analysis of agroindustrial products, Chemometr. Intell. Lab. Syst. 83 (2006) 83–90, http://dx.doi.org/10.1016/j.chemolab.2006.01.007.

[46] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[47] S. Roccas, L. Sagiv, S.H. Schwartz, A. Knafo, The big five personality factors and personal values, Pers. Soc. Psychol. Bull. 28 (2002) 789–801, http://dx.doi.org/10.1177/0146167202289008.

[48] C. Rahe, M.E. Czira, H. Teismann, K. Berger, Associations between poor sleep quality and different measures of obesity, Sleep Med. 16 (2015) 1225–1228, http://dx.doi.org/10.1016/j.sleep.2015.05.023.
[49] M. Fabbri, A. Beracci, M. Martoni, D. Meneo, L. Tonetti, V. Natale, Measuring subjective sleep quality: a review, Int. J. Environ. Res. Public Health 18 (2021) 1082, http://dx.doi.org/10.3390/ijerph18031082.
[50] N. Breslau, T. Roth, L. Rosenthal, P. Andreski, Sleep disturbance and psychiatric disorders: a longitudinal epidemiological study of young adults, Biological psychiatry 39 (1996) 411–418, http://dx.doi.org/10.1016/0006-3223(95)00188-3.
[51] M.A. Al-Abri, Sleep deprivation and depression: A bi-directional association, Sultan Qaboos Univ. Med. J. 15 (2015) e4.
[52] W. Su, X. Han, H. Yu, Y. Wu, M.N. Potenza, Do men become addicted to internet gaming and women to social media? a meta-analysis examining gender-related differences in specific internet addiction, Comput. Hum. Behav. 113 (2020) 106480, http://dx.doi.org/10.1016/j.chb.2020.106480.
[53] D. Blackwell, C. Leaman, R. Tramposch, C. Osborne, M. Liss, Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction, Pers. Individ. Differ. 116 (2017) 69–72, http://dx.doi.org/10.1016/j.paid.2017.04.039.
[54] T. Bowden-Green, J. Hinds, A. Joinson, How is extraversion related to social media use? a literature review, Pers. Individ. Differ. 164 (2020) 110040, http://dx.doi.org/10.1016/j.paid.2020.110040.
[55] R. Garett, S. Liu, S.D. Young, The relationship between social media use and sleep quality among undergraduate students, Inf. Commun. Soc. 21 (2018) 163–173, http://dx.doi.org/10.1080/1369118X.2016.1266374.
[56] Z. Križan, G. Hisler, Personality and sleep: Neuroticism and conscientiousness predict behaviourally recorded sleep years later, Eur. J. Pers. 33 (2019) 133–153, http://dx.doi.org/10.1002/per.2191.
[57] Y. Stephan, A.R. Sutin, S. Bayard, Z. Križan, A. Terracciano, Personality and sleep quality: Evidence from four prospective studies, Health Psychol. 37 (2018) 271–281, https://psycnet.apa.org/doi/10.1037/hea0000577.
[58] B. Adenekan, A. Pandey, S. McKenzie, F. Zizi, G.J. Casimir, G. Jean-Louis, Sleep in america: role of racial/ethnic differences, Sleep Med. Rev. 17 (2013) 255–262, http://dx.doi.org/10.1016/j.smrv.2012.07.002.
[59] L. Tomfohr, M.A. Pung, K.M. Edwards, J.E. Dimsdale, Racial differences in sleep architecture: the role of ethnic discrimination, Biol. Psychol. 89 (2012) 34–38, http://dx.doi.org/10.1016/j.biopsycho.2011.09.002.
[60] H. Fritz, M. Tang, K. Kinney, Z. Nagy, Evaluating machine learning models to classify occupants' perceptions of their indoor environment and sleep quality from indoor air quality, J. Air Waste Manage. Assoc. 72 (2022) 1381–1397, http://dx.doi.org/10.1080/10962247.2022.2105439, pMID: 35939653.