Automatic Classification of Sleep Stage from an ECG Signal Using a Gated-Recurrent Unit

Urtnasan Erdenebayar[1]* , Yeewoong Kim[1]* , Joung-Uk Park[1] , SooYong Lee[2], and Kyoung-Joung Lee[1]

[1] Department of Biomedical Engineering, College of Health Science, Yonsei University, Wonju, Korea
[2] Department of Liberal Education, Yonsei University, Wonju, Korea

ljfis

Abstract

A healthy sleep structure is clinically very important for overall health. The sleep structure can be represented by the percentage of different sleep stages during the total sleep time. In this study, we proposed a method for automatic classification of sleep stages from an electrocardiogram (ECG) signal using a gated-recurrent unit (GRU). The proposed method performed multiclass classification for three-class sleep stages such as awake, light, and deep sleep. A deep structured GRU was used in the proposed method, which is a common recurrent neural network. The proposed deep learning (SleepGRU) model consists of a 5-layer GRU and is optimized by batch-normalization, dropout, and Adam update rules. The ECG signal was recorded during nocturnal polysomnography from 112 subjects, and was normalized and segmented into units of 30-second duration. To train and evaluate the proposed method, the training set consisted of 80,316 segments from 89 subjects, and the test set used 20,079 segments from 23 subjects. We achieved good performances with an overall accuracy of 80.43% and F1-score of 80.07% for the test set. The proposed method can be an alternative and useful tool for sleep monitoring and sleep screening, which have previously been manually evaluated by a sleep technician or sleep expert.

Keywords: Sleep stage classification, Gated-recurrent unit, Deep learning, Electrocardiogram

1.   Introduction

On average, humans spend one-third of their lives sleeping. Sleep has a complex structure and cyclic rhythm that can define its quality and efficiency. The phases of sleep consist of wake, stages 1-4, and rapid eye movement (REM) sleep. Stages 3 and 4 constitute the slow-wave sleep (SWS) also known as delta sleep, and Stages 1-4 combine to form the non-rapid eye movement (NREM) stages [1]. During sleep, the brain organizes learned contents, eliminates toxins, and recharges energy [2]. Sleep disorders have a number of negative effects such as daytime sleepiness [3], headaches [4], cardiovascular diseases [5], decreased cognitive function [6], and decreased immunity [7]. The prevalence of sleep disorders, such as insomnia, sleep fragmentation, and sleep apnea, is increasing. Therefore, it is necessary to diagnose these problems properly through systematic sleep analysis [8].

Polysomnography (PSG) is the gold standard diagnostic test for sleep structure and sleep fragmentation. The patients attach equipment to their bodies to measure various bio-signals,

such as electroencephalogram (EEG), electrooculogram (EOG), electrocardiogram (ECG), and electromyogram (EMG). They then sleep at the sleep center. Based on the bio-signals obtained from PSG, not only can sleep fragmentation and structure be identified, but sleep disorders can also be diagnosed objectively. Among the electrophysiological measurements, EEG and EOG can be used to assess sleep stages.

ECG is an alternative bio-signal for automatic identification of sleep stages in home healthcare. Recently, various studies have proposed different methods for automatic classification of sleep stages using ECG signals. Initially, these studies extracted intermediate vital signs such as heart rate (HR), beat-to-beat (RR) interval, and heart rate variability (HRV) from the raw ECG signal. Adnane et al. [9] proposed a method that can classify sleep and wake, and also calculate sleep efficiency using a detrended fluctuation analysis of the HRV. Xiao et al. [10] proposed an alternative method for sleep stage classification based on a random forest using the HRV signal. Singh et al. [11] investigated a method for differentiating REM from NREM sleep using the RR interval. Yucelbas et al. [12] identified wake, REM, and NREM sleep based on nonlinear and morphological feature sets extracted from ECG signals. However, in all these studies, it was necessary to obtain intermediate vital signs from the ECG signal, including HR, HRV, and RR interval, and extract a number of features by using high-dimensional domain transformation. Furthermore, it was also necessary to select high-discriminative features to reduce the number of features for classifier training.

In this study, we proposed a novel method for automatic classification of sleep stages from a single-lead ECG signal using a deep gated-recurrent unit. We called the proposed method a SleepGRU model because the constructed model was designed and optimized by a gated-recurrent unit (GRU) to analyze the complex structure and cyclic rhythm of human sleep. The single-lead ECG was used without extracting any of the abovementioned intermediate vital signs required in similar studies, or using any hand-crafted features. The SleepRGU model was constructed, trained, and evaluated using clinical PSG datasets obtained from normal subjects and sleep apnea patients.

2. Material and Method

2.1 Participants and Datasets

For this study, 112 participants were enrolled to obtain the ECG dataset for automatic classification of sleep stages. The

Table 1. Information of the participants

	Training set	Test set	Total
Sex	89	23	112
Male	55	16	71
Female	34	7	41
Age (yr)	53.1±10.5	55.1±10.4	53.5±10.5
BMI (kg/m^2)	24.2±3.0	24.2±3.0	24.2±3.0
AHI (/hr)	10.6±9.5	10.0±8.5	10.4±9.3
TST (hr)	6.2±1.0	6.2±0.7	6.2±0.9
SE (%)	84.8±13.1	86.0±8.7	85.1±12.3

BMI, body mass index; AHI, Apnea-Hypopnea Index; TST, total sleep time; SE, sleep efficiency.

participants consisted of 52 control (Apnea-Hypopnea Index [AHI],: 2.3 ± 2.2) and 60 apnea (AHI, 17.5 ± 6.8) subjects. Subjects with severe sleep apnea were excluded from this study. The ECG dataset comprised the nocturnal PSG recordings of all participants. All the PSG recordings were obtained using a polygraphic amplifier (Model N7000; Embla, Reykjavik, Iceland) over an average period of 7.4 hour at the Samsung Medical Center, Seoul, Korea (Table 1). The ECG signals were recorded at a sampling rate of 200 Hz and segmented into episodes of 30 seconds. The sleep stages were labeled according to the criteria of the American Association of Sleep Medicine (AASM) [13]. All PSG procedures were approved by the Institutional Review Board of Samsung Medical Center (No. 2012-01-063).

The ECG signal of the subjects in each group was collected using a single-lead transducer and resampled at 100 Hz, resulting in 3,000 samples per episode. A total of 100,395 episodes were obtained after combining the signals obtained from all participants. The training and test datasets for the constructed SleepGRU model were randomly selected from each subject group (Table 1). The training set consisted of 80,316 episodes from 89 subjects (control 42, apnea 47), and the test set comprised 20,079 episodes from 23 subjects (control 10, apnea 13).

2.2 SleepGRU Model

We constructed a SleepGRU model taking the characteristics of human sleep, such as complex structure and cyclic rhythm, into consideration. Thus, the SleepGRU model consists of 5-layer GRU. The GRU is a very robust gating mechanism used in recurrent neural networks [14]. For the optimization of the SleepGRU model, batch normalization [15], dropout [16], and

rectified linear unit (ReLU) [17] were selected following trials of multiple optimization methods.

The GRU was introduced by Chung et al. [14] as a more efficient mechanism than the long short-term memory model. A GRU only has two gates: an update gate z and a reset gate r. The reset gate can capture short-term dependencies in the sleep stages, whereas the update gate captures long-term dependencies. The gating mechanism of the GRU is expressed as follows:

$$z_t = g(W_z \cdot [h_{t-1}, x_t]), \qquad (1)$$

$$r_t = g(W_r \cdot [h_{t-1}, x_t]), \qquad (2)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t \cdot h_{t-1}, x_t]), \qquad (3)$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t, \qquad (4)$$

where z, r, and h are, respectively, the update gate, the reset gate, and cell activation vectors, all of which are the same size as vector h, which defines the hidden value. Terms σ and τ represent nonlinear and hyperbolic tangent functions, respectively. Term x_t is the input to the memory cell layer at time t.

2.3 Model Optimization

Batch normalization was applied to the input ECG signal before training the constructed SleepGRU model, as shown in Eq. (5):

$$x_b = \alpha \cdot \left(\frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \right) + \beta, \qquad (5)$$

where ε is a small random noise, μ is the mini-batch mean, σ is the mini-batch variance, α is a scale parameter, and β is a shift parameter. Both α and β are trainable and updated in an epoch-wise manner [15].

Dropout is a technique that refers to randomly eliminating nodes to reduce overfitting in the network model by preventing complex adaptations on the training data [16].

A ReLU was used as the activation function of each layer of the SleepGRU model. It can be represented as follows:

$$f(x) = \max(0, wx + b), \qquad (6)$$

where x represents the feature map, w is the weight, and b denotes the bias. The ReLU demonstrates robust training performance and produces consistent gradients that aid gradient-based learning [17].

Table 2. Detailed structure of the constructed SleepGRU model

No	Layers	Units	Dropout	Feature maps	Params
1	bNorm	-	-	3000×1	4
2	GRU1	24	0.5	3000×24	1,920
3	GRU2	20	0.5	3000×20	2,760
4	GRU3	16	0.4	3000×16	1,824
5	GRU4	8	0.3	3000×8	642
6	GRU5	4	0.2	3000×4	168
7	MLP0	3	-	15×3	45

2.4 Model Structure

Table 2 shows the detailed characteristics of the final architecture of the constructed SleepGRU model. The SleepGRU model is a structure consisting of a 5-layer GRU. Each recurrent layer contains (24, 20, 16, 8, 4) hidden nodes in the GRU. Finally, we used a fully connected multilayer perceptron with softmax activation for the final discrimination of the sleep stage classification from the ECG datasets.

2.5 Model Implementation

For this study, the PSG data were processed using MATLAB (R2018b). The SleepGRU model was constructed using the Keras library with a TensorFlow backend [18]. A workstation with an Intel CPU (i9-9900X @3.5GHz) and NVIDIA GPU (GeForce RTX 2080 Ti) was used for deep learning.

2.6 Model Evaluation

The F1-score was used to evaluate the constructed SleepGRU model, which evaluates the classification accuracy of each class according to class equality. To obtain the F1-score, two evaluation measures, precision and recall, were combined. These are defined as follows:

$$precision = \frac{TP}{TP + FP}, \qquad (7)$$

$$recall = \frac{TP}{TP + FN}, \qquad (8)$$

where TP, FP, and FN represent the true positives, false positives, and false negatives, respectively. These values were determined for each sleep stage event. The F1-score, better known as the unbalanced data set, is computed based on the

Table 3. The performance of the SleepGRU model

Datasets	Classes	Precision	Recall	F1	Accuracy (%)
Training set	Wake	0.86	0.68	0.76	
	REM	0.88	0.67	0.77	84.01
	NREM	0.86	0.90	0.89	
Test set	Wake	0.58	0.42	0.51	
	REM	0.77	0.81	0.73	80.43
	NREM	0.86	0.84	0.83	

sample proportion of precision and recall as follows:

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}. \tag{9}$$

3. Experimental Results

The results of the SleepGRU model for the automatic classification of the sleep stage based on the single-lead ECG signal are shown in Table 3. The SleepGRU model was evaluated based on precision, recall, F1-score, and accuracy for multiclass sleep stage classification comprising wake, REM, and NREM. For this study, sleep stages N1, N2, and N3 were integrated into the NREM stage.

For the automatic classification of sleep stages (Table 3), the SleepGRU model demonstrated robust performance with an accuracy of 84.01% for the training set and 80.43% for the test set. In particular, the SleepGRU model was most effective in the NREM stage and least effective in the wake stage. Although the REM stage had the largest number of events, the SleepGRU model did not demonstrate remarkable accuracy for this stage.

4. Discussion

In this study, a SleepGRU model was constructed for the automatic classification of sleep stages using a single-lead ECG signal. To ensure that the model was well suited to the complex characteristics of human sleep, the GRU was integrated into the SleepGRU model. We obtained a robust performance with a classification accuracy of 80.43% for the three-class sleep stages using the single-lead ECG signal. In addition, the Sleep-GRU model was evaluated using the ECG dataset obtained from the control and sleep apnea groups.

Table 4 compares and analyzes the existing studies that perform sleep stage scoring using either the ECG signal or the features obtained from the HRV and RR intervals. Most of these studies used a shallow learning classifier as the SVM,

rather than deep learning models; many of the features were extracted from the ECG signal. In addition, binary (sleep or wake) and multiclass (wake, REM, and NREM) classification was performed using extracted hand-crafted feature sets. First, intermediate vital signs were extracted, including the HR [9], HRV [10], and RR interval [11]. Then, a number of features were extracted by analyzing them in the time, frequency, and nonlinear domain [12]. Finally, the top features were selected to reduce the number of features for training the random forest classifier. However, in these studies, the results cannot be generalized, and nonlinear features are required for complex calculations.

Some studies have been based on the deep learning framework for automatic sleep stage scoring using the ECG signal [19–21]. In these studies, deep learning frameworks (DNN [19], CNN [20], RNN [20], and LSTM [21]) were used as feature extractors or classifiers in previous studies based on the ECG signal. In addition, they demonstrated less accuracy in the three-class sleep stage classification than the current study.

We therefore constructed a SleepGRU model that can perform automatic classification of the sleep stage based on sequential feature extraction and discrimination using the ECG signal. The constructed SleepGRU model outperforms conventional methods because it considers the complex and cyclic characteristics of sleep. Another advantage of the SleepGRU model for sleep stage scoring is that it does not require the intermediate biosignals (HRV, RR interval, and ECG derived respiration) or any hand-crafted feature sets extracted through domain transformation analysis. Furthermore, the SleepGRU model has a simpler structure than the other deep learning models designed for sleep stage scoring. In addition, it was trained and tested on a clinical dataset consisting of normal and sleep apnea groups.

Table 4. Comparison with previous studies

Author	Signal	Classes	Accuracy (%)
SVM [9]	HRV	2	79.9
Randon Forest [10]	HRV	3	72.5
SVM [11]	RR interval	2	72.8
Random Forest [12]	ECG	3	78.0
DNN [19]	ECG	3	77.8
CNN [20]	ECG	3	73.0
RNN [21]	HR, Activity	3	66.6
GRU	ECG	3	80.4

Our results demonstrated that the constructed SleepGRU model has the potential to accurately perform multiclass classification using only a single-lead ECG signal.

However, this study has some limitations. First, we used a small dataset for the model evaluation. Larger and more diverse datasets are needed. Second, we only covered sleep apnea patients in this study, but some other datasets from common sleep disorders should be covered for a more robust model. Finally, the constructed SleepGRU model has a higher computational cost than conventional methods. These are the challenges that should be addressed in future studies.

5. Conclusion

In this study, a SleepGRU model was constructed for the automatic classification of sleep stages based on a single-lead ECG signal. In most ECG-based studies on sleep stage scoring, classification was performed using a binary class, but the SleepGRU model can perform multiclass classification, as it demonstrates efficiency for the three-class sleep stages. In addition, the Sleep-GRU model can automatically extract the feature maps and classify the sleep stages simultaneously based on the ECG signal. We achieved an overall high accuracy of 80.43% for the three-class sleep stages (Wake, REM, and NREM). Therefore, the SleepGRU model is suitable for sleep stage classification using a single-lead ECG signal without any feature extraction. In future research, the SleepGRU model based on the single-lead ECG signal should be validated using larger and more diverse datasets.

Conflict of Interest

No potential conflicts of interest relevant to this article were reported.

Acknowledgements

References

[1] A. Rechtschaffen and A. Kales, *A Manual for Standardized Terminology, Techniques and Scoring System for Sleep Stages in Human Subjects*. Bethesda, MD: US Department of Health, Education, and Welfare, 1968.

[2] P. Peigneux, S. Laureys, X. Delbeuck, and P. Maquet, "Sleeping brain, learning brain: the role of sleep for memory systems," *Neuroreport*, vol. 12, no. 18, pp. A111-A124, 2001. https://doi.org/10.1097/00001756-200112210-00001

[3] E. Stepanski, J. Lamphere, P. Badia, F. Zorick, and T. Roth, "Sleep fragmentation and daytime sleepiness," *Sleep*, vol. 7, no. 1, pp. 18-26, 1984. https://doi.org/10.1093/sleep/7.1.18

[4] P. Jennum and R. Jensen, "Sleep and headache," *Sleep Medicine Reviews*, vol. 6, no. 6, pp. 471-479, 2002. https://doi.org/10.1053/smrv.2001.0223

[5] M. P. Hoevenaar-Blom, A. M. Spijkerman, D. Kromhout, J. F. van den Berg, and W. M. Verschuren, "Sleep duration and sleep quality in relation to 12-year cardiovascular disease incidence: the MORGEN study," *Sleep*, vol. 34, no. 11, pp. 1487-1492, 2011. https://doi.org/10.5665/sleep.1382

[6] G. Curcio, M. Ferrara, and L. De Gennaro, "Sleep loss, learning capacity and academic performance," *Sleep Medicine Reviews*, vol. 10, no. 5, pp. 323-337, 2006. https://doi.org/10.1016/j.smrv.2005.11.001

[7] C. E. Gamaldo, A. K. Shaikh, and J. C. McArthur, "The sleep-immunity relationship," *Neurologic Clinics*, vol. 30, no. 4, pp. 1313-1343, 2012. https://doi.org/10.1016/j.ncl.2012.08.007

[8] C. A. Kushida, A. Chang, C. Gadkary, C. Guilleminault, O. Carrillo, and W. C. Dement, "Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients," *Sleep Medicine*, vol. 2, no. 5, pp. 389-396, 2001. https://doi.org/10.1016/s1389-9457(00)00098-8

[9] 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. https://doi.org/10.1016/j.eswa.2011.08.022

[10] 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. https://doi.org/10.1016/j.bspc.2013.06.001

[11] J. Singh, R. K. Sharma, and A. K. Gupta, "A method of REM-NREM sleep distinction using ECG signal for unobtrusive personal monitoring," *Computers in Biology and Medicine*, vol. 78, pp. 138-143, 2016. https://doi.org/10.1016/j.compbiomed.2016.09.018

[12] S. Yucelbas, C. Yucelbas, G. Tezel, S. Ozsen, and S. Yosunkaya, "Automatic sleep staging based on SVD, VMD, HHT and morphological features of single-lead ECG signal," *Expert Systems with Applications*, vol. 102, pp. 193-206, 2018. https://doi.org/10.1016/j.eswa.2018.02.034

[13] 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, IL: The American Academy of Sleep Medicine, 2012.

[14] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, Available https://arxiv.org/abs/1412.3555.

[15] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of Machine Learning Research (PMLR)*, vol. 37, pp. 448-456, 2015.

[16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929-1958, 2014.

[17] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML)*, Haifa, Israel, 2010, pp. 807-814.

[18] Keras API, Available https://keras.io/.

[19] 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. https://doi.org/10.1007/s13534-017-0044-1

[20] QQ. Li, Q. Li, C. Liu, S. P. Shashikumar, S. Nemati, and G. D. Clifford, "Deep learning in the cross-time frequency domain for sleep staging from a single-lead electrocardiogram," *Physiological Measurement*, vol. 39, no. 12, article no. 124005, 2018. https://doi.org/10.1088/1361-6579/aaf339

[21] X. Zhang, W. Kou, I. Eric, C. Chang, H. Gao, Y. Fan, and Y. Xu, "Sleep stage classification based on multi-level feature learning and recurrent neural networks via wearable device," *Computers in Biology and Medicine*, vol. 103, pp. 71-81, 2018. https://doi.org/10.1016/j.compbiomed.2018.10.010

Urtnasan Erdenebayar received his B.S. in Computer Science from Huree University, Ulaanbaatar, Mongolia, in 2007 and M.S. in Electronic Engineering from Inha University, Incheon, Korea, in 2010, respectively. He also received his Ph.D. in Biomedical Engineering from Yonsei University, Seoul, Korea, in 2018. Since 2018, he is a Postdoc researcher at the Department of Biomedical Engineering, Yonsei University. His current research interests are artificial intelligence, deep learning, machine learning, digital healthcare, digital medicine, data science, and biosignal processing.

Yeewoong Kim received his B.S. and M.S. in Biomedical Engineering from Yonsei University, Wonju, Korea, in 1995 and 2002, respectively. He is currently a Ph.D. candidate at the Department of Biomedical Engineering from Yonsei University. He has been working on research related to sleep signal analysis, algorithm development, and signal processing.

Joung-Uk Park received his B.S. in Biomedical Engineering from Konyang University, Daejeon, Korea, in 2008 and M.S. in Biomedical Engineering from Yonsei University, Seoul, Korea, in 2012. He is currently a Ph.D. candidate at the Department of Biomedical Engineering from Yonsei University. He has been working on research related to sleep signal analysis, algorithm development, and signal processing.

SooYong Lee received his Ph.D. in Mathematics from Kyunghee University, Seoul, Korea, in 1992. He also received his Ph. D. in Computer Science from Yonsei University, Seoul, Korea, in 2004. He is a faculty member at Yonsei University, Wonju, Korea. since 2004. He has been working on research related to artificial intelligence, machine learning, and data mining.

Kyoung-Joung Lee received his B.S., M.S., and Ph.D. in electric engineering from Yonsei University, Seoul, Korea, in 1981, 1983, and 1988, respectively. He was an international fellow at Case Western Reserve University, USA, in 1993. He joined Yonsei University, Wonju, Korea, as a faculty member in 1989. His research interests include medical instruments, biosignal processing, and biosystem modeling.