

Article

A Smart Health (sHealth) Centric Method Toward Estimation of Sleep Deficiency Severity from Wearable Sensor Data Fusion

Md Juber Rahman ¹, Bashir I. Morshed^{2, *} and Chrysanthé Preza ¹

¹ Electrical and Computer Engineering Department, The University of Memphis

² Department of Computer Science, Texas Tech University; bmorshed@ttu.edu

* Correspondence: bmorshed@ttu.edu; Tel.: +1-806-834-4898

Abstract: Sleep deficiency impacts the quality of life and may have serious health consequences in the long run. Questionnaire-based subjective assessment of sleep deficiency has many limitations. On the other hand, objective assessment of sleep deficiency is challenging. In this study, we propose a polysomnography-based mathematical model for computing baseline sleep deficiency severity score and then investigated the estimation of sleep deficiency severity using features available only from wearable sensor data including heart rate variability and single-channel electroencephalography for a dataset of 500 subjects. We used Monte-Carlo Feature Selection (MCFS) and inter-dependency discovery for selecting the best features and removing multi-collinearity. For developing the Regression model we investigated both the frequentist and the Bayesian approaches. An Artificial Neural Network achieved the best performance of RMSE = 5.47 and an R-squared value of 0.67 for sleep deficiency severity estimation. The developed method is comparable to conventional methods of Functional Outcome of Sleep Questionnaire and Epworth Sleepiness Scale for assessing the impact of sleep apnea on sleep deficiency. Moreover, the results pave the way for reliable and interpretable sleep deficiency severity estimation using a wearable device.

Keywords: Artificial Neural Network, Bayesian Regression, Electroencephalography, Monte-Carlo Method, Regression, Sleep deficiency severity, Smart Health, Wearable Sensor

1. Introduction

Sleep is an important biological process and plays a key role in restoring energy, solidifying and consolidating memories, and repairing body cells. It also helps in metabolism and cardiovascular function [1]. The regulation of sleep is controlled by the circadian biological clock and sleep/wake homeostasis. With the rise of obesity, excessive usage of personal gadgets, rapid urbanization, and other socio-economic changes sleep/wake homeostasis is getting impacted disrupting the normal circadian rhythm and healthy sleep. Good quality sleep is essential for good health and improved quality of life. Neither body nor the brain can function properly without enough sleep. A previous study suggests that complete sleep deprivation impairs attention and working memory [2]. Moreover, it also affects other functions, such as long-term memory and decision-making. Even partial sleep deprivation negatively impacts attention, and vigilance in the long run [3]. Moreover, Sleep deficiency can lead to physical and mental health problems, injuries, loss of productivity, even greater risk of life-threatening diseases [4-5].

The common practice to evaluate sleep health is to use standard questionnaires along with a sleep test where a polysomnogram is captured. The questionnaire-based approach has many limitations including high bias, long evaluation period, etc. Polysomnogram is expensive, has limited availability, and less user-friendly. In recent years, there has been a significant expansion in the development and use of multi-modal sensors and technologies to monitor sleep which includes sleep patterns monitoring, wellness ap-

plications, sleep coaching of individuals with chronic conditions, etc. [6]. Kuo *et al.* developed an actigraphy-based wearable device for sleep quality assessment [7]. Mendonca *et al.* proposed a method for sleep quality estimation using electrocardiogram by cardiopulmonary coupling analysis [8]. Azimi *et al.* reported an objective IoT-based longitudinal study for sleep quality assessment [9]. They estimated the sleep quality average to classify sleep into good or poor quality. Bsoul *et al.* developed a Sleep Efficiency Index based on ECG features using Support Vector Machines [10]. Additionally, some commercial devices e.g. the Fitbit Charge smart band (Fitbit Inc., USA), the Apple Watch (Apple Inc., USA), the Oura sleep ring (Oura Health Ltd., Finland) have attempted the estimation of sleep scores from non-polysomnographic measures. However, most of the previous approaches were focused on sleep quality assessment or sleep score estimation. Sleep deficiency provides a more specific evaluation of sleep disorder than sleep score or the sleep quality, as sleep deficiency includes lack of enough sleep (sleep deprivation), not getting all types of sleep that the human body needs, and poor quality of sleep [11]. Early detection of sleep deficiency is beneficial to avoid many linked chronic health problems, including heart disease, kidney disease, high blood pressure, diabetes, stroke, obesity, and depression. A method for objective assessment of sleep deficiency from wearable sensor data only is not well-established yet requiring further investigation.

The main objective of this work was to develop a physiological sensor data-based objective sleep deficiency assessment method that can be integrated with user-friendly wearables e.g. smartwatch, smart band, etc. We proposed a mathematical model to facilitate a quantitative evaluation of sleep deficiency based on polysomnogram features. Then we addressed the same problem of estimating sleep deficiency severity when polysomnogram data is not available, with a machine learning-driven model using ECG/EEG based features that can be captured by wearables. The estimated sleep deficiency severity using the machine learning-based method has been validated against the ground truth established from polysomnography measurements. The results pave the way for automated sleep deficiency severity assessment using single-channel EEG.

2. Materials and Methods

2.1.1 Smart Health (sHealth) Framework

We define “Smart Health (sHealth)” as a system that uses embedded artificial intelligence such as edge computing, machine learning, etc. and aims to deliver improved healthcare using users’ smart devices, wearables, and the Internet of Things (IoT) centric solutions. It not only benefits and monitors individual user’s health but also collects spatiotemporal community-wide data for collective and social well-being and informed policy makings. It is an emerging paradigm for efficient processing, sharing, and visualization of healthcare data, which is coming from different IoT devices, and wearable sensors. sHealth can be perceived as an upgraded and extended version of Mobile Health (mHealth). We previously reported a framework for sHealth and conducted a pilot study to evaluate the technical feasibility of the framework [12]. The system architecture for the sHealth framework is shown in Fig. 1. The main components in the framework are various sensors, such as battery-less body-worn passive sensors with a scanner (i.e. reader for the passive sensors), commercial wearables, a custom smartphone app (SCC-Health app), and a custom web server (SCC-Health server). Details of the design and functionality of the sensors and scanner can be found elsewhere [13]. For physiological data collection, we utilized novel inkjet-printed (IJP) sensors in addition to commercial wearables such as a smart wristband (Mi Band 2, Xiaomi) and a fingertip pulse oximeter (CMS50E, Crucial Medical Systems) [14]. The IJP sensors were zero-power, analog, wireless, and fully passive. Data collected in the IJP sensors were pre-processed and digitized by a custom-made scanner. Data from

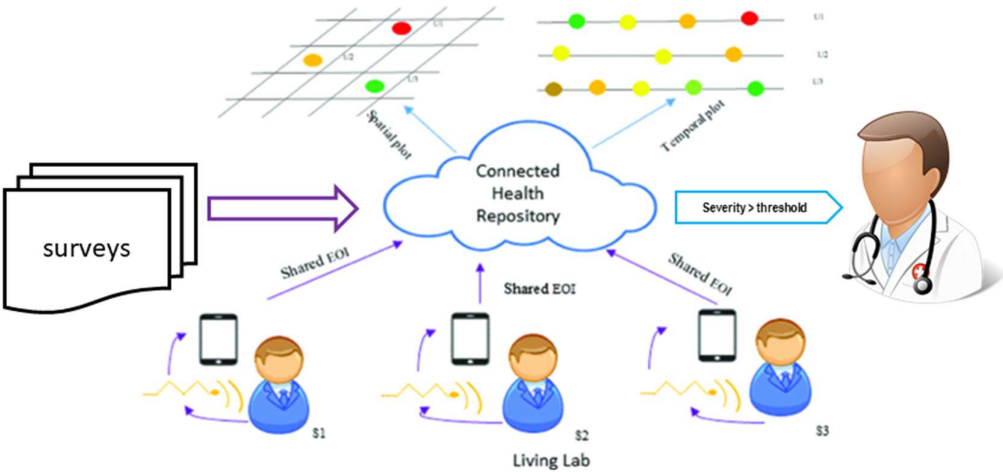


Fig. 1 Workflow for Events of interest capture and spatiotemporal visualization in sHealth

the scanner is transmitted to the smartphone via Bluetooth. Data reliability check, feature extraction, and classification or regression analysis using pre-trained machine learning models are performed in the smartphone for disease detection and severity assessment. The computed severity of the disease is then visualized in the smartphone as well as shared with the webserver using Wi-Fi (or cellular network) for observing the temporal and spatial distribution of the diseases. The webserver is accessible at <http://sccmobilehealth.com>. The app was developed using Android studio 3.1 with build tool version 25.0.2 and the minimum SDK level of 19. The pre-trained machine learning models integrated with the app for Events of Interest (EOI) computation were developed and evaluated in WEKA. Additionally, electronic and paper-based surveys were conducted to collect user-reported symptoms and user feedback [14].

The process of spatiotemporal visualization is fully automated and near real-time. JavaScript object notation (JSON) has been used to share data from the android smartphone to the database in the webserver [15]. The shared data contains participants' anonymized user ID, area code (hashed), computed EoI, the algorithm name used for EoI computation, and timestamp of data collection. For personalized monitoring of diseases, temporal trends of disease severity for a participant can be visualized using a time plot graph [16]. Flow graph has been used for community health trend monitoring over. In addition to that, a spatial plot was used to visualize the severity of the disease in different areas at a period interval (averaged) [16]. Color coding has been used to indicate severity where red indicates the highest severity and green indicates the lowest severity.

2.1.2 Sleep Health Assessment Overview

As shown in Fig.2 conventional methods of sleep health assessment fall under two broad categories- subjective assessment and objective assessment of sleep. Subjective assessment of sleep deficiency using standard questionnaires is well investigated and is widely used in clinical practice. Some of the well-accepted and popular methods for subjective sleep quality assessment are- the Pittsburgh Sleep Quality Index (PSQI), the Epworth Sleepiness Scale (ESS), and the Functional Outcome of Sleep Questionnaire (FOSQ).

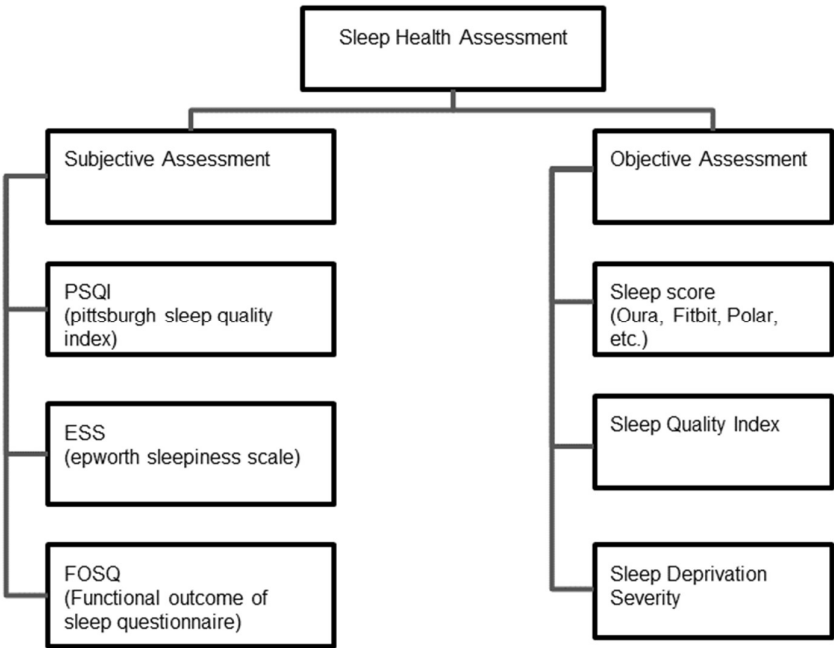


Fig. 2 Subjective and objective methods for sleep health assessments

PSQI uses a 7 component questionnaire and the subject assigns a score of 0-3 for each component [17]. The components are - subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction. A global score >5 indicates poor sleep quality. FOSQ has 21 questions related to activity levels, vigilance, intimacy and relationships, general productivity, and social outcomes [18]. The potential range of scores for each subscale is 1 – 4 with higher scores indicating greater insomnia severity. Similarly, in ESS the subject assigns a score of 0-3 for 8 questions aimed at assessing daytime sleepiness. A total score of 16-24 indicates excessive daytime sleepiness suggesting the need for medical attention [19]. The Karolinska Sleep Diary (KSD) is another questionnaire that was developed to assess subjective sleep quality [20].

Subjective reports of sleep quality are important in the clinical setting and can help determine whether further screening and/or treatment for a sleep complaint might be warranted [21]. However, subjective methods suffer from high bias, require active user participation, and a longer period (2 weeks - 1 month) for sleep deficiency evaluation. Objective sleep quality consists not only of the total duration of sleep, but also of the architecture of sleep (amount of the different sleep stages across the sleep episode), the amount of wake time during the sleep episode, and the frequency and duration of awakenings across the night [22]. Prominent quantitative metrics that are used for objective sleep assessment are- the Sleep Quality/Efficiency Index and the Sleep Score.

Sleep Quality/Efficiency Index (SQI) –Several quantitative metrics have been developed to measure the quality of sleep from physiological sensor data. However, there is still a lack of standard and well-established definitions for the term ‘Sleep Quality’ [23]. Usually, it is used to refer to a score computed from the collection of quantitative sleep measures i.e. sleep duration, sleep onset time, degree of fragmentation, etc. [24].

Sleep Score- The concept of sleep score has been introduced mainly by commercial entities i.e. Fitbit, Polar, Oura, Apple, etc. Sleep score is usually tracked by smartphone apps based on data collected using a smart band or a smartwatch during sleep. There is no well-accepted and standard definition for Sleep score. Fitbit computes overall sleep score as a sum of individual scores in sleep duration, sleep quality, and restoration, for a total score of up to 100. Most people have a sleep score between 72 and 83. Sleep score ranges

are: Excellent: 90-100, Good: 80-89, Fair: 60-79, Poor: Less than 60 [25]. A previous validation study showed that Fitbit smart bands performance is promising in detecting sleep-wake states and sleep stage composition relative to gold standard polysomnogram [26]. Oura sleep ring measures sleep using sensors that capture the body signals including resting heart rate (RHR), heart rate variability (HRV), body temperature, respiratory rate, and movement, to determine sleep patterns and compute the sleep score [27]. Sleep score-based monitoring has been criticized due to a lack of consistency in measurement and the impact of a sleep-related disease on sleep score is not well-investigated [28]. Scientific study suggests that current consumer sleep tracking technologies are immature for diagnosing sleep disorders, but they may be reasonably satisfactory for general purpose and non-clinical use [29].

Sleep Deficiency Severity (SDS)- A new metric proposed by the authors aimed at pre-clinical early evaluation of sleep deficiency and is based on a fusion of features from ECG, EEG, SpO₂, and other wearable sensors. Details of the metric are described in the following sections.

2.1.2 Dataset

Sleep Health Heart Study (SHHS) is a dataset available from the National Sleep Research Resource [30]. SHHS was implemented as a multi-center cohort study in two phases by the US National Heart Lung & Blood Institute. Unattended home polysomnograms were obtained for both the phases of SHHS by certified and trained technicians. The polysomnogram data was saved in European Data Format (EDF). Data processing and initial scoring were done using Compumedics software (Compumedics Ltd., Australia) as part of SHHS. Two manual scorings were included to annotate the database with sleep duration, sleep efficiency, arousal index, sleep stages, oxygen saturation level, etc. A dataset of 500 subjects containing good quality data for both ECG and EEG is available from the dataset provider and is recommended for use in a research study. In our study, for developing the regression models we used this dataset of 500 subjects. The gender distribution of records in the dataset is as follows: male- 231, female- 269. The age of the subjects ranges from 44 to 89 years old with a mean of 65 years old and a standard deviation of 10.41 years. The body mass index (BMI) of the subjects ranges from 18 – 46 with a mean of 27.51 kilograms per square meter and a standard deviation of 4.11 kilograms per square meter.

2.1.3 Mathematical Model for Baseline SDS Score

Guidelines for computing a composite sleep health score from polysomnographic measures have been developed and reported in previous research studies [31-32]. In this study, we used a generalized mathematical model for computing the baseline SDS score. The model has been described by equation (1).

$$\text{Sleep Deprivation Severity} = \frac{1}{m+n} \{ \sum_{i=1}^m X_{neg(i)} - \sum_{j=1}^n X_{pos(j)} \} \times 100 \quad (1)$$

where X_{neg} is the sleep attributes that increases sleep deficiency, i.e. higher is responsible for more sleep deficiency, X_{pos} is the sleep attributes that reduce sleep deficiency, i.e. higher is better. m is the total number of negative attributes and n is the total number of positive attributes. The positive attributes available from SHHS dataset are as follows:

Sleep time- Duration of entire sleep.

Sleep efficiency - Percentage of time in bed that was spent sleeping, or the ratio of total sleep time to total time in bed, expressed as a percentage.

Time deep sleep (%) - Percent time in sleep stages 3 and 4.

Time REM sleep (%) - Percent Time in rapid eye movement sleep (REM).

SpO₂ (%) - Average oxygen saturation (SpO₂) level in sleep.

The negative attributes available from SHHS is as follows:

Sleep Fragmentation Index (SFI)- Total number of arousals per hour of sleep i.e. ratio of the count of arousals to total sleep time in hours.

In computing the sleep deficiency severity, all the attributes have been normalized on a scale of 0-1. To achieve a consistent “higher is better” rule the value of each negative attribute is subtracted from 1. Then, the attribute values have been summed up to develop a composite score. The composite score has been multiplied by 100 and divided by the total no. of positive and negative attributes to obtain the SDS in the range of 0-100. Statistical analysis has been conducted to investigate the relationship of baseline SDS with age, gender, and BMI. Additionally, the partial correlation of SDS (controlling for age and BMI) with HRV and EEG features has been investigated.

2.1.4 Machine-learning driven method for SDS estimation

For developing the machine learning-based regression model we extracted features from wearable sensor data, used Monte Carlo Feature Selection for selecting the best features, and explored machine learning and deep learning methods as described below:

Feature Extraction - The recording montage for polysomnograms consisted of data from 14 channels which include- ECG, EEG, electrooculogram (EOG), Electromyogram (EMG), nasal airflow, thoracic and abdominal movement signal, SpO2, sleep hypnogram, etc. Hardware filters have been used for preliminary noise reduction. The cutoff frequency for hardware filters had been as follows: ECG-0.15 Hz, EOG-0.15 Hz, EMG-0.15 Hz, EEG-0.15 Hz, Thoracic respiration signal of 0.05 Hz, Abdominal respiration signal-0.05 Hz. The sampling rate is 125 Hz for EEG, ECG, and EMG signals. For EOG, the sampling rate is 50 Hz. In investigating a minimalistic approach, we considered the use of features from ECG, EEG, SpO2 signals considering the sensors are more user-friendly and widely used. For RR interval correction we used *malik's* rule followed by a cubic interpolation for the determination of Normal-to-Normal (NN) intervals [33].

From the NN interval series time domain and frequency domain features have been extracted following the HRV guidelines using the HRV Toolkit available from Physionet [34]. For the power spectrum estimation, we used Lomb's periodogram method. The entire ECG record has been divided into 5-minute epochs to estimate short-term components of HRV. In total 20 HRV features have been extracted. From EEG, we computed spectral features as described in Table 1. The EEG signal was collected using two channels from the central region of the brain. One of the channels was C4-A1 and the other one was C3-A2. The power spectral densities for these two channels are very similar. In our study, we have only used the signal from the C4 channel as it has been designated as the primary EEG channel in the movement (NREM) power, and Total power at each frequency band. Also, 102 EEG spectral features i.e. REM, N-REM power at single frequencies have been computed for 51 frequencies from 0 to 25 Hz with a 0.5 Hz gap i.e. 0 Hz, 0.5 Hz, 1 Hz, 1.5 Hz, ..., 24.5 Hz, 25 Hz.

Table 1. Spectral Features Extracted from EEG

EEG Band	Frequency (Hz)	Features
Slow OSC	0.5 -1	Power- REM , NREM, Total
Delta	0.5 – 4	Power- REM , NREM, Total
Theta	4 – 8	Power- REM , NREM, Total
Alpha	8 - 13	Power- REM , NREM, Total
Sigma	12 – 14	Power- REM , NREM, Total
Beta	13 – 30	Power- REM , NREM, Total
Gamma	36 – 90	Power- REM , NREM, Total

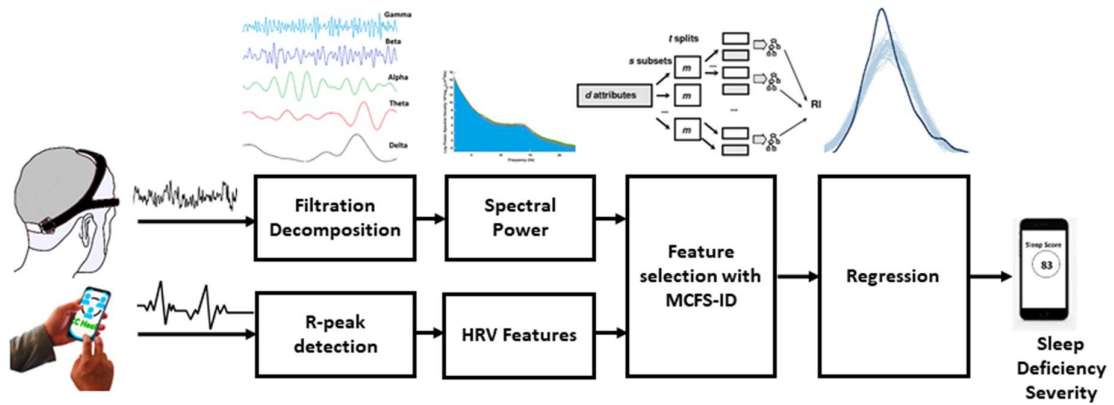


Fig. 3 Method for feature extraction, feature selection, and regression for sleep deficiency severity

Monte Carlo Feature Selection and Inter-dependency Discovery- Feature selection has been done primarily to compare the relative importance of ECG and EEG-based features for SDS estimation. Monte-Carlo Feature Selection (MCFS) and inter-dependency discovery has been used for ranking the feature importance. In MCFS the relative importance of features is estimated by building hundreds of trees for a randomly selected subset of features [35]. In a mathematic notion, i subsets of m randomly selected features are constructed where $m \ll n$, n being the total number of features and for each subset, k trees are constructed and their performance is assessed for classification/ regression. Finally, $i \times m$ trees are constructed and evaluated. The procedure has been illustrated in Fig. 3. Weighted accuracy of a tree as defined by eq. (2) is used as a metric to assess the classification or regression ability of the tree.

$$Wac = \frac{1}{c} \sum_{i=1}^c \frac{n_{ij}}{n_{i1} + n_{i2} + \dots + n_{ic}} \quad (2)$$

Where c = number of classes, $i, j = 1, 2, \dots, c$;

n_{ij} is the number of samples from class i classified as class j

$\sum n_{ij} = n$ is the number of all samples

The Relative Importance (RI) of feature g_d denoted by RI_{gd} is defined by eq. (3)

$$RI_{gd} = \sum_{x=1}^{m,k} Wac_x^u \sum_{r_{gd}(x)} IG(r_{gd}(x)) \left(\frac{\text{no. in } r_{gd}(x)}{\text{no. in } x} \right)^v \quad (3)$$

Where Wac stands for the weighted accuracy for x^{th} tree, $IG(r_{gd}(x))$ stands for the Information Gain for node $r_{gd}(x)$, $(\text{no. in } r_{gd}(x))$ denotes the number of samples in the node $r_{gd}(x)$, $(\text{no. in } x)$ denotes the number of samples in the root of the x^{th} tree, and u and v are fixed positive reals. Information Gain (IG) is measured by Gini Index or Gain Ratio [36].

Both ECG and EEG have correlated features that introduce the problem of multicollinearity. To deal with this, the inter-dependency discovery was used to remove features with strong pairwise interactions. *rmcfs* package from R has been used for feature ranking using Monte-Carlo Feature Selection and Interdependency Discovery (MCFS-ID) method [36]. The steps of preprocessing, feature extraction, feature selection, and regression has been shown in Fig. 3.

Regression Model- For developing the regression model, we investigated the Bayesian Regression method, and Artificial Neural Network (ANN). Bayesian inference facilitates overcoming insufficient data or poorly distributed data as it allows to put a prior on the coefficients and the noise so that in the absence of data, the priors can take over. In a Bayesian framework, the regression model is stated in a probabilistic manner where the Bayesian sampling algorithm returns a probability distribution (known as the posterior of

the effect) that is compatible with the observed data instead of a point estimate. The posterior distribution is obtained by the product of the prior distribution and the likelihood function. The model for Bayesian Linear Regression can be represented by eq. (4)

$$y \sim N(W^T X, \sigma^2 I) \quad (4)$$

Where response data points y is sampled from a multivariate Gaussian distribution that has a mean equal to the product of W coefficients and the predictors X and variance of σ^2 . I – is the $N \times N$ Identity matrix [37]. In this work, we used Markov Chain Monte Carlo Method (MCMC) sampling and weakly informative prior for Bayesian regression. To verify convergence potential scale reduction statistic R-hat was used [38].

Artificial Neural Network (ANN) is capable of approximating any linear or non-linear relationship including multi-dimensional regression mapping problems quite well. However, the ANN must have enough neurons in the hidden layers and the data distribution should be consistent. During the training process, an ANN fits a function on a set of inputs to produce a set of associated outputs. Once training is finished the network forms a generalization of the input-output relationship and can be utilized to generate outputs for unseen inputs. The structure of ANN has multiple layers with interconnected artificial neurons as the building blocks for each layer. Each neuron has weights that are adjusted during the training process. Training stops when any of these conditions occur: the maximum number of epochs (repetitions) is reached, or the maximum amount of time is exceeded or performance is minimized to the goal or the performance gradient falls below minimum gradient. The ANN used in this study is of feed-forward type and has 3 layers- input, output, and hidden layer. The number of neurons in each layer is input-117, hidden- 10, output-1. The used activation functions are- *relu* for the hidden layer and *softmax* for the output layer. Levenberg-Marquardt optimization with backpropagation was used as the training algorithm [39]. The hyperparameters used for the ANN are as follows: *max epochs* = 1000, *min gradient* = 1e-7, *momentum (Mu)* = 0.001, *Mu decrease ratio* = 0.1, *Mu increase ratio* = 0.1. To facilitate proper training and evaluation the input data was randomly divided into training (80%), and test set (20%). Root Mean Squared Error (RMSE) and R-squared (R^2) values were used for performance evaluation of both the Bayesian model and ANN. Additionally, Pareto smoothed importance sampling (PSIS) diagnostic plot was used for the Bayesian model. Good Pareto k estimates ($k < 0.5$) in the PSIS diagnostic plot show that the model fits the data. The version of PSIS used in this work corresponds to the algorithm presented in Vehtari, Simpson, Gelman, Yao, and Gabry [40].

Assessment of Obstructive Sleep Apnea Impact- It is well perceived that Obstructive Sleep Apnea (OSA) has a negative consequence on sleep and is a reason for sleep deficiency. OSA induces behavioral sleep problems, bedtime resistance, and a significantly shortened sleep duration [40]. Apnea-hypopnea Index (AHI) is used to quantify the degree of OSA. We investigated the correlation of ESS, FOSQ, and SDS with AHI to examine which one better captures the impact of OSA on sleep deficiency.

3. Results

The probability density plot of SDS computed using the eq. (1) has been shown in Fig. 4. The histogram of SDS follows a Gaussian distribution with a mean of 60 ($N=500$) and a standard deviation of 22. A boxplot comparison between the sleep deficiency severities of males and females has been shown in Fig. 5. No significant ($p\text{-value} > 0.05$) difference was observed between the average SDS of males with that of females. SDS shows a moderate ($r=-0.35$, $p=0.0$) correlation with age, i.e. higher the age, the higher is the sleep deficiency. The scatterplot of age and SDS with a trend line has been visualized in Fig. 6. Additionally, SDS shows a weak ($r=-0.21$, $p=0.0$) positive correlation with Body Mass Index (BMI). Boxplots of SDS for normal and overweight categories are shown in Fig. 7. The overweight category has a higher SDS.

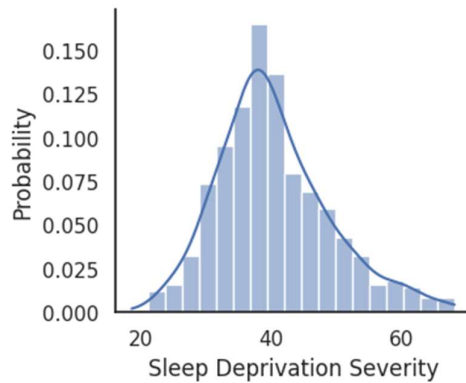


Fig. 4 Probability density plot for SDS distribution

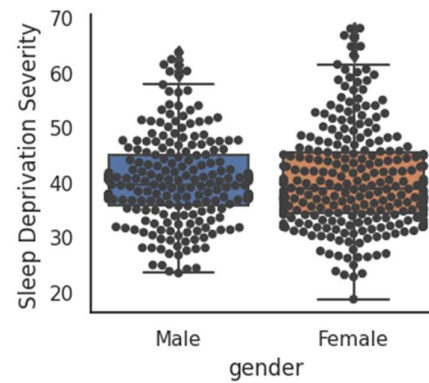


Fig. 5 Comparison of SDS between males and females

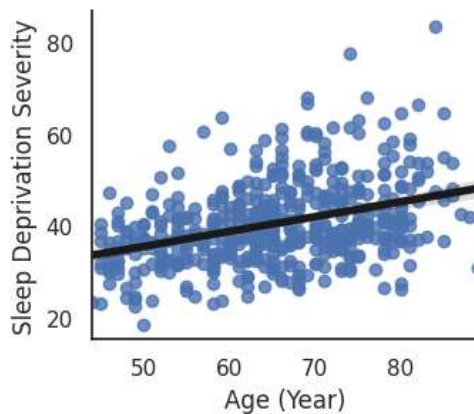


Fig. 6 Correlation of SDS with Age

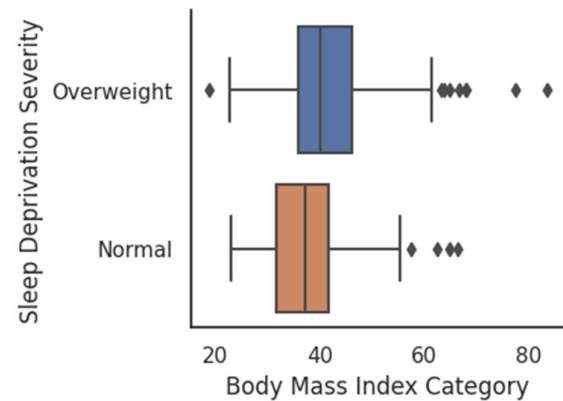


Fig. 7 SDS for BMI Categories

The partial correlation (controlled for age and BMI) of HRV and EEG features with baseline SDS computed using eq (1) indicated a significant correlation for several features. with have been performed to investigate the relationship of SDS with these features when controlled for age and BMI. The best 5 features from each sensor showing a significant correlation with SDS have been listed in Table 2

Table 2 Partial Correlation of SDS with HRV and EEG features

HRV Features			EEG Features		
Feature	r	p-value	Feature	r	p-value
AVNN	0.09	0.04	slowosc_nrem	0.39	0.01
pNN10	0.08	0.04	delta_nrem	0.39	0.01
HR	-0.11	0.02	slowosc_sleep	0.33	0.01
VLF	-0.11	0.02	delta_sleep	0.29	0.01
LF/HF	-0.13	0.01	theta_rem	0.25	0.01

* Variables in the table have been described in the Appendix A

Although both HRV and EEG features show a significant partial correlation with SDS, the correlation for EEG features is much stronger than HRV features indicating that EEG features have relatively higher importance than HRV features in estimating SDS. Hence, in developing the regression method, only EEG and anthropometric measures have been used.

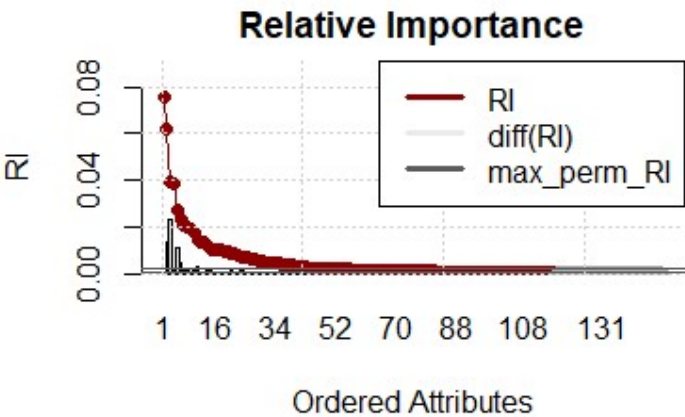


Fig. 8 Relative importance of features

MCFS results indicate that out of 150 features 117 are important based on the cut-off value of feature relative importance (RI) as shown in Fig. 8. The line with red/gray dots gives RI values, the vertical bar plot gives the difference δ between consecutive RI values. Informative features are separated from non-informative ones by the cutoff value and are presented in the plot as red and gray dots, respectively. The convergence of MCFS-ID algorithm has been shown in Fig. 9. The distance function (red line) shows the difference between two consecutive rankings – zero means no changes between two rankings (see the left y-axis). The common part (in blue color) gives the fraction of features that overlap for two different rankings (see the right y-axis). The ranking stabilizes after some iterations: the distance tends to zero and the common part tends to 1. β_1 shows the slope of the tangent of a smoothed distance function. If β_1 tends to 0 (the right y-axis) then the distance is given by a flat line. The top-ranked 20 features based on normalized relative importance by MCFS-ID have been shown in Fig. 10.

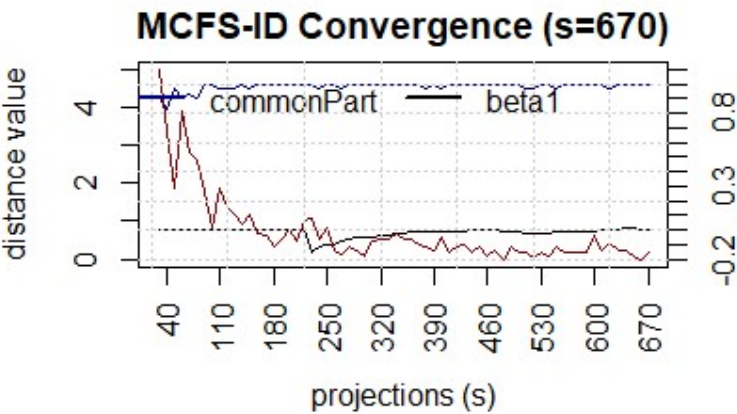


Fig. 9 Convergence of Monte-Carlo feature selection

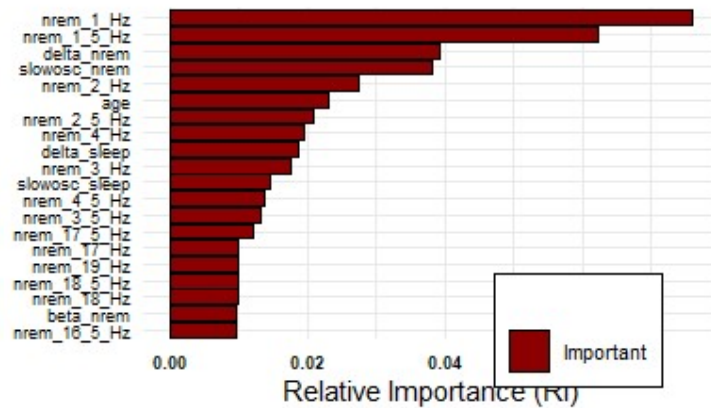


Fig. 10 Top 20 features by MCFS-ID algorithm

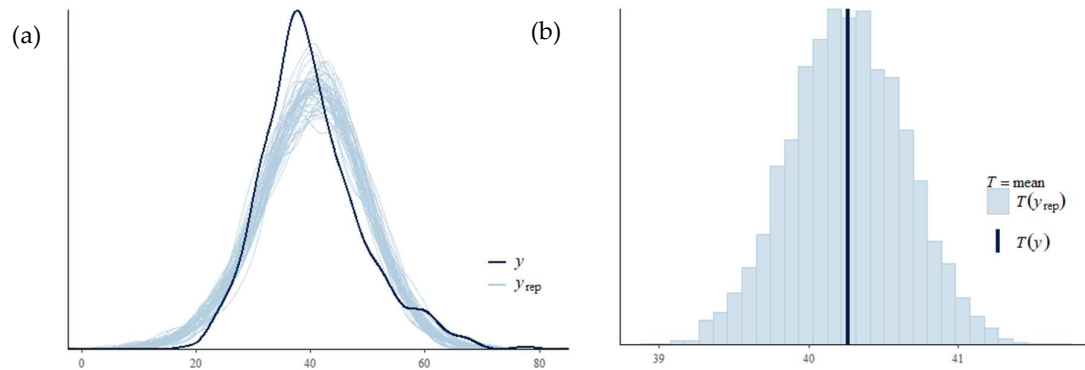


Fig. 11 (a) posterior predictive check on MCMC sampler (b) density plot of Bayesian model estimated SDS including the point estimate

The distribution of posterior R^2 for estimated SDS using Bayesian regression indicates an approximately normally distributed pattern. In MCMC diagnostics R -hat values for all parameters were less than 1.1. A posterior predictive check on MCMC sampler has been shown in Fig. 11(a). The dark blue line shows the observed data while the light blue lines are simulations from the posterior predictive distribution. The patterns for both distributions are in agreement with some deviations for the peak. Fig. 11(b) shows the probability density plot for the estimated SDS using Bayesian regression where the solid line indicates the point estimate from the Ordinary least squares method. The plot-of-fit for the Bayesian regression model is shown in Fig. 12 where R -squared value = 0.60 and RMSE = 5.63. The PSIS diagnostics plot for the Bayesian model has been shown in Fig. 13 which reveals that only a few points are outside the acceptable threshold. The estimated shape parameter k for each observation is used as a measure of the observation's influence on the posterior distribution of the model.

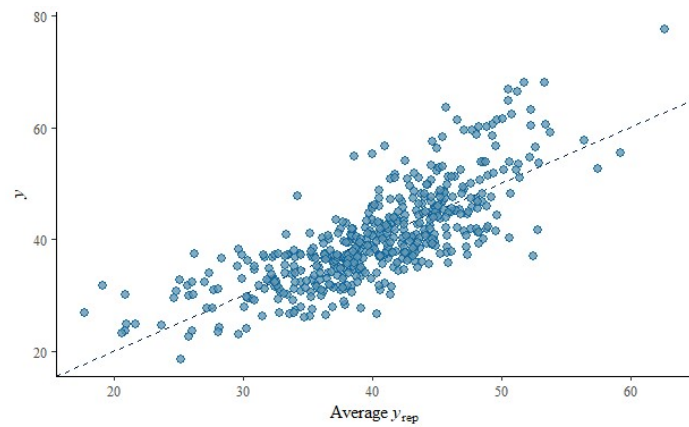


Fig. 12 Plot-of-fit for the Bayesian regression model

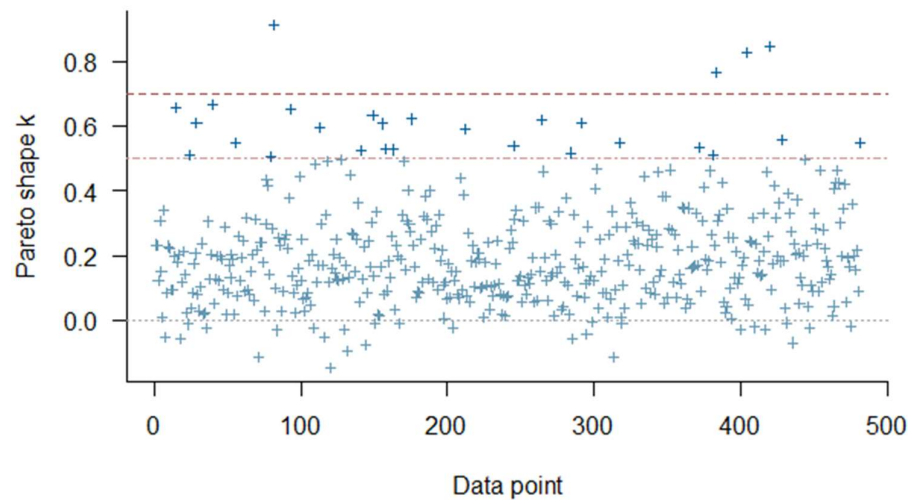


Fig. 13 PSIS diagnostic plot and regression plot for Bayesian method

For SDS estimation, ANN achieved a performance of RMSE of 4.65 and R-squared value of 0.86 in the training set, and an RMSE of 5.47 and R-squared value of 0.67 in the test set. The fit of the regression plot for the ANN model has been shown in Fig. 14. The dashed line indicates the ideal trend line and the solid line indicates the fitted trend line for the actual versus predicted values. The histogram of prediction error showed symmetrically skewed and almost normally distributed patterns with a higher frequency in the error bin ± 2 . The residual plot for the regression analysis shows a random scattering around the zero lines.

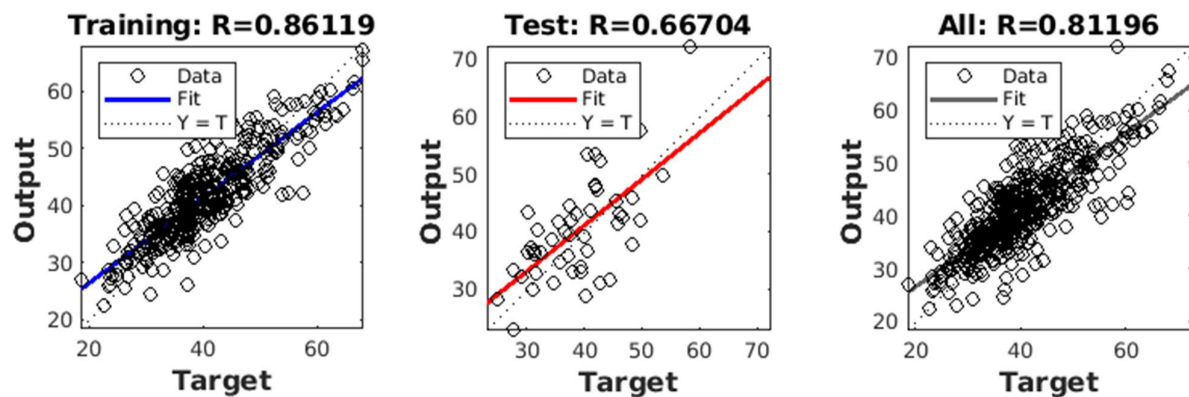


Fig. 14 Performance of Regression model in the train, validation, and test set

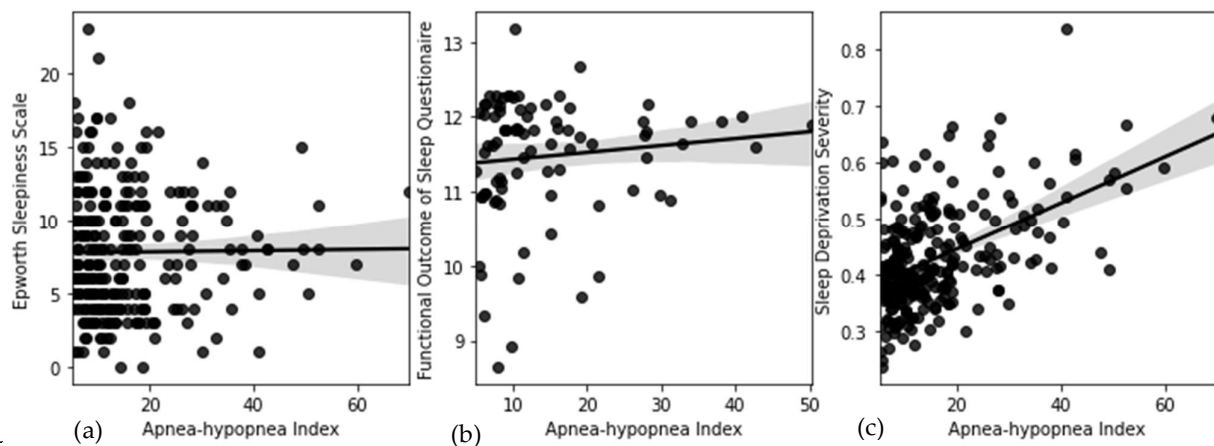


Fig. 15 Correlation of (a) ESS with AHI (b) FOSQ with AHI (c) SDS with AHI

Fig. 15(a) shows the impact of OSA as captured by the Epworth Sleepiness Scale (ESS). It can be seen that ESS does not reveal an informative trend and failed ($r < 0.1$) to capture the impact of severe OSA on the sleep deficiency of OSA patients. Similarly, Fig. 15(b) shows the correlation of the Functional Outcome of Sleep Questionnaire (FOSQ) with the apnea-hypopnea index. The trend in this case also fails ($r < 0.19$) to capture the impact of OSA severity on sleep deficiency. Fig. 15(c) shows that SDS, as computed using the proposed method, shows a good positive correlation ($r = 0.31$) with AHI. As OSA severity increases, SDS also proportionately increases.

4. Discussion

While quantification and longitudinal monitoring of sleep deficiency are beneficial for early diagnosis and continuous monitoring of sleep disorder and may facilitate corrective habitual actions and practices that adversely affect good sleep, it is noteworthy that sleep deficiency is not only linked with the physiological disorder but also emotional stress and other factors. To reduce the variability in everyday measurement a moving average over a week or longer period as well as sleep pattern visualization may provide better insights when added to the SDS score. Signal quality and data reliability also impact the measurements and hence a data reliability metric may be used to enhance the usability of the method. Moreover, it is noteworthy that we could not directly compare the utility of SDS with that of sleep score as sleep score formulae used by commercial entities are not publicly available to the best of our knowledge.

5. Conclusions

In this study, we analyzed SDS and its relationship with HRV, and EEG-based features. We performed a feature ranking using MCFS-ID for identifying the most informative features for SDS estimation. Finally, we developed a regression method using ANN for SDS score estimation from spectral features of single-channel EEG. The findings of this study increased the interpretability of SDS and paves way for the usage of SDS as a potential indicator for automated sleep disorder checks using wearables. In future studies, we are aiming a large scale deployment of the model for longitudinal monitoring of SDS with wearables.

Supplementary Materials: A Smart Health(sHealth) framework was developed previously for implementing the sleep deficiency severity algorithm. The framework is available online at www.sccmobilehealth.com. Associated codes and resources are available at our GitHub repo at <https://github.com/esarplab>

Author Contributions: Md Juber Raman and Bashir I. Morshed conceived of the presented idea. Md Juber Rahman developed the algorithm and performed the computations including the data analytics. Bashir I. Morshed and Chrysanthé Preza supervised the findings of this work and verified the analytical methods. All authors discussed the results and contributed to the final manuscript.

Funding: This material is based upon work supported by the US National Science Foundation under Grant No. 1637250.

Informed Consent Statement: Approval has been obtained from an institutional review board at the University of Memphis (IRB# PRO-FY2017-474)

Data Availability Statement: The data used in this study was obtained from the National Sleep Research Resource. Anyone interested may apply here <https://sleepdata.org/data/requests/shhs/start> for getting access to the dataset.

Acknowledgments: We acknowledge the contributions of B.E. Harmon, Md Sabbir bin Zaman, Sharmin Afroz, and Mamun Rahman in developing the overall Smart Health (sHealth) framework.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Feature	Description
AVNN	Mean of Normal to Normal interval
pNN1x	NNx count divided by the total no. of all NN intervals where NNx is the number of pairs of adjacent NN intervals differing by more than x ms.
HR	Mean heart rate in beat per minute
VLF	Power in very low frequency range (≤ 0.04 Hz)
LF/HF	Ratio of LF/HF power
slowosc_nrem	N-REM EEG power in slow oscillations band
delta_nrem	N-REM EEG power in delta band
slowosc_sleep	Total EEG power in slow oscillations band
delta_sleep	Total EEG power in delta band
theta_rem	N-REM EEG power in theta band

References

1. Mukherjee, Sutapa, Sanjay R. Patel, Stefanos N. Kales, Najib T. Ayas, Kingman P. Strohl, David Gozal, and Atul Malhotra. "An official American Thoracic Society statement: the importance of healthy sleep. Recommendations and future priorities.", *American journal of respiratory and critical care medicine*, vol.191, no. 12, pp: 1450-1458, 2015.
2. Dai, Cimin, Ying Zhang, Xiaoping Cai, Ziyi Peng, Liwei Zhang, Yongcong Shao, and Cuifeng Wang. "Effects of sleep deprivation on working memory: change in functional connectivity between the dorsal attention, default mode, and fronto-parietal networks." *Frontiers in Human Neuroscience* 14 (2020).
3. Alhola, Paula, and Päivi Polo-Kantola. "Sleep deprivation: Impact on cognitive performance." *Neuropsychiatric disease and treatment* vol. 3,5 (2007): 553-67.
4. Spiegel, Karine, Esra Tasali, Rachel Leproult, and Eve Van Cauter. "Effects of poor and short sleep on glucose metabolism and obesity risk.", *Nature Reviews Endocrinology*, vol. 5, no. 5, pp: 253, 2009.
5. Sharma, Monika, J. P. S. Sawhney, and Samhita Panda. "Sleep quality and duration–Potentially modifiable risk factors for Coronary Artery Disease?.", *Indian heart journal*, vol. 66, no. 6, pp.: 565-568, 2014.
6. Perez-Pozuelo, I., Zhai, B., Palotti, J. et al. The future of sleep health: a data-driven revolution in sleep science and medicine. *npj Digit. Med.* 3, 42 (2020). <https://doi.org/10.1038/s41746-020-0244-4>.
7. C. Kuo, Y. Liu, D. Chang, C. Young, F. Shaw and S. Liang, "Development and Evaluation of a Wearable Device for Sleep Quality Assessment," in *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1547-1557, July 2017.
8. F. Mendonça, S. S. Mostafa, F. Morgado-Dias and A. G. Ravelo-García, "Sleep Quality Estimation by Cardiopulmonary Coupling Analysis," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 12, pp. 2233-2239, Dec. 2018, doi: 10.1109/TNSRE.2018.2881361.
9. I. Azimi et al., "Personalized Maternal Sleep Quality Assessment: An Objective IoT-based Longitudinal Study," in *IEEE Access*, vol. 7, pp. 93433-93447, 2019, doi: 10.1109/ACCESS.2019.2927781.
10. Bsoul, Majdi, Hlaing Minn, Mehrdad Nourani, Gopal Gupta, and Lakshman Tamil. "Real-time sleep quality assessment using single-lead ECG and multi-stage SVM classifier." In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 1178-1181. IEEE, 2010.
11. Sleep Deprivation and Deficiency, <https://www.nhlbi.nih.gov/health-topics/sleep-deprivation-and-deficiency>, National Institute of Health, US Dept. of Health and Human Services, retrieved on July 4, 2021.
12. M. J. Rahman and B. I. Morshed, "SCC Health: A Framework for Online Estimation of Disease Severity for the Smart and Connected Community," *2019 IEEE International Conference on Electro Information Technology (EIT)*, 2019, pp. 373-378, doi: 10.1109/EIT.2019.8834189.
13. Zaman, Md Sabbir, and Bashir I. Morshed. "A low-power portable scanner for body-worn Wireless Resistive Analog Passive (WRAP) sensors for mHealth applications." *Measurement* 177 (2021): 109214.
14. M. J. Rahman, B. I. Morshed and B. Harmon, "A Field Study to Capture Events of Interest (EoI) from Living Labs Using Wearables for Spatiotemporal Monitoring Towards a Framework of Smart Health (sHealth)," *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 5943-5947, doi: 10.1109/EMBC44109.2020.9175771.
15. Morshed, Bashir I., Brook Harmon, Md Sabbir Zaman, Md Juber Rahman, Sharmin Afroz, and Mamunur Rahman. "Inkjet printed fully-passive body-worn wireless sensors for smart and connected community (SCC)." *Journal of Low Power Electronics and Applications* 7, no. 4 (2017): 26.
16. Afroz, Sharmin, and Bashir I. Morshed. "Web Visualization of Temporal and Spatial Health Data from Smartphone App in Smart and Connected Community (SCC)." In *2018 IEEE International Smart Cities Conference (ISC2)*, pp. 1-6. IEEE, 2018.
17. Buysse, Daniel J., Charles F. Reynolds III, Timothy H. Monk, Susan R. Berman, and David J. Kupfer. "The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research." *Psychiatry research*, vol. 28, no. 2, pp: 193-213, 1989.
18. Weaver, Terri E., Andréa M. Laizner, Lois K. Evans, Greg Maislin, Deepak K. Chugh, Kerry Lyon, Philip L. Smith *et al.* "An instrument to measure functional status outcomes for disorders of excessive sleepiness." *Sleep*, vol. 20, no. 10, pp: 835-843, 1997.
19. Johns, Murray W. "A new method for measuring daytime sleepiness: the Epworth sleepiness scale." *Sleep*, vol. 14, no. 6, pp: 540-545, 1991.
20. Åkerstedt, Torbjörn, K. E. N. Hume, David Minors, and J. I. M. Waterhouse. "The subjective meaning of good sleep, an intraindividual approach using the Karolinska Sleep Diary." *Perceptual and motor skills* 79, no. 1 (1994): 287-296.
21. O'donnell, Deirdre, Edward J. Silva, Mirjam Münch, Joseph M. Ronda, Wei Wang, and Jeanne F. Duffy. "Comparison of subjective and objective assessments of sleep in healthy older subjects without sleep complaints." *Journal of sleep research* 18, no. 2 (2009): 254-263.
22. Zhang, Lin, and Zhong-Xin Zhao. "Objective and subjective measures for sleep disorders." *Neuroscience Bulletin* 23, no. 4 (2007): 236-240. A. D. Krystal and J. D. Edinger, "Measuring sleep quality", *Sleep Med.*, vol. 9, pp. S10-S17, Sep. 2008.
23. Sun, Shuyu, Xianchao Zhao, Jiafeng Ren, Jinxiang Cheng, Junying Zhou, and Changjun Su. "Characteristics of objective sleep and its related risk factors among Parkinson's disease patients with and without restless legs syndrome." *Frontiers in Neurology* 12 (2021).
24. What's sleep score in the Fitbit app? https://help.fitbit.com/articles/en_US/Help_article/2439.htm. Retrieved on June 25, 2021.

25. de Zambotti, Massimiliano, *et al.* "A validation study of Fitbit Charge 2™ compared with polysomnography in adults." *Chronobiology international* 35.4 (2018): 465-476.
26. How Oura Measures Your Sleep, Retrieved on June 25, 2021 from <https://ouraring.com/blog/sleep-score/>
27. That Sleep Tracker Could Make Your Insomnia Worse, New York Times. Retrieved on June 25, 2021 from <https://www.nytimes.com/2019/06/13/health/sleep-tracker-insomnia-orthosomnia.html>. A version of this article appears in print on June 17, 2019, Section B, Page 3 of the New York edition of with the headline: Sleep Trackers Could Make Your Insomnia Worse
28. Liang, Zilu, and Mario Alberto Chapa Martell. "Validity of consumer activity wristbands and wearable EEG for measuring overall sleep parameters and sleep structure in free-living conditions." *Journal of Healthcare Informatics Research* 2.1 (2018): 152-178.
29. Quan, Stuart F., Barbara V. Howard, Conrad Iber, James P. Kiley, F. Javier Nieto, George T. O'Connor, David M. Rapoport *et al.* "The sleep heart health study: design, rationale, and methods." *Sleep* 20, no. 12 (1997): 1077-1085.
30. Rosipal, Roman, Achim Lewandowski, and Georg Dorffner. "In search of objective components for sleep quality indexing in normal sleep." *Biological psychology* 94, no. 1 (2013): 210-220.
31. Landry, Glenn J *et al.* "Measuring sleep quality in older adults: a comparison using subjective and objective methods." *Frontiers in aging neuroscience* vol. 7 166. 7 Sep. 2015, doi:10.3389/fnagi.2015.00166
32. Malik, M., Farrell, T., Cripps, T. and Camm, A.J., "Heart rate variability in relation to prognosis after myocardial infarction: selection of optimal processing techniques." *European heart journal*, vol. 10 , no.12, pp.1060-1074, 1989.
33. Goldberger, Ary L., Luis AN Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals." *circulation* 101, no. 23 (2000): e215-e220.
34. Damiński, Michał, Alvaro Rada-Iglesias, Stefan Enroth, Claes Wadelius, Jacek Koronacki, and Jan Komorowski. "Monte Carlo feature selection for supervised classification." *Bioinformatics* 24, no. 1, pp.: 110-117, 2008.
35. Damiński, Michał, and Jacek Koronacki. "rmcfs: an R package for Monte Carlo feature selection and interdependency discovery." *Journal of Statistical Software* 85, no. 1, pp.: 1-28, 2008.
36. Bishop, Christopher M., and Michael E. Tipping. "Bayesian regression and classification." *Nato Science Series sub Series III Computer And Systems Sciences* 190 (2003): 267-288.
37. Goodrich B, Gabry J, Ali I, Brilleman S (2020). "rstanarm: Bayesian applied regression modeling via Stan." R package version 2.21.1, <https://mc-stan.org/rstanarm>.
38. Marquardt, Donald W. "An algorithm for least-squares estimation of nonlinear parameters." *Journal of the society for Industrial and Applied Mathematics* 11, no. 2, pp.: 431-441, 1963.
39. Vehtari, Aki, Daniel Simpson, Andrew Gelman, Yuling Yao, and Jonah Gabry. "Pareto smoothed importance sampling." *arXiv preprint arXiv:1507.02646*, 2015.