Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:110
https://doi.org/10.28919/cmbn/8840
ISSN: 2052-2541

DEEP LEARNING-BASED SLEEP APNEA DETECTION USING SINGLE-LEAD ECG SIGNALS FROM THE PHYSIONET APNEA-ECG DATABASE

PANDU WICAKSONO*, REZKI YUNANDA

Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480

Abstract: Sleep apnea, a significant medical concern affecting many individuals, is the focal point of this research, which investigates the potential of machine learning and deep learning techniques, particularly in detecting Obstructive Sleep Apnea (OSA). The study's results highlight the effectiveness of models like 1D-CNN in processing electrocardiogram (ECG) signals without requiring manual feature extraction. Notably, our 1D-CNN model achieves an impressive accuracy of 88.36% using 6,000 features, demonstrating strong recall and F1 score performance. Conversely, our Random Forest (RF) model attains a commendable accuracy of 82.23% with just seven features, showcasing high precision and F1 score. However, the K-Nearest Neighbors (KNN) model, characterized by high precision, displays lower recall and specificity, indicating a propensity to classify all data as positive cases. These findings underscore the potential of machine learning and deep learning techniques to enhance sleep apnea detection, offering valuable insights for the diagnosis and management of this critical medical condition.

Keywords: 1DCNN; deep learning; ECG; machine learning; sleep apnea.

2020 AMS Subject Classification: 92C50.

---

*Corresponding author
E-mail address: pandu.wicaksono005@binus.ac.id
Received August 15, 2024

1. INTRODUCTION

Sleep apnea is a pathological disease that causes disturbances in regular breathing patterns during sleep. People afflicted with this ailment may encounter frequent disruptions in their sleep due to difficulties breathing. There exist three distinct categories of sleep apnea: obstructive sleep apnea (OSA), central sleep apnea (CSA), and mixed sleep apnea (MSA). Obstructive sleep apnea (OSA) is the predominant sleep condition, predominantly affecting elderly adults and those who are overweight. Polysomnography (PSG) is the most reliable technique for identifying sleep apnea. However, acquiring this data is time-consuming [1] and expensive, ranging from $3000 to $6000 for each test [2].

The health industry has experienced substantial breakthroughs in recent years as a result of the rapid development of artificial intelligence (AI). As a result, we have integrated AI into various facets of life, including healthcare. The integration of artificial intelligence (AI) into the healthcare sector, specifically in the field of sleep apnea diagnosis, has the potential to provide a cost-effective and effective method for resolving a variety of concerns. Several studies have investigated the identification of sleep apnea using various sophisticated machine learning and deep learning methods. Traditional machine learning algorithms such as Support Vector Machines (SVM) [3] and k-Nearest Neighbors (kNN) [4], as well as deep learning models like Convolutional Neural Networks (CNN) [5], Long Short-Term Memory (LSTM) [6], and Artificial Neural Networks (ANN) [7], exhibit strong performance in sleep apnea detection due to their ability to effectively handle binary classification tasks with two distinct classes: apnea and normal.

The early identification of obstructive sleep apnea (OSA) has the potential to significantly impact mortality rates and mitigate the cost of expensive medical treatments. In their study, Sheta et al. [2] developed a novel Computer-aided Diagnosis (CAD) approach aimed at efficient detection of Obstructive Sleep Apnea (OSA). Computer-aided design (CAD) incorporates several models derived from machine learning (ML) and deep learning (DL) methodologies. The CNN+LSTM model achieves the highest possible accuracy, with a training accuracy of 90.75% and a validation accuracy of 86.25%. Banluesombatkul et al. [8] attempted to develop a filter during the preprocessing phase. Filtering is a commonly employed technique aimed at eliminating distracting noise. In the context of the present investigation, the MrOS sleep study dataset (Visit 1) was utilized to assess the correctness of the filtering process, yielding a value of 79.45%.

Deep learning algorithms are subject to significant demands in correctly detecting sleep apnea. The model must consistently deliver excellent and dependable performance when carrying out its responsibilities. By using datasets that professionals have gathered and annotated, it is possible to train models to achieve this goal. The performance of ML and DL models can be influenced by carefully selecting key features. Previous studies have utilized unprocessed electrocardiogram signals, video data [9], picture data [10], and other relevant data sources.

In their study, Almazaydeh et al. [11]employed ten manually derived features from the PhysioNet Apnea-ECG dataset. These features comprised seven primary features [12] and three additional features [13]. To classify the data, they utilized an SVM algorithm, which yielded a maximum accuracy rate of 96.5%.

In addition, Feng et al. [14] engage in manual feature extraction by extracting the RR interval from the PhysioNet Apnea-ECG dataset. The researchers employed SVM-HMM, a combination of Support Vector Machine (SVM) and Hidden Markov Model (HMM), which yielded an accuracy of 84.70% in their studies.

Almutairi et al. [15] employ the RR interval and QRS amplitude as input features in their research. They explore the effectiveness of several models, including 1D-CNN, 1D-CNN+LSTM, and 1D-CNN+GRU. The 1D-CNN+LSTM model demonstrates the highest level of accuracy, with an accuracy rate of 89.11%.

Hassan et al. [16] introduce an alternative approach for feature extraction from electrocardiogram (ECG) signals, employing the Tunable-Q Factor Wavelet Transform (TQWT). The ECG data is segmented per minute and subsequently subjected to TQWT processing. In addition, the AdaBoost algorithm was employed to identify sleep apnea, resulting in an accuracy rate of 87.33%.

Mukherjee et al. [17] conducted a study where they manually extracted three features, namely RRI, EDR, and RAMP. These features were then utilized in conjunction with the ensemble learning approach and the 1D-CNN model [15], [18], [19]. The study reported an accuracy of 85.58%, which was the highest attained accuracy.

The objective of this study is to investigate the potential of deep learning and machine learning models in addressing challenges within the healthcare industry. Specifically, the research aims to accomplish two main goals: (i) developing a reliable sleep apnea detection model by leveraging existing ECG signal data and (ii) conducting a comparative analysis of different models using a dataset comprising 6000 features and 7 ECG signal features. This study employs many models, including 1D-CNN, LSTM, Random Forest, KNN, and Decision Tree.

This research provides the following contributions:

- This study focuses on the implementation and assessment of deep learning and machine learning techniques using pre-processed feature sets.
- This study aims to compare and validate several models that have been trained using electrocardiogram (ECG) data from a homogeneous population.

2. METHODS

2.1. DATASET

In this study, single-lead electrocardiogram signals were taken from the PhysioNet Apnea-ECG [20] databases. The Apnea-ECG database consists of 70 recordings, with each recording being sampled at a frequency of 100 Hz. Out of these records, 35 are specifically labeled as training records, while the remaining 35 are set aside as test records. The duration of each recording displays variety, spanning from 7 to 10 hours. The study included a cohort of 35 individuals, comprising 30 males and 5 females, aged between 27 and 63. In addition, the apnea ECG recording is annotated at intervals of 60 seconds. A clinical specialist conducted the annotations, precisely identifying obstructive sleep apnea, central apnea, and hypopnea, both with and without arousal. Researchers commonly utilize these databases to detect sleep apnea. This study used a dataset that included various data samples, specifically a01-a20, b01-b05, and c01-c10 from the Apnea-ECG dataset.

2.2. FILTERING

Table 1. Variations of ECG Filters

Filter Name	Freq.	Description
Bandpass	5-11 Hz	Eliminate noise from the raw signal.
Notch	60 Hz	Eliminate noise from the raw signal.
Bandpass second order	5 and 35 Hz	Eliminate noise from the raw signal.
Butterworth filter	0.7 Hz	Eliminate noise from the respiratory signal.

During this phase, the dataset will go through a filtering process. Filtering is employed to eliminate unwanted noises or disturbances from the recorded audio. Table 1 presents many examples of filters that apply to electrocardiogram data [21]. The present study employed a second-order Butterworth Bandpass filter, characterized by a lowpass frequency of 5 Hz and a highpass frequency of 35 Hz. The following procedures were implemented:

- Apply a filtering process to all the records.
- The recorded data is partitioned into 60-second segments, as designated by the expert's labeling.
- Each segment of the Apnea-ECG dataset consists of a maximum of 6000 data points.
- Find the peak in each segment.
- Correct the peak on every segment.

Following the segmentation process, the dataset produced 17,062 segments. The segment will be partitioned into two categories, namely normal and apnea data, according to the assigned label. Once the filtering procedure has been finalized, the subsequent feature extraction step will be done manually.

2.3. FEATURE EXTRACTION

The signal is partitioned into segments of 60 seconds, which are subsequently subjected to filtering in the earlier stage, followed by feature extraction. Feature extraction is used to get helpful information from the electrocardiogram signal. It can effectively capture the unique characteristics of apnea or its absence. The electrocardiogram signal will be analyzed, and seven distinct features will be extracted, as outlined in Table 2.

Table 2. Extracted Features

Name	Description
Total Peaks	The number of peaks observed during a time frame of 60 seconds.
AvgHR	The average heart rate every minute.
MeanNN	The mean of the RR intervals.
RMSSD	The square root of the average of the cumulative differences between consecutive RR intervals.
pNN50	The ratio of RR intervals that exceed 50ms to the overall count of RR intervals.
Age	The age of each patient.
Gender	The gender of each patient.

$$AvgHR = \sum_{h=1}^{h} \frac{1}{h} \tag{1}$$

$$MeanNN = \frac{\sum_{r=1}^{n_r} d_{r+1} - d_r}{n_r} \tag{2}$$

$$RMSSD = \sqrt{\frac{(d_{rr})^2}{n_{r-1}}} \tag{3}$$

$$pNN50 = (\forall(n_r)\ (NN50{+}{+}) \leftarrow \sum_{r=1}^{n_r} d_{r+1} - d_r > 50ms) \times 100 \tag{4}$$

Where:

- h = Number of heart rate
- n_r = Number of r peaks
- $=NN50 = \left(\forall (n_r)(NN50{+}{+}) \leftarrow \sum_{r=1}^{n_r} d_{r+1} - d_r > 50ms\right)$
- $d_{rr} = \sum_{r=1}^{n_r} d_{r+1} - d_r$

Features such as ‘total peaks’, ‘avgHR’, ‘meanNN’, ‘RMSSD’, and ‘pNN50’ are extracted from each segment. The parameters ‘age’ and ‘gender’ are allocated to each segment based on the information provided in the dataset.

2.4. DATA SPLITTING

The data splitting process starts with a distribution ratio of 70:15:15, wherein 70% is allocated for training purposes, 15% is designated for validation, and the remaining 15% is utilized as test data. The purpose of this separation scheme is to effectively segregate testing and validation data to prevent any leakage of information during the training of the model. The test data is employed to evaluate the model's performance once the training phase has been completed. At the same time, the validation data is utilized to optimize and fine-tune the training process.

2.5. Classification

The dataset that was previously divided will be utilized for training and testing the classifier model. The total number of training, validation, and testing data was 11,943, 2,559, and 2,560, respectively. Then, selecting models considered appropriate for sleep apnea data problems will be undertaken.

The 1D-CNN approach put forth by Almutairi et al. [15] and Wang et al. [18] is where the model developed in this study finds its inspiration. Notably, this study introduces slight adjustments to specific parameters. Nevertheless, the findings suggest that the model proposed by Almutairi et al. is better suited for pre-processed data in the preceding step, leading to the decision to employ this model. The other classifier models used for comparison are LSTM, RF (Random Forest) [22], KNN [4], and Decision Tree (DT) [23].

- 1D-CNN: By increasing the convolutional layers and neurons, Almutairi et al. [15] inspired the development of this 1D-CNN model.
- LSTM: The LSTM model uses four layers with 64, 128, 32, and 16 neurons.

All machine learning models (RF, KNN, and DT) use default parameters provided by the library; for 1D-CNN and LSTM models, use ReLu and Adam Optimizer with a learning rate of 0.0005.

2.6. Performance Measures

This experiment utilizes five metrics, specifically precision, recall, f1-score, accuracy, and specificity, to evaluate the model's performance. Given the goal of this categorization job, it is evident that there are two clearly defined categories: positive and negative. A sample is judged true when its projected class matches the actual class and false otherwise.

- Accuracy is defined as the proportion of accurate predictions, both positive and negative, in relation to the total amount of data.

$$Accuracy= \frac{TP+TN}{TP+FP+TN+FN} \tag{5}$$

- Precision is defined as the proportion of accurate positive predictions in relation to the total number of positive outcomes projected.

$$Precision= \frac{TP}{TP+FP} \tag{6}$$

- Recall is the quotient obtained by dividing the number of correct positive predictions by the total number of actual positive data points.

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

- Specificity is defined as the accuracy of forecasting negative data in relation to all negative data.

$$Specificity = \frac{TN}{TN+FP} \tag{8}$$

- The F1-Score is a statistic that combines precision and recall into a single value.

$$F1\text{-}Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{9}$$

3. RESULT AND DISCUSSION

The results of this research are divided into two categories: using 6,000 features and 7 features. We use binary cross-entropy as the loss function in the training process. The results achieved during model training and tested using test data are presented in Table 3. The models showing better performance are highlighted in bold, as seen in Table 3.

Table 3. Comparisson of the proposed model with other previous studies.

Model	Dataset	Accuracy (%)	Recall (%)	Specificity (%)
CNN-LSTM [2]	Apnea-ECG	86.25	88.79	-
CNN-LSTM [17]	MrOS sleep study (Visit 1)	84.08	82.94	86.15
AlexNet CNN + Decision Fusion [24]	Apnea-ECG	86.22	90.00	83.8
SVM-HMM [14]	Apnea-ECG	84.70	68.80	**94.50**
Our Proposed (RF) (7 Features)	Apnea-ECG	82.23	85.14	77.35
Our Proposed (1D-CNN) (6000 Features)	Apnea-ECG	**88.36**	**91.36**	83.75

The 1D-CNN model achieved the best accuracy of 88.36% while utilizing 6,000 features, whereas the RF model achieved the highest accuracy of 82.23% with just seven features. The 1D-CNN model has a higher recall and F1-score, which indicates that this model is good at recognizing most positive cases. Accuracy and precision are also quite good, and high specificity indicates the model's ability to classify the negative (normal) class. Meanwhile, the RF (Random Forest) model has high precision and an F1-score, as well as good accuracy. This model also has a good balance between precision and recall, as well as an excellent ability to classify positive (apnea) and negative (normal) classes.

The KNN model was found to have a very high precision value but a low recall value and low specificity in an experiment employing 6,000 features. This demonstrates that this model has a propensity to classify all data as positive, producing high precision but poor predictive quality. The KNN model, which should not be employed with high-dimensional data, results in a 0% specificity value. In large dimensions, it is difficult to distinguish between points that are close to and those that are far from a particular query point [25].

This study finds that deep learning models perform better when using many features, whereas machine learning models are better suited for employing numerous features. This makes feature selection crucial to determine whether the features employed can accurately depict the characteristics of apnea symptoms.

The model's performance developed in this study will be evaluated by comparing it to several references from earlier studies. The findings from Table 4 demonstrate that our suggested model outperforms other prior research regarding accuracy and recall.

Table 4. Comparison of the proposed model with other previous studies.

Model	Accuracy (%)		Recall (%)		Precision (%)		Specificity (%)		F1-Score (%)	
	6,000 Features	7 Features	6,000 Features	7 Features	6,000 Features	7 Features	6,000 Features	7 Features	6,000 Features	7 Features
1D-CNN	**88.36**	76.02	**91.36**	76.61	89.63	**88.05**	**83.75**	74.56	**90.49**	81.93
LSTM	69.69	76.21	73.80	77.55	78.94	86.53	61.68	73.24	76.28	81.79
RF	64.14	**82.23**	65.43	**85.14**	88.93	86.27	57.42	**77.35**	75.39	**85.71**
KNN	61.72	74.65	61.74	79.99	**99.94**	78.62	0	66.40	76.33	79.30
DT	56.29	78.91	65.00	82.59	63.31	82.59	43.14	72.79	64.15	83.01

Our Proposed Model (1D-CNN) (6000 Features) on the Apnea-ECG dataset dominates in terms of the highest accuracy and recall due to the use of 1D-CNN architecture, which is very effective in extracting features from long sequential data. With 6000 features generated, this model can identify very subtle patterns in sleep data, which supports better classification.

4. CONCLUSION

This study demonstrates that machine learning and deep learning models have high efficacy in predicting obstructive sleep apnea. Models such as one-dimensional convolutional neural networks (1D-CNN) are well-suited for processing signals with a duration of one minute, as they eliminate the need for manual feature extraction. This can be explained by the fact that deep learning models can naturally find and understand the most critical features of the electrocardiogram data on their own. In the context of sleep apnea detection, machine learning models like Random Forest (RF) exhibit a high degree of suitability. These models are particularly effective when employed with manually retrieved and selected features.

The 1D-CNN and RF models consistently demonstrate strong performance on both data sets, which include 6000 and 7 features, respectively. This constant performance underscores the robustness of these models in effectively handling varying numbers of features. Moreover, it is seen that the LSTM and Decision Tree models exhibit superior performance when applied to the dataset containing seven features. This implies that employing dimensionality reduction techniques could enhance the model's efficacy. In the context of the given datasets, it can be observed that the K-nearest neighbors (KNN) model has favorable recall rates but at the expense of compromised accuracy and specificity. This implies that additional refinements may be necessary for KNN to yield more optimal outcomes.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] F.Z. Göğüş, G. Tezel, S. Özşen, et al. Identification of apnea-hypopnea index subgroups based on multifractal detrended fluctuation analysis and nasal cannula airflow signals, Traite. Signal 37 (2020), 145–156. https://doi.org/10.18280/ts.370201.

[2] A. Sheta, H. Turabieh, T. Thaher, J. Too, M. Mafarja, M.S. Hossain, S.R. Surani, Diagnosis of Obstructive Sleep Apnea from ECG Signals Using Machine Learning and Deep Learning Classifiers, Applied Sciences 11 (2021) 6622. https://doi.org/10.3390/app11146622.

[3] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, in: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM, Pittsburgh Pennsylvania USA, 1992: pp. 144–152. https://doi.org/10.1145/130385.130401.

[4] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inform. Theory 13 (1967), 21–27. https://doi.org/10.1109/tit.1967.1053964.

[5] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015), 436–444. https://doi.org/10.1038/nature14539.

[6] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

[7] F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the brain, Psychol. Rev. 65 (1958), 386–408. https://doi.org/10.1037/h0042519.

[8] N. Banluesombatkul, T. Rakthanmanon, T. Wilaiprasitporn, Single channel ECG for obstructive sleep apnea severity detection using a deep learning approach, in: TENCON 2018 - 2018 IEEE Region 10 Conference, IEEE, Jeju, Korea (South), 2018: pp. 2011–2016. https://doi.org/10.1109/TENCON.2018.8650429.

[9] A.S. Rajawat, O. Mohammed, P. Bedi, FDLM: Fusion deep learning model for classifying obstructive sleep apnea and type 2 diabetes, in: 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), IEEE, Palladam, India, 2020: pp. 835–839. https://doi.org/10.1109/I-SMAC49090.2020.9243553.

[10] U. Erdenebayar, Y.J. Kim, J.U. Park, et al. Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram, Computer Methods Progr. Biomed. 180 (2019), 105001. https://doi.org/10.1016/j.cmpb.2019.105001.

[11] L. Almazaydeh, K. Elleithy, M. Faezipour, Detection of obstructive sleep apnea through ECG signal features, in: 2012 IEEE International Conference on Electro/Information Technology, IEEE, Indianapolis, IN, USA, 2012: pp. 1–6. https://doi.org/10.1109/EIT.2012.6220730.

[12] P. de Chazal, T. Penzel, C. Heneghan, Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram, Physiol. Meas. 25 (2004), 967–983. https://doi.org/10.1088/0967-3334/25/4/015.

[13] B. Yılmaz, M.H. Asyalı, E. Arıkan, et al. Sleep stage and obstructive apneaic epoch classification using single-

lead ECG, BioMed. Eng. OnLine 9 (2010), 39. https://doi.org/10.1186/1475-925x-9-39.

[14] K. Feng, G. Liu, Obstructive sleep apnea detection based on unsupervised feature learning and hidden markov model, in: BIBE 2019; The Third International Conference on Biological Information and Biomedical Engineering, Hangzhou, China, 2019, pp. 1–4.

[15] H. Almutairi, G.M. Hassan, A. Datta, Detection of obstructive sleep apnoea by ECG signals using deep learning architectures, in: 2020 28th European Signal Processing Conference (EUSIPCO), IEEE, Amsterdam, Netherlands, 2021: pp. 1382–1386. https://doi.org/10.23919/Eusipco47968.2020.9287360.

[16] A.R. Hassan, Computer-aided obstructive sleep apnea detection using normal inverse Gaussian parameters and adaptive boosting, Biomed. Signal Process. Control 29 (2016), 22–30. https://doi.org/10.1016/j.bspc.2016.05.009.

[17] D. Mukherjee, K. Dhar, F. Schwenker, et al. Ensemble of deep learning models for sleep apnea detection: an experimental study, Sensors 21 (2021), 5425. https://doi.org/10.3390/s21165425.

[18] X. Wang, M. Cheng, Y. Wang, et al. Obstructive sleep apnea detection using ecg-sensor with convolutional neural networks, Multimed. Tools Appl. 79 (2018), 15813–15827. https://doi.org/10.1007/s11042-018-6161-8.

[19] R.V. Sharan, S. Berkovsky, H. Xiong, et al. ECG-derived heart rate variability interpolation and 1-D convolutional neural networks for detecting sleep apnea, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, Montreal, QC, Canada, 2020: pp. 637–640. https://doi.org/10.1109/EMBC44109.2020.9175998.

[20] T. Penzel, G.B. Moody, R.G. Mark, A.L. Goldberger, J.H. Peter, The apnea-ECG database, in: Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163), IEEE, Cambridge, MA, USA, 2000: pp. 255–258. https://doi.org/10.1109/CIC.2000.898505.

[21] S.S. Mostafa, F. Mendonça, A.G. Ravelo-García, et al. A systematic review of detecting sleep apnea using deep learning, Sensors 19 (2019), 4934. https://doi.org/10.3390/s19224934.

[22] L. Breiman, Random forests, Mach. Learn. 45 (2001), 5–32. https://doi.org/10.1023/a:1010933404324.

[23] J.R. Quinlan, Induction of decision trees, Mach. Learn. 1 (1986), 81–106. https://doi.org/10.1007/bf00116251.

[24] S.A. Singh, S. Majumder, A novel approach OSA detection using single-lead ECG scalogram based on deep neural network, J. Mech. Med. Biol. 19 (2019), 1950026. https://doi.org/10.1142/s021951941950026x.

[25] N. Kouiroukidis, G. Evangelidis, The effects of dimensionality curse in high dimensional kNN search, in: 2011 15th Panhellenic Conference on Informatics, IEEE, Kastoria, Greece, 2011: pp. 41–45. https://doi.org/10.1109/PCI.2011.45.