Journal of Medical Systems (2018) 42:104
https://doi.org/10.1007/s10916-018-0963-0

MOBILE & WIRELESS HEALTH

Automated Detection of Obstructive Sleep Apnea Events from a Single-Lead Electrocardiogram Using a Convolutional Neural Network

Erdenebayar Urtnasan[1] · Jong-Uk Park[1] · Eun-Yeon Joo[2] · Kyoung-Joung Lee[1]

Received: 28 February 2018 / Accepted: 16 April 2018

Abstract
In this study, we propose a method for the automated detection of obstructive sleep apnea (OSA) from a single-lead electrocardiogram (ECG) using a convolutional neural network (CNN). A CNN model was designed with six optimized convolution layers including activation, pooling, and dropout layers. One-dimensional (1D) convolution, rectified linear units (ReLU), and max pooling were applied to the convolution, activation, and pooling layers, respectively. For training and evaluation of the CNN model, a single-lead ECG dataset was collected from 82 subjects with OSA and was divided into training (including data from 63 patients with 34,281 events) and testing (including data from 19 patients with 8571 events) datasets. Using this CNN model, a precision of 0.99%, a recall of 0.99%, and an F_1-score of 0.99% were attained with the training dataset; these values were all 0.96% when the CNN was applied to the testing dataset. These results show that the proposed CNN model can be used to detect OSA accurately on the basis of a single-lead ECG. Ultimately, this CNN model may be used as a screening tool for those suspected to suffer from OSA.

Keywords Obstructive sleep apnea · Single-lead ECG · Convolutional neural network

Introduction

Obstructive sleep apnea (OSA) is a major sleep-disordered breathing (SDB) syndrome that is prevalent in the middle-aged population [1]. OSA is defined as breathing cessations during sleep, accompanied by oxygen desaturation or arousal for 10-s or more [2]. This condition is associated with repetitive airflow limitation and sleep fragmentation, which decrease the sleep time and degrade the sleep quality of those with OSA [3]. Undiagnosed and untreated OSA is a risk factor for excessive fatigue, daytime sleepiness [4], and even drowsy driving, which can result in traffic accidents [5]. OSA may also lead to serious and tragic consequences, such as heart attacks and sudden death [6, 7].

Polysomnography (PSG) is regarded as a standard tool for the accurate diagnosis of OSA. It can be used to characterize a subject's sleep, including the sleep cycle, sleep efficiency, apnea–hypopnea index, and snoring. However, PSG is expensive, requires well-controlled equipment and the attachment of multiple sensors, and the results must be interpreted by sleep specialists, which is very time-consuming, labor-intensive, and prone to errors. Therefore, reliable alternative methods that use fewer sensors and provide better usability for the detection of OSA should be developed.

In recent decades, various alternative methods have been proposed to replace PSG and minimize the number of biosignals required to detect OSA. These studies were based on signals such as electrocardiogram (ECG) [8–10], SpO_2 [11, 12], respiratory [13], and snoring [14, 15] signals. These studies focused on extracting the time domain, frequency domain, and other nonlinear features from the physiological signals and applying feature selection to reduce the number of dimensions comprising the feature space. In addition, various machine learning methods have been employed to improve the OSA detection performance on the basis of the extracted features.

This article is part of the Topical Collection on *Mobile & Wireless Health*

✉ Kyoung-Joung Lee
lkj5809@yonsei.ac.kr

[1] Department of Biomedical Engineering, College of Health Science, Yonsei University, 1, Yeonsedae-gil, Wonju-si, Gangwon-do 26493, South Korea

[2] Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

Published online: 23 April 2018

Several types of classifiers have been applied, including multilayer perceptron (MLP) [11], support vector machines (SVMs) [13], artificial neural networks (ANNs) [16], fuzzy logic [17], the linear discriminant analysis (LDA) [18], and AdaBoost [19]. Most of these approaches can be classified as canonical machine learning, which involves data preprocessing, feature extraction or selection, and classification. Deriving the most informative features is essential for performance; however, this process can be labor-intensive and requires domain knowledge and is particularly limiting for high-dimensional data [20].

Recent advances in machine-learning-based automated feature extraction based on learning a suitable representation of the data have been made with the use of deep neural networks [21]; this approach is also called deep learning. Deep learning provides the ability to automate the extraction of relevant features and merge the process with the classification procedure. This technique has been shown to improve the performance of applications of image and speech recognition [22, 23], natural language processing [24], and biomedical engineering [25]. A convolutional neural network (CNN) is one of the most notable deep-learning approaches and involves training multiple layers robustly. CNNs are designed to proceed input data, such as two-dimensional images with RGB color channels [26] and one-dimensional (1D) biological signals with single-channel vital signs [27], as multidimensional arrays. In this study, we propose a method for the automated detection of OSA from a single-lead ECG signal using a CNN model. The performance of the CNN model was compared with those methods proposed in previous studies to identify the best approaches to use in OSA screening systems.

Materials & methods

Subjects and data acquisition

Standard full-night PSG data from 82 subjects (63 males and 19 females) diagnosed with OSA were acquired using an Embla N7000 amplifier system (Embla Systems Inc., USA) located at the Sleep Center of Samsung Medical Center (Seoul, Korea). Overnight diagnostic PSG study, the following physiological signals were collected; electroencephalogram (EEG), electro-oculogram (EOG), electrocardiogram (ECG), chin−leg electromyogram (EMG), oxygen saturation (SpO_2), airflow, thoracic and abdominal breathing, and snoring. The PSG data were scored by a sleep specialist according to the standards of the American Academy of Sleep Medicine (AASM) guidelines [2]. The single-lead ECG signals were measured simultaneously at a sampling rate of 200 samples/s and stored at a 16-bit resolution. Central sleep apnea, mixed sleep apnea, and cardiovascular disorders were the exclusion criteria for this study. All subjects provided written informed consent before participating in the study. The study protocol was authorized by the Institutional Review Board (IRB) of Samsung Medical Center (IRB number 2012–01-063).

The subjects were randomly allocated to two groups, which were used to make up the training and testing datasets. The training dataset group comprised 17 subjects with mild OSA, 23 subjects with moderate OSA, and 23 subjects with severe OSA, and the testing dataset group comprised 5 subjects with mild OSA, 7 subjects with moderate OSA, and 7 subjects with severe OSA, as outlined in Table 1.

Data processing and datasets

A bandpass filter (5–11 Hz) was applied to the single-lead ECG data to remove the undesirable noise. The data was segmented into events, each 10-s long, to train the CNN model. The OSA datasets contained training and testing events from the subjects of the respective training and testing groups. The training dataset comprised balanced numbers of events extracted from the training subject group, including 17,092 normal breathing events and 17,189 apnea events, to avoid unbalanced training of the CNN model. The testing dataset that was used to evaluate the performance of the CNN model comprised 4334 normal breathing events and 4237 apnea events from the testing subject group.

Architecture of the CNN model

CNNs were originally inspired by a cognitive neuroscience model of a cat's visual cortex. In general, CNNs comprises a convolution, a pooling, and a classification layer. The convolution layer extracts a feature map by applying a convolutional filter to the input data. The pooling layer makes features more distinct and reduces the amount of data. Final discrimination of the input data is conducted in the classification layer [21, 22]. The proposed CNN model comprises several layers, including convolution, pooling, and activation layers, as shown in Fig. 1. Each component of the CNN is explained in detail as follows.

Input. The input signal for the CNN model is a 1D time series obtained from a single-lead ECG signal. This input signal is segmented into 10-s epochs, and each epoch is formatted as 1 × 2000 (Fig. 1).

CNN. The CNN comprises three sublayers, such as convolution, activation, and pooling layers, followed by dropout. The input signal is normalized by batch normalization, as described in Eq. (1), prior to training the CNN model.

$$\mathrm{BN}(x_i) = \alpha \cdot \left(\frac{x_i - \mu_\mathrm{B}}{\sqrt{\sigma_\mathrm{B}^2 + \epsilon}} \right) + \beta, \tag{1}$$

Table 1 Demographic and anthropometric characteristics of the subjects

All subjects	Mild [*mean*±*SD*]	Moderate [*mean*±*SD*]	Severe [*mean*±*SD*]	*p*-value	Total
Subjects (M:F)	22 (15:7)	30 (23:7)	30 (25:5)	=	82 (63:19)
Age (years)	58.08 ± 13.78	59.50 ± 10.11	56.93 ± 10.09	*NS*	58.19 ± 11.07
BMI (kg/m^2)	24.59 ± 2.28	25.49 ± 2.96	26.75 ± 2.91	*NS*	25.71 ± 2.88
AHI (per hour)	10.07 ± 2.82	21.21 ± 3.75	47.78 ± 13.78	< 0.01	27.94 ± 18.01
TRT (hour)	7.66 ± 0.68	7.21 ± 0.80	7.38 ± 0.64	*NS*	7.39 ± 0.72
Training dataset	Mild [*mean*±*SD*]	Moderate [*mean*±*SD*]	Severe [*mean*±*SD*]	*p*-value	Total
Subjects (M:F)	17 (12:5)	23 (18:5)	23 (19:4)	=	63 (49:14)
Age (years)	59.12 ± 11.91	60.26 ± 11.15	56.57 ± 9.82	*NS*	58.24 ± 10.84
BMI (kg/m^2)	24.48 ± 2.42	25.63 ± 3.08	26.96 ± 3.18	*NS*	25.88 ± 3.06
AHI (per hour)	11.03 ± 2.44	22.98 ± 3.03	52.09 ± 12.79	< 0.01	30.27 ± 19.04
TRT (hour)	7.73 ± 0.55	7.26 ± 0.60	7.47 ± 0.63	*NS*	7.46 ± 0.62
Testing dataset	Mild [*mean*±*SD*]	Moderate [*mean*±*SD*]	Severe [*mean*±*SD*]	*p*-value	Total
Subjects (M:F)	5 (3:2)	7 (5:2)	7 (6:1)	=	19 (14:5)
Age (years)	53.75 ± 21.85	57.00 ± 5.39	58.14 ± 11.56	*NS*	56.72 ± 12.06
BMI (kg/m^2)	24.96 ± 1.91	25.06 ± 2.72	26.07 ± 1.84	*NS*	25.41 ± 2.16
AHI (per hour)	6.80 ± 0.69	16.56 ± 1.17	33.64 ± 3.78	< 0.01	20.28 ± 11.43
TRT (hour)	7.44 ± 1.06	7.06 ± 1.31	7.08 ± 0.59	*NS*	7.17 ± 0.98

Mild = (5 ≤ AHI < 15); Moderate = (15 ≤ AHI < 30); Severe = (AHI ≥ 30); N = numbers; M = male; F = female; BMI = body mass index; AHI = apnea hypopnea index; TRT = total recording time; *NS*: no significant difference between patient groups (*p*-value >0.01)

where ϵ is a random noise (for stability), μ_B 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 are updated in an epoch-wise manner [28].

In the convolution layer, 1D convolution is used to extract feature maps from the input signal. Different number of 1D convolutional filters of different sizes are applied in each layer, as shown in Fig. 2. The 1D convolution can be expressed by

$$x_k^l = b_k^l + \sum_{i=1}^{N} w_k^{l-1} * y_i^{l-1} \quad (2)$$

where x_k^l denotes the kth feature map in layer l, b_k^l is the bias of the kth feature map in layer l, y_i^{l-1} represents the ith feature map in layer l–1, w_k^{l-1} is the kth convolutional kernel from all features in layer l–1 to the kth feature map in layer l, N denotes the total number of features in layer l–1, and (*) denotes vector convolution [27].

Rectified linear units (ReLU) are used as the function of the activation layer that follows each convolution layer. The use of ReLU has been shown to have a robust training performance compared to other activation functions in advanced studies [20, 29]. This approach produces large and consistent gradients, thereby aiding in gradient-based learning. The ReLU can be expressed by the following equation:

$$f(x) = \max(0, wx + b), \quad (3)$$

where x denotes the feature maps of convolution layer, w is the weight factor, and b represents the bias.

A pooling layer follows the convolution layer and is used to reduce the number of dimensions in the feature maps and network parameters. Max pooling is the most commonly used strategy. This function divides the input data into a set of non-overlapping rectangles and outputs the maximum value of each of these subsets [24]. The proposed model uses max pooling with a size of 1 × 2 after activating the convolution layer.

Dropout is a technique by which node connections are randomly ignored or dropped in each training epoch; when training is complete, all node connections are used for the classification process. This way, overfitting can be avoided and the generalizability of the network can be retained [30].

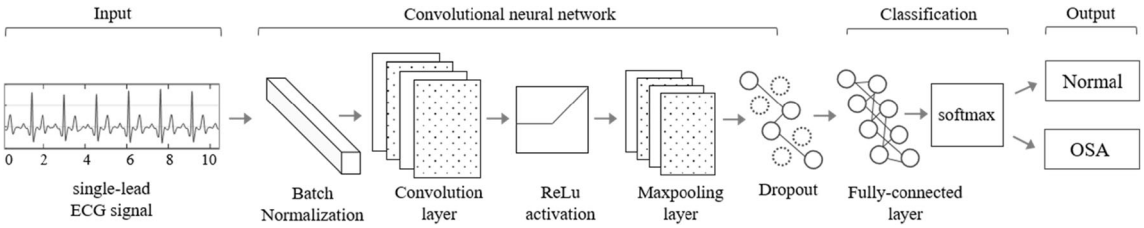

Fig. 1 Schematic of the proposed CNN model for the automated detection of OSA

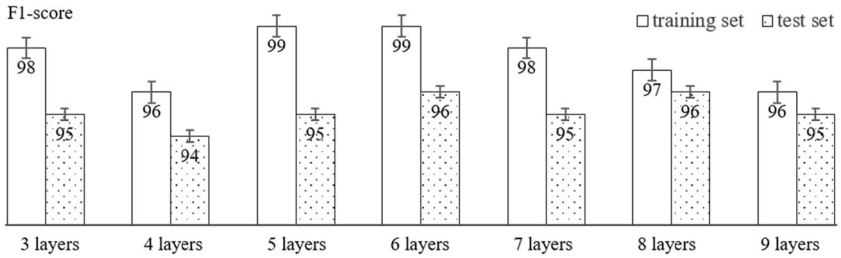

Fig. 2 Comparison of the F1-scores of the CNN models for automated OSA detection

Classification Finally, binary classification is performed in the fully connected layer by softmax activation. At this stage, all neurons are fully connected and learning proceeds by feed-forward and back-propagation algorithms [31].

Implementation

The CNN model proposed here was initially developed on Keras [32], a lightweight library used to build and train deep-learning algorithms, with a TensorFlow [33] background. The model was built on a workstation with a GeForce GTX 1080 Ti graphics processing unit (GPU) with 3584 GPU cores and 11 GB of GDDR5X memory in a Windows environment.

The CNN model was trained in a fully supervised manner, and the gradients were back-propagated from the softmax layer to the convolution layers [31]. The network parameters were optimized by minimizing the cross-entropy loss function based on the gradient descent with the Adam updating rule [34] with a learning rate of $10e^{-3}$ and a decay factor of $r = 0.9$. For the optimization, during training and testing, the data was segmented into mini-batches, each containing 256 data segments [35].

Performance analysis

To evaluate the performance of the CNN model, several accepted measures and formulas were used as indices: accuracy, precision, recall, and the F_1-score (which is the harmonic mean of the precision and recall). The measures are defined as follows:

$$accuracy = (\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}), \quad (4)$$

$$precision = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}), \quad (5)$$

$$recall = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}), \quad (6)$$

where true positive (TP), true negative (TN), false positive (FP), and false negative (FN) refer to the numbers of events in which normal is classified as normal, abnormal is classified as abnormal, abnormal is classified as normal, and normal is classified as abnormal, respectively.

$$F_1 = \sum_i 2 \cdot w_i \frac{precision_i \cdot recall_i}{precision_i + recall_i}, \quad (7)$$

where i is the class index and $w_i = n_i/N$ is the proportion of the number of samples in the ith class, n_i, and the total number of samples, N.

Experiments

Optimization of the CNN architecture

To test the proposed CNN model, training and testing were conducted with the OSA dataset described in "Data processing and datasets" section. The results demonstrate the complexity and significance of the hyperparameter selection that is required to achieve an optimal performance in the automated detection of OSA. An important objective of these experiments was to identify the optimal architecture that provides the highest performance while incurring the lowest computational cost for real-time implementation.

Table 2 Comparison of different architectures of the CNN model for automated OSA detection

Layers	Parameters	Training time (epoch/s)	Training loss (%)	Precision (%)	Recall (%)	F_1-score (%)
3 layers	15,420	3.0	0.123	0.95	0.95	0.95
4 layers	21,294	3.0	0.151	0.95	0.94	0.94
5 layers	45,272	4.0	0.112	0.95	0.95	0.95
6 layers	**61,832**	**4.0**	**0.110**	**0.96**	**0.96**	**0.96**
7 layers	73,112	4.0	0.113	0.95	0.95	0.95
8 layers	84,608	6.0	0.114	0.96	0.96	0.96
9 layers	97,628	8.0	0.126	0.95	0.95	0.95

The architecture was bolded that shown the best performances.

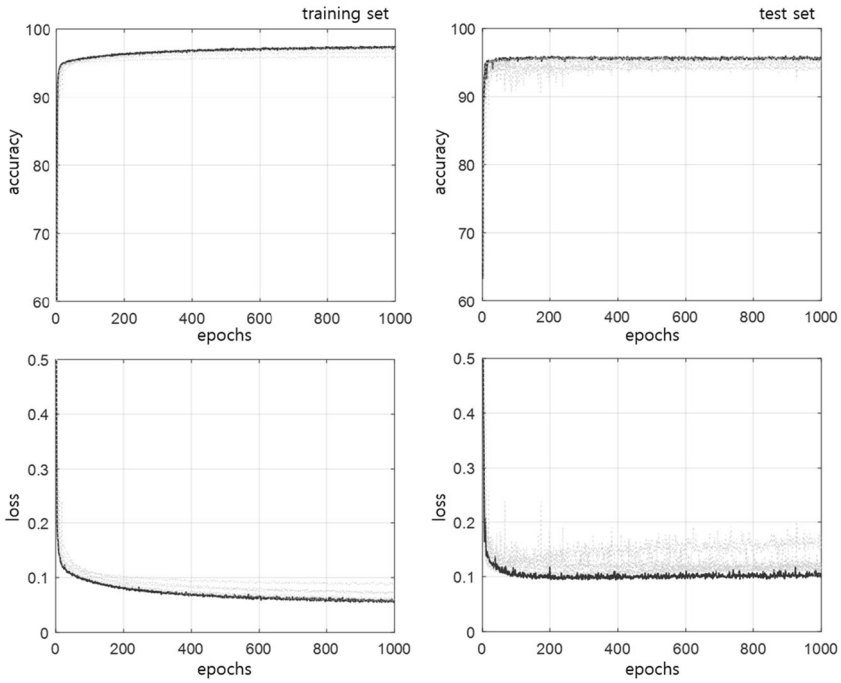

Fig. 3 Accuracy and loss of the proposed CNN model for automated OSA detection

Therefore, experiments were performed while varying the depth of the convolution layer to identify the optimal architecture for automated OSA detection. The hyperparameter conditions of the activation function (ReLU) [29], dropout rate ($p = 0.25$) [30], and optimizer function (Adam) [34] were also used herein. The trainable parameters, training time for each epoch, training loss, and average performance for the testing dataset were considered in the selection of the optimal architecture for the CNN model. The results are shown in Table 2; architectures with more than nine layers exhibited overfitting and underfitting and were therefore omitted.

As shown in Table 2, as the convolution layer depth and the number of trainable parameters increase, the training time increases and the training loss decreases. The results show that a six-layer architecture provides the highest performance and the lowest training time and loss. A comparison of the F_1-scores for different architectures when applied to the training and testing datasets is shown in Fig. 2. These results show that architectures with five and six layers result in the highest F_1-scores of 0.99% for the training dataset and those with six and eight layers show the highest F_1-scores of 0.96% for the testing dataset. However, the six-layer architecture has a smaller number of trainable parameters and a faster training time compared with the eight-layer architecture.

Figure 3 shows the learning results in terms of accuracy and loss obtained as the number of epochs was varied. The results show that the accuracy and loss reach stable values after several iterations of learning when applied to the testing dataset. The accuracy of the six-layer architecture is higher and the loss is lower compared to the other architectures for both the training and testing datasets. From these results, it can be concluded that the CNN model has been designed and optimized appropriately to detect apnea events from single-lead ECG signals. The optimal architecture of the CNN model for automated OSA detection was therefore selected on the basis of this performance analysis of all CNN model

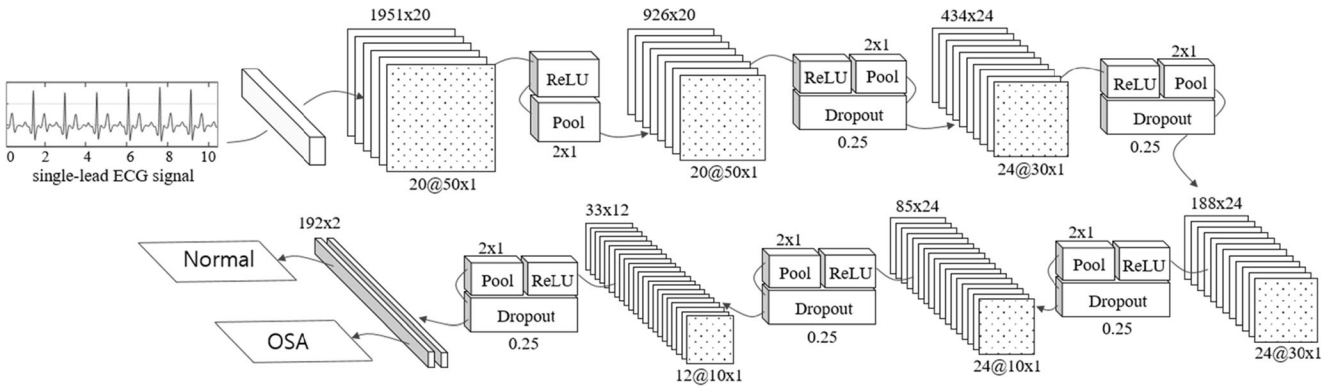

Fig. 4 The architecture of the proposed CNN model for automated OSA detection

Table 3 The performance of the proposed CNN model for automated OSA detection

Dataset	Events		Precision	Recall	F_1-score
Training dataset	Normal	17,092	0.99	0.98	0.99
	OSA	17,189	0.98	0.99	0.99
	Average		0.99	0.99	0.99
Testing dataset	Normal	4334	0.95	0.96	0.96
	OSA	4237	0.96	0.95	0.96
	Average		0.96	0.96	0.96

architectures, as presented in Table 2 and Figs. 2 and 3. The detailed characteristics and performance of the optimal architecture of the CNN model are described in "Performance of the proposed CNN model" section.

Performance of the proposed CNN model

In this section, we summarize the evaluation that was conducted to identify the optimal architecture for the CNN model for automated OSA detection from a single-lead ECG signal. The optimal architecture is a six-layer CNN model: layers 1 and 2 comprise 20 filters with a kernel size of 50 × 1, layers 3 and 4 comprise 24 filters with a kernel size of 30 × 1, layer 5 comprises 24 filters with a kernel size of 10 × 1, and layer 6 comprises 12 filters with a kernel size of 10 × 1. ReLU was used as the activation function in each layer, and dropout was applied at layers 2 through 6 (Fig. 4).

Table 3 shows the performance of the CNN model with the optimal architecture for automated detection of OSA from a single-lead ECG signal. When applied to the training dataset, a precision of 0.99%, a recall of 0.98%, and an F_1-score of 0.99% were obtained for the normal breathing events, and these parameters were 0.98%, 0.99%, and 0.99%, respectively, for the apnea events. When applied to the testing dataset, a precision of 0.95%, a recall of 0.96%, and an F_1-score of 0.96% were obtained for the normal breathing events, and these parameters were 0.96%, 0.95%, and 0.96%, respectively, for apnea events.

Discussion & conclusion

In this study, a novel method was proposed for the automated detection of OSA from single-lead ECG data using a CNN model. The CNN model was designed with 1D convolution kernels, and various architectures were compared to identify the optimal architecture for automated OSA detection. The CNN model was trained and tested using OSA datasets from 82 subjects with OSA. This CNN model achieved a precision of 0.96, a recall of 0.96, and an F_1-score of 0.96 when applied to the testing datasets, representing an excellent level of performance.

The results of this study were compared with those of previous studies, as shown in Table 4. A previous study [19] used heart rate variability (HRV), oxygen saturation, and respiratory effort data; extracted feature sets; and applied an SVM classifier with linear and second-order polynomial kernels to achieve automated detection of OSA. Apnea was detected from an ECG signal using a combinational feature set that includes nonlinear and frequency-domain features using an SVM [13, 16, 18]. A study [9, 36–38] proposed an automated screening approach for apnea based on the low- and high-frequency ratio of RR intervals in a single-lead ECG signal. In each of these studies, at least two vital signs were extracted from the single-lead ECG signal, including HRV, ECG-derived respiratory, and RR intervals; then, a number of features were predefined. However, the proposed CNN model can efficiently detect OSA events without the requirement to extract vital signs from a raw signal or to conduct calculations to predefine features.

This comparison indicated that the proposed CNN model performed better compared with the other methods based on single-lead ECG signals that have been previously presented. A one-sample *t*-test was used to identify statistically significant differences in the accuracies obtained with the proposed CNN model and the other methods; the results are shown in Table 4. In all of the *t*-tests, the null hypothesis was rejected at a 5% significance level, indicating that the proposed CNN model achieved a significantly higher performance compared with all the previous studies. In addition, the CNN model exhibited a stable and robust performance without being

Table 4 Comparison of the results of the proposed method with those reported in other studies

Author	Year	Subject	Method	Acc. (%)	Sen. (%)	Spe. (%)
Álvarez-Estévez [17]	2009	12	Fuzzy reasoning module	=	87.0	89.0
Mendez [18]	2010	25	LDA, QDA	89.7	90.4	86.7
Al-Angari [19]	2012	100	SVM	82.4	69.9	91.4
Xie [36]	2012	25	AdaBoost, Bagging REPTree	87.0	85.8	84.4
Jafari [37]	2013	35	SVM	94.8	94.1	95.4
Chen [38]	2015	35	Kernel density classifier	82.1	83.2	80.2
Our method	2017	82	CNN	96.0	96.0	96.0

Acc. accuracy, *Sen.* sensitivity, *Spe.* specificity

affected by the severity of OSA or sleep stages of the subject. According to the AASM guidelines [2], SpO_2 and respiratory signals are necessary to diagnose OSA. However, the proposed CNN model exhibited the ability to accurately and automatically detect OSA events based on a single-lead ECG signal. This approach is also applicable to ECG signals obtained from sleep measurement systems based on PSG, such as those used in hospitals, and continuous positive airway pressure (CPAP), which are commonly used at homes. Our novel approach for accurately diagnosing and detecting OSA can ultimately be utilized for stable and accurate detection in the hospital and home environments.

There are some limitations in this study. At first, the proposed CNN model only classifies OSA event and normal, but not for the central, mixed, and hypopnea events. The central and mixed sleep apnea occurs very rarely, but hypopnea is prevalent in SDB patients. False detection could be shown for patients who hypopnea events are dominant. Second, our reference PSG data are scored by one certified clinician, so there is no cross-checking, which is intra-scorer and inter-scorer variability, for reference annotation. Third, the proposed CNN model didn't detect the starting and end time of events, we only detect the presence or absence of OSA during 10 s event in this study. Fourthly, the single-lead ECG signals that were used in this study were not fully separated from other events, such as snoring, movement, and flow limitation. Finally, a small population was investigated in this study. Thus, further studies should be conducted to address these limitations to facilitate the development of more robust deep-learning models.

In conclusion, we proposed a method for the automated detection of OSA based on a single-lead ECG signal using a CNN model. The proposed CNN model was trained and tested on ECG recording datasets obtained from 82 subjects with OSA. Based on the results, this CNN model can be a helpful tool for sleep technicians to annotate OSA data as an alternative to manually scoring OSA events according to the criteria specified by the AASM guidelines. In addition, this model may be helpful for OSA screening, particularly when applied to standard PSG and CPAP systems.

Acknowledgments This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Ministry of Education. It is a product of the Local Innovative Creative Human Resource Training Project (NRF-2014H1C1A1063845).

Compliance with ethical standards

Conflict of interest All authors declares that he or she has no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Young, T., Evans, L., Finn, L., and Palta, M., Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep* 20(9):705–706, 1997.
2. Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Marcus, C., and Vaughn, B., AASM manual for the scoring of sleep and associated events. Rules, terminology and technical specifications. Darien, IL: AASM, 2012.
3. Engleman, H. M., and Douglas, N. J., Sleep 4: Sleepiness, cognitive function, and quality of life in obstructive sleep apnoea/hypopnea syndrome. *Thorax* 59(7):618–622, 2004. https://doi.org/10.1136/thx.2003.015867.
4. Vgontzas, A. N., Papanicolaou, D. A., Bixler, E. O., Hopper, K., Lotsikas, A., Lin, H.-M., Kales, A., and Chrousos, G. P., Sleep apnea and daytime sleepiness and fatigue: Relation to visceral obesity, insulin resistance, and hypercytokinemia. *J. Clin. Endocrinol. Metab.* 85(3):1151–1158, 2000.
5. Barbé, F., Pericas, J., Munoz, A., Findley, L., Anto, J. M., and Agusti, A. G., Automobile accidents in patients with sleep apnea syndrome: An epidemiological and mechanistic study. *A J. Res. Crit. Care Med.* 158(1):18–22, 1998.
6. Bhattacharjee, R., Kheirandish-Gozal, L., Pillar, G., and Gozal, D., Cardiovascular complications of obstructive sleep apnea syndrome: Evidence from children. *Prog. Cardiovasc. Dis.* 51(5):416–433, 2009.
7. Lopez-Jimenez, F., Sert Kuniyoshi, F. H., Gami, A., and Somers, V. K., Obstructive sleep apnea: Implications for cardiac and vascular disease. *Chest* 133(3):793–804, 2008. https://doi.org/10.1378/chest.07-0800.
8. Penzel, T., The apnoea-ECG database. *Comput. Cardiol.* 27:255–258, 2000.
9. Penzel, T., McNames, J., de Chazal, P., Raymond, B., Murray, A., and Moody, G., Systematic comparison of different algorithms for apnoea detection based on ECG recordings. *Med. Biol. Eng. Comput.* 40(4):402–407, 2002.
10. Yildiz, A., Akın, M., and Poyraz, M., An expert system for automated recognition of patients with obstructive sleep apnea using electrocardiogram recordings. *Expert Syst. Appl.* 38(10):12880–12890, 2011.
11. Marcos, J. V., Hornero, R., Alvarez, D., Aboy, M., and Del Campo, F., Automated prediction of the apnea-hypopnea index from nocturnal oximetry recordings. *IEEE Trans. Biomed. Eng.* 59(1):141–149, 2012. https://doi.org/10.1109/TBME.2011.2167971.
12. Park, J. U., Lee, H. K., Lee, J., Urtnasan, E., Kim, H., and Lee, K. J., Automatic classification of apnea/hypopnea events through sleep/wake states and severity of SDB from a pulse oximeter. *Physiol. Meas.* 36(9):2009–2025, 2015.
13. Koley, B. L., and Dey, D., Automatic detection of sleep apnea and hypopnea events from single channel measurement of respiration signal employing ensemble binary SVM classifiers. *Measurement* 46(7):2082–2092, 2013.
14. Solà-Soler, J., Fiz, J. A., Morera, J., and Jané, R., Multiclass classification of subjects with sleep apnoea–hypopnoea syndrome through snoring analysis. *Med. Eng. Phys.* 34(9):1213–1220, 2012.
15. Erdenebayar, U., Park, J. U., Jeong, P., and Lee, K. J., Obstructive sleep apnea screening using a piezo-electric sensor. *J. Korean Med. Sci.* 32(6):893–899, 2017.
16. Khandoker, A. H., Gubbi, J., and Palaniswami, M., Automated scoring of obstructive sleep apnea and hypopnea events using

short-term electrocardiogram recordings. *EEE Trans Inf Technol Biomed* 13(6):1057–1067, 2009. https://doi.org/10.1109/TITB.2009.2031639.

17. Álvarez-Estévez, D., and Moret-Bonillo, V., Fuzzy reasoning used to detect apneic events in the sleep apnea-hypopnea syndrome. *Expert Syst. Appl.* 36(4):7778–7785, 2009. https://doi.org/10.1016/j.eswa.2008.11.043.
18. Mendez, M. O., Corthout, J., Van Huffel, S., Matteucci, M., Penzel, T., Cerutti, S., and Bianchi, A. M., Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis. *Physiol. Meas.* 31(3):273–289, 2010. https://doi.org/10.1088/0967-3334/31/3/001.
19. Al-Angari, H. M., and Sahakian, A. V., Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier. *IEEE Trans. Inf. Technol. Biomed.* 16(3):463–468, 2012. https://doi.org/10.1109/TITB.2012.2185809.
20. Angermueller, C., Pärnamaa, T., Parts, L., and Stegle, O., Deep learning for computational biology. *Mol. Syst. Biol.* 12(7):878, 2016.
21. LeCun, Y., Bengio, Y., and Hinton, G., Deep learning. *Nature* 521(7553):436–444, 2015. https://doi.org/10.1038/nature14539.
22. Krizhevsky, A., Sutskever, I., Hinton, G. E., Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process Syst.* 1:1097–1105, 2012
23. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., and Kingsbury, B., Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process. Mag.* 29(6):82–97, 2012.
24. Sutskever, I., Vinyals, O., Le, Q. V., Sequence to sequence learning with neural networks. *Adv. Neural Inf. Process Syst.* 2:3104–3112, 2014
25. Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Perez, J. A., Lo, B., and Yang, G. Z., Deep learning for health informatics. *IEEE J. Biomed. Health Inform.* 21(1):4–21, 2016. https://doi.org/10.1109/JBHI.2016.2636665.
26. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., and Lew, M. S., Deep learning for visual understanding: A review. *Neurocomputing* 187:27–48, 2016. https://doi.org/10.1016/j.neucom.2015.09.116.
27. Kiranyaz, S., Ince, T., and Gabbouj, M., Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Trans. Biomed. Eng.* 63(3):664–675, 2016.
28. Ioffe, S., Szegedy, C., Batch normalization: accelerating deep network training by reducing internal covariate shift. International Conference on Machine Learning 448–456, 2015.
29. Rosasco, L., Vito, E. D., Caponnetto, A., Piana, M., and Verri, A., Are loss functions all the same? *Neural Comput.* 16(5):1063–1076, 2004.
30. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15:1929–1958, 2014.
31. Yu, X., Efe, M. O., and Kaynak, O., A general backpropagation algorithm for feedforward neural networks learning. *IEEE Trans. Neural Netw.* 13(1):251–254, 2002.
32. Chollet, F, Keras, 2015. http://keras.io/.
33. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., and Kudlur, M., TensorFlow: A system for large-scale machine learning. *OSDI* 16:265–283, 2016.
34. Kingma, D. P., Ba, J., Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
35. Chien, C., Batch size selection for the batch means method. Proceedings of the 26th Conference on Winter Simulation, Society for Computer Simulation International, 345–352, 1994.
36. Xie, B., and Minn, H., Real-time sleep apnea detection by classifier combination. *IEEE Trans. Inf. Technol. Biomed.* 16(3):469–477, 2012. https://doi.org/10.1109/TITB.2012.2188299.
37. Jafari, A., Sleep apnoea detection from ECG using features extracted from reconstructed phase space and frequency domain. *Biomed Signal Process Control* 8(6):551–558, 2013.
38. Chen, L., Zhang, X., and Song, C., An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram. *IEEE Trans. Autom. Sci. Eng.* 2:106–115, 2015.