An Automatic Multi-Head Self-Attention Sleep Staging Method Using Single-Lead Electrocardiogram Signals

Songlu Lin, Yuzhe Wang , Zhihong Wang*

Instrument Science and Electrical Engineering, Jilin University, Changchun, China

Abstract

Automated sleep grading is an alternative to the time-consuming gold standard manual scoring system. Most existing methods use convolutional layers and recurrent-based architectures to build sleep-scoring systems. In this paper, we proposed a sleep stage classification method using the single-channel ECG, which is called ECG-SleepNet. After data preprocessing, we designed CNN layers, SE blocks, and LSTM layer for feature extraction, and multi-head attention mechanism for temporal context encoder, then we used weighted cross entropy as class-aware loss of the classifier. Finally, feedback regulation mechanism to perform the three classification of sleep stages – wake, non-rapid eye movement, and rapid eye movement. We conducted experiments on the Haaglanden Medisch Centrum sleep staging dataset which achieved overall accuracy of ***79.98%*** *and kappa score of 0.61. Our comparative experiments with similar studies showed that our model was superior to most other studies. The performance of our method also confirms that the features captured by multi-head self-attention and class-aware loss function are useful for sleep staging.*

1. Introduction

Sleep stage is important to both sleep professionals and researchers. According to the American Academy of Sleep Medicine (AASM), human sleep is divided into three stages: Wake (W), Non-Rapid Eye Movement (NREM), and Rapid Eye Movement (REM) [1]. Sleep staging has long relied on sleep specialists to perform this labor-intensive and time-consuming task manually [2]. More recently, many studies have used machine learning models to automatically classify ECG periods into appropriate stages, such as the Support Vector Machine (SVM) [3, 4], Random Forest (RF) [5-7], and XGBoost [8], etc. These methods use empirical mode decomposition or wavelet transform to design representative features for model classification. However, this technique requires a deep understanding of sleep-related functions, which may require attention and time [9]. Deep learning has become the state of the art in all fields, including automatic sleep staging [2, 8, 10]. The existing deep learning models for sleep staging are primarily composed of convolutional neural networks [11] (CNNs). For example, [12] used a CNN architecture to extract features from the raw data, and then exploited the long short-term memory (LSTM) to learn transition rules through sleep stages. In [13], heart rate variability (HRV) and respiratory signals were calculated and 25 features were extracted, and then the LSTM model was used for sleep stage classification. However, CNNs and RNNs also have limitations due to their recurrent nature, i.e., they usually have high model complexity, so it is difficult to train them in parallel, in addition, most of the existing research always need ECG experts to manually engineer features such as HRV, RR interval and R-peak amplitude (RA), and so on.

In addition to the selection of classification models, sleep staging must also address the problem of data imbalance since humans spend different amounts of time in each stage. Oversampling is a common strategy to deal with this problem. However, oversampling techniques expand the training data and thus increase the training time.

To address the above issues, we propose a novel deep learning framework that can automatically extract features from the raw ECG signals through the constructed ECG-SleepNet, which combines CNN layers, squeeze-and-excitation (SE) blocks, LSTM layer, and multi-head attention mechanism. In addition, we use cross-validation to enhance the stability of the model, and a class-aware loss function to handle the class imbalance without introducing additional computations, which outperforms the traditional cross-entropy (CE) loss function in both classification performance and handling imbalanced data.

2. Methodology and Experiments

2.1. Data source

The Haaglanden Medisch Centrum (HMC) sleep center in the Netherlands collected a diverse dataset of 151 whole-night polysomnographic (PSG) sleep recordings in 2018 [14]. This collection includes data from 85 male and 66 female participants with a mean age of 53.9 ± 15.4 years.

These patient recordings were randomly selected from a heterogeneous population referred to the HMC Sleep Center for PSG evaluation for a variety of sleep disorders. For our analysis, we use the ECG channel from the recordings, which were sampled at a frequency of 256 Hz.

2.2. Model and Experimental Design

2.2.1 Data preprocessing

For data preprocessing, our focus is limited to the ECG channel to critically evaluate the ECG-SleepNet model. The preprocessing protocols for all datasets involve segmenting the ECG data into 30-second epochs to ensure synchronization with the accompanying labels. Epochs which labels are "movement" and "unscored" are appropriately excluded. Following AASM guidelines, we assimilate the "N1", "N2", "N3", and "N4" categories into “NREM” category. To sharpen our analysis of distinct sleep cycles, our study only includes the 30 minutes of wakefulness that preceded each sleep session.

2.2.2 ECG-SleepNet Design

Fig. 1 illustrates the overall framework of our ECG-SleepNet model. It consists of three main blocks, namely, 1) feature extraction, 2) temporal context encoder and 3) classification.

In each convolutional layer of the CNN layers, there is a batch normalization layer, succeeded by an activation function in the form of a Gaussian error linear unit (GELU) [15]. Using the first convolutional layer as an illustration, Conv1D (64, 50, 1) is a one-dimensional convolutional layer indicating 64 channels, a kernel size of 50, and a stride of 1. Likewise, Maxpool1D (2, 2) is a max pooling layer specifying a kernel size of 2 with a stride of 2. Dropout layers with a dropout rate of 0.5 are used to mitigate network overfitting. The feature maps obtained from the CNN layers undergo subsequent processing in an LSTM layer with 128 units.

In each SE block, global mean pooling is first applied to the input feature map, resulting in a channel-wise descriptor. This descriptor is then processed through two fully connected layers, where the first layer reduces the dimensionality by a factor r (reduction ratio) and the second layer restores the original dimensionality. The output of the second fully connected layer is passed through a sigmoid activation function to produce a set of channel-wise weights. These weights are then used to scale the input feature map, enhancing the representation of important features while suppressing less useful features. This adaptive recalibration process allows the model to focus on the most informative features of the task at hand, thereby improving model performance.

(2) Temporal context encoder

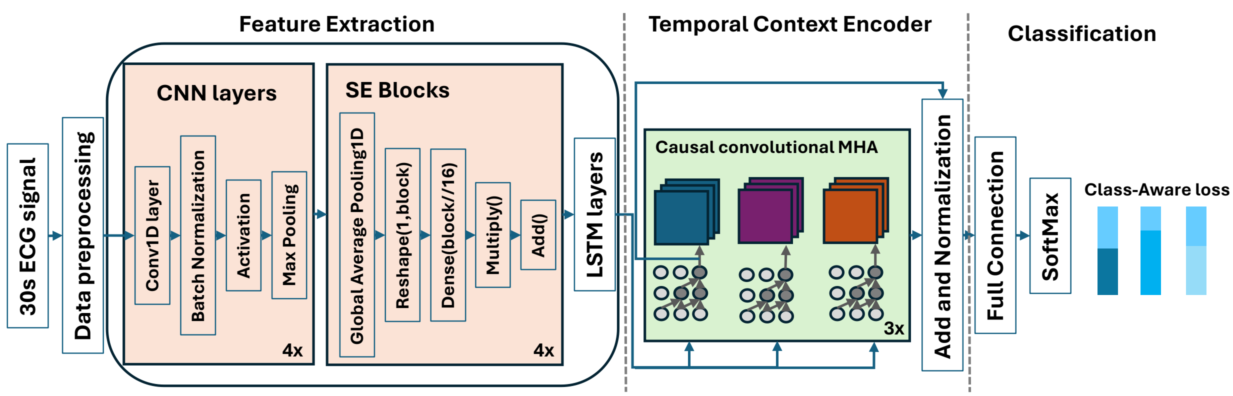


Figure 1. Overall framework of the proposed ECG-SleepNet model for automatic sleep stage classification.

First, the original ECG signals after data preprocessing, we designed a 4-convolutional-layer CNNs, SE Blocks and LSTM for feature extraction. Second, we develop a TCE module to capture the long-term dependencies in the input features. The core component of TCE is the multi-head attention supported by causal convolutions. Third, the classification decision is done by a fully connected layer with a SoftMax activation function. We also leverage a class-aware cost-sensitive loss function to handle the data imbalance issue.

(1) Feature Extraction

MHA improves the self-attention in two main aspects. First, it expands the model’s capability to focus on different positions, as the encoding of each head knows about the encodings of the other heads as well. This improves the model ability to learn temporal dependencies. Second, splitting the input features into different partitions increases the representation subspaces. Therefore, the generated attention weights for each subspace are more representative to the importance of each partition, and concatenating these representations produces better overall representation, which enhances the classification accuracy. In our ECG-SleepNet model, MHA leverages the causal convolutions to encode the positional information of input features and capture their temporal relations. The causal convolutions have an advantage of fast and parallel processing.

(3) Loss Function and Optimization

We use the weighted cross-entropy loss as the loss function of the model. The function below is the weighted cross-entropy loss (J_{WCE}):

$$J_{WCE} = W \times J_{CE}\ (y, p), \tag{1}$$

where $J_{CE}(y, p) =$

$$-\frac{1}{N}\{\sum_{n=1}^{N}\sum_{c=1}^{C}[y_n^c \log(\hat{y}_n^c) + (1 - y_n^c) \times \log(1 - \hat{y}_n^c)]\}. \tag{2}$$

In Eq. (2), J_{CE} is the traditional cross-entropy loss function, y represents the true label, and p represents the probability that the model predicts that the sample belongs to each category. N is the number of samples. y_n^c and $\hat{y}_n^c$ represent the real and the predicted label, respectively. W is a set of class weights.

To evaluate the effectiveness of our model, we use a subject-independent evaluation method. Specifically, we use k-fold cross-validation [16] for both training and testing. Given a dataset with N_d subjects, during each fold, the recordings of $N_d - (N_d/k)$ subjects serve as the training set, while recordings from (N_d/k) subjects are designed for testing. We have consistently set the value of k at 20. After repeating the process k times, the predicted results from all test sets across all folds are combined to provide a comprehensive assessment.

To enable parameter optimization, training, and evaluation of the deep learning model, we choose Python 3.10 as the programming language and TensorFlow 2.13.0 as the deep learning framework; moreover, two NVIDIA RTX A5000 GPUs are used to accelerate the computations. During the training process, we employ Early Stopping to prevent overfitting, setting the patience parameter to 5 to stop training and restore the best weights when the performance on the validation set no longer improved. We use the "ReduceLROnPlateau" method to adjust the learning rate during training. Finally, we use Adam [17] as the optimizer to minimize our class-aware loss and learn model parameters.

2.3. Evaluation Metrics

To assess the classification performance per class, we employ precision (PR), recall (RE), F1-score (F1), and G-mean (GM) as metrics. Moreover, accuracy (ACC), micro F1-score (MF1), Cohen's Kappa (κ) [18], and macro-averaged G-mean (MGm) are utilized to evaluate the overall classification performance. Assuming that the true positive, false positive, true negative, and false negative of i th class are represented by TP_i, FP_i, TN_i, and FN_i respectively.

3. Results and Discussions

3.1. Data analysis and pre-processing

In HMC dataset, It has 137243 epochs, including 13686 wake epochs, 92302 NREM epochs, and 21255 REM epochs. A comparison of the resulting overnight sleep statistics for both our ECG-SleepNet model and manual scoring is shown in Table 1. In this table we show the mean and standard deviation as calculated over the recordings. It can be observed that the estimated overnight sleep statistics using our proposed methodology are very close to those calculated from the human scorers, especially for the time spent in the N1 stage.

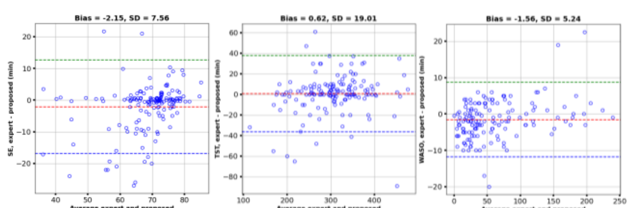
Table 1. Sleep measures summary assessed with manual annotation and ECG-SleepNet annotation (Means ± standard deviations).

Annotation	Manual	Automatic
TWT(min) *	78.4 ± 64.0	88.9 ± 59.7
TST (min) *	376.0 ± 87.5	394.6 ± 82.2
SE (%)*	82.8 ± 13.9	84.7 ± 16.6
WASO (min)	60.0 ± 53.6	67.2 ± 64.5

Abbreviations: TWT — total wake time; TST — total sleep time; SE — sleep efficiency; WASO — wake after sleep onset; W — wake; and R — rapid eye movement sleep stage.

* Due to we did the unified data processing, the TWT, TST, SE and WASO in four datasets are not representing the real-life scenario, but simply for comparison.

The results of the Bland-Altman graph, in Fig.2, showed that the sleep indices obtained using the manual scoring and the proposed method were similar in terms of the SE and WASO but considerably different in terms of the TST. Because body movement indirectly affects the assessment of the sleep indices, if a subject's movement during sleep was unusual (e.g., with low sleep efficiency), the WASO and TST could not be accurately reported using the proposed method, and further examinations would be required.



(a) (b) (c)

Fig. 2. Bland-Altman graphs for each sleep index in HMC dataset. (a) SE, (b) WASO, and (c) TST. The X axis and Y axis represent the mean of the measurements of the experts and those of the proposed method and the difference between the results of the manual and the automatic method, respectively.

3.2. Sleep Staging Performance

Shown in Table 2 is the evaluation result of the ECG-SleepNet model on the HMC dataset. The results include the confusion matrix, as well as the per-class evaluation metrics, including PR, RE, and F1. In our experiments, the rows of the confusion matrix represent the true labels, and the columns represent the predicted labels. The ECG-SleepNet model shows commendable performance in identifying the "W", "NR", and "R" sleep stages. And the overall performance of accuracy, MF1, kappa score and MGm are 79.98%, 88.87%, 0.61 and 74.32%, respectively.

Table 2. confusion matrix and per-class metrics of ECG-SleepNet sleep staging model on HMC dataset.

	Predicted			Per-class Metrics			
	W	NR	R	RE (%)	PR (%)	F1 (%)	GM (%)
W	17167	6431	88	72.48	62.02	66.84	81.10
NR	8478	83264	560	90.21	83.61	86.79	75.80
R	2037	9886	9332	43.90	93.51	59.75	66.08

4. Conclusion

In conclusion, we have presented a comprehensive approach to sleep stage classification using a deep learning model that combines the strengths of CNN, SE block, LSTM, and multi-head attention mechanism. Our model also has excellent performance on accuracy and Cohen's Kappa in terms of HMC dataset, which achieved 79.98% accuracy and 0.61 kappa score. The proposed ECG-SleepNet model effectively captures the complex patterns and features of raw ECG data to accurately classify sleep stages.

References

[1] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, and B. V. Vaughn, "The AASM manual for the scoring of sleep and associated events," *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine,* vol. 176, pp. 2012, 2012.

[2] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "Automatic sleep stage classification using single-channel eeg: Learning sequential features with attention-based recurrent neural networks." pp. 1452-1455.

[3] E. Alickovic, and A. Subasi, "Ensemble SVM method for automatic sleep stage classification," *IEEE Transactions on Instrumentation and Measurement,* vol. 67, no. 6, pp. 1258-1265, 2018.

[4] V. Gurrala, P. Yarlagadda, and P. Koppireddi, "A novel single channel EEG-based sleep stage classification using SVM," *International Journal of Biomedical Engineering and Technology,* vol. 36, no. 2, pp. 119-132, 2021.

[5] S. Santaji, S. Santaji, and V. Desai, "Automatic sleep stage classification with reduced epoch of EEG," *Evolutionary Intelligence*, pp. 1-8, 2022.

[6] M. Xiao, H. Yan, J. Song, Y. Yang, and X. Yang, "Sleep stages classification based on heart rate variability and random forest," *Biomedical Signal Processing and Control,* vol. 8, no. 6, pp. 624-633, 2013.

[7] T. L. da Silveira, A. J. Kozakevicius, and C. R. Rodrigues, "Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain," *Medical & biological engineering & computing,* vol. 55, pp. 343-352, 2017.

[8] L. Fiorillo, A. Puiatti, M. Papandrea, P.-L. Ratti, P. Favaro, C. Roth, P. Bargiotas, C. L. Bassetti, and F. D. Faraci, "Automated sleep scoring: A review of the latest approaches," *Sleep medicine reviews,* vol. 48, pp. 101204, 2019.

[9] V. Gerla, V. Kremen, M. Macas, D. Dudysova, A. Mladek, P. Sos, and L. Lhotska, "Iterative expert-in-the-loop classification of sleep PSG recordings using a hierarchical clustering," *Journal of neuroscience methods,* vol. 317, pp. 61-70, 2019.

[10] H. Phan, and K. Mikkelsen, "Automatic sleep staging of EEG signals: recent development, challenges, and future directions," *Physiological Measurement,* vol. 43, no. 4, pp. 04TR01, 2022.

[11] Z. Mousavi, T. Y. Rezaii, S. Sheykhivand, A. Farzamnia, and S. Razavi, "Deep convolutional neural network for classification of sleep stages from single-channel EEG signals," *Journal of neuroscience methods,* vol. 324, pp. 108312, 2019.

[12] J. Zhang, Z. Tang, J. Gao, L. Lin, Z. Liu, H. Wu, F. Liu, and R. Yao, "Automatic detection of obstructive sleep apnea events using a deep CNN-LSTM model," *Computational intelligence and neuroscience,* vol. 2021, 2021.

[13] Y. Wei, X. Qi, H. Wang, Z. Liu, G. Wang, and X. Yan, "A multi-class automatic sleep staging method based on long short-term memory network using single-lead electrocardiogram signals," *IEEE Access,* vol. 7, pp. 85959-85970, 2019.

[14] D. Alvarez-Estevez, and R. Rijsman, "Haaglanden medisch centrum sleep staging database (version 1.0. 1)," *PhysioNet*, 2021.

[15] D. Hendrycks, and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.

[16] Y. Bengio, and Y. Grandvalet, "No unbiased estimator of the variance of k-fold cross-validation," *Advances in Neural Information Processing Systems,* vol. 16, 2003.

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

[18] T. O. Kvålseth, "Note on Cohen's kappa," *Psychological reports,* vol. 65, no. 1, pp. 223-226, 1989.

Address for correspondence:

Zhihong Wang
#938 West Minzhu Street, Jilin University, Changchun, China.
Email: wzh@jlu.edu.cn