Predicting Subjective Sleep Quality Using Recurrent Neural Networks

Julien Boussard[1], *Mykel J. Kochenderfer*[2], *Jamie M. Zeitzer*[3]

[1]Department of Statistics, [2]Department of Aeronautics and Astronautics, [3]Department of Psychiatry and Behavioral Sciences
Stanford University, Stanford, CA, USA
{jzeitzer, mykel, julienb}@stanford.edu

***Abstract*—Our goal is to predict subjective sleep quality (SSQ) from objective sleep data and identify the causes and markers of the variances within "normal" sleep. Such information would increase our understanding of the causes of variation in SSQ and potentially improve our ability to improve SSQ. Previous approaches rely on human annotation of the electroencephalographic (EEG) brain signals, to deal with the noisy, high dimensional nature of the EEGs. We aim to use recurrent neural networks to directly analyze and extract useful information from EEG brain signals. We analyze population-based overnight sleep polysomnography data obtained from 4885 community-dwelling adults. We use convolutional and recurrent neural networks to process the EEGs and combine them with information related to health and lifestyle to predict subjective depth and restfulness of sleep. We compare the coefficient of determination to the ones obtained with regression methods and technician annotations of the EEGs in previous studies. Predicting SSQ from our data set of community-dwelling adults using RNNs to analyze the whole EEG signals appear to be less accurate than previous approaches predictions. It might be necessary to acquire more data, possibly with new variables that might be better correlated with SSQ. RNNs are, however, able to extract variables correlated with SSQ from EEG signals. Our results provide insights into how RNNs can be used to extract information from brain signals and how methods such as hierarchical clustering analysis can help neural networks predict subjective variables from polysomnography data.**

I. Introduction

Sleep medicine is largely focused on the treatment of major sleep disorders such as insomnia or sleep apnea, but relatively little work has been done on understanding and explaining the changes in usual sleep. Many individuals have poor sleep and, while not rising to the level of "pathological", this poor sleep can still have a negative impact on overall mental and physical well-being, especially over long durations [1]. Objective sleep measures might correlate better with particular health outcomes, but most patients actually care about the subjective nature of sleep – how do they feel about their sleep. If one knew the objective elements of subjective sleep quality (SSQ), one could design and test interventions that could specifically target SSQ and incidentally improve objective aspects of sleep. It is difficult to convince people to modify their behavior based on objective measures and medical outcomes in the distant future. Understanding variables that are markers for, if not causally involved in, the subjective quality of sleep is fundamental to improving sleep. Only a few studies, such as Kaplan et al. ([2], [3]) have been conducted on predicting SSQ. We aim to predict SSQ and identify such markers from the brain EEG signals using recurrent neural networks (RNNs).

Previous research has been mostly unsuccessful at determining the objective correlates of SSQ. Most prior work on predicting sleep quality from objectively recorded sleep variables used small data sets, typically examining healthy individuals or individuals with insomnia ([4], [5], [6], [7], [8], [9], [10], [11]). Kaplan et al. ([2], [3]) have previously reported SSQ on large data sets of community-dwelling adults, where sleep was recorded for a single night and SSQ was assessed on the subsequent morning. The data sets, constituting data recorded from 4885 adults, contain information about the patients health, lifestyle, the recordings of the brain signals, and variables derived from the signals by a technician, such as the proportion of time asleep or the time the patient was awake at the end of the recording. Kaplan et al. analysis rely on these derived variables that allowed them to use supervised machine learning algorithms, random forest and lasso regression, to find correlates of sleep quality.

EEG signals are subject to diverse noise or artifacts contaminating the signal obtained from the scalp during EEG recording ([12], [13]), such as eye movements, muscular contractions and variations in scalp surface caused by sweat and hair. The temporal resolution (EEGs are sampled at 125Hz) combined with the length of each recording (typically around 8 hours of recordings) makes it very hard to take into account the raw signal. RNNs have recently emerged as an effective model in a wide variety of applications that involve sequential data. RNNs have not been widely used in the field of EEG, but previous work ([14], [15]) has demonstrated that neural networks constitute an appropriate network architecture for categorizing EEG data into stages of sleep, and that a deep neural network sleep stager can be almost indistinguishable in performance from a human annotator. To take advantage of the time-dependent structure and the high temporal resolution of the signals, our goal is to use RNNs to detect patterns in the EEG structure that impact the subjective sleep quality, and predict SSQ without using the technician's annotations or variables derived from the signal.

II. Methods

We examined EEG recordings of a single night of sleep obtained in the Sleep Heart Health Study (SHHS; https://sleepdata.org/datasets/shhs) [16]. The SHHS data set consists of 6441 overnight polysomnographic (PSG) recordings obtained from healthy, community-dwelling men and women aged 40 and older [17]. For our analysis, we only kept 4885 out of the 6441 PSG recordings, which had reported SSQ and whose EEG recording was of sufficient quality, and did not use the 1556 others. Among the PSG data, we specifically included filtered EEG data (collected at 125 Hz with a high pass filter at 0.15 Hz) collected from a central (C3/A2) lead. There were 4 electrodes used in SHHS, which are bilaterally placed (2 on each side, but in the same location). Several phenomena of sleep are localized in different brain regions, and more EEG derivations would be interesting, but these were not available as it was not done in SHHS. For specific analyses, we included technician annotations of these EEG signals, which parsed each 30 second epoch into a stage of wake, non-rapid eye movement (NREM) sleep stage 1 (N1), NREM sleep stage 2 (N2), NREM sleep stage 3 (N3), and rapid eye movement sleep (REM). We also included mathematical rearrangement of this staging as total sleep time, wake after sleep onset (WASO), and sleep efficiency, among others. These variables can be directly derived from the staging with simple formulas. For example, sleep efficiency is equal to the total time asleep divided by the time of the recording. Additional variables included in the models were related to demographics (race, ethnicity, gender, education, marital status), body habitus (age, height, hip and waist circumference, body mass index, weight), and habitual amount and type of caffeine and alcohol intake. Our goal was to incorporate the EEG signals using RNNs, instead of using the technician annotations and the derived variables, and to examine whether these data could be used to predict the SSQ obtained on the subsequent morning. SSQ was determined through two questions that examined sleep depth (1 to 5 on a Likert-like scale, anchored by deep and light) and sleep restfulness (1 to 5 on a Likert-like scale, anchored by restful and restless). It is part of the SHHS dataset. These ratings of sleep quality have been used in previous research ([2], [3]). Participants also reported the length of their sleep (how long they estimate the total time they were asleep during the night) and how well they slept compared to their usual sleep night. Following the work of Kaplan et al. ([2]) we focus on predicting subjective depth and restfulness, as the quality of sleep compared to usual is highly dependent on the participants sleep habits, and the information needed to predict it won't necessary be contained in the EEG signals themselves.

Previous approaches use the coefficient of determination R^2 to ascertain the quality of prediction [2]. R^2 represents the proportion of variance explained by the statistical analysis. We use R^2 as our main metric to compare our results to the ones of previous studies and to quantify the importance of derived variables and EEG signals. For example, if we are able to explain 70% of the variance using two variables x and y, and 60% when using only x, we can say that y explains 10% of the variance that are unexplained by x. When fitting a regression model to n observations $y_1, ..., y_n$, with $\hat{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ the mean of the observed data, and $\bar{y_1}, ..., \bar{y_n}$ the predicted values associated with each observation, the coefficient of determination is

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \bar{y_i})^2}{\sum_{i=1}^{m}(y_i - \hat{y})^2} \tag{1}$$

We created train, development, and test sets to tune hyper-parameters on the development set and calculate R^2 on the test set prediction. All the results will be given on the test set.

We tried different preprocessing techniques to analyze EEG and feed the signal into our RNN model. We trained each model with data preprocessed in the way described below. We first performed a Fast Fourier Transform (FFT) [18] before determining the relative power in traditionally-determined spectral bands [19] : delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), sigma (12–15 Hz), and beta (15–30 Hz). All bands are normalized to the overall power in the data set. We finally performed a wavelet transform of the EEG. We found that only performing FFT led to the best performance for the different tasks, on our test set. We thus chose to present, in the next section, the results obtained when training our model with data preprocessed using FFT.

In order to assess the performance of RNNs, we applied several regression models [20] to predict subjective depth and restfulness from participant information and derived PSG variables. We used random forest [21], AdaBoost, gradient boosting regression, and support vector machine. We compared their performance to the performance of fully connected neural networks. The regression models take as input the categorical variables and the variables derived by technicians from the EEG signals. We also trained these models to predict the self-reported length of sleep and quality of sleep compared to usual.

To take the EEG signals into account, we then experimented with different architectures combining convolutional layers, bi-directional long short term memory (LSTM) layers [22], and fully connected dense layers. We tried different losses such as mean-squared error or cross-entropy loss, with relu or sigmoid activation functions to train the neural networks. The architecture that gave the best prediction is represented in Fig. 1. This model was trained without taking the derived variables or technician annotations into account. The results presented in the next section correspond to the best performance on the test set, after comparing models different architectures.

We calculated the coefficient of determination obtained when predicting with EEG and categorical variables related to the participants' health and lifestyle combined, or just EEG alone, to quantify how much information is contained in the EEG themselves.

To guide and help the network detect useful features from the signals throughout the learning, we used transfer learning. Transfer learning is a method that focuses on

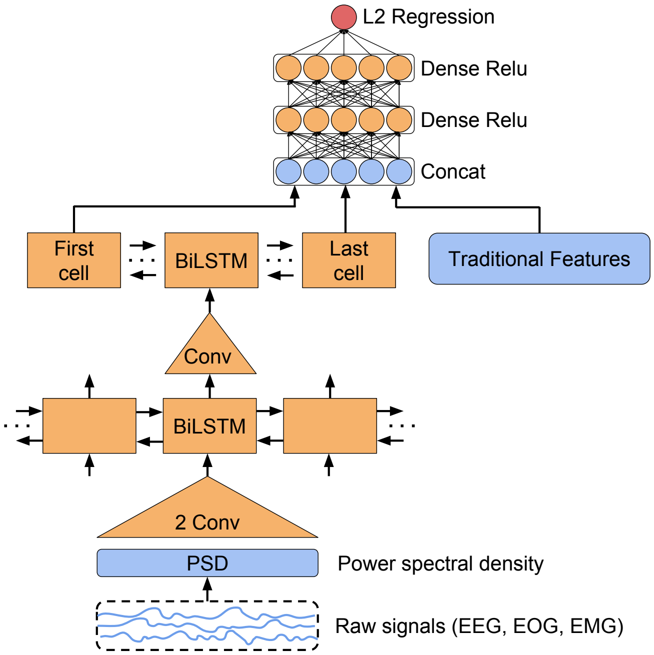

Figure 1. Neural network architecture that gave the best prediction.

storing knowledge gained while solving one problem and applying it to a different but related problem. If a variable x is very correlated with the variable y we want to predict, a model that first learns how to predict x can be trained to learn y very efficiently. Our idea was to ensure that the network learns features of the EEG data that is representative of the SSQ. We pre-trained the network to learn variables correlated with SSQ according to Kaplan et al. [2], before training it to learn to predict SSQ. It also allowed us to examine how well the network can predict simple variables from EEG signals and gain insight into how RNNs understand EEG structure.

Following the work of Kaplan et al. [2], we decided to start by training the network to learn sleep efficiency (amount of sleep time divided by total time in bed) and wake after sleep onset (WASO, amount of wake sleep that occurs interspersed with sleep throughout the night), the two variables that are most correlated with SSQ. We then trained the neural network to predict the percent of sleep time in REM sleep (tmremp), and the number of shifts between N3 and N1 or N2 per hour of sleep (hstg342p). We chose these variables given their direct link to EEG structure and hypothesized link to SSQ. Given the results of these predictions, to perform transfer learning, we chose to pre-train the convolutional and recurrent layers of the neural network predicting fraction of sleep time in REM sleep before predicting SSQ. Technician annotation of sleep stages was not used in this analysis.

After experimenting with transfer learning, we wanted to account for specific individual differences between participants. We used hierarchical cluster analysis (HCA) [23] to define groups of participants who will rate their sleep quality the same way (e.g., a group that will be very sensitive to the total sleep time). HCA works as follows. Once a distance between participants is defined, it first finds the two most similar participants (P_1, P_2) to create the first cluster C_1, and then recursively adds the participant that is the closest to $C_n = (P_1, P_2, \ldots, P_n)$ to create cluster C_{n+1}. We chose to use the complete linkage method to define the distance between the merged cluster and the other participants as the maximum of the pair of dissimilarities in each case. For the distance between participants, we used the Euclidean distance between vectors composed of a set of meaningful variables representing the participant's health and lifestyle. We tried different sets of variables (including more or less information) and will present the results obtained with the best combination of variables we found.

After performing HCA with different variable sets and defining subgroups of similar participants, we applied Lasso regression and Random Forests to the different clusters to find the subgroups of participants who have specific objective correlates that are more predictive and who will be mostly responsive to the same variable found in the data set. Once such groups are defined, it should be easier for our model to predict SSQ of participants within each subgroup since SSQ would be more correlated with specific variables. Thus, we trained, fit the parameters, and tested our regression and neural network models on each clusters, with and without taking into account the technicians annotations and derived variables.

III. Results

A. Predicting Subjective Sleep Quality

The results of the different regression models and the multi-layer perceptron (a fully connected dense neural network) can be seen in Fig. 2. The performance on predicting sleep depth is significantly higher than in prior work on the same data set [2]. Kaplan et al. only used a training and a test set, so we compare their results with our results on the development set. For sleep depth our methods achieve $R^2 \in [0.13 : 0.22]$ over $R^2 \in [0.07 : 0.09]$. For sleep restfulness we achieve $R^2 \in [0.13 : 0, 21]$ over $R^2 \in [0.09 : 0.13]$. The best performance on our test set, $R^2 = 0.13$ when predicting depth and $R^2 = 0.11$ when predicting restfulness, was obtained using gradient boosting regression.

Regression methods achieve good results on predicting self-reported sleep length, with a coefficient of determination close to 0.3 on the test set. The technicians annotations contain sleep efficiency (percent of time asleep) and total time of recording, that easily gives the total time asleep during the night, which is highly correlated with the self-reported time asleep. The prediction of quality of sleep compared to usual is, for most models, better than subjective depth or restfulness. We believe that it is due to the high correlation with sleep habits variables contained in the data set.

With recurrent and convolutional neural networks, we achieved at best $R^2 = 0.04$ on the test set. This is

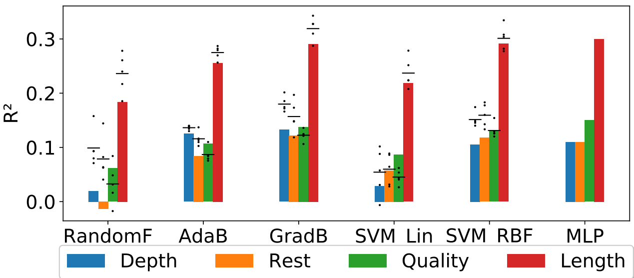

Figure 2. R^2 scores for different models when predicting the sleep quality from derived variables and patient information. The bars indicate the performance on the test set. The black dots are the scores from a 5-fold cross validation on the training set, the black bar is the average score. The blue bar represents the prediction of sleep depth, the orange bar the prediction of sleep restfulness, the green bar the prediction of quality of sleep compared to usual and the red bar the prediction of self-reported sleep length.

Table I
R^2 RESULTS.

	EEGs alone	EEGs and categorical variables
Sleep Efficiency	0.72	0.7
WASO	0.73	0.72
tmremp	0.60	0.40
hstg342p	0.29	0.21
Depth	0.007	0.04
Restfulness	0.01	0.05

surprising because the traditional feature branch generally proved quite robust to changes, but it can be explained by the fact that the fully connected dense layers have more features in the input layer to overfit on.

B. Predicting Intermediate Variables for Transfer Learning

We then performed transfer learning by training the network to predict intermediate variables such as WASO, sleep efficiency, the time spent in REM sleep and the number of shifts between sleep stages 2, 3 and 4. The results obtained for different steps of transfer learning are shown in Table I.

We trained the network to learn these variables from the EEG signals alone, and from the EEG signals combined with categorical variables. For the variables directly related to the EEG structure, we obtain better results when predicting form the EEGs alone. Categorical variables don't contain any information about the EEG structure and make the prediction task more difficult.

The network is able to learn relatively well to predict sleep efficiency and WASO. It learns how to distinguish awake from non-wake stages. It predicts WASO a little bit better, as the neural network has to understand when the participant is awake only at the end of the recording.

The results for percent of REM sleep are also promising. Being able to explain 60% of the variance when the inputs are only EEG signals shows that the neural network can learn from the EEG to predict the time spent in REM sleep. As we believed that part of the information needed to predict SSQ is the transitions between REM and non-REM sleep, we decided to perform transfer learning by pre-training the convolutional and recurrent layers of the neural network (layers that only take the signal as input) to learn how to predict fraction of REM sleep, before predicting SSQ.

Predicting shifts between N3 and N1 or N2 per hour was, however, more difficult. The poorer results show that it is still difficult for the neural network to learn sleep structure and its complex variables. For both percent of REM sleep and shifts between N3 and N1 or N2, the prediction is better when given EEG as only inputs. This is due to the derived categorical variables not being correlated with the fraction of REM sleep or shifts between N3 and N1 or N2, and the neural network tends to over-fit them. It makes the prediction worse on the development set.

When predicting depth and restfulness, we did not get better results than with regression models, with our best R^2 coefficient being 0.05 for restfulness and 0.04 for depth, when given as inputs EEG combined with categorical variables. We also tried predicting subjective depth and restfulness of sleep from the EEGs only, resulting in a very low R^2 coefficient (less than 0.01). The prediction of depth and restfulness is significantly better when given categorical variables. Contrary to sleep efficiency and WASO, these variables are subjective and cannot be directly inferred from the EEG structure itself. Information about the participant is needed in order to capture the specificity of this rating.

C. Hierarchical Cluster Analysis

We then performed the HCA to account for individual characteristics of the participants. Random Forests and lasso regression algorithms revealed several clusters of participants who seem to be responsive to specific variables. We chose to analyze two of these groups, G_1, mostly constituted by participants having high caffeine intake (determined either by the self-reported number of cups of coffee the participant had during the day or before going to bed) who appear to be more responsive to the total sleep time, and G_2, mostly constituted by participants having high body mass index. We then performed prediction on these subgroups and their complements G_1^C and G_2^C.

On each of these subgroups, Random Forest best predicts the subjective sleep restfulness. Table II presents the results of Random Forest and the RNNs model, which takes as input the EEG signals and SHHS variables.

The prediction of all models on the clusters G_1 and G_2 are much better than the prediction on the whole data set. This can be explained by the sensitivity of participants in each group to similar hidden variables that the models are capable to predict. However, the random forests model taking as input the categorical and derived variables still explain more variance than the recurrent neural network model taking as input the EEG signal and categorical variables. The information needed to predict SSQ might be contained in the derived variables, and extracting this

Table II
R^2 RESULTS AT PREDICTING SLEEP RESTFULNESS AFTER HCA.

Cluster	RNNs	Random Forest Model
G_1 High Caffeine Intake	0.16	0.33
G_1^C Low Caffeine Intake	0.07	0.18
G_2 High BMI	0.13	0.28
G_2^C Low BMI	0.04	0.17

information from the signal itself before predicting SSQ is a more complicated task. We can see that the RNNs model is able to extract useful information from the EEG recordings and understand what each subgroup of patients is sensitive to. It obtains a R^2 score of 0.16 on cluster G_1, which is significantly higher than the $R^2 = 0.05$ obtained on the whole dataset.

The result on the complement group of G_1 is still better than the prediction on the total data set. The cluster G_1 is comprised of 2244 participants over 4885. By being able to calculate the distance of an individual to G_1 and G_1^C from categorical variables, we can use the trained Random Forest model on the corresponding group for all individuals, leading to a total coefficient of determination $R^2 = \frac{0.33 \cdot 2244 + 0.18 \cdot 2641}{4885} = 0.25$. This coefficient of determination is much higher than the one obtained on the test set by our best model. For the neural network model, the total coefficient of variation is $R^2 = \frac{0.16 \cdot 2244 + 0.7 \cdot 2641}{4885} = 0.11$. We expect that we would get much better results at predicting sleep depth after performing the same clustering analysis.

IV. Discussion

The different results allow us to consider that RNNs are able to capture some information and understand relatively simple variables such as sleep efficiency, WASO, or proportion of REM sleep from EEG signals alone. However, RNNs have difficulty extracting more complex variables such as the number of shifts between N3 and N1 or N2 per hour.

The difference in performance between the RNNs model and the regression model may not be due to the ability of the RNNs to extract useful information from the data but from the difficulty of the task itself. We were requiring the RNNs model to predict SSQ from EEG signals and information about the participants. It then needed to extract information from very long, noisy recordings at the same time as understanding how the participants rate their night of sleep and what hidden variables are correlated with SSQ. It makes sense that the performance is lower than when using categorical variables and the derived variables extracted from the signal by a technician, that, if they contain less information, are much easier to understand and analyze.

The performance of RNNs for predicting SSQ is very low despite the fact that it is able to predict sleep efficiency and WASO, the two variables most correlated with SSQ. Furthermore, its performance significantly improves when predicting on specific subgroups of participants (0.11 compared to 0.05). RNNs are not able to identify and extract from EEG signals the information explaining the variation in SSQ between individuals. The difference in results for predicting sleep restfulness after performing the HCA shows that, in order to predict SSQ, understanding the participants' subjectivity and way of rating their sleep may be crucial. In this objective, it would be very useful to have recordings of several nights per person. Being able to sub-categorize the participants might change the way doctors treat their patients, as they could find the precise element that causes sleep disorders.

Participants rate their SSQ immediately after awakening. It is therefore possible that SSQ could be more dependent on the events at the end of the night. Future work could take into account the time course of the signal and determine the best temporal segment to consider to get the most accurate prediction.

Based on our current results, RNNs were not significantly better at determining SSQ from EEG than more traditional machine learning techniques. However, because they can learn information from the EEG signals, RNNs might still be very useful in future research trying to predict SSQ or understand what other variables individuals react to, using new data involving several recordings per person or more extensive environmental data that could influence SSQ.

V. Conclusion

Using RNNs to predict SSQ from the EEG signal is not as accurate as previous approaches that predict SSQ using annotated data and variables derived from the signal by a technician. We showed that RNNs are able to accurately learn sleep variables from the signal such as sleep efficiency or sleep time spent in REM sleep. These variables are correlated with SSQ but learning to extract such variables from EEGs in order to predict SSQ is a very complicated task. We showed that defining subgroups of adults whose SSQ depend on particular variables with Hierarchical Analysis leads to much better results, for both our model and previous approaches. It suggests that finding better ways to categorize individuals, understand what variables each patient is sensitive to and how they rate their own sleep quality, may be a promising direction for future work.

Acknowledgement

This work is partially supported by the National Science Foundation grants DRK-12 DRL: 1417769, ITEST DRL: 1614085, and RET Site EEC: 1542286, NY Space Grant Consortium grant 76156-10488; and the Stanford University Precision Health and Integrated Diagnostics Center.

References

[1] B. DJ., "Sleep health: Can we define it? does it matter?" *Sleep*, vol. 37(1), pp. 9–17, 2014.
[2] K. Kaplan, P. Hardas, S. Redline, and J. Zeitzer, "Correlates of sleep quality in midlife and beyond: a machine learning analysis," *Sleep Medicine*, vol. 34, p. 162167, 2017.

[3] K. Kaplan, J. Hirshman, B. Hernandez, M. Stefanick, A. Hoffman, S. Redline, S. Ancoli-Israel, K. Stone, L. Friedman, and J. Zeitzer, "When a gold standard isn't so golden: Lack of prediction of subjective sleep quality from sleep polysomnography," *Biological Psychology*, vol. 123, p. 3746, 2017.

[4] D. O'Donnell, E. Silva, M. Munch, J. Ronda, W. Wang, and J. Duffy, "Comparison of subjective and objective assessments of sleep in healthy older subjects without sleep complaints." *Sleep*, vol. 18(2), p. 254263, 2009.

[5] F. Baekeland and P. Hoy, "Reported vs recorded sleep characteristics." *Arch Gen Psychiatry*, vol. 24(6), p. 548551, 1971.

[6] A. Westerlund, Y. Lagerros, G. Kecklund, J. Axelsson, and T. Akerstedt, "Relationships between questionnaire ratings of sleep quality and polysomnography in healthy adults." *Behav Sleep Med.*, p. 115, 2014.

[7] C. Hoch, C. Reynolds, D. Kupfer, S. Berman, P. Houck, and J. Stack, "Empirical note: Self-report versus recorded sleep in healthy seniors." *Psychophysiology*, vol. 24(3), p. 293299, 1987.

[8] B. Riedel and K. Lichstein, "Objective sleep measures and subjective sleep satisfaction: How do older adults with insomnia sefine a good night's sleep?" *Psychol Aging.*, vol. 13(1), p. 159163, 1998.

[9] M. Bonnet and L. Johnson, "Relationship of arousal threshold to sleep stage distribution and subjective estimates of depth and quality of sleep." *Sleep*, vol. 1(2), p. 161168, 1978.

[10] T. Akerstedt, K. Hume, D. Minors, and J. Waterhouse, "The meaning of good sleep: a longitudinal study of polysomnography and subjective sleep quality." *J Sleep Res.*, vol. 3(3), p. 152158, 1994.

[11] A. Laffan, B. Caffo, B. Swihart, and N. Punjabi, "Utility of sleep stage transitions in assessing sleep continuity." *Sleep*, vol. 33(12), p. 16811686, 2010.

[12] J. Li, C. Li, N. Thakor, A. Cichocki, and A. Bezerianos, "A new perspective of noise removal from eeg." *8th International IEEE EMBS Conference on Neural Engineering*, pp. 501–504, 2017.

[13] C. Q. Lai, H. Ibrahim, M. Z. Abdullah, J. M. Abdullah, S. A. Suandi, and A. Azman, "Artifacts and noise removal for electroencephalogram (eeg): A literature review," *IEEE Symposium on Computer Applications and Industrial Electronics*, pp. 326–332, 2018.

[14] G.-M. G. Bresch E, Groekathfer U, "Recurrent deep neural networks for real-time sleep stage classification from single channel eeg." *Front Comput Neurosci.*, vol. 12:85, 2018. [Online]. Available: https://doi.org/10.3389/fncom.2018.00085

[15] J. B. Stephansen, A. N. Olesen, M. Olsen, A. Ambati, E. B. Leary, H. E. Moore, O. Carrillo, L. Lin, F. Han, H. Yan, Y. L. Sun, Y. Dauvilliers, S. Scholz, L. Barateau, B. Hogl, A. Stefani, S. C. Hong, T. W. Kim, F. Pizza, G. Plazzi, S. Vandi, E. Antelmi, D. Perrin, S. T. Kuna, P. K. Schweitzer, C. Kushida, P. E. Peppard, H. B. D. Sorensen, P. Jennum, and E. Mignot, "Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy," *Nature Communications*, vol. 9, p. 5229, 2018. [Online]. Available: https://doi.org/10.1038/s41467-018-07229-3

[16] D. A. Dean, A. Goldberger, R. Mueller, M. Kim, M. Rueschman, D. Mobley, S. S. Sahoo, C. P. Jayapandian, L. Cui, M. G. Morrical, S. Surovec, G. Zhang, and S. Redline, "Scaling up scientific discovery in sleep medicine: the national sleep research resource." *Sleep*, vol. 39(5), p. 11511164, 2016.

[17] S. F. Quan, B. V. Howard, C. Iber, J. P. Kiley, J. Nieto, G. T. O'Connor, D. M. Rapoport, S. Redline, J. Robbins, J. M. Samet, and P. W. Wahl, "The sleep heart health study: Design, rationale, and methods." *Sleep*, vol. 20, p. 10771085, 1997.

[18] E. Brigham and R. Morrow, "The fast fourier transform," *IEEE Spectrum*, vol. 4(12), pp. 63 – 70, 1967.

[19] I. G. Campbell, "Eeg recording and analysis for sleep research." *Current Protocols in Neuroscience*, 2009.

[20] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[21] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[22] S. Hochreiter, , and J. Schmidhuber, "Long short-term memory." *Neural Computation*, vol. 9(8), pp. 1735–1780, 1997.

[23] S. Patel, S. Sihmar, and A. Jatain, "A study of hierarchical clustering algorithms." in *International Conference on Computing for Sustainable Global Development*, 2015, pp. 537–541.