Blinking patterns (EOG) vs. brain waves (EEG) in automated sleep-stage scoring

Bruna C. S. Karez[1]

[1] *Promotion 140, ESPCI Paris, PSL Research University, Paris, France. Electronic address: bruna.correa-santiago-karez@espci.fr*

Submitted June 10, 2024

1. INTRODUCTION

Sleep is a core function of human life. Sleep disorders have been associated with several health problems, such as depression, difficulty in concentration and memory consolidation, as well as with neurodegenerative diseases such as Parkinsons, or Huntington's disease [1–7]. The most prevailing sleep disorder is insomnia [8], but other sleep disorders include hypersomnias (narcolepsy, idiopathic hypersomnia, Kleine-Lewin syndrome) and parasomnias (sleepwalking/sleep terrors, REM sleep behaviour disorder) [9, 10]. Some of these pathologies are linked to changes in the sleep architecture, for instance in the continuity and efficiency of sleep, as well as the alteration of the duration and overall characteristics of the different sleep stages [11].

Therefore, classifying sleep stages can be a crucial step for accurately diagnosing and treating sleep disorders. This process typically involves trained sleep scorers manually evaluating polysomnography (PSG) data recorded during sleep. This data is most commonly composed of an electroencephalogram (EEG), which is the recording of brain electric signals, an electro-oculogram (EOG), to track eye movement, an electromyogram (EMG), to for muscle activity monitoring, and an electrocardiogram (ECG), to check the heartbeat, among other physiological signals [12]. Analyzing and labeling such signals can be laborious, time-intensive, and susceptible to human bias. Thus, leveraging machine learning techniques for sleep stage scoring can be a powerful way of obtaining faster, easier, and more robust diagnosis.

Sleep is mainly categorized into two stages, namely rapid eye movement (REM) sleep and non-rapid eye movement (NREM) sleep. NREM sleep stage accounts for 75–80 % and REM sleep stage usually lasts for around 20–25 % of total sleep duration [8]. NREM stage is further categorized into four sleep stages namely stage-1, 2, 3 and 4. Thus, a total of five sleep stages are known (1, 2, 3, 4 and REM) [13].

Here, we used a labelled dataset from the Sleep-EDF Database Expanded, which contains extensive polysomnography (PSG) recordings including electroencephalography (EEG), electromyography (EMG), and electrooculography (EOG) [14]. All EEG and EOG had the sampling rate of 100 Hz. This database comprises 197 full-night PSG sleep recordings, each manually scored by trained technicians for wake (W), non-rapid eye movement N1, N2, N3 stages, and rapid eye movement (REM) sleep stages, following the Rechtschaffen and Kales manual [13]. The scoring is based on Fpz-Cz/Pz-Oz EEGs, as recommended by previous studies. These recordings are sourced from the Sleep Cassette Study, consisting of 153 files collected between 1987 and 1991. The study focuses on investigating the effects of age on sleep among healthy Caucasians aged 25-101, without any sleep-related problems. The data are available in Physionet [15].

Our study aimed to forecast the hypnogram (representation of the sleep stages) by harnessing either EEG or EOG data. To achieve this objective, we designed a neural network architecture to classify sleep stages from EEG or EOG data. The model leverages a combination of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to capture both spatial and temporal features inherent in the electrophysiology signals.

2. METHODS

Data processing

The Sleep Cassette data from the Sleep-EDF Database was downsampled to 20kHz and divided into 30-second segments, in a script adapted from [16]. This segment length was determined because certified technicians assign sleep stages to each 30-second segment during visual inspection on a standardized screen. The annotations corresponded to the sleep stages: 0 (Wake), 1 (N1), 2 (N2), 3 (N3), 4 (REM), and 5 (Unknown), with only annotated segments retained, thus excluding any unknown signals. The recorded physiological signals included EEG Fpz-Cz, EEG Pz-Oz, horizontal EOG, oronasal respiration, submental EMG, and rectal temperature. Here, we focus on the EOG and EEG Fpz-Cz and Pz-Oz channels, comparing their performance to determine if a single channel could accurately predict sleep stages. The data was organized into tensors containing the signals (Fig. 1) and their corresponding labels (Fig. 2) and then split: 90% was used for training (175,996 signals) and 10% for validation (19,485 signals).

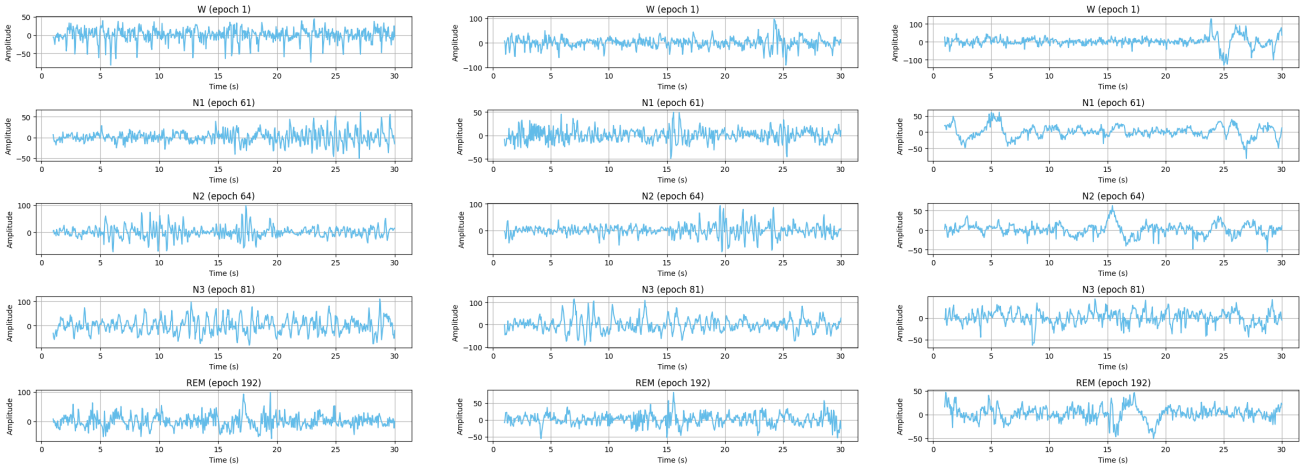
Model components

Input

The input to the model is a single-channel EEG or EOG channel, extracted from the four-channel PSG data, pre-processed and split into 30-s sequences.

Convolutional Layers

Our neural network's first part consists of representation learning, and comprises two convolutional layers. This step is to extract time-invariant features from the raw single-channel 30-second EEG or EOG epochs. A first convolutional layer uses 16 filters with a kernel size of 5 and stride 1, padded to maintain the input dimensions. This layer captures basic spatial features from the input signal. Following the first, a second convolutional layer uses 32 filters with the same kernel size and stride, further refining the spatial features extracted from the EEG data. After each convolutional layer, a 1D-max pooling layer with a kernel size of 8 and stride 8 is applied to reduce the dimensionality and retain the most prominent features. Between the CNN and the Long-Short Term Memory (LSTM) layer, we use fully connected layers to reduce dimensionality. A first layer



(a) EEG Fpz-Cz (b) EEG Pz-Oz (c) EOG

Fig. 1. Epochs corresponding to the five sleep stages. The three different signals (EEG Fpz-Cz, EEG Pz-Oz and EOG) correspond to the three PSG channels used for this study. Each plot corresponds to a 30-s epoch labeled as the sleep stage indicated above (in order, W, N1, N2, N3 and REM). **a.** Electroencephalogram (EEG) in the Fpz-Cz electrode configuration. **b.** Electroencephalogram (EEG) in the Pz-Oz electrode configuration. **c.** Electro-oculogram (EOG).

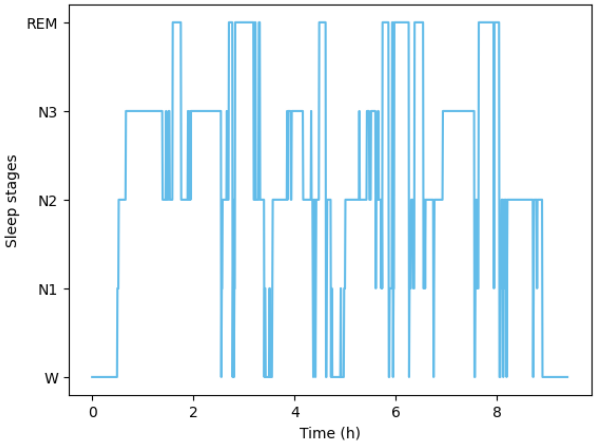


Fig. 2. Hypnogram representing the sleep stages progression during the night. This corresponds to all of the labels attributed to the 30-s segments of PSG data.

reduces the dimensionality of the output from the convolutional layers to 1024 neurons, providing a dense representation of the spatial features. For regularization purposes, we add a dropout layer with a dropout rate of 0.5, which prevents overfitting by randomly setting half of the activations to zero during training. A second fully connected layer further reduces the dimensionality to 32 neurons, preparing the data for the LSTM layer.

Recurrent Layer

The core of the temporal processing, this layer consists of an Long Short-Term Memory (LSTM) network with 512 hidden units, operating bidirectionally (bi-directional LSTM). It captures the sequential dependencies in the electrophysiology data, crucial for understanding the temporal patterns of sleep stages. The output from the LSTM layer is fed into a fully connected layer with neurons equal to the number of sleep stage classes (in this case, 5). This layer produces the final class probabilities for each sleep stage.

Activation

The Rectified Linear Unit (ReLU) activation function is used after each convolutional and fully connected layer to introduce non-linearity and enable the model to learn complex patterns.

Training and Evaluation

The model is trained using the Adam optimizer. Several learning rates and batch sizes were tested (see **Results**). We use the cross-entropy loss to measure the agreement between the predicted class probabilities and the target sleep stages in both of these training steps.

Performance Metrics

The model's performance is evaluated using accuracy, and a confusion matrix is plotted to visualize the classification results across different sleep stages. We also computed the accuracy score with the sklearn library, which corresponds to the number of correct predictions divided by the number of total predictions, and the f_1 score, which is the harmonic mean of the precision and the recall. By definition, precision is the ratio of correctly predicted positive observations to the total predicted positives, or the answer to the question, "Of all the instances that were predicted as positive, how many were truly positive?". On the other hand, recall, or sensitivity, is the ratio of correctly predicted positive observations to the all observations in the actual class. It answers the question, "Of all the instances that were truly positive, how many were predicted correctly?"

Summary

SleepStageNet effectively combines the strengths of CNNs for spatial feature extraction and LSTMs for temporal sequence modeling, making it well-suited for the task of sleep stage classification from EEG data. The architecture is designed to capture the complex patterns in EEG signals, providing accurate and reliable sleep stage predictions.

Implementation

The scripts for this study were written in PyTorch and executed in Spyder (Python 3.11). The computations were performed in an

NVIDIA GeForce RTX 3060 GPU with a driver version of 545.84 and a CUDA version of 12.3.

3. RESULTS AND DISCUSSION

We tested a wide range of parameters (Table 1) on the EOG data, and we later assessed their performance (Fig. 3), to retain only one set of parameters for the comparison between EOG and EEG datasets. We executed the model best-performing model (epochs = 150, $l_r = 10^{-6}$, batch size = 10) in both EOG and two EEG datasets (Fpz-Cz and Pz-Oz), and compared the results. In addition, we reproduced the double EEG-EOG comparison on another set of parameters (epochs = 150, $l_r = 10^{-6}$, batch size = 100) for control (not shown here, since the results were essentially the same).

With both EEG and EOG datasets, we have the same order of most accurately predicted labels, which is (in decreasing accuracy order): W, N2, N3, REM and N1 (Fig. 4). Interestingly, the N1 seems to be also the most difficult to predict by humans, showing the highest interrater variability in different analysis [17, 18]. The N1 stage is an unstable sleep stage, heavily affected by the transitions between wakefulness and sleep onset, as well as by changes between sleep stages during arousals [19]. Studies indicate that scorers frequently struggled to stage epochs during the transitions from wakefulness to N1 and from N1 to N2. Additionally, it was noted that scorers had difficulty distinguishing between N2 and N3.

Consistent with existing literature, our results show that training with EEG data achieves superior performance compared to training with EOG data (Fig. 5). The accuracy achieved is of 0.745 for the model trained using EEG Fpz-Cz signals, and of 0.699 for its EOG counterpart. The reasoning behind this is that EEG patterns offer the primary information for interpreting EEG recordings, relevant for both human sleep technicians and intelligence algorithms. This could also mean that the primary data in EOG useful to extract sleeping stages would typically arise from contamination by EEG signals [20]. Nonetheless, it is important to note that EEG signal acquisition is relatively complex and can even disturb natural sleep patterns. In a clinical application, it is thus crucial to consider if the gain in performance is significant enough to counter-balance the complications. Researchers have created EOG-based sleep scoring models, using inspiration from EEG models and adapting them to the EOG signal characteristics.

Another result that comes out as striking, is the underperformance of the EEG Pz-Oz channel using two different parameter sets. The accuracy of the model when using this channel is of 0.677. However, this finding is coherent with studies showing that outcomes appear to be affected by the selection of the EEG channel, with frontal leads demonstrating improved performance [19].

4. CONCLUSION

Several aspects of the model that could be enhanced, from the data pre-processing to the model architecture itself. Some studies suggest that combining EEG and EOG data can be used for obtaining more accurate predictions [21], others exploit both temporal and frequential components of the signals [22] to enhance feature-extraction. Other architectures, such as GRUs or transformers [21, 23], have also shown interesting results in sleep scoring tasks. Nonetheless, with a relatively simple architecture consisting of a CNN layer and a bi-directional LSTM layer, we were able to predict sleep stages with an accuracy of 74,5%. Finally, even though some stages were less accurately predicted than others, this phenomenon that can also occur with human-scored data, and enriching the model can help to homogeneize the performance among classes.

Table 1. Different training parameters tested in the model. # corresponds to the number identifying the set of parameters, epochs corresponds to the number of epochs used to train and evaluate the model, time corresponds to the total time to train and evaluate the entirety of the epochs, and f_1 corresponds to the f_1 score calculated with the sklearn library, by computing the harmonic mean of the precision and recall scores. "NA" (Non Available) is indicated when a parameter was not computed for a given set of parameters.

#	epochs	l_r	batch size	time	f_1
0	150	10^{-4}	10	NA	NA
1	150	10^{-5}	10	NA	NA
2	100	10^{-6}	10	1h 58 min	.69
3	150	10^{-6}	10	2 h 43 min	.7
4	100	10^{-6}	100	12 min 31 s	.68
5	150	10^{-6}	100	18 min 39 s	.68
6	250	10^{-6}	100	31 min 32 s	.69
7	100	10^{-6}	30	42 min 39 s	.69

ABBREVIATIONS

CNN = Convolutional Neural Network
EEG = Electroencephalogram
EOG = Electro-oculogram
LSTM = Long Short-Term Memory
NREM = Non Rapid Eye-Movement
PSG = Polysomnography
REM = Rapid Eye-Movement
RNN = Recurrent Neural Network

REFERENCES

1. D. Nutt, S. Wilson, and L. Paterson, "Sleep disorders as core symptoms of depression," Dialogues Clin. Neurosci. **10**, 329–336 (2008).
2. M. T. Acosta, "Sleep, memory and learning," Med. (B Aires.) **79 Suppl 3**, 29–32 (2019).
3. I. Arnulf, J. Nielsen, E. Lohmann, J. Schiefer, E. Wild, P. Jennum, E. Konofal, M. Walker, D. Oudiette, S. Tabrizi, and A. Durr, "Rapid eye movement sleep disturbances in huntington disease," Arch. Neurol. **65**, 482–488 (2008).
4. V. C. De Cock, M. Vidailhet, and I. Arnulf, "Sleep disturbances in patients with parkinsonism," Nat. Clin. Pract. Neurol. **4**, 254–266 (2008).
5. J. Massicotte-Marquez, A. Décary, J.-F. Gagnon, M. Vendette, A. Mathieu, R. B. Postuma, J. Carrier, and J. Montplaisir, "Executive dysfunction and memory impairment in idiopathic REM sleep behavior disorder," Neurology **70**, 1250–1257 (2008).
6. R. Silvestri, A. Gagliano, I. Aricò, T. Calarese, C. Cedro, O. Bruni, R. Condurso, E. Germanò, G. Gervasi, R. Siracusano, G. Vita, and P. Bramanti, "Sleep disorders in children with Attention-Deficit/Hyperactivity disorder (ADHD) recorded overnight by video-polysomnography," Sleep Med. **10**, 1132–1138 (2009).
7. M. T. Acosta, "[sleep, memory and learning]," Medicina **79 Suppl 3**, 29—32 (2019).
8. M. Sharma, H. S. Dhiman, and U. R. Acharya, "Automatic identification

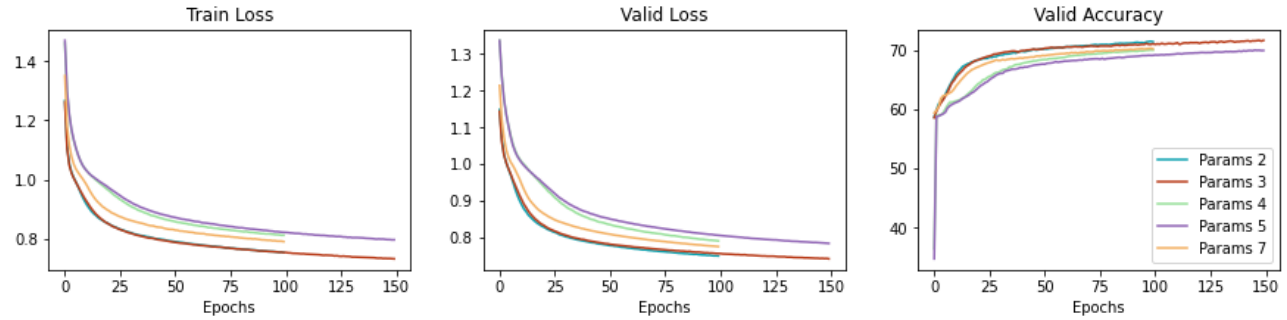
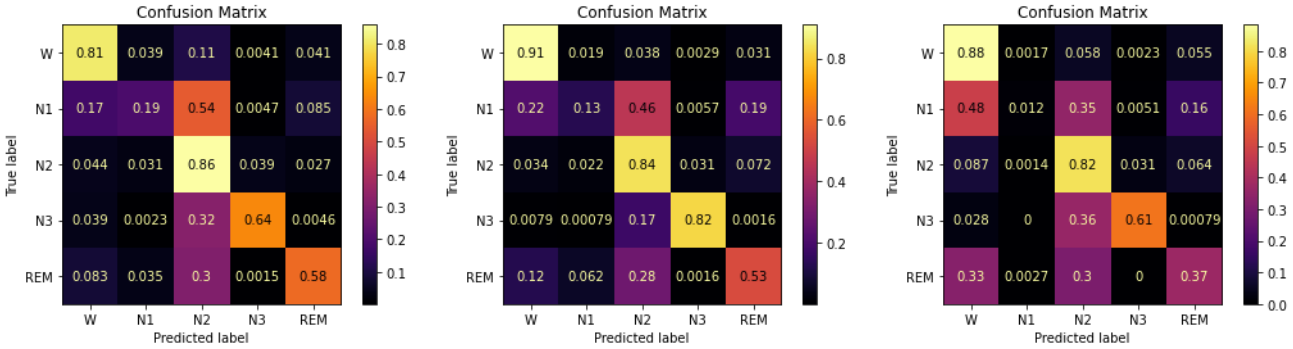


Fig. 3. Comparison of train loss, valid loss and valid accuracy for some of the sets of parameters tested. The sets of parameters correspond to the ones detailed in Table 1.



(a) EOG (b) EEG (Fpz-Oz) (c) EEG (Pz-Oz)

Fig. 4. Confusion matrixes for the model trained using the set of parameters #3, for three single channels (EOG, EEG Fpz-Cz, and EEG Pz-Oz). Each element of the matrix is a scalar between 0 and 1. The diagonal elements of the confusion matrix indicate the number of instances where the predicted label matches the true label, while the off-diagonal elements represent the misclassified instances. Higher values along the diagonal suggest better performance, as they reflect a greater number of correct predictions. **a.** EOG **b.** EEG (Fpz-Cz) **c.** EEG (Pz-Oz).

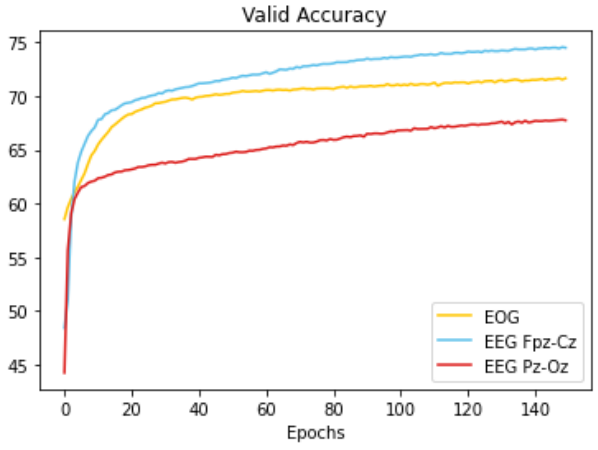


Fig. 5. Accuracy evolution with epochs for EOG, EEG (Fpz-Cz) and EEG (Pz-Oz) data.

of insomnia using optimal antisymmetric biorthogonal wavelet filter bank with ECG signals," Comput. Biol. Med. **131**, 104246 (2021).

9. M. K Pavlova and V. Latreille, "Sleep disorders," Am. J. Med. **132**, 292–299 (2019).
10. I. Arnulf, L. Lin, N. Gadoth, J. File, M. Lecendreux, P. Franco, J. Zeitzer, B. Lo, J. H. Faraco, and E. Mignot, "Kleine-Levin syndrome: a systematic study of 108 patients," Ann. Neurol. **63**, 482–493 (2008).
11. R. M. Benca, "Sleep and psychiatric disorders," Arch. Gen. Psychiatry **49**, 651 (1992).
12. J. V. Rundo and R. Downey, III, "Polysomnography," in *Clinical Neurophysiology: Basis and Technical Aspects,* (Elsevier, 2019), Handbook of clinical neurology, pp. 381–392.
13. A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. Public Health Service, US Government Printing Office* (Washington DC, 1968).
14. M. S. Mourtazaev, B. Kemp, A. H. Zwinderman, and H. A. Kamphuisen, "Age and gender affect different characteristics of slow waves in the sleep EEG," Sleep **18**, 557–564 (1995).
15. A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," Circulation **101**, E215–20 (2000).
16. A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," IEEE Trans. Neural Syst. Rehabil. Eng. **25**, 1998–2008 (2017).
17. R. S. Rosenberg and S. Van Hout, "The american academy of sleep medicine inter-scorer reliability program: sleep stage scoring," J. Clin. Sleep Med. **9**, 81–87 (2013).
18. Y. J. Lee, J. Y. Lee, J. H. Cho, and J. H. Choi, "Interrater reliability of sleep stage scoring: a meta-analysis," J. Clin. Sleep Med. **18**, 193–202 (2022).
19. I. Lambert and L. Peter-Derex, "Spotlight on sleep stage classification based on EEG," Nat. Sci. Sleep **15**, 479–490 (2023).
20. J. Fan, C. Sun, M. Long, C. Chen, and W. Chen, "EOGNET: A novel deep learning model for sleep stage classification based on single-channel

EOG signal," Front. Neurosci. **15**, 573194 (2021).
21. I. Niroshana, X. Zhu, Y. Chen, and W. Chen, "Sleep stage classification based on EEG, EOG, and CNN-GRU deep learning model," in *2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST),* (IEEE, 2019).
22. P. Jadhav and S. Mukhopadhyay, "Automated sleep stage scoring using time-frequency spectra convolution neural network," IEEE Trans. on Instrum. Meas. **71**, 1–9 (2022).
23. Q. Zhu, X. Zhao, J. Zhang, Y. Gu, C. Weng, and Y. Hu, "Eeg2vec: Self-supervised electroencephalographic representation learning," (2023).