ICIC Express Letters
Part B: Applications
Volume **14**, Number **5**, May **2023**

ICIC International ©2023 ISSN 2185-2766
pp. **467–474**

SLEEP APNEA DETECTION FROM ECG SIGNAL USING GOOGLENET-BILSTM

Aditya Samagan Cahyo Bagaskoro and Sani Muhamad Isa*

BINUS Graduate Program – Master of Computer Science
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
aditya.bagaskoro@binus.ac.id; *Corresponding author: sani.m.isa@binus.ac.id

Received November 2022; accepted January 2023

Abstract. *Sleep apnea is a sleep disorder caused by a failure in the respiratory process. There are two types of sleep apnea: central sleep apnea and obstructive sleep apnea, which are distinguished by respiratory failure. Sleep apnea detection has been a topic of discussion for a long time. Currently, sleep apnea detection uses ECG signals, which are then used for machine learning or deep learning. According to previous research, the best algorithm for detecting sleep apnea is a combination of CNN and LSTM. As a result, this study aims to compare GoogLeNet and GoogLeNet-BiLSTM to see how well the two algorithms perform. The results show that in validation and testing, GoogLeNet got an accuracy of 84.19% and 85.83%, and GoogLeNet-BiLSTM got an accuracy of 83.23% and 85.75%. However, GoogLeNet-BiLSTM outperforms GoogLeNet in terms of the loss function with a difference of roughly 29%-30% for validation and 13%-15% for testing, and the number of epochs required to achieve convergence. So, if you want to get fast and decent results while sacrificing a little accuracy, GoogLeNet-BiLSTM is the best choice; otherwise, GoogLeNet can be used.*

Keywords: Sleep apnea, RR-Interval, R-Peak amplitude, Electrocardiogram signal, Deep learning

1. **Introduction.** Sleep apnea is classified into two types: Central Sleep Apnea (CSA) and Obstructive Sleep Apnea (OSA). CSA is a sleep problem that occurs as a result of the breathing muscles not receiving signals from the brain, and OSA is a prevalent sleep problem characterized by recurring pharyngeal failure [1], yet CSA instances are relatively rare in sleep apnea cases since CSA is frequently encountered in conjunction with OSA cases [2]. The average population percentage of patients with sleep apnea was 17% for women and 22% for men, according to eleven epidemiological studies done between 1993 and 2013 [3]. People with OSA may also have low oxygen saturation. Therefore, every 10% drop in oxygen saturation increases the risk of sudden cardiac death by 14% [4].

Since it is a crucial first step in providing patients with effective therapy for sleep apnea, the accuracy of sleep apnea diagnosis is vital. The "gold standard" of diagnosis for OSA is Polysomnography (PSG), which captures brain waves, heart rate, oxygen levels, and breathing; nevertheless, PSG requires overnight evaluation in a hospital or laboratory [5]. As a result, various research has discovered a novel method for detecting sleep apnea, precisely by utilizing the Electrocardiogram (ECG) signal and its derivatives such as Heart Rate Variability (HRV) or RR-Interval (beat by heartbeat) [6,7], which will subsequently be detected via machine learning or deep learning. Because ECG signals capture variations in the heart's electrical activity during each cardiac cycle, they can be utilized to detect various conditions [8]. These conditions include arrhythmias [9], sinus bradycardia and ventricular flutter [10], sleep apnea, etc.

DOI: 10.24507/icicelb.14.05.467

Previous research has used machine learning and deep learning algorithms such as Linear Discriminant Analysis (LDA) [2], k-Nearest Neighbor (kNN), Naïve-Bayes, and Support Vector Machine (SVM) [6], as well as deep learning or hybrid algorithms such as Convolutional Neural Network (CNN) [11], CNN combined with Long Short-Term Memory (LSTM) [12], CNN-LSTM combined with Deep Neural Network (DNN) [5], and multiscale deep neural network [13]. According to prior studies, the best algorithm for detecting sleep apnea is the CNN-LSTM hybrid algorithm. On the other hand, previous research exclusively used the CNN architecture, which has a small number of layers and a network that is not complicated, deep, or wide.

Therefore, in this paper, we will conduct research to identify sleep apnea by utilizing ECG signals using a hybrid deep learning algorithm, namely the CNN-BiLSTM. The experiment in this study will combine the CNN GoogLeNet architecture with BiLSTM. The usage of GoogLeNet is based on its capabilities, mainly a highly deep or wide network, which indicates that the more profound or broader the network, the better the model's performance. Furthermore, the capability of BiLSTM in handling time-series data, such as ECG signals, influences their utilization. In addition, the GoogLeNet architecture represents a considerable improvement over ZFNet (ILSVRC 2013 Winner) and AlexNet (Winner of ILSVRC 2012), and it also has a reduced error rate compared to VGGNet (Runner-up of ILSVRC 2014). Theoretically, when paired with BiLSTM, it can produce a better result. As a result, the contributions this article may provide are understanding the performance of GoogLeNet architecture or those integrated with BiLSTM.

This paper is categorized as follows: Section 2 describes the related works, Section 3 describes the methodology, Section 4 describes the result and discussion, and Section 5 describes the conclusion.

2. **Related Works.** There have been numerous earlier studies that used ECG signals to detect sleep apnea. Chazal et al. conducted research in 2004 that used Linear Discriminant Analysis (LDA) and reached an accuracy of roughly 90% [2]. However, Chazal et al.'s research has limitations, such as the need for more information regarding the sleep stage in the dataset utilized. Regardless of sleep stage information, Isa et al. did research in 2011 to diagnose sleep apnea using multiple classification algorithms, including k-Nearest Neighbour (k-NN), Naïve-Bayes, and Support Vector Machine (SVM) [6]. Because there is no automated feature selection in machine learning classification, Isa et al. used Principal Component Analysis (PCA) as the feature selection. The accuracy attained from the research was around 70% for k-NN, 70% for Naïve-Bayes, and 78% for SVM.

Following research that employs a machine learning technique to diagnose sleep apnea, Banluesombatkul et al. did a study in 2018 utilizing a deep learning approach that uses CNN-LSTM-DNN and achieved an accuracy of 79.45% [5]. Banluesombatkul et al. developed the architecture comprising three convolution layers, three LSTM layers, and six fully connected layers before reaching the output layers. The activation function utilized in the CNN layer is Rectified Linear Unit (ReLU), and at the fully connected layer, it is hyperbolic tangent (tanh), which then outputs layers using SoftMax.

There is additional research that exclusively utilizes CNN, such as Wang et al.'s research in 2019 [11]. Wang et al. employed the LeNet-5 architecture, which has been modified by changing the convolutional layer from two dimensions to one, adding a dropout layer with a parameter of 0.8 before the fully connected layer, and employing only one fully connected layer with a parameter of ten units, which will then divide the output layer into two classifications. The modified LeNet-5 architecture achieves an accuracy of 87.6%.

To expand knowledge about previous research in research using hybrid algorithms, there is also research conducted by Shen et al. in 2021 [13]. Shen et al. diagnosed sleep apnea using Multiscale Dilation Attention (MSDA)-1DCNN, which subsequently applied a Weighted-Loss Time-Dependent (WLTD) classification. The implementation of dilated convolution is focused on balancing the relationship between parameters and performance, and WLTD is used to overcome imbalanced data and increase the classifier accuracy. The proposed method achieves an accuracy of 89.8%.

After that, in 2021, Faust et al. conducted research simply utilizing the LSTM algorithm [14]. The LSTM architecture consists of 2 forward and backward LSTM layers, and the output is generated using the sigmoid activation function. The obtained accuracy when using held-out data is 81.30%. Bahrami and Forouzanfar conducted research utilizing the LeNet-LSTM hybrid algorithm in the same year [15]. In the performed research, the hybrid algorithm employs LeNet for feature extraction and LSTM for classification. The LeNet-LSTM hybrid algorithm yields an accuracy of 80.67 percent.

The most recent research is a study undertaken by Bahrami and Forouzanfar in 2022 that employs classification utilizing various machine learning and deep learning methods [12]. In contrast to the research conducted by Bahrami and Forouzanfar in 2021, several machine learning classification algorithms were utilized in this work, including LDA, Quadratic Discriminant Analysis (QDA), Logistic Regression (LR), Gaussian Naïve-Bayes (GNB), SVM, and others. Meanwhile, the CNN-LSTM, CNN-BiLSTM, and CNN-GRU hybrid algorithms used in this research include AlexNet, ZFNet, VGG16, and VGG19 CNN architectures. The best algorithm in this research was ZFNet-BiLSTM, which had an accuracy of around 88.13%.

There are several approaches for detecting sleep apnea, ranging from machine learning to hybrid deep learning. The proposed method from earlier research's accuracy and performance is also improving, allowing for further development. Previous studies have shown that CNN-BiLSTM is the best approach, although the design of the CNN utilized does not yet use a complex CNN architecture. As a result, this research will try to use a complex, broad, and deep CNN architecture like GoogLeNet. The study's implementation of GoogleNet is based on research conducted by Kim et al. in 2019 [9]. The issue was raised by Kim et al. regarding the evaluation of ECG signals for arrhythmias using GoogleNet. This study yielded a very high degree of accuracy, almost 98%. Therefore, in order to classify sleep apnea, this study aims to employ either pure GoogleNet or a combination of GoogleNet and BiLSTM.

3. **Methodology.** This section will describe how the research was conducted. K-Fold Cross-Validation will generally be used in each model in the research. Details regarding the data collection, data processing as well as the proposed model's structure are included in this section.

3.1. **Data collection.** The apnea-ECG database (apnea-ecg) [16] was used in this research, and it comprises 70 recordings of ECG signals with durations ranging from 401 to 578 minutes. The ECG signal has a frequency of 100 Hz. The sleep apnea data set includes 57 males and 13 women with various data outcomes. Every minute of this data is divided into one epoch, which has 6000 samples and is annotated. Because the data utilized is from the 2000 challenge, it is separated into two equal parts: data for training and data for testing. Unlike most research, which uses just 35 records data training, the research that will be undertaken will use all existing records, exactly 70 records.

3.2. **Data preprocessing.** In this study, we will employ two data obtained by the extraction procedure from the dataset used, specifically RR-Interval and R-Peak Amplitude. This data has been used since Wang et al.'s research found that the two data were quite useful in detecting sleep apnea [11]. In Bahrami and Forouzanfar's research, it is also mentioned that using either the RR-Interval or R-Peak Amplitude might reduce the model's performance [12].

The extraction process will be carried out utilizing Wang et al.'s study method of utilizing the Hamilton Algorithm to obtain the R-Peak, which will be used to compute

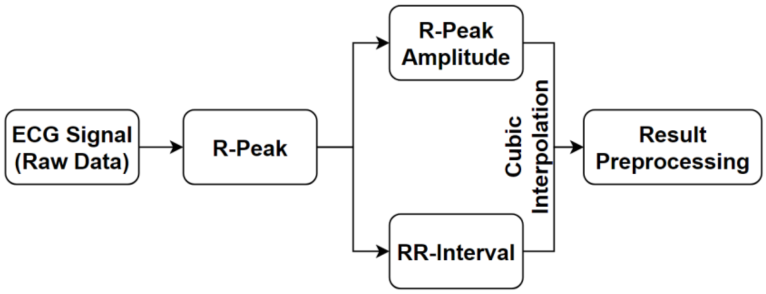

FIGURE 1. Data preprocessing

the distance between the R-Peaks known as the RR-Interval. The two data sets are also adjusted using cubic interpolation [11]. Figure 1 represents the data processing schema used in this study.

3.3. **Proposed model.** A hybrid model, CNN-BiLSTM, with the CNN architecture utilizing the GoogLeNet architecture, will be employed in this study. GoogLeNet is a CNN architecture developed by Szegedy et al. that won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014 classification task [17]. GoogLeNet is a deep CNN architecture with 22 layers, including nine inception layers. Because it is excessively deep, two auxiliary classifiers are used to avoid overfitting and to manage diminishing gradient descent, resulting in GoogLeNet having three outputs. Auxiliary classifier loss is weighed, which means that the estimated loss is multiplied by 0.3.

Afterward, the BiLSTM algorithm will be implemented on last output of the GoogLeNet architecture following the pooling layer. In this research, the parameter value for the hidden unit is 2048. The final step is the dense layer with an activation function, SoftMax, and parameters corresponding to the desired output, which is 2. Table 1 represents the proposed method's detail used in this study.

TABLE 1. Detail of the proposed model for GoogLeNet-BiLSTM

Layer	**Configuration**
Input	Size = (180, 1, 2)
GoogLeNet	Modified all kernel sizes, e.g., (n, n) to (n, 1)
BiLSTM	Filter = 2048
Fully Connected (Output)	Dense = 2, Activation = SoftMax

4. **Result and Discussion.** This section contains detailed information regarding the experimental setup, the results of the experiments, and a discussion of the comparison between previous research and the proposed model.

4.1. **Experimental setup.** To achieve optimal and consistent outcomes in this research, k-Fold Cross Validation with k = 5 will be performed. The most common values for k are 3, 5, and 10. The basic concept of the number k is that a small k is less expensive but more biased, while a big k is more expensive but has a lower bias. Consequently, k = 5 was chosen for this study in an effort to make it cheaper and less biased. In addition, this study only uses a dataset that is neither too huge nor too small. The experiment begins by dividing the dataset into two portions, 70% for training data, 20% for data testing, and 10% for data validation. The parameter optimizer in the model used is ADAM with a loss function using Categorical Cross Entropy (CCE). Then, the epoch for this research is 50, the learning rate is determined using a schedule with a beginning learning rate of 0.001, and every 10 epochs are divided by 10.

4.2. **Result.** Several popularly used metrics from previous research, including accuracy, sensitivity, specificity, and F-Score, were used to evaluate the proposed model's performance. Nevertheless, this study will also include two more metrics involving loss functions that use CCE and precision. The metric loss function is used in the evaluation to determine how effectively the model generates the desired result. According to Demirkaya et al. in 2020, loss functions are easier to train and more generic [18]. For calculating the metric evaluation, the confusion metric containing True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) will be utilized. The following formula will be used to calculate the evaluation metrics:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{2}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{3}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{4}$$

$$F\text{-}Score = 2 * \frac{(Precision * Sensitivity)}{(Precision + Sensitivity)} \tag{5}$$

$$CCE = -\sum_{i=1}^{n} t_i \log(p_i), \text{ for } n \text{ classes} \tag{6}$$

t is the vector of the actual label and p is the probability of the predicted label in Equation (6) regarding CCE. Table 2 shows the outcomes of the experiments conducted on the proposed models, consisting of performance metrics between validation (Val.) and testing (Test.). In validation, the performance of the traditional GoogLeNet method beats the performance of the GoogLeNet-BiLSTM hybrid algorithm in metrics other than loss function.

Table 2. Result experiment of CNN GoogLeNet architecture

Name	Type	Accuracy	Sensitivity	Specificity	Precision	F-Score	CCE
GoogLeNet-BiLSTM	Val.	±83.23%	±75.53%	±87.88%	±79.02%	±77.23%	±0.54
	Test.	85.75%	80.97%	88.79%	82.07%	81.51%	0.46
GoogLeNet	Val.	±84.19%	±77.40%	±88.30%	±80.03%	±78.68%	±0.77
	Test.	85.83%	81.31%	88.69%	82.00%	81.65%	0.54

GoogLeNet-BiLSTM and GoogLeNet on validation differed in accuracy, sensitivity, specificity, precision, and F-Score by 0.96%, 1.87%, 0.42%, 1.01%, and 1.45%, respectively. However, in testing, GoogLeNet-BiLSTM outperforms the evaluation metrics of precision, specificity, and loss function with the difference in accuracy, sensitivity, specificity, precision, and F-Score, respectively are 0.08%, 0.34%, 0.1%, 0.07%, and 0.14%.

In the metric loss function, GoogLeNet-BiLSTM outperforms GoogLeNet by roughly 29%-30% for validation and 13%-15% for testing. According to Figure 2(a) loss function curve, GoogLeNet-BiLSTM performs better than GoogLeNet because, as can be observed from the loss function curve shown, GoogLeNet-BiLSTM achieves its lowest value and convergence first.

Additionally, by examining the accuracy curve from Figure 2(b), it is still possible to conclude that GoogLeNet-BiLSTM and GoogLeNet both achieved competitive accuracy throughout the epoch. Because GoogLeNet-BiLSTM can provide results with a minimal

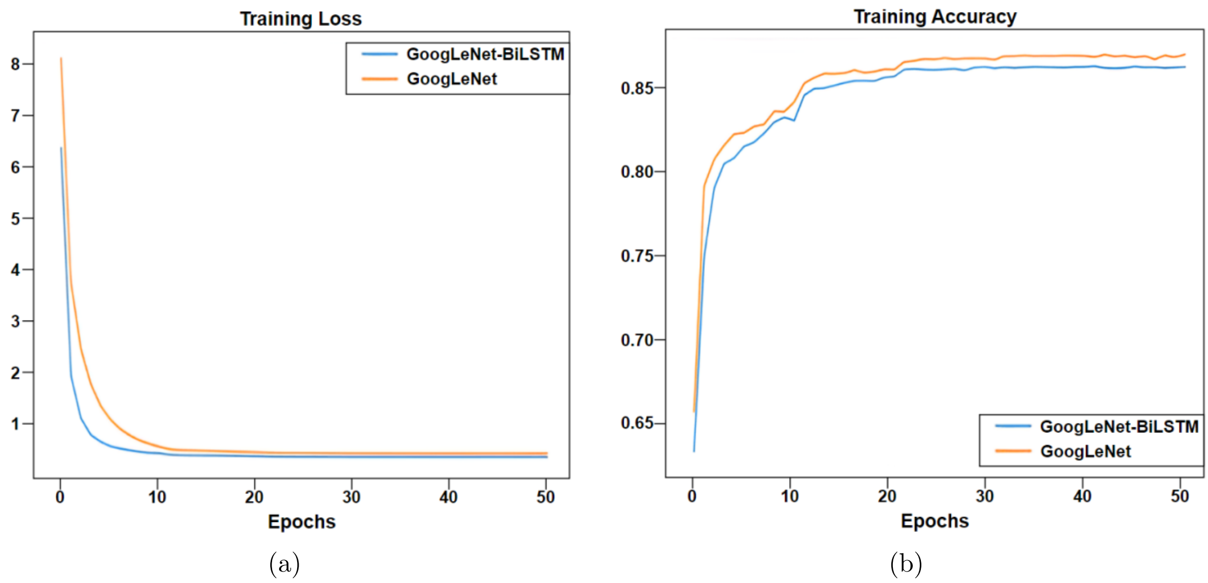

FIGURE 2. Loss function curve (a) and accuracy curve (b) from training GoogLeNet-BiLSTM and GoogLeNet

error rate and accuracy comparable to GoogLeNet with fewer epochs, it can also be faster at providing appropriate results.

4.3. **Discussion.** Because there are several approaches for detecting sleep apnea, this research will also compare it to models given by other researchers whose models are still relatively new. The research that became the comparison was likewise chosen based on using the same dataset. The comparison of the proposed model in this research with state-of-the-art detection of sleep apnea may be shown in Table 3.

TABLE 3. Comparison of GoogLeNet-BiLSTM and GoogLeNet with state-of-the-art sleep apnea detection

Reference	Model	Accuracy	Sensitivity	Specificity	Description
Faust et al. [14]	LSTM	81.30%	59.90%	91.75%	Using 35 of the 70 records and no mention of the metric loss function.
Shen et al. [13]	MSDA-1DCNN + WLTD	89.80%	89.10%	89.40%	Using 35 records for training and 35 records for validation, but no mention of the metric loss function.
Bahrami and Forouzanfar [12,15]	LeNet-BiLSTM	79.98%	68.38%	87.11%	Using 35 records for training and 35 records for testing, but no mention of the metric loss function.
	ZFNet-BiLSTM	88.13%	81.49%	92.27%	Using 70 records with k-Fold Cross Validation, but no mention of the metric loss function.
Our model	GoogLeNet-BiLSTM	85.75%	80.97%	88.79%	–
	GoogLeNet	85.83%	81.31%	88.69%	–

Based on Table 3, accuracy and sensitivity from GoogLeNet-BiLSTM and GoogLeNet are higher than LSTM and LeNet-BiLSTM, but lower than MSDA-1DCNN + WLTD and ZFNet-BiLSTM with specificity just higher than LeNet-BiLSTM. However, it should be noted that just one study, which used the ZFNet-BiLSTM model, showed experimental setup similarities. The ZFNet-BiLSTM model makes use of all accessible records (70 in total) and utilizes k-Fold Cross Validation with k = 5 with only 35 records from the available 70 records or 70 records divided equally into 35 training and testing records were employed in research with other models.

One critical point is missing from all existing papers: the loss function evaluation metric. Knowing the error rate of the model used is very helpful in determining whether the model is feasible or not because accuracy alone is insufficient.

5. **Conclusions.** It can be concluded that in validation, GoogLeNet performs exceptionally well in metrics other than loss functions, such as accuracy, sensitivity, specificity, precision, and F-Score from GoogLeNet-BiLSTM. However, in testing, GoogLeNet-BiLSTM can outperform GoogLeNet in metrics of precision, specificity, and loss function. Therefore, GoogLeNet-BiLSTM is the best option if you want the lowest loss function possible and be faster at providing decent results because there are only slight differences with GoogLeNet.

Due to the deep design of GoogLeNet alone, overfitting is potential when paired with BiLSTM. Therefore, some adjustments must be made so that there is a possibility of reducing the performance of GoogLeNet-BiLSTM, and that is the limitation of this study. As a result, in the future, we can look for ways to improve the performance of GoogLeNet-BiLSTM.

REFERENCES

[1] D. P. White, Sleep apnea, *Proc. Am. Thorac. Soc.*, vol.3, no.1, pp.124-128, 2006.

[2] P. de Chazal, T. Penzel and C. Heneghan, Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram, *Physiol. Meas.*, vol.25, no.4, pp.967-983, 2004.

[3] K. A. Franklin and E. Lindberg, Obstructive sleep apnea is a common disorder in the population – A review on the epidemiology of sleep apnea, *J. Thorac. Dis.*, vol.7, no.8, pp.1311-1322, 2015.

[4] J. N. Blackwell, M. Walker, P. Stafford, S. Estrada, S. Adabag and Y. Kwon, Sleep apnea and sudden cardiac death, *Circ. Reports*, vol.1, no.12, pp.568-574, 2019.

[5] N. Banluesombatkul, T. Rakthanmanon and T. Wilaiprasitporn, Single channel ECG for obstructive sleep apnea severity detection using a deep learning approach, *TENCON 2018-2018 IEEE Region 10 Conference*, pp.2011-2016, 2018.

[6] S. M. Isa, M. I. Fanany, W. Jatmiko and A. M. Arymurthy, Sleep apnea detection from ECG signal: Analysis on optimal features, principal components, and nonlinearity, *2011 5th International Conference on Bioinformatics and Biomedical Engineering*, pp.1-4, 2011.

[7] T. Penzel, J. McNames, P. De Chazal, B. Raymond, A. Murray and G. Moody, Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings, *Med. Biol. Eng. Comput.*, vol.40, no.4, pp.402-407, 2002.

[8] W. Zhang, P. Ju, T. Li, B. Du, X. Lang, W. Wang and J. Wu, Research on risk management methods in big data based on statistic process control and deep learning, *International Journal of Innovative Computing, Information and Control*, vol.18, no.6, pp.1815-1828, doi: 10.24507/ijicic.18.06.1815, 2022.

[9] J.-H. Kim, S.-Y. Seo, C.-G. Song and K.-S. Kim, Assessment of electrocardiogram rhythms by GoogLeNet deep neural network architecture, *J. Healthc. Eng.*, vol.2019, 2019.

[10] H. Ueno, Y. Totoki and T. Matsuo, ECG characterization of sinus bradycardia and ventricular flutter using malthusian parameter and recurrence plot, *ICIC Express Letters, Part B: Applications*, vol.9, no.1, pp.23-30, 2018.

[11] T. Wang, C. Lu, G. Shen and F. Hong, Sleep apnea detection from a single-lead ECG signal with automatic feature-extraction through a modified LeNet-5 convolutional neural network, *PeerJ*, vol.7, e7731, 2019.

[12] M. Bahrami and M. Forouzanfar, Sleep apnea detection from single-lead ECG: A comprehensive analysis of machine learning and deep learning algorithms, *IEEE Trans. Instrum. Meas.*, 2022.

[13] Q. Shen, H. Qin, K. Wei and G. Liu, Multiscale deep neural network for obstructive sleep apnea detection using RR interval from single-lead ECG signal, *IEEE Trans. Instrum. Meas.*, vol.70, pp.1-13, 2021.

[14] O. Faust, R. Barika, A. Shenfield, E. J. Ciaccio and U. R. Acharya, Accurate detection of sleep apnea with long short-term memory network based on RR interval signals, *Knowledge-Based Syst.*, vol.212, 106591, 2021.

[15] M. Bahrami and M. Forouzanfar, Detection of sleep apnea from single-lead ECG: Comparison of deep learning algorithms, *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp.1-5, 2021.

[16] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger and J. H. Peter, The apnea-ECG database, *Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163)*, pp.255-258, 2000.

[17] C. Szegedy et al., Going deeper with convolutions, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-9, 2015.

[18] A. Demirkaya, J. Chen and S. Oymak, Exploring the role of loss functions in multiclass classification, *2020 54th Annual Conference on Information Sciences and Systems (CISS)*, pp.1-5, 2020.