A Novel Method for Sleep Score Estimation Using Wearable Sensors with a Deep Sequential Neural Network

2021 IEEE International Conference on Electro Information Technology (EIT) | 978-1-6654-1846-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/EIT51626.2021.9491896

Md Juber Rahman
Department of Electrical and Computer Engineering
University of Memphis
Memphis, USA
e-mail: mrahman8@memphis.edu

Bashir I. Morshed
Department of Computer Science
Texas Tech University
Lubbock, USA
e-mail: bmorshed@ttu.edu

Abstract—**The use of sleep score as a measure of fitness and wellness is getting popular in Smart Health as it provides an objective assessment of sleep quality. However, reliable estimation of sleep scores from wearable sensor data only is challenging. In this study, we investigated the estimation of sleep score using only features available from single-channel ECG or single-channel EEG data. We used partial correlation and conditional permutation importance for feature selection; then compared extreme gradient boosting, artificial neural network, and sequential neural network for developing a regression model for sleep score estimation. TabNet- an attention-based deep sequential learning model achieved the best performance of RMSE = 5.47 and R-squared value of 0.59 in the test set for sleep score estimation using only spectral features of single-channel EEG. The results pave the way for reliable and interpretable sleep score estimation using a wearable device.**

Keywords- ***Attention Model, Electroencephalography, Regression, Sleep Score, Smart Health, TabNet, Wearable Sensor***

I. INTRODUCTION

Sleep is an important biological process and plays a key role in restoring energy, solidifying and consolidating memories, and repairing body cells. It also helps in metabolism and cardiovascular function [1]. The regulation of sleep is controlled by the circadian biological clock and sleep/wake homeostasis. Good quality sleep is essential for good health and improved quality of life. Poor sleep is linked to depression, obesity, daytime drowsiness, less productivity, and a greater risk of coronary artery disease and stroke [2-4].

Subjective assessment of sleep quality using standard questionnaires is well investigated and is widely used in clinical practice. Some of the well-accepted and popular methods for subjective sleep quality assessment are- Pittsburgh Sleep Quality Index (PSQI), Epworth Sleepiness Scale (ESS), and Functional Outcome of Sleep Questionnaire (FOSQ). PSQI uses a 7 component questionnaire and the subject assigns a score of 0-3 for each component [5]. The components are - subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction. A global score >5 indicates poor sleep quality. FOSQ has 21 questions related to activity levels, vigilance, intimacy and relationships, general productivity, and social outcomes [6]. The potential range of scores for each subscale is 1 – 4 with higher scores indicating greater insomnia severity. Similarly, in ESS the subject assigns a score of 0-3 for 8 questions aimed at assessing daytime sleepiness. A total score of 16-24 indicates excessive day time sleepiness suggesting the need for medical attention [7].

Subjective methods suffer from high bias, require active user participation, and a longer period (2 weeks - 1 month). To overcome these limitations objective sleep assessment methods have been developed. The gold standard for objective sleep assessment is based on polysomnogram- a complex test usually performed in a sleep lab and capture 14 different physiological signals during sleep. Polysomnogram is expensive, not user friendly, and not readily available everywhere. Hence, there is a growing need for reliable sleep assessment using wearables. Kuo *et al*. developed an actigraphy based wearable device for sleep quality assessment [8]. Mendonca *et al*. proposed a method for sleep quality estimation using electrocardiogram by cardiopulmonary coupling analysis [9]. Azimi *et al*. reported an objective IoT-based longitudinal study for sleep quality assessment [10].

Recently the concept of sleep score has been introduced. A reliable estimation of sleep score is achieved by combining sleep efficiency, sleep time in deep sleep stages, frequency of arousals, oxygen saturation level, resting heart rate during sleep, etc. Sleep score provides valuable information regarding the fitness and wellness of a person and may play a crucial role in Smart Health as a key health indicator. Although, some commercial initiatives e.g. Fitbit Charge smart band (Fitbit Inc., USA), Apple Watch (Apple Inc., USA), Oura sleep ring (Oura Health Ltd., Finland) as well as research studies have attempted the estimation of sleep score from non-polysomnographic measures, development of a well-accepted standard methodology is yet to achieve and needs further investigation [11]. In this work, we investigated a method for reliable estimation of sleep score from non-polysomnographic measures using an attention-based deep sequential neural network. Priority has been given to ECG and EEG based features so that sleep scores can be estimated using user-friendly wearable devices. The developed sleep score has been validated against ground truth established from polysomnography measurements. The results pave the way for reliable sleep score assessment using single-channel EEG.

II. Materials and Methods

A. Dataset

Sleep Health Heart Study (SHHS) is a dataset available from the National Sleep Research Resource [12]. SHHS was implemented as a multi-center cohort study in two phases by the US National Heart Lung & Blood Institute. Unattended home polysomnograms were obtained for both the phases of SHHS by certified and trained technicians. The polysomnogram data was saved in European Data Format (EDF). Data processing and initial scoring were done by Compumedics software (Compumedics Ltd., Australia). Two manual scorings were included to annotate the database with sleep duration, sleep efficiency, arousal index, sleep stages, oxygen saturation level, etc. A dataset of 500 subjects containing good quality data for both ECG and EEG is available from the dataset provider and is recommended for use in a research study. In our study, for developing the regression models we used this dataset of 500 subjects. The distribution of records in the dataset is as follows: male- 231, female- 269. The age of the subjects ranges from 44 to 89 years with a mean of 65 years and a standard deviation of 10.41 years. The body mass index (BMI) of the subjects ranges from 18 – 46 with a mean of 27.51 kilograms per square meter and a standard deviation of 4.11 kilograms per square meter.

B. Computation of Baseline Sleep Score

Guidelines for computing a composite sleep health score from polysomnographic measures have been developed and reported in previous research studies [13-14]. In this study, we used a generalized mathematical model for computing the baseline sleep score. The model has been described by equation (1).

$$Sleep\ Score = \frac{1}{m+n}\left\{\sum_{i=1}^{m} X_{pos(i)} + \sum_{j=1}^{n}(1 - X_{neg(j)})\right\} \quad (1)$$

where X_{pos} are the sleep attributes that impact sleep score positively, i.e. higher is better, X_{neg} are the sleep attributes that impact sleep score negatively, i.e. lower is better. m is the total number of positive attributes and n is the total number of negative attributes.

The positive attributes available from SHHS dataset are as follows:

Sleep time- Duration of entire sleep.

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

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

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

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

The negative attributes available from SHHS is as follows:

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

In computing the sleep score, all the attributes have been normalized on a scale of 0-1. To achieve a consistent "higher is better" rule the value of each negative attribute is subtracted from 1. Then, the attribute values have been summed up to develop a composite score. The composite score has been multiplied by 100 and divided by the total no. of positive and negative attributes to obtain the sleep score in the range of 0-100.

C. Feature Extraction from Wearable Sensor Data

The recording montage for polysomnogram consisted of data from 14 channels which include- ECG, EEG, electrooculogram (EOG), Electromyogram (EMG), nasal airflow, thoracic and abdominal movement signal, SpO2, sleep hypnogram, etc. Hardware filters have been used for preliminary noise reduction. The cutoff frequency for hardware filters had been as follows- ECG-0.15 Hz, EOG-0.15 Hz, EMG-0.15 Hz, EEG-0.15 Hz, Thoracic respiration signal-0.05 Hz, Abdominal respiration signal-0.05 Hz. The sampling rate is 125 Hz for EEG, ECG, and EMG signals. For EOG, the sampling rate is 50 Hz. In investigating a minimalistic approach, we considered the use of features from ECG, EEG, SpO2 signals considering the sensors are more user-friendly and widely used. For RR interval correction we used maliks rule followed by a cubic interpolation for the determination of Normal-to-Normal (NN) intervals [15].

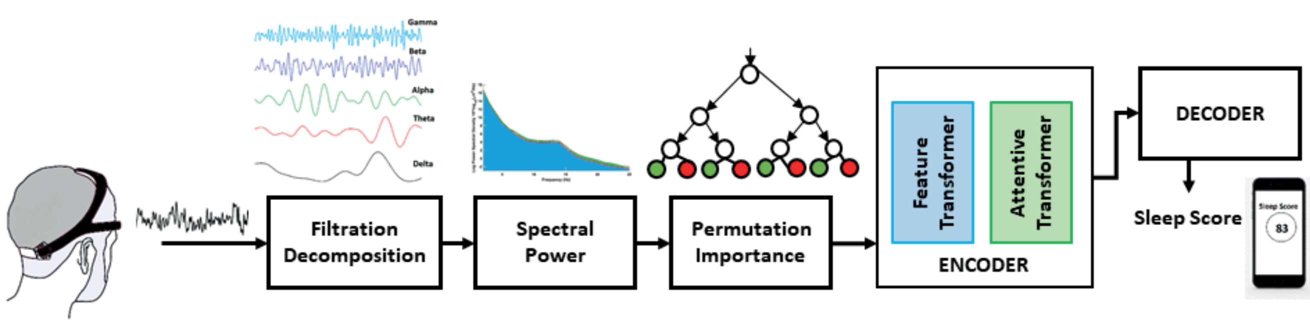

Fig. 1 Method of feature extraction, feature selection and regression for sleep score estimation

From the NN interval series time domain and frequency domain features have been extracted HRV guidelines using HRV Toolkit available from Physionet [16]. For the power spectrum estimation, we used Lomb's periodogram method. The entire ECG record has been divided into 5-minute epochs to estimate short-term components of HRV. In total 20 HRV features have been extracted. From EEG we computed spectral features as shown in Fig. 1. The EEG signal was collected using two channels from the central region of the brain. One of the channels was C4-A1 and the other one was C3-A2. The power spectral densities for these two channels are very similar. In our study, we have only used the signal from the C4 channel as it has been designated as the primary EEG channel in SHHS. EEG spectral analysis was performed using the SpectralTrainFig App in MATLAB [17]. We have extracted 21 spectral band features from the decontaminated EEG signal as shown in Table. I which includes rapid eye movement (REM) power, Non-rapid eye movement (NREM) power, and Total power at each frequency band. Also, 102 EEG spectral features i.e. REM, N-REM power at single frequencies have been computed for 51 frequencies from 0 to 25 Hz with a 0.5 Hz gap i.e. 0 Hz, 0.5 Hz, 1 Hz, 1.5 Hz, …., 24.5 Hz, 25 Hz.

TABLE I. SPECTRAL FEATURES FROM EEG

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

D. *Feature Selection and Regression for Sleep Score Estimation*

Feature selection has been done primarily to compare the relative importance of ECG based features with EEG based features for sleep score estimation. Permutation importance has been used with Random Forest for ranking the feature importance [18]. Both ECG and EEG have correlated features that introduce the problem of multi-collinearity. To deal with this, conditional permutation has been used. *party* package from R has been used for the feature ranking [19].

For developing the regression model, the entire dataset was divided into train and test set in a ratio of 80:20 following a random shuffle. We investigated Extreme Gradient Boosting (XGBoost), Artificial Neural Network (ANN), and a sequential neural network called TabNet for regression. TabNet is an attention-based deep neural network optimized for tabular data and uses an encoder-decoder architecture [20]. The encoder part consists of a feature transformer, an attentive transformer, and feature masking at each decision step. The decoder part has a feature transformer block at each step. For hyper-parameter selection Bayesian search has been used.

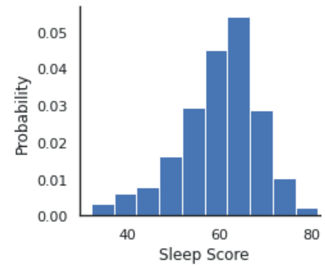

Fig. 2 Distribution of sleep score

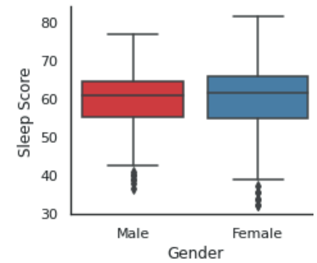

Fig. 3 Boxplot comparison for male and female sleep scores

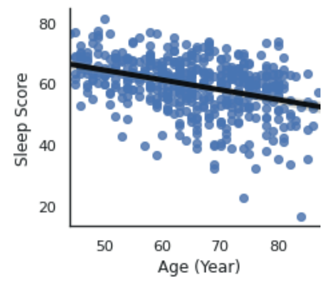

Fig. 4 Scatterplot of sleep score and age

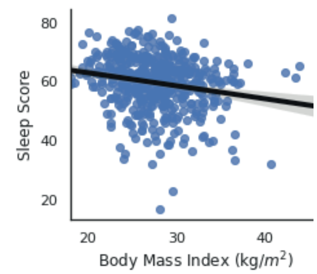

Fig. 5 Scatterplot of sleep score and body mass index.

III. RESULTS

The probability density plot of sleep score has been shown in Fig. 2. The histogram of sleep score follows a Gaussian distribution with a mean of 60 (N=500) and a standard deviation of 22. A boxplot comparison between the sleep scores of males and females has been shown in Fig. 3. No significant (p-value>0.05) difference was observed between the average sleep score of males with that of females. Sleep score shows a moderate (r=-0.35, p=0.0) negative correlation with age. The scatterplot of age and sleep score with a trendline has been visualized in Fig. 4. Similarly, sleep score shows a weak (r=-0.21, p=0.0) inverse correlation with Body Mass Index (BMI). The scatterplot of BMI and sleep score with a trendline has been visualized in Fig. 5. The trendline shows a downward slope indicating an inverse relationship i.e. people with higher BMI tend to have lower sleep scores. The partial correlation of sleep score with HRV and EEG features have been performed to investigate the relationship of sleep score with these features when controlled for age and BMI. Features (best 5 from each sensor) showing significant correlation has been listed in Table II.

TABLE II. PARTIAL CORRELATION OF FEATURES WITH SLEEP SCORE

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

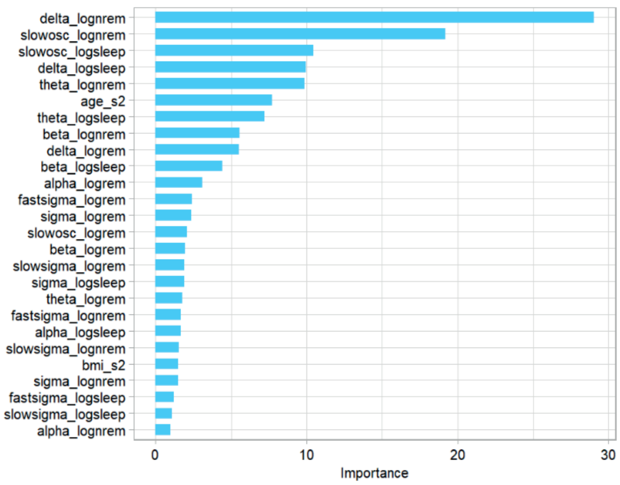

Fig. 6 Top 20 features based on permutation importance

Although both HRV and EEG features show significant partial correlation with sleep score, the correlation for EEG features is much stronger than HRV features. Since, EEG features have relatively higher importance compared to HRV features, in developing the regression method only EEG and anthropometric measures have been used. The ranking of top 20 EEG features (log transformed) based on conditional permutation importance have been shown in Fig. 6. The results of regression performance for estimation of sleep score using XGBoost, ANN and TabNet model have been shown in Table III. TabNet achieved the best performance with a Room Mean Squared Value (RMSE) of 4.65 and R-squared(R2) value of 0.64 in the training set, and an RMSE of 5.47 and R2 value of 0.59 in the test set. The finalized hyper-parameter for the TabNet model with bayesian search has been shown in Table IV. Adam has been used as the optimizer function and the masking function type was 'sparsemax'. The fit of the regression plot for the TabNet model has been shown in Fig. 7. The dashed line indicates the ideal and the solid line indicates the achieved trend line for actual versus predicted values. The histogram of prediction error in Fig. 8 shows symmetrically skewed and almost normally distributed patterns with a higher frequency in the error bin ± 2. The residual plot in Fig. 9 for the regression analysis shows a random scattering around the zero lines.

TABLE III. PERFORMANCE OF REGRESSION MODELS

Model	Performance			
	Train(80%)		Test(20%)	
	RMSE	R^2	RMSE	R^2
XGBoost	6.56	0.46	7.34	0.35
ANN	6.59	0.61	7.17	0.49
TabNet	4.65	0.64	5.47	0.59

TABLE IV. PARAMETERS OF TABNET MODEL

Parameter	**Value**
Width of the decision prediction layer	10
Width of the attention embedding for each mask	10
Number of steps in the architecture	3
gamma	1.3
epsilon	1e-15

IV. CONCLUSION

In this study, we analyzed sleep score and its relationship with anthropometric, HRV, and EEG based features. We performed a feature ranking for identifying the most informative features for sleep score estimation. Finally, we developed a regression method using a TabNet model for sleep score estimation from spectral features of single-channel EEG. The findings of this study increased the interpretability of sleep score and paves way for the usage of sleep score as a potential indicator for automated routine health check using wearables. In future studies, we aim to investigate the relationship of different diseases i.e. sleep apnea, insomnia, etc. with sleep score and implement the method in an edge device i.e. smartphone for online estimation of sleep scores.

ACKNOWLEDGMENT

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

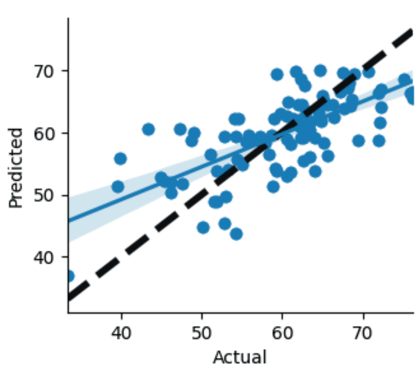

Fig. 7 Fit of regression plot for TabNet model

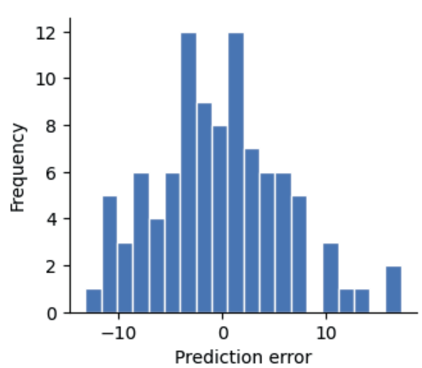

Fig. 8 Error Histogram for TabNet model

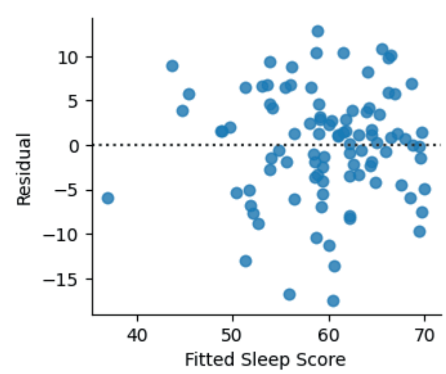

Fig. 9 Residual plot for TabNet model

REFERENCES

[1] Mukherjee, Sutapa, Sanjay R. Patel, Stefanos N. Kales, Najib T. Ayas, Kingman P. Strohl, David Gozal, and Atul Malhotra. "An official American Thoracic Society statement: the importance of healthy sleep. Recommendations and future priorities.", *American journal of respiratory and critical care medicine,* vol.191, no. 12, pp: 1450-1458, 2015.

[2] Spiegel, Karine, Esra Tasali, Rachel Leproult, and Eve Van Cauter. "Effects of poor and short sleep on glucose metabolism and obesity risk.", *Nature Reviews Endocrinology, vol.* 5, no. 5, pp: 253, 2009.

[3] O'Leary, Kimberly *et al*. "Why might poor sleep quality lead to depression? A role for emotion regulation." *Cognition & emotion,* vol. 31, no. 8, pp: 1698-1706., 2017. doi:10.1080/02699931.2016.1247035.

[4] Sharma, Monika, J. P. S. Sawhney, and Samhita Panda. "Sleep quality and duration–Potentially modifiable risk factors for Coronary Artery Disease?.", *indian heart journal,* vol. 66, no. 6, pp.: 565-568, 2014.

[5] Buysse, Daniel J., Charles F. Reynolds III, Timothy H. Monk, Susan R. Berman, and David J. Kupfer. "The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research." *Psychiatry research*, vol. 28, no. 2, pp: 193-213, 1989.

[6] Weaver, Terri E., Andréa M. Laizner, Lois K. Evans, Greg Maislin, Deepak K. Chugh, Kerry Lyon, Philip L. Smith *et al*. "An instrument to measure functional status outcomes for disorders of excessive sleepiness." *Sleep*, vol. 20, no. 10, pp: 835-843, 1997.

[7] Johns, Murray W. "A new method for measuring daytime sleepiness: the Epworth sleepiness scale." *Sleep, vol.* 14, no. 6, pp: 540-545, 1991.

[8] C. Kuo, Y. Liu, D. Chang, C. Young, F. Shaw and S. Liang, "Development and Evaluation of a Wearable Device for Sleep Quality Assessment," in *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1547-1557, July 2017.

[9] F. Mendonça, S. S. Mostafa, F. Morgado-Dias and A. G. Ravelo-García, "Sleep Quality Estimation by Cardiopulmonary Coupling Analysis," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 26, no. 12, pp. 2233-2239, Dec. 2018, doi: 10.1109/TNSRE.2018.2881361.

[10] I. Azimi *et al*., "Personalized Maternal Sleep Quality Assessment: An Objective IoT-based Longitudinal Study," in *IEEE Access*, vol. 7, pp. 93433-93447, 2019, doi: 10.1109/ACCESS.2019.2927781.

[11] Haghayegh, Shahab *et al*. "Accuracy of Wristband Fitbit Models in Assessing Sleep: Systematic Review and Meta-Analysis." *Journal of medical Internet research* vol. 21,11 e16273. 28 Nov. 2019, doi:10.2196/16273.

[12] Quan, Stuart F., Barbara V. Howard, Conrad Iber, James P. Kiley, F. Javier Nieto, George T. O'Connor, David M. Rapoport *et al*. "The sleep heart health study: design, rationale, and methods." *Sleep* 20, no. 12 (1997): 1077-1085.

[13] Rosipal, Roman, Achim Lewandowski, and Georg Dorffner. "In search of objective components for sleep quality indexing in normal sleep." *Biological psychology* 94, no. 1 (2013): 210-220.

[14] Landry, Glenn J *et al*. "Measuring sleep quality in older adults: a comparison using subjective and objective methods." *Frontiers in aging neuroscience* vol. 7 166. 7 Sep. 2015, doi:10.3389/fnagi.2015.00166

[15] Malik, M., Farrell, T., Cripps, T. and Camm, A.J.,"Heart rate variability in relation to prognosis after myocardial infarction: selection of optimal processing techniques." *European heart journal*, vol. 10 , no.12, pp.1060-1074, 1989.

[16] Goldberger, Ary L., Luis AN Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals." *circulation* 101, no. 23 (2000): e215-e220.

[17] Dennis A. Dean, II, SpectralTrainFig , MATLAB Central File Exchange. Retrieved February 14, 2020.

[18] Trang T. Le, Weixuan Fu and Jason H. Moore "Scaling tree-based automated machine learning to biomedical big data with a feature set selector". *Bioinformatics*, vol. 36(1), pp: 250-25, 2020.

[19] Package, R. "party." Repository CRAN. Available online: https://cran. r-project. org/web/packages/party/party. pdf (accessed on Aug 12 2020).

[20] Arik, Sercan O., and Tomas Pfister. "Tabnet: Attentive interpretable tabular learning." *arXiv* preprint arXiv:1908.07442, 2019.