Hardware and Software supporting Physiological Measurement (HSPM-2022)

Hardware and Software Supporting Physiological Measurement Workshop

October 27-28, 2022

Konstanz, Germany

Edited by

Ángel Serrano Alarcón and Andrei Boiko

Apnea-Hypopnea Index Using Deep Learning Models with Whole and Window-Based Time Series

Ángel Serrano Alarcón[1] [0000-0003-0841-4382], Natividad Martínez Madrid[1] [0000-0003-1965-9481], Ralf Seepold[2] [0000-0002-5192-2537] and Juan Antonio Ortega[3] [0000-0001-9095-7241]

Abstract. Today many scientific works are using deep learning algorithms and time series, which can detect physiological events of interest. In sleep medicine, this is particularly relevant in detecting sleep apnea, specifically in detecting obstructive sleep apnea events. Deep learning algorithms with different architectures are used to achieve decent results in accuracy, sensitivity, etc. Although there are models that can reliably determine apnea and hypopnea events, another essential aspect to consider is the explainability of these models, i.e., why a model makes a particular decision. Another critical factor is how these deep learning models determine how severe obstructive sleep apnea is in patients based on the apnea-hypopnea index (AHI). Deep learning models trained by two approaches for AHI determination are exposed in this work. Approaches vary depending on the data format the models are fed: full-time series and window-based time series.

Keywords -- Deep Learning, Obstructive Sleep Apnea, OSA, Precision Medicine, AHI

I. Introduction

The classification of physiologically relevant events using artificial intelligence is a flourishing topic in medicine. Concretely, in sleep medicine, there are numerous scientific works in the field of obstructive sleep apnea (OSA) classification [1–4]. Some works use a single time series collected from the patient that represents only one physiological signal to use several physiological signals. [5–7]. Most publications focus on the performance of machine learning models (in terms of the accuracy of the model, sensitivity, etc.). However, not all of them include two crucial aspects when working with biomedical time series of patients suffering from apnea: the explainability of the model and the calculation of the apnea-hypopnea index. Calculating the apnea-hypopnea index is crucial to get to know the severity of a patient suffering from obstructive sleep apnea. Not all developed models take into account the AHI calculation. However, if the models are to be used in a real healthcare environment (in this case, a sleep laboratory), the models should allow the AHI calculation.

In this study, two techniques are explained to calculate the AHI depending on the deep learning model used. The focus is based on the type of input the neural network receives, either window of a certain duration or complete time series. The objective of both approaches is the same, however they present several differences that will be explained throughout this scientific work.

II. Methods

There are a large number of factors to take into account when working with deep learning and time series models. These range from the selection of the model architecture to the dataset used to train the algorithm. Although a priori, it may seem that these factors are not relevant, a correct dataset selection with good quality data and a large number of instances is crucial for success when it comes to deep learning model development tasks.

A. Data set

The dataset is usually determined by the natural setting where the model will be deployed. However, datasets from external data repositories are commonly used. One widely used repository is the Sleep Heart Health Study (SHHS). The data contains time series of patients who may suffer from sleep disorders like sleep apnea. In this case, a time series of seven hours duration can be selected for sleep apnea events classification. The data can be divided into two sets of signals according to the for-mat of the dataset that fed the models during the training:

- Window-based time series classification
- Whole-time series classification

After several tests, the 60-second windows seem to work best for classifying sleep apnea events and especially for calculating the AHI.

B. Model architecture

After a thorough literature analysis, convolutional neural networks are the best per-formers when classifying time series, in this case, OSA events. Several factors should be considered:

[1] Reutlingen University, Alteburgstr. 150, 72762 Reutlingen, Germany
{angel.serrano_alarcon,natividad.martinez}@reutlingen-university.de
[2] HTWG Konstanz, Alfred-Wachtel-Str. 8, 78462 Konstanz, Germany
ralf.seepold@htwg-konstanz.de
[3] University of Seville, San Fernando 4, 41004 Sevilla, Spain
jortega@us.es

the neural network complexity, the computational cost in training and inference, and the explainability and performance of the models. Based on the two approaches followed in this study, we found the following architectures:

- Window-based time series classification: Fully-Convolutional Neural Net-works (FCNNs)
- Whole time series classification: convolutional neural network for segmentation

C. Explainability of the model

The explainability of the model is crucial when it comes to developing algorithms that will be used in an actual medical environment. The class activation map (CAM) or gradient-weighted class activation mapping (Grad-CAM) technique are some of the methods used for explainability [8,9]. These techniques are easier to implement with window-based time series classification, as it is easier to visualize in 60 seconds than with the full-time series. The explainability of the complete time series classification needs to be further evaluated in order to be able to provide the clinician with all the information about a given decision quickly. However, this is not the scope of this study.

D. Apnea-hypopnea index (AHI)

The widely known formula is used to calculate the AHI:

$$AHI = \frac{Number\ of\ apnea\ and\ hypopnea\ events}{Total\ sleep\ time\ (hours)} \quad (1)$$

Two methods are used to obtain the number of apnea and hypopnea events depending on whether windows or full-time series are used. These methods are described in section 3

III. Results

This work does not pretend to show how to develop a deep learning model to work with time series and to show architectures that yield good results. Once the whole model training process, evaluation of the neural network performance, and model explainability are finished, it is time to calculate the AHI. To get the most accurate AHI possible, the deep learning model must classify as many apnea events correctly as possible. This paper proposes a method to calculate the AHI using the two most typical approaches when working with time series in machine learning, the window-based time series classification and the whole-time series classification.

A. Window-based tune series classification

When working with time series divided into windows of a specific duration, it is al-ways necessary to consider whether overlapping is applied. In this case, where the aim is to classify apnea and hypopnea events in addition to knowing the number per hour, the absence of overlapping is crucial. This way, counting how many events have occurred in the whole sleep time is possible. In this case, selecting 60-second windows helps to include complete apnea events longer than 10 seconds. Information can be lost when dividing the time series into windows. Still, this 60 seconds allows us to classify the patient's sleep minute by minute, and considering that we have time series of several hours, the AHI can be very close to the real one. Therefore, for this approach, the number of events (60-second windows) classified as apnea would be counted and divided by the total sleep time. The fact of not including overlapping allows not counting the same event several times.

B. Whole time series classification

With the segmentation technique, it is possible to identify apnea events with second-by-second accuracy. Regardless of the apnea or hypopnea event duration, such events will be classified without selecting time windows. This fact avoids the loss of information by dividing the time series into windows of a particular period in order to be classified by the deep learning algorithm. However, when it comes to calculating the AHI, it is more complex than windowing. To do so, it will be necessary to identify all the parts of the signal that have been classified as an apnea event (in seconds), along with obtaining the sets where there are more than 10 seconds of an event (if an apnea event is classified as '1' then it would be necessary to look for more than ten consecutive values of '1s'). Once all the sets have been obtained, regardless of their duration, such sets can be counted and divided by the total sleep time. This method will have higher accuracy than the windowing approach. However, it also requires higher accuracy of the neural network so that all the second-by-second events are classified correctly as we can see in Figure 1.

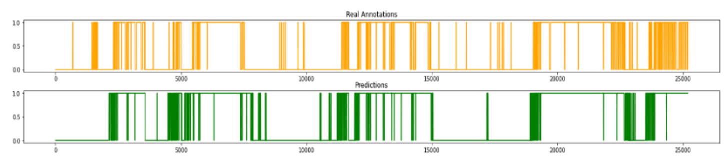

Fig. 1. Real annotations performed by the clinicians and predictions generated by the deep learning model.

IV. Conclusion and Future Work

Although different alternatives exist to work with time series of patients suffering from obstructive sleep apnea and to train deep learning models to classify sleep apnea events, the AHI calculation is limited. In particular, following the two approaches explained in previous sections: window-based and whole time series classification. The techniques described in this paper allow a simple and concise calculation of AHI and help doctors and clinicians to understand the results derived from deep learning models. The aim is to make the models as understandable and accurate as possible so they can be easily implemented in real healthcare settings.

V. Acknowledgment

This research was partially funded by (a) Carl Zeiss Foundation and the MORPHEUS-Project “Non-invasive system for measuring parameters relevant to sleep quality” (project number: P2019-03-003); and (b) by the German Academic Exchange Service – DAAD within the project "Portable system for detecting and alerting driver fatigue" supported (project number 57562477).

VI. References

1. Piorecky, M., Bartoň, M., Koudelka, V., Buskova, J., Koprivova, J., Brunovsky, M., Piorecka, V.: Apnea Detection in Polysomnographic Recordings Using Machine Learning Techniques. Diagnostics (Basel, Switzerland). 11, (2021). https://doi.org/10.3390/DIAGNOSTICS11122302
2. Ramachandran, A., Karuppiah, A.: A Survey on Recent Advances in Machine Learning Based Sleep Apnea Detection Systems. Healthc. 2021, Vol. 9, Page 914. 9, 914 (2021). https://doi.org/10.3390/HEALTHCARE9070914
3. JeyaJothi, E.S., Anitha, J., Rani, S., Tiwari, B.: A Comprehensive Review: Computational Models for Obstructive Sleep Apnea Detection in Biomedical Applications. Biomed Res. Int. 2022, 1–21 (2022). https://doi.org/10.1155/2022/7242667
4. Drzazga, J., Cyganek, B.: An lstm network for apnea and hypopnea episodes detection in respiratory signals. Sensors. 21, (2021). https://doi.org/10.3390/s21175858
5. Dey, D., Chaudhuri, S., Munshi, S.: Obstructive sleep apnoea detection using convolutional neural network based deep learning framework. Biomed. Eng. Lett. 8, 95–100 (2018). https://doi.org/10.1007/s13534-017-0055-y
6. Haidar, R., McCloskey, S., Koprinska, I., Jeffries, B.: Convolutional Neural Networks on Multiple Respiratory Channels to Detect Hypopnea and Obstructive Apnea Events. Proc. Int. Jt. Conf. Neural Networks. 2018-July, (2018). https://doi.org/10.1109/IJCNN.2018.8489248
7. Pathinarupothi, R.K., Dhara Prathap, J., Rangan, E.S., Gopalakrishnan, A.E., Vinaykumar, R., Soman, K.P.: Single Sensor Techniques for Sleep Apnea Diagnosis Using Deep Learning. Proc. - 2017 IEEE Int. Conf. Healthc. Informatics, ICHI 2017. 524–529 (2017). https://doi.org/10.1109/ICHI.2017.37
8. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Int. J. Comput. Vis. 128, 336–359 (2020). https://doi.org/10.1007/s11263-019-01228-7
9. Vijayarangan, S., Murugesan, B., Vignesh, R., Preejith, S.P., Joseph, J., Sivaprakasam, M.: Interpreting Deep Neural Networks for Single-Lead ECG Arrhythmia Classification. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS. 2020-July, 300–303 (2020). https://doi.org/10.48550/arxiv.2004.05399