



Automatic detection of sleep-disordered breathing events using recurrent neural networks from an electrocardiogram signal

Erdenebayar Urtnasan¹ · Jong-Uk Park¹ · Kyoung-Joung Lee¹

Received: 27 June 2018 / Accepted: 19 October 2018 / Published online: 28 October 2018
© Springer-Verlag London Ltd., part of Springer Nature 2018

Abstract

In this study, we propose a novel method for automatically detecting sleep-disordered breathing (SDB) events using a recurrent neural network (RNN) to analyze nocturnal electrocardiogram (ECG) recordings. We design a deep RNN model comprising six stacked recurrent layers for the automatic detection of SDB events. The proposed deep RNN model utilizes long short-term memory (LSTM) and a gated-recurrent unit (GRU). To evaluate the performance of the proposed RNN method, 92 SDB patients were enrolled. Single-lead ECG recordings were measured for an average 7.2-h duration and segmented into 10-s events. The dataset comprised a training dataset (68,545 events) from 74 patients and test dataset (17,157 events) from 18 patients. The proposed method achieved high performance with an F1-score of 98.0% for LSTM and 99.0% for GRU. The results demonstrate superior performance over conventional methods. The proposed method can be used as a precise screening and diagnosing tool for patients with SDB disorders.

Keywords Sleep-disordered breathing · Recurrent neural network · Long short-term memory · Gated-recurrent unit

1 Introduction

Sleep-disordered breathing (SDB) is the most common sleep disorder, and it includes sleep apnea and sleep hypopnea. SDB is characterized by repetitive cessations (apnea) or decreases (hypopnea) of breathing for at least 10 s during sleep. SDB is known to degrade the quality of sleep and that of life by causing excessive sleepiness, fatigue, irritability, and inattention [1]. Undiagnosed SDB can exacerbate risk factors of coronary artery disease [2], cardiac arrhythmias [3], hypertension [4], stroke [5], diabetes [6], cognitive dysfunction [7], and depression [8]. The treatment of these factors has become a high-cost burden on the healthcare system [9].

A nocturnal polysomnography (PSG) is considered the standard method for objectively evaluating sleep disorders, including SDB. However, PSG requires uncomfortable diagnostic equipment with multiple sensors, trained attendees, and great expense. Additionally, manual annotation

by sleep specialists is particularly time-consuming and labor-intensive. Different results can be produced or errors can occur depending on the experience and subjective judgment of the specialist.

Over the last two decades, there have been several studies of novel methods using a single-lead electrocardiogram (ECG) to replace PSG for SDB detection. Initially, Penzel et al. used a single-lead ECG for automatic detection of sleep apnea in early 2000 [10]. Since then, many studies have used a single-lead ECG signal for minimizing the sensors for signal measurement and easy implementation. For those studies, it is important to extract the discriminative features, select the optimal features, and apply them to the various machine learning methods. Heart rate variability [11], inter-beat (RR) interval, and ECG-derived respiratory were used to extract the discriminative feature sets [12]. Those signals were analyzed using advanced signal processing techniques in different analytic domains (e.g., time, frequency, nonlinear domain) [13, 14]. Optimal features were selected through statistical evaluation [14], wrapper methods [15], and principal component analysis [12] from extracted feature sets to reduce the dimensions and improve performance. Finally, robust classifiers such as artificial neural networks [16] and

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

¹ Department of Biomedical Engineering, College of Health Science, Yonsei University, Wonju 26493, Korea

support vector machines [17] were employed for SDB detection. However, those studies had some drawbacks because of numerous calculations and computation, hand-crafted feature sets, and lower detection rates. To solve these problems, some studies [18, 19] used deep learning in the form of convolutional neural networks (CNN), which have shown high performances. However, CNN is designed for image recognition and requires high computational power.

Recurrent neural networks (RNN) are extensions of conventional feedforward neural networks, which handle variable-length sequences and time-series data [20]. They are known to have enhanced the performance of speech recognition [21], natural language processing [22], and biomedical engineering activities [23]. Furthermore, SDB has a repetitive temporal occurrence that can be regarded as a “time series.” RNN can be more useful and appropriate for detection of SDB than conventional machine learning and/or CNN-based methods.

In this study, we propose a novel method for automatic detection of SDB events based on deep RNN using a single-lead ECG signal. We utilize two major RNN models: long short-term memory (LSTM) and gated-recurrent unit (GRU). The LSTM is used as the main memory cell, and GRU is used for performance comparison. Finally, we compare performances between the conventional and proposed methods.

2 Materials and methods

2.1 Subjects and data processing

We collected recordings of nocturnal PSGs from 92 subjects (74 males and 18 females) suffering SDB. The PSG recordings were conducted using the Embla N7000 amplifier system (Embla System Inc., USA) in the Sleep Center of the Samsung Medical Center (Seoul, Korea). In accordance with the American Academy of Sleep Medicine (AASM) guidelines [24], all PSG recordings were annotated by certified sleep technicians and verified by sleep specialists. The institutional review board (No. 2012-01-063) of Samsung Medical Center approved this study and waived the patient consent requirement. All patients were provided written informed consent for participating in this study (Table 1). The exclusion criteria were patients with central sleep apnea, mixed sleep apnea, and cardiovascular disorders.

A single-lead ECG signal was recorded by a lead II transducer at 200 samples/s during the nocturnal PSG. A bandpass filter (5–11 Hz) was applied for data preprocessing to remove undesired noise from the ECG signal. Then, all preprocessed ECG signals were segmented at

10-s duration events. The segmentation was performed by specialists with no overlap. If more than half of a segment is annotated as normal, it is considered a normal event, and vice versa. Of all events, 182,642 were normal, 21,426 were apneas, and 34,841 were hypopneas.

2.2 The proposed method

An RNN is ideally suited to sequential information and is excellent for time-series data because it also has memory. RNN is a looped-back architecture of interconnected neurons and current input; the last hidden state affects the output of the next hidden state.

The proposed method for automatic detection of SDB events, based on RNNs from a single-lead ECG, is illustrated in Fig. 1. The proposed method comprises four parts: input, RNN, classification, and output. The input of RNN is a single-lead ECG signal, including the physiological signs (e.g., RR interval, heart rate, and respiration). Input signals were normalized before applying the RNN (Fig. 1a). The architecture consists of a 6-layer RNN, and each has a different number of memory cells. We experimentally found an optimal architecture of the deep RNN model for automatic detection of SDB. Additionally, LSTM and GRU memory cells were applied to the proposed deep RNN model to compare their performances (Fig. 1b).

LSTM is a modification of RNN that allows the influence of time steps to be passed farther along a sequence than is possible with a simple RNN [22]. LSTM is an extension of a simple RNN with memory cells to make learning temporal relationships easy over time. In LSTM, each memory cell contains three major gates: an input gate, an output gate, and a forget gate [23]. LSTM is expressed as follows.

The input gate controls the flow of input activations into the memory cell.

$$i_t = \sigma(W^{xi}x_t + W^{hi}h_{t-1} + b_i). \quad (1)$$

The output gate controls the output flow of cell activations into the rest of the network.

$$o_t = \sigma(W^{xo}x_t + W^{ho}h_{t-1} + b_o). \quad (2)$$

The forget gate scales the internal state of the cell before adding it as input through the self-recurrent connection of the cell. Therefore, it adaptively forgets or resets the cell's memory.

$$f_t = \sigma(W^{xf}x_t + W^{hf}h_{t-1} + b_f), \quad (3)$$

$$g_t = \sigma(W^{xc}x_t + W^{hc}h_{t-1} + b_c), \quad (4)$$

$$c_t = f_t * c_{t-1} + i_t * g_t, \quad (5)$$

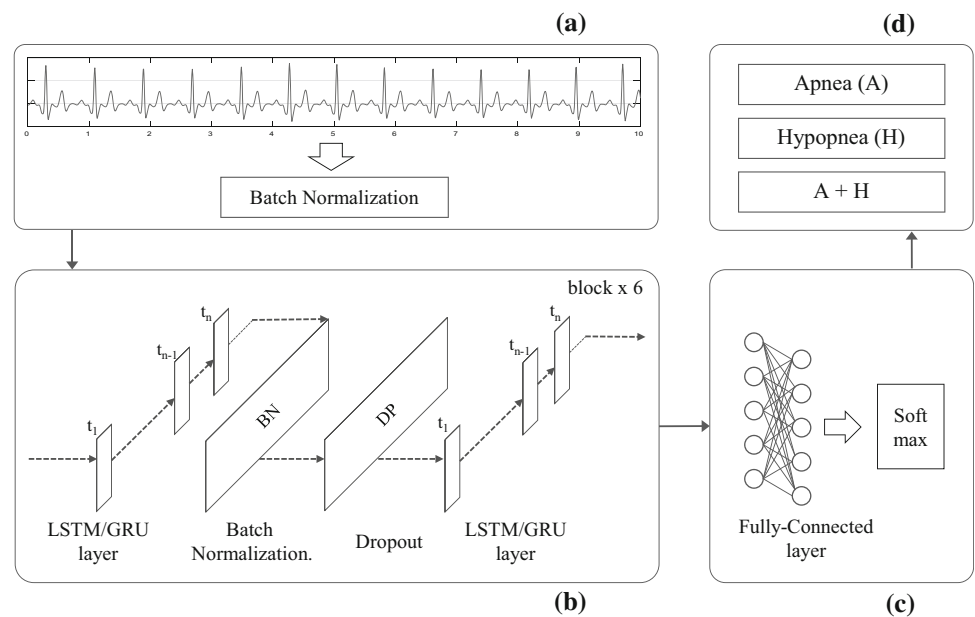
$$h_t = o_t * \phi(c_t), \quad (6)$$

Table 1 Demographic and anthropometric characteristics of the subject groups

All subjects	Mild (mean ± SD)	Moderate (mean ± SD)	Severe (mean ± SD)	<i>p</i> value	Total (mean ± SD)
<i>N</i> Gender (M:F)	22 (15:7)	38 (31:7)	32 (27:5)		92 (73:19)
Age (years)	58.1 ± 13.8	58.6 ± 9.7	56.3 ± 10.2	<i>NS</i>	57.7 ± 10.8
BMI (kg/m ²)	24.6 ± 2.3	26.2 ± 3.4	26.8 ± 3.0	<i>NS</i>	26.0 ± 3.1
AHI (per hour)	10.1 ± 2.8	21.8 ± 4.3	48.0 ± 13.8	< 0.01	28.1 ± 17.5
TST (hour)	6.0 ± 0.9	5.9 ± 1.0	5.7 ± 1.2	<i>NS</i>	5.8 ± 1.0
<i>Training set</i>					
<i>N</i> Gender (M:F)	18 (12:6)	30 (26:4)	26 (22:4)		74 (57:14)
Age (years)	59.5 ± 11.7	58.6 ± 10.3	55.4 ± 9.9	<i>NS</i>	56.9 ± 12.6
BMI (kg/m ²)	24.6 ± 2.4	26.3 ± 3.5	27.1 ± 3.1	<i>NS</i>	25.6 ± 3.2
AHI (per hour)	10.8 ± 2.5	22.9 ± 4.2	51.5 ± 12.8	< 0.01	30.0 ± 12.5
TST (hour)	6.1 ± 0.9	5.9 ± 0.9	5.7 ± 1.2	<i>NS</i>	5.5 ± 1.3
<i>Test set</i>					
<i>N</i> Gender (M:F)	4 (3:1)	8 (6:2)	6 (5:1)		18 (14:4)
Age (years)	49.7 ± 24.8	55.0 ± 5.9	60.2 ± 11.4	<i>NS</i>	55.9 ± 12.2
BMI (kg/m ²)	24.7 ± 2.1	26.0 ± 3.4	25.6 ± 1.9	<i>NS</i>	25.6 ± 2.6
AHI (per hour)	6.6 ± 0.7	16.3 ± 1.0	32.5 ± 2.7	< 0.01	19.6 ± 10.3
TST (hour)	5.5 ± 0.7	6.3 ± 1.2	5.4 ± 1.2	<i>NS</i>	5.8 ± 1.1

N Numbers; *M* male; *F* female; *BMI* body mass index; *AHI* apnea–hypopnea index; *TST* total sleep time; *SE* sleep efficiency; *NS* = no significant difference among patient groups (*p* value > 0.01)

Fig. 1 Schematic diagram of the proposed deep RNN model. **a** Input gets single-lead ECG and normalized. **b** RNN model consists of the LSTM and GRU. **c** Classification contains a fully connected multilayer perceptron with softmax activation. **d** Output is calculated for apnea (A) events, hypopnea (H) events, and A + H events



where *i*, *f*, *o*, and *c* are, respectively, the input gate, forget gate, output gate, and cell activation vectors, all of which are the same size as vector *h*, defining the hidden value. Terms σ and τ represent nonlinear and hyperbolic tangent functions, respectively (Fig. 2).

GRU is a relatively new type of RNN. It is a simplified version of LSTM that combines the cell and hidden states and uses an update gate instead of a forget gate and an

input gate. GRU use has boomed in recent years, turning into a strong competitor of LSTM [25, 26]. However, The GRU has gate units that modulate the flow of information inside the unit. However, they do not have separate memory cells [27].

$$z_t = \sigma(W^{xz}x_t + W^{hz}h_{t-1} + b_z), \tag{7}$$

Fig. 2 Structure of the LSTM memory cell

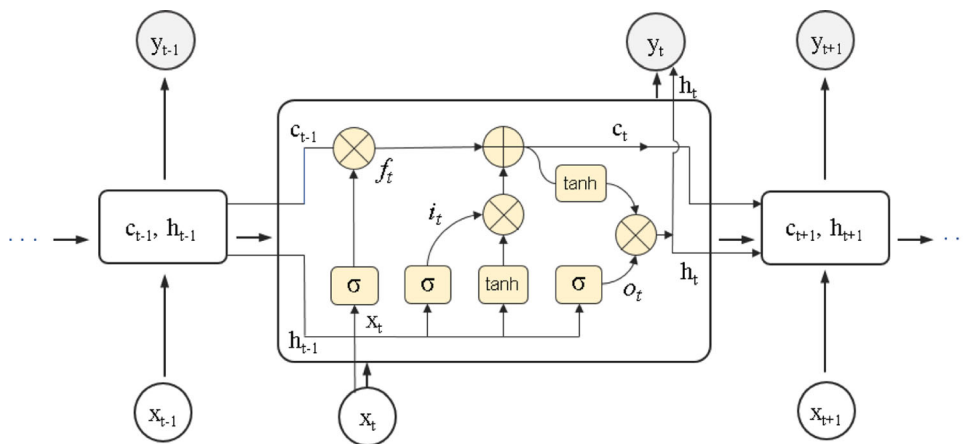
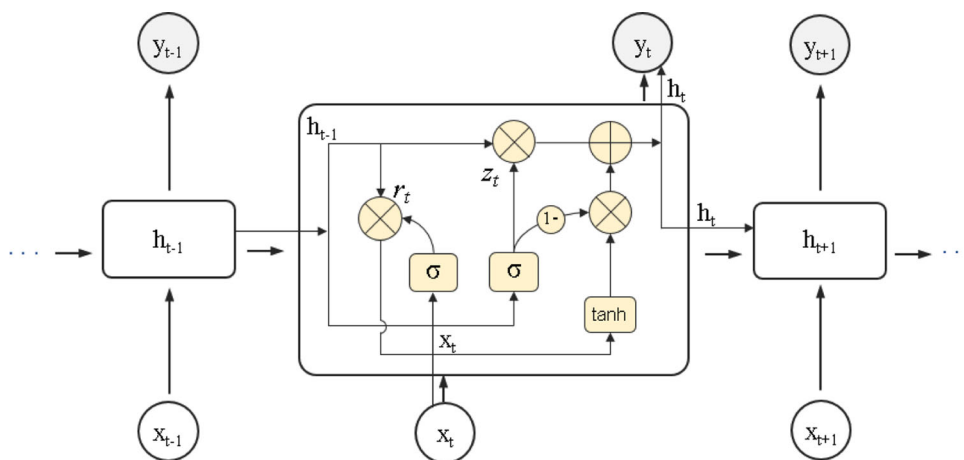


Fig. 3 GRU memory cell



$$r_t = \sigma(W^{xr}x_t + W^{hr}h_{t-1} + b_r), \tag{8}$$

$$h'_t = \tanh(Wx_t + Wh_{t-1} * r_t), \tag{9}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * h'_t, \tag{10}$$

where z , r , and h are, respectively, the input gate, the forget gate, the output gate, and cell activation vectors, all of which are the same size as vector h , defining 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 (Fig. 3).

After the LSTM/GRU layer, output feature maps endure batch normalization and dropout layers to avoid overfitting and divergence. Classification is performed by the fully connected network using softmax regression (Fig. 1c). The outputs of the proposed method are evaluated by the apnea (A), hypopnea (H), and A + H events (Fig. 1d).

2.3 Implementation and training

The proposed deep RNN model for automatic detection of SDB events was implemented by Keras' [28] platform using a TensorFlow [29] background. Keras is a library that

can easily build and evaluate deep learning models. It was trained and evaluated on a graphical processing unit (GeForce GTX1080 Ti) and a central processing unit (Intel E5-1620 v2 3.50 GHz, 8 CPUs). RNNs are trained in a fully supervised way, back-propagating the gradients from the softmax layer through to the recurrent units. The network parameters are optimized by minimizing the cross-entropy loss function using mini-batch gradient descent with the Adam update rule [30].

We performed heuristic experiments to find the optimal architecture of the proposed deep RNN model. All ECG segments had same 10-s duration and were shaped as 2000×1 . Finally, we found optimal architecture of the deep RNN model for our dataset. The model architecture was optimized by batch normalization, dropout, and multilayer perceptron (MLP) as presented Table 2.

2.4 Performance measures

We evaluate the proposed deep RNN model using the F-measure (F1-score), one that considers the correct classification of each class equally. The F1-score combines two measures as precision and recall. Additionally, accuracy

Table 2 The parameters and characteristics of the proposed deep RNN model

Layer type	Layer name	Activation	# Memory cells	Parameters	
				GRU	LSTM
Input	batchnorm_0	=		4	4
RNN model	rnn_1	sigmoid	120	43,920	58,560
	batchnorm_1			480	480
	rnn_2	sigmoid	100	66,300	88,400
	batchnorm_2			400	400
	rnn_3	sigmoid	80	43,440	57,920
	batchnorm_3			320	320
	dense_1			6480	6480
	rnn_4	sigmoid	60	25,380	33,840
	batchnorm_4			240	240
	rnn_5	sigmoid	40	12,120	16,160
	batchnorm_5			160	160
	rnn_6	sigmoid	20	3660	4880
	batchnorm_6			80	80
	dense_2			420	420
Classification	dense_3	softmax	2	42	42
Total	6 rnn layers		422	203,446	268,386

rnn Recurrent neural layer, *batchnorm* batch normalization layer, *dense* MLP

was calculated for performance comparison with other studies. These are defined as follows.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}), \tag{11}$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}), \tag{12}$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}), \tag{13}$$

where TP and FP are the number of true and false positives, respectively. TN and FN correspond to the number of true and false negatives.

$$\text{F1} = \sum_i 2 \cdot w_i \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}, \tag{14}$$

where i is the class index and $w_i = n_i/N$ is the proportion of samples of class i , with n_i being the number of samples of the i th class; N is the total number of samples.

3 Results

3.1 SDB datasets

We used the SDB datasets collected from the 92 subjects to train and evaluate the proposed deep RNN model. The dataset consisted of the balanced and randomly selected events from the total segmented events, including normal, apnea, and hypopnea. The LSTM and GRU models were trained on the dataset of 74 subjects and tested on the dataset of 18 subjects (Table 3).

3.2 Results of LSTM model

The results of performance evaluation of the LSTM model are shown in Table 4. For the test set, the LSTM model showed a precision of 98.0%, a recall of 98.0%, and an F1-score of 98.0% for apnea events; 97.0%, 97.0%, and 97.0%, for hypopnea events; and 97.0%, 96.0%, and 96.0%, for $A + H$ events, respectively. The LSTM model achieved a very stable and robust performance for the SDB events.

Accuracy and loss rates of the LSTM model were determined, as shown in Fig. 4. The accuracy graphs show that there were two inspiration points after 5 and 25 iterations. Learning accuracy stabilized after 40 iterations. The LSTM model took many iterations to achieve stability and inspiration because of its complex structure. Additionally, there were some spikes and variations in the hypopnea event for test set. That result demonstrates that it is challenging to learn for hypopnea events using single-lead ECG signal. Loss gradually decreases in the training set, but it stabilized after 40 iterations in the test set for all SDB events. There are some fluctuations in the curves of accuracy and loss in the test phase of the LSTM model. Those fluctuations were caused not only by the similar patterns of apnea and hypopnea events, but also by the motion artifacts and other noises of the single-lead ECG signal since ECG signal was not preprocessed.

Table 3 Detailed SDB dataset information

Dataset	Normal events	Hypopnea events	Apnea events	Total
Training set	34,281	17,140	17,124	68,545
Test set	8570	4285	4302	17,157
Total	42,852	21,426	21,426	85,702

Table 4 The LSTM model performance for SDB events

Dataset	Events	Precision	Recall	F1-score
Training set	Apnea	98.0	98.0	98.0
	Hypopnea	98.0	98.0	97.0
	A + H	97.0	97.0	97.0
Test set	Apnea	98.0	98.0	98.0
	Hypopnea	97.0	97.0	97.0
	A + H	97.0	96.0	96.0

3.3 GRU model results

GRU performance is presented in Table 5. For the test set, the GRU model had a precision of 99.0%, a recall of 99.0%, and an F1-score of 99.0% for apnea events; 97.0%, 97.0%, and 97.0% for hypopnea events; and 96.0%, 95.0%, and 95.0% for A + H events, respectively. The GRU model had a precise and high performance for the SDB events.

Figure 5 shows how accuracy and losses of the GRU model changed according to the number of iterations

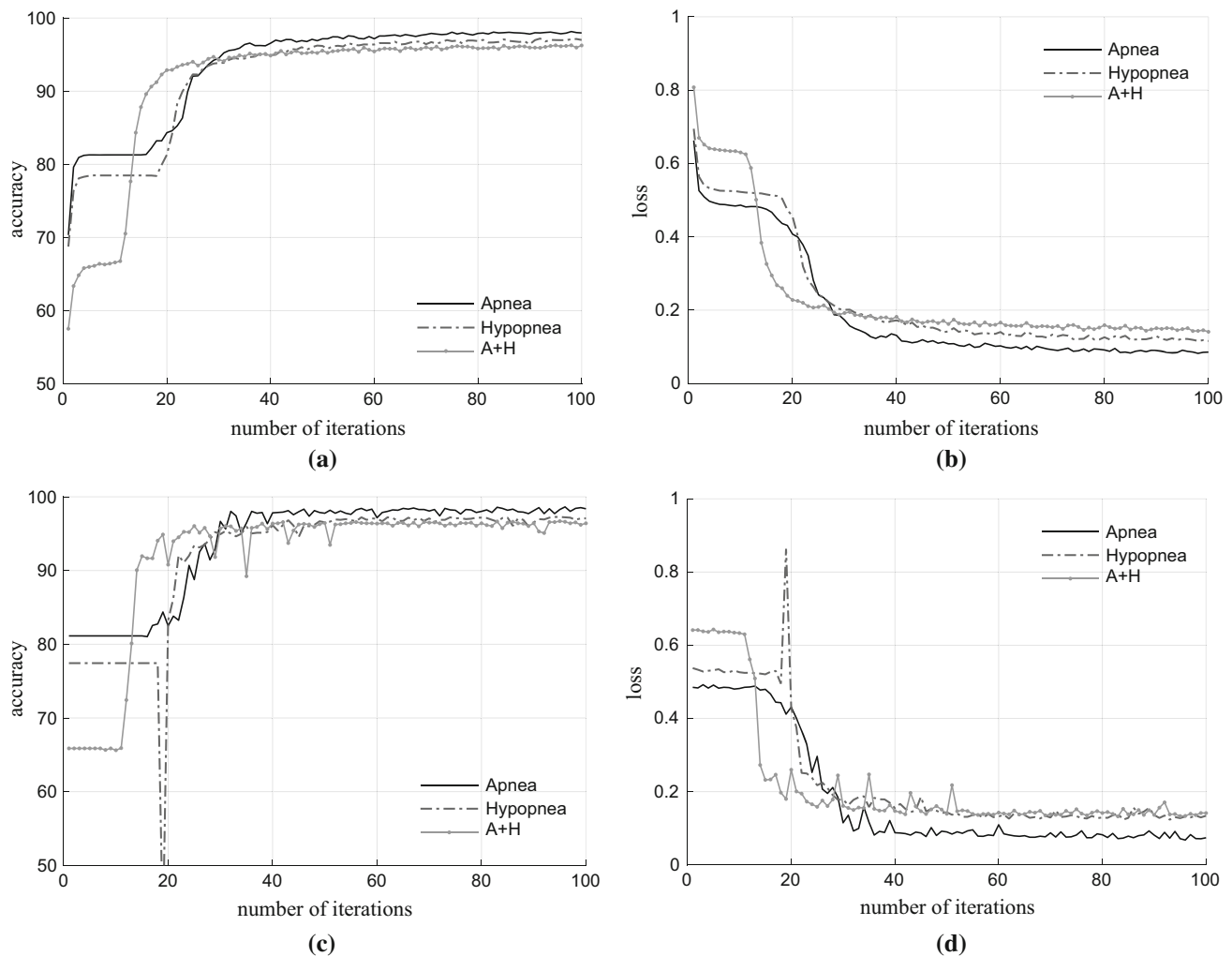


Fig. 4 LSTM model accuracy and losses for SDB event detection. **a** Training set accuracy, **b** training set losses, **c** test set accuracy, **d** test set losses

Table 5 GRU model performance for SDB events

Dataset	Events	Precision	Recall	F1-score
Training	Apnea	99.0	99.0	99.0
	Hypopnea	97.0	97.0	97.0
	A + H	96.0	96.0	96.0
Test	Apnea	99.0	99.0	99.0
	Hypopnea	97.0	97.0	97.0
	A + H	96.0	95.0	95.0

performed. The GRU model showed faster inspiration and better performance than the LSTM model. The GRU model’s learning accuracy stabilized after 20 iterations (Fig. 5a), which was almost twice as fast as the LSTM model. Additionally, GRU showed robust performance

over the LSTM model. However, some spikes occurred when processing the apnea and hypopnea event datasets.

4 Discussion

This study proposed a novel method for automatic detection of SDB events based on deep RNN from a single-lead ECG signal. The proposed deep RNN model was designed on an LSTM model, and a GRU model was applied for performance comparison. Each model was trained and evaluated using SDB datasets from 92 patients with SDB. The LSTM model achieved a high performance with an F1-score of 98.0% for apnea events, 97.0% for hypopnea events, and 96.0% for A + H events, for the test set. The GRU model showed a precise performance with an F1-

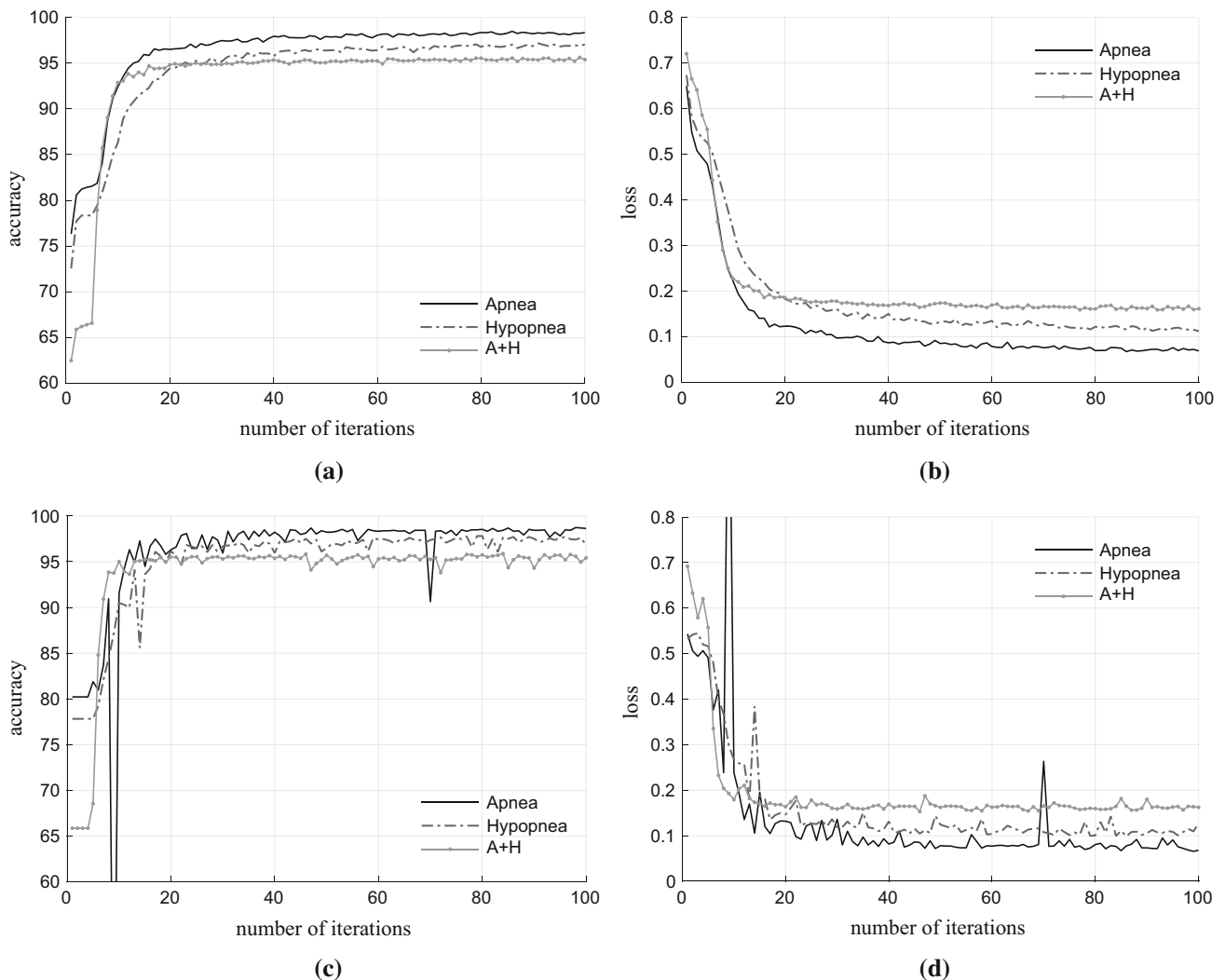


Fig. 5 GRU model accuracy and losses for SDB event detection. **a** Training set accuracy, **b** training set losses, **c** test set accuracy, **d** test set losses

score of 99.0% for apnea events, 97.0% for hypopnea events, and 95.0% for A + H events.

Several studies that proposed methods for automatic detection of sleep apnea using a single-lead ECG signal, as listed Table 6. Mendez et al. [12] used RR interval and ECG-derived respiratory signal from a single-lead ECG signal. These were analyzed by empirical mode decomposition and wavelet analysis to extract two feature sets containing 10 and 20 features. Linear and quadratic discriminant classifier (QDA) were used for sleep apnea classification. Al-Angari and Sahakian [17] used a non-linear measure of synchronous signals that presented a phase-locking value between respiratory, ECG, and SpO2 signals. The phase-locking values were applied to a support vector machine (SVM) for sleep apnea classification. Xie and Minn [15] used SpO2 and ECG signals as input and extracted 150 features from those two inputs. Finally, 39 features were selected through feature selection and used classifier combinations. However, all of those studies proposed methods that conducted complicate signal processing, feature extraction, and feature selection. Additionally, the results showed performances under 90%. In contrast, Jafari [13] and Chen et al. [14] showed higher performances for sleep apnea detection, but complex nonlinear feature sets and multivariable statistical analyses were used. The proposed deep RNN method can eliminate those complex calculations for signal processing, feature extraction, and feature selection. This is shown to be superior than all conventional methods listed in Table 6.

Dey et al. [18] and Urtnasan et al. [19] used a CNN model to classify sleep apnea using the single-lead ECG, as listed in Table 6. In those studies, they designed and found the optimal architecture of the CNN model, and their results demonstrated high performances for only classification of apnea events but not for classification of SDB, events including the hypopnea and A + H events. Additionally, their results were higher than previous studies, which used the conventional machine learning algorithms. Finally, they also used fewer numbers of subjects than ours. In Dey et al. [18], the population did not contain the

mild and moderate SDB patients; they only consisted of 12 normal and 23 severe SDB patients. However, the proposed deep RNN model showed superior performance because we used a bigger dataset of several types of SDB patients. Urtnasan et al. [19] used a dataset from all groups of SDB patients and found an optimal CNN architecture for sleep apnea detection using a single-lead ECG signal. However, their performances were lower than that of the proposed LSTM and GRU model for apnea events.

The proposed RNN model for automatic detection of SDB events obtained more robust performance than conventional methods. In addition, it can discriminate hypopnea events using a single-lead ECG signal that was very challenging task for conventional methods [11, 12]. The main reasons to reach at the result are deep architecture of RNN model and recurrent memory cells such as LSTM/GRU. The proposed deep RNN model used basic memory cell of LSTM and GRU. Particularly, forget gate of LSTM and update gate of GRU played a main role for automatic detection of SDB events. Not only memory cells can recognize the characteristics of SDB dataset, but it can strongly represent the long-term dependencies of apnea and hypopnea events. Also, deep architecture as a good supporter of LSTM and GRU memory cells function as the performance enhancer of the proposed RNN model.

From the result of our deep RNN models designed for automatic detection of SDB events, we received some insights for suitability of the RNN model in diagnosing and screening SDB. In terms of engineering, first is the enhancement of the feature extraction process performed by high-dimensional data abstraction. Second is the increase in discrimination power for precise classification of the events, which is rarely seen in conventional classification methods. From a clinical perspective, deep RNN models can provide more robust performances for SDB event detection and can distinguish the hidden events including hypopnea and A + H using fewer input signals. Thus, the proposed deep RNN model can possibly serve as a helpful and alternative tool for the PSG method.

Table 6 Performance comparison with other studies

Study	Years	Subject	Method	Accuracy	Sensitivity	Specificity
Mendez [12]	2010	50	QDA	89.7	90.4	86.7
Al-Angari [17]	2012	100	SVM	82.4	69.9	91.4
Xie [15]	2012	25	Bagging REPTree	79.0	84.6	83.3
Jafari [13]	2013	35	SVM	94.8	95.4	94.1
Chen [14]	2015	90	SVM	97.4	98.9	92.9
Dey [18]	2016	35	CNN	98.9	97.8	99.2
Urtnasan [19]	2017	82	CNN	96.0	96.0	96.0
Our method	2018	92	LSTM	98.5	98.0	98.0
			GRU	99.0	99.0	99.0

5 Limitations and conclusion

There are some limitations in our study. We did not consider the central and mixed sleep apnea events because of their rarity. The proposed deep RNN model is unaware of the starting and ending point of apnea events because of performing event-based detection that only can detect the presence or absence of apnea events. The reference annotation of the PSG recordings was labeled by one certified clinician and not by cross-checking. We did not remove the noise events (e.g., snoring, movements). We used only basic memory cell of LSTM and GRU, and did not use any variation of LSTM/GRU and bi-directional RNNs. Finally, a small number of subjects were used for the proposed method. Further studies resolving these limitations and thereby facilitating the development of more robust deep learning models should be conducted. In addition, use of another class of methods such as Gaussian process should be considered [31].

In this study, the deep RNN models demonstrated automatic detection of SDB events using a single-lead ECG. Their performance was evaluated for LSTM and GRU models. Each model showed excellent performance. The LSTM model demonstrated an F1-score of 98.0% for apnea events, and the GRU model showed an F1-score of 99.0% for apnea events. The results of these models are applicable to ECG signals obtained from sleep measurement systems. Finally, a new approach was proposed for accurately diagnosing and detecting SDB events. A GRU model can be a helpful tool for sleep technicians to annotate SDB because they manually annotate SDB events according to their preferred criteria within the AASM guidelines. Additionally, the model can be more valuable for SDB screening, particularly with standard PSG and CPAP systems.

Acknowledgements This work was supported by the Yonsei University Research Fund of 2018.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

References

- Engleman HM, Douglas NJ (2004) Sleep-4: sleepiness, cognitive function, and quality of life in obstructive sleep apnoea/hypopnoea syndrome. *Thorax* 59(7):618–622
- Arzt M, Hetzenecker A, Steiner S, Buchner S (2015) Sleep-disordered breathing and coronary artery disease. *Can J Cardiol* 31(7):909–917
- Flemons WW, Remmers JE, Gillis AM (1993) Sleep apnea and cardiac arrhythmias: is there a relationship? *Am Rev Respir Dis* 148(3):618–621
- Grote L, Ploch T, Heitmann J, Knaack L, Penzel T, Peter JH (1999) Sleep-related breathing disorder is an independent risk factor for systemic hypertension. *Am J Respir Crit Care Med* 160(6):1875–1882
- Mohsenin V (2001) Sleep-related breathing disorders and risk of stroke. *Stroke* 32(6):1271–1278
- Botros N, Concato J, Mohsenin V, Selim B, Doctor K, Yaggi HK (2009) Obstructive sleep apnea as a risk factor for type 2 diabetes. *Am J Med* 122(12):1122–1127
- Fulda S, Schulz H (2001) Cognitive dysfunction in sleep disorders. *Sleep Med Rev* 5(6):423–445
- Peppard PE, Szklo-Coxe M, Hla KM, Young T (2006) Longitudinal association of sleep-related breathing disorder and depression. *Arch Intern Med* 166(16):1709–1715
- Kapur V, Blough DK, Sandblom RE, Hert R, de Maine JB, Sullivan SD, Psaty BM (1999) The medical cost of undiagnosed sleep apnea. *Sleep* 22(6):749–755
- Penzel T (2000) The apnoea-ECG database. *Comput Cardiol* 27:255–258
- Penzel T, McNames J, de Chazal P, Raymond B, Murray A, Moody G (2002) Systematic comparison of different algorithms for apnoea detection based on ECG recordings. *Med Biol Eng Comput* 40(4):402–407
- Mendez MO, Corthout J, Van Huffel S, Matteucci M, Penzel T, Cerutti S, Bianchi AM (2010) Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis. *Physiol Meas* 31(3):273–289. <https://doi.org/10.1088/0967-3334/31/3/001>
- Jafari A (2013) Sleep apnoea detection from ECG using features extracted from reconstructed phase space and frequency domain. *Biomed Signal Proc Control* 8(6):551–558
- Chen L, Zhang X, Song C (2015) An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram. *IEEE Trans Autom Sci Eng* 2:106–115
- Xie B, Minn H (2012) Real-time sleep apnea detection by classifier combination. *IEEE Trans Inf Technol Biomed* 16(3):469–477. <https://doi.org/10.1109/TITB.2012.2188299>
- Khandoker AH, Gubbi J, Palaniswami M (2009) Automated scoring of obstructive sleep apnea and hypopnea events using short-term electrocardiogram recordings. *IEEE Trans Inf Technol Biomed* 13(6):1057–1067. <https://doi.org/10.1109/TITB.2009.2031639>
- Al-Angari HM, Sahakian AV (2012) Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier. *IEEE Trans Inf Technol Biomed* 16(3):463–468. <https://doi.org/10.1109/TITB.2012.2188209>
- Dey D, Chaudhuri S, Munshi S (2018) Obstructive sleep apnoea detection using convolutional neural network based deep learning framework. *Biomed Eng Lett* 8(1):95–100
- Urtnasan E, Park JU, Joo EY, Lee KJ (2018) Automated detection of obstructive sleep apnea events from a single-lead electrocardiogram using a convolutional neural network. *J Med Syst* 42:1–8
- Zhang H, Cao X, Ho JK, Chow TW (2017) Object-level video advertising: an optimization framework. *IEEE Trans Ind Inform* 13(2):520–531
- Zhang H, Ji Y, Huang W, Liu L (2018) Sitcom-star-based clothing retrieval for video advertising: a deep learning framework. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-018-3579-x>
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neur Comput* 9(8):1735–1780

23. Sak H, Senior A, Beaufays F (2014) Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. [arXiv:1402.1128](https://arxiv.org/abs/1402.1128)
24. Berry RB, Brooks R, Gamaldo CE, Harding SM, Marcus C, Vaughn B (2012) AASM manual for the scoring of sleep and associated events. Rules, terminology and technical specifications. AASM, Darien, IL
25. Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. [arXiv:1412.3555](https://arxiv.org/abs/1412.3555)
26. Zhang H, Li J, Ji Y, Yue H (2017) Understanding subtitles by character-level sequence-to-sequence learning. *IEEE Trans Ind Inform* 13(2):616–624
27. Gao, Y, Glowacka D (2016) Deep gate recurrent neural network. In: Asian conference on machine learning, pp 350–365
28. Chollet F (2015) Keras. <http://keras.io/>
29. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Kudlur M (2016) TensorFlow: a system for large-scale machine learning. *OSDI 16*:265–283
30. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. *arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)*
31. Das D, Lee CS (2018) Cross-scene trajectory level intention inference using gaussian process regression and naive registration. *Purdue ECE Technical Report* (2018)