



Extended Abstract

# A Convolutional Network for the Classification of Sleep Stages †

Isaac Fernández-Varela , Elena Hernández-Pereira and Vicente Moret-Bonillo

CITIC, Universidade da Coruña, 15071 A Coruña, Spain; elena@udc.es (E.H.-P.); vicente.moret@udc.es (V.M.-B.)

\* Correspondence: isaac.fvarela@udc.es

† Presented at the XoveTIC Congress, A Coruña, Spain, 27–28 September 2018.

Published: 14 September 2018



**Abstract:** The classification of sleep stages is a crucial task in the context of sleep medicine. It involves the analysis of multiple signals thus being tedious and complex. Even for a trained physician scoring a whole night sleep study can take several hours. Most of the automatic methods trying to solve this problem use human engineered features biased for a specific dataset. In this work we use deep learning to avoid human bias. We propose an ensemble of 5 convolutional networks achieving a kappa index of 0.83 when classifying 500 sleep studies.

**Keywords:** sleep staging; convolutional neural network; classification

## 1. Introduction

Sleep disorders are a common problem: insomnia has a prevalence of 20% and daytime sleepiness between 12% and 15% [1,2]. Sleep disorders can be diagnosed analysing a set of bio-signals recorded during the sleep period, a technique called polysomnography. This analysis is expensive, uncomfortable for the patient and difficult to interpret. Thus, it is usually presented as an hypnogram, a graph showing the evolution of the sleep stages.

The gold standard for the hypnogram construction is the American Academy of Sleep Medicine (AASM) guide, which includes how to identify sleep stages and associated events such as arousals, movements and cardiac and respiratory events. This guide identifies 5 sleep stages: Awake (W), Rapid Eye Movements (REM), and 3 Non REM known as N1, N2, and N3. A well built hypnogram allows a quickly and accurate diagnosis. Yet, the agreement between two experts trying to build the same hypnogram is lower than 90% (with a kappa index between 0.48 and 0.89 [3]), with even less agreement for specific stages such as N1.

The aforementioned reasons motivated several works that automate the sleep stages classification. Traditionally, these works were based on feature extraction and later classification [4–8], solutions commonly biased towards the available dataset. To solve the bias problem we propose the use of Deep Learning, an option already explored by some authors [9–13].

Particularly, we use a convolutional network that learns the relevant features for the classification by itself. Following the AASM guides we use multiple channels, namely two electroencephalogram (EEG), one electromyogram (EMG) and both electrooculogram (EOG). Furthermore, our signals are filtered to reduce noise and remove artefacts induced by the electrocardiogram (ECG).

## 2. Materials

Our experiments were carried out using real polisomnographs (PSG) from the Sleep Heart Health Study (SHHS) [14]. These PSG were scored by several experts following the AASM rules [15] and include 2 EEG, both EOG, EMG and ECG.

From the database we randomly selected 3 datasets for training, testing and validation containing 400, 100, and 500 registers respectively, or 288,000, 119,121, and 606,981 samples. Most of the samples belong to class N2 (36%) or W (38%), and the less represented one is N1 (3%). Imbalanced classes is a typical problem in sleep medicine.

### 3. Method

We use a convolutional network that is fed with 5 filtered signals simultaneously: two EEG derivations, both EOG and one EMG. The filtering pipeline includes a Notch filter in 60 Hz for all the signals, and a high pass in 15 Hz for the EMG signal. We also remove ECG artefacts using an adaptive filter [16].

Following clinical procedure, the network input are 30 s windows (usually called epochs) with the signals re-sampled (if needed) to 125 Hz, resulting in a sample dimension of  $3750 \times 5$ . Each signal is normalised to mean 0 and deviation 1, using as reference the training dataset.

Figure 1 represents the proposed convolutional network. The convolutional block presented in the figure is a set of four layers including: 1D convolution, batch normalization, ReLu activation and average pool. This block is repeated  $n$  times. All the 1D convolutions have the same kernel size but layer  $i$  has twice the filters of layer  $i - 1$ .

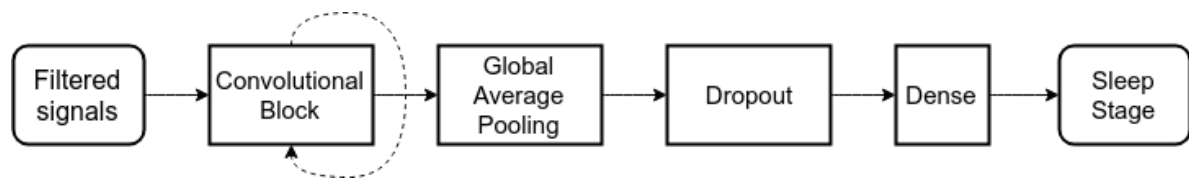


Figure 1. Proposed Convolutional Network.

The network was trained using Adam optimiser with 64 samples per batch and early stopping with a patience of 10, monitoring the loss made in the validation dataset.

To select the hyper-parameters:  $n$ , the number of filters for the first layer, kernel size, and learning rate; we used a Tree-structured Parzen Estimator (TPE), which is sequential model-based optimisation (SMBO) approach. We trained 50 models using the TPE and selected the 5 best to build an ensemble.

### 4. Results

The ensemble built with the 5 best models was used to carry out experiments with the test dataset, obtaining the performance measures and confusion matrix shown in Figure 2. The class with the best classification is W and then N2, N3, REM show similar values regarding the F1 score, although there are significant differences in the sensitivity. As expected, N1 is the class with the worst classification, with values lower than 0.4. Apart from the problems classifying class N1, most of the errors happen between classes N2 and N3.

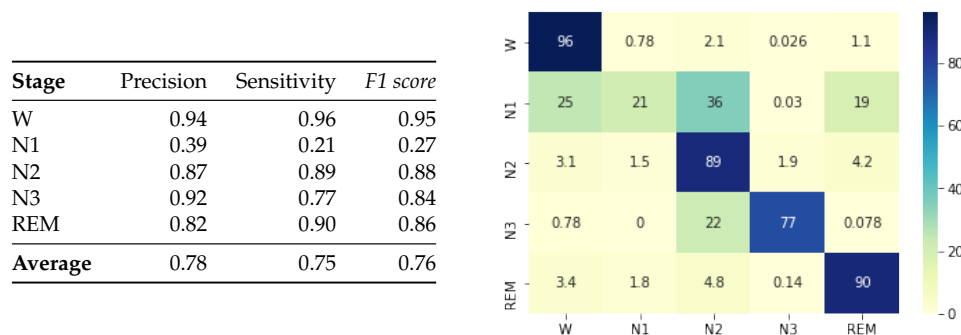


Figure 2. Confusion matrix for the test set classification using the 5 model ensemble.

## 5. Discussion and Conclusions

In this work we propose an ensemble of convolutional networks to classify sleep stages. The main reason is to avoid introducing human bias in our solution with a method that learns the relevant features by itself.

To configure the network hyper-parameters we used a tree-structured parzen estimator, evaluating 50 different models and selecting the best 5 to build an ensemble. This ensemble achieves an average precision, sensitivity, and F1 score of 0.78, 0.75 and 0.76 with a kappa index of 0.83. Yet it shows difficulties to classify class N1 and a bias towards class N2.

Given the lack of standards or benchmarks related to this problem, it is difficult to compare our solutions against previous works. Some references are shown in Table 1. Our values are competitive, achieving the highest kappa index and the best classification for class W.

**Table 1.** Performance achieved in previous works.

Work	Dataset	Kappa	F1 Score				
			W	N1	N2	N3	REM
Biswal et al. [12]	Massachusetts General Hospital, 1000 recordings	0.77	0.81	<b>0.70</b>	0.77	0.83	<b>0.92</b>
Långkvist et al. [9]	St Vincent’s University Hospital, 25 recordings	0.63	0.73	0.44	0.65	<b>0.86</b>	0.80
Sors et al. [13]	SHHS, 1730 recordings	0.81	0.91	0.43	0.88	0.85	0.85
Supratak et al. [11]	MASS dataset, 62 recordings	0.80	0.87	0.60	<b>0.90</b>	0.82	0.89
Supratak et al. [11]	SleepEDF, 20 recordings	0.76	0.85	0.47	0.86	0.85	0.82
Tsinalis et al. [10]	SleepEDF, 39 recordings	0.71	0.72	0.47	0.85	0.84	0.81
Tsinalis et al. [17]	SleepEDF, 39 recordings	0.66	0.67	0.44	0.81	0.85	0.76
This work	SHHS, 500 recordings	<b>0.83</b>	<b>0.95</b>	0.27	0.88	0.84	0.86

Results are promising and the approach should be easily extended to other PSG sources. Moreover, if we could have PSG acquired in different conditions during the training phase, regularisation should improve.

As future work, we need to understand how and why the model makes the decisions instead of treating it as a black box.

**Funding:** This research was partially financed by the Xunta de Galicia [ED431G/01] and the European Union through the ERDF fund.

**Acknowledgments:** We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

- Ohayon, M.M.; Sagales, T. Prevalence of insomnia and sleep characteristics in the general population of Spain. *Sleep Med.* **2010**, *11*, 1010–1018, doi:10.1016/j.sleep.2010.02.018.
- Marin, J.; Gascon, J.M.; Carrizo, S.; Gispert, J. Prevalence of sleep apnoea syndrome in the Spanish adult population. *Int. J. Epidemiol.* **1997**, *26*, 381–386, doi:10.1093/ije/26.2.381.
- Stepnowsky, C.; Levendowski, D.; Popovic, D.; Ayappa, I.; Rapoport, D.M. Scoring accuracy of automated sleep staging from a bipolar electroocular recording compared to manual scoring by multiple raters. *Sleep Med.* **2013**, *14*, 1199–1207, doi:10.1016/j.sleep.2013.04.022.
- Liang, J.; Lu, R.; Zhang, C.; Wang, F. Predicting Seizures from Electroencephalography Recordings: A Knowledge Transfer Strategy. In Proceedings of the 2016 IEEE International Conference on Healthcare Informatics (ICHI 2016), Chicago, IL, USA, 4–7 October 2016; pp. 184–191, doi:10.1109/ICHI.2016.27.
- Hassan, A.R.; Bhuiyan, M.I.H. A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. *J. Neurosci. Meth.* **2016**, *271*, 107–118, doi:10.1016/J.JNEUMETH.2016.07.012.

6. Sharma, R.; Pachori, R.B.; Upadhyay, A. Automatic sleep stages classification based on iterative filtering of electroencephalogram signals. *Neural Comput. Appl.* **2017**, *28*, 2959–2978, doi:10.1007/s00521-017-2919-6.
7. Lajnef, T.; Chaibi, S.; Ruby, P.; Aguera, P.E.; Eichenlaub, J.B.; Samet, M.; Kachouri, A.; Jerbi, K. Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines. *J. Neurosci. Meth.* **2015**, *250*, 94–105, doi:10.1016/J.JNEUMETH.2015.01.022.
8. Huang, C.S.; Lin, C.L.; Ko, L.W.; Liu, S.Y.; Su, T.P.; Lin, C.T. Knowledge-based identification of sleep stages based on two forehead electroencephalogram channels. *Front. Neurosci.* **2014**, *8*, 263, doi:10.3389/fnins.2014.00263.
9. Långkvist, M.; Karlsson, L.; Loutfi, A. Sleep Stage Classification Using Unsupervised Feature Learning. *Adv. Artif. Neural Syst.* **2012**, *2012*, 1–9, doi:10.1155/2012/107046.
10. Tsinalis, O.; Matthews, P.M.; Guo, Y.; Zafeiriou, S. Automatic Sleep Stage Scoring with Single-Channel EEG Using Convolutional Neural Networks. *arXiv* **2016**, arXiv:1610.01683.
11. Supratak, A.; Dong, H.; Wu, C.; Guo, Y. DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 1998–2008, doi:10.1109/TNSRE.2017.2721116.
12. Biswal, S.; Kulas, J.; Sun, H.; Goparaju, B.; Westover, M.B.; Bianchi, M.T.; Sun, J. SLEEPNET: Automated Sleep Staging System via Deep Learning. *arXiv* **2017**, arXiv:1707.08262.
13. Sors, A.; Bonnet, S.; Mirek, S.; Vercueil, L.; Payen, J.F. A convolutional neural network for sleep stage scoring from raw single-channel EEG. *Biomed. Signal Process. Control* **2018**, *42*, 107–114, doi:10.1016/j.bspc.2017.12.001.
14. Quan, S.F.; Howard, B.V.; Iber, C.; Kiley, J.P.; Nieto, F.J.; O'connor, G.T.; Rapoport, D.M.; Redline, S.; Robbins, J.; Samet, J.M.; et al. The Sleep Heart Health Study: Design, Rationale, and Methods. *Sleep* **1997**, *20*, 1077–1085, doi:10.1093/sleep/20.12.1077.
15. Bonnet, M.H.; Carley, D.; Carskadon, M. EEG arousals: Scoring rules and examples: a preliminary report from the Sleep Disorders Atlas Task Force of the American Sleep Disorders Association. *Sleep* **1992**, *15*, 173–184.
16. Fernández-Varela, I.; Alvarez-Estevéz, D.; Hernández-Pereira, E.; Moret-Bonillo, V. A simple and robust method for the automatic scoring of EEG arousals in polysomnographic recordings. *Comput. Biol. Med.* **2017**, *87*, 77–86, doi:10.1016/J.COMPBIOMED.2017.05.011.
17. Tsinalis, O.; Matthews, P.M.; Guo, Y. Automatic Sleep Stage Scoring Using Time-Frequency Analysis and Stacked Sparse Autoencoders. *Ann. Biomed. Eng.* **2016**, *44*, 1587–1597, doi:10.1007/s10439-015-1444-y.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).