ECG-Derived Respiration for Sleep-Wake Stage Classification

Roneel V. Sharan
Australian Institute of Health Innovation
Macquarie University
Sydney, Australia
roneel.sharan@mq.edu.au

***Abstract*—Sleep disorders affect millions of people worldwide. Polysomnography (PSG) is a sleep study that is commonly used to diagnose sleep disorders, such as using sleep staging. However, PSG can be labor intensive, time consuming, expensive, and may not be easily available. Sleep and wake cycles can cause variation in heart rate and respiration which can be estimated using electrocardiogram (ECG), available as wearable sensors. As such, this work studies the use of single-lead ECG for detecting sleep and wake stages, in particular, using the heart rate variability (HRV) and ECG-derived respiration (EDR) signals. Various temporal and spectral descriptors are extracted from the HRV and EDR signals for this purpose. Sequential backward feature selection is employed to select the discriminative features for classification using logistic regression. The proposed method is evaluated on a dataset of more than 85 hours of ECG recordings from 16 subjects in leave-one-subject-out cross-validation. An accuracy of 75% (AUC=0.83) is achieved using the EDR features in classifying sleep and wake stages. This increased to an accuracy of 80% (AUC=0.88) when combined with HRV features. The proposed method demonstrates potential to be used for screening sleep disorders using ECG.**

Keywords—electrocardiogram, feature selection, logistic regression, respiration, sleep disorders

I. Introduction

Approximately one-third of adults in the United States reported short sleep in 2017, an increase of about 15% since 2004 [1]. Similarly, results of a 2016 Australian survey show that 33-45% of adults suffer from inadequate sleep and its daytime consequences [2]. Sleeping habits are intrinsically related to mental and physical health [3]. Lack of sleep is associated with decline in cognitive performance, motor vehicle accidents, poor quality of life, heightened socioeconomic burden, and various other mental and physical conditions [3]. Sleep disorders, such as insomnia and obstructive sleep apnea, are a common cause of sleep deficiency [2].

Early diagnosis and treatment of sleep disorders may help regain the restorative function of sleep, reduce daytime sleepiness, reduce the risk of accidents and cardiovascular diseases, and improve quality of life [4]. Polysomnography (PSG) is a type of sleep study that is widely used in the diagnosis of sleep disorders. It is a multi-parametric test measuring the biophysiological changes, including brain activity, heart rhythm, eye movements, and muscle activity, during sleep. The recorded data is divided into small time windows called epochs and multiple physiological signals are analyzed in each epoch to determine the sleep (rapid eye movement (REM) sleep and non-REM (NREM) sleep) and wake stages. Sleep staging is used to determine sleep latency, total sleep time, sleep efficiency, sleep onset, amongst others [5]. These, along with other metrics such as the apnea-hypopnea index (AHI), are useful in determining the sleep disorder.

However, overnight in-laboratory PSG is labor intensive, time consuming, expensive, and not readily available [6, 7]. Sleep and wake stages cause variation in the heart rate which can be analyzed using the heart rate variability (HRV) [8, 9]. The HRV can be computed from a single-lead electrocardiogram (ECG) signal and descriptors of the HRV have shown promise in differentiating between sleep and wake stages [10-12]. ECG is available as wearable technology and ECG-based sleep-wake stage classification has the potential to act as a screening tool for sleep disorders. Respiration is also affected during sleep-wake stages and combined analysis of ECG and respiration signals has shown improvement in detecting sleep-wake stages [13, 14]. The respiration signal can, however, be estimated from a single-lead ECG signal, referred as ECG-derived respiration (EDR) [15]. While EDR has been well studied, such as in detecting sleep apnea, its application in detecting sleep stages has received limited attention [16, 17].

This study explores the use of single-lead ECG signal for classification of sleep-wake stages. In particular, it focuses on temporal and spectral analysis of the heart rate and respiration signals derived from the ECG for this purpose. HRV and EDR are unevenly sampled data. Unlike earlier work [12, 13], where frequency analysis is performed using Fourier transform, this work utilizes Lomb-Scargle periodogram [18] which has shown to better estimate the power spectral density of unevenly sampled signal than Fourier transform based methods [19]. The extracted features are classified using logistic regression and the discriminative features are selected using sequential backward feature selection. In addition, instead of mixing epochs from all subjects in cross-validation [17, 20], subject independent cross-validation is performed in this work to evaluate the performance of the proposed method in detecting sleep-wake stages.

978-1-6654-2899-6/21/$31.00 ©2021 IEEE

II. METHOD

A. Dataset

This work utilizes the MIT-BIH PSG Database [21, 22]. The database has a collection of 18 recordings of multiple physiologic signals, sampled at 250 Hz, during sleep from 16 subjects who were monitored in Boston's Beth Israel Hospital Sleep Laboratory for evaluation of obstructive sleep apnea and to test the effects of constant positive airway pressure (CPAP).

An illustration of the PSG signals in the dataset is given in Fig. 1 and an overview of the database is provided in Table I. The ECG signal in each recording has been annotated beat-by-beat together with sleep stage annotation of 30 second epochs using EEG and respiration signals. PSG records *slp01a* and *slp01b* belong to the same subject, separated by a gap of about 1 hour, and records *slp02a* and *slp02b* are from another subject, separated by a gap of about 10 minutes. The other 14 records are from different subjects. The sleep apnea annotations for records *slp41* and *slp45* were unavailable; their AHI was estimated using visual review.

The database contains over 85 hours of PSG recordings and the duration of the recordings varies from 1.17 to 6.30 hours. All 16 subjects are male, aged 32 to 56 years (average age of 42.24 years) with weight in the range of 89 to 152 kg (average weight of 118.63 kg). The AHI of the subjects is in the range of 0.7 to 100.8; one subject with AHI ≤ 5, one subject with $5 \leq$ AHI < 15, three subjects with $15 \leq$ AHI < 30, and eleven subjects with AHI ≥ 30. The number of epochs per subject is in the range of 154 to 780 with a total of 10,197 epochs.

B. Proposed Method

This work considers the binary classification task of sleep epochs against wake epochs. As such, all sleep stages, REM and NREM, are considered as a single class. While sleep staging in the dataset has been performed in 30 second epochs, in classifying the current epoch, a 5 minute window, current plus adjacent epochs, is used for feature extraction, as recommended for analysis by the Task Force of the European Society of Cardiology and the North American Society of Pacing Electrophysiology [23]. After removing epochs with missing beat annotations, the usable dataset has a total of 9,977 epochs; 6,978 sleep epochs and 2,999 wake epochs. Time and frequency domain features are extracted from the HRV or RR interval signal, computed from the beat annotations, together with frequency domain features from the EDR signal.

1) RR Interval Time-Domain Features: The following features are extracted from the RR interval signal in each 5 minute window: *AVNN* – mean of the RR intervals, *SDNN* – standard deviation of RR intervals, *SKNN* – skewness of RR intervals, *KUNN* – kurtosis of RR intervals, *RMSSD* – root mean square of successive RR interval differences, *SDSD* – standard deviation of successive RR interval differences, *NN50* – number of pairs of successive RRs that differ by more than 50ms, *pNN50* – fraction of RR intervals that differ by more than 50ms, *NN20* – number of pairs of successive RRs that differ by more than 20ms, and *pNN20* – fraction of RR intervals that differ by more than 20ms.

Fig. 1. Illustration of the first 10 seconds of electrocardiogram (ECG), blood pressure (BP), electroencephalogram (EEG), respiration (Resp), stroke volume (SV), and oxygen saturation (SO2) signals from a PSG recording.

TABLE I. OVERVIEW OF THE DATABASE USED IN THIS WORK

Record	Duration (Hrs)	AHI	Age (Years)	Weight (kg)	Num. of Epochs
slp01a	2:00	17	44	89	240
slp01b	3:00	22.3	44	89	360
slp02a	3:00	34	38	145	360
slp02b	2:15	22.2	38	145	270
slp03	6:00	43	51	152	720
slp04	6:00	59.8	40	108	720
slp14	6:00	30.7	37	152	714
slp16	6:00	53.1	35	118	694
slp32	5:20	22.1	54	92	640
slp37	5:50	100.8	39	125	698
slp41	6:30	60	45	145	780
slp45	6:20	5	42	133	760
slp48	6:20	46.8	56	-	760
slp59	4:00	55.3	41	111	458
slp60	5:55	59.2	49	108	710
slp61	6:10	41.2	32	91	720
slp66	3:40	65.5	33	95	439
slp67x	1:17	0.7	-	-	154

2) RR Interval Frequency Domain Features: A 512-point Lomb-Scargle periodogram was computed for the RR interval signals, up to a frequency of 0.4 Hz, from which the frequency domain features, 32 equally spaced subband energies [24], were computed.

3) EDR Frequency Domain Features: The EDR algorithm derives a sample of a respiratory signal for each QRS complex by projecting that axis onto the lead axis. The EDR signal was estimated using [15] and frequency analysis was performed same as for the RR interval signal resulting in 32 EDR frequency-domain features.

As such, the final feature set has 10 time-domain RR interval features, 32 frequency domain RR interval features, and 32 EDR features – a total of 74 features. The features are standardized using z-score and the discriminant features are identified using sequential backward feature selection [25] for classification using logistic regression.

TABLE II. SLEEP AND WAKE EPOCH CLASSIFICATION RESULTS USING VARIOUS FEATURE SETS

Features	Sensitivity	Specificity	Accuracy	AUC	F1-Score
RR Time	0.7094	0.7129	0.7104	0.7740	0.7741
RR Frequency	0.6662	0.8083	0.7089	0.7929	0.7620
EDR Frequency	0.7463	0.7723	0.7541	0.8274	0.8094
RR Time + RR Frequency	0.7091	0.8033	0.7374	0.8217	0.7907
RR Time + RR Frequency + EDR Frequency	**0.7939**	**0.8086**	**0.7983**	**0.8811**	**0.8463**

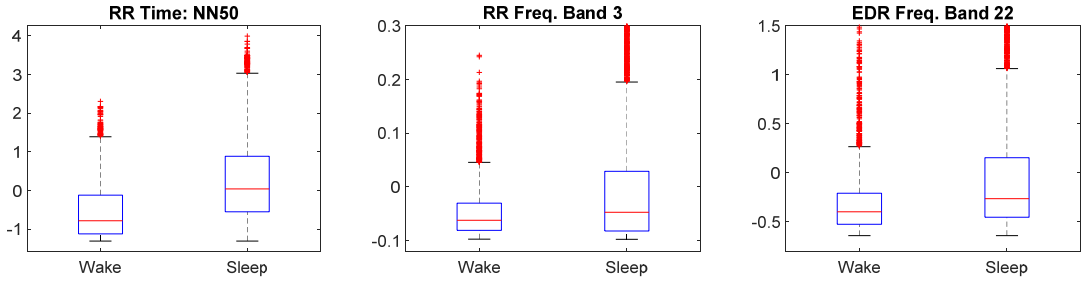

Fig. 2. Box plot of the most discriminative RR interval and EDR features.

C. Evalution Metrics

The performance of the proposed method is evaluated using sensitivity, specificity, accuracy, area under the curve (AUC) of the receiver operating characteristic (ROC) curve, and F1-score, where sensitivity and specificity are the proportion of sleep and wake epochs that are correctly classified, respectively. The optimal threshold on the ROC curve is determined as the point on the ROC that minimizes the distance to the point (0,1).

III. EVALUATION RESULTS

A. Experimental Setup

The performance of the model is evaluated in leave-one-subject-out cross-validation whereby, in each fold, the epochs from 15 subjects are used to train the classifier and epochs from the remaining subject are used for validation. The features are standardized and selected using sequential backward feature selection in each fold.

The performance is evaluated on the three individual feature sets: RR interval time-domain features, RR interval frequency domain features, and the EDR frequency domain features. In addition, to gauge the effectiveness of the EDR features on the combined feature set, the performance is first evaluated on the combined RR interval features (time + frequency) and then with the combined RR interval and EDR features.

B. Results

The sleep vs wake epoch classification results using the different feature sets are given in Table II. An accuracy of 0.7104 (AUC=0.7740) is achieved using RR interval time-domain features and an accuracy of 0.7089 (AUC=0.7929) using RR interval frequency domain features. As such, there is slight improvement in the AUC using the frequency domain features. An accuracy of 0.7541 (AUC=0.8274) is achieved using the EDR features which is the best results of the three individual feature sets.

The boxplot of the most discriminative feature from each feature group is illustrated in Fig. 2. These are the *NN50* feature from the RR interval time-domain feature set, subband energy in frequency band 3 (corresponding to frequency 0.025-0.0375 Hz) from the RR interval frequency-domain feature set, and subband energy in frequency band 22 (corresponding to frequency 0.2625-0.2750 Hz) from the EDR feature set.

Furthermore, an accuracy of 0.7374 (AUC=0.8217) is achieved using the combined RR interval features which is an improvement over the individual RR interval feature sets but lower than what is achieved using the EDR features. The inclusion of EDR features in the combined RR interval features improves the accuracy to 0.7983 (AUC=0.8811) together with a sensitivity, specificity, and F1-score of 0.7939, 0.8086, and 0.8463, respectively. These are the best results of all the individual and combined feature sets considered in this work.

The ROC curves using the different feature sets are illustrated in Fig. 3 which further demonstrates the superiority of the EDR features and the combined RR interval and EDR features. In addition, the ROC curves for the four different AHI groups using the best feature set are given in Fig. 4. The highest AUC is achieved for the group $15 \leq \text{AHI} < 30$ and lowest for $\text{AHI} \geq 30$. However, the number of subjects in the AHI groups is unevenly distributed with some groups having only one subject, making a conclusive interpretation difficult.

IV. DISCUSSION AND CONCLUSION

A method for classification of sleep stages using ECG signal analysis in presented in this paper. An accuracy of 80%

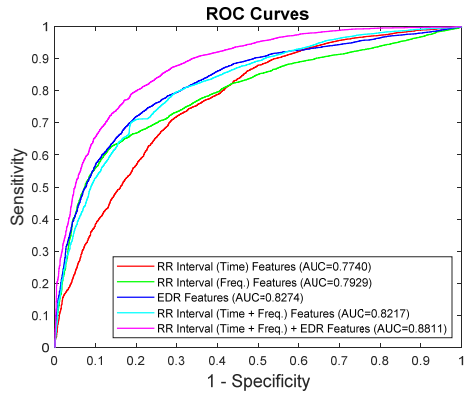

Fig. 3. ROC curves for sleep-wake stage classification using various feature sets.

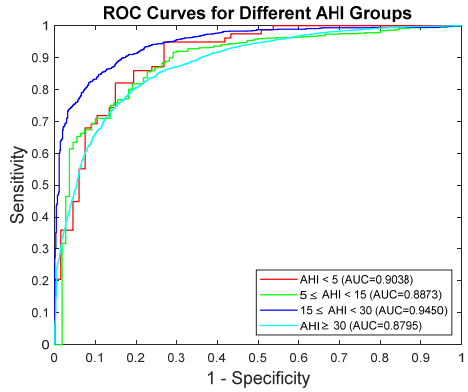

Fig. 4. ROC curves for sleep-wake stage classification for different AHI groups.

(AUC=0.8811) is achieved using combined HRV and EDR features, a relative improvement of 8.26% in accuracy and 7.23% in AUC over the HRV features. This work, however, has some limitations. In particular, the database is relatively small, there are no female subjects, and there are lack of healthy subjects and subjects with other types of sleep disorders. Therefore, it is hoped to evaluate the performance of the proposed method on a larger and more diverse dataset in future.

REFERENCES

[1] C. M. Sheehan, S. E. Frochen, K. M. Walsemann, and J. A. Ailshire, "Are U.S. adults reporting less sleep?: Findings from sleep duration trends in the National Health Interview Survey, 2004–2017," *Sleep,* vol. 42, no. 2, 2019.

[2] R. J. Adams *et al.*, "Sleep health of Australian adults in 2016: Results of the 2016 Sleep Health Foundation national survey," *Sleep Health,* vol. 3, no. 1, pp. 35-42, 2017.

[3] M. Tahmasian *et al.*, "The interrelation of sleep and mental and physical health is anchored in grey-matter neuroanatomy and under genetic control," *Communications Biology,* vol. 3, no. 1, p. 171, 2020.

[4] J. T. Maurer, "Early diagnosis of sleep related breathing disorders," *GMS Current Topics in Otorhinolaryngology - Head and Neck Surgery,* vol. 7, p. Doc03, 2008.

[5] D. Shrivastava, S. Jung, M. Saadat, R. Sirohi, and K. Crewson, "How to interpret the results of a sleep study," *Journal of Community Hospital Internal Medicine Perspectives,* vol. 4, no. 5, p. 24983, 2014.

[6] S. Chokroverty, M. Bhatt, and T. Goldhammer, "1 - Polysomnographic Recording Technique," in *Atlas of Sleep Medicine*, S. Chokroverty, M. Bhatt, and R. J. Thomas Eds. Philadelphia: Butterworth-Heinemann, 2005, pp. 1-28.

[7] B. Prasad, D. W. Carley, and J. J. Herdegen, "Continuous positive airway pressure device-based automated detection of obstructive sleep apnea compared to standard laboratory polysomnography," *Sleep and Breathing,* vol. 14, no. 2, pp. 101-107, 2010.

[8] S. Elsenbruch, M. J. Harnish, and W. C. Orr, "Heart rate variability during waking and sleep in healthy males and females," *Sleep,* vol. 22, no. 8, pp. 1067-1071, 1999.

[9] T. Penzel, J. W. Kantelhardt, C.-C. Lo, K. Voigt, and C. Vogelmeier, "Dynamics of heart rate and sleep stages in normals and patients with sleep apnea," *Neuropsychopharmacology,* vol. 28, no. 1, pp. S48-S53, 2003.

[10] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias, and T. Penzel, "A review of obstructive sleep apnea detection approaches," *IEEE Journal of Biomedical and Health Informatics,* vol. 23, no. 2, pp. 825-837, 2019.

[11] F. Bozkurt, M. K. Uçar, C. Bilgin, and A. Zengin, "Sleep–wake stage detection with single channel ECG and hybrid machine learning model in patients with obstructive sleep apnea," *Physical and Engineering Sciences in Medicine,* vol. 44, no. 1, pp. 63-77, 2021.

[12] R. Wei, X. Zhang, J. Wang, and X. Dang, "The research of sleep staging based on single-lead electrocardiogram and deep neural network," *Biomedical Engineering Letters,* vol. 8, no. 1, pp. 87-93, 2018.

[13] W. Karlen, C. Mattiussi, and D. Floreano, "Sleep and wake classification with ECG and respiratory effort signals," *IEEE Transactions on Biomedical Circuits and Systems,* vol. 3, no. 2, pp. 71-78, 2009.

[14] P. Fonseca, X. Long, M. Radha, R. Haakma, R. M. Aarts, and J. Rolink, "Sleep stage classification with ECG and respiratory effort," *Physiological Measurement,* vol. 36, no. 10, pp. 2027-2040, 2015.

[15] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, "Derivation of respiratory signals from multi-lead ECGs," in *Computers in Cardiology*, Linköping, Sweden, 1985, vol. 12, pp. 113-116.

[16] S. J. Redmond and C. Heneghan, "Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea," *IEEE Transactions on Biomedical Engineering,* vol. 53, no. 3, pp. 485-496, 2006.

[17] A. Rahimi, A. Safari, and M. Mohebbi, "Sleep stage classification based on ECG-derived respiration and heart rate variability of single-lead ECG signal," in *26th National and 4th International Iranian Conference on Biomedical Engineering (ICBME)*, Tehran, Iran, 27-28 Nov 2019, pp. 158-163.

[18] N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophysics and Space Science,* vol. 39, no. 2, pp. 447-462, 1976.

[19] P. Laguna, G. B. Moody, and R. G. Mark, "Power spectral density of unevenly sampled data by least-square analysis: Performance and application to heart rate signals," *IEEE Transactions on Biomedical Engineering,* vol. 45, no. 6, pp. 698-715, 1998.

[20] M. Adnane, Z. Jiang, and Z. Yan, "Sleep–wake stages classification and sleep efficiency estimation using single-lead electrocardiogram," *Expert Systems with Applications,* vol. 39, no. 1, pp. 1401-1413, 2012.

[21] Y. Ichimaru and G. B. Moody, "Development of the polysomnographic database on CD-ROM," *Psychiatry and Clinical Neurosciences,* vol. 53, no. 2, pp. 175-177, 1999.

[22] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," (in eng), *Circulation,* vol. 101, no. 23, pp. e215-e220, 2000.

[23] Task Force of the European Society of Cardiology and the North American Society of Pacing Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Circulation,* vol. 93, no. 5, pp. 1043-1065, 1996.

[24] P. de Chazal and N. Sadr, "Sleep apnoea classification using heart rate variability, ECG derived respiration and cardiopulmonary coupling parameters," in *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, 16-20 Aug 2016, pp. 3203-3206.

[25] R. V. Sharan and T. J. Moir, "Pseudo-color cochleagram image feature and sequential feature selection for robust acoustic event recognition," *Applied Acoustics,* vol. 140, pp. 198-204, 2018.