

Original Article
Biomedical Engineering



Identification of Sleep Apnea Severity Based on Deep Learning from a Short-term Normal ECG

Erdenebayar Urtnasan ,¹ Jong-Uk Park ,² Eun Yeon Joo ,³ and
Kyoung Joung Lee ²

¹Institute of AI and Big Data in Medicine, Wonju College of Medicine, Yonsei University, Wonju, Korea

²Department of Biomedical Engineering, College of Health Science, Yonsei University, Wonju, Korea

³Department of Neurology, Samsung Medical Center, School of Medicine, Sungkyunkwan University, Seoul, Korea



Received: Jan 28, 2020

Accepted: Sep 14, 2020

Address for Correspondence:

Kyoung Joung Lee, PhD

Department of Biomedical Engineering,
College of Health Science, Yonsei University, 1
Yonseidae-gil, Wonju 26493,
Republic of Korea.
E-mail: lkj5809@yonsei.ac.kr

© 2020 The Korean Academy of Medical Sciences.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs

Erdenebayar Urtnasan
<https://orcid.org/0000-0002-3493-9724>

Jong-Uk Park
<https://orcid.org/0000-0002-6358-5777>

Eun Yeon Joo
<https://orcid.org/0000-0003-1233-959X>

Kyoung Joung Lee
<https://orcid.org/0000-0002-2704-456X>

Funding

This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE) and the Korea Institute for Advancement of Technology (KIAT) through the National Innovation Cluster R&D Program (P0006697, Development of a Cardiopulmonary Monitoring System Using Wearable Device).

ABSTRACT

Background: This paper proposes a novel method for automatically identifying sleep apnea (SA) severity based on deep learning from a short-term normal electrocardiography (ECG) signal.

Methods: A convolutional neural network (CNN) was used as an identification model and implemented using a one-dimensional convolutional, pooling, and fully connected layer.

An optimal architecture is incorporated into the CNN model for the precise identification of SA severity. A total of 144 subjects were studied. The nocturnal single-lead ECG signal was collected, and the short-term normal ECG was extracted from them. The short-term normal ECG was segmented for a duration of 30 seconds and divided into two datasets for training and evaluation. The training set consists of 82,952 segments (66,360 training set, 16,592 validation set) from 117 subjects, while the test set has 20,738 segments from 27 subjects.

Results: F1-score of 98.0% was obtained from the test set. Mild and moderate SA can be identified with an accuracy of 99.0%.

Conclusion: The results showed the possibility of automatically identifying SA severity based on a short-term normal ECG signal.

Keywords: Automatic Prediction; Sleep Apnea; Short-term Normal ECG; Convolutional Neural Network; Deep Learning

INTRODUCTION

Sleep apnea (SA) is one of the most common sleep disorders, characterized by the presence of repeated apneas and hypopneas. SA is a respiratory event annotated by a complete cessation of airflow (i.e., decreasing amplitude of at least 90%) for at least 10 seconds and continued respiratory effort. A hypopnea is similar to an SA, although the decrease in airflow amplitude is 30%–90% during sleep.¹

Healthy and adequate sleep is essential for people of all ages, because sleep is a basic human physiological need that occupies one-third of a person's life. During sleep, the brain organizes learned contents, eliminates toxins, and the body recharges itself.² Therefore, the quality and quantity of sleep are very important for maintaining good health. However, SA

Disclosure

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Urtnasan E. Data curation: Urtnasan E, Joo EY. Formal analysis: Park JU. Software: Urtnasan E. Investigation: Lee KJ, Joo EY. Methodology: Urtnasan E, Park JU. Validation: Park JU, Joo EY. Writing - original draft: Urtnasan E. Writing - review & editing: Urtnasan E, Park JU, Joo EY, Lee KJ.

causes oxygen deficiency and disrupts the normal structure of the sleep.³ It is regarded as the main factor influencing sleep fragmentation and sleep efficiency. Furthermore, SA can lead to a variety of illnesses, including fatigue,⁴ headache,⁵ cardiovascular disease,⁶ and decreases in cognitive function⁷ and immunity.⁸ Accurate and advance prediction of SA is necessary for people with diagnosed or undiagnosed sleep-related breathing disorders.^{9,10}

Nocturnal laboratory-based polysomnography (PSG) is commonly used as a tool to diagnose sleep disorders and to prescribe sleep medication. PSG is a multi-signal recorded measurement of the brain, cardiopulmonary, and muscle activity via electroencephalography (EEG), electrooculography (EOG), electrocardiography (ECG), respiratory, pulse oximetry (SpO₂), and electromyography (EMG). PSG can provide an objective assessment of sleep disorders, including details on airflow, respiratory effort, and oximetry, and can be used to predict SA.¹¹ However, PSG is not user-friendly; inconvenience arises from the number of sensors required for multichannel recordings. PSG is also costly and requires trained attendants. Furthermore, manual interpretation of PSG recordings is cumbersome and labor-intensive, requiring licensed technicians.

ECG has been used as an alternative physiological signal source for detecting or classifying SA events since the early 2000s.¹² SA directly affects ECG through the response of the sympathetic nervous system during sleep. Because of sleep apnea there are some ECG morphology differences including very small variations in amplitude of the R peaks and baseline fluctuation due to respiration, and inter-beat interval decreasing.¹³ In addition, ECG is relatively easy to measure and can provide vital signs including heartbeat, beat-to-beat intervals, QRS complex, and ECG-derived respiration. Therefore, ECG, especially the single-lead signal, has been used in many studies for automatically detecting, classifying, and screening SA events based on machine learning techniques.

To date, studies have focused on finding and extracting the most informative features from the ECG signal to achieve better performance. For instance, Mendez et al.^{14,15} have proposed a method based on nonlinear and frequency domain analyses for SA screening. They extracted 20 features and applied 10 of them to four different classifiers. Chen et al.¹⁶ proposed a systematic route of the conventional signal processing algorithm for the automatic screening of SA events. This method requires only beat-to-beat intervals derived from the ECG signals of suspected SA patients. In addition, these studies not only analyzed and extracted features from normal ECG but apneic ECG signals also. These methods require the applied ECG signal to be only recorded at night. Additionally, the RR interval and heart rate variability (HRV) are derived from the raw ECG signal for at least 5 minutes.

The convolutional neural network (CNN) is one of the most common neural networks derived from the cognitive model of human vision.¹⁷ Recently, the CNN has drawn extensive interest because of its demonstrated efficiency as a detector and classifier in the fields of image recognition,¹⁸ computer vision,¹⁹ and biomedicine.²⁰ Several studies on the detection of SA events or the classification of apnea and hypopnea events have been conducted using the CNN model from a single-lead ECG. For instance, Dey et al.²¹ demonstrated a deep learning framework for SA detection based on the CNN algorithm using a single-lead ECG signal. In our previous study, we designed a CNN model to classify multiclass SA events such as apnea and hypopnea from a single-lead ECG.²² In these studies, the apneic segment or abnormal segments of the ECG signal were used for the automatic detection and classification of SA.

In this paper, we propose a novel method of automatically identifying SA severity based on a short-term normal ECG signal using a deep learning model. To this end, a CNN model with a deep structure was designed for identifying the severity of SA using the short-term normal ECG. Because of its automatic, high-dimension feature extraction, a deep CNN was selected as the automatic identifier for SA severity from a short-term normal ECG signal. Finally, we trained and evaluated the designed CNN model by using the diagnostic nocturnal PSG dataset containing the training and test sets.

METHODS

The proposed methodology is composed of four main parts: the short-term normal ECG datasets, input signal, deep learning model, and outputs (Fig. 1). A detailed description of each part is presented in subsequent sections.

Study population

A total of 144 subjects enrolled in the study: fifty-two in a normal group (male: 25, female: 27) and ninety-two in the OSA group (mild: 22, moderate: 38, and severe: 32) (Table 1). All subjects underwent an overnight PSG using a polysomnographic amplifier (Embla N7000; Natus, Kópavogur, Iceland). The test involved EEG, EOG, EMG, ECG, SpO₂, chest and abdomen respiration, nasal pressure, airflow, and snoring recordings. The average recording time was 7.2 hours, and mean sleep time was 5.7 hours.

The apnea-hypopnea index (AHI) is compiled based on the averaged frequency of apnea and hypopnea per hour of sleep. In accordance with the criteria proposed by the American Academy of Sleep Medicine (AASM),²³ all sleep events from the PSG recordings were manually annotated by licensed and experienced technicians.

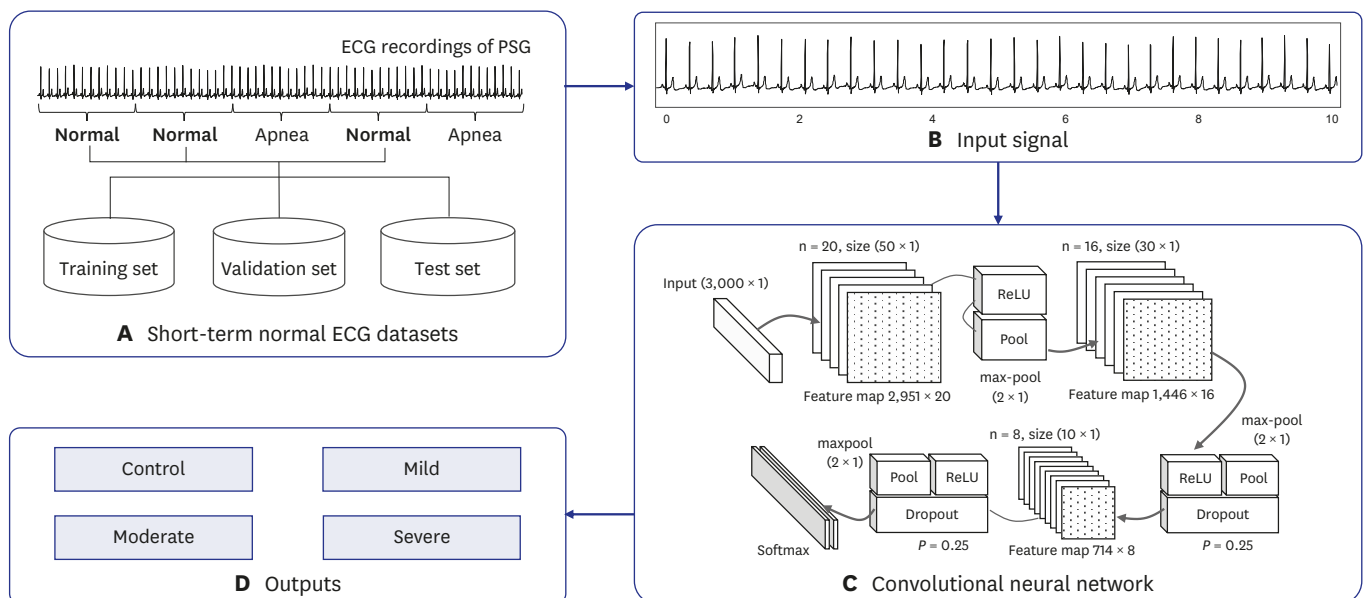


Fig. 1. Proposed method for SA severity identification based on deep learning using a short-term normal ECG. It consists of four main parts: (A) short-term normal ECG datasets, (B) input signal, (C) deep learning model, and (D) outputs. SA = sleep apnea, ECG = electroencephalography, PSG = polysomnography, ReLU = rectified linear unit.

Table 1. Participants' information in detail

Measures	Normal	Mild	Moderate	Severe	Total
Total					
Subjects (female:male)	52 (27:25)	22 (7:15)	38 (7:31)	32 (5:27)	144 (46:98)
Age, yr	47 ± 6	58 ± 14	59 ± 10	56 ± 10	54 ± 10
Body mass index, kg/m ²	22.6 ± 1.8	24.6 ± 2.3	26.2 ± 3.4	26.8 ± 3.0	24.8 ± 3.2
Apnea-hypopnea index, events/hr	2.3 ± 2.3	10.1 ± 2.8	21.8 ± 4.3	48.0 ± 13.8	18.8 ± 18.7
Sleep efficiency, %	90.1 ± 10.9	78.9 ± 12.9	81.7 ± 11.1	76.7 ± 14.2	83.2 ± 13.1
Training set					
Subjects (female:male)	42 (20:22)	18 (7:11)	31 (6:25)	26 (4:22)	117 (37:80)
Age, yr	48 ± 6	61 ± 13	59 ± 10	56 ± 11	55 ± 11
Body mass index, kg/m ²	22.7 ± 1.9	24.4 ± 2.3	26.5 ± 3.6	26.4 ± 2.9	24.8 ± 3.2
Apnea-hypopnea index, events/hr	2.6 ± 2.4	9.7 ± 2.5	22.1 ± 4.3	46.7 ± 13.2	18.7 ± 18.2
Sleep efficiency, %	90.8 ± 11.8	77.9 ± 13.5	81.2 ± 10.7	74.5 ± 14.5	82.6 ± 14.1
Test set					
Subjects (female:male)	10 (7:3)	4 (0:4)	7 (1:6)	6 (1:5)	27 (13:14)
Age, yr	49 ± 6	46 ± 9	55 ± 6	56 ± 5	52 ± 8
Body mass index, kg/m ²	22.3 ± 1.2	25.5 ± 1.4	24.9 ± 1.7	28.5 ± 2.1	24.8 ± 2.9
Apnea-hypopnea index, events/hr	0.8 ± 0.7	11.6 ± 3.1	20.6 ± 3.8	53.4 ± 13.6	19.2 ± 21.4
Sleep efficiency, %	87.2 ± 4.2	83.4 ± 5.7	84.0 ± 11.5	86.5 ± 4.1	85.7 ± 7.3

Data are presented as mean ± standard deviation or number.

Table 2. The short-term normal ECG dataset for SA severity identification

Measures	Normal	Mild	Moderate	Severe	Total
Training set	28,922	11,765	15,966	9,707	66,360
Validation set	7,201	2,882	4,026	2,483	16,592
Test set	8,980	3,625	5,041	3,092	20,738
Total	45,103	18,272	25,033	15,282	103,690

ECG = electrocardiography, SA = sleep apnea.

Data processing and ECG dataset

The short-term normal ECG datasets (**Table 2**) were extracted from the study population presented in **Table 1**. All participants were randomly divided into two subject groups for the training and test sets. A total of 103,690 segments for the short-term normal ECG dataset of 144 participants was obtained to train and evaluate the proposed method. The training set consisted of 82,952 segments (66,360 training set, 16,592 validation set) from 117 subjects, while the test set had 20,738 segments from 27 subjects (**Fig. 1A**).

The short-term normal ECG segment was used as the input data. The ECG signal was divided into 30 seconds segments, and the corresponding annotations of the sleep apnea episodes were stored together (**Fig. 1B**). Because the single-lead ECG signal was recorded at a sampling rate of 200 Hz, each segment consisted of 6,000 samples per segment.

Deep learning model

The CNN is a type of deep learning technique that can extract high-dimensional features from various types of datasets; it demonstrates excellent performance in both learning and testing stages.²⁴ The basic structure of the CNN is a convolutional layer, a pooling layer, and a fully-connected layer. The convolution layer is used for extracting various feature maps from the input signal. The pooling layer performs the logical functions to increase the discriminative power of features and reduce the data dimensions. Finally, the fully-connected layer is the stage at which the CNN completes the learning process, and the final discrimination of the input data is determined. It has a structure in which all the neurons of the neural network are completely connected.¹⁷ This layer discriminates among the data based on the probability values obtained through the softmax function. The designed deep

CNN model for the automatic identification of SA severity from a short-term normal ECG was represented in Fig. 1C.

Model optimization

The optimization of the model was performed using trial and error, as shown in Table 3. To optimize the model, various techniques, including rectified linear unit (ReLU), dropout, and batch normalization, were used. Batch normalization was used to the normalization of input data for model optimization at the initial stage of the model.²⁵ The dropout was used with $P = 0.25$ immediately after each max-pooling layer to reduce overfitting and divergence.²⁶ ReLU performed the activation function for the designed deep CNN model, and thereby robustly improved prediction performance.²⁷ Finally, the performance comparison between the models was based on a model constructed with repeated learning of batch sizes of 64 and 256 epochs.²⁸ Using the results of the experiments, we identified the optimum structure of the designed deep CNN model for the automatic identification of SA severity from a short-term normal ECG, as shown in Table 3.

Implementation

In this study, PSG recordings were preprocessed using MATLAB (Mathworks, Natick, MA, USA). The designed CNN model was implemented by Keras²⁹ with background TensorFlow.³⁰ In addition, the designed deep CNN model was trained and evaluated using hardware with a graphics processing unit, GeForce GTX 1080 (8 GB, GDDR5X).

Data analysis

The F-measure was used to evaluate the proposed automatic identification model for SA severity. It evaluates the correct classification of classes according to class equality. To calculate the F-measure, two evaluation measures are combined, such as precision and recall. These are defined as follows: Precision = $TP/(TP + FP)$; Recall = $TP/(TP + FN)$, where TP denotes the true positives, FP denotes the false positives, and FN denotes the false negatives. They represent the number of each events.

The F-measure can present the F1-score; it is better known for appropriating the unbalanced dataset. It can be computed based on the sample proportion of precision and recall as follows:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Table 3. Structure of the designed deep CNN model

Layers	Filter size	Output shape	Parameters
batchnorm	=	3,000 × 1	4
conv1D_1	20@50 × 1	2,951 × 20	1,020
maxpool_1	2 × 1	1,475 × 20	1,020
conv1D_2	16@30 × 1	1,446 × 16	9,616
maxpool_2	2 × 1	723 × 16	9,616
dropout_2	$P = 0.25$	723 × 16	9,616
conv1D_3	8@10 × 1	714 × 8	1,288
maxpool_3	2 × 1	357 × 8	1,288
dropout_3	$P = 0.25$	357 × 8	1,288
flatten_1		2,856 × 1	
dense_1	4 × 1	2,856 × 4	11,428
3 Conv. layers			23,354

Stride size is 1 at the convolutional layer.
CNN = convolutional neural network.

Table 4. Results of the designed deep CNN model for SA severity identification

Dataset	Performance	Normal	Mild	Moderate	Severe
Training set	Precision	1.00	0.97	0.99	1.00
	Recall	0.99	1.00	1.00	0.99
	F1-score	0.99	0.98	1.00	0.99
	Balanced accuracy	0.99	1.00	1.00	1.00
Validation set	Precision	1.00	0.97	0.99	0.99
	Recall	0.98	1.00	1.00	0.99
	F1-score	0.99	0.98	0.99	0.99
	Balanced accuracy	0.99	1.00	1.00	0.99
Test set	Precision	1.00	0.96	0.99	0.99
	Recall	0.98	1.00	1.00	0.99
	F1-score	0.99	0.98	0.99	0.99
	Balanced accuracy	0.99	0.99	1.00	0.99

CNN = convolutional neural network, SA = sleep apnea.

The balanced accuracy was used to the comparison with similar previous studies, and it was represented as the balanced value of the accuracy of the training and test phases.

Ethics statement

The protocol of PSG study was carried out at the sleep clinic of the Samsung Medical Center, Korea, having obtained permission from the Center's Institutional Review Board (approval No. 2012-01-063). Informed consent was submitted by all subjects when they were enrolled.

RESULTS

The results of the designed deep CNN for the automatic identification of SA severity from a short-term normal ECG are presented in **Table 4**. The performance was evaluated using measuring indexes such as precision, recall, F1-score and balanced accuracy. We obtained very high performances, with an accuracy of 99% for all subject groups in the training set, validation set, and test set. Furthermore, the F1-score was greater than 99% not only for the training set but also for the validation set and test set for all groups.

The confusion matrix of the performance of the proposed method for the automatic identification of SA severity is shown in **Fig. 2**. All datasets consisted of the equally distributed SA severity events as shown in confusion matrix. In addition, the results of the

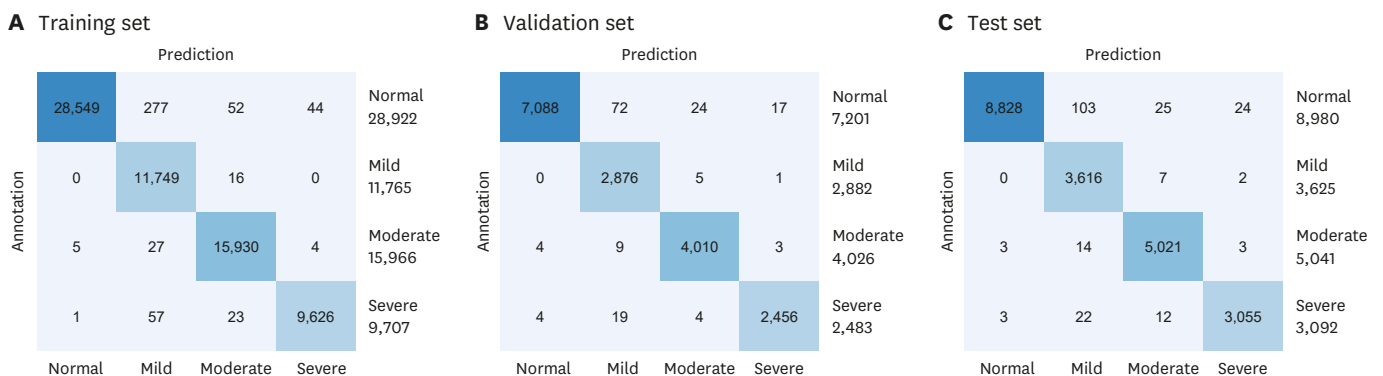


Fig. 2. Confusion matrix of the proposed method for SA severity identification based on deep learning using a short-term normal ECG signal. Confusion matrix for (A) the training set, (B) validation set, and (C) test set. SA = sleep apnea, ECG = electroencephalography.

designed deep CNN model showed very high performances with accuracy of 0.99% for all SA severity groups such as normal, mild, moderate, and severe in all datasets. The accuracy curve of the learning and evaluation phase to demonstrate the over and underfitting of the designed deep CNN model was presented in **Fig. 3A**. There is no overfitting in the designed deep CNN model and we can see that it was well learned and optimized after 50 epochs. Lastly, the receiver operating characteristic (ROC) curve and area under the curve (AUC) values for each SA severity groups in the training set, validation set and test set are presented in **Fig. 3B-D**. We obtained a very high performance of the ROC curve and AUCs more than 99% for all SA severity groups.

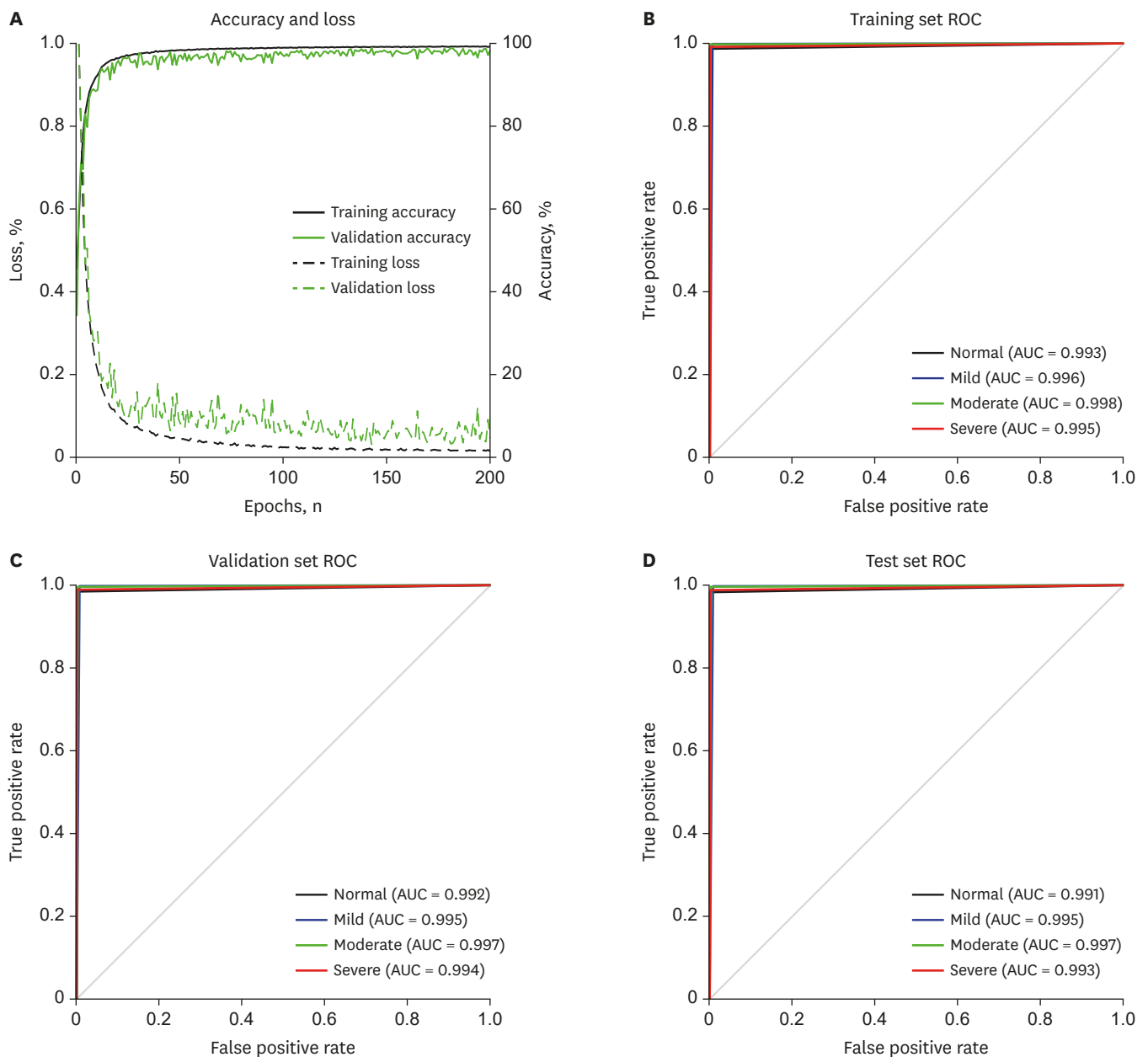


Fig. 3. Accuracy and ROC curve of the proposed method for SA severity identification based on deep learning using a short-term normal ECG signal. **(A)** Accuracy and loss, **(B)** training set ROC, **(C)** validation set ROC, and **(D)** test set ROC. ROC = receiver operating characteristic, SA = sleep apnea, ECG = electrocardiography, AUC = area under the curve.

DISCUSSION

This study demonstrated a novel method for automatically identifying SA severity from a short-term normal ECG based on a CNN model. A designed deep CNN model was used as an identifier for estimating the severity of SA using short-term normal ECG. We achieved a very high prediction rate, with a mean accuracy of 96.0% for all subject groups. In addition, the method accurately predicted the normal and severe groups with an F1-score of 99.0%. Lastly, the performance values are almost equal in the training set and test set, demonstrating that the designed deep CNN model was well trained and optimized. The ability to identify undetected SA with an easy, widely available point-of-care test using a short-term normal ECG has important practical implications, particularly for the diagnosis of patients with SA. This may be regarded as one of the most important outcomes of this study.

From an engineering perspective, there are many studies on SA screening based on ECG signals. Among them, Mendez et al.¹⁴ conducted OSA screening based on autoregressive models using a single-lead ECG signal. The QRS complexes and RR intervals were calculated from the ECG, and these intermediate vital signs were analyzed in the time and frequency domains to extract 72 features. The top 10 features were selected based on sequential forward selection and applied to the k-nearest neighbor (KNN) and neural network (NN) classifiers. They achieved good results, with an accuracy of 88% for KNN and NN. Furthermore, Mendez et al.¹⁵ proposed a method for SA screening from ECG based on nonlinear and spectral analyses. They extracted 20 features using canonical machine learning and applied the top 10 features to the linear and quadratic discriminant classifiers. Recently, Chen et al.¹⁶ suggested a novel approach for automatically screening for SA based on conventional signal processing. They used only beat-to-beat intervals derived from ECG signals and analyzed the spectral domain to extract features. They also used the support vector machine as a classifier for suspected SA patients by using extracted spectral domain features. With an accuracy of 92.8%, their method outperformed that of previous studies.^{14,15} In these studies (Table 5), the ECG signal was used to extract physiological parameters, including beat-to-beat intervals and QRS complexes. Further, a number of features were extracted using various signal processing methods including discrete wavelet analysis (DWA) and empirical mode decomposition (EMD). Thus, previous studies used manually extracted or hand-crafted features from intermediate vital signs, such as the beat-to-beat intervals and QRS complex. In addition, features were extracted from both normal and apneic ECG signals. Finally, ECG segments were analyzed for at least 5 minutes because of HRVs. In this study, no feature extraction or feature selection process was required. Using our methodology, only 30 seconds short-term normal ECG is necessary to identify SA severity. Therefore, the proposed method can be applied to undiagnosed subjects to predict and screen the severity of SA in advance.

Table 5. Comparison of the current study with previous studies

Author	Year	Dataset	Signal	Method	Accuracy
Mendez et al. ¹⁴	2009	PhysioNet, Apnea-ECG	RR, QRS	KNN, NN	88.0
Mendez et al. ¹⁵	2010	PhysioNet, Apnea-ECG	HRV, QRS	LDA, QDA	89.0
Chen et al. ¹⁶	2015	PhysioNet, Apnea-ECG SVUH/UCB, Sleep Apnea Dataset	RR	SVM	92.8
Current study	2019	SMC, Sleep Apnea Dataset	ECG	CNN	99.0

ECG = electrocardiography, KNN = k-nearest neighbor, NN = neural network, HRV = heart rate variability, LDA = linear discriminant analysis, QDA = quadratic discriminant analysis, SVUH = St. Vincent's University Hospital, UCB = University College of Dublin, SVM = support vector machines, SMC = Samsung Medical Center, CNN = convolutional neural network.

In this study, the deep CNN model was used as an automatic predictor that contained a feature map extractor and classifier throughout the layers. We have illustrated the entire working process of the deep CNN model from input to output for each subject group (Fig. 4). From this, we can deduce the difference between outputs, not only in each layer but also within the subject groups. We can see that it was difficult to find the difference between the groups of subjects in the output of the second convolution layer (conv1D_2). However, the difference between the morphology and amplitude of the output signals of the conv1D_3 layer was demonstrated. Furthermore, it can be expected that the output signal does not reflect a number of features or morphology differences. Some outputs of the conv1D_3 layer have similar morphology with the output of EMD and DWA for the ECG signal in previous studies.¹⁸ From this, we can infer that deep learning methods, including the CNN model, do not only cover conventional features but also contain novel feature maps. Their reliable and superior performances in comparison to conventional methods may be attributed to this.

The deep learning framework based on the CNN model has been used in several studies to detect SA using a single-lead ECG. In those studies, the proposed CNN models performed binary or multiclass classification for OSA events based on the ECG signal. Dey et al.²¹ conducted the binary classification of SA using ECG signal from only mild group subjects. In contrast, our previous study performed the multiclass classification of SA events including apnea and hypopnea based on a CNN model using a single-lead ECG signal from diverse patient groups.²² However, it neither covered the prediction of SA cases nor detected the

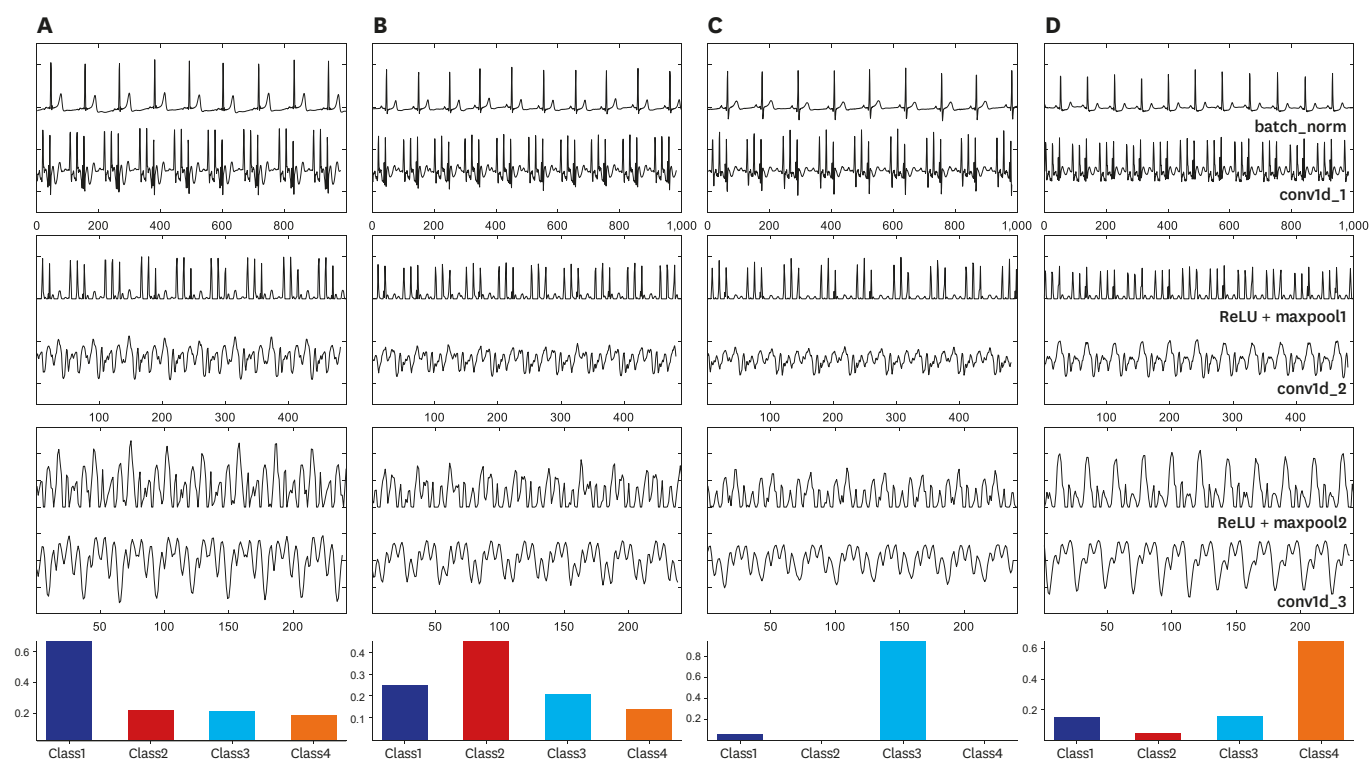


Fig. 4. Example of the output of the designed CNN model for the automatic prediction of SA using normal sinus rhythm. Intermediate features of (A) normal, (B) mild, (C) moderate, and (D) severe cases. (batch_norm—output signal of the batch-normalization layer; conv1d_1—output of first convolutional layer; ReLU+maxpool1—output of ReLU activation and max-pooling layers; conv1d_2—output of second convolutional layer; ReLU+maxpool2—output of second ReLU activation and max-pooling layers; and conv1d_3—output of last convolutional layer). The bottom bar graph is the final probability value after discrimination has occurred in the fully connected layer (class1— normal group; class2— mild group; class3— moderate group; and class4— severe group). CNN = convolutional neural network, SA = sleep apnea, ReLU = rectified linear unit.

severity of OSA. In this study, we were able to predict the severity of OSA using a 30-second normal ECG signal.

The study has the following limitations. First, we did not use the external validation set of other institutions, therefore it can lead to overfitting of the designed deep CNN model. It is, therefore, necessary to conduct a further validation study of this method using another clinical dataset such as PhysioNet, Apnea-ECG Dataset. A more diverse and larger population should be used to reduce the overfitting problem in the following study. Second, we used a short-term normal ECG from the recording of nocturnal PSG. The proposed method could underestimate when the short-term normal ECG that was measured during daytime was applied. Third, we excluded the subjects with any cardiovascular diseases and central or mixed sleep apnea. The proposed method could therefore show lower performance for the excluded groups. Finally, we cannot calculate the AHI from a short-term normal ECG signal. Currently, we can only classify SA into normal, mild, moderate, or severe based on the short-term (30 seconds) normal ECG signal.

In conclusion, we demonstrated a novel method for automatically identifying SA severity based on the deep learning model from a short-term normal ECG signal. Our suggested method achieved robust performances with a balanced accuracy of 99.0% for normal, mild, moderate, and severe groups. The results demonstrate the possibility of reliable sleep screening and monitoring using the short-term normal ECG signal without any feature extraction and preprocessing. Thus, the method we propose is a potentially helpful approach for sleep studies.

REFERENCES

1. Thorpy M. International classification of sleep disorders. In: Chokroverty S, editor. *Sleep Disorders Medicine: Basic Science, Technical Considerations and Clinical Aspects*. 4th ed. New York, NY: Springer; 2017, 475-84.
2. Šušmáková K. Human sleep and sleep EEG. *Meas Sci Rev* 2004;4(2):59-74.
3. Banno K, Kryger MH. Sleep apnea: clinical investigations in humans. *Sleep Med* 2007;8(4):400-26.
[PUBMED](#) | [CROSSREF](#)
4. Chervin RD. Sleepiness, fatigue, tiredness, and lack of energy in obstructive sleep apnea. *Chest* 2000;118(2):372-9.
[PUBMED](#) | [CROSSREF](#)
5. Graff-Radford SB, Newman A. Obstructive sleep apnea and cluster headache. *Headache* 2004;44(6):607-10.
[PUBMED](#) | [CROSSREF](#)
6. Lattimore JD, Celermajer DS, Wilcox I. Obstructive sleep apnea and cardiovascular disease. *J Am Coll Cardiol* 2003;41(9):1429-37.
[PUBMED](#) | [CROSSREF](#)
7. Lal C, Strange C, Bachman D. Neurocognitive impairment in obstructive sleep apnea. *Chest* 2012;141(6):1601-10.
[PUBMED](#) | [CROSSREF](#)
8. Freire AX, Kadaria D, Avecillas JF, Murillo LC, Yataco JC. Obstructive sleep apnea and immunity: relationship of lymphocyte count and apnea hypopnea index. *South Med J* 2010;103(8):771-4.
[PUBMED](#) | [CROSSREF](#)
9. Kapur V, Strohl KP, Redline S, Iber C, O'Connor G, Nieto J. Underdiagnosis of sleep apnea syndrome in U.S. communities. *Sleep Breath* 2002;6(2):49-54.
[PUBMED](#) | [CROSSREF](#)
10. Patil SP, Schneider H, Schwartz AR, Smith PL. Adult obstructive sleep apnea: pathophysiology and diagnosis. *Chest* 2007;132(1):325-37.
[PUBMED](#) | [CROSSREF](#)

11. Douglas NJ, Thomas S, Jan MA. Clinical value of polysomnography. *Lancet* 1992;339(8789):347-50.
[PUBMED](#) | [CROSSREF](#)
12. de Chazal P, Penzel T, Heneghan C. Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram. *Physiol Meas* 2004;25(4):967-83.
[PUBMED](#) | [CROSSREF](#)
13. Bacharova L, Triantafyllou E, Vazaios C, Tomeckova I, Paranicova I, Tkacova R. The effect of obstructive sleep apnea on QRS complex morphology. *J Electrocardiol* 2015;48(2):164-70.
[PUBMED](#) | [CROSSREF](#)
14. Mendez MO, Bianchi AM, Matteucci M, Cerutti S, Penzel T. Sleep apnea screening by autoregressive models from a single ECG lead. *IEEE Trans Biomed Eng* 2009;56(12):2838-50.
[PUBMED](#) | [CROSSREF](#)
15. Mendez MO, Corthout J, Van Huffel S, Matteucci M, Penzel T, Cerutti S, et al. Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis. *Physiol Meas* 2010;31(3):273-89.
[PUBMED](#) | [CROSSREF](#)
16. Chen L, Zhang X, Song C. An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram. *IEEE Trans Autom Sci Eng* 2015;12(1):106-15.
[CROSSREF](#)
17. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform* 2018;19(6):1236-46.
[PUBMED](#) | [CROSSREF](#)
18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*; 2016 June 27-30; Las Vegas, NV. Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2016. p. 770-8.
[CROSSREF](#)
19. Wu R, Yan S, Shan Y, Dang Q, Sun G. Deep image: scaling up image recognition. *arXiv* 2015:1501.02876.
20. Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Trans Biomed Eng* 2016;63(3):664-75.
[PUBMED](#) | [CROSSREF](#)
21. Dey D, Chaudhuri S, Munshi S. Obstructive sleep apnoea detection using convolutional neural network based deep learning framework. *Biomed Eng Lett* 2017;8(1):95-100.
[PUBMED](#) | [CROSSREF](#)
22. Urtnasan E, Park JU, Lee KJ. Multiclass classification of obstructive sleep apnea/hypopnea based on a convolutional neural network from a single-lead electrocardiogram. *Physiol Meas* 2018;39(6):065003.
[PUBMED](#) | [CROSSREF](#)
23. Berry RB, Quan SF, Abreu A. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Darien, IL: American Academy of Sleep Medicine; 2012.
24. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;61:85-117.
[PUBMED](#) | [CROSSREF](#)
25. van Laarhoven T. L2 regularization versus batch and weight normalization. *arXiv* 2017:1706.05350.
26. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;15(56):1929-58.
27. Zeiler MD, Ranzato M, Monga R, Mao M, Yang K, Le QV, et al. On rectified linear units for speech processing. In: *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*; 2013 May 26-31; Vancouver, Canada. Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2013. p. 3517-21.
[CROSSREF](#)
28. Zou F, Shen L, Jie Z, Sun J, Liu W. Weighted AdaGrad with unified momentum. *arXiv* 2018:1808.03408.
29. Keras: the Python deep learning API. <https://keras.io/>. Updated 2015. Accessed March 24, 2019.
30. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. *arXiv* 2016:1603.04467.