Jurnal Ilmiah KURSOR
Menuju Solusi Teknologi Informasi
Vol. 11, No. 4, Desember 2022

ISSN 0216 – 0544
e-ISSN 2301– 6914

SLEEP DISORDER IDENTIFICATION FROM SINGLE LEAD ECG BY IMPROVING HYPERPARAMETERS OF 1D-CNN

[a]Iman Fahruzi

[a]Department of Electrical Engineering, Politeknik Negeri Batam, Indonesia
E-mail: iman@polibatam.ac.id

Abstract

Disruption of the flow of breathing during sleep will result in significant heart problems if not treated seriously. An electrocardiogram (ECG) recording is one of the most used methods for detecting sleep disorders early on. An ECG is a representation of electrical activity in the heart while it is beating. The irregularities of the morphology and the complexity of the recordings have clinical significance that can be used as a tool for diagnosing sleep disorders. This study uses engineering to obtain features from ECG recordings that are carried out automatically using deep learning machine learning with a Convolutional Neural Network (CNN) model approach. The ECG recordings were processed to remove noise before being used in the CNN model. Tests are carried out on the most optimal model to get good accuracy by applying two scenarios. The test results of the two scenarios show that scenario one has an accuracy of 83.03% compared to scenario two with an accuracy of 76.88%. Meanwhile, the precision, sensitivity, cohens kappa and ROC UAC levels were 81.78%, 87.78%, 65.73% and 82.68% in scenario one testing on the CNN model with the most optimal parameter settings, respectively.

Key words: sleep disorder, sleep diagnosing, heart problems, cnn, ecg.

INTRODUCTION

The presence of snoring even if they do not feel tired all day, feel tired every time they wake up, are not refreshed, wake up feeling tired, have a headache when they wake up, have a dry mouth, feel sleepy during the day, and their ability to concentrate is reduced are all symptoms that are common in people with sleep disorders [1], [2].

According to previous study, humans spend roughly a third of their lives sleeping [3]. Sleep apnea is the most frequent sleep problem that many people suffer from. Sleep apnea is divided into three types based on the diagnosis: Obstructive Sleep Apnea (OSA) or Obstructive Sleep Apnea Hypopnea Syndrome (OSAHS), Central Sleep Apnea (CSA), and Mixed Sleep Apnea (MSA) [4]. These sleep apnea categories can be identified by observing the frequency of occurrence of the Apnea/Hypopnea Index (AHI) per hour. The criteria for AHI scores based on the standard are AHI classification < 5 indicated as normal OSA condition, AHI between 5-14 as mild OSA, AHI between 15-30 as moderate OSA, and AHI > 30 classified as severe OSA [5]. Furthermore, the prevalence of OSA is believed to be 9 percent to 24 percent of the entire population, with up to 90% of cases going misdiagnosed [6].

The dependence on human needs and actions is a major concern in studies on sleep disorders. Visual identification of sleep interruptions for very large datasets is one of them, which will make diagnosis more difficult. The inspection process is inefficient and prone to misdiagnosis since it takes so long. ECG readings are commonly utilized to detect computerized sleep disorders [7]. As a result, the feasibility of software and hardware is a concern to be developed in order for it to be utilized anytime, anyplace, and by anybody without much intervention from third parties. Currently, two methodologies are employed to extract characteristics from ECG recordings: manual extraction and automatic extraction. The difference between the two methods has an impact on the final process by simplifying the procedure and improving the classification process's performance. Increasing classification accuracy was initially extensively demonstrated by using various extraction methodologies, with a recent approach based on images with one minute ECG recordings [8]. Sleep problems can be predicted using ECG Scalograms and Spectrograms [9], or the current conventional approach which is better known as handcraft features, one of which is by increasing the frequency features of Spectrograms [10], and diagnosis that grounded in the complexity of recordings based on fractal analysis of a signal [11].

The diagnosis of sleep disorders from ECG records continues to advance the technique of diagnosing sleep disorders. In recent years, research focusing on enhancing recognition performance through deep learning-based feature engineering processes has begun to supplant handcraft feature engineering techniques. The manual feature engineering procedure necessitates a significant amount of time and resources. Computational results are frequently suboptimal since many features are useless and have no bearing on the identification or classification results.

This study introduces a novel way to engineering features by utilizing one-dimensional of single-lead ECG recordings for automatic identification of sleep disorder through deep learning engineering techniques with the Convolutional Neural Network (CNN) model approach. Single lead means using the electrodes in only one particular position on the body. Furthermore, this one-dimensional recording is trained utilizing two scenarios to achieve the best outcomes by adjusting a number of factors. Finally, the employment of deep learning-based engineering techniques with the appropriate parameter settings can improve performance throughout the identification process.

MATERIAL AND METHODS

Dataset

The present study reports on testing a deep learning model to detect sleep disorders using the ECG-apnea dataset sourced from the physionet database [12]. Recording duration ranges from 7 to 10 hours. This study used ECG recordings that were segmented into 10 seconds and 20 seconds. The dataset used was then normalized aiming to improve performance by accelerating the training process and not always relying on dropout parameters to reduce overfitting. The use of Batch Normalization (BN) basically produces a distribution of data points for each sample with the best standard normal distribution through a linear transformation process. The implementation of BN on an artificial intelligence network was developed by Ioffe and Szegedy [13].

Preprocessing on Baseline Drift from Single Lead ECG

Preprocessing has a pivotal role to obtain accuracy in detecting the R wave without being affected by various kinds of noise. Mathematically, the ECG signal *y[n]* is defined by referring to Equation 1. The variable *s[n]* is a representation of the ECG signal and *w[n]* is the value obtained from the overall noise that affects the ECG signal. Whilst α is the signal attenuation parameter.

$$y[n] = \alpha\, s[n] + w[n] \tag{1}$$

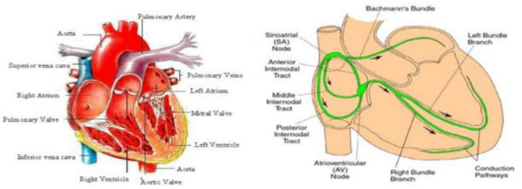
Overall, complex QRS is a wave whose shape does not change despite a lot of noise. When recording an ECG signal, in general, the frequency range for obtaining the main information is from 200-500Hz. In this section, no high frequency is required to get the right information. Specifically, the frequency range of QRS waves is between 5 to 25 Hz [14]. While the other waves have a frequency range below the frequency range of the QRS wave. The following are sources of noise in the ECG signal:

Baseline drift or baseline wander: This noise effect produces a low-frequency variation from the ECG baseline. This type of noise severely limits the usefulness of the recorded EKG. Hence, it needs to be reduced for better clinical evaluation. Several methods have been developed for noiseless ECG elevation, including the use of the wavelet method and empirical mode decomposition [15]. This filtering process can reduce high-frequency noise and baseline drift simultaneously without losing important information from the ECG signal.

Power line interference (PLI): is a significant source of noise during bio-potential measurements. The interference is at a frequency of 50 to 60 Hz. This condition reduces signal quality and makes it difficult to obtain small features that may be important and have clinical information for diagnosis.

Electrocardiogram Signal Morphology

ECG is a very common instrument used to detect the heart's electrical activity over time. The heart is enclosed by a membrane of cells, each of which contains a charge that gets disconnected during each heartbeat. Figure 1 shows the anatomy and electrical system of human heart. The cardiac cycle refers to a complete heartbeat from the beginning to the start of the next beat and has a series of waves labeled P, QRS, and T as shown in Figure 2. Such definition of each recording in the ECG is as follows:



(a) (b)

Fig 1. (a) Anatomy of the human heart; (b). Electrical system of the human heart [16]

P wave is a process of atrial depolarization. When the valves between the atria and ventricles open, 70% of the blood in the atria is extracted by the ventricles as they expand. The contractions applied by the atria are required for the next 30%. Accordingly, the work carried out by the muscles requires only low tension;

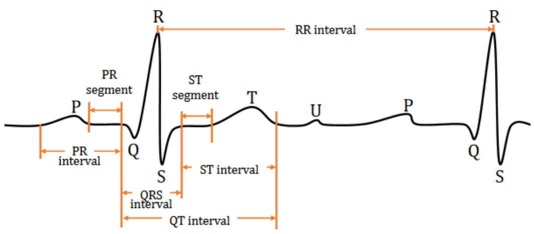


Fig 2. Characteristics waveform of P, QRS dan T [16]

- PQ segment represents the stage before the initial contraction where blood moves into the ventricles;
- QRS wave represents ventricular depolarization by ventricular muscle contraction;
- The ST segment connects the QRS wave and T wave. This segment is the interval between ventricular depolarization and repolarization whereby irregularity will lead to infarction or myocardial ischemia;
- The T wave corresponds to the final phase of the ventricular muscle cell potential action. The same potential handles contraction and repolarization, but one is an upstroke and the other is a downstroke. For this reason, T waves may be associated with unspecified events.

To get the value of the RR interval as a basis for calculating other parameters as a representation of HRV, each beat on the ECG recording displays the electrical activity of the heart. In the complex R waveform, QRS is always pointing upwards and therefore is usually used to identify each beat and its duration which is called the RR interval (from one R to successive R waves) [16]. Figure 2 shows that two R beats produce one RR interval, and so on to get the average value of the RR interval.

Characteristics of Heart rate Variability (HRV)

Normal biological systems exhibit heart rate patterns with complex variability and can be explained mathematically. This variability describes the change in the time interval between successive heartbeats called interbeat intervals (IBIs). In this section, an overview of changes in heart rate oscillations will be described based on the HRV time domain, namely the frequency domain and the nonlinear metric domain[17]. A time domain will measure the number of HRV observed over a certain period. Changes in the frequency domain to determine the absolute or relative value of signal energy in the component spectrum band range. Meanwhile, nonlinear measurements are used to measure the uncertainty and complexity of the observed series of IBIs. HRV measurements examine how these RR intervals evolve over time [18].

Heart rate is the number of heartbeats counted per minute. An HRV is basically a fluctuation in the time interval between adjacent heartbeats. High HRV values are not always better because pathological conditions can produce HRV. When there is an abnormality in heart function, it increases the HRV measurement. It is strongly associated with an increased risk of death. Therefore, an in-depth examination of ECG morphology and periodic patterns of ECG signals can reveal whether the increase in HRV values is caused by problems such as breathing problems due to obstruction during sleep.

Modern Deep Learning Method for Classification of Sleep Disorders

In this study, one-lead ECG recordings were also tested on modern machine learning such as deep learning, specifically using the Convolutional Neural Network (CNN) architecture with a framework as shown in Figure 3. The main objective was to test the different stages and classification results based on the features that previously have been extracted. In the era of artificial intelligence in recent years, it has developed rapidly and has become a solution in solving various cases, including those related to image processing, signal processing, text processing, voice recognition, video processing, and case resolution in the biomedical field. The growing use of deep learning increases the capabilities of computing devices, as well as the development of algorithmic capabilities, such as the implementation of activation functions, weight initialization schemes, optimizations schemes, batch normalization, and depth-wise separable convolutions.

RESULT AND DISCUSSION

In this section, the use of deep learning is to get the best accuracy results based on a one-dimensional CNN architecture (1D-CNN) on a dataset that has been segmented by 10 seconds and 20 seconds. The purpose of this division is to find out which dataset has the best results based on settings on several parameters to improve the performance of the selected CNN model. The number of records used is adjusted to the scenario to be applied. Testing according to this scenario is based on the number of samples, optimization of parameters, and implementation of different architectures.

Scenario 1: This Schema Uses a 10 sec Dataset

Prior to use, the dataset is preprocessed to reduce the effect of noise. Thereafter, the preprocessed dataset is normalized using zscore to obtain a feature value scaling between 0 and 1, but without destroying or losing information because it is

still within its limits. The proposed model is shown in Figure 4. The main objective of the test is to find out whether CNN is able to classify apnea and non-apnea with better accuracy results. Parameter settings are carried out in each scenario starting from testing using standard settings to adjusting parameters. Parameter settings are repeated until the best accuracy results are obtained during the training, validation, and testing processes. To see performance measures during the training process, it can be a graph in the form of accuracy and errors during the process. The goal is to analyze and evaluate a model, if the results are less promising, the model can be improved and parameter settings can be made until promising results are obtained with the appropriate model. Furthermore, if CNN model has been selected, then, the next step is to test it with new data as testing data. The composition of the dataset is divided into three parts, i.e. training data set with 38,344 samples (80%), validation data for 3634 samples and test data set with 8925 samples (20%).

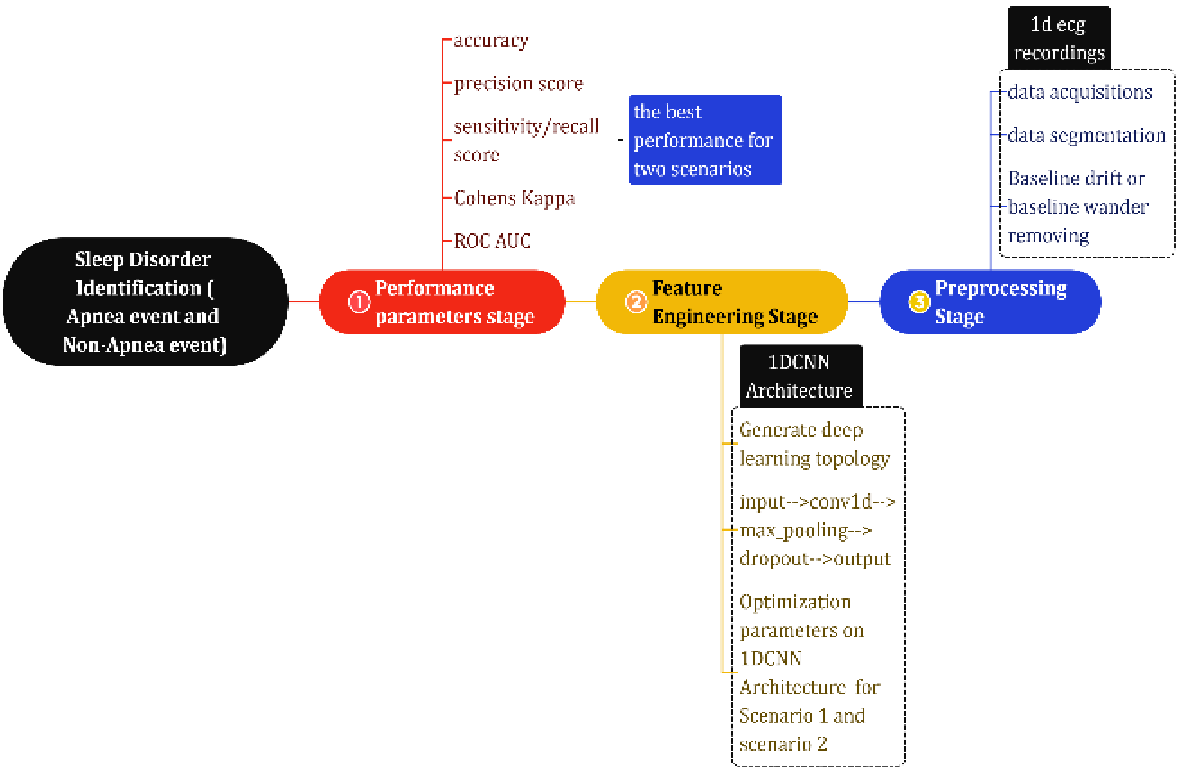


Fig 3. Deep learning framework for identification sleep disorder.

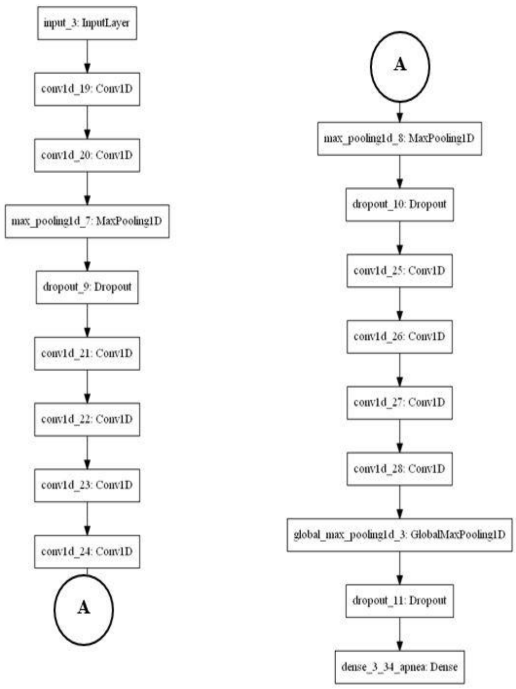


Fig 4. 1D-CNN architecture on 10 second dataset

Test data contains a completely separate data set and is not part of the training data. This is to avoid biased results and do not reflect the superiority of a model over a new data set. However, before carrying out testing process, the selected model is validated using validation data taken from training data. The purpose of model validation is to ensure that the architecture is able to produce good performance without overfitting.

The implementation of 1D-CNN uses a configuration consisting of a convolutional layer (conv layer), max-pooling, dropout, and fully-connected (FC) with different amounts. All convs use Relu activation function, but the output part of the FC uses sigmoid activation. In the compilation section, we use sparse_categorical_crossentropy as a loss function because it is used to classify two classes. Meanwhile, for CNN parameter optimization, we use Adam's optimization function of 0.0001 with 50 epochs and batch size=32.

Table 1 is the CNN architecture settings for scenario 1. Grounded in the summary of the model, a more complete architecture with more layers and the appropriate parameter settings results in a better performance measure than scenario 2.

The CNN architecture of scenario 1 uses 655,730 configurable parameters. Table 2 shows the results of the CNN scenario 1 performance measurement with an accuracy of 83.03%, a precision score of 81.78%, a sensitivity/recall score of 87.78%, a Cohens

Kappa score of 65.73%, and a ROC AUC calculation of 82. ,68%.

Table 1. Here are the parameters of 1D-CNN model for scenario 1

Layer	Filter/nodes/neurons	Kernel	Activation
Input layer(1000x1)	-	-	-
conv1d (334x256)	256	8x8	ReLU
conv1d (334x128)	128	8x8	ReLU
Max pooling_1 (167x128)	128	2x2	-
Dropout_1(167 x128)	128	-	-
conv1d (167x128)	128	7x7	ReLU
conv1d (167x64)	64	7x7	ReLU
Max pooling_1 (83x64)	64	2x2	-
Dropout_1(83x 64)	64	-	-
conv1d (83x64)	64	6x6	ReLU
conv1d (83x128)	128	6x6	ReLU
Max pooling_1 (41x128)	128	2x2	-
Dropout_1(41x 128)	128	-	-
conv1d (41x128)	128	5x5	ReLU
conv1d (41x64)	64	5x5	ReLU
Max_pooling_1 (20x64)	64	-	-
Dropout_1(20x 64)	64	-	-
conv1d (20x64)	64	5x5	ReLU
conv1d (20x64)	32	5x5	ReLU
GlobalMax_ pooling_1(20x3 2)	32	-	-
Dropout_1(20x 32)	32	-	-

Table 2. Performance results of 1D-CNN model for scenario 1 and scenario 2

Performance parameters	Scenario 1	Scenario 2
accuracy (%)	83,03	76,88
precision score (%)	81,78	78,17
sensitivity/recall score (%)	87,78	80,34
Cohens Kappa (%)	65,73	53,17
ROC AUC (%)	82,68	76,50

Scenario 2: This Schema Uses a 20 sec Dataset

In line with the implementation of scenario 1, the dataset used has been preprocessed. The test uses an architecture as shown in Figure 5. This test uses a dataset with divisions, i.e. 12,903 samples of training data, 2581 samples of validation data which are part of the training data, and 3141 samples of test data. Each data has 2000 samples or a duration of 20 seconds. The result of the model construction shows that the parameters that can be set are 168,892.

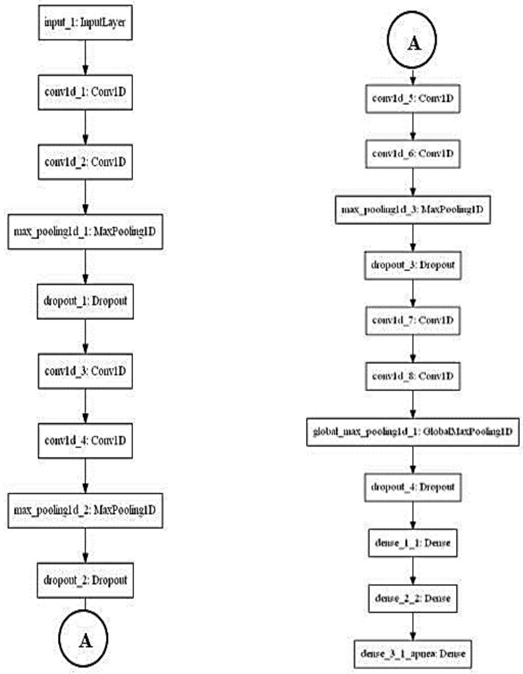


Fig 5. CNN architecture on 20 second dataset

Hereafter, the parameter setting for model optimization is using the Adam function with a learning rate of 0.001 and the number of epochs = 100. Meanwhile, the architectural construction for scenario 2 is translated into a summary model as shown in Table 3.

Table 3. Here are the parameters of 1D-CNN model for scenario 2

Layer	Filter/nodes/neurons	Kernel	Activation
Input layer(2000x1)	-	-	-
conv1d (2000x50) x 2	50	7x7	ReLU
Max pooling_1 (1000x50)	-	2x2	-
Dropout_1(100(x50)	-	-	-
conv1d (1000x50) x 2	50	7x7	ReLU
Max pooling_1 (500x50)	-	2x2	-
Dropout_1(500: 50)	-	-	-
conv1d (500x64 x 2	64	5x5	ReLU

Layer	Filter/nodes/neurons	Kernel	Activation
Max pooling_1 (250x64)	-	2x2	-
Dropout_1(250x 64)	-	-	-
conv1d (250x128) x 2	128	3x3	ReLU
GlobalMax_ pooling_1(128)	-	-	-
Dropout_1(128)	-	-	-

Based on the test, the application of scenarios in each tested model produces different accuracy. This is due to many factors, including the number of data and the number of samples, dataset complexity, data normalization and standardization, different architectures, and parameter optimization. The task of classifying, apnea and non-apnea of the ECG recording is a challenging one. However, what is required is the right architecture configuration with various scenarios such as setting the learning rate, the use of the adaptive gradient method, a large-scale distributed training process, and the use of iteration algorithms that do not converge quickly during the training process and model testing. Parameter setting is undertaken by repeated testing until the most optimal results are obtained. Changes to one parameter do not necessarily produce the best accuracy. In one condition the results of the settings require a very long computation time. For instance, setting the learning rate is very important to optimize inasmuch as it can improve CNN performance, but if the value is too large it will make CNN work harder to extract information so that the process is unstable. On the other hand, if it is too small, the learning process will take longer.

In addition, the architectural complexity and the number of datasets also affect CNN's performance. Scenario 1 produces better accuracy than scenario 2 because the number of datasets is larger, the number of layers is more complex, and the parameter settings are more in line with the characteristics of the data. Furthermore, to get better results, such corrective actions are needed as increasing the amount of data using several schemes, e.g. augmenting the dataset using the oversampling method and Synthetic Minority Over-sampling Technique (SMOTe) which maximizes the minority class to generate synthetic data. In addition, it is also necessary to try other normalization and standardization techniques such as maximum absolute scaling, min-max scaling, and robust scaling so as to overcome differences in feature values to the same scale by re-scaling the dataset. Figure 6 shows the curve of the relationship between accuracy and error performance. The lower the error level will result in better accuracy. In order to maintain the shape of the curve with high accuracy and low error level, attention must be paid to parameter settings and data validation so that there is no overfitting, whereby performance during training is better than performance during testing.

This is because the model is not able to recognize the test data optimally so that the level of accuracy does not increase at the same time the error rate is still high.

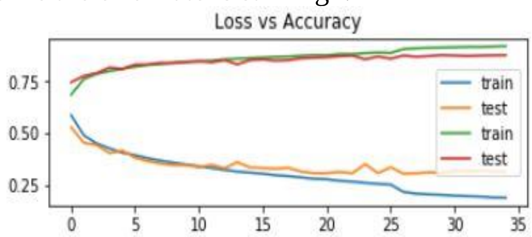


Fig 6. Relationship between accuracy and error performance.

CONCLUSION

In this work, the testing shows that the CNN model with parameter settings according to the first scenario has the highest performance value compared to the second scenario. Scenario 1 sequentially has detection performance with accuracy, precision, sensitivity, Cohens kappa, and ROC UAC of 83.03%, 81.78%, 87.78%, 65.73%, and 82.68%. To increase accuracy in deep learning, additional datasets with sufficient balance of training data, validation data, and test data for both apnea and non-apnea classes are required. In addition, deep learning requires innovation through transfer learning techniques that function to complete a specific task using part or all of a model that was pre-trained on a different task. This approach is expected to improve the performance of the classification method. The use of large datasets in deep learning causes more attention to datasets, including how to add data, how to make existing data more abundant, how to rescaling data, how to normalize data, and how to select data. The scenario in the dataset aims to increase accuracy without overfitting.

REFERENCES

[1] M. F. AkŞahİn, B. Oltu, and B. K. Karaca, "Sleep apnea detection using blood pressure signal," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, May 2018, pp. 1–4.

[2] An American Academy of Sleep Medicine Review, M. S. Wise, D. L. Arand, R. R. Auger, S. N. Brooks, and N. F. Watson, "Treatment of Narcolepsy and other Hypersomnias of Central Origin," *Sleep*, vol. 30, no. 12, pp. 1712–1727, Dec. 2007.

[3] A. R. Fekr, M. Janidarmian, K. Radecka, and Z. Zilic, "Respiration Disorders Classification With Informative Features for m-Health Applications," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 3, pp. 733–747, May 2016.

[4] "The Pathogenesis of Central and Complex Sleep Apnea | SpringerLink." https://link.springer.com/article/10.1007/s11910-022-01199-2 (accessed May 31, 2022).

[5] D. W. Hudgel, "Sleep Apnea Severity Classification — Revisited," *Sleep*, vol. 39, no. 5, pp. 1165–1166, May 2016.

[6] M. Singh, P. Liao, S. Kobah, D. N. Wijeysundera, C. Shapiro, and F. Chung, "Proportion of surgical patients with undiagnosed obstructive sleep apnoea," *Br. J. Anaesth.*, vol. 110, no. 4, pp. 629–636, Apr. 2013.

[7] X. Lin *et al.*, "Contactless sleep apnea detection in snoring signals using hybrid deep neural networks targeted for embedded hardware platform with real-time applications," *Biomed. Signal Process. Control*, vol. 77, p. 103765, Aug. 2022.

[8] S. M. I. Niroshana, X. Zhu, K. Nakamura, and W. Chen, "A fused-image-based approach to detect obstructive sleep apnea using a single-lead ECG and a 2D convolutional neural network," *PLOS ONE*, vol. 16, no. 4, p. e0250618, Apr. 2021.

[9] H. Nasifoglu and O. Erogul, "Convolutional Neural Networks based OSA Event Prediction from ECG Scalograms and Spectrograms," In Review, preprint, Apr. 2021.

[10] Y. Wu, X. Pang, G. Zhao, H. Yue, W. Lei, and Y. Wang, "A novel approach to diagnose sleep apnea using enhanced frequency extraction network," *Comput. Methods Programs Biomed.*, vol. 206, p. 106119, Jul. 2021.

[11] S. Chatterjee, S. Pratiher, and R. Bose, "Multifractal detrended fluctuation analysis based novel feature extraction technique for automated detection of focal and non-focal electroencephalogram signals," *IET Sci. Meas. Technol.*, vol. 11, no. 8, pp. 1014–1021, 2017.

[12] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, Jun. 2000.

[13] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *ArXiv150203167 Cs*, Mar. 2015, Accessed: Aug. 06, 2020. [Online].

[14] M. Elgendi, B. Eskofier, S. Dokos, and D. Abbott, "Revisiting QRS Detection Methodologies for Portable, Wearable, Battery-Operated, and Wireless ECG Systems," *PLoS ONE*, vol. 9, no. 1, p. e84018, Jan. 2014. 10.1371/journal.pone.0084018.

[15] M. Blanco-Velasco, B. Weng, and K. E. Barner, "ECG signal denoising and baseline wander correction based on the empirical mode decomposition," *Comput. Biol. Med.*, vol. 38, no. 1, pp. 1–13, Jan. 2008.

[16] R. Gutiérrez-Rivas, "Real-time early detection of allergic reactions based on heart rate variability," Doctoral thesis, University College Cork, 2016. Accessed: Sep. 12, 2020. [Online].

[17] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Front. Public Health*, vol. 5, Sep. 2017.

[18] S. Rezaei, S. Moharreri, S. Ghiasi, and S. Parvaneh, "Diagnosis of sleep apnea by evaluating points distribution in poincare plot of RR intervals," in *2017 Computing in Cardiology (CinC)*, Sep. 2017, pp. 1–4.