Human Interaction and Emerging Technologies (IHIET-AI 2025), Vol. 161, 2025, 478–484
https://doi.org/10.54941/ahfe1005940

Bidirectional Long Short-Term Memory (Bi-LSTM) With Convolutional Neural Networks (CNN) Based Obstructive Sleep Apnea Detection Using ECG Signals

Yinxian He[1], Amy M. Kwon[2], and Kyungtae Kang[2]

[1]Dept. of Computer Science and Engineering, Major in Bio Artificial Intelligence, Ansan, Gyeonggi-do 04763, Republic of Korea
[2]Dept. of Artificial Intelligence, Ansan, Gyeonggi-do 04763, Republic of Korea

ABSTRACT

Recently, advances in artificial intelligence (AI) have enhanced the performance of classification tasks in the medical domain, including signal analysis such as electrocardiogram (ECG). As a respiratory disorder, obstructive sleep apnea (OSA) presents distinct features in ECG signals that allow for detection. However, these patterns are typically irregular and highly variable, especially in the condition's early stages, which often leads to poor performance even with deep learning techniques. Based on temporal dependency and non-stationary features of ECGs, our study proposed a model that integrates Bi-directional Long Short-Term Memory (Bi-LSTM) with Convolutional Neural Networks (CNN) to classify OSA disorders by ECG recordings. By the experiment results, our approach achieved 88.68% accuracy, 86.94% sensitivity, 90.38% specificity, and an F1 score of 0.895. The results demonstrate superior performance compared to the model with traditional LSTM and the potential to detect OSA disorders by ECG signals. The model is also particularly beneficial for applications in home health monitoring.

Keywords: Deep learning, Electrocardiogram, Obstructive sleep apnea, Medical diagnoses

INTRODUCTION

Obstructive sleep apnea (OSA) is a condition involving repeated upper airway obstructions, either partial or complete, that occur while sleeping (Eckert and Malhotra, 2008). but the symptoms are often subtle and not easily recognizable (Ghassemi et al., 2018). Electrocardiogram (ECG) is a standardized procedure to diagnose OSA, which is a low-cost, non-invasive procedure corresponding to the heart's electrical activity. Because OSA causes cyclical changes in some characteristics, ECG signals capture the subtle differences between heart rate variability (HRV) and RR interval (RRI). However, ECG signals are noisy and exhibit significant variability in symptoms among individuals, making it difficult to detect the electrical abnormalities of OSA. Some studies have focused on extracting features such as frequency bands from power distribution (Babaeizadeh, 2010) or Q, R, and S waves from the QRS complex (Sharma and Sharma, 2016). Although these features can provide more distinctive patterns for artificial intelligence (AI) models, detecting OSA remains challenging, largely due to the high costs involved in current diagnostic methods with data collection and the diverse range of symptoms exhibited by individuals.

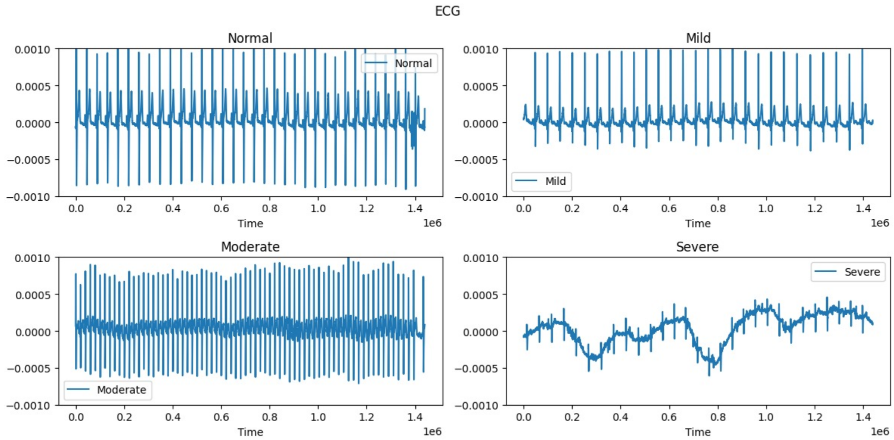

Figure 1: Example of specific difference between four severities in ECG.

However, ECG signals are typical time-series data, characterized by high temporal dependency and non-stationary features. This means that the features of ECG signals are not limited to individual time points but are closely related to the dynamic changes in the signal over time. Therefore, when analysing and processing ECG data, extracting and modeling temporal features is crucial. Based on this idea, Shoeb and Sridhar (2018) employed a hybrid model combining Long Short-Term Memory (LSTM) with Convolutional Neural Network (CNN) to capture latent temporal features in multiple physiological signals.

Bi-directional LSTM (Bi-LSTM) is a bi-directional recurrent neural network that extends LSTM by processing temporal data in both directions (Schuster and Paliwal, 1997). While LSTM processes information sequentially from past time steps, Bi-LSTM processes data in both the forward (signal front-to-back) and backward (signal back-to-front) directions, capturing more contextual dependencies and extracting richer features from the data.

Our study takes into account the temporal features embedded in ECG signals and accordingly combines Bi-LSTM with CNNs as our model. The proposed model has been validated in detecting OSA disorders, demonstrating its effectiveness in capturing the ECG signal's spatial and temporal characteristics. This hybrid model enhances the accuracy and reliability of OSA diagnosis, providing a more comprehensive approach compared to traditional methods, by ensuring computational efficiency, this method is also ideal for real-time applications.

Database and Preprocessing

We used the ECG dataset of the PSG-Audio dataset, which was simultaneously recorded with others by the given protocols at the Sismanoglio-Amalia Fleming General Hospital of Athen for the apnea study (Korompili et al., 2021). According to the protocols, 278 subjects participated in the study, which took approximately 5 hours per subject to record. The distribution of the disease severity was regarded as three based on the apnea-hypopnea index (AHI), guided by the National Institutes of Health (NIH) for OSA: mild, moderate, and severe, the differences are illustrated in Figure 1.

For the experiments, we excluded five duplicated records from the dataset, based on the lowest number of individuals at each severity level by gender, we selected a balanced sample of 72 subjects from a total dataset of 273 subjects, representing both genders and each severity level.

For segmentation, first we resampled the original 72 ECG records into 200 Hz. then we take the median of the OSA event with a duration greater than 20 seconds as the center to intercept samples with a length of 30 seconds and keep the original label unchanged. The normal samples are extracted as the same length of 30 seconds, same as apnea samples. 2382 samples were extracted, 1667 for training, and 715 for testing.

For noise removal, very high frequencies were removed using low-pass filters with cutoffs of 30 Hz, while very low frequencies were removed using high-pass filters with a cutoff frequency of 0.5 Hz (Zhou et al., 2020). The notch interference is filtered out from the ECG signal based on a frequency of 50 Hz (Charlton et al., 2016).

We employed Empirical Mode Decomposition (EMD) to process ECG, an adaptive decomposition method that can decompose a signal into several Intrinsic Mode Functions (IMFs) and a residual (Kwon and Kang, 2022). After the complex ECG signals are decomposed, the various frequency components and features for additional ECG processing will be extracted more effectively. The difference between the preprocessed signal and the original data is shown in Figure 2.

After the dimension reduction. The global standardization was applied to standardize all ECG samples, it differs from local standardization in that it uses the mean and standard deviation of the entire dataset, which were defined as follows:

$$z_i = \frac{x_i - \mu_{global}}{\sigma_{global}} \tag{1}$$

In this equation, x_i is the original data point, μ_{global} is the mean of the entire dataset, σ_{global} is the standard deviation of the entire dataset. After completing all the steps above, the data for each sample had a length of 6000.

Experiments

In this work, we proposed a network with five convolutional layers and one Bi-LSTM layer, followed by three fully connected layers to classify

OSA from normal. The model is implemented with the following parameter configuration:

- A stack of 1D CNNs with 32, 64, 128, 256, and 512 units. All convolution kernels' sizes were set to 3, to capture local features effectively.
- Each CNN layer is followed by batch normalization, which reduces the internal deviation of variable distributions and accelerates model convergence. The Rectified Linear Unit (ReLU) activation function is applied, which resolves the vanishing gradient problem and has a faster computation speed compared to sigmoid and tanh. Additionally, a max pooling process with a pool size of 2 is utilized to extract important features from the previous layer's output.

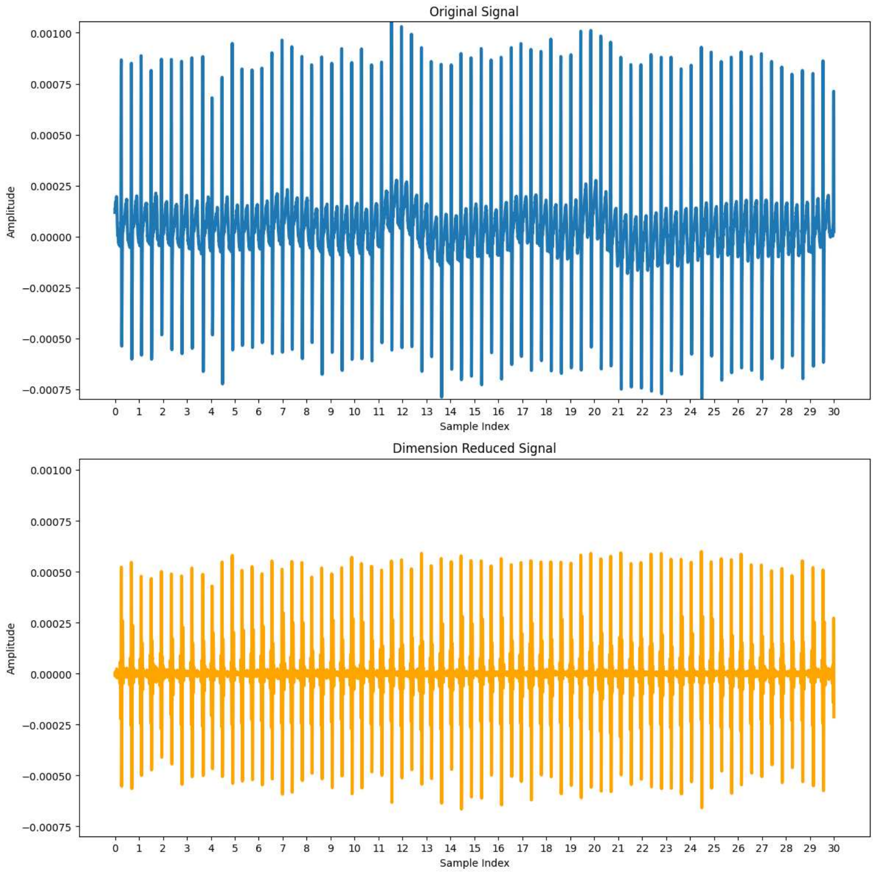

Figure 2: Comparison of ECG signals before and after empirical mode decomposition.

- After the five CNNs, the Bi-LSTM layer with 128 units is employed, followed by three dense layers with units of 128, 64, and 1. The final layer is used to classify and compute the probability for the normal and OSA categories.
- A dropout rate of 0.2 is applied after each CNN layer and the second dense layer. The dropout mechanism helps prevent overfitting during training.

The full architecture is shown in Figure 3. In the experiment, we used 715 pieces of samples in a test set to evaluate the performance of the Bi-LSTM and CNNs combined model. To optimize model performance, an adaptive learning rate was employed. Specifically, we monitor the validation loss and decrease the learning rate to 50% of its previous value if there is no progress in the validation loss for five consecutive epochs. The binary cross-entropy function was adopted with a batch size of 64, and the Adam optimizer was used.

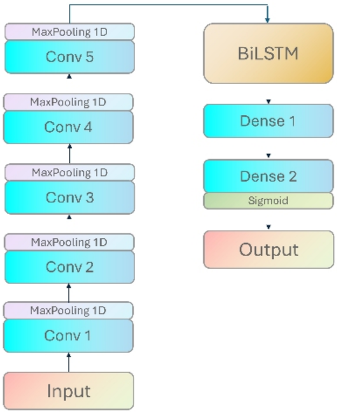

Figure 3: Architecture of the network.

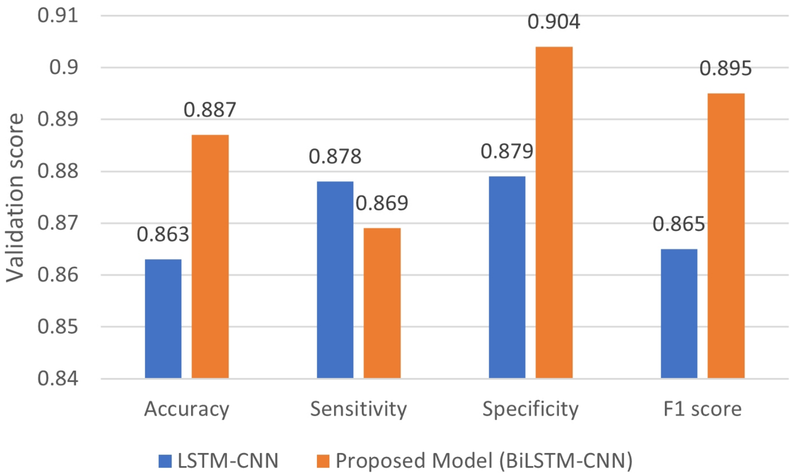

Figure 4: Distributions of validation scores for the proposed model and LSTM-CNN model.

As illustrated in Figure 4, the method we proposed achieved the performance with 88.68% accuracy, 86.94% sensitivity, 90.38% specificity, and a 0.895 F1 score. The overall performance of our proposed model is better than that of the LSTM-CNNs combined method. Based on the results, we conclude that this method effectively detects apnea from normal.

Compared to the LSTM-CNNs combined model, our model with Bi-LSTM enhances the model's performance and feature representation capabilities by simultaneously processing input sequences from past and future directions, allowing for better capture of contextual information. Compared to standard LSTM, Bi-LSTM exhibits superior performance in handling long-term dependencies and varying sequence lengths.

CONCLUSION

Our proposed model effectively integrates Bi-LSTM and CNN architectures to leverage the temporal features inherent in ECG signals. The model has proven to be highly effective and demonstrates great potential in detecting OSA symptoms from ECG signals. The model's architecture prioritizes the extraction of critical features making it suitable for real-time applications, which cannot be ignored in healthcare monitoring. Future studies will aim to enhance the performance of the model through further optimization such as utilizing the Bi-LSTM to extract more effective temporal features, while maintaining an optimized model complexity for practical use in real-time monitoring.

ACKNOWLEDGMENT

This work was partly supported by a grant from the Institute of Information & Communications Technology Planning & Evaluation (IITP), funded by the Korean government (MSIT) (No. RS-2022-00155885, Artificial Intelligence Convergence Innovation Human Resources Development Program at Hanyang University ERICA), and by the Global Research Support Program in the Digital Field (No. RS-2024-00431388).

REFERENCES

Babaeizadeh, S., White, D. P., Pittman, S. D., Zhou, S. H., 2010. Automatic detection and quantification of sleep apnea using heart rate variability. Journal of Electrocardiology 43(6), 535–541. Elsevier.

Charlton, P. H., Bonnici, T., Tarassenko, L., Clifton, D. A., Beale, R., Watkinson, P. J., 2016. An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram. Physiological Measurement 37(4), 610.

Eckert, D. J., Malhotra, A., 2008. Pathophysiology of adult obstructive sleep apnea. Proceedings of the American Thoracic Society 5(2), 144–153.

Ghassemi, M. M., Moody, B. E., Lehman, L.-W. H., Song, C., Li, Q., Sun, H., Mark, R. G., Westover, M. B., Clifford, G. D., 2018. You snooze, you win: The PhysioNet/Computing in Cardiology Challenge 2018. 2018 Computing in Cardiology Conference (CinC) 45, 1–4.

Korompili, G., Amfilochiou, A., Kokkalas, L., Mitilineos, S. A., Tatlas, N.-A., Kouvaras, M., Kastanakis, E., Maniou, C., Potirakis, S. M., 2021. PSG-Audio, a scored polysomnography dataset with simultaneous audio recordings for sleep apnea studies. Scientific Data 8(1), 197.

Kwon, A. M., Kang, K., 2022. A temporal dependency feature in lower dimension for lung sound signal classification. Scientific Reports 12(1), 7889. Nature Publishing Group UK London.

Schuster, M., Paliwal, K. K., 1997. Bi-directional recurrent neural networks. IEEE Transactions on Signal Processing 45(11), 2673–2681. IEEE.

Sharma, H., Sharma, K. K., 2016. An algorithm for sleep apnea detection from single-lead ECG using Hermite basis functions. Computers in Biology and Medicine 77, 116–124.

Shoeb, A., Sridhar, N., 2018. Evaluating convolutional and recurrent neural network architectures for respiratory-effort related arousal detection during sleep. 2018 Computing in Cardiology Conference (CinC) 45, 1–4. IEEE.

Zhou, G., Li, R., Zhang, S., Wang, J., Ma, J., 2020. Multimodal sleep signals-based automated sleep arousal detection. IEEE Access 8, 106157–106164. IEEE.