Biomedical Signal Processing and Control 69 (2021) 102906

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Classification of Obstructive Sleep Apnoea from single-lead ECG signals using convolutional neural and Long Short Term Memory networks

Haifa Almutairi *, Ghulam Mubashar Hassan, Amitava Datta

Department of Computer Science and Software Engineering, The University of Western Australia, Australia

ARTICLE INFO

Keywords:
Obstructive Sleep Apnoea
ECG
Deep learning
Classification
Convolutional Neural Networks
Long Short Term Memory

ABSTRACT

Obstructive Sleep Apnoea (OSA) is a breathing disorder that happens during sleep. Polysomnography (PSG) is typically used as a reference standard for the diagnosis of OSA which uses different physiological signals such as Electrocardiography (ECG), Electroencephalogram (EEG) and Electromyogram (EMG) in a sleep laboratory. This procedure is time-consuming, expensive and inconvenient. However, detection of OSA by using a wearable sensor to collect Electrocardiography (ECG) signals is a practical and effective alternative. Previous studies of OSA classification from ECG signals focused on feature engineering methods which involves extracting specific features from ECG signals and using the extracted feature as inputs to the machine learning methods. In this study, we propose a novel method of OSA classification of ECG signal where deep learning methods automatically extract the features from the ECG signals and classify them. Our deep learning approach uses a hybrid model involving Convolution Neural Networks (CNN) and Long Short Term Memory (LSTM) networks. PhysioNet Apnea-ECG database is used for training and evaluation of our proposed deep learning model. For the released training dataset, our proposed model achieves the accuracy of 94.27%, sensitivity of 94.57%, specificity of 93.93% and F1 score of 95.41%. While for the testing dataset, the achieved accuracy, sensitivity, specificity and F1 score for the proposed model are 90.92%, 91.24%, 90.36% and 92.76% respectively. The performance of our model is compared with state of the art techniques and we found our model to achieve the best performance to classify OSA and health ECG signals.

1. Introduction

Sleep Apnoea is a prevalent sleep disease that has three types: Obstructive Sleep Apnoea (OSA), Central Sleep Apnoea and Mixed Sleep Apnoea [1,2]. OSA is a common sleep disorder characterized by repeated episodes during sleep. It occurs when the upper airway is obstructed despite attempts to breathe. When the upper airflow is partially closed, the amount of air entering into the lungs is reduced, this is called Hypopnea [3]. The typical symptom of a person suffering from OSA after a full night of sleep as a result of these episodes is that they experience excessive daytime sleepiness. Other typical symptoms include a headache in the morning, exhaustion and fatigue throughout their day. Snoring is considered a sign of OSA [4]. Detection of OSA can protect patients from other disease such as: cardiovascular diseases, recurrent heart attacks, diabetes, stroke and neurocognitive deficits.

Conventionally, *polysomnography* (PSG) [5] is a clinical procedure used for the diagnosis of OSA which uses different physiological signals including Electrocardiography (ECG) [6], Electroencephalogram (EEG) [7] and Electromyogram (EMG) [8] in a sleep laboratory. It is an expensive as well as inconvenient clinical procedure. Typically the patient is required to sleep in a laboratory at a hospital for two to three days which require appropriate medical staff including experts to monitor and analyse the physiological signals. These signals are manually analyzed by expert physicians to detect OSA and other heart diseases. This manual task is time consuming and requires a high level of skills. It requires patient's body to be connected with multiple electrodes and 22 wires to record 16 major physiological signals. Furthermore physiological signals quality is another challenging issue due to its low amplitude. Artefacts caused by muscle contractions, body motion and breathing introduce noise into the signals undermining ECG signal quality [9]. The large number of wires and electrodes connected to the patient's body and being in a sleep laboratory tend to disturb sleep and in turn affect the signals being recorded [10]. This complicates the task of manually analyzing the physiological signals and may cause human errors that can lead to misdiagnosis. Therefore, detection of OSA from a single-lead ECG signal is a low-cost practical alternative that can be used

* Corresponding author.
E-mail address: haifa.almutairi@research.uwa.edu.au (H. Almutairi).

https://doi.org/10.1016/j.bspc.2021.102906
Received 9 September 2020; Received in revised form 3 June 2021; Accepted 19 June 2021
Available online 30 June 2021

at patient's home [11]. Few studies [12,13] used oxygen saturation (SpO2) signals to detect OSA. However, other studies [14–17] demonstrate that ECG signals are more promising to detect OSA.

ECG records the electrical activity of the heart muscle. The three primary waves P, QRS complex and T are contained from each heartbeat [18]. Fig. 1 presents 1-min ECG segments of an OSA and a normal subject. A regular pattern of ECG waves can be observed for the normal subject which shows a normal heart cycle.

2. Related works

Few studies suggested to extract features from consecutive R interval and ECG-Derived Respiration (EDR) signals, and using machine learning classifiers to detect OSA. For instance, Bsoul et al. used Daubechies wavelet features as inputs which were based on R-R intervals and ECG-Derived Respiration (EDR) signals [14]. The study adopted a Gaussian SVM method to detect OSA. The accuracy was reported to be 89.09%. Similarly, Chazal et al. [19] used Quadratic Discriminants (QD) classifier to detect OSA from ECG signals. The study used R-R intervals and EDR signals as inputs and the accuracy was reported to be 90%. Recently, Varon et al. [20] used LS-SVM to classify OSA from ECG signals. The study suggested to extract statistical features such as standard deviation and principal component analysis (PCA) of R-R intervals and EDR signals. The accuracy was reported to be 84.74%.

Similarly, few studies focused on feature selection from raw data by applying different wavelet methods and then using machine learning classifiers to detect OSA. Babaeizadeh et al. [21] proposed a Quadratic classifier to detect OSA. The study used spectral power distribution features as inputs and the per segment classification accuracy was 84.70%. Hassan et al. [22] used Adaptive Boosting method to detect OSA. The study proposed to use Tunable Q-factor Wavelet transform (TQWT) features as inputs. The classification accuracy was 87.33% over all segments. Nguyen et al. [23] used SVM classifier to detect OSA from ECG signals. The study used Recurrence Quantification Analysis (RQA) features as inputs. The segments' classification accuracy was 85.26%. Sharma et al. [24] proposed SVM, K-Nearest Neighbours (KNN) and multilayer perceptron neural networks (MPNN) classifier models to detect OSA. A filter bank method was used to extract features and best accuracy of 85% was achieved for the MPNN model. Another study by Sharma et al. [25] proposed an KNN model to detect OSA. The proposed technique used filter bank features as inputs. The classification accuracy on all segments was 84.9%. Recently, Zarei et al. [10] used SVM method to detect OSA. The technique used a wavelet decomposition method to extract entropy based features of the ECG signals. The segment classification accuracy was 92.98%.

Aforementioned researches share two common limitations. Firstly, they use handcrafted methods to extract features which are normally complicated and computationally expensive. Secondly, creation of these relevant handcrafted features require considerable domain knowledge of ECG-based signals and OSA. These limitations can be solved by using *Deep Learning* (DL) methods that can automatically extract features from raw data.

Deep Learning (DL) is a subfield of machine learning inspired by the human brain functions called *artificial neural networks*(ANN). It contains multiple layers that extract features from input data and classify them with high accuracy [26]. Recently, DL methods have achieved high performance for extracting and classifying data in clinical applications [27–32]. Due to the technological advances in the health care system, DL is found to have high potential for learning from massive amount of available data. For example, recently several studies proposed to

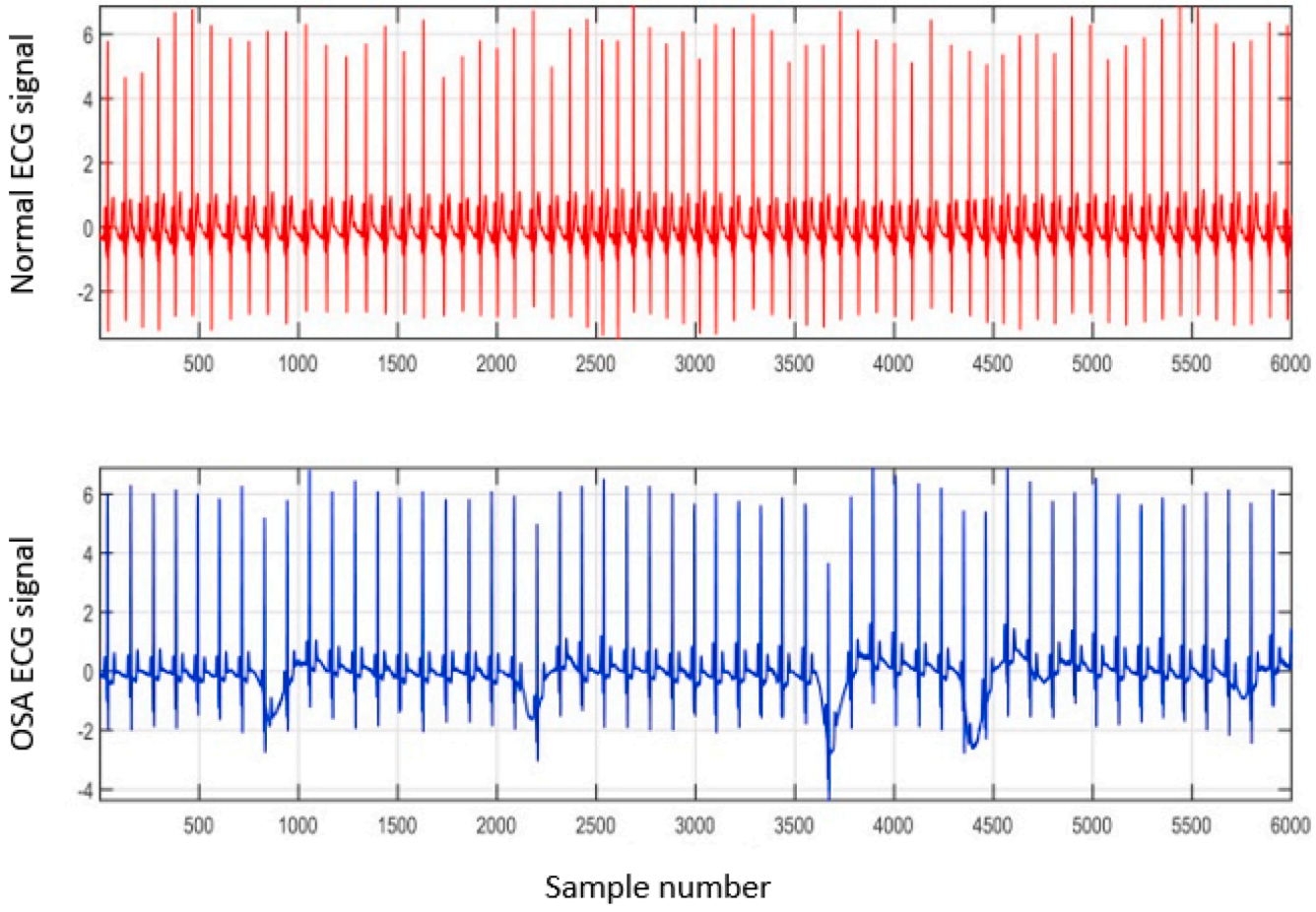

Fig. 1. 1-min ECG signals of an OSA and a normal subject.

diagnose the condition of the heart patient in real-time using ECG signals showing promising results [33–37].

Due to the success of DL methods in different medical problem and large amount of available data, researchers focused on detecting OSA using ECG signals. For instance, Li et al. [38] used SVM classifiers to detect OSA. Sparse auto-encoder method was used to represent features from unlabelled ECG signals. The model achieved an accuracy of 84.7%, which indicated that unsupervised learning can not extract features effectively. Similarly, Feng et al. [39] proposed unsupervised feature learning by using auto-encoder model. They used Hidden Markov Model (HMM) as a classifier, which achieved an accuracy of 85%. Another study by Wang et al. [40] developed CNN model to extract features from R-R intervals and R amplitudes. The model contained two CNN layers, and each CNN layer followed by a max-pooling layer. The model used LeNet-5 as a classifier to detect OSA. The accuracy was reported to be 87.6%.

The main objective of this paper is to develop a novel deep learning model which is a hybrid architecture involving Convolutional Neural Networks (CNNs) and Long Short Term Memory (LSTM) networks. Recently, CNN and LSTM have shown promising results with classification [41–44]. The idea is encouraged by successful application of deep learning models having combination of CNN and LSTM architecture in the detection of Coronary Artery Disease (CAD) and automatic diagnosis of Arrhythmias [45,46]. Similarly, in Speech Recognition and Precipitation Nowcasting problems, the architectures based on CNN and LSTM achieved promising results [43,47]. The proposed model extracts useful features automatically from R-R intervals and R amplitudes of ECG signals without feature engineering.

3. Proposed approach

3.1. Dataset

In this study we used PhysioNet Apnea-ECG database [48] which is available online to detect OSA from ECG signals. The total number of the ECG recordings are 70 which consists of two sets: released set for training and withheld set for testing as advised in the guidelines of dataset. The released dataset contain 35 recordings (a01-a020, b01-b05 and c01-c10) while the withheld dataset contain 35 recordings for testing (x01-x035). The variable length of recordings ranges from 6 to 8 h which were recorded from 32 subjects (25 males and 7 females). The ages of subjects range between 27 and 63 years, and their weights range between 35 and 135 kilograms. The ECG signals were sampled at 100 Hz with 16-bit resolution. Specialists segmented and labelled the ECG recordings into 1-min segment and labelled them as healthy or OSA. The total number of labelled segments for released and withheld sets are 17,125 (6,514 OSA and 10,611 healthy) and 17,303 (6,552 OSA and 10,751 healthy) respectively. Table 1 presents PhysioNet Apnea-ECG database details.

3.2. Preprocessing

The time between two consecutive R peaks is defined as R-R interval, and R peaks amplitude is known as ECG-Derived Respiration (EDR) signal. These studies [20,38,49,50] demonstrated that R-R intervals and R amplitude capture substantial information for the task of detecting OSA events. Fig. 2 shows pre-processing steps of ECG signals as suggested by Wang et al. [40]. We found R-peaks location by using

Table 1
PhysioNet Apnea-ECG database description.

Dataset	OSA	Healthy	The total
Released set	6514	10611	17,125
Withheld set	6552	10751	17,303
The total	13,066	21,362	34,428

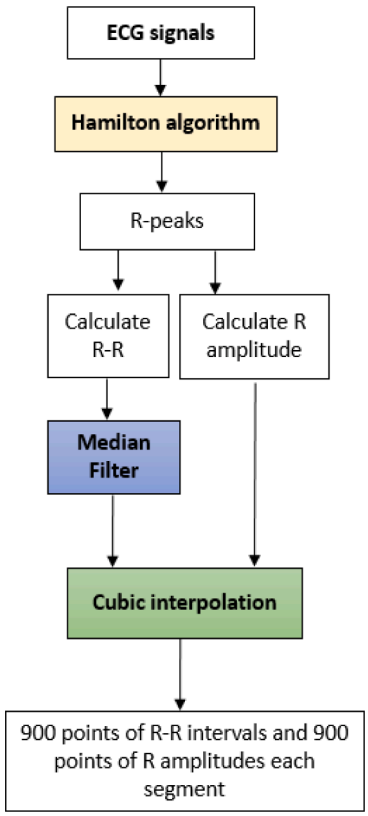

Fig. 2. PhysioNet Apnea-ECG dataset preprocessing.

Hamilton algorithm [51] which is an open source code and able to extract R-peaks values accurately from the noisy ECG signals. We extracted the amplitude of the extracted R-peaks and calculated the intervals between the two R-peaks. A median filter [52] was applied as we expected to have non-interpretable points which were generated from R-R intervals. This provided a robust estimate of the generated value from R-R-intervals. The variations of generated values of R-R intervals led to mark them as a suspect R-R interval which happened because of false R peaks detections, or missed R peaks. Cubic interpolation was used to detect false R peaks by comparing the sum of adjacent R-R intervals with a robust estimate of the generated value from the R-R-intervals. When robust estimation of the generated value from R-R-intervals was implemented, false R peaks were numerically closer to the sum of adjacent R-R intervals than the individual R-R intervals. We obtained 900 points of R-R intervals and 900 points for R amplitudes for each 1-min segment similar to Wang et al. [40]. At the end, the signal was normalised by Z-score as shown in the Eq. (1) [53].

$$z_{score} = \frac{(S - E_S)}{\alpha_S} \tag{1}$$

where E and α are the mean and the standard deviation of the signal S respectively.

The total number of segments after pre-processing are 33,854 with the released set totalling 16,909 min and the withheld set 16,945 min. The total number of healthy and OSA segments are 20,861 and 12,993, respectively. After pre-processing PhysioNet Apnea-ECG database, 574 (1-min segments) are removed due to false R peak detection, which are less than 2% from the dataset at the stage of cubic interpolation. Table 2 shows PhysioNet Apnea-ECG database description after pre-processing.

3.3. Deep Learning model

We propose a deep learning architecture as shown in Fig. 3 to classify

Table 2
PhysioNet Apnea-ECG database description after preprocessing.

Dataset	OSA	Healthy	The total
Released set	6503	10406	16,909
Withheld set	6490	10455	16,945
The total	12,993	20,861	33,854

each segment of the signal as OSA or healthy. The proposed architecture consists of four blocks. The input of ECG signals fed into the first and second block having 1D-CNN, batch normalisation, max-pooling and dropout layers to extract automatic features from ECG signals. The extracted features are input to block 3, which has two consecutive LSTMs and dropout layers. The last block is the output layer which uses two classifier sigmoid and SVM.

The 1D-CNN layer of the first block aims to extract features from the inputs. It is composed of several different filters called *Kernels* of size 10. It is used to calculate the input features in the form of a feature map. Each feature map has neurons which are connected to the neighbour neurons in the previous layer. The mathematical operation of the CNN layer is represented as [54].

$$z^l_{i,j,k} = W^l_k x^l_{i,j} + b^l_k \tag{2}$$

where l is the layer, (i,j) is the location of the feature value in the kth feature map. W and b are the weight vector and bias respectively, and x is the input.

The 64 features map is produced by convolving the inputs with filters. The Nonlinear activation function ReLU is applied to the convolved results. The mathematical operation of ReLU is shown in Eq. (3) [55]:

$$max(0,x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{otherwise} x > 0 \end{cases} \tag{3}$$

where x is the input signal.

The output of the 1D-CNN layer is fed into a batch normalization layer that normalizes the parameters between the layers. Batch normalization can affect the training phase performance by reducing the number of epochs [26]. Then 1D-max pooling layer with 5 parameter is applied to reduce the feature map size. The primary purpose of max-pooling is to reduce the computational cost of the architecture. After that, dropout layer is added to reduce overfitting during the training phase. The process of block 1 is repeated in block 2 with different feature map parameters as listed in Table 3.

The third block has two consecutive LSTM networks with 64 and 128 parameters respectively which is a type of Recurrent Neural Networks (RNN). The main purpose of LSTM is to solve two problems: the gradient vanishing problem and long-term dependency problem. In general, LSTM contains three gate layers: Input Gate layer, Forget Gate layer and Output Gate layer. Forget Gate layer f_t decides to set some information out of the cell state by using a sigmoid layer. The mathematical representation of forget gate layer is presnted in Eq. (4) [56]. The Input Gate layer i_t decides about the new information that will be stored in the cell state C_t as presented in Eq. (5). It has two steps: firstly to determine the values that will be updated by using the sigmoid layer, and secondly to create a vector of new candidate values $C^{\sim}_t$ which will be added to the cell state by utilising a *tanh* layer as shown in Eq. (6). Both of these steps create an update to the old cell state C_{t-1} as shown in Eq. (7) [56]. The Output Gate layer h_t also has two steps, one is to use a sigmoid layer and the other is to use a *tanh* layer as represented in Eq. (9) [56]. The sigmoid layer o_t filters the information in the cell state as mentioned in Eq. (8), while the *tanh* layer normalizes the values in the cell between 1 and −1. Multiplying the result from the sigmoid layer with the *tanh* layer is the result of the output gate layer.

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{4}$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{5}$$

Table 3
Parameters of the proposed deep learning model.

Block	Layer name	Feature map	Activation	Output shape	Parameters
Block1	1D-CNN	64	ReLU	(None, 1991, 64)	1984
	Batch normalization			(None, 1991, 64)	256
	1D-Max-pooling	5		(None, 398, 64)	0
	Dropout	0.2		(None, 398, 64)	0
Block2	1D-CNN	32	ReLU	(None, 389, 32)	20512
	Batch normalization			(None, 389, 32)	128
	1D-Max-pooling	5		(None, 77, 32)	0
	Dropout	0.2		(None, 77, 32)	0
Block3	LSTM	64	Tanh	(None, 77, 64)	24832
	Dropout	0.2		(None, 77, 64)	0
	LSTM	128	Tanh	(None, 20)	6800
	Dropout	0.2		(None, 20)	0
Block4	Output layer	1	Sigmoid	(None, 1)	21

Total parameters: 54,533.
Trainable parameters: 54,341.
Non-trainable parameters: 192.

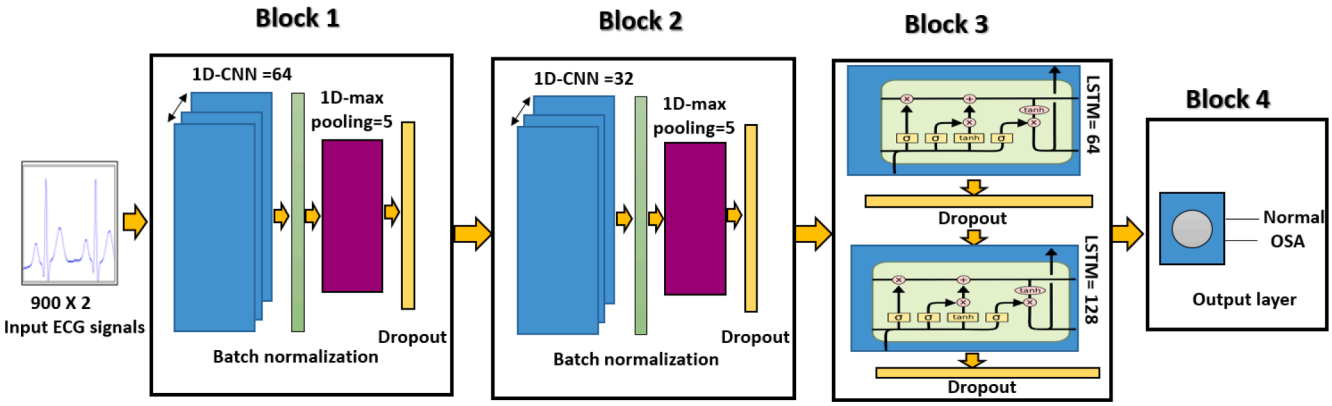

Fig. 3. The architecture of proposed deep learning model.

$$C_t^{\sim} = tanh(W_c.[h_{t-1}, x_t] + b_c) \quad (6)$$

$$C_t = f_t \ * \ C_{t-1} + i_t \ * \ C_t^{\sim} \quad (7)$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t \ * \ tanh(C_t) \quad (9)$$

where h_{t-1} and x_t are the number of hidden units and input features at the time t respectively while W_f, W_i, W_c and W_o are the weights of the inputs and b_f, b_i, b_c and b_o are the bias vectors. σ represents the non-linear hyperbolic function.

The last block in our proposed architecture is the output layer. We used two different classifiers to predict the final result. The first classifier is the fully connected layer with a sigmoid activation function, while the second classifier is a Support Vector Machine (SVM) using a RBF kernel. We integrated SVM classifier with our proposed model by replacing the last layer, which is a fully connected layer with sigmoid function. The final layer classifies a segment to be either OSA or healthy. We employed the grid-search method to set the optimum parameters of SVM with a RBF kernel. The learning rate and the optimum value of the parameters of the two classifiers are presented in Table 4, and batch size selected is 128.

The complexity of the proposed model is dependent on the total number of parameters of each layer. CNN Layers have different numbers of the feature map n, kernel size k, number of channels m and bias b, which are considered as the model complexity of CNN layer. The mathematical operation for the parameters of CNN layer, CNN_p is shown in Eq. (10) as in [57].

$$CNN_p = ((k * m) + b) * n) \quad (10)$$

In our model, Block 1 and Block 2 have CNN layers with 1984 and 20512 parameters respectively. The complexity of batch normalization layer is computed by the equation of batch normalization, $Batch_p$ as presented in Eq. (11), as in [58].

$$Batch_p = 4 * m \quad (11)$$

The total number of parameters of batch normalization in Block 1 and Block 2 are 256 and 128 respectively. The complexity of LSTM layer is calculated by the equation of calculating the parameters of LSTM layer $LSTM_p$ as shown in Eq. (12) as in [59].

$$LSTM_p = 4(nm + n^2 + n) \quad (12)$$

Block 3 has two LSTM layers with parameters 24832 and 6800. The complexity of the dense layer is 21, resulting from the number of channels of the previous layer and bias. The equation for the dense layer $Dense_p$ is presented in Eq. (13) as in [58]. The total number of trainable parameters in our model is 54,341.

$$Dense_p = (m + b) * n \quad (13)$$

4. Evaluation metrics

There are many different metrics to measure the proposed network performance. In this study, we used Sensitivity (SE) or Recall, Specificity (SP), Accuracy (ACC), Matthews Correlation Coefficient (MCC) and F1 measures. These metrics are described in Eqs. (14)–(19).

$$\text{SE} = \frac{TP}{TP + FN} \quad (14)$$

Table 4
The learning parameters of Sigmoid and SVM classifiers.

Classifiers	Learning parameters
Sigmoid classifier	optimizer Adam with learning rate 0.01
SVM classifier	Kernal type = RBF, C = 1, gamma = 0.001

$$\text{SP} = \frac{TN}{TN + FP} \quad (15)$$

$$\text{ACC} = \frac{TN + TP}{TN + TP + FN + FP} \quad (16)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (17)$$

$$\text{F1} = \frac{2(SE \times Precision)}{SE + Precision} \quad (18)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (19)$$

where TP represents the number of True Positives, TN represents the number of True Negative, FN represents the number of False Negatives and FP represents the number of False Positives.

5. Results

The proposed architecture was trained and tested on Intel (R) Core (TM) 3.6 GHz (i7-7700) processor having 8 GB RAM. The architecture was implemented in Python 3.7 with Keras and Scikit-learn libraries. The pre-processing of the input recording takes approximately 30 min depending on the duration of recordings, which is normally more than 8 h in the available dataset. The training of our proposed model took 40 epochs for each fold of 10 k-fold cross-validation where each epoch takes 5 min for training. Once the model is trained then the decisions are taken instantly by the model. However all the reported times depend on the processing capacity and memory of the used computer.

Fig. 4 presents a flowchart of our experiment. In the training phase, we used 10-fold cross validation as suggested in [60]. The released dataset was split into 9 folds for training and determining the classifier parameters, and the remaining last fold for evaluation of the model. The procedure was repeated 10 times, and the overall accuracy of the proposed model was calculated. In the testing phase, we selected the best classifier based on the performance obtained from the training phase and used it to classify withheld set as an independent test to evaluate the performance of the proposed model.

Table 5 presents the results with both classifiers and bold values in the table represent the best results between the two output layer

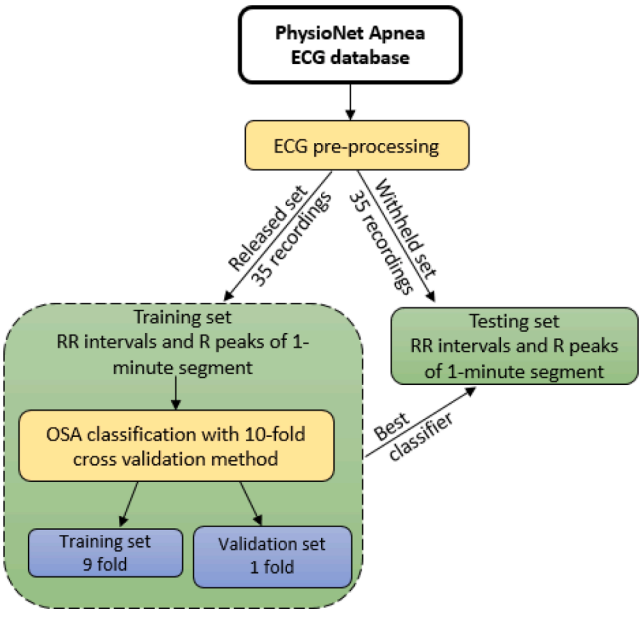

Fig. 4. Flowchart of experimental procedure.

classifiers. The best performance in the training phase was, using SVM as the output classifier, with an accuracy of 94.27%, sensitivity of 94.57%, specificity of 93.98%, F1 of 95.41% and MCC of 88%. Fig. 5 presents the boxplots of the accuracy obtained by the proposed model with sigmoid and SVM classifiers. Fig. 6 shows the accuracy and loss of training and validation set of one fold of 10-fold. The accuracy of the proposed model with SVM output layer ranges from 93.35% to 95%, while with sigmoid output layer ranges from 93.35 to 94.91%. In the testing phase, the performance of the proposed model with SVM classifier is found to be slightly better with an accuracy of 90.92%, sensitivity of 91.24%, specificity of 90.36% F1 92.76% and MCC of 80.67% as compared to sigmoid output layer classifier. The high accuracy for both classifiers show that the proposed model works efficiently in extracting the features from ECG signal which can be easily classified.

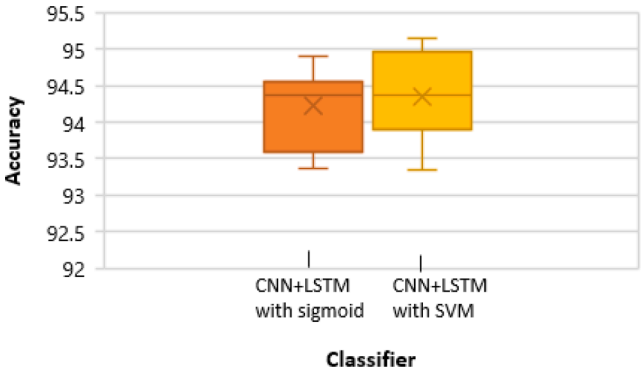

Fig. 5. Boxplots of OSA classification accuracy for two classifiers.

6. Discussion

In this part, we compare our proposed model with the most relevant state-of-the-art research studies which used entire released set for training and withheld set for testing. We acknowledge that other existing research studies also achieved good results with deep learning models. However, due to the difference in the dataset selection size used in their experiments, a comparison of our work with their work is not suitable [61–64]. Our proposed model uses features extracted by a pre-processing method. The pre-processing method to clean the data from noises to extract the accurate values from the raw data without using complicated procedures. Feature engineering is a method used to create features depending on the prior knowledge of signals with complicated feature extraction procedures. When these features are extracted, additional procedures are implied to select important features to include them in their model. Some studies [20,65] used both pre-processing and feature engineering methods in their proposed techniques.

Our proposed architecture achieves promising results with an accuracy of 94.27% with the released training set and 90.92% with the withheld testing set using SVM classifier to classify OSA and healthy segments. Table 6 lists the comparison of our results with the state of the art results with the same dataset. All the presented techniques except Wang et al.'s and Li et al.'s technique [40,66] used different feature engineering techniques which involve wavelet decomposition methods to create features from ECG signals. As mentioned earlier, Feature Engineering process require prior knowledge of OSA and is a time consuming process. Bengio et al. [67] mentioned in their study that machine learning techniques learns the feature patterns better than the traditional Feature Engineering techniques. Therefore, we employed an alternate approach to use deep learning to extract and select important features from ECG signals. The approach is simple and reduce human efforts.

It can be observed from results presented in Table 6 that our proposed model outperforms the state of the art techniques which have been evaluated on PhysioNet Apnea-ECG database except Zarie et al. [10]. However, Zarie et al. [10] used Feature Engineering based on wavelet method which is complicated and computationally expensive technique, and its performance is highly dependant on the selection of the mother wavelet. The results of our proposed model were close to

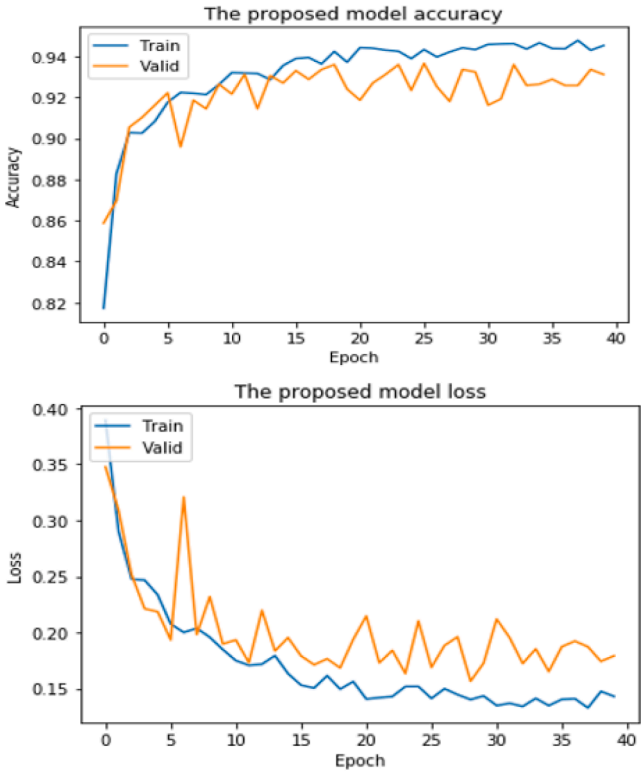

Fig. 6. The accuracy and loss of Training and Valition set of one fold.

Zarie et al.'s [10] results however our F1 score is better which is discussed in detail later. Comparing to state of the art machine learning techniques [38,40] mentioned in the Table 6, our proposed technique achieves better results.

Most of the existing OSA classification techniques in literature presents accuracy as the measuring metric which is based on correct classification of both classes (OSA and healthy). Therefore, the accuracy

Table 5
The results obtained by the proposed deep learning model with two different classifiers.

Evaluation	Proposed model	ACC (%)	SE (%)	SP (%)	F1 (%)	TNS	CCS	MCC(%)
Training	Sigmoid classifier	94.217	94.40	93.93	94.71	16709	15743	87.7
	SVM classifier	**94.27**	**94.57**	**93.98**	**95.41**	**16709**	**15737**	**88**
Testing	Sigmoid classifier	90.88	91.38	90	92.72	16945	15401	80.60
	SVM classifier	**90.92**	**91.24**	**90.36**	**92.76**	**16945**	**15407**	**80.67**

The font bold is the best results.
*ACC = Accuracy, SE = Sensitivity, SP = Specificity, AUC = Area Under the Curve, TNS = Total number of segments, CCS = Correctly classified segments, MCC= Matthews Correlation Coefficient.

Table 6
Comparison between per segment OSA detection performance of our model and existing works.

Reference	Feature (num)	Classifier	ACC (%)	SE (%)	SP (%)	F1 (%)
Feng et al. [39]	Auto encoder	HMM	85	86	84.40	76.6
Li et al. [38]*	Auto encoder	Decision fusion	84.7	88.9	82.1	81
Mendez et al. [70]	Feature Engineering (30)	ANN	88	85	90	–
Nguyen et al. [23]	Feature Engineering (32)	SVM, NN	85.26	86.37	83.47	–
Sharma et al. [24]	Feature Engineering (18)	LS-SVM	83.3	79.5	88.4	–
Sharma et al. [25]	Feature Engineering (36)	LS-SVM	87.5	84.9	88.2	–
Chazal et al. [71]*	Feature Engineering (88)	LD	90	86.4	92.3	87.59
Varon et al. [20]	Feature Engineering (28)	LS-SVM	84.74	84.71	84.69	–
Wang et al. [40]	CNN	LeNet-5	87.6	83.1	90.3	–
Zarie et al. [10]	Feature Engineering (108)	SVM	92.98	91.74	93.75	91
Proposed model	**CNN and LSTM**	**SVM**	**90.92**	**91.24**	**90.36**	**92.76**

The font bold is the best results.
* F1 scores are calculated from the confusion matrix reported in the studies.

metric is dependent on the distribution of the classes where higher represented class contribute more in the final accuracy. Studies suggest that F1 score is a better performance metric for binary classification problem having imbalanced dataset [68,69] because it is the harmonic mean of precision and recall and thus is not affected by imbalanced class distribution. PhysioNet Apnea-ECG dataset is an imbalanced dataset with 6,514 OSA segments and 10,611 healthy segments. OSA segments are representing less than 38% of the entire dataset and accuracy metric is not the best performance metric. We compared the F1 scores of our proposed technique with the available F1 scores of the state-of-the-art techniques. Some studies [20,23–25,40,70,71] did not provide F1 scores in their articles, while few studies [38,71] provided the confusion matrix which helped us to calculate F1 scores using Eq. (18). Our proposed model achieved F1 score of 92.76%, which is the best among the compared models [10,38,39,71].

Noise or artifacts contained in the ECG signal recordings is an issue which lowers the accuracy. These artifacts are introduced due to the body movements during sleep. It is challenging to distinguish these artifacts from the actual signal and leads to inaccurate classification and misdiagnosis. The state of the art technique [10,20] presented in Table 6 removed segments from the dataset which has artifacts. This results in improving the accuracy of their results. In this study, we removed less than 2% of the total number of ECG segments which could not be processed irrespective of the fact whether they had artifacts or not. Therefore, we can observe that our proposed model robustly deals with the noise in signals and performs best.

7. Conclusion

In this paper, we proposed a deep learning architecture based on CNN and LSTM networks to classify OSA and healthy ECG signals. Two classifiers are used at the output layer. Both classifiers performed well which shows that our proposed deep learning architecture works well in extracting feature automatically. The best results were obtained by SVM classifier which achieved the accuracy of 94.27% and 90.92% for training and testing sets respectively. Due to the imbalance class distribution of PhysioNet Apnea-ECG dataset, F1 score is considered to be the optimal parameter in addition to accuracy, sensitivity and specificity results. The results are compared with state of the art techniques and found that the proposed technique performs better than the existing techniques.

PhysioNet Apnoea-ECG dataset labels the signals as apnoea or healthy, while hypopnea events are discarded. In further studies, we will use a different dataset in order to create multi-class classifiers for sleep apnoea classes. Future work may also include data collected by other sensors such as EEG, EMG and SpO2. Also, to improve our model, we suggest to balance the segments of each class by using oversampling techniques such as Synthetic Minority Over-sampling Technique (SMOTE).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W.A. De Backer, Central sleep apnoea, pathogenesis and treatment: an overview and perspective, European Respiratory Journal 8 (8) (1995) 1372–1383.
[2] James M. Parish, Virend K. Somers, Obstructive sleep apnea and cardiovascular disease, in: Mayo Clinic Proceedings, vol. 79, Elsevier, 2004, pp. 1036–1046.
[3] Ahsan H Khandoker, Jayavardhana Gubbi, Marimuthu Palaniswami, Automated scoring of obstructive sleep apnea and hypopnea events using short-term electrocardiogram recordings, IEEE Transactions on Information Technology in Biomedicine 13 (6) (2009) 1057–1067.
[4] J.R.D. Espiritu, Health consequences of obstructive sleep apnea, Journal of Sleep Disorders and Therapy 8 (307) (2019) 2167–0277.
[5] Konrad E Bloch, Polysomnography: a systematic review, Technology and Health Care 5 (4) (1997) 285–305.
[6] Kevin Wise, Electrocardiography (ecg), in: The International Encyclopedia of Communication Research Methods, 2017, pp. 1–15.
[7] D. Puthankattil Subha, Paul K Joseph, Rajendra Acharya, Choo Min Lim, Eeg signal analysis: a survey, Journal of Medical Systems 34 (2) (2010) 195–212.
[8] Wim H.J.P. Linssen, Dick F. Stegeman, Ed. M.G. Joosten, Servaas L.H. Notermans, Martin A. van't Hof, Rob A. Binkhorst, Variability and interrelationships of surface emg parameters during local muscle fatigue, Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine 16 (8) (1993) 849–856.
[9] Matteo D'Aloia, Annalisa Longo, Maria Rizzi, Noisy ecg signal analysis for automatic peak detection, Information 10 (2) (2019) 35.
[10] Asghar Zarei, Babak Mohammadzadeh Asl, Automatic detection of obstructive sleep apnea using wavelet transform and entropy-based features from single-lead ecg signal, IEEE Journal of Biomedical and Health Informatics 23 (3) (2018) 1011–1021.
[11] J. Lee Garvey, Ecg techniques and technologies, Emergency Medicine Clinics 24 (1) (2006) 209–225.
[12] Daniel Álvarez, Roberto Hornero, María García, Félix del Campo, Carlos Zamarrón, Improving diagnostic ability of blood oxygen saturation from overnight pulse oximetry in obstructive sleep apnea detection by means of central tendency measure, Artificial Intelligence in Medicine 41 (1) (2007) 13–24.
[13] Félix del Campo, Roberto Hornero, Carlos Zamarrón, Daniel E Abasolo, Daniel Álvarez, Oxygen saturation regularity analysis in the diagnosis of obstructive sleep apnea, Artificial Intelligence in Medicine 37 (2) (2006) 111–118.
[14] Majdi Bsoul, Hlaing Minn, Lakshman Tamil, Apnea medassist: real-time sleep apnea monitor using single-lead ecg, IEEE Transactions on Information Technology in Biomedicine 15 (3) (2010) 416–427.
[15] Oliver Faust, U. Rajendra Acharya, E.Y.K. Ng, Hamido Fujita, A review of ecg-based diagnosis support systems for obstructive sleep apnea, Journal of Mechanics in Medicine and Biology 16 (01) (2016) 1640004.
[16] Christian Guilleminault, Roger Winkle, Stuart Connolly, Kenneth Melvin, Ara Tilkian, Cyclical variation of the heart rate in sleep apnoea syndrome: mechanisms, and usefulness of 24 h electrocardiography as a screening technique, The Lancet 323 (8369) (1984) 126–131.
[17] Alan Jovic, Nikola Bogunovic, Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features, Artificial Intelligence in Medicine 51 (3) (2011) 175–186.
[18] Leif Sörnmo, Pablo Laguna, Electrocardiogram (ecg) signal processing, in: Wiley Encyclopedia of Biomedical Engineering, 2006.
[19] Philip De Chazal, Maria O'Dwyer, Richard B Reilly, Automatic classification of heartbeats using ecg morphology and heartbeat interval features, IEEE Transactions on Biomedical Engineering 51 (7) (2004) 1196–1206.

[20] Carolina Varon, Alexander Caicedo, Dries Testelmans, Bertien Buyse, Sabine Van Huffel, A novel algorithm for the automatic detection of sleep apnea from single-lead ecg, IEEE Transactions on Biomedical Engineering 62 (9) (2015) 2269–2278.

[21] Saeed Babaeizadeh, David P White, Stephen D Pittman, Sophia H Zhou, Automatic detection and quantification of sleep apnea using heart rate variability, Journal of Electrocardiology 43 (6) (2010) 535–541.

[22] A. Hassan, computer-aided obstructive sleep apnea detection using normal inverse gaussian parameters and adaptive boosting, Biomedical Signal Processing and Control 29 (1) (2016) 22–30.

[23] Hoa Dinh Nguyen, Brek A Wilkins, Qi Cheng, Bruce Allen Benjamin, An online sleep apnea detection method based on recurrence quantification analysis, IEEE Journal of Biomedical and Health Informatics 18 (4) (2013) 1285–1293.

[24] Hemant Sharma, K.K. Sharma, An algorithm for sleep apnea detection from single-lead ecg using hermite basis functions, Computers in Biology and Medicine 77 (2016) 116–124.

[25] Sharma Hemant, Kumar Sharma Kamalesh, Sleep apnea detection from ecg using variational mode decomposition, Biomedical Physics & Engineering Express (2020).

[26] Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT press, 2016.

[27] U Rajendra Acharya, Hamido Fujita, Oh Shu Lih, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network, Information Sciences 405 (2017) 81–90.

[28] U. Rajendra Acharya, Hamido Fujita, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, Application of deep convolutional neural network for automated detection of myocardial infarction using ecg signals, Information Sciences 415 (2017) 190–198.

[29] U. Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Hojjat Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals, Computers in Biology and Medicine 100 (2018) 270–278.

[30] Oh Shu Lih, V. Jahmunah, Tan Ru San, Edward J. Ciaccio, Toshitaka Yamakawa, Masayuki Tanabe, Makiko Kobayashi, Oliver Faust, U. Rajendra Acharya, Comprehensive electrocardiographic diagnosis based on deep learning, Artificial Intelligence in Medicine 103 (2020) 101789.

[31] Sherin Mary Mathews, Dictionary and deep learning algorithms with applications to remote health monitoring systems, PhD thesis, University of Delaware, 2017.

[32] Sherin M. Mathews, Chandra Kambhamettu, Kenneth E. Barner. A novel application of deep learning for single-lead ecg classification, Computers in Biology and Medicine 99 (2018) 53–62.

[33] Rasmus S Andersen, Abdolrahman Peimankar, Sadasivan Puthusserypady, A deep learning approach for real-time detection of atrial fibrillation, Expert Systems with Applications 115 (2019) 465–473.

[34] Dakun Lai, Xinshu Zhang, Bu. Yuxiang, Su. Ye, Chang-Sheng Ma, An automatic system for real-time identifying atrial fibrillation by using a lightweight convolutional neural network, IEEE Access 7 (2019) 130074–130084.

[35] Angshul Majumdar, Rabab Ward, Robust greedy deep dictionary learning for ecg arrhythmia classification, in: 2017 International Joint Conference on Neural Networks (IJCNN), IEEE, 2017, pp. 4400–4407.

[36] Sherin M Mathews, Chandra Kambhamettu, Kenneth E Barner, Maximum correntropy based dictionary learning framework for physical activity recognition using wearable sensors, in: International Symposium on Visual Computing, Springer, 2016, pp. 123–132.

[37] Sherin M. Mathews, Luisa F. Polania, Kenneth E. Barner, Leveraging a discriminative dictionary learning algorithm for single-lead ecg classification, in: 2015 41st Annual Northeast Biomedical Engineering Conference (NEBEC), IEEE, 2015, pp. 1–2.

[38] Kunyang Li, Weifeng Pan, Yifan Li, Qing Jiang, Guanzheng Liu, A method to detect sleep apnea based on deep neural network and hidden markov model using single-lead ecg signal, Neurocomputing 294 (2018) 94–101.

[39] Kaicheng Feng, Hengji Qin, Wu Shan, Weifeng Pan, Guanzheng Liu, A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram, IEEE Transactions on Instrumentation and Measurement (2020).

[40] Tao Wang, Lu. Changhua, Guohao Shen, Feng Hong, Sleep apnea detection from a single-lead ecg signal with automatic feature-extraction through a modified lenet-5 convolutional neural network, PeerJ 7 (2019), e7731.

[41] Colin Lea, Rene Vidal, Austin Reiter, Gregory D. Hager, Temporal convolutional networks: A unified approach to action segmentation, in: European Conference on Computer Vision, Springer, 2016, pp. 47–54.

[42] Jin Wang, Bo Peng, Xuejie Zhang, Using a stacked residual lstm model for sentiment intensity prediction, Neurocomputing 322 (2018) 93–101.

[43] Yue Xie, Ruiyu Liang, Zhenlin Liang, Chengwei Huang, Cairong Zou, Björn Schuller, Speech emotion classification using attention-based lstm, IEEE/ACM Transactions on Audio, Speech, and Language Processing 27 (11) (2019) 1675–1685.

[44] Gan Zhou, Zhi Li, Meng Fu, Yanjun Feng, Xingyao Wang, Chengwei Huang, Sequence-to-sequence load disaggregation using multi-scale residual neural network, IEEE Transactions on Instrumentation and Measurement (2020).

[45] Jen Hong Tan, Yuki Hagiwara, Winnie Pang, Ivy Lim, Shu Lih Oh, Muhammad Adam, Ru San Tan, Ming Chen, U. Rajendra Acharya, Application of stacked convolutional and long short-term memory network for accurate identification of cad ecg signals, Computers in Biology and Medicine 94 (2018) 19–26.

[46] Zhenyu Zheng, Zhencheng Chen, Hu Fangrong, Jianming Zhu, Qunfeng Tang, Yongbo Liang, An automatic diagnosis of arrhythmias using a combination of cnn and lstm technology, Electronics 9 (1) (2020) 121.

[47] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, Wang-chun Woo, Convolutional lstm network: A machine learning approach for precipitation nowcasting, Advances in Neural Information Processing Systems 28 (2015) 802–810.

[48] Thomas Penzel, George B. Moody, Roger G. Mark, Ary L. Goldberger, J. Hermann Peter, The apnea-ecg database. In Computers in Cardiology 2000, vol. 27 (Cat. 00CH37163), IEEE, 2000, pp. 255–258.

[49] Ahsan H Khandoker, Marimuthu Palaniswami, Chandan K Karmakar, Support vector machines for automated recognition of obstructive sleep apnea syndrome from ecg recordings, IEEE Transactions on Information Technology in Biomedicine 13 (1) (2008) 37–48.

[50] Changyue Song, Kaibo Liu, Xi Zhang, Lili Chen, Xiaochen Xian, An obstructive sleep apnea detection approach using a discriminative hidden markov model from ecg signals, IEEE Transactions on Biomedical Engineering 63 (7) (2015) 1532–1542.

[51] Pat Hamilton, Open source ecg analysis, in: Computers in Cardiology, IEEE, 2002, pp. 101–104.

[52] Lili Chen, Xi Zhang, Changyue Song, An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram, IEEE Transactions on Automation Science and Engineering 12 (1) (2014) 106–115.

[53] Sun I. Kim, Tae S. Suh, World Congress of Medical Physics and Biomedical Engineering 2006: August 27-September 1, 20006 COEX Seoul, Korea, vol. 14, Springer Science & Business Media, 2007.

[54] Gu. Jiuxiang, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al., Recent advances in convolutional neural networks, Pattern Recognition 77 (2018) 354–377.

[55] Jian Xue, Jinyu Li, Yifan Gong, Restructuring of deep neural network acoustic models with singular value decomposition, in: Interspeech, 2013, pp. 2365–2369.

[56] Urtnasan Erdenebayar, Yoon Ji Kim, Jong-Uk Park, Eun Yeon Joo, Kyoung-Joung Lee, Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram, Computer Methods and Programs in Biomedicine 180 (2019) 105001.

[57] Kaiming He, Jian Sun, Convolutional neural networks at constrained time cost, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5353–5360.

[58] Antonio Gulli, Sujit Pal, Deep learning with Keras, Packt Publishing Ltd, 2017.

[59] Felix A Gers, Jürgen Schmidhuber, Fred Cummins, Learning to forget: Continual prediction with lstm, Neural Computation 12 (10) (2000) 2451–2471.

[60] Haifa Almutairi, Ghulam Mubashar Hassan, Amitava Datta, Detection of obstructive sleep apnoea by ecg signals using deep learning architectures (accepted), in: European Signal Processing Conference (EUSIPCO, 2020, pp. 1–5.

[61] Maowei Cheng, Worku J. Sori, Feng Jiang, Adil Khan, Shaohui Liu, Recurrent neural network based classification of ecg signal features for obstruction of sleep apnea detection, in: 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), vol. 2, IEEE, 2017, pp. 199–202.

[62] Debangshu Dey, Sayanti Chaudhuri, Sugata Munshi, Obstructive sleep apnoea detection using convolutional neural network based deep learning framework, Biomedical Engineering Letters 8 (1) (2018) 95–100.

[63] Oliver Faust, Ragab Barika, Alex Shenfield, Edward J. Ciaccio, U. Rajendra Acharya, Accurate detection of sleep apnea with long short-term memory network based on rr interval signals, Knowledge-Based Systems (2020) 106591.

[64] Xiaowei Wang, Maowei Cheng, Yefu Wang, Shaohui Liu, Zhihong Tian, Feng Jiang, Hongjun Zhang, Obstructive sleep apnea detection using ecg-sensor with convolutional neural networks, Multimedia Tools and Applications 79 (23) (2020) 15813–15827.

[65] D. Novák, K. Mucha, Tarik Al-Ani, Long short-term memory for apnea detection based on heart rate variability, in: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2008, pp. 5234–5237.

[66] Taiyong Li, Min Zhou, Ecg classification using wavelet packet entropy and random forests, Entropy 18 (8) (2016) 285.

[67] Yoshua Bengio, Aaron Courville, Pascal Vincent, Representation learning: A review and new perspectives, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (8) (2013) 1798–1828.

[68] Shameem Puthiya Parambath, Nicolas Usunier, Yves Grandvalet, Optimizing f-measures by cost-sensitive classification, in: Advances in Neural Information Processing Systems, 2014, pp. 2123–2131.

[69] Roghayeh Soleymani, Eric Granger, Giorgio Fumera, F-measure curves: A tool to visualize classifier performance under imbalance, Pattern Recognition 100 (2020), 107146.

[70] Martin O Mendez, Anna Maria Bianchi, Matteo Matteucci, Sergio Cerutti, Thomas Penzel, Sleep apnea screening by autoregressive models from a single ecg lead, IEEE Transactions on Biomedical Engineering 56 (12) (2009) 2838–2850.

[71] Redmond B. Shouldice, Louise M. O'Brien, Ciara O'Brien, Philip de Chazal, David Gozal, and Conor Heneghan. Detection of obstructive sleep apnea in pediatric subjects using surface lead electrocardiogram features, Sleep 27 (4) (2004) 784–792.