



HAL
open science

ECG-based deep convolutional recurrent network with attention mechanism for sleep apnea detection

Faustine Faccin, El-Hadi Djermoune, Laurent Bougrain, Pauline Guyot

► **To cite this version:**

Faustine Faccin, El-Hadi Djermoune, Laurent Bougrain, Pauline Guyot. ECG-based deep convolutional recurrent network with attention mechanism for sleep apnea detection. *Computing in Cardiology*, Sep 2024, Karlsruhe, Germany. hal-04730104

HAL Id: hal-04730104

<https://hal.science/hal-04730104v1>

Submitted on 10 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

ECG-based Deep Convolutional Recurrent Network with Attention Mechanism for Sleep Apnea Detection

Faustine Faccin^{1,2,3}, El-Hadi Djermoune², Laurent Bougrain³, Pauline Guyot¹

¹ Noviga, Nancy, France

² Université de Lorraine, CNRS, CRAN, Nancy, France

³ Université de Lorraine, CNRS, LORIA, Nancy, France

Abstract

Sleep apnea syndrome (SAS) is a nocturnal respiratory disorder that can be associated with long-term cardiovascular complications. Alternative screening solutions are currently being developed to overcome the limitations of reference in-lab polysomnography. As the respiratory signal can be reconstructed from the electrocardiogram (ECG), the latter is all the more interesting as its recording is easy and non-invasive for the patient. The application of deep learning algorithms using ECGs has proved effective in classifying sleep-related pathological events. In this paper, we propose a novel hybrid architecture to detect apneic episodes using single-lead ECGs. Following a pre-processing step, morphological and temporal components of interest are extracted through convolutional and recurrent blocks, respectively. Additional mechanisms are further integrated to enhance the classification. Models were trained and validated on a dataset derived from STAGES and Apnea-ECG databases. Influence of patient phenotype on classification was estimated by comparing the performance between several groups of patients with different clinical information. Overall, a model we have developed performs competitively with the best current methods by accurately classifying patients to different degrees of severity with average sensitivity, specificity and accuracy of 93.38%, 75.46% and 86.66%, respectively.

1. Introduction

Sleep apnea is a breathing disorder that goes undiagnosed in 80% of patients. Different types of apnea exist: Central Sleep Apnea and Obstructive Sleep Apnea (OSA) [1]. The latter is defined in The International Classification of Sleep Disorders by the occurrence of repetitive episodes of upper airway obstruction during sleep [2]. Such episodes can be characterized by a partial reduction (hypopnea) or a complete cessation (apnea) of airflow for more than 10 seconds [3]. Because sleep apnea events are

usually ended by unconscious micro-arousals, the individual's sleep is disturbed. As described by the American Academy of Sleep Medicine, according to the severity of the pathology, patients rapidly experience daytime symptoms such as excessive sleepiness and poor concentration [4]. If not treated, the patient is also exposed to significant cardiovascular risks on the long-term. Early detection of this pathology is therefore essential.

The gold-standard diagnostic of sleep apnea is polysomnography. The patient is required to spend one or several nights in a medical facility during which multiple physiological signals are recorded, including among others electroencephalogram (EEG), oronasal airflow and electrocardiogram (ECG) [5]. Based on the nocturnal signals, abnormal respiratory events can be easily detected by physicians who can then evaluate the severity of the disease. Severity is commonly quantified by the Apnea-Hypopnea Index (AHI) which computes the number of pathological events (hypopnea or apnea) per hour of sleep. Based on this score, the syndrome can be classified as mild ($5 \leq \text{AHI} < 15$), moderate ($15 \leq \text{AHI} < 30$) or severe ($\text{AHI} \geq 30$).

Unfortunately polysomnography is costly, troublesome and with a long waiting time for the patient. That is why new researches are focusing on new screening solutions, like ECG-based diagnosis. The latter offers a non-invasive diagnosis alternative, which is less restrictive for the patient who can carry out his examination at home. Combination of ECG and deep learning algorithms have already shown promising results [6].

In this article, we propose a hybrid deep architecture, combining convolutional and recurrent layers as well as attention mechanisms in order to extract the morphological and temporal patterns characteristic of sleep apnea from a long-term single-lead ECG recording. Such solution would be a faster and more comfortable mean of screening for the patient, which would also encourage the early diagnosis of this often undiagnosed pathology.

2. Materials and methods

The overall approach of the proposed method is composed of 4 steps: (1) preprocessing of raw ECG signals; (2) ECG-derived signals (EDRs) generation from cleaned signals; (3) morphological and temporal features extraction; (4) classification. Framework described in this article is presented in Figure 1.

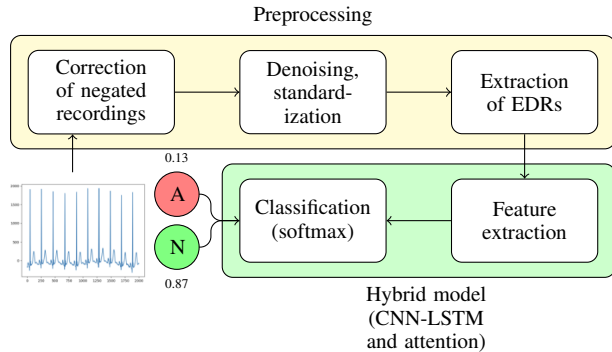


Figure 1: Flowchart of the proposed approach for classifying apneic (A) and normal (N) ECG sequences

2.1. Database description

Data used in this work are ECGs extracted from the Stanford Technology Analytics and Genomics in Sleep (STAGES) study available on the National Sleep Research Resource registry [7] and the Apnea-ECG Database from PhysioNet [8, 9].

STAGES database gathers sleep data of more than 1,500 patients who visited 6 different sleep clinics across the United States and Canada. Sleep was scored according to the clinical protocol of each collection site. Provided data comprise nocturnal sleep recordings, continuous actigraphs, scans, medical record data and questionnaires.

Apnea-ECG database is composed of 70 annotated recordings of patients from two earlier studies. Sleep was firstly scored according to standard criteria, and secondly to indicate whether the ECG segment contains a disordered breathing event or not.

2.2. Data preprocessing and datasets

Inputs to the algorithm were EDR signals from 1-min segments, sampled at 4 Hz, and extracted according to the method described in [10]. Among them were computed R-peaks' amplitudes and intervals.

Before extracting EDRs, ECGs were first preprocessed. Indeed, it is quite common during the placing of electrodes that the position of the latter is reversed, resulting in negative ECG signals. Several negated recordings have been

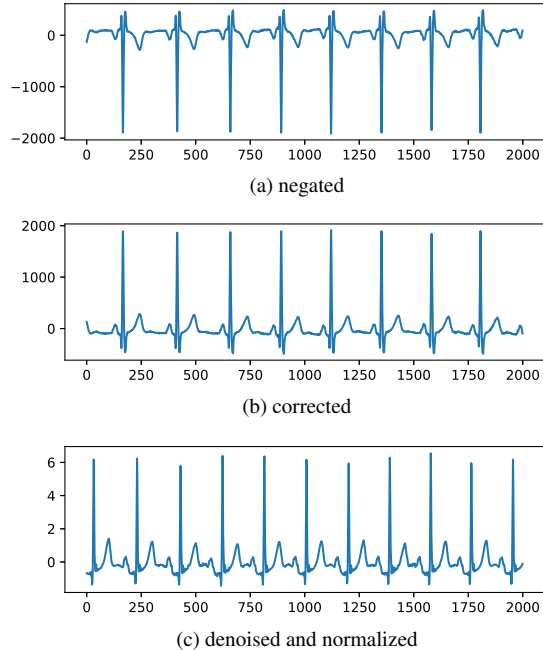


Figure 2: Example of the preprocessing steps applied to one ECG segment

found in STAGES datasets and have therefore been corrected. Then, a notch and moving median filters have been applied to remove the powerline interference and baseline wander respectively. To finish, all data were scaled using z-score normalization. An example of the preprocessing is depicted in Figure 2.

Since the Apnea-ECG is a reference database in the literature, this work proposes to augment the original learning set composed of 35 patients by adding 20 more patients from STAGES. 5-fold cross-validation was performed on this new dataset (26,108 ECG segments in total). The best model was finally tested on Apnea-ECG database original test set comprising 35 additional patients (16,901 samples). Inter-patient paradigm, where each patient only belongs to one dataset and his signals are not shared between datasets is used in this work [11]. Particular attention has been paid to check that various clinical profile based on the patients' age, gender, BMI, and AHI are found in the three datasets.

Annotation files originally include five classes in STAGES: normal, hypopnea, central apnea, obstructive apnea and mixed apnea. In order to match PhysioNet's annotation system, classification has been reduced to a binary problem: thus, all pathological classes have been considered as positive.

3. Model architecture

The proposed architecture is a modified version of Almutairi et al. architecture [12] presented in Mukherjee et al. paper [10], to which were added attention mechanisms more suitable when dealing with long sequences. The model is divided in two parts: feature extractor and classification. First part is composed of 3 convolutional blocks ended by a fully connected layer, followed by a Long-Short Term Memory (LSTM) to which was added an attention layer. Convolutional Neural Network (CNN) are particularly useful to extract spatial patterns while recurrent layers can highlight temporal dependencies between the data. Together, these two form a powerful set of features, emphasized by dot-product attention mechanism which prioritizes and adapts the extraction of relevant patterns in the features. At the end-level of the architecture, features are classified through a dense layer with softmax activation function. The model outputs the probability of the sample to belong to the normal or pathological class.

4. Results

4.1. Evaluation metrics

Evaluation was based on sensitivity (recall), specificity, accuracy, precision and F1-score (harmonic mean of recall and precision). All these measures are based on true, false positive and negative rates:

$$\begin{aligned} \text{Sensitivity} &= \frac{TP}{TP + FN}, \\ \text{Specificity} &= \frac{TN}{TN + FP}, \\ \text{Accuracy} &= \frac{TP + TN}{TP + TN + FN + FP}, \end{aligned}$$

where TP, TN, FP and FN represent the true/false positive and negative rates respectively. The pathological class is the positive one here. While all metrics should be maximized, a special attention was given to the sensitivity as the intended goal was to detect all positive patients and reduce false negatives.

4.2. Overall results

Model performance was evaluated based on signal-wise and subject-wise strategies. The model presented in this paper was able to discriminate positive and negative class samples with a sensitivity of 79.48%, a specificity of 85.30% and an accuracy of 83.09% on the test dataset. Subject-wise evaluation estimates the model's ability to accurately predict the severity of the patient's pathology. The best model, combining a CNN-LSTM and attention, was able to accurately detect each patient syndrome's severity

with average sensitivity of 93.38%, specificity of 75.46% and accuracy of 86.66%. Table 1 gathers metrics' values according to the patient's syndrome severity.

Table 1: Results (in %) for different scores of severity

Severity	Sensitivity	Specificity	F1-score
Mild	100	58.33	90.20
Moderate	94.44	82.35	89.47
Severe	85.71	85.71	82.76

Results reported in [10] for Almutairi *et al.* architecture that inspired the model proposed in this work show, for a 5-fold cross-validation, a sensitivity, specificity and F1-score of 83.7%, 88.1% and 83%. These results must be compared with caution since learning and evaluation strategies, including datasets and results computation, are not exactly the same between the two studies.

4.3. Performance based on patients clinical information

More and more studies are questioning the ability of AHI to fully characterize OSA syndrome. In this sense, many authors have been interested in finding additional parameters from the patient's medical spectrum that could improve the understanding, management and especially the diagnosis of this pathology. A phenotyping approach includes gathering risk factors and clinical, physiopathological, biological and genetic attributes to better define the risks and impacts of such disease [13]. While one of the main risk factor remains the obesity, age and gender were also found to be important attributes [1]. Consequently, the model was tested on different subgroups of patients, depending on the above-mentioned factors in order to identify eventual disparities and biases that will need further improvements in the model's performance.

In view of Table 2, it appears that depending on the patient's weight, the classification results can greatly change as patients suffering from obesity or overweight are significantly better classified than those with a non-pathological weight.

4.4. Discussion

Since apnea events often appear only briefly and very sporadically, only a minority of ECG segments belong to the positive class, making the datasets unbalanced. Mixing databases helped reducing data imbalance particularly important in STAGES while maintaining good results although using different databases can badly impact the performance. Limitations were also faced regarding computing resources. Deep architecture, especially when involv-

Table 2: Mean results (in %) over the different degrees of severity for different clinical attributes

Clinical attribute	Number	Sensitivity	F1-score
Age (18-45 yo)	20	100	84.63
Age (45-60 yo)	12	90.48	88.97
Age (\geq 60 yo)	3	88.89	93.33
Gender (female)	8	100	85.71
Gender (male)	27	92.36	87.90
BMI (normal)	12	66.67	42.86
BMI (overweight)	12	90.48	84.13
BMI (obesity)	11	95.83	97.78

ing recurrent layers, was found to be time-consuming during the training. A trade-off had to be made by training on less data while reducing the number of configuration tests.

5. Conclusion and future work

In this work, the proposed method introduces a deep learning model combining the advantages of both convolutional and recurrent architectures to predict sleep apnea from a long-term Holter ECG. This new model also shows the edges of using attention mechanisms. It has high performance and detects mild and moderate cases particularly well. With further improvements, this study could benefit both doctors and patients by providing a reliable yet simpler and faster alternative to traditional screening examinations of sleep apnea. As only powerline interference and baseline wander were removed from the signals although other types of noises such as myoelectric interferences can also alter the ECGs, some additional preprocessing steps need to be done. In addition, future work will also extend the binary classification to a multi-class problem in order to detect precisely hypopnea, central apnea and obstructive apnea episodes.

Acknowledgments

This research has been conducted using the STAGES - Stanford Technology, Analytics and Genomics in Sleep Resource funded by the Klarman Family Foundation. The investigators of the STAGES study contributed to the design and implementation of the STAGES cohort and/or provided data and/or collected biospecimens, but did not necessarily participate in the analysis or writing of this report. The full list of STAGES investigators can be found at the project website.

The National Sleep Research Resource was supported by the U.S. National Institutes of Health, National Heart Lung and Blood Institute (R24 HL114473, 75N92019R002).

References

- [1] Banno K, Kryger MH. Sleep apnea: clinical investigations in humans. *Sleep Medicine* 2007;8(4):400–426.
- [2] Sateia MJ. International classification of sleep disorders. *Chest* 2014;146(5):1387–1394.
- [3] Redline S, Strohl KP. Recognition and consequences of obstructive sleep apnea hypopnea syndrome. *Clinics in Chest Medicine* 1998;19(1):1–19.
- [4] The Report of an American Academy of Sleep Medicine Task Force. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep* 1999;22:667–689.
- [5] McNicholas WT. Diagnosis of obstructive sleep apnea in adults. *Proceedings of the American Thoracic Society* 2008;5(2):154–160.
- [6] Chang HY, Yeh CY, Lee CT, Lin CC. A sleep apnea detection system based on a one-dimensional deep convolution neural network model using single-lead electrocardiogram. *Sensors* 2020;20(15):4157.
- [7] Zhang GQ, Cui L, Mueller R, Tao S, Kim M, Rueschman M, Mariani S, Mobley D, Redline S. The national sleep research resource: towards a sleep data commons. *Journal of the American Medical Informatics Association* 2018; 25(10):1351–1358.
- [8] Penzel T, Moody GB, Mark RG, Goldberger AL, Peter JH. The apnea-ecg database. In *Computers in Cardiology 2000*. Vol. 27 (Cat. 00CH37163). IEEE, 2000; 255–258.
- [9] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 2000;101(23):e215–e220.
- [10] Mukherjee D, Dhar K, Schwenker F, Sarkar R. Ensemble of deep learning models for sleep apnea detection: an experimental study. *Sensors* 2021;21(16):5425.
- [11] Garcia G, Moreira G, Menotti D, Luz E. Inter-patient ECG heartbeat classification with temporal VCG optimized by PSO. *Scientific Reports* 2017;7(1):10543.
- [12] Almutairi H, Hassan GM, Datta A. Detection of obstructive sleep apnoea by ecg signals using deep learning architectures. In *28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021; 1382–1386.
- [13] Zinchuk AV, Gentry MJ, Concato J, Yaggi HK. Phenotypes in obstructive sleep apnea: a definition, examples and evolution of approaches. *Sleep Medicine Reviews* 2017; 35:113–123.

Address for correspondence:

Faustine Faccin
 NOVIGA
 faustine.faccin@univ-lorraine.fr