2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Wavelet-based CNN for Predicting PAP Adherence Using Overnight Polysomnography Recordings: a Pilot Study

Mingxi Lei[1], Tom Maxim[2], Edwin M. Valladares[3], Eric Kezirian[4] and B. Keith Jenkins[1]

Abstract— **Obstructive sleep apnea (OSA) is a common sleep disorder. Positive airway pressure (PAP) therapy is the first-line treatment, while its effectiveness is significantly limited by incomplete adherence in many patients. This work aims to find a predictive association between data from in-laboratory sleep studies during treatment (PAP titration polysomnogram, or PSG) and PAP adherence. Based on a PAP titration PSG database, we present a pipeline to develop a wavelet-based deep learning model and address two challenges. First, to tackle the problem of extremely long overnight PSG signals, it randomly draws segments and extracts features locally. The global representation for the entire signal is achieved by local feature P-norm pooling. Second, to tackle the problem of limited dataset size, the pre-trained EfficienNet-B7 is used as an unsupervised feature extractor to transfer ImageNet knowledge to PSG signals in the wavelet domain. The trained pipeline achieves 78% balanced accuracy and 83% AUC on the test set using airflow and frontal EEG signals, which, we believe, is a compelling result as a pilot study.**

Clinical relevance— **Polysomnogram signals may improve clinical treatment of OSA by identifying patients with low likelihood of PAP adherence, enabling intensive efforts to improve adherence or consider alternative therapies.**

I. INTRODUCTION

Obstructive sleep apnea (OSA) is a common disorder characterized as the symptomatic, repeated blockage of breathing during sleep. Symptoms of OSA include daytime sleepiness, fatigue, and decreases in cognitive function. OSA also is associated with cardiovascular disease and other health-related consequences.

OSA diagnosis is established with an overnight sleep study, and historically the most common type of sleep study has been the polysomnogram (PSG). The PSG includes collection of multiple data signals during sleep, including electroencephalogram (EEG), electrooculogram (EOG), electrocardiogram (ECG), airflow and oxygen saturation (SpO2). The PSG determines the presence or absence of OSA as well as the severity of OSA.

Positive airway pressure therapy (PAP) is considered first-line OSA treatment because of its low risks and high efficacy [1]. PAP delivers positive pressure through the nose and/or mouth, functioning as a pneumatic splint to maintain an open upper airway. One of the challenges of PAP is the need to wear it comfortably during sleep. Although many patients do achieve success with PAP, the most common clinical care pathway involves patients receiving PAP at home to use for the first time on their own. Patients undergoing PSG for diagnosis of OSA may show clear evidence of OSA in the first portion of the night, such that the team in the sleep center may introduce PAP on the same night. This combination of PSG for diagnosis and initiation of treatment is termed a split-night PSG and allows a sleep center to evaluate a patient's clinical response to PAP initiation and changes in PAP settings (for example, the level of PAP pressure).

Although the goal is for patients to use PAP all night, every night, current clinical criteria for adequate PAP adherence is usage on at least 4 hours a night for at least 70% of all nights. Over the past decade, technological advances have enabled remote monitoring of PAP adherence. Substantial efforts are devoted to improving PAP adherence, but unfortunately approximately 30% of patients do not tolerate PAP and must consider other options. Because of the important negative consequences of untreated OSA, understanding early in the course of treatment whether individual patients were unlikely to tolerate PAP would allow clinicians to initiate aggressive attempts to enhance PAP adherence or to pursue alternative therapies and treat patients more effectively.

Recently, deep learning has been gaining tremendous attention due to their outperformance for various tasks, including healthcare applications. Regarding OSA, many previous studies have developed deep learning systems to diagnose OSA using PSG signals, while there have been no previous studies examining PAP adherence using PSG signals during PAP treatment. The aim of this study is to investigate whether there are associations between PAP titration PSG signals (treatment portion of the split-night PSG) and PAP adherence using a deep learning approach, and whether the former is predictive of the latter. We typically face two challenges. First, PSG data are limited to train a deep network, especially given that PSGs with PAP titration are not performed in all patients. Secondly, overnight PSG signals are extremely long, so that regular deep learning architectures are not feasible for them. The method for addressing these challenges are inspired from three previous works about processing extremely large histopathology images [2], [3], and long documents classification [4]. The sections below are organized as follows. Section II reviews some related previous work. Section III shows the pipeline for processing

[1]Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, 3740 McClintock Avenue, Los Angeles, CA, 90089, USA mingxile@usc.edu, bjenkins@usc.edu

[2]Department of Head & Neck Surgery, UCLA School of Medicine, 10833 Le Conte Avenue, Los Angeles, CA, 90095, USA tmaxim@mednet.ucla.edu

[3]USC Sleep Disorders Center, Keck Medicine of USC, 1500 San Pablo Street, Los Angeles, CA 90033, USA Edwin.Valladares@med.usc.edu

[4]USC Caruso Department of Otolaryngology – Head and Neck Surgery, Keck School of Medicine of USC, 1500 San Pablo Street, Los Angeles, CA 90033, USA Eric.Kezirian@med.usc.edu

overnight signals and developing predictive models from them. Section IV presents the results of 5 investigated PSG channels. Section V discusses and compares the method in this work with previous studies and conclusions are given in Section VI.

II. RELATED WORK

Previous studies about PAP adherence are limited, even though its overall status among OSA patients did not improve over more than 20 years. Early investigations identified predictive factors using conventional statistical analysis. The severity of OSA was identified to be related to PAP adherence using t-test, Mann-Whitney U-test, and χ^2-test; patients with less severe OSA are more likely to abandon PAP treatment [5]. Another study revealed that age is related to PAP adherence using Pearson correlation coefficients and 2-tailed t-test [6].

Machine (deep) learning might be an effective approach with available "big-data." However, a thorough search also only yields limited relevant studies. Three studies [7], [8], [9] are found and will be discussed in Section V.

III. METHODOLOGY

A. Definition of PAP Adherence

We adopt the common clinical definition of PAP Adherence as $>$4 hours/night on $\geq$5 nights/week during patients' overall treatment periods indicated in their PAP usage report. because it generally supports significant improvement for OSA [1], [10]. Let y_+ be the notation of good adherence; y_- be the notation of poor adherence.

B. Dataset

This study was approved by the University of Southern California institutional review board. A retrospective cohort database was assembled, comprising data from 202 study participants with split-night PSGs performed at the Keck Medicine of USC Sleep Disorders Center and with PAP adherence tracking using the AirView system (Resmed, San Diego, California, USA). Eighty-nine and 113 patients were labeled as y_+ and y_-, respectively, according to their PAP usage report. All patients received PAP treatment at least for 30 days, commencing after a medical evaluation and split-night PSG (described below). The split-night PSG consisted of two portions: diagnostic (without PAP) during the first portion, followed by PAP titration (with PAP). PSG data was collected during both stages. This study used data from the PAP titration portion only.

The dataset is randomly divided into Training Set (80%) and Test Set (20%). The training set is used to select optimal models and tune model parameters; the test set is used only for reporting the performance of models trained using the training set. We investigated 5 PSG channels: 1) SpO2, 2) frontal EEG, 3) central EEG, 4) occipital EEG, 5) airflow; the sample rates for them are 16Hz, 64 Hz, 256 Hz, 256 Hz, and 256 Hz, respectively.

C. Proposed framework overview

An overview of the proposed pipeline for developing models from PSG signals is given in Fig. 1, generally including 4 stages: 1) random subsequence sampling; 2) wavelet-based EfficientNet activated feature extraction; 3) feature engineering; 4) final classification.

D. Random subsequence sampling

Since overnight sleep studies produced extremely long-length signals, it is inevitable to employ deep learning architecture locally first. The random sampling method is adopted to generate local segments from the entire overnight treatment-stage signals. We refer to the local segments as subsequences in the following analysis. The number of subsequences for each signal is determined adaptively according to the length of entire signals, calculated by a simple equation as follows:

$$N = \left\lfloor \frac{T}{T_{sub}} \right\rfloor \tag{1}$$

Where T is the length of the entire signal, T_{sub} is the length of the subsequence, and N is the number of subsequences. This is designed as the trade-off between the computation cost and coverage of sampled subsequences. The entire signals can be generally covered without losing too much information, using the equation above (Fig. 2).

E. Downsample

Useful information typically lies within low-frequency ranges. Hence, downsample is used after random sampling to remove high-frequency components, and also reduce the data dimensionality. Multi-level discrete wavelet transform (DWT) is adopted to decimate subsequences. DWT has the advantage of dealing with non-stationary time series. Bioelectric time series usually includes sudden transitions, such as heartbeats in ECG. Comparing to classical Fourier-based filters, DWT can preserve the peak locations and their shape.

The choice of wavelets and level of DWT decomposition need to be determined for this step. We use the Daubechies-5 (db5) wavelet for the trade-off between its vanishing moment and oscillation of its wavelet function. Since SpO2 channel is already under a very low sample rate, no DWT is needed; DWT decomposition levels for the other 4 channels are specified proportional to their sample rate so that processed signals are under identical Nyquist sample rates. For instance, after using 2-level DWT decomposition, the Nyquist sample rate of downsampled airflow signal is reduced to $\frac{64}{2\times2} = 16$Hz; using 4-level DWT decomposition, the Nyquist sample rate of downsampled EEG signal is reduced to $\frac{256}{2\times2\times2\times2} = 16$Hz.

F. Wavelet-based feature extraction

A 1-dimensional (downsampled) subsequence is transformed into the 2-dimensional time-frequency domain using continuous wavelet transform (CWT), i.e., the two axes of the response image correspond to time and frequency resolutions. We employ the complex Morlet wavelet, the default choice for wavelet analysis [11] for its simplicity

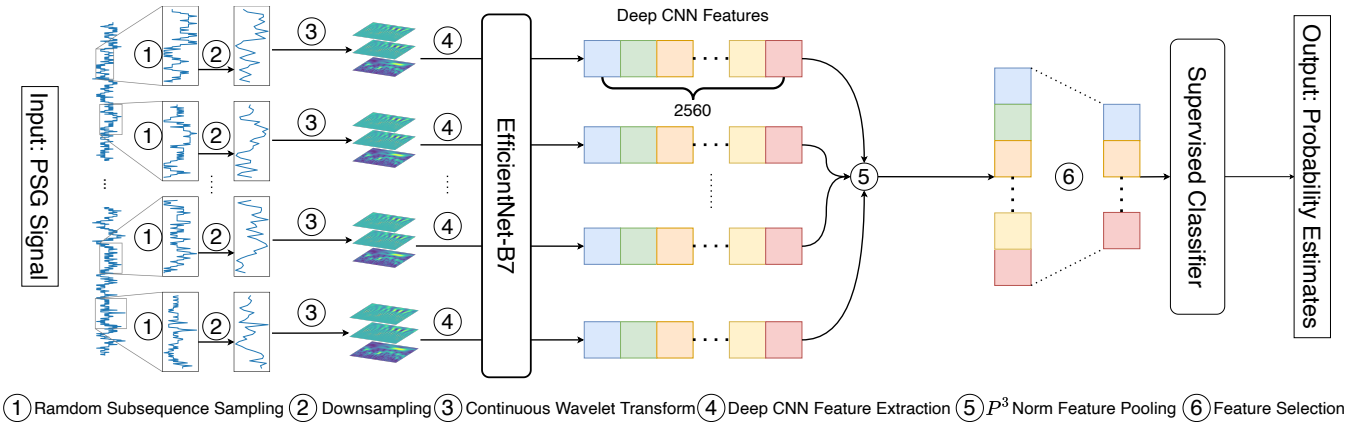
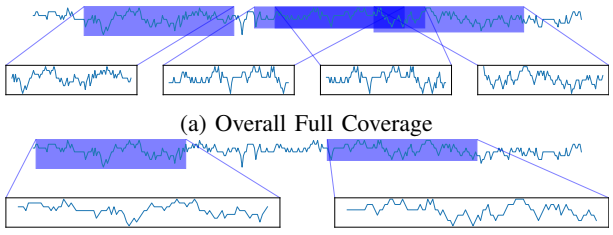


Fig. 1: Model Pipeline



(a) Overall Full Coverage

(b) Improper Sparse Coverage

Fig. 2: Subsequence Sampling

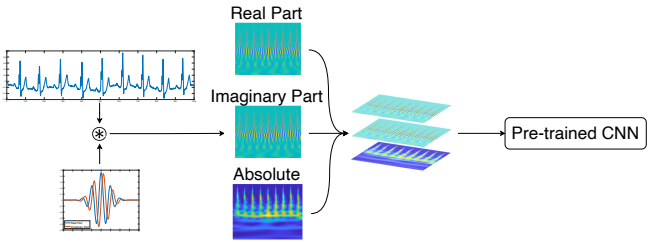


Fig. 3: Input Volume Preparation

and equal spread in time and frequency. It is constructed by a complex exponential multiplied by a Gaussian envelope:

$$\psi(t) = \pi^{-\frac{1}{4}} e^{-\frac{t^2}{2}} e^{j5t} \tag{2}$$

The complex Morlet wavelet produces a complex-value response image. The absolute value of it is called a scalogram, which provides visual characteristics of the subsequence about time, frequency, and amplitude (Fig. 3). The real part and imaginary part of the complex-value response also provide information of phase.

Visual features are then extracted using a convolutional neural network (CNN). We choose the pre-trained EfficientNet-B7 as the feature extractor for two reasons. First, it receives a large size of input (600×600), which is able to cover longer subsequences. Secondly, according to a recent study, better ImageNet models transfer better on other tasks [12]. Considering the ImageNet leaderboard and model size, EfficientNet-B7 is a good choice.

Frequencies of interest are set logarithmically spaced between 0.1 Hz to 4 Hz. The response images of all subsequences are fed into EfficientNet-B7 to obtain 2560 CNN-activated features. Since the CWT responses are complex-valued and CNN accepts real-valued inputs only, each complex-valued input image is transformed into a 3-channel real-valued volume to make it compatible with the pre-trained EfficientNet-B7, where the 3 channels represent the real part, imaginary part, and absolute value, respectively. Such a data format also mimics the structure of colored images with RGB (red, green, blue) channels, each of which provides insights of subsequences individually, but also provides fully joint insights together with other channels about time, frequency, amplitude, and phase in the wavelet domain (Fig. 3).

G. Feature engineering

Each overnight signal is represented by a feature matrix $\mathbf{F} \in \mathbb{R}^{N \times 2560}$ after feature extraction. All subsequence-level feature vectors should be combined and aggregated to obtain the global patterns of the entire overnight signals. P-norm pooling is adopted as follows:

$$f_d = \left(\frac{1}{N}\sum_{i=1}^{N}(F_{i,d})^p\right)^{\frac{1}{p}} \tag{3}$$

where $F_{i,d}$ is the d-th feature value at the i-th subsequence, f_d is the pooled d-th feature value, and the parameter $P = 3$ as suggested in a theoretical study [13]. Each overnight signal then is represented as a pooled feature vector $\mathbf{f} \in \mathbb{R}^{1 \times 2560}$.

Feature selection is further conducted to select a subset of informative features, including 3 sequential analyses (Table I): 1) feature stability assessment, 2) univariate feature analysis, and 3) multivariate feature analysis.

Feature stability refers to the robustness with respect to data sampling and to its stochastic nature [14]. Specifically, in this work, features are expected to be consistent across different random seeds for the random sampling. Unstable features that vary easily by the changing of subsequences positions should be excluded to avoid unreliable predictions.

TABLE I: Number of Selected Features in Each Step

	SpO2	Airflow	Frontal EEG	Cenral EEG	Occipital EEG
Feature Stability Assessment	1930	2546	2555	2547	2548
Univariate Feature Analysis	1414	1387	591	1314	1225
Multivariate Feature Analysis	56	63	49	63	57

This is accessed using the intraclass correlation coefficient (ICC). We generated 2 reference feature sets by replicating the feature extraction twice with different random seeds. Intuitively, stable features should not change a lot, comparing to the 2 reference sets. The ICC for each feature is calculated from 3 feature sets. Features with ICC < 0.75 are considered to be unstable and are excluded. Stable features help improve the model performance and reproducibility.

Univariate feature selection examines every feature individually to assess the relevance to PAP adherence. Features are scored using the area under the curve (AUC) from 5-fold cross-validation of univariate logistic regression. We exclude irrelevant features with AUC ≤ 0.5.

Multivariate features analysis not only assesses the relevance to the PAP adherence but also considers the interaction between features in the meantime and removes redundant ones. We used the minimum-redundancy-maximum-relevance (mRMR) algorithm [15] to select the final feature subset. Specifically, a parallel ensemble version of it [16] was adopted to mitigate the drawback of its greedy heuristic property and find a more robust feature subset. Twenty ensembles are implemented in parallel. Each ensemble selects the top 5‰ for feature sets extracted from SpO2, airflow, central EEG, and occipital EEG channels, while for frontal EEG the percentage is relaxed to 10‰ because its feature set is much smaller after the prior two steps (Table I).

H. Final classification

A LASSO regression model is fit using the selected features. The glmnet package in R is adopted to tune the model efficiently. Regularization parameter λ is determined by 5-fold cross-validation on the training set.

I. Inference Strategy

Unseen test data samples pass through the trained pipeline to infer its predicted output, while we increase the number of subsequences extracted from the entire signal to produce a more accurate result. The pipeline is run 5 times in parallel with different random seeds for each data sample to output 5 predictions. The final prediction is given by their average.

IV. RESULTS

Five models are trained using the 5 investigated PSG channels, where each model is trained with a specific parameter set as described in Section III. Performance of them is reported using multiple metrics (Table II): precision, recall, F1-score, balanced accuracy, and AUC.

The overnight SpO2 and occipital EEG signal are not likely to be related to the PAP adherence since the AUC scores are around 0.5. The central EEG signal shows weak relation to the PAP adherence, reporting an AUC of 61% on the test set. Both airflow and frontal EEG signal achieve an AUC of 76%, indicating a strong predictive association with the PAP adherence. We further develop an ensemble model of the airflow and frontal EEG signal, i.e., the predictions of the 2 PSG channels are combined by calculating the average of them to obtain a better prediction. The ensemble model reports the highest AUC of 83% and balanced accuracy of 78% (Table II).

Besides measuring the metrics above, probability estimates for test samples provide insights into model performance as well. The LASSO regression is a probabilistic model, where the probability of y_+ is given by the sigmond function

$$P(y_+|\mathbf{f}) = \frac{1}{1+\exp^{-\beta\cdot\mathbf{f}}} \tag{4}$$

where β are the model parameters for LASSO regression. Ideally, predictions for patients with class y_- should be close to 0, while those with class y_+ should be close to 1. Visualizations of the distribution of $P(y_+|\mathbf{f}) - 0.5$ (i.e., distance to the decision boundary) for the test set is given in Fig. 4. Negative class samples (y_-) received better predictions since they are far away from the decision boundary ($P(y_+|\mathbf{f}) = 0.5$).

V. DISCUSSION

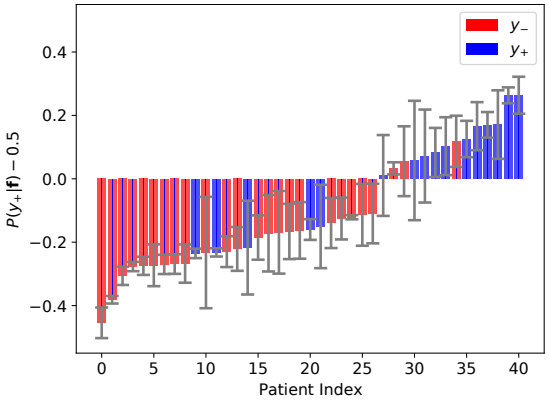
This study has demonstrated that PAP adherence is associated with a combination of airflow and frontal EEG signals during PAP titration. To our knowledge, this is the first study of its kind. PSG signals are complex and noisy data, with considerable physiological variation during the overnight study.

Previous PAP adherence studies using machine learning have focused on early PAP usage patterns over months to detect long-term PAP non-adherence [7], [8] and driving performance data for its association with the prior night's PAP usage [9]. Our work broadens the scope of machine learning research to offer another tool (at the absolute first use of PAP) to identify individual patients who were unlikely to tolerate PAP. As described earlier, this could allow clinicians to initiate aggressive attempts to enhance PAP adherence (avoiding the wasted patient and clinician resources on patients who will have no difficulties with PAP adherence) or to pursue alternative therapies and treat patients more effectively.

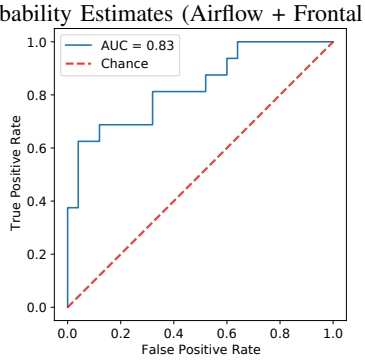
This study has important limitations. First, many patients receive PAP therapy without a PAP titration PSG, so wide-scale use of this approach would require a change in practice patterns or else collection of airflow (often collected by PAP machines) and frontal EEG data (not currently collected by PAP machines). Future research may examine signals from diagnostic PSG or home sleep apnea tests. Second,

TABLE II: Model Performance

	Precision	Precision(-)	Recall	Recall(-)	F1	F1(-)	Balanced Accuracy	AUC
SpO2	0.50	0.65	0.31	0.80	0.38	0.71	0.56	0.46
Airflow	0.62	0.71	0.50	0.80	0.55	0.75	0.65	0.76
Frontal EEG	0.73	**0.81**	**0.69**	0.84	0.71	0.82	0.76	0.76
Central EEG	0.44	0.65	0.50	0.60	0.47	0.63	0.55	0.61
Occipital EEG	0.42	0.62	0.31	0.72	0.36	0.67	0.52	0.52
Airflow + Frontal EEG	**0.79**	**0.81**	**0.69**	**0.88**	**0.73**	**0.85**	**0.78**	**0.83**



(a) Probability Estimates (Airflow + Frontal EEG)



(b) ROC (Airflow + Frontal EEG)

Fig. 4: Assessment of Probability Estimates. Error bar represents the standard deviation of outputs of 5 parallel runs with different random seeds

this research comes from a single center, so the work would benefit from inclusion of larger datasets from multiple centers to refine the algorithm and evaluate generalizability.

VI. CONCLUSION

This work investigates the predictive association between PAP titration PSG signals and PAP adherence in OSA. We report a pipeline to process overnight PSG signals using Wavelet-based Deep CNN feature representation. The airflow and frontal EEG combined show predictive power for PAP adherence, evaluated using multiple metrics. Further study can investigate whether diagnostic PSG signals or other sleep study data predict PAP adherence.

REFERENCES

[1] F. Campos-Rodriguez, M. Martinez-Alonso, M. S. de-la Torre, and F. Barbe, "Long-term adherence to continuous positive airway pressure therapy in non-sleepy sleep apnea patients," *Sleep Medicine*, vol. 17, pp. 1–6, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389945715020018

[2] Y. Xu, Z. Jia, L.-B. Wang, Y. Ai, F. Zhang, M. Lai, I. Eric, and C. Chang, "Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features," *BMC bioinformatics*, vol. 18, no. 1, pp. 1–17, 2017.

[3] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," in *International Conference Image Analysis and Recognition*. Springer, 2018, pp. 737–744.

[4] L. Liu, K. Liu, Z. Cong, J. Zhao, Y. Ji, and J. He, "Long length document classification by local convolutional feature aggregation," *Algorithms*, vol. 11, no. 8, p. 109, 2018.

[5] C. Janson, E. Nöges, S. Svedberg-Brandt, and E. Lindberg, "What characterizes patients who are unable to tolerate continuous positive airway pressure (cpap) treatment?" *Respiratory medicine*, vol. 94, no. 2, pp. 145–149, 2000.

[6] R. Budhiraja, S. Parthasarathy, C. L. Drake, T. Roth, I. Sharief, P. Budhiraja, V. Saunders, and D. W. Hudgel, "Early cpap use identifies subsequent adherence to cpap therapy," *Sleep*, vol. 30, no. 3, pp. 320–324, 2007.

[7] M. Araujo, R. Bhojwani, J. Srivastava, L. Kazaglis, and C. Iber, "Ml approach for early detection of sleep apnea treatment abandonment: A case study," in *Proceedings of the 2018 International Conference on Digital Health*, 2018, pp. 75–79.

[8] M. Araujo, L. Kazaglis, C. Iber, and J. Srivastava, "A data-driven approach for continuous adherence predictions in sleep apnea therapy management," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 2716–2725.

[9] A. D. McDonald, J. D. Lee, N. S. Aksan, J. D. Dawson, J. Tippin, and M. Rizzo, "Highway healthcare: How naturalistic driving data index adherence to cpap therapy in obstructive sleep apnea," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2013, pp. 1859–1863.

[10] P. A. Cistulli, J. Armitstead, J.-L. Pepin, H. Woehrle, C. M. Nunez, A. Benjafield, and A. Malhotra, "Short-term cpap adherence in obstructive sleep apnea: a big data analysis using real world data," *Sleep Medicine*, vol. 59, pp. 114–116, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389945718307974

[11] M. P. Wachowiak, R. Wachowiak-Smolíková, M. J. Johnson, D. C. Hay, K. E. Power, and F. M. Williams-Bell, "Quantitative feature analysis of continuous analytic wavelet transforms of electrocardiography and electromyography," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2126, p. 20170250, 2018.

[12] S. Kornblith, J. Shlens, and Q. V. Le, "Do better imagenet models transfer better?" in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2661–2671.

[13] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 111–118.

[14] S. Nogueira, K. Sechidis, and G. Brown, "On the stability of feature selection algorithms." *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 6345–6398, 2017.

[15] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[16] N. De Jay, S. Papillon-Cavanagh, C. Olsen, N. El-Hachem, G. Bontempi, and B. Haibe-Kains, "mrmre: an R package for parallelized mrmr ensemble feature selection," *Bioinformatics*, vol. 29, no. 18, pp. 2365–2368, 2013.