2020

Predicting Subjective Sleep Quality Using Objective Measurements in Older Adults

Reza Sadeghi
Wright State University

PREDICTING SUBJECTIVE SLEEP QUALITY USING OBJECTIVE MEASUREMENTS IN OLDER ADULTS

A Dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

by

REZA SADEGHI
B.S. Isfahan University of Technology, 2012
M.S. International University of Imam Reza, 2015

2020
Wright State University

Wright State University
COLLEGE OF GRADUATE STUDIES

April 2020

I HEREBY RECOMMEND THAT THE DISSERTATION PROPOSAL PREPARED UNDER MY SUPERVISION BY REZA SADEGHI ENTITLED PREDICTING SUBJECTIVE SLEEP QUALITY USING OBJECTIVE MEASUREMENTS IN OLDER ADULTS BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

Tanvi Banerjee, Ph.D.
Dissertation Director

Yong Pie, Ph.D.
Director, Computer Science
and Engineering Ph.D. Program

Barry Milligan, Ph.D.
Interim Dean of the Graduate School

Committee on
Final Examination

Tanvi Banerjee, Ph.D.

Mateen M. Rizki, Ph.D.

Krishnaprasad Thirunarayan, Ph.D.

William Romine, Ph.D.

Ali Azarbarzin, Ph.D.

ABSTRACT

Sadeghi, Reza. PhD, Department of Computer Science and Engineering, Wright State University, 2020. *PREDICTING SUBJECTIVE SLEEP QUALITY USING OBJECTIVE MEASUREMENTS IN OLDER ADULTS*.

Humans spend almost a third of their lives asleep. Sleep has a pivotal effect on job performance, memory, fatigue recovery, and both mental and physical health. Sleep quality (SQ) is a subjective experience and reported via patients' self-reports. Predicting subjective SQ based on objective measurements can enhance diagnosis and treatment of SQ defects, especially in older adults who are subject to poor SQ. In this dissertation, we assessed enhancement of subjective SQ prediction using an easy-to-use E4 wearable device, machine learning techniques and identifying disease-specific risk factors of abnormal SQ in older adults.

First, we designed a clinical decision support system to estimate SQ and feeling refreshed after sleep using data extracted from an E4 wearable device. Specifically, we processed four raw physiological signals of heart rate variability (HRV), electrodermal activity, body movement, and skin temperature using distinct signal processing methodologies. Following this, we extracted signal-specific features and selected a subset of the features using recursive feature elimination cross validation strategy to maximize the accuracy of SQ classifiers in predicting the SQ of older caregivers.

Second, we investigated discovering more effective features in SQ prediction using HRV features which are not only effortlessly measurable but also can reflect sleep stage transitions and some sleep disorders. Evaluation of two interpretable machine learning methodologies and a convolutional neural network (CNN) methodology demonstrated the CNN outperforms by an accuracy of 0.6 in predicting light, medium, and deep SQ. This outcome verified the capability of using HRV features measurable by easy-to-use wearable devices, in predicting SQ.

Finally, we scrutinized daytime sleepiness risk factors as a sign of abnormal SQ from four perspectives: sleep fragmented, sleep propensity, sleep resilience, and non-restorative sleep. The analysis demonstrates distinguishability of the main risk factors of excessive daytime sleepiness (EDS) between patients suffering from fragmented sleep (e.g. apnea) and sleep propensity. We identified the average area under oxygen desaturation curve corresponds to apnea/hypopnea event as a disease-specific risk factor of abnormal SQ. Our further daytime sleepiness prediction demonstrated the significant role of the founded disease-specific risk factor as well.

Table of contents

List of Figures

List of Tables

Acknowledgment

I would like to express my sincere appreciation to my advisor Dr. Tanvi Banerjee for her guidance and support. She always helped me push the boundary of my abilities during this dissertation. I will fondly remember every moment of our weekly meetings that made my PhD experience productive.

I send my deepest gratitude to my dissertation committee members Drs. Mateen M. Rizki, Krishnaprasad Thirunarayan, and William Romine. Their productive comments improved the quality of this dissertation. They generously dedicated hours of their time to share their unbiased thoughts about my findings and scrutinize my dissertation content.

I also extend my gratitude to my external committee member, Dr. Ali Azarbarzin. His assistance and contributions significantly enhanced my research studies. While I worked under his supervision at Brigham and Women's Hospital, he taught me distinct perspectives of obstructive sleep apnea. He was so generous with his time as he guided me in analyzing polysomnography data.

I am immensely grateful to Professor Susan Redline, whose advocacy enabled me to join the Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Harvard Medical School. It was my honor to work under her supervision and become familiar with top researchers in the field of sleep medicine. I also thank National Sleep Research Resource (NSRR) and the sleep division staff who provided access to large cohorts of sleep studies and aided me in continuing my sleep research.

It is my pleasure to acknowledge practical advice of Drs. Larry Lawhorne, John Hughes, and Derek Doran. They helped me have an insight into complex clinical and computational analyses. Also, I was fortune to use the invaluable knowledge of Drs. Andrew Wellman, Scott Sands, and Daniel Vena in extending my knowledge on sleep disorders. Finally, my biggest thanks goes to my wife whose love and encouragement supported me throughout my PhD. I am thankful to my family for their endless support in all aspects of my life.

1.Introduction

1.1. Overview

Humans spend almost a third of their life asleep [1]. This significant portion of life has pivotal effects on job performance [2], memory [3], fatigue recovery [4], and both mental [5] and physical health [6]. The discharge of anabolic hormones (e.g., prolactin, testosterone, luteinizing hormone) during sleep in addition to physical restoration lead to feeling refreshed after sleep [7]. When people experience difficulty sleeping, the detrimental effects range from daytime sleepiness to performance reduction and lack of attention [8]. As a result, sleep assessment has attracted much attention in recent decades.

Sleep assessment methods can be categorized using distinct criteria such as subjective vs. objective measurements [9]. In subjective measurements, individual persons describe their own sleep quality (SQ). Sleep questionnaires and diaries are an organized way to provide a discrete approximation of real sleep quality based on individual reports. They are the first diagnostic test used in primary care, and they are cheap, self-administered, and validated through large statistical studies [10].

In contrast, many objective measurements are proposed to provide pure quantitative SQ measurements extracted from either Polysomnogram (PSG), wearable devices, or contactless devices [11]. PSG measurements include electroencephalogram (EEG) signals, oximetry, and cardiovascular and respiratory measurements. As sleep is a complex and mysterious phenomenon, sleep doctors sometimes monitor sleep using audio, actigraphy, video, or temperature.

Predicting SQ based on objective measurements can enhance diagnosis and treatment of SQ defects. Despite the existence of a great deal of sleep studies, predicting SQ based on objective measurements is still an evolving science [12]. This dissertation proposes solutions to the three existing challenges in SQ assessment detailed in Section 1.2 by addressing three research questions described in Section 1.3. Next, the thesis statement and contributions sections summarize the general ideas and main contributions of this study. Chapter 2 provides the preliminary knowledge for SQ assessment. Following this, the proposed methodologies to address the research questions are detailed in Chapter 3. Finally, Chapter 4 concludes the main findings and the limitations of this study.

1.2. Existing challenges

Currently, there are several challenges to sleep assessment:

1. **Polysomnography (PSG) is expensive and destructive to the sleep routine.**

   PSG is the primary method to assess sleep, but PSG is expensive and requires attaching electrodes to the bodies of participants, and should be conducted under medical surveillance. As a result, using PSG is probably destructive to the sleep routine, especially in older adults, who are susceptible to poor sleep quality as a result of aging and age-related comorbidities.

2. **The current knowledge regarding sleep quality prognostication factors is limited.**

   The emergence of wearable technologies enables non-invasive, long-term monitoring of physiological signals during sleep, especially in older adults who may not be able to join PSG studies. The current knowledge regarding sleep quality explains how the recorded physiological signals have different characteristics in different sleep stages, and life style has a direct effect on sleep quality. However, the currently known sleep-related features extracted from both physiological signals and medical history do not completely describe an individual's sleep quality.

3. **Abnormal sleep quality has distinct disease-specific risk factors.**

   The raw physiological signals can be collected by a wearable device without changing the sleep routine. Machine learning techniques process these data and describe the corresponding SQ. Developing effective features can lead to an efficient way of predicting SQ by reducing both computational complexity and the number of required signals to be monitored during sleep. Such features are effective as they potentially point to the risk factors of abnormal SQ. However, objective sleep-related risk factors of abnormal SQ are affected by conditions like stress level, falling and fractures, and mental and respiratory diseases.

1.3. Motivation

In this dissertation, we endeavored to enhance SQ prediction using an easy-to-use and non-invasive E4 wearable device in older adults. Using a wearable device enables us to inexpensively monitor required physiological signals to evaluate the SQ of different subjects sleeping in their usual home environments. Additionally, we extracted significant sleep-related risk factors from PSG data that can lead to an efficient way of predicting SQ. Identifying the risk factors of abnormal SQ can enhance both diagnosis [13] and treatment [14] of SQ defects. We addressed the three existing challenges in SQ prediction detailed in

Section 1.2 by addressing the following three research questions in estimating SQ using objective measurements, which are extracted from physiological signals and patients' medical history. The main research questions for this study are described as follows:

1. **Can we predict sleep quality using non-invasive and easy-to-use wearable devices instead of PSG in family caregivers of people with dementia?**

   A growing group of older adults, family caregivers of people with dementia (CPWD), are susceptible to poor SQ due to the high degree of stress and the demands of providing care to their loved ones. As SQ can be a measure of readiness for starting CPWD duties, SQ monitoring can be beneficial for the caregivers to identify potential concerns and provide potential interventions before reaching a detrimental level. However, using PSG is cost prohibitive and should be conducted under medical surveillance. As a result, we introduced a clinical decision support system to predict both SQ and feeling refreshed in caregivers using a wearable E4 wristband in our published work in *Computers in Biology and Medicine* [15]. This easy-to-use device enabled us to monitor important physiological signals during sleep while not interrupting regular sleep patterns of CPWD. We recorded 100 sleep nights of eight CPWD, who aged 65+, for a period of two weeks each. As distinct physiological signals have different characteristics in deep sleep stage, our proposed method extracts signal-specific features from each of the four raw recorded physiological signals: heart rate variability (HRV), electrodermal activity, body movement, and skin temperature. Following this, our method selects a subset of the extracted features using recursive feature elimination cross validation strategy to maximize the accuracy of SQ classifiers in prediction of the SQ of CPWD.

2. **Can machine learning techniques enhance the current knowledge of sleep quality prediction?**

   Using effective features extracted from physiological signals can encourage using wearable devices by reducing the number of required signals to be monitored during sleep. However, the computed features are highly dependent on feature engineering and sleep expert knowledge. This preprocessing limits the computational space to the expert knowledge and may remove the critical features in the decision-making process. We examined the capability of using machine learning techniques in pushing the boundaries for predicting SQ in our work submitted to *Conference of the IEEE Engineering in Medicine and Biology Society* [16]. We specifically assessed predicting SQ by investigating three methodologies based on electronic health records and HRV. Heart activities are effortlessly measurable by easy-to-use wearable devices. In addition, they can reflect sleep stage transitions [17] and some sleep disorders [8]. We scrutinized heart activity based on electrocardiogram (ECG) data extracted from 792 PSG samples of men

aged 65+ who participated in the Osteoporotic Fractures in Men (MrOS) sleep study. As obstructive sleep apnea (OSA) is a leading factor of sleep fragmentation and daytime sleepiness in older adults [18], the baseline methodology utilized nine features highly correlated to OSA extracted from electronic health records. The second methodology utilized 23 HRV features extracted from R-peaks annotation files. Annotated R-peaks are necessary in any heart variability analysis to remove ventricular ectopic beats, atrial ectopic beats, and artifacts. The third method utilized a convolutional neural network (CNN) to predict SQ based on ECG images. In the first two methodologies, the SQ prediction power were limited to the segregation power of engineered features while in the last methodology we fed the raw ECG images and let the CNN analyze and predict the outcomes.

3. **What are the disease-specific risk factors for abnormal sleep quality?**

   Abnormal SQ or sleep quantity causes excessive daytime sleepiness (EDS) [19], [20], which is a highly prevalent condition in the older adult society [21]. EDS is a symptom of several diseases, such as neurological disorders, e.g. dementia [22], and sleep breathing disorders [19], [20], e.g. apnea [23]. Distinguishing disease-specific risk factors of EDS can both reveal underlying reasons of abnormal SQ and enhance its prediction. To do so, we scrutinized EDS risk factors in our accepted paper in *Annals of the American Thoracic Society* [24]. We specifically investigate EDS risk factors in two groups of patients one group with and one group without severe sleep apnea where both groups may suffer from dementia. We evaluated sleep records from four distinct perspectives of sleep fragmentation, sleep propensity, sleep resilience to disruptive stimuli, and non-restorative sleep. Several statistical and classification analyses were performed on 4445 complete sleep records derived from the Sleep Heart Health Study (SHHS), which is the largest publicly available PSG dataset of people aged >=40. We also predicted daytime sleepiness using four simple and interpretable classifiers: logistic regression, naïve Bayes, decision tree, and K-nearest neighbors (KNN). We utilized F1-score as the objective function, weighted samples, and 10-times 10-fold cross-validation to reduce the chance of biasing majority samples and taking effects from time dependency between sample records.

1.4. Thesis statement

Sleep quality (SQ) is predictable using objective measurements of medical history and physiological signals (including heart rate variability, electrodermal activity, skin temperature, body movement, and blood oxygen saturation). Using easy-to-use wearable sensors, we can accurately and noninvasively acquire these physiological signals for a robust SQ prediction by leveraging machine learning techniques and extracting disease-specific risk factors.

1.5. Contributions

This dissertation introduces both computer science and biomedical contributions in the process of solving existing challenges in predicting SQ as follows:

1.5.1. Computer science contributions

CS1. Our work in predicting sleep quality (SQ) of caregivers of people with dementia (CPWD) published in *Computers in Biology and Medicine* [15] was a novel application of varied signal processing techniques to enable interpretable machine learning techniques to predict SQ. Also, our proposed clinical decision support system (CDSS) provides a prediction that is robust against the malfunction of a sensor on the wearable device and nonlinear relation among features by selecting the specific subset of divergent features.

CS2. In the process of predicting SQ from HRV in our work accepted in Conference of the *IEEE Engineering in Medicine and Biology Society* [16], we came up with a CDSS that processes the raw electrocardiogram signals independently from the prior knowledge of sleep experts. This CDSS employs a convolutional neural network (CNN) to predict SQ based on heart activities during each night by analyzing images of two ECG signals during Polysomnography studies. To our knowledge, this is one of the first studies to predict SQ using HRV.

1.5.2. Bioinformatics contributions

BI1. Our work published in *Computers in Biology and Medicine* [15] was a pioneering study for predicting SQ in older adults using easy-to-use wearable devices compared to PSG, which is probably sleep destructive. We predicted the SQ of 100 sleep nights of older caregivers of people with dementia with an accuracy of 75%. We found that the most important features in the process of predicting SQ using the recorded physiological signals by E4 wearable device are sleep efficiency (ratio of amount of time asleep to the amount of time in bed) and skin temperature.

BI2. In our accepted paper in *Annals of the American Thoracic Society* [16], we introduced a new metric of average hypoxia-specific area under the oxygen desaturation curve as a hypoxemia measurement. The high value of this metric is a significant risk factor of excessive daytime sleepiness in people who are suffering from obstructive sleep apnea (OSA) or hypopnea. This metric does not rely on a specific threshold in comparison to other common hypoxemia measurements, like oxygen desaturation index. Also, the metric reveals the duration and depth of hypoxia caused by partial or complete upper airway obstruction.

2. Preliminary knowledge

Sleep study is a vast area that increasingly attracts much attention. For instance, the number of publications in PubMed [25] with sleep in their title shown in the figure below follows an exponential growth. This section reviews some of these studies from three aspects: sleep architecture, physiology, and measurement tools.

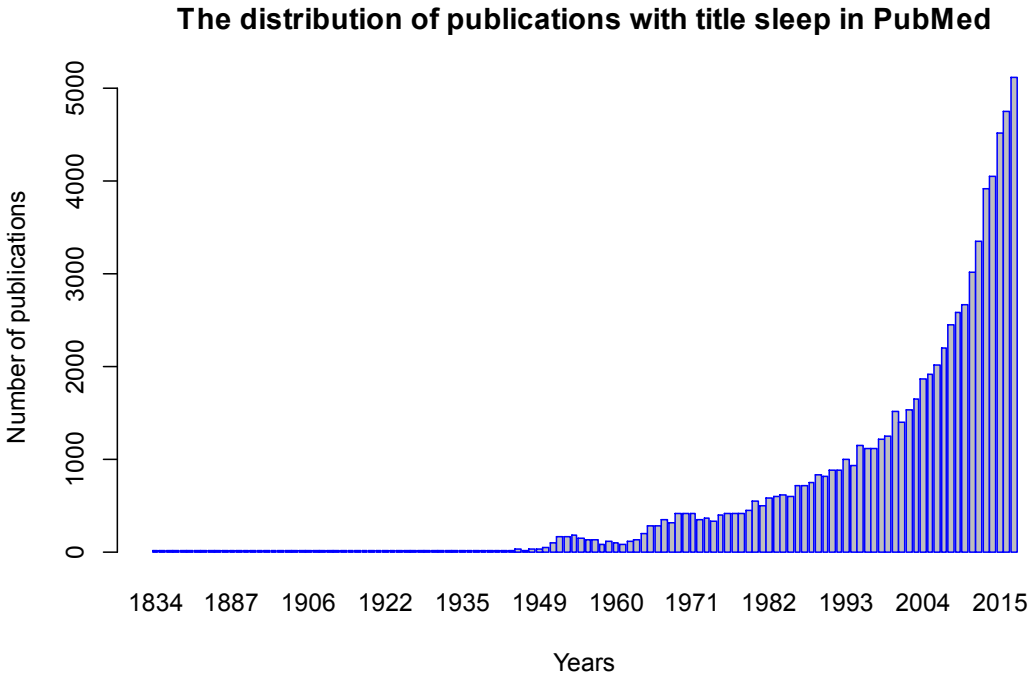

Figure 1.5-1 The distribution of publications in PubMed with sleep in their title from 1834 to 2017

2.1. Sleep architecture

Sleep architecture comprises two broad parts of rapid eye movement (REM) and non-REM (NREM) [26]. Since about 80% of the sleep time of an adult is NREM sleep [7], most of the sleep studies focus on NREM sleep. NREM sleep can be further split into four stages (stages 1–4) according to the R-K scoring manual [27]. Due to the similarity of NREM stage 3 and stage 4, they are considered as one stage of slow wave sleep (SWS) based on the current American Academy of Sleep Medicine (AASM) scoring manual [28], primarily on the basis of electroencephalogram (EEG) criteria.

Sleep stages are recognized and scored based on characteristic rhythms and events observed in the brain waves (EEG). As shown in the following figure, relaxed wakefulness is associated with alpha waves seen as a rhythm with peaks in the 8- to 13-Hz range. In addition, drowsiness coincides with slow, rolling eye movements that may persist into light sleep. The lightest stage of NREM sleep, NREM1, is characterized by a loss of alpha rhythm and presence of theta waves, which are characterized by frequencies of 4–7 Hz. Stage NREM 2 is described by the expression of spindles, burst-like trains of waves in the 11- to 16-Hz range with a total duration ≥0.5 seconds, and K-complexes, biphasic waves lasting ≥0.5 seconds and usually maximal over the frontal cortex. The deepest NREM sleep stage, NREM 3 and 4, is marked with large (≥75 μV) slow (0.5–3 Hz) waves known as delta waves. Typically, REM sleep is associated with the lowest skeletal muscle tone and with either sharp theta waves, smooth waves, or wake-like EEG patterns.

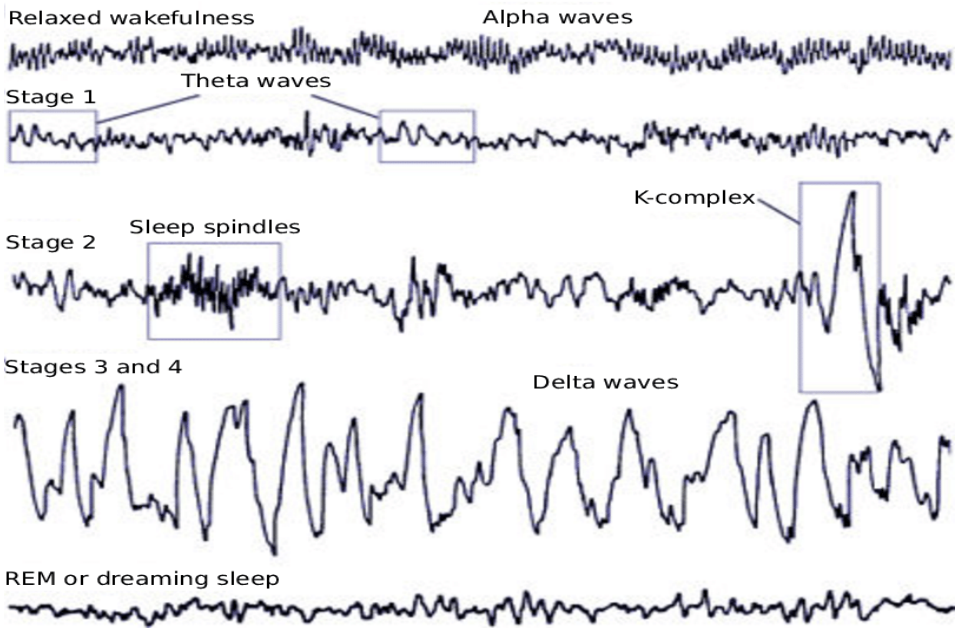

Figure 2.1-1 Sleep stage identifiers based on EEG features [29]

Among these sleep stages, SWS (deep sleep) has special characteristics relevant to SQ. For example, human growth hormone that corresponds to tissue repair is released during the first SWS episode [30], [31]. Also, it has been shown that taking sleeping pills is accompanied by extended deep sleep in patients suffering from insomnia [32]. Furthermore, patients who are suffering from sleep disorders like obstructive sleep apnea (OSA), periodic limb movement syndrome (PLMS), and insomnia experience less SWS than healthy subjects [33]. Moreover, the lessening of SQ in older adults is accompanied by the loss of deep sleep [34]. Since distinct sleep stages play different roles in SQ and its health outcomes, identification of length and specific physiological signal trends in distinct sleep stages can provide valuable information regarding the SQ prediction.

2.2. Sleep physiology

During NREM, most of the physiological processes in terms of brain activity, heart rate, blood pressure, sympathetic nerve activity, respiration, and body temperature diminish from their usual amount during wakefulness. In contrast, there is an increase in these physiological processes in REM as compared to NREM [8]. These fluctuations are well tolerated in healthy subjects; however, they may break the fragile balance of individuals who suffer severe diseases [26]. Also these fluctuations help us distinguish sleep stages based on physiological signals as described below:

The occurrence of intensive fluctuations in brain activities at stage 2, called K-complex, leads to a burst of sympathetic activities in NREM [26]. The sympathetic activity can be described by electrical changes of skin surface, called electrodermal activity (EDA) [35]. EDA recorded in overnight PSG has proven that EDA is more strongly associated with deep sleep than other stages [34]. In fact, various EDA studies have shown that people in SWS experience the highest level of both EDA values [36] and the number of local EDA peaks [37]. Also, [37] exhibits the stability of EDA properties in SWS by testing different places for wearable device attachment and different threshold selections for EDA peak definition; however, each individual has a different pattern and varied magnitude of EDA during sleep.

Parasympathetic activity increases as sleep goes from stage 1 to SWS. Parasympathetic activity leads to a continuous reduction in cardiovascular output [38]. This decreasing pattern continues such that the heart rate reaches its lowest point at SWS [39]. Body movement is another physiological measurement that has a strong relation with sleep stages [40]. Short movements appear over all sleep stages; however, the frequency of their occurrence in SWS is significantly lower [41]. Like body movement, body temperature is also influenced by sleep stages. For instance, adults experience a reduction in temperature during their deep sleep [42]. The speed and amount of this temperature reduction has a strong relationship with SQ [43]. Sleep physiology aids sleep experts in identifying poor SQ risk factors and their underlying pathology, as a first step of an effective treatment. The sleep physiology enables us to extract signal-specific features from the physiological signals that point to SQ risk factors, such as: irregular sleep stage patterns, respiratory diseases, and the aging process.

2.3. Sleep measurement tools

PSG is the primary method to both assess SQ and diagnose sleep disorders [9]. It provides holistic information about sleep physiology by summarizing the results of various tests, listed in Table 2.3-1. PSG contains additional hardware in the case of treatment, like continuous positive airway pressure (CPAP) machines [44]. Also, physicians may reduce the number of PSG tests due to examining patients for special diseases or conducting sleep

studies in patients' homes. For instance, the number of PSG channels in the home sleep test (HTS) can be reduced to airflow, respiratory effort, and oximetry [45].

Table 2.3-1 PSG tests

Test name	**Monitoring target**
Electroencephalogram	Brainwave activity
Electrooculogram	Eye movement
Electromyogram	Muscle activity
Electrocardiogram	Heart rate and rhythm
Pulse oximetry	Oxygen saturation
Respiratory monitor	Respiratory effort
Capnography	Inhaled and exhaled CO2 concentrations
Transcutaneous monitors	Diffusion of O2 and CO2 through the skin
Microphone	Snoring volume
Video camera	Identify body motion and position
Thermometer	Core body temperature
Light intensity tolerance	Influence of light intensity on sleep
Nocturnal penile tumescence	Identifying physiological erectile dysfunctions
Esophageal	Acidity test
Nasal and oral airflow	airflow and breathing rate
Gastroesophageal monitor	Detect gastroesophageal reflux disease
Blood pressure	Blood pressure change

Many other sleep assessment tools are proposed to enhance both affordability and continuity of sleep studies. These gadgets can be categorized into three categories: contact devices, contactless devices, and sleep questionnaires and diaries. Contact devices shape the fastest-growing group of sleep trackers since the compact size of wearables enables the subjects to monitor their sleep-related physiological signals while continuing their usual lifestyle. Table 2.3-2 provides detailed information about some of the current wearable sleep trackers.

Most sleep trackers are based on actigraphy such that some sleep surveys [46] divided sleep trackers into three categories of research-based devices, commercial devices based on actigraphy and others. All the information about some actigraphy devices, like FitBit Flex and Charge 2, is easily accessible; however, the information for others, like actiwatch 64 and GT3X+, is accessible only by contacting their producers.

Table 2.3-2 Details about some contact sleep trackers

Name	Price	Category	Description
FitBit Flex	$55 (Amazon)	Wristband	Tracks steps, distance, calories burned and active minutes (total sleep time and time in bed)
FitBit Charge2	$112 (Amazon)	Wristband	Activity tracker to total sleep time, time in bed
OURA	$197 (Amazon)	Ring	Steps and distance tracking, heart rate, activity, calories burned, sleep monitoring (measures light, deep, and REM sleep).
Dream 2	$500 (https://shop.dreem.com)	Headband	Accelerometer, 4 EEG sensor, and oximetry
Zmachine	$500 (https://www.generalsleep.com/zmachine-insight.html)	Three skin sensors	Three skin sensors placed behind ears and back of the neck for recording EEG.
E4 Empatica	$1,690 (https://store.empatica.com)	Wristband	It provides Electrodermal activity, blood volume pressure, heart rate, temperature, and accelerometer
Re-timer	$119 (Amazon)	Eyewear	Program to improve winter blues, sleep, reduce jet lag or prepare for night shift work
ResMed	$189.99 (https://www.isleephst.com/)	Home-based PSG	Home Sleep Apnea Test
Lofta	$189.00 (https://lofta.com/)	Wristband	Home Sleep Apnea Test

Not only do most sleep trackers report their outputs using smartphone applications, but also many sleep applications utilize smartphone sensors to analyze sleep duration, sleep efficiency, and quality of sleep [47]. Choi et. al [48] reported the existence of 2,431 sleep-related smartphone apps, of which 73 apps offered sleep support self-management. Sleep as Android Unlock is the most popular app [9], [48], which estimates light and deep sleep stages, the level of snoring and environmental noise, and sleep duration. In addition to smartphone apps, video cameras [49] and pillow or mattress accelerometers [50] are employed to track sleep quality using contactless devices.

Sleep questionnaires and diaries are sleep assessment methods that record subjective SQ in an organized way. Sleep questionnaires summarize information about a period of time, while sleep diaries are more accurate since they are not dependent upon the patient's memory [10]. Sleep questionnaires and diaries with general goals contain a high number of questions and a point scale, while the others that are designed for a specific goal or participants have fewer questions and a smaller range of answers. For instance, the sleep disorders questionnaire [51], designed for initial sleep disorder identification, contains 175 items with a 5-point scale, while the STOP-BANG questionnaire [52], predictor of severity of Apnea, includes only 8 items with a 2-point scale.

Only five percent of the wearable sleep devices and applications are formally validated [53]. Most sleep tracker providers do not share the main algorithm, utilize one type of signals for decision making, and only reveal the processed data, such as the length of sleep stages [46]. However, using wearable devices that offer measurements of several raw physiological signals may lead to more reliable and reproducible sleep assessment

methodologies. As a result, in this dissertation, we used E4 wristband and PSG to provide the accurate synchronized multiple physiological signals.

3. The proposed methods

This section introduces three proposed methods to enhance predicting sleep quality (SQ). The first method inexpensively predicted SQ of older adults using an easy-to-use and non-invasive wearable device. This method tracked the physiological trends by processing the raw physiological signals measurable with an easy-to-use E4 wristband [54]. Using a wearable device for monitoring sleep enabled us to inexpensively evaluate the SQ of different subjects sleeping in their usual home environments. The second method employed machine learning techniques to enhance SQ prediction by selecting more effective features and reducing the dependability to the expert knowledge in feature computation. We evaluated SQ by applying a great range of machine learning techniques over electronic health records, heart rate variability (HRV) features, and the raw Electrocardiogram (ECG) signals. The third method predicted excessive daytime sleepiness (EDS) from physiological signals as a means to predict abnormal SQ. Additionally, we scrutinized the association of new sleep-related biomarkers and EDS with the aim of identifying disease-specific risk factors of abnormal SQ.

3.1. Predicting sleep quality using wearable data (Contributions: CS1 and BI1)

3.1.1. Data

We monitored the sleep of eight CPWD for a period of two weeks each. Two of the caregivers were adult children of dementia patients, and six were patient spouses. However, using the wearable E4 device, we were only able to record 100 nights of sleep (as opposed to 112 nights). The caregivers were requested to wear the wristband approximately 15 min before sleep and remove the device immediately after waking up. The wearable device reports the physiological signals during sleep via four raw signals: heart rate variability, electrodermal activity, body movement (accelerometer), and skin temperature.

CPWD reported their SQ of the previous night through an Android tablet application, which contained the daily use caregiver sleep survey (DUCSS) [55]. As shown in contingency Table 3.1-1, 65% of the caregivers reported feeling tired after their sleep.

However, 82% of the sleep records are labeled as very good, good, or okay SQ. Indeed, 98% of sleep records reported fragmented sleep as a result they had lower chance to experience restful sleep and deep sleep. Various conditions can cause sleep fragmentations. Studies specific to caregiver sleep disturbance have shown that the act of providing care to loved ones with dementia-related illness results in chronic stress [56]. CPWD can experience sleep difficulties because of both overwhelming caregiving responsibilities and unpredictable dementia patient behaviors.

Table 3.1-1 Contingency table of Sleep quality and rest feeling after sleep in CPWD

		Sleep quality					
		Very good	**Good**	**Okay**	**Fairly bad**	**Bad**	**Sum**
Mood	**Feeling tired**	10	13	25	15	2	65
	Feeling rested	7	25	2	1	0	35
	Sum	17	38	27	16	2	100

3.1.2. Methods

We proposed a clinical decision support system to predict SQ based on trends of physiological signals in the deep sleep stage (SWS). As shown in the following figure, this sleep analysis strategy reuses the sleep investigations which have been conducted mainly using the expensive and accurate PSG. Then, it provides transparent outcomes based on proven medical evidence, and it is applicable for a wide range of users, especially for individuals experiencing stress or burden like CPWD.

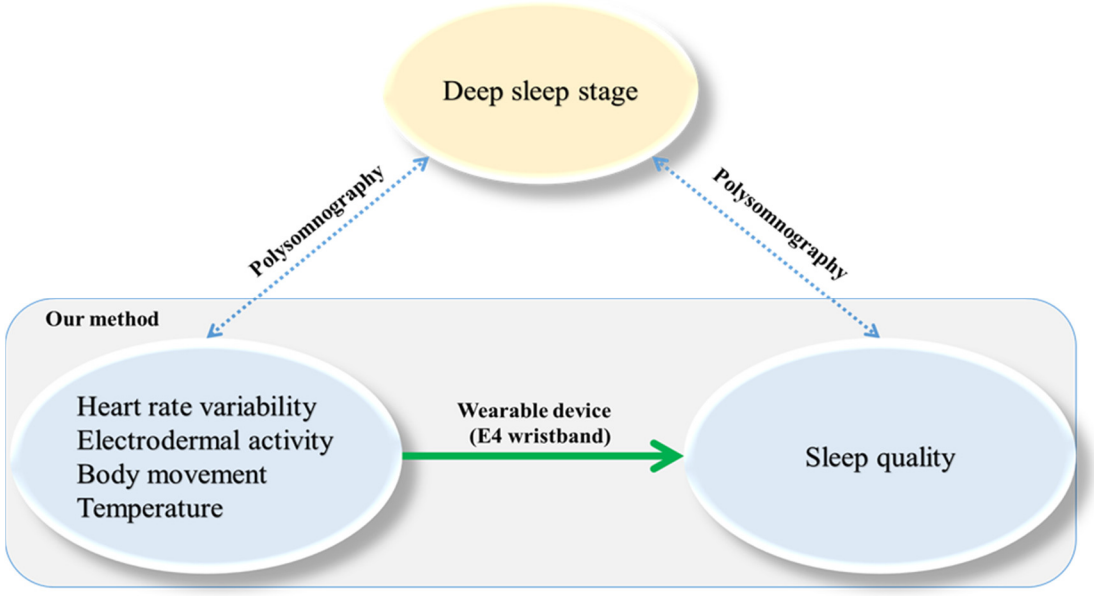

Figure 3.1-1 The proposed method

This methodology enables us to estimate SQ according to physiological signals, which are accessible by wearable devices like E4 wristbands. As shown in Figure 3.1-2, our sleep analysis system processes the recorded physiological signals through a four-state procedure, which includes signal preprocessing, feature extraction, feature selection, and classification.

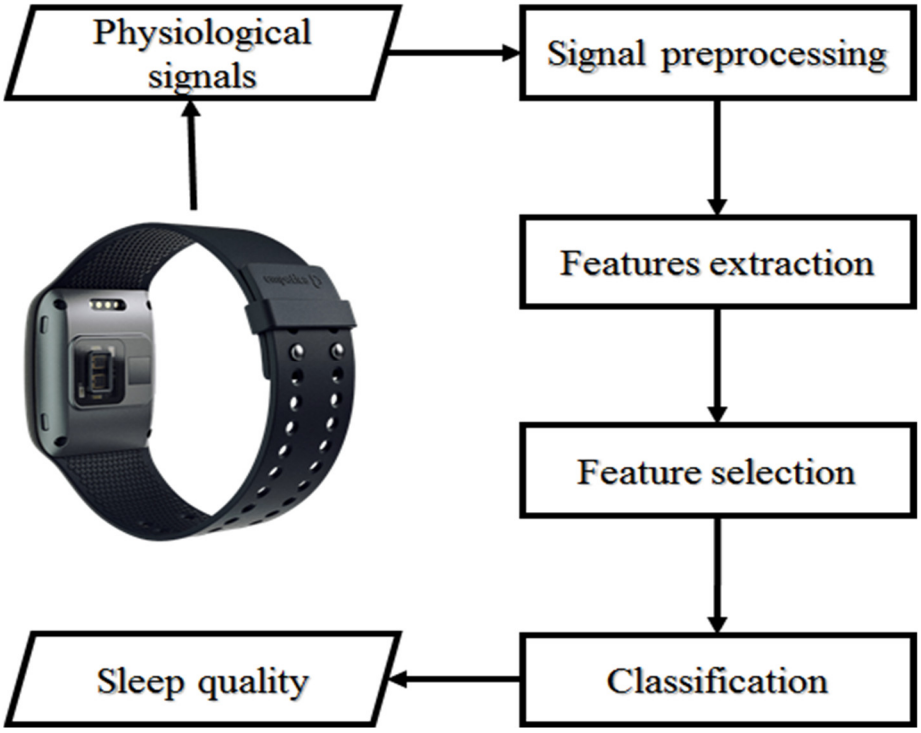

Figure 3.1-2 The proposed sleep analysis system for predicting sleep quality using E4 wearable device

Since distinct signals have different characteristics in SWS, each signal was processed in different ways with the aim of highlighting the SWS portion of sleep. We preprocessed signals using Poincare plot [57], filtering methods, Cole's function [58], and first derivative as described below:

PPG (also known as blood volume pulse), computed from pulse oximetry [59], is one of the four physiological signals recorded by E4 wristbands. We computed inter-beat intervals (IBIs) from PPG by taking the difference between two consecutive diastolic points, which indicates the lowest blood pressure in arteries when the heart rests between beats. IBI (RR intervals) describes the irregularity among two successive heart beats by measuring the variation in the beat-to-beat interval. To detect the SWS during sleep, we computed rRR (correlation coefficient of consecutive RR intervals) [57] in Poincare plot based on computed IBIs. This transformation distinguished intervals belonging to SWS if their rRR value was less than 0.1 units below the mean rRR of the initial 4 h for at least 10 min [57].

In electrodermal activity (EDA) signal, we utilized the validated algorithm of Cole's function [58] to identify EDA signals recorded during sleep. Next, the low-pass finite impulse response filtering with cutoff frequency 0.4 Hz and 32nd order was applied over these EDA time series to remove possible artifacts [60]. Next, the first derivative of EDA provided a map of EDA fluctuations during sleep stages. To define the intensity level of EDA fluctuation based on [37], we detected EDA peaks, which had values higher than 0.01 in a 30 s interval. The higher occurrence of peaks demonstrates a higher intensity level and a deeper sleep stage [37].

To investigate descending trends of accelerometer and skin temperature, we applied two transformation and filtering methods. We utilized the Poincare plot as described before to highlight the part of sleep showing a declining pattern in body movement and skin temperature signals. For the filtering method, a central moving average filter with a window length of 1 min was conducted over the two signals. Then, the portions of signals with negative slope are considered as decreasing trends of the time series.

In order to predict sleep quality, we estimated SWS time and length from different processed signals with twenty quantitative features (listed in Appendix A). For instance, the processed HRV signal using Poincare plot of rRR distinguished SWS segments during sleep as shown in the following figure. The number of occurrences of these segments can be considered as the SWS time feature. Similarly, the accumulated sum of the lengths of these segments provided an estimation for the SWS length feature.

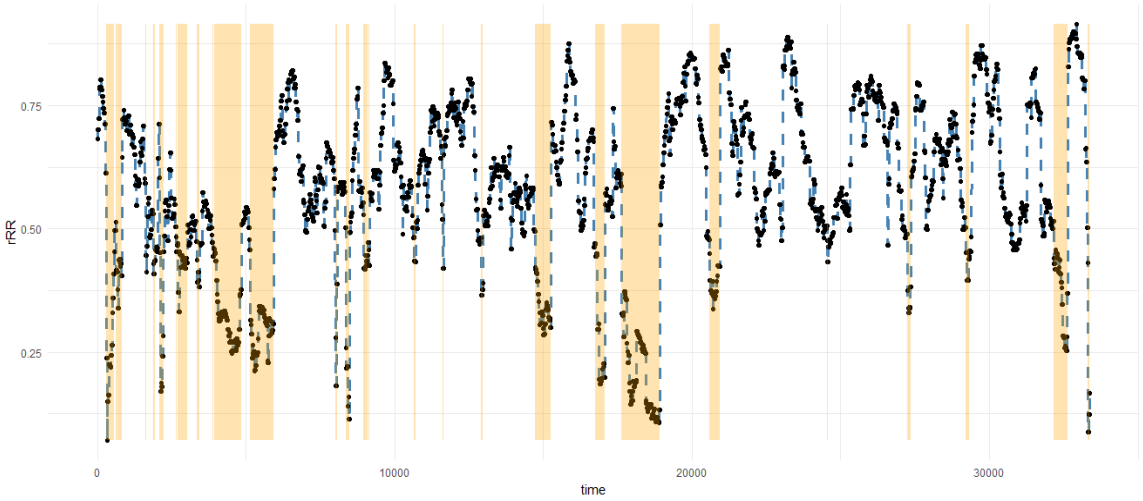

Figure 3.1-3 Prediction of SWS occurrence from HRV signal based on the Poincare plot

We employed a wrapper feature selection method called recursive feature elimination (RFE) [61] to reduce the complexity of models and faster training of the classifiers. To handle nonlinear relation between features, as well as reduce the risk of overfitting, we used a version of RFE called RFECV model, which incorporates resampling of k-fold cross-validation strategy [62]. 10-fold RFECV selects the optimal combination of features while maximizing the performance of the chosen classifiers. We selected three classifiers: naïve Bayes, random forest, and bagged tree with varying complexity.

The overall accuracy of classifiers was highly sensitive to the number of input features. As shown in Figure 3.1-4, the classifiers' performance does not follow a linear increasing trend with the growth of the number of features selected by RFECV. RFECV utilized the importance of features in each classifier to reduce the state space from 20! (20 factorial) possible subsets of features to only 20 cases. In this way, random forest and bagged tree gained the highest performance for feeling refreshed and sleep quality, respectively. The best model for feeling refreshed utilized 12 features while the sleep quality one employed 18 features.

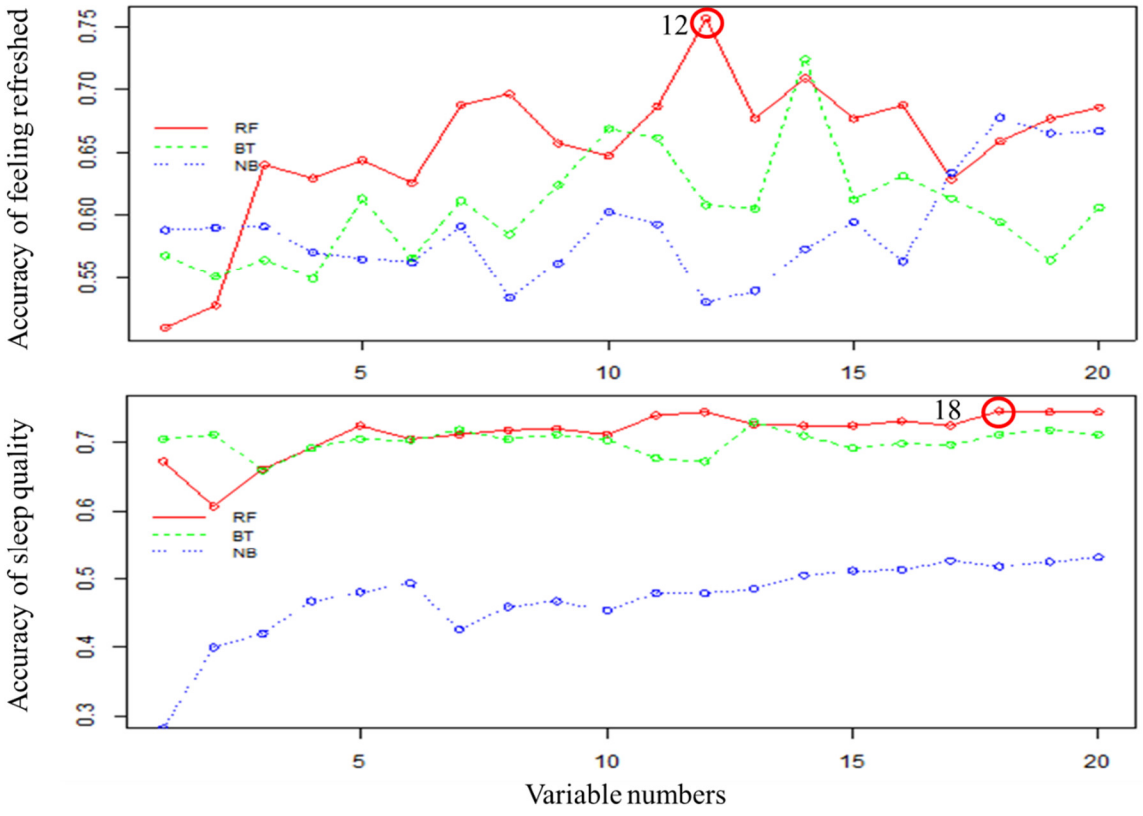

Figure 3.1-4 10-fold cross-validation recursive feature elimination (10 RFECV) process over random forest (RF), Bagged tree (BT), and naïve bays (NB) classifiers

3.1.3. Results

To evaluate reliability of best predicted models, we scrutinized performance of classifiers in terms of sensitivity, specificity, precision, and accuracy. These metrics provided valuable information about precise patients' diagnosis in medical assistant systems. By evaluating sleep records referring to feeling "refreshed" or "tired" after sleep, we considered it as a binary class problem. On the other hand, distinguishing SQs can be a multi-class problem. In this case, we employed one-vs-all evaluation methodology [63]. Finally, the average of all values per metric is reported as shown in Table 3.1-2.

Table 3.1-2 Comparison of the performance of classifiers with the best possible features

Labeling	Metric	Classifier		
		Random forest	Bagged tree	Naïve Bayes
Feeling refreshed	Accuracy	**0.73**	0.69	0.61
	Precision	**0.75**	0.58	0.42
	Sensitivity	0.34	**0.42**	0.34
	Specificity	**0.94**	0.83	0.75
Sleep quality	Accuracy	**0.75**	0.73	0.52
	Precision	**0.75**	0.74	0.52
	Sensitivity	**0.75**	0.70	0.52
	Specificity	**0.92**	0.91	0.84

Random forest outperformed in both labeling approaches for all metrics. Also, bagged tree and naïve Bayes were the second and third best classifiers according to performance, respectively. The order of classifiers' performance revealed there is a nonlinear relation among features such that more complex classifiers like random forest gain higher performance. Also, random forest obtains high specificity, which means the proposed method can detect records of tiredness in caregivers with high probability. As a result, our sleep recognition system can be a reliable system to alert the caregivers about the potential increase in caregiving-related stress or burnout.

To investigate the feature importance in the best classifier, we employed the Gini index, which is an important impurity criterion [64]. As shown in Figure 3.1-5, the feeling refreshed model utilized a lower number of features for making decisions compared to the sleep quality model. This fact is as a result of, the feeling refreshed model having lower variability of possible cases compared to sleep quality. Also, sleep efficiency was an important factor in both the sleep quality and feeling refreshed models. These results coincide with previous results relevant to CPWD sleep characteristics showing they experienced less total sleep time while taking longer to fall asleep [65].

Moreover, features relevant to temperature have the substantial effect on decision making in both models, especially in sleep quality. Indeed, the nocturnal temperature dysregulation as an age-related sleep disturbance contributes to fragmentation of sleep, which is a common predisposing factor for sleep complaints in caregivers [66]. Since importance features provided by models agree with well-established CPWD sleep studies, the proposed models can be applicable to CPWD.

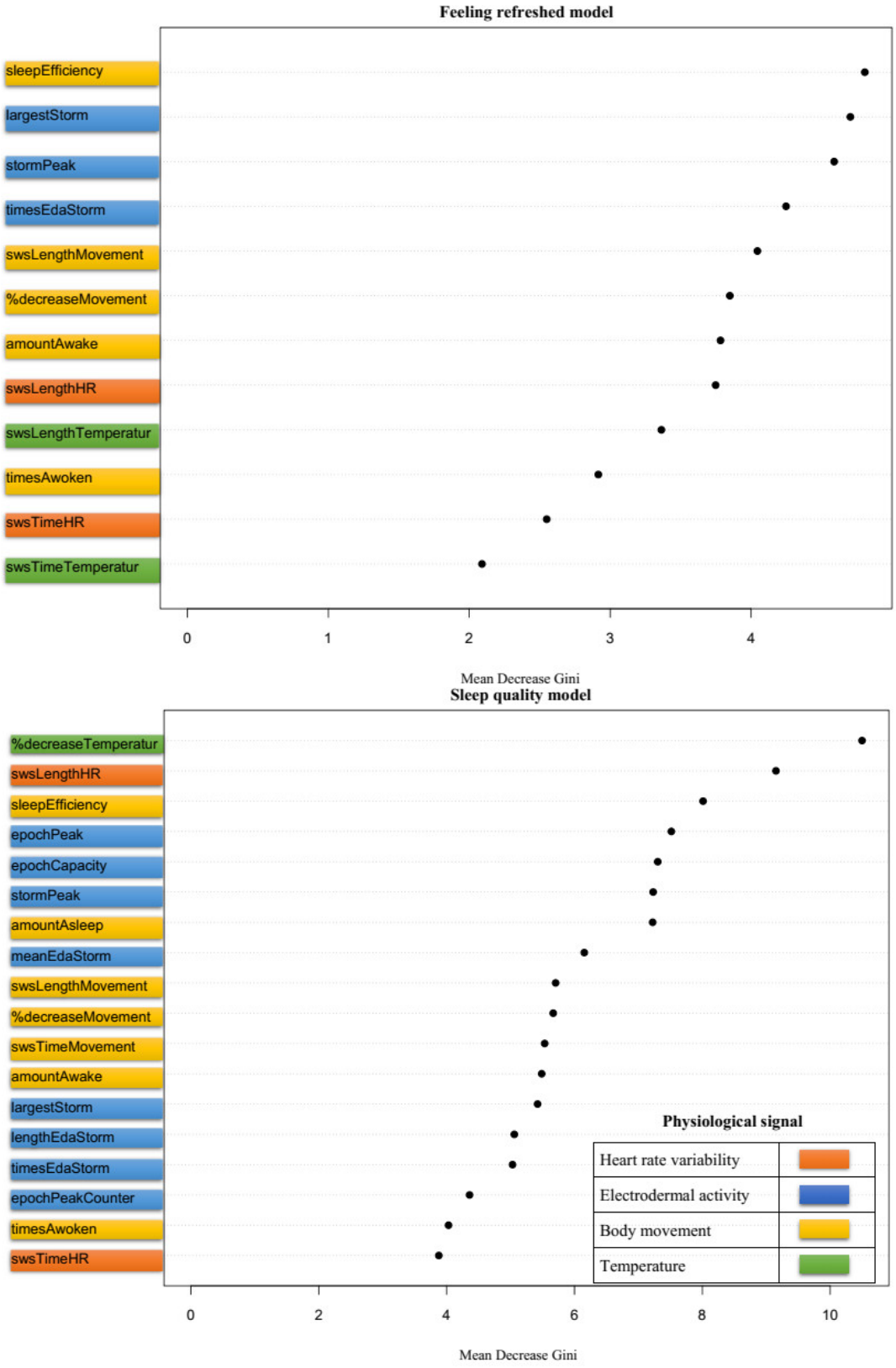

Figure 3.1-5 Feature Importance in the best models

3.2. Enhancing sleep quality prediction by machine learning techniques (Contribution: CS2)

3.2.1. Data

We utilized the PSG sleep records of Osteoporotic Fractures in Men (MrOS) sleep cohort [67]–[69]. Approval for the use of the MrOS sleep dataset in this study was obtained from the National Sleep Research Resource (NSRR). MrOS is a multi-center observational study of 5,994 men aged equal or greater than 65 years, who are the most susceptible group to sleep disorders, such as obstructive sleep apnea (OSA) [52], [70], [71]. Of the 3,933 provided MrOS PSG data records in the national sleep research resource [72], [73], 847 of the sleep records contain their R-beat annotations. The annotated R-beats described normal heart activity by excluding ectopic beats and artifacts [74]. Analysis of R-beat annotations using PhysioNet Cardiovascular Signal Toolbox [75] demonstrated that 792 sleep records satisfy preliminary requirements for computing HRV features, such as having length greater than or equal to five-min [74].

We extracted three types of features from the selected sleep records. In the first category of features, we selected nine features (listed in Appendix B) highly correlated to OSA, the most prevalent sleep disorder, based on the STOP-BANG questionnaire [7]. As the MrOS dataset is comprised of sleep data of men aged equal to or greater than 65, we did not consider feature numbers 6 and 8 in STOP-BANG, which investigate people older than 65 years and the gender of subjects, respectively. In the second category of features, we computed 23 HRV features (listed in Appendix C) based on the R-annotated files. Finally, the pictures of two ECG signals recorded during sleep nights formed the third category of features. To capture different lengths of sleep as another important feature, all ECG plots are drawn on the same axes such that they are plotted in the y ranges of -1 to 1 and x ranges of 1 to the length of the longest sleep night.

3.2.2. Methods

In this study, we investigated predicting SQ using three different methods: Two interpretable methodologies and one Deep Neural Network (DNN). The two interpretable methods were designed based on proven medical facts and transparent machine learning methods. The DNN method employs one of the current most powerful machine learning classifiers called Convolutional Neural Networks (CNN) with the aim of discovering the highest potential accuracy in explaining SQ through HRV.

In the first method, we tried to predict SQ by estimating the severity of sleep apnea, episodic obstruction of the upper airway during sleep [76]. The occurrence of this condition

equal or higher than 30 times per hour is known as obstructive sleep apnea (OSA). It is a common condition in the US [77] that is a leading factor of many health issues such as sleep fragmented, excessive daytime sleepiness, hypertension, and stroke [18]. Since OSA can be screened with the STOP-BANG questionnaire [52], we predicted SQ based on the features specified with STOP-BANG (listed in Appendix B). We fed the selected features to well-known and easy-to-interpret classifiers: Naïve Bayes, k nearest neighbors (kNN), decision tree, random forest, bagged tree, and linear discriminant analysis (LDA).

In the second method, SQ was predicted using estimated sleep stages extracted from HRV. As we elaborated in Section 2, the relation between physiological signals, like HRV, and sleep stages has been proven through abundant previous literature. In the case of HRV, this relation is so strong such that sleep stages can be distinguished using HRV [78]. Additionally, the transition and length of sleep stages are associated with SQ. For example, the reduction of SQ in the elderly is accompanied by the loss of slow wave sleep (SWS) stage [34]. Therefore, we predicted SQ from the estimated transition of sleep stages based on HRV during sleep in this method. We utilized annotated R-peaks for computing HRV. Since five-min is standard duration to ensure stationarity of the ECG time-series [74], HRV features were computed for each five-min window that slides 20 seconds of R-peaks using ten temporal, six frequency-based, two entropy, and five other features (listed in appendix C). The computed HRV features were fed into interpretable classifiers to estimate the dominant sleep stage, corresponding to each sliding window. Then the best classifier produced a sequence of predicted sleep stages for a recorded night. The generated sleep stages' sequences comprised five possible stages of weak, non-rapid eye movement (NREM) 1, NREM 2, SWS, and rapid eye movement (REM). Each sequence of sleep stages was encoded into the following features:

- The frequency of each sleep stage
- The number of series of each sleep stage
- The standard deviation of each sleep stage
- The number of series (length>1) of each sleep stage
- The longest series of each sleep stage
- The number of transitions among sleep stages
- Total length of sleep

The 27 extracted features from the predicted sleep stages were fed to six classifiers: Naïve Bayes, kNN, decision tree, random forest, bagged tree, and LDA.

Finally, we estimated SQ by feeding the raw electrocardiogram (ECG) signals of participants to a CNNs architecture. We fed the images of two ECG signals belonging to one night as the input. As shown in Figure 3.2-1, the proposed CNN architecture is comprised of two sections of feature extraction and classification. The feature extraction section extracted the complex relations between different regions of the input figures. This section is comprised of convolutional filters, max pooling, and dropout layers. The convolutional filters were employed to investigate the same patterns in different regions of

signals belonging to distinct sleep nights [79]. We utilized the pooling layers to reduce the dimension of data using non-linear down-sampling. The max pooling layers maximized values from each cluster of neurons at the prior layer [80]. The dropout layers reduced the chance of overfitting by randomly removing some of the neurons from the training process. The flattening layer connected the sections of feature extraction and classification by transforming two dimensional features to one vector as a required input of a dense layer (often called fully connected layer). In the dense layer, receptive fields, which are the inputs of neurons, are constructed from all the entire computed features. As a result, this layer investigated the relation between features processed via previous layers without any assumption. Following this, the softmax layer mapped all the computed relations to three types of SQ.

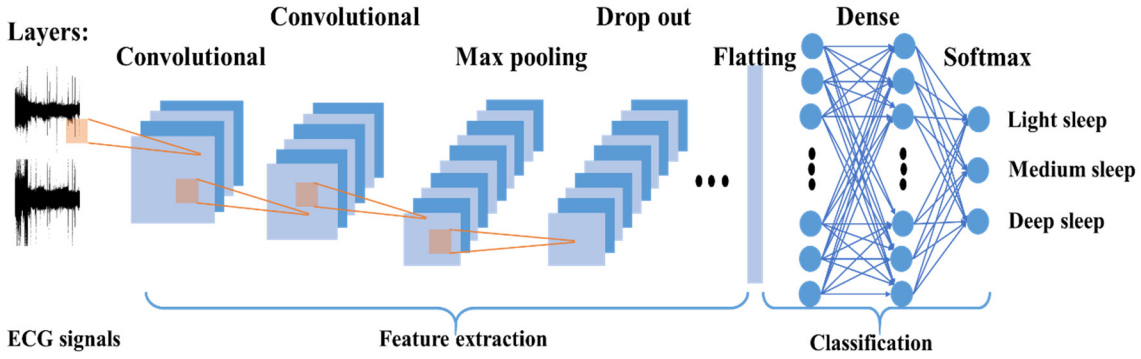

Figure 3.2-1 The proposed convolutional neural network architecture

3.2.3. Results

The cohort data was divided into two sets of train and test from 80% and 20% of all data, respectively. Then three different methodologies as described in the previous section were trained. The average performance of each classifier in predicting light, medium, and deep SQ using STOP-BANG features are reported in Table 3.2-1. We ordered the performance of different classifiers based on the F1 score to take both false positives and false negatives into account. Bagged tree with an F1 value of 0.35 gained the highest performance in SQ prediction using STOP-BANG features.

We conducted a suggested wrapping methodology called mean decrease accuracy (MDA) to estimate the variable importance in bagged tree [81], [82]. In this permutation importance, the feature whose ignorance in the training process led to the highest reduction was selected as the most important feature. As shown in Figure 3.2-2, feeling usual tiredness, BMI, and neck circumference are the most important features for distinguishing deep, medium, and light sleep records based on the STOP-BANG features. Usual tiredness is a common condition that forces patients who suffer from OSA to refer themselves to medical centers. High BMI and large neck circumference are two proven predictors of severe obesity, OSA, fragmented sleep, and low sleep quality [83].

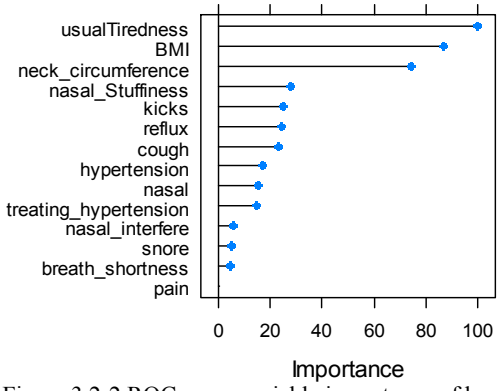

Figure 3.2-2 ROC curve variable importance of bag tree in STOP-BANG features

Table 3.2-1 Performance of six interpretable classifiers in STOP-BANG method

	F1	Accuracy	Sensitivity	Specificity
Decision tree	0.27	0.35	0.33	0.67
Naïve Bayes	0.30	0.38	0.37	0.68
LDA	0.32	0.40	0.37	0.69
Random forest	0.34	0.35	0.34	0.67
KNN	0.34	0.37	0.35	0.68
Bagged tree	**0.35**	**0.36**	**0.36**	**0.68**

In the second methodology, the LDA classifier with an accuracy of 0.41 outperformed the other classifiers in predicting SQ. To scrutinize the impacts of HRV features in separating sleep stages, we also investigated variable importance of separating sleep stages using HRV. This analysis demonstrated that ultra-low frequency (ULF) was the most important feature in separating sleep stages. This finding coincided with the fact that the value of very low frequency was significantly lower in deep sleep stage than in the other sleep phases [84].

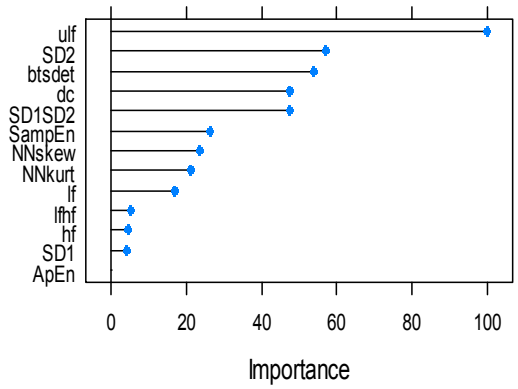

Figure 3.2-3 ROC curve variable importance of separating sleep stages using HRV

Table 3.2-2. Performance of six interpretable classifiers in HRV method

	F1	Accuracy	Sensitivity	Specificity
Naïve Bayes	0.24	0.29	0.34	0.67
Random forest	0.44	0.28	0.34	0.67
Decision tree	0.44	0.28	0.33	0.67
Bagged tree	0.44	0.28	0.33	0.67
KNN	0.46	0.36	0.34	0.67
LDA	**0.47**	**0.41**	**0.38**	**0.69**

Finally, we fed ECG pictures into the CNN architecture with the aim of finding the SQ of participants. The evaluation results demonstrated that the CNN can achieve an average accuracy of 0.6 for separating SQ in three categories of light, medium, and deep sleep. These promising outcomes from a CNN demonstrated the capability of using AI methodology in processing raw data in contrast to using pre-engineered features. Also, it can enhance the speed of processing PSG data for predicting SQ by reducing the number of signals and removing the detecting R-peaks step.

3.3. Identifying the disease-specific risk factors of daytime sleepiness (Contribution: BI2)

3.3.1. Data

These study data were derived from Sleep Heart Health Study (SHHS) [85], [86] as a large home-based PSG database of people aged 40 years or older without a history of OSA treatment. Approval for the use of the SHHS sleep dataset in this study was obtained from the National Sleep Research Resource (NSRR). The baseline examination, visit 1, of the SHHS recruited 6441 individuals between 1995 and 1998. During the home visit, the technicians interviewed the participants using a standardized questionnaire to collect information on medical history and health-related characteristics.

PSG was performed with 12-lead Compumedics PS equipment. This portable, unattended monitoring setup enabled recording of this large in-home PSG. The multiple channels recorded several physiological signals, including electroencephalogram, electrocardiogram, airflow, chin electromyogram, abdominal and thoracic excursions, oxyhemoglobin saturation, and body position. These leads were connected to a small monitor worn in a vest pocket so that participants could be fully ambulatory while awake. The PSG records and a great deal of covariates are publicly available at National Sleep Research Resource (NSRR) [72], [73].

Our analytical records included 5804 SHHS participants who successfully completed baseline PSG. Excluding sleep records with missing values in ESS left 5583 records with valid labels. Sleep efficiency and alcohol usage features contained 2281 and 403 missing values, respectively. Since these two features did not provide information for many records, we excluded them from further analyses. In addition, we narrowed our analyses to 4445 complete records by removing all records that contained any missing value. All utilized features are listed in Appendix D.

We employed the well-known Epworth sleepiness scale (ESS) [87] to measure subjective daytime sleepiness. ESS measures the probability of falling asleep in distinct situations, such as watching TV, reading, or laying down to rest. This metric reports the severity levels of sleepiness by an integer number in the range of 0-24 (The higher values show more severe daytime sleepiness). In the medical domain, this scale usually is dichotomized and converted to a binary category of Non-EDS (0-10) and EDS (11-24) [12], [88], [89].

To investigate EDS prevalence, we analyzed the distribution of ESS using three plots that do not use cutting points as shown in Figure 3.3-1. As plot C demonstrates, the two tails of ESS distribution in Q-Q (quantile-quantile) plot violate the normal distribution pattern. Indeed, the black circles did not fall on the two tails of red lines of the Q-Q plot. This fact is also reflected in plot B such that the right whisker of box plot has nearly twice the length

of other three quantiles. Such a left skewed distribution of ESS, also visualized in plot A, implied an imbalanced dataset of daytime sleepiness. The box plot illustrates that only one quantal (25% of sleep records) belongs to people who suffer EDS (ESS ≥11). As a result, we used F1-score in addition to sensitivity, specificity, accuracy metrics to describe the performance of classifiers in distinguishing sleep records of people with EDS condition from non-EDS. Using F1-score guaranteed that the probable high performance of a classifier in predicting EDS is a result of identifying real pattern of data without being disturbed by many false alarms.

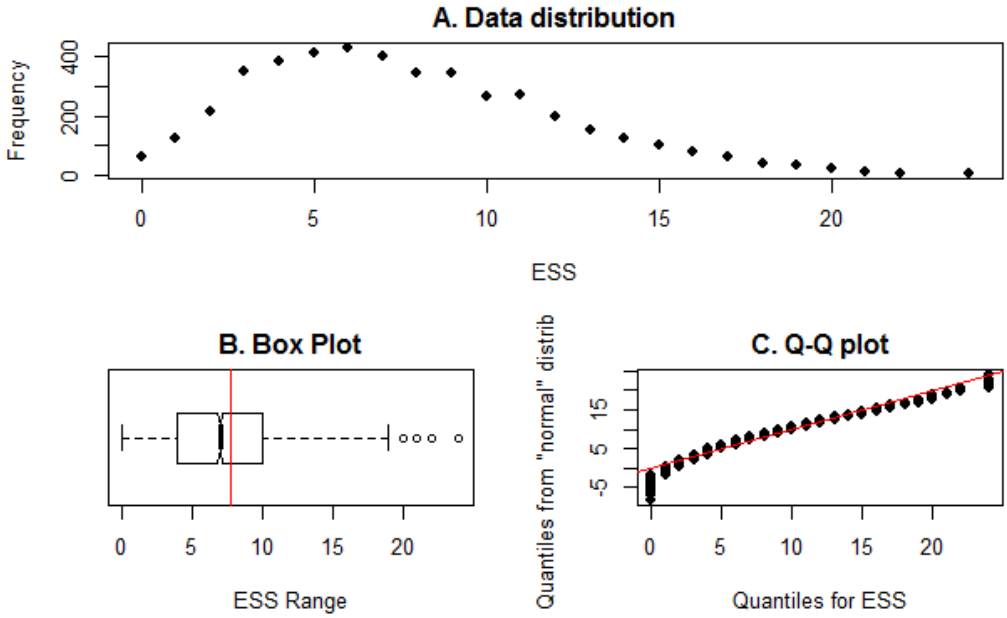

Figure 3.3-1 Epworth sleepiness scale (ESS) distribution in SHHS visit one across 4445 completed sleep records

To better characterize and predict EDS, we investigated describing EDS in terms of several probable disease-specific risk factors of sleep fragmentations (measured by averaged respiratory event-related desaturation (HB_{EV}) [24]), sleep propensity (calculated by odds ratio product (ORP) of EEG [90], [91] and sleep awakening index [20]), sleep resilience to disruptive stimuli (estimated using spindles activity [93]), and non-restorative sleep (observed by alpha intrusion [94], [95]).

To measure HB_{EV}, we detected significant reduction of nasal flow or abdominal respiration as a sign of airflow obstruction, known as obstructive sleep apnea and hypopnea. As shown in the following figure, these respiratory events cause significant reduction of oxygen desaturation in blood. We computed HB_{EV} as the average area under the curve of oxygen desaturations corresponding to respiratory events.

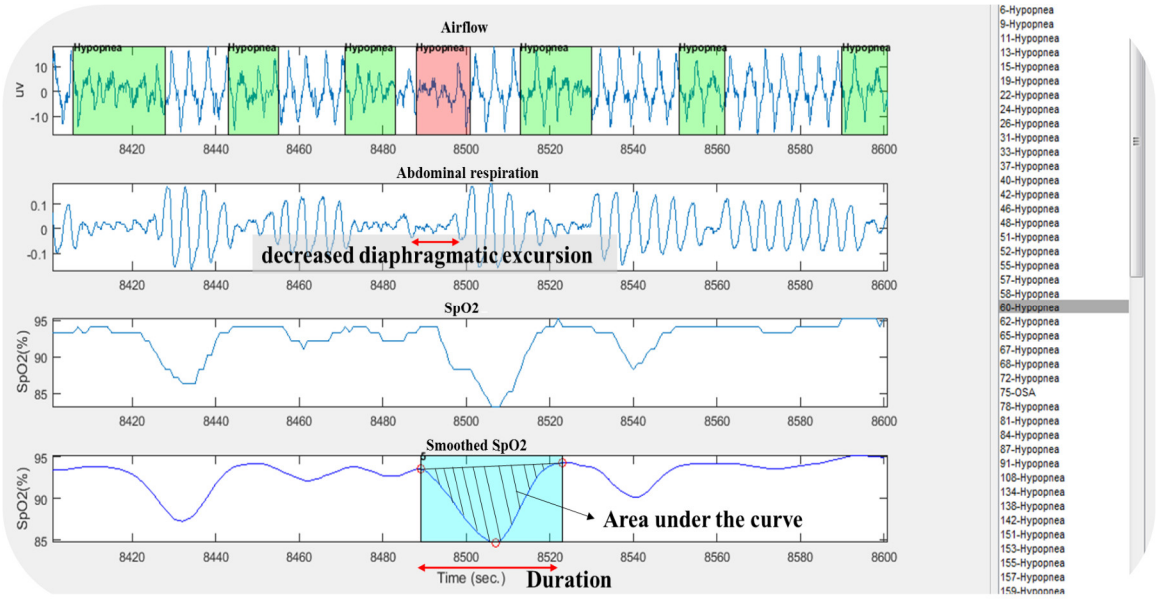

Figure 3.3-2 Blood oxygen desaturation as a result of obstructive sleep apnea or hypopnea

Additionally, we utilized both clinical and common PSG features as covariates. The clinical features are easily obtainable from a clinical interview, including: age, gender, race, Body mass index (BMI), systolic and diastolic blood pressures, diabetes, cardiovascular disease (CVD), depression, smoking, consumption of caffeinated beverages, having insomnia, and average sleep duration on working days.

3.3.2. Methods

We performed several statistical analyses in three levels to investigate the effect size of each PSG feature, a probable disease-specific risk factor, in predicting the target variable of ESS. In the first level, ESS was modeled by nine logistic regressions with new individual PSG features. Following this, the adjustment of level 1 was modeled using the clinical features from the models of level 2. Finally, the level 2 models were adjusted by common PSG features. Tracking the change of odds ratio and statistical significance of new PSG features by adjusting the other covariates can explain the direction and strength of each new individual feature in modeling EDS. Additionally, we performed several regression analyses using interaction terms and stratifying data based on different levels of sleep apnea severity.

To control both type I and II errors in our statistical analyses, we utilized a recommended error control criteria for health studies called false discovery rate [96]. This criteria is defined as the expected proportion of falsely rejected hypotheses among all rejections [97]. We used the Benjamini and Yekutieli FDR [98], which can be used under general dependence among multiple tests.

To predict subjective daytime sleepiness via objective measurements, we employed four simple classifiers of logistic regression, naïve Bayes, decision tree, and K-nearest neighbors (KNN). The simplicity of these classifiers helped us to easily interpret and visualize the decision-making process in addition to investigating the nonlinear relationship among the predictive factors. We evaluated the performance of classifiers by 10-times 10-fold cross-validation strategy to avoid predicting by chance or time dependency among records. The final reported metrics from our evaluation methodology were the results of 100 independent experiments.

3.3.3. Results

3.3.3.1. Identification of the disease-specific risk factors

Since frequency features (HB_{EV} and Awakening index) are affected by the apnea-hypopnea index (AHI) severity level, we compared frequency features and physiological trait features separately. As shown in Figure 3.3-3, our statistical tests for the physiological trait features revealed that ORPTRT, the sleep depth during all recording time, was an independent stimulator of daytime sleepiness even when adjusting for both clinical and conventional PSG features. Since lower ORP shows higher sleep depth, this outcome revealed that the patients with EDS had drowsiness in all the recording time in addition to sleep time.

Increasing some brain activities in terms of spindles (C34_Power) and alpha intrusion (Average_alpha) may decrease daytime sleepiness; however, these features lose their statistical significance when adjusting for other features. Also, we found there was no relation between spindle duration and daytime sleepiness.

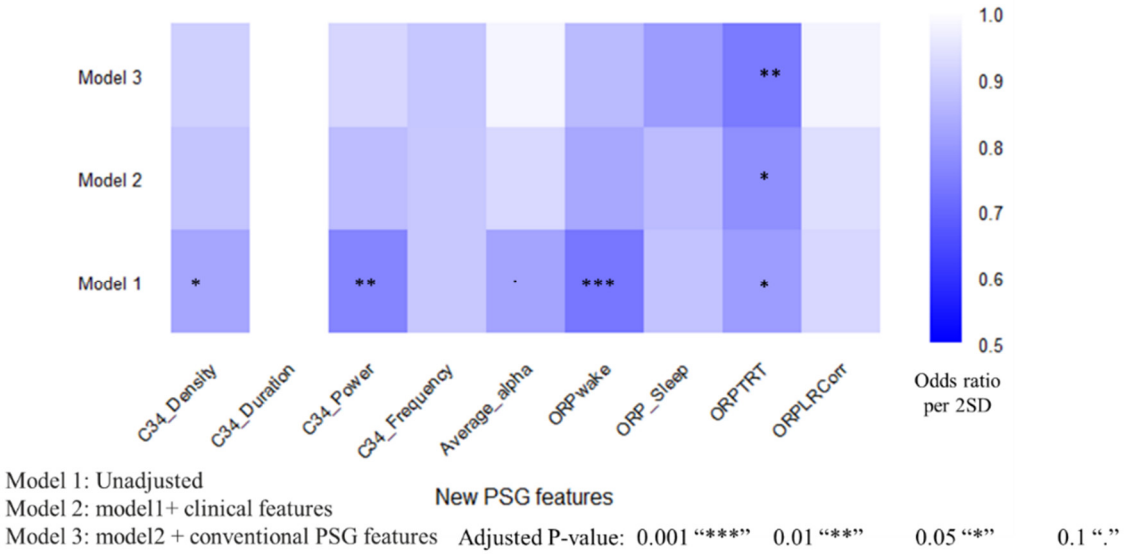

Figure 3.3-3 Effect size of the trait features over all SHHS. The blue color (OR<1) is associated with lower probability of EDS

We examined the roles of frequency features (HB_{EV} and awakening index) in developing EDS in different severity levels of obstructive sleep apnea (OSA). We segregated AHI, the severity measurement of OSA, into four categories of normal (AHI<5), mild (5=<AHI<15), moderate (15=<AHI<30), and severe (AHI>=30). Using the interaction terms in the logistic models, we evaluated the association of the frequency features in different levels of sleep apnea. As shown in the following figure, both HB_{EV} and awakening index odds ratios increased in response to the higher level of AHI. The effect size of these two features was more than twice in OSA in contrast to normal participants. Since both these two frequency features, AHI severity levels, and their interactions were statistically significant, both HB_{EV} and awakening index in addition to AHI should be considered in daytime sleepiness prediction models.

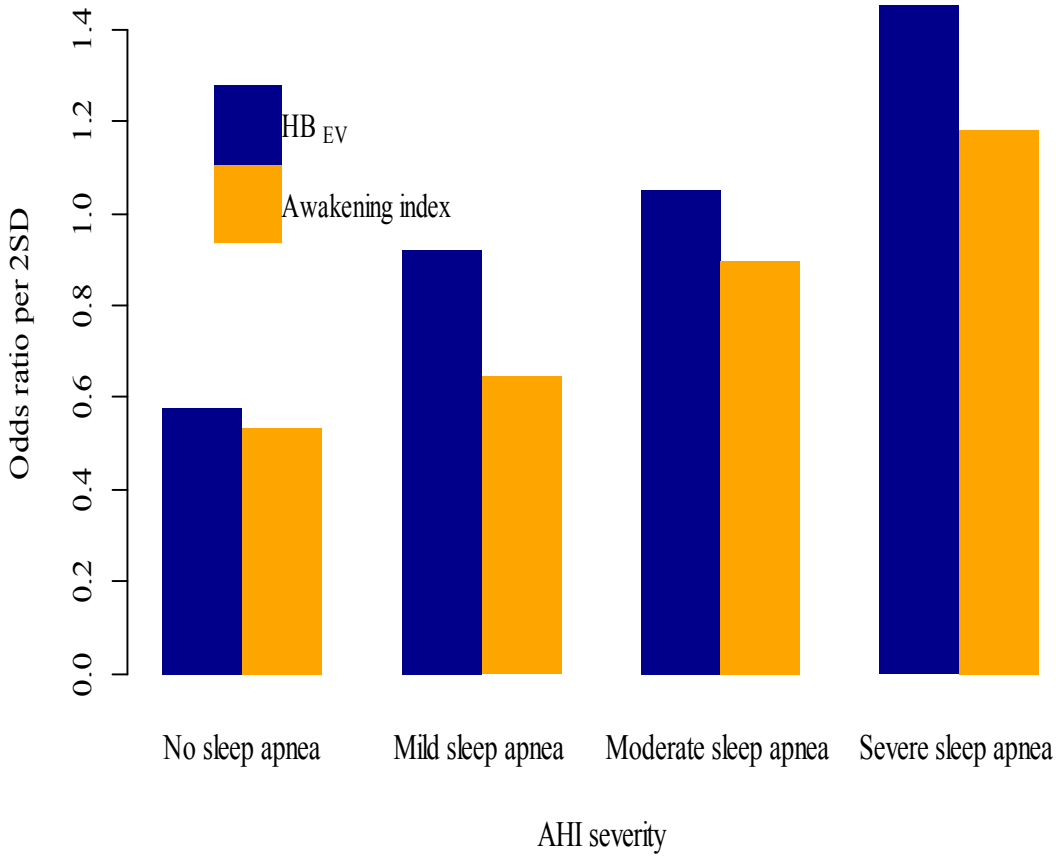

Figure 3.3-4 The statistically significant effects of HB_{EV} and Awakening index at different levels of AHI severity

The two previous analyses demonstrated the significance of ORPTRT, HB_{EV} and awakening index in daytime sleepiness. To further assess the separate associations between these three risk factors of daytime sleepiness, we stratified the samples into those with moderate to severe OSA (AHI≥15 events/hour) and those with mild or Non-OSA (AHI<15 events/hour). As shown in Table 3.3-1, HB_{EV}, ORPTRT, and awakening index are distinct risk factors of EDS in patients suffering from OSA and Non-OSA. In those with moderate to severe, HB_{EV} was a significant predictor of daytime sleepiness. For a one-unit increase in HB_{EV}, the odds of having EDS increase by 43% (P-value<0.001) in Moderate/severe OSA patients. HB_{EV} was no longer a significant predictor of sleepiness in those with

AHI<15. In contrast, ORPTRT and awakening index statistically significantly described sleepiness in those with mild or Non-OSA. As higher ORPTRT and Awakening index reversely describe the sleep propensity, the occurrence of EDS is directly associated with sleep propensity. These findings revealed that EDS has distinct disease-specific risk factors in OSA patients and Non-OSA people. The severer OSA-specific hypoxia, measured by HB_{EV}, in OSA patients' causes higher fragmented sleep which consequently leads to EDS. However, the high sleep propensity may be due to genetic bases [99] continuously presenting the sleepiness feeling in Non-OSA subjects.

Table 3.3-1 The odds ratios per 2SD of the three significant risk factors of daytime sleepiness in stratified logistic regressions; P-value<0.001 '***'

	Mild or Non-OSA N= 2897	Moderate or severe OSA N= 1548
HB_{EV}	0.81	1.43***
ORPTRT	0.70***	0.86
Awakening index	0.62***	1.09

3.3.3.2. Prediction of daytime sleepiness

We examined the predictability of daytime sleepiness on SHHS dataset using both clinical and PSG features, especially the disease risk factors. The detailed definition of all the features is listed in Appendix D. As shown in Figure 3.3-5, Logistic regression and decision tree reported higher sensitivity while K-nearest neighbors (KNN) and Naïve Bayes (NB) provided higher specificity. In the first point of view, we would suggest using logistic regression with average sensitivity of 0.57 for the task of identifying people who suffer EDS, and KNN with average specificity of 0.9 for ruling out people who do not have the disease.

However, physicians are more interested in utilizing a clinical decision support system (CDSS) that flags anything that could be dangerous with an acceptable range of false alarms. We used the common metric of accuracy to evaluate the performance of classifiers in detecting both EDS and non-EDS sleep records. The accuracy metric identified KNN and NB as the best classifiers with an accuracy of 0.70. However, this value was reached at the expense of very low average sensitivity values (0.13 for KNN and 0.24 for NB). The reason for such poor sensitivity is that SHHS is an imbalanced dataset from EDS prevalence point of view. By considering both false negative and false positive, F1-score reported the quality of classification for the both categories better than the accuracy. Logistic regression with an average F1-score of 0.42 outperformed the other classifiers. As logistic regression utilized a hyperplane to separate EDS and non-EDS sleep records, it managed to provide a generalized decision boundary in this large dataset with low variance in all metrics of sensitivity, specificity, accuracy, and F1-score.

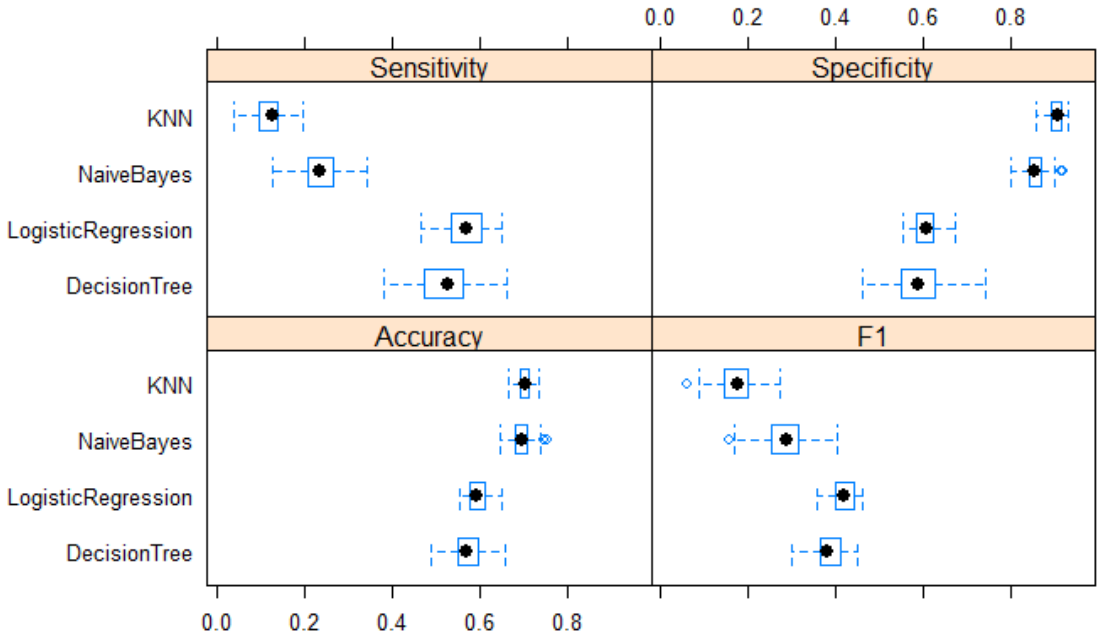

Figure 3.3-5 The outcomes of classifiers in separating EDS and non-EDS records of SHHS visit one across 4445 completed sleep records

To examine the role of each feature in the best classifier, we computed the odds ratios of all the features in the logistic regression model. The odds ratio reveals both the direction and strength of association between exposures and daytime sleepiness. The odds ratios of value equal to, greater than, and less than one demonstrate neutral, positive, and negative effects of the predictors, respectively. Additionally, the descending order of |1-odds ratios| represents the features based on their importance. Figure 3.3-6 illustrates the features importance in logistic regression using odds ratios. In the following, we investigated the five most important features that have either positive or negative direct relation to daytime sleepiness.

NTCA (non-tricyclic antidepressants), HB (hypoxic burden), and timest2p (the NREM2 percentage) with odds ratios of 1.51, 1.48, and 1.34 are the three most important features that have a positive direct relation to daytime sleepiness. This analysis demonstrates that taking NTCA pills is associated with more daytime sleepiness. Indeed, some NTCA antidepressants disrupt the sleep as a result of activating side effects, including anxiety, agitation, and akathisia [100].

The odds ratio higher than one reveals that increasing the HB elevates the chance of suffering EDS. As HB demonstrates the total area under oxygen desaturation curve in association with airway obstruction, this metric describes both frequency of the airway obstructions, measured by AHI, and the severity of hypoxia, measured by HB_{EV}. Therefore, the higher value of HB demonstrates the more fragmented sleep as a result of the airway obstructions and less time to have restorative sleep. The significant role of HB in predicting EDS is another sign that shows the need for considering this disease-specific risk factor in daytime sleepiness analyses.

The higher NREM2 percentage (timest2p), which forms roughly half of the night [7], reduces the chance of longer SWS, which is the most restorative sleep stage. A one-unit increase in NREM2 sleep stage elevates 34% odds of having EDS.

The fourth and fifth important features have negative direct relation to daytime sleepiness. HrsWD02, the average sleep duration on working days, with odds ratios of 0.69 has a reverse connection to EDS such that the more sleep duration is accompanied by less daytime sleepiness. This fact is aligned with our finding regarding higher sleep propensity in the normal people as a risk factor of higher daytime sleepiness. Finally, females' odds of having daytime sleepiness are smaller by a factor of 0.73. As men are at the higher risk of suffering OSA [52], [101], the prevalence of daytime sleepiness is higher in men rather than women.

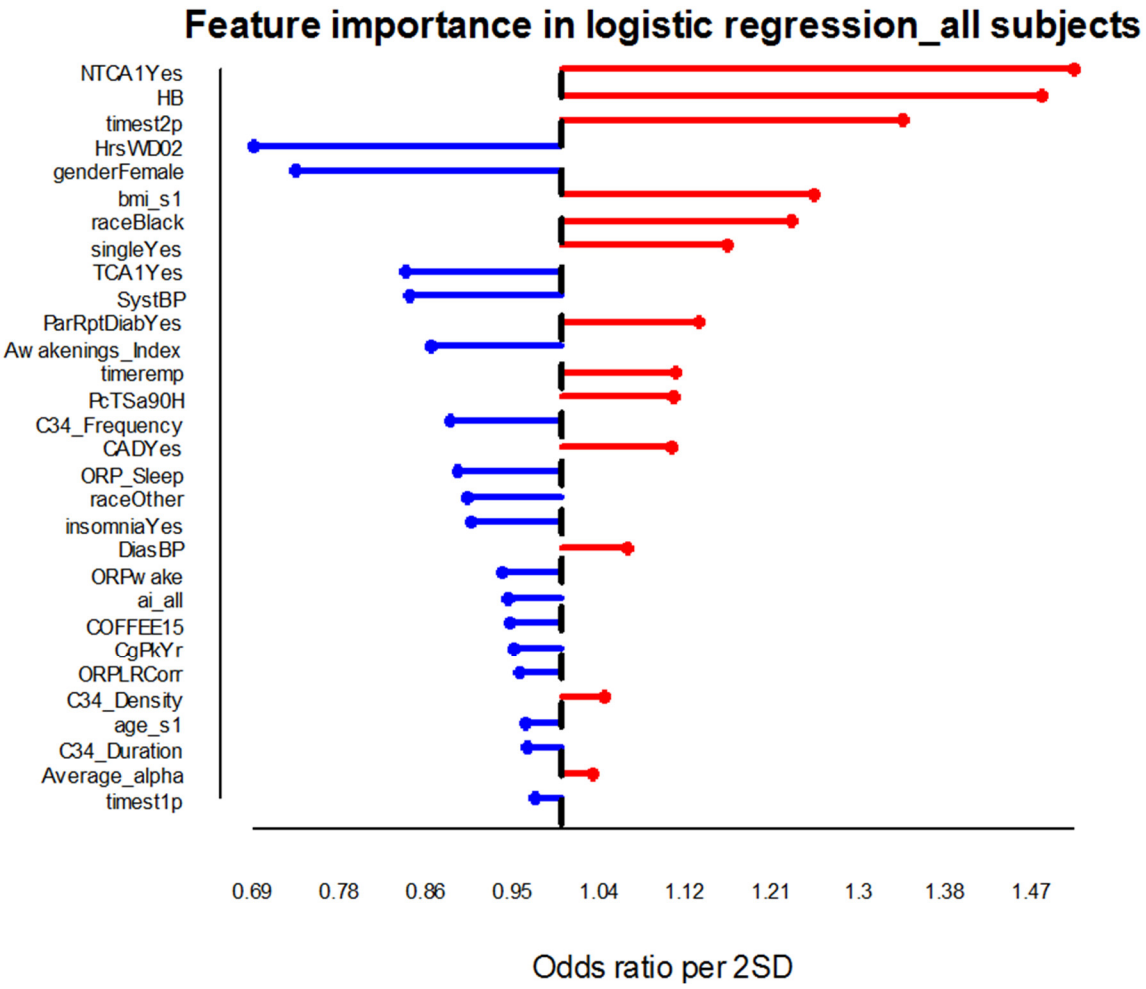

Figure 3.3-6 Feature importance in logistic regression
The red color (OR>1) and blue color (OR<1) are associated with higher probability of EDS and Non-EDS, respectively

4.Conclusion

In this dissertation, we presented our work in enhancing sleep quality (SQ) prediction using an easy-to-use wearable device, machine learning techniques, and identifying significant disease-specific sleep-related risk factors in older adults. We demonstrated how using wearable devices and machine learning techniques can enhance the prediction of SQ by accessing the raw physiological signals without changing the sleep routine and analyzing the recorded signals independently from sleep-expert knowledge. Also, we identified disease-specific sleep-related risk factors in PSG data that can lead to an efficient way for SQ estimation. In this section, I review the main findings from the proposed methods and their limitations. Following this, I present my future work and the list of my publications throughout my Ph.D.

4.1. The main findings

Contribution CS1 enabled us to evaluate the SQ of older caregivers of people with dementia (CPWD) by incorporating easy-to-interpret time-domain features extracted from multiple physiological signals. Our proposed models for SQ and feeling refreshed selected 18 and 12 computed features, respectively. Our analyses of the selected features demonstrated that sleep efficiency and skin temperature are significant features in predicting SQ and feeling refreshed in CPWD, respectively.

The convolutional neural network (CNN)-based method proposed in contribution CS2 outperforms the other interpretable methods in predicting SQ from medical history and heart rate variability (HRV). The developed CNN-based method predicts SQ using two electrocardiogram signals with an accuracy of 0.6. However, the interpretable methodologies based on STOP-BANG features (listed in Appendix B) and HRV features (listed in Appendix C) reached accuracies of 0.41 and 0.36, respectively. These outcomes demonstrated how using machine learning techniques enhances the current sleep expert knowledge of SQ prediction.

Contribution BI1 leads to the introduction of a clinical decision support system that predicts both SQ and feeling refreshed in CPWD using an easy-to-use and non-invasive wearable E4 wristband. We managed to monitor the physiological signals of CPWD for two-week periods in their home environment. We reached an accuracy of 75% in predicting the SQ of CPWD based on the extracted features from the raw recorded signals. This outcome reveals the capability of using wearable devices in assessing the SQ of older adults.

In contribution BI2, we identified a disease-specific sleep-related risk factor of excessive daytime sleepiness (EDS) as a sign of abnormal SQ. We showed that the average areas under the oxygen desaturation curve in association with apnea/hypopnea is a significant predictor of EDS in patients with OSA. We also illustrated that sleep depth measured by the odds ratio product of electroencephalogram is a significant risk factor for EDS in people with no history of OSA.

4.2. Limitations

Through all the analyses, we demonstrated how using wearable devices and machine learning methods can aid older adults to track their own sleep quality. As a result, all our findings in this dissertation are limited to older adult communities. Although we performed our analyses on various sleep cohorts and reported the outcomes as accurately as possible, it is still essential to evaluate the proposed methods and reexamine our findings on other sleep cohorts. The introduced causalities and risk factors are based on cross-sectional data, and there is a need to assess these findings using longitudinal data. All the proposed methods were performed based on a predefined algorithm while sleep pattern can change during a night; therefore, there is a need to investigate real time algorithms that track these changes during a sleep. Finally, we emphasize that all our proposed methods should be considered as tools for helping sleep physicians in a process of sleep assessment not as a replacement.

4.3. Future work

This work was the pioneering study of predicting sleep quality using a great deal of machine learning techniques. The interpretable techniques provided easy-to-understand decision-making process based on the pre-engineered features. These were introduced by domain experts and extracted from both clinical and physiological signals. But, these techniques were limited to evaluating the segregation power of pre-engineered features. On the other hand, the black-box techniques provided higher performance in segregating sleep data by considering non-affine relationships. However, such techniques did not provide a clear explanation regarding the causality of its outcomes.

To examine the capability of distinct physiological signals in predicting sleep quality, we will use deep learning techniques that are the most current powerful black-box techniques. Following this, we will evaluate how considering a group of physiological signals can elevate the overall performance of segregating sleep data. To bring light to the internal logic of black-box techniques, we will visualize different layers in the final model by reverse engineering of the deep learning architecture.

Finally, we also plan on identifying the strength of the current physiological-based features using the interpretable classifiers. The direction and effect size of each feature in multiple datasets will demonstrate the reliability of current features in predicting sleep quality. The instability of the direction and effect size of features in distinct datasets highlight uncomplimentary of the current features.

4.4. Publications

The list of publication directly related to sleep quality:

- **R. Sadeghi**, T. Banerjee, J. C. Hughes, & L. W. Lawhorne (2019), "Predicting sleep quality of caregivers using physiological signals", *Computers in Biology and Medicine,*110, 276-288. (Impact factor: 2.115)
- **R. Sadeghi**, T. Banerjee, J. Hughes (2020), "Predicting sleep quality using heart rate variability", Accepted in 42nd *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBC2020)
- **R. Sadeghi**, M. Younes, S. Sands, L. Taranto-Montemurro, S. Bertisch, A. Wellman, S. Redline, T. Banerjee, A. Azarbarzin (2020), "Sleep Propensity And Sleep Apnea-specific Hypoxia Are Associated With Excessive Daytime Sleepiness", Accepted in *Annals of the American Thoracic Society* (ATS2020)

The list of publication in analyzing physiological signals:

- **R. Sadeghi**, T. Banerjee, W. Romine (2018), "Early Hospital Mortality Prediction using Vital Signals", *Smart Health*, 9, 265-274.
- **R. Sadeghi**, T. Banerjee, W. Romine (2020), "Predicting alcohol withdrawal in intensive care units", presented at Symposium of Student Research, Scholarship, and Creative Activities Materials 2020.

References

[1] C. B. Saper, "Introduction to the Topic," in *Sleep Disorders Medicine*, Springer, 2017, pp. 3–4.

[2] J. S. Ruggiero and N. S. Redeker, "Effects of napping on sleepiness and sleep-related performance deficits in night-shift workers: a systematic review," *Biological research for nursing*, vol. 16, no. 2, pp. 134–142, 2014.

[3] A. J. Phillips *et al.*, "Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing," *Scientific reports*, vol. 7, no. 1, p. 3216, 2017.

[4] H. Yoon, S. H. Hwang, J.-W. Choi, Y. J. Lee, D.-U. Jeong, and K. S. Park, "Slow-Wave Sleep Estimation for Healthy Subjects and OSA Patients Using R–R Intervals," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 119–128, 2018.

[5] S. Dobing, N. Frolova, F. McAlister, and J. Ringrose, "Sleep quality and factors influencing self-reported sleep duration and quality in the general internal medicine inpatient population," *PloS one*, vol. 11, no. 6, p. e0156735, 2016.

[6] H. A. Beydoun *et al.*, "Sex and age differences in the associations between sleep behaviors and all-cause mortality in older adults: results from the National Health and Nutrition Examination Surveys," *Sleep medicine*, vol. 36, pp. 141–151, 2017.

[7] S. Chokroverty, "Overview of normal sleep," in *Sleep Disorders Medicine*, Springer, 2017, pp. 5–27.

[8] S. Chokroverty, "Erratum to: Sleep Disorders Medicine," in *Sleep Disorders Medicine*, Springer, 2017, pp. E1–E1.

[9] V. Ibáñez, J. Silva, and O. Cauli, "A survey on sleep assessment methods," *PeerJ*, vol. 6, p. e4849, 2018.

[10] V. Ibáñez, J. Silva, and O. Cauli, "A survey on sleep questionnaires and diaries," *Sleep medicine*, vol. 42, pp. 90–96, 2018.

[11] A. Roebuck *et al.*, "A review of signals used in sleep analysis," *Physiological measurement*, vol. 35, no. 1, p. R1, 2013.

[12] N. A. Eiseman, M. B. Westover, J. E. Mietus, R. J. Thomas, and M. T. Bianchi, "Classification algorithms for predicting sleepiness and sleep apnea severity," *Journal of sleep research*, vol. 21, no. 1, pp. 101–112, 2012.

[13] D. J. Buysse, "Sleep health: can we define it? Does it matter?," *Sleep*, vol. 37, no. 1, pp. 9–17, 2014.

[14] A. D. Krystal and J. D. Edinger, "Measuring sleep quality," *Sleep medicine*, vol. 9, pp. S10–S17, 2008.

[15] R. Sadeghi, T. Banerjee, J. C. Hughes, and L. W. Lawhorne, "Sleep quality prediction in caregivers using physiological signals," *Computers in Biology and Medicine*, 2019.

[16] R. Sadeghi, T. Banerjee, and J. Hughes, "Predicting sleep quality using heart rate variability," Accepted in the 42*nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBC2020).

[17] E. Tobaldini, L. Nobili, S. Strada, K. R. Casali, A. Braghiroli, and N. Montano, "Heart rate variability in normal and pathological sleep," *Frontiers in physiology*, vol. 4, p. 294, 2013.

[18] S. Redline *et al.*, "Obstructive sleep apnea–hypopnea and incident stroke: the sleep heart health study," *American journal of respiratory and critical care medicine*, vol. 182, no. 2, pp. 269–277, 2010.

[19] G. Slater and J. Steier, "Excessive daytime sleepiness in sleep disorders," *Journal of thoracic disease*, vol. 4, no. 6, p. 608, 2012.

[20] C. Guilleminault and S. N. Brooks, "Excessive daytime sleepiness: a challenge for the practising neurologist," *Brain*, vol. 124, no. 8, pp. 1482–1491, 2001.

[21] A. G. Andrade, O. M. Bubu, A. W. Varga, and R. S. Osorio, "The relationship between obstructive sleep apnea and Alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 64, no. s1, pp. S255–S270, 2018.

[22] D. Foley *et al.*, "Daytime sleepiness is associated with 3-year incident dementia and cognitive decline in older Japanese-American men," *Journal of the American Geriatrics Society*, vol. 49, no. 12, pp. 1628–1632, 2001.

[23] E. O. Bixler, A. N. Vgontzas, H.-M. Lin, S. L. Calhoun, A. Vela-Bueno, and A. Kales, "Excessive daytime sleepiness in a general population sample: the role of sleep apnea, age, obesity, diabetes, and depression," *The Journal of Clinical Endocrinology & Metabolism*, vol. 90, no. 8, pp. 4510–4515, 2005.

[24] R. Sadeghi *et al.*, "Sleep Propensity And Sleep Apnea -specific Hypoxia Are Associated With Excessive Daytime Sleepiness," Accepted in *Annals of the American Thoracic Society* (ATS2020).

[25] PubMed, URL: https://www.ncbi.nlm.nih.gov/pubmed/ (accessed Jul. 03, 2018).

[26] B. M. Altevogt and H. R. Colten, *Sleep disorders and sleep deprivation: an unmet public health problem*. National Academies Press, 2006.

[27] A. Rechtschaffen, "A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects," *Brain information service*, 1968.

[28] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. L. Marcus, and B. V. Vaughn, "The AASM manual for the scoring of sleep and associated events," *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 2012.

[29] S. Hitziger, "Modeling the variability of electrical activity in the brain," PhD Thesis, 2015.

[30] J. F. Sassin, D. C. Parker, J. W. Mace, R. W. Gotlin, L. C. Johnson, and L. G. Rossman, "Human growth hormone release: relation to slow-wave sleep and sleep-waking cycles," *Science*, vol. 165, no. 3892, pp. 513–515, 1969.

[31] G. Brandenberger and L. Weibel, “The 24-h growth hormone rhythm in men: sleep and circadian influences questioned,” *Journal of sleep research*, vol. 13, no. 3, pp. 251–255, 2004.
[32] D. Schneider-Helmert, “Why low-dose benzodiazepine-dependent insomniacs can’t escape their sleeping pills,” *Acta Psychiatrica Scandinavica*, vol. 78, no. 6, pp. 706–711, 1988.
[33] J. K. Walsh, “Enhancement of slow wave sleep: implications for insomnia,” *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, vol. 5, no. 2 Suppl, p. S27, 2009.
[34] D.-J. Dijk, J. A. Groeger, N. Stanley, and S. Deacon, “Age-related reduction in daytime sleep propensity and nocturnal slow wave sleep,” *Sleep*, vol. 33, no. 2, pp. 211–223, 2010.
[35] A. S. R. W. Picard, “Understanding Ambulatory and Wearable Data for Health and Wellness.”
[36] J. A. Onton, S. C. Matthews, D. Y. Kang, and T. P. Coleman, “In-Home Sleep Recordings in Military Veterans With Posttraumatic Stress Disorder Reveal Less REM and Deep Sleep< 1 Hz,” *Frontiers in human neuroscience*, vol. 12, 2018.
[37] A. Sano, R. W. Picard, and R. Stickgold, “Quantitative analysis of wrist electrodermal activity during sleep,” *International Journal of Psychophysiology*, vol. 94, no. 3, pp. 382–389, 2014.
[38] P. A. Lanfranchi and V. K. Somers, “Cardiovascular physiology: autonomic control in health and in sleep disorders,” in *Principles and Practice of Sleep Medicine: Fifth Edition*, Elsevier Inc., 2010.
[39] M. G. Miglis, “Autonomic dysfunction in primary sleep disorders,” *Sleep medicine*, vol. 19, pp. 40–49, 2016.
[40] J. Wilde-Frenz and H. Schulz, “Rate and distribution of body movements during sleep in humans,” *Perceptual and motor skills*, vol. 56, no. 1, pp. 275–283, 1983.
[41] F. Giganti, G. Ficca, S. Gori, and P. Salzarulo, “Body movements during night sleep and their relationship with sleep stages are further modified in very old subjects,” *Brain research bulletin*, vol. 75, no. 1, pp. 66–69, 2008.
[42] J. P. Libert, “Thermal regulation during sleep,” *Revue neurologique*, vol. 159, no. 11 Suppl, pp. 6S30–4, 2003.
[43] R. J. Raymann, D. F. Swaab, and E. J. Van Someren, “Skin deep: enhanced sleep depth by cutaneous temperature manipulation,” *Brain*, vol. 131, no. 2, pp. 500–513, 2008.
[44] M. Berthon-Jones, “Feasibility of a self-setting CPAP machine,” *Sleep*, vol. 16, no. 8 Suppl, pp. S120–1, 1993.
[45] S. De Cruz, M. R. Littner, and M. R. Zeidler, “Home Sleep Testing for the Diagnosis of Obstructive Sleep Apnea—Indications and Limitations,” in *Seminars in respiratory and critical care medicine*, 2014, vol. 35, pp. 552–559.
[46] F. Mendonça, S. S. Mostafa, F. Morgado-Dias, A. G. Ravelo-García, and T. Penzel, “A Review of Approaches for Sleep Quality Analysis,” *IEEE Access*, vol. 7, pp. 24527–24546, 2019.
[47] A. A. Ong and M. B. Gillespie, “Overview of smartphone applications for sleep analysis,” *World journal of otorhinolaryngology-head and neck surgery*, vol. 2, no. 1, pp. 45–49, 2016.
[48] Y. K. Choi *et al.*, “Smartphone applications to support sleep self-management: review and evaluation,” *Journal of Clinical Sleep Medicine*, vol. 14, no. 10, pp. 1783–1790, 2018.
[49] J. Lee, M. Hong, and S. Ryu, “Sleep monitoring system using kinect sensor,” *International Journal of Distributed Sensor Networks*, vol. 11, no. 10, p. 875371, 2015.
[50] W. Chen, X. Zhu, T. Nemoto, Y. Kanemitsu, K. Kitamura, and K. Yamakoshi, “Unconstrained detection of respiration rhythm and pulse rate with one under-pillow sensor during sleep,” *Medical and Biological Engineering and Computing*, vol. 43, no. 2, pp. 306–312, 2005.
[51] D. J. Buysse, C. F. Reynolds III, T. H. Monk, S. R. Berman, and D. 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.
[52] F. Chung, H. R. Abdullah, and P. Liao, “STOP-Bang questionnaire: a practical approach to screen for obstructive sleep apnea,” *Chest*, vol. 149, no. 3, pp. 631–638, 2016.
[53] J. M. Peake, G. Kerr, and J. P. Sullivan, “A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations,” *Frontiers in physiology*, vol. 9, p. 743, 2018.
[54] E4 sensors, URL: https://www.empatica.com/research/e4 (accessed Dec. 07, 2017).
[55] J. Hughes, T. Banerjee, W. Romine, G. Goodman, A. Edwards, and L. Lawhorne, “Interdisciplinary Development of a Daily Use Caregiver Sleep Survey (DUCSS): A Mixed Methods Design,” *SocArXiv*, Feb. 2019, doi: 10.31235/osf.io/6v2sr.
[56] C. L. McKibbin *et al.*, “Sleep in spousal caregivers of people with Alzheimer’s disease,” *Sleep*, vol. 28, no. 10, pp. 1245–1250, 2005.
[57] D. Herzig, P. Eser, X. Omlin, R. Riener, M. Wilhelm, and P. Achermann, “Reproducibility of heart rate variability is parameter and sleep stage dependent.,” *Frontiers in Physiology*, vol. 8, p. 1100, 2017.
[58] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, “Automatic sleep/wake identification from wrist activity,” *Sleep*, vol. 15, no. 5, pp. 461–469, 1992.
[59] J. E. Sinex, “Pulse oximetry: principles and limitations,” *The American journal of emergency medicine*, vol. 17, no. 1, pp. 59–66, 1999.
[60] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.
[61] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” *Machine learning*, vol. 46, no. 1–3, pp. 389–422, 2002.
[62] M. Kuhn, “Variable selection using the caret package,” URL:http://cran. cermin. lipi. go. id/web/packages/caret/vignettes/caretSelection. pdf.

[63] R. Rifkin and A. Klautau, "In defense of one-vs-all classification," *Journal of machine learning research*, vol. 5, no. Jan, pp. 101–141, 2004.
[64] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," in *Advances in neural information processing systems*, 2013, pp. 431–439.
[65] M. A. Rowe, C. S. McCrae, J. M. Campbell, A. P. Benito, and J. Cheng, "Sleep pattern differences between older adult dementia caregivers and older adult noncaregivers using objective and subjective measures," *Journal of Clinical Sleep Medicine*, vol. 4, no. 04, pp. 362–369, 2008.
[66] S. M. McCurry, R. G. Logsdon, L. Teri, and M. V. Vitiello, "Sleep disturbances in caregivers of persons with dementia: contributing factors and treatment implications," *Sleep medicine reviews*, vol. 11, no. 2, pp. 143–153, 2007.
[67] J. B. Blank *et al.*, "Overview of recruitment for the osteoporotic fractures in men study (MrOS)," *Contemporary clinical trials*, vol. 26, no. 5, pp. 557–568, 2005.
[68] E. Orwoll *et al.*, "Design and baseline characteristics of the osteoporotic fractures in men (MrOS) study—a large observational study of the determinants of fracture in older men," *Contemporary clinical trials*, vol. 26, no. 5, pp. 569–585, 2005.
[69] T. Blackwell *et al.*, "Associations between sleep architecture and sleep-disordered breathing and cognition in older community-dwelling men: the Osteoporotic Fractures in Men Sleep Study," *Journal of the American Geriatrics Society*, vol. 59, no. 12, pp. 2217–2225, 2011.
[70] F. Chung *et al.*, "Stop questionnairea tool to screen patients for obstructive sleep apnea," *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 108, no. 5, pp. 812–821, 2008.
[71] M. Nagappa *et al.*, "Validation of the STOP-Bang questionnaire as a screening tool for obstructive sleep apnea among different populations: a systematic review and meta-analysis," *PloS one*, vol. 10, no. 12, p. e0143697, 2015.
[72] D. A. Dean *et al.*, "Scaling up scientific discovery in sleep medicine: the National Sleep Research Resource," *Sleep*, vol. 39, no. 5, pp. 1151–1164, 2016.
[73] G.-Q. Zhang *et al.*, "The National Sleep Research Resource: towards a sleep data commons," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1351–1358, 2018.
[74] D. Raman *et al.*, "Polysomnographic heart rate variability indices and atrial ectopy associated with incident atrial fibrillation risk in older community-dwelling men," *JACC: Clinical Electrophysiology*, vol. 3, no. 5, pp. 451–460, 2017.
[75] A. N. Vest *et al.*, "An open source benchmarked toolbox for cardiovascular waveform and interval analysis," *Physiological measurement*, vol. 39, no. 10, p. 105004, 2018.
[76] S. Chokroverty, *Sleep Disorders Medicine E-Book: Basic Science, Technical Considerations, and Clinical Aspects*. Elsevier Health Sciences, 2009.
[77] "Sleep Apnea | National Heart, Lung, and Blood Institute (NHLBI)." https://www.nhlbi.nih.gov/health-topics/sleep-apnea (accessed Apr. 30, 2019).
[78] X. Zhang *et al.*, "Sleep stage classification based on multi-level feature learning and recurrent neural networks via wearable device," *Computers in biology and medicine*, vol. 103, pp. 71–81, 2018.
[79] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.
[80] J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3642–3649.
[81] M. Kuhn, "Variable importance using the caret package," 2012.
[82] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
[83] C. Zammit, H. Liddicoat, I. Moonsie, and H. Makker, "Obesity and respiratory diseases," *International journal of general medicine*, vol. 3, p. 335, 2010.
[84] E. Vanoli, P. B. Adamson, Ba-Lin, G. D. Pinna, R. Lazzara, and W. C. Orr, "Heart rate variability during specific sleep stages: a comparison of healthy subjects with patients after myocardial infarction," *Circulation*, vol. 91, no. 7, pp. 1918–1922, 1995.
[85] S. F. Quan *et al.*, "The sleep heart health study: design, rationale, and methods," *Sleep*, vol. 20, no. 12, pp. 1077–1085, 1997.
[86] S. Redline *et al.*, "Methods for obtaining and analyzing unattended polysomnography data for a multicenter study," *Sleep*, vol. 21, no. 7, pp. 759–767, 1998.
[87] M. W. Johns, "A new method for measuring daytime sleepiness: the Epworth sleepiness scale," *sleep*, vol. 14, no. 6, pp. 540–545, 1991.
[88] P. H. Gander, N. S. Marshall, R. Harris, and P. Reid, "The Epworth Sleepiness Scale: influence of age, ethnicity, and socioeconomic deprivation. Epworth Sleepiness scores of adults in New Zealand," *Sleep*, vol. 28, no. 2, pp. 249–254, 2005.
[89] C. W. Whitney, P. L. Enright, A. B. Newman, W. Bonekat, D. Foley, and S. F. Quan, "Correlates of daytime sleepiness in 4578 elderly persons: the Cardiovascular Health Study," *Sleep*, vol. 21, no. 1, pp. 27–36, 1998.
[90] M. Younes *et al.*, "Odds ratio product of sleep EEG as a continuous measure of sleep state," *Sleep*, vol. 38, no. 4, pp. 641–654, 2015.

[91] M. Dres *et al.*, "Sleep and Pathological Wakefulness at the Time of Liberation from Mechanical Ventilation (SLEEWE). A Prospective Multicenter Physiological Study," *American journal of respiratory and critical care medicine*, vol. 199, no. 9, pp. 1106–1115, 2019.
[92] S. Qanash, E. Giannouli, and M. Younes, "Assessment of intervention-related changes in non-rapid-eye-movement sleep depth: importance of sleep depth changes within stage 2," *Sleep medicine*, vol. 40, pp. 84–93, 2017.
[93] S. M. Purcell *et al.*, "Characterizing sleep spindles in 11,630 individuals from the National Sleep Research Resource," *Nature communications*, vol. 8, p. 15930, 2017.
[94] M. Younes, D. Mazzotti, and S. Redline, "RANGE AND REPRODUCIBILITY OF EEG BIOMARKERS FOR COGNITION ACROSS FIVE YEARS," in *SLEEP*, 2019, vol. 42.
[95] M. Younes, "The case for using digital EEG analysis in clinical sleep medicine," *Sleep Science and Practice*, vol. 1, no. 1, p. 2, 2017.
[96] M. E. Glickman, S. R. Rao, and M. R. Schultz, "False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies," *Journal of clinical epidemiology*, vol. 67, no. 8, pp. 850–857, 2014.
[97] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.
[98] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *The annals of statistics*, vol. 29, no. 4, pp. 1165–1188, 2001.
[99] H. Wang *et al.*, "Genome-wide association analysis of self-reported daytime sleepiness identifies 42 loci that suggest biological subtypes," *Nature communications*, vol. 10, no. 1, pp. 1–12, 2019.
[100] A. Wichniak, A. Wierzbicka, M. Walęcka, and W. Jernajczyk, "Effects of antidepressants on sleep," *Current psychiatry reports*, vol. 19, no. 9, p. 63, 2017.
[101] C. M. Lin, T. M. Davidson, and S. Ancoli-Israel, "Gender differences in obstructive sleep apnea and treatment implications," *Sleep medicine reviews*, vol. 12, no. 6, pp. 481–496, 2008.

Appendices

Appendix A

The list of extracted features from wearable data

Number	Abbreviation Name	Feature description	Physiological signal
1	swsLengthHR	The predicted fraction of sleep belonging to the deep sleep portion based on the heart signal	Heart rate variability
2	swsTimeHR	The number of transitions to the deep sleep stage according to the heart signal	
3	epochCapacity	The ratio of the number of the epochs in a sleep to the total possible epochs	Electrodermal activity
4	epochPeak	The mean number of peaks in all epochs	
5	epochPeakCounter	The number of epochs during a sleep	
6	stormPeak	The percentage of peaks which occur in the storms	
7	largestStorm	The number of epochs that construct the largest storm	
8	timesEdaStorm	The number of distinct storms	
9	meanEdaStorm	The average number of epochs comprising each EDA storm	
10	lengthEdaStorm	The number of whole epochs shaping the storms	

Number	Abbreviation Name	Feature description	Physiological signal
11	timesAwoken	The number of times people awake from sleep	Body movement
12	sleepEfficiency	The proportion of sleep time to the time a person tries to sleep	
13	amountAwake	The length of night time subjects are awake	
14	amountAsleep	The length of time subjects are asleep	
15	swsTimeMovement	The predicted fraction of sleep belonging to the deep sleep portion based on body movement	
16	swsLengthMovement	The number of transitions to the deep sleep stage according to the body movement	
17	%decreaseMovement	The percentage of sleep time in which the body movement has a decreasing pattern	
18	swsLengthTemperature	The predicted fraction of sleep belonging to the deep sleep portion based on the skin temperature	Temperature
19	swsTimeTemperature	The number of transitions to the deep sleep stage according to the skin temperature	
20	%decreaseTemperature	The percentage of sleep time which participants experienced temperature reduction	

Appendix B

STOP-Bang features extracted from PSG Morning Survey of MrOS study

Row	Abbreviation	Definition	STOP-Bang
1	POXNASAL	Did you have nasal stuffiness, obstruction, or discharge last night? Y/N	3
2	POXINTER	Did this interfere with your sleep last night? Y/N	3
3	POXCOUGH POXSNORT POXCPAIN POXSBRE POXSTUFF POXHBURN POXLEGK	During the PAST MONTH, how often have you had trouble sleeping because of… Coughing Snorting or gasping Chest pain or discomfort Shortness of breath Nasal stuffiness Heartburn or reflux Leg jerks or kicks	1, 3
4	mhbp	Has a doctor or other health care provider ever told you that you had hypertension of high blood pressure?	4
5	mhbpt	Are you currently being treated for hypertension of high blood pressure by a doctor?	4
6	hwbmi	Body mass index	5
7	hwneck	Average neck circumference	7
8		The average of 26 questions in Functional Outcomes of Sleep survey range (1-4)	2

Appendix C

The list of computed HRV features from the annotated R-peaks

Row	Abbreviation	Definition	Type
1	NNmean	mean distance value of two normal consecutive R-peaks (NN)	Temporal
2	NNmode	mode of NN intervals	
3	NNmedian	median value of NN intervals	
4	NNskew	skweness of NN intervals	
5	NNkurt	kurtosis of NN intervals	
6	NNiqr	interquartile range of NN intervals	
7	SDNN	standard deviation of all NN intervals	
8	RMSSD	The square root of the mean of the sum of the squares of differences between adjacent NN intervals	
9	pnn50	NN>=50ms count divided by the total number of all NN intervals	
10	btsdet	average number of beats detected in 5 min intervals	
11	ulf	power in the ultra-low frequency range (less than 0.003 Hz)	Frequency -based
12	vlf	power in very low frequency range (0.003 <= vlf < 0.04 Hz)	
13	lf	power in low frequency range (0.04Hz <= lf < 0.15 Hz)	
14	hf	power in high frequency range (0.15 <= hf < 0.4 Hz)	
15	lfhf	Ratio LF /HF	
16	ttlpwr	total spectral power	
17	SampEn	Sample entropy	Entropy-based
18	ApEn	approximate entropy	
19	ac	acceleration capacity	Others
20	dc	deceleration capacity	
21	SD1	standard deviation of projection of the Poincare Plot (PP) on the line perpendicular to the line of identity	
22	SD2	standard deviation of the projection of the PP on the line of identity	
23	SD1SD2	Ratio SD1/SD2	

Appendix D

The list of utilized features in excessive daytime sleepiness analyses

Number	Name	Definition
1	Hypoxic Burden (HB)	The total area under the respiratory-event-related desaturation curve (SpO2)
2	HB_{EV}	The average event-related area under desaturation curve
3	C34_Density	The number of spindles/min. Spindles: Bursts of sigma frequency band activity (11–15 Hz) of EEG signal
4	C34_Duration	The average duration of spindles (Seconds).
5	C34_Power	The average power of spindles (μV)
6	C34_Frequency	The average frequency of spindles (Hz)
7	Average_alpha	The intrusion of alpha activity, 8–12 Hz, in NREM sleep can be measured via the percentage of 3-second NREM epochs with alpha power equal to or greater than 30 μV^2
8	ORPWake	The average of continuous sleep depth metric (the odds ratio product (ORP) of sleep EEG) observed in awake sleep stage.
9	ORP_Sleep	The average of sleep depth during sleep
10	ORPTRT	The average of sleep depth during total recording time
11	ORPLRCorr	The correlation between right and left hemispheres sleep depth
12	ai_all	The number of arousals during sleep. Arousals: an abrupt shift in EEG to a higher frequency for at least 3 seconds and proceeded with greater or equal 10 seconds of stable sleep
13	Awakening index	The number of awakenings during the night. Awakening: An arousal last more than 15 seconds
14	Apnea-hypopnea index (AHI)	The per-hour frequency of all apneas (complete obstruction) and hypopneas (partial obstruction) associated with 3% oxygen desaturation during sleep.

15	Nocturnal Hypoxemia (NH/PcTSa90H)	The percentage of the night below 90% oxygen saturation
16	timest1p	The percentage of sleep that belongs to NREM 1
17	timest2p	The percentage of sleep that belongs to NREM 2
18	timest34p	The percentage of sleep that belongs to NREM 3 and 4
19	timeremp	The percentage of sleep that belongs to REM
20	NTCA1Y	Using non-tricyclic antidepressants drugs
21	TCA1Y	Using tricyclic antidepressants drugs
22	bmi_s1	Body mass index
23	HrsWD02	An average sleep duration on working days
24	SystBP	Systolic blood pressures
25	DiasBP	Diastolic blood pressures
26	SingleYes	Being single
27	ParRptDiabYes	Having diabetes
28	CVDYes	Having cardiovascular disease
29	COFFEE15	Consumption of caffeinated beverage four hours before sleep
30	CgPkYr	Number of cigaret packets
31	age_s1	The age of participants
32	gender	Female/Male
33	race	White, Black, or Others