ECG-Derived Heart Rate Variability Interpolation and 1-D Convolutional Neural Networks for Detecting Sleep Apnea

Roneel V. Sharan, *Senior Member, IEEE*, Shlomo Berkovsky, Hao Xiong, and Enrico Coiera

Abstract—Feature extraction from ECG-derived heart rate variability signal has shown to be useful in classifying sleep apnea. In earlier works, time-domain features, frequency-domain features, and a combination of the two have been used with classifiers such as logistic regression and support vector machines. However, more recently, deep learning techniques have outperformed these conventional feature engineering and classification techniques in various applications. This work explores the use of convolutional neural networks (CNN) for detecting sleep apnea segments. CNN is an image classification technique that has shown robust performance in various signal classification applications. In this work, we use it to classify one-dimensional heart rate variability signal, thereby utilizing a one-dimensional CNN (1-D CNN). The proposed technique resizes the raw heart rate variability data to a common dimension using cubic interpolation and uses it as a direct input to the 1-D CNN, without the need for feature extraction and selection. The performance of the method is evaluated on a dataset of 70 overnight ECG recordings, with 35 recordings used for training the model and 35 for testing. The proposed method achieves an accuracy of 88.23% (AUC=0.9453) in detecting sleep apnea epochs, outperforming several baseline techniques.

I. INTRODUCTION

Sleep apnea is a sleep disorder that occurs when breathing is interrupted during sleep. There are three main types of sleep apnea: obstructive sleep apnea (OSA), central sleep apnea (CSA), and mixed sleep apnea. OSA is the most common type sleep apnea and is characterized by partial (hypopnea) or full (apnea) obstruction of the upper airways during sleep which limits airflow to the lungs.

The apnea-hypopnea index (AHI), calculated as the number of apnea and hypopnea events per hour of sleep, is used to indicate the severity of sleep apnea. It is estimated that among 30-70 year old adults, 13% of men and 6% of women have moderate to severe sleep-disordered breathing (AHI $\geq$ 15) and 14% of men and 5% of women have AHI $\geq$ 5 with symptoms of daytime sleepiness [1]. In a study of 223 adults suspected of sleep-related breathing disorder, OSA was prevalent in 84% of the subjects [2].

Treatment of sleep apnea can reverse symptoms such as daytime sleepiness, improve cognitive performance and quality of life [3], and reduce cardiovascular risk [4]. Polysomnography is a type of sleep study used as a diagnostic tool to determine sleep disorders. This multi-parametric test is normally performed overnight and provides recordings of various physiological changes, such as brain activity, eye movement, heart rhythm (ECG), muscle activity, respiratory effort, nasal pressure, and blood oxygen saturation levels, during sleep. The multiple hours of polysomnography data are then analyzed subjectively to indicate the presence or absence of OSA.

The occurrence of apnea and hypopnea events leads to heart rate variations. The heart rate decreases during apnea and increases during recovery [5]. This characteristic behavior, heart rate variability (HRV), has been used to objectively detect apnea events using signal processing and machine learning techniques, such as in [6-8].

Techniques for HRV based apnea classification could be divided into time- and frequency-domain analyses of the HRV with frequency domain techniques generally producing better results [6]. Various classification methods have been utilized in literature with support vector machines (SVM) seen to be popular [7, 8]. In addition, feature combination is a commonly used technique for improving the classification performance. The most common technique for this is to concatenate multiple feature vectors into a single combined feature vector.

However, more recently, deep learning techniques have shown to outperform these complex feature extraction and selection methods in various tasks. This work explores the use of one such deep learning method, convolutional neural networks (CNN), for detecting apnea and non-apnea events. CNN has produced encouraging results in image classification tasks and has been increasingly applied in various signal classification applications. In this work, CNN is used for classification of the HRV signal computed from a single-lead ECG signal. In particular, it is proposed to resize the unevenly sampled one-dimensional HRV data to a common dimension using cubic interpolation and feed it directly into a one-dimensional CNN (1-D CNN), without any time- and/or frequency-domain feature extraction and without the need for feature selection.

Various techniques have been proposed for apnea detection using heart rate variability. However, it is difficult to directly compare the performance of these techniques unless the experimental setup is identical. For this reason, our paper utilizes the default preprocessing techniques applied to a publicly available dataset used in this work. The baseline techniques are then implemented using the same experimental setup for valid comparison. The performance of the proposed technique is evaluated against previously used time domain features, frequency domain features, a combined feature set, and logistic regression model (LRM) and SVM classifiers. The proposed method is seen to outperform these baseline methods in detecting apnea and non-apnea segments.

R. V. Sharan, S. Berkovsky, H. Xiong, and E. Coiera are with the Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2109, Australia (e-mail: roneel.sharan@mq.edu.au, shlomo.berkovsky@mq.edu.au, hao.xiong@mq.edu.au, and enrico.coiera@mq.edu.au).
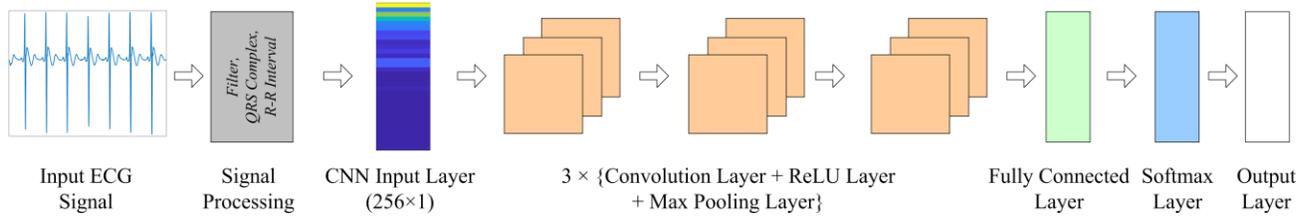
Figure 1. Overview of the proposed method.

The rest of this paper is organized as follows. Section II describes the dataset used in this work and overviews the proposed and baseline methods. The results are presented in Section III and discussions and conclusions are given in Section IV.

II. METHOD

A. Dataset

The dataset used in this work contains 70 overnight ECG recordings sampled at 100 Hz which vary in duration from 401 minutes to 578 minutes [9, 10]. Each 60 second epoch is annotated by human experts as having or not having apnea based on simultaneously recorded respiration and other related signals. The dataset is divided into a training set of 35 subjects and a test set of 35 subjects. The age of the subjects in both the training and test dataset varies from 27 to 63 years. There are 30 male subjects in the training dataset and 27 male subjects in the test dataset.

The recordings in this dataset are grouped into three classes. Class A (apnea) contains at least one hour with an AHI $\geq$ 10 and at least 100 minutes of apnea during the recording. Class B (borderline) contains at least one hour with AHI $\geq$ 5 and between 5 and 99 minutes with apnea during the recording. Class C (control) contains fewer than 5 minutes with apnea during the recording.

B. Proposed Method

An overview of the proposed ECG based apnea classification algorithm is shown in Fig. 1 and discussed in more detail in the following subsections.

Signal Processing

Preprocessed ECG signals, obtained from [10], were used in this work. Utilizing data from adjacent epochs in predicting the label of the current epoch has shown to be useful [11]. In this work, HRV computation for the ith epoch included the ECG signal data from the ith epoch and two immediately preceding and following epochs but the final class label was for the ith epoch only, as depicted in Fig. 2.

The QRS complex detection algorithm which is used to compute the R-R interval is based on the algorithm described in [12]. All 5 minute segments with artifacts were removed from the study. As a result, the training dataset contains 16,817 epochs and the test dataset - 16,996 epochs. The R-R interval varies, depending on the heart rate. This means HRV is an unevenly sampled data and the number of R-R intervals in each 5-minute segment is different. Since CNN requires a fixed size input data, the HRV data was resized to a dimension of 256 using cubic interpolation [13]. This resized data of size 256×1 forms the input layer of the 1-D CNN.
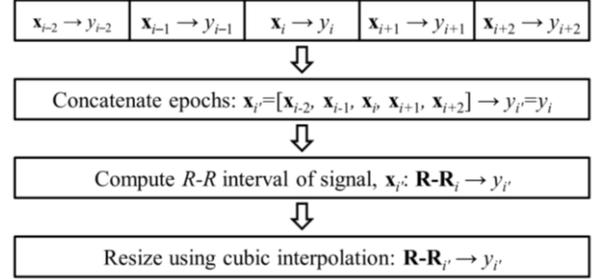

Figure 2. Overview of epoch combination for classifying the ith epoch; $\mathbf{x}_i$ represents the ECG signal in the ith epoch and y_i is the class label (apnea/non-apnea) of the ith epoch.

1-D Convolutional Neural Network (1-D CNN)

The layout of the CNN, shown in Fig. 1, was determined by several considerations and offline experiments. The HRV was computed from the single-lead ECG signal and resized to a dimension of 256 using cubic interpolation. The resulting 1-D vector or 1-D image-like representation forms the input layer of the CNN. The CNN model was trained using adaptive moment estimation [14]. The network consists of three convolution layers, each including a rectified linear unit (ReLU) [15] followed by a max pooling layer [16]. The filter size of all the convolution layers was 10×1 and stride 1×1. The number of filters in each layer was set to 64, 128, and 256, respectively. Similarly, the max pooling layer size was 2×1 and stride 2×1. This was followed by two fully connected layers of size 256 (followed by ReLU) and 2, respectively, a softmax layer [17], and an output layer.

The settings for other parameters were as follows: initial learn rate = 0.005, learn rate schedule = piecewise, learn rate drop factor = 0.6, learn rate drop period = 5, L2 regularization = 0.2, mini batch size = 1/10 of training data, data shuffle = every-epoch, and max epochs = 50. The parameters were optimized based on the training performance. The training stopped after the maximum number of epochs was reached.

C. Baseline Methods

Various time- and frequency- domain features are considered as baseline feature sets in this work. These are summarized in Tables I and II, respectively. All features were computed over 5 minute intervals.

The spectral energy was computed from the Lomb-Scargle periodogram [18]. Lomb-Scargle periodogram has shown to better estimate the power spectral density (PSD) of unevenly sampled data, such as HRV, then fast Fourier transform (FFT) based methods [19]. The Lomb-Scargle

638

TABLE I. SUMMARY OF TIME-DOMAIN FEATURES [20, 21]

Feature	Description
AVNN	Mean of 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
pNN20	Fraction of RR intervals that differ by more than 20ms

TABLE II. SUMMARY OF FREQUENCY-DOMAIN FEATURES [20, 21]

Feature	Description
ULF	Ultra low frequency (0.0001Hz ≥ ULF < 0.003Hz)
VLF	Very low frequency (0.003Hz ≥ VLF<0.04Hz)
LF	Low frequency (0.04Hz ≥ LF<0.15Hz)
HF	High frequency (0.15Hz ≥ HF<0.4Hz)
LF/HF	Ratio of LF and HF

periodogram was computed from the *R-R* intervals. A 1024 point periodogram was computed over 5 epochs (5 minute intervals), as depicted in Fig. 2, up to a frequency of 0.5 Hz from which the frequency domain features, four subband energies and subband energy ratio given in Table II, were computed.

The baseline classifiers used in this work are LRM and SVM with radial basis function (RBF) kernel.

D. Performance Evaluation

The performance of the proposed and baseline techniques is evaluated on the task of apnea and non-apnea epoch classification on the test dataset. The following metrics are used: accuracy, sensitivity, specificity, and area under the ROC curve (AUC).

III. RESULTS

The results for apnea and non-apnea epoch classification using the baseline and proposed methods are given in Table III. The classifier in each case was trained using the 35 training recordings. The results in Table III are for the predictions of the 35 test recordings. The optimal cut-off point on the ROC curve was determined as the point of maximum accuracy.

The results for the baseline features, time and frequency domain features as defined in Tables I and II, respectively, are given using LRM and SVM classification methods. The SVM results for both the baseline feature sets are better than LRM. Interestingly, with an accuracy of 78.15%, sensitivity of 66.02%, and specificity of 85.67%, the best baseline

TABLE III. RESULTS FOR APNEA AND NON-APNEA EPOCH CLASSIFICATION USING BASELINE AND PROPOSED TECHNIQUES

Feature Set	Feature Dim.	Classifier	Acc. (%)	Sens. (%)	Spec. (%)
Time-domain features (Table I)	10	LRM	70.00	58.53	77.11
	10	SVM	78.15	66.02	85.67
Frequency-domain features (Table II)	5	LRM	73.20	61.10	80.70
	5	SVM	74.56	52.05	88.50
Combined features (time + frequency)	15	LRM	75.19	63.31	82.54
	15	SVM	80.91	72.33	86.23
HRV (R-R Intervals)	256	CNN	88.23	82.74	91.62

results using a single feature set are achieved using time-domain features. The combination of time and frequency domain features is seen to further improve the classification performance for both the LRM and SVM classifiers. With an accuracy of 80.91%, sensitivity of 72.33%, and specificity of 86.23%, the best overall results are achieved using SVM.

However, with an accuracy of 88.23%, sensitivity of 82.74%, and specificity of 91.62%, the results using the proposed interpolated HRV and 1-D CNN are the highest. This technique does not utilize any feature combination and is seen to derive meaningful characteristics from the HRV signal without the need for feature engineering.

The ROC curve using the baseline and proposed methods are given in Fig. 3. AUC of 0.8673 is achieved for the best baseline technique, a combination of time and frequency domain features with SVM. The AUC increases to 0.9453 for the proposed interpolated HRV and 1-D CNN approach, which again shows the superiority of the proposed technique over the baseline methods. In addition, using the proposed method, we achieved an accuracy of 97.14%, sensitivity of 100%, and specificity of 93.33% in predicting apnea subjects (recordings) from the rest combined as per the criteria described in Section II-A.

IV. DISCUSSION AND CONCLUSION

This work proposes the use of interpolated ECG derived HRV as a direct input to a 1-D CNN for detecting sleep apnea epochs. With an accuracy of 88.23%, sensitivity of 82.74%, and specificity of 91.62%, the proposed method is seen to perform better than previously used time and frequency domain features and LRM and SVM classifiers.

Furthermore, in Table IV we compare our results against earlier studies which utilized the same dataset. Our method is seen to produce slightly better classification performance which strengthens its advantages. It should be noted, however, that the experimental setup of prior works may be slightly different from ours, such as different signal filtering and QRS detection techniques and different number of epochs. In [22, 23], conventional feature extraction, feature selection, and classification strategies have been utilized which produce lower accuracy than our approach. The closest method to our work is [11] but it utilizes a much larger CNN input, 900×2, and a two-dimensional (2-D) CNN, while still demonstrating slightly lower accuracy.
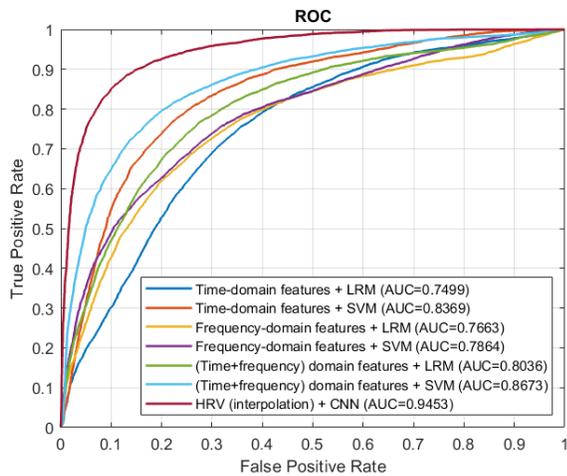
Figure 3. ROC curve for apnea and non-apnea epoch classification using baseline and proposed techniques.

TABLE IV.   SUMMARY OF RESULTS FROM SOME OTHER STUDIES UTILIZING THE SAME DATASET

Reference	Method	Acc. (%)	Sens. (%)	Spec. (%)
Varon et al. 2015 [22]	Feature extraction, feature selection, LS-SVM	84.74	84.71	84.69
Song et al. 2016 [23]	Feature extraction, feature selection, HMM+SVM	84.50	74.00	90.80
Wang et al. 2019 [11]	RR intervals + Amplitude (900×2) + CNN	87.60	83.10	90.30
This work	RR intervals (256×1) + 1-D CNN	88.23	82.74	91.62

The training dataset used in this work has only 35 subjects of which only 5 are female. Also, 20 subjects in the training dataset have apnea and only 10 are normal. Considering sleep apnea affects millions of people worldwide, a small imbalanced dataset may not produce generalizable models. Also, our method utilizes default preprocessing techniques as available in the dataset [10].

In the future, we plan to train and evaluate the proposed method on a larger dataset with more balanced gender and class representations. In addition, we plan to employ improved signal processing techniques, such as using an improved QRS complex detection algorithm and further preprocessing to remove artifacts, as shown to be useful in [24].

REFERENCES

[1] P. E. Peppard, T. Young, J. H. Barnet, M. Palta, E. W. Hagen, and K. M. Hla, "Increased prevalence of sleep-disordered breathing in adults," *American Journal of Epidemiology,* vol. 177, no. 9, pp. 1006-1014, 2013.

[2] T. I. Morgenthaler, V. Kagramanov, V. Hanak, and P. A. Decker, "Complex sleep apnea syndrome: Is it a unique clinical syndrome?," *Sleep,* vol. 29, no. 9, pp. 1203-1209, 2006.

[3] H. M. Engleman and N. J. Douglas, "Sleep · 4: Sleepiness, cognitive function, and quality of life in obstructive sleep apnoea/hypopnoea syndrome," *Thorax,* vol. 59, no. 7, pp. 618-622, 2004.

[4] N. J. Buchner, B. M. Sanner, J. Borgel, and L. C. Rump, "Continuous positive airway pressure treatment of mild to moderate obstructive sleep apnea reduces cardiovascular risk," *American Journal of Respiratory and Critical Care Medicine,* vol. 176, no. 12, pp. 1274-1280, 2007.

[5] C. Guilleminault, R. Winkle, S. Connolly, K. Melvin, and A. Tilkian, "Cyclical variation of the heart rate in sleep apnoea syndrome: Mechanisms, and usefulness of 24 h electrocardiography as a screening technique," *The Lancet,* vol. 323, no. 8369, pp. 126-131, 1984.

[6] T. Penzel, J. McNames, P. de Chazal, B. Raymond, A. Murray, and G. Moody, "Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings," *Medical and Biological Engineering and Computing,* vol. 40, no. 4, pp. 402-407, 2002.

[7] L. Almazaydeh, K. Elleithy, and M. Faezipour, "Obstructive sleep apnea detection using SVM-based classification of ECG signal features," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, USA, 2012, pp. 4938-4941.

[8] H. Sharma and K. K. Sharma, "An algorithm for sleep apnea detection from single-lead ECG using Hermite basis functions," *Computers in Biology and Medicine,* vol. 77, pp. 116-124, 2016.

[9] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, "The apnea-ECG database," in *Computers in Cardiology*, Cambridge, MA, USA, 2000, pp. 255-258.

[10] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark*, et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation,* vol. 101, no. 23, pp. e215-e220, 2000.

[11] T. Wang, C. Lu, G. Shen, and F. Hong, "Sleep apnea detection from a single-lead ECG signal with automatic feature-extraction through a modified LeNet-5 convolutional neural network," *PeerJ,* vol. 7, p. e7731, 2019.

[12] W. A. H. Engelse and C. Zeelenberg, "A single scan algorithm for QRS-detection and feature extraction," in *Computers in Cardiology*, Geneva, Switzerland, 1979, pp. 37-42.

[13] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 29, no. 6, pp. 1153-1160, 1981.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[15] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *27th International Conference on Machine Learning*, Haifa, Israel, 2010, pp. 807-814.

[16] K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?," in *IEEE 12th International Conference on Computer Vision*, Kyoto, Japan, 2009, pp. 2146-2153.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning.* New York: Springer, 2006.

[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] A. Delane, J. Bohórquez, S. Gupta, and M. Schiavenato, "Lomb algorithm versus fast Fourier transform in heart rate variability analyses of pain in premature infants," in *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, USA, 2016, pp. 944-947.

[20] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in Public Health,* vol. 5, pp. 1-17, 2017.

[21] G. D. Clifford, "Signal processing methods for heart rate variability," PhD thesis, Oxford University, UK, 2002.

[22] C. Varon, A. Caicedo, D. Testelmans, B. Buyse, and S. V. Huffel, "A novel algorithm for the automatic detection of sleep apnea from single-lead ECG," *IEEE Transactions on Biomedical Engineering,* vol. 62, no. 9, pp. 2269-2278, 2015.

[23] C. Song, K. Liu, X. Zhang, L. Chen, and X. Xian, "An obstructive sleep apnea detection approach using a discriminative Hidden Markov Model from ECG signals," *IEEE Transactions on Biomedical Engineering,* vol. 63, no. 7, pp. 1532-1542, 2016.

[24] J. N. McNames and A. M. Fraser, "Obstructive sleep apnea classification based on spectrogram patterns in the electrocardiogram," in *Computers in Cardiology*, Cambridge, MA, USA, 2000, pp. 749-752.