IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 February 2023, accepted 17 March 2023, date of publication 31 March 2023, date of current version 11 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3263477

RESEARCH ARTICLE

Sleep-Energy: An Energy Optimization Method to Sleep Stage Scoring

BRUNO ARISTIMUNHA[1,2], ALEXANDRE JANONI BAYERLEIN[1], M. JORGE CARDOSO[2], WALTER HUGO LOPEZ PINAYA[2], AND RAPHAEL YOKOINGAWA DE CAMARGO[1]

[1]Center for Mathematics, Computing, and Cognition (CMCC), Federal University of ABC (UFABC), São Paulo 09210-580, Brazil
[2]Department of Biomedical Engineering, School of Biomedical Engineering and Imaging Sciences, King's College London, WC2R 2LS London, U.K.

Corresponding author: Walter Hugo Lopez Pinaya (walter.diaz_sanz@kcl.ac.uk)

This work was supported in part by the Wellcome Trust (Wellcome Innovations) under Grant WT213038/Z/18/Z. The work of Bruno Aristimunha was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES) under Grant 001. The work of M. Jorge Cardoso and Walter Hugo Lopez Pinaya was supported by Wellcome Innovations under Grant WT213038/Z/18/Z.

ABSTRACT Sleep is essential for physical and mental health. Polysomnography (PSG) procedures are labour-intensive and time-consuming, making diagnosing sleep disorders difficult. Automatic sleep staging using Machine Learning (ML) - based methods has been studied extensively, but frequently provides noisier predictions incompatible with typical manually annotated hypnograms. We propose an energy optimization method to improve the quality of hypnograms generated by automatic sleep staging procedures. The method evaluates the system's total energy based on conditional probabilities for each epoch's stage and employs an energy minimisation procedure. It can be used as a meta-optimisation layer over the sleep stage sequences generated by any classifier that generates prediction probabilities. The method improved the accuracy of state-of-the-art Deep Learning models in the Sleep EDFx dataset by 4.0% and in the DRM-SUB dataset by 2.8%.

INDEX TERMS Automated sleep staging, energy optimization, deep learning, electroencephalogram.

I. INTRODUCTION

Sleep is one of the fundamental cognitive tasks performed by the brain for keeping physical and mental health [1]. Sleep disorders can be associated with multiple health problems, including psychiatric, immune, cardiovascular, metabolic, and sexual dysfunctions [1], [2], [3], [4], [5], [6], [7], [8]. The standard gold test for diagnostic sleep disorders is the polysomnography (PSG), where the subject spends a whole night with several electrodes and sensors attached to the body (*i.e.*, Electroencephalogram - EEG, Electrooculogram - EOG, Electromyography - EMG, Electrocardiogram - ECGs, airflow, and blood oxygenation) [9]. A well-trained technician or physician then uses the captured time series to categorise each 30-second epoch in one of the multiple sleep stages, in a process called sleep scoring or sleep staging [10]. This labelling process is labour-intensive, time-consuming, and subject to errors and variability. This laborious process is a significant bottleneck that prevents more widespread testing in the population, resulting in the underdiagnosis of several sleep conditions.

Given the characteristic brain patterns governing the sleep stages, automatic sleep staging based on EEG has been studied extensively [9], [11], [12]. The first step is pre-processing the PSG signals, applying normalization, detrending, and band-pass filters, and performing artefact removal [13]. The next step is feature extraction, which generates a set of temporal, spectral, time-frequency, spatial, and other features. Finally, one can use different machine learning (ML)-based methods over the extracted features, such as decision trees [14], Hidden Markov Models [15], Support Vector Machines [16], Self-Organizing Maps [17], Random Forests [18], and more recently, Deep Learning Models [19], [20].

Deep learning models learn to identify important features in the EEG signals and use these features to classify the sleep stages when provided with sufficient training examples [19], [20]. The recent increase in available public datasets [21] enabled these methods to achieve state-of-the-art performance on sleep staging [22].

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno M. Garcia.

Adoption in clinical practice requires automatic sleep staging methods to have similar reliability levels to trained technicians and physicians. Most state-of-the-art models classify each 30-second epoch in isolation or considering only the immediate neighbouring epochs. However, in clinical practice, a broader context regarding the sleep stages around each epoch is also important and should be considered by models. For instance, Yang et al. [23] propose the use of a Hidden Markov Model (HMM) to improve the predictions made by a single electrode convolutional neural network by exploring sleep stage transition probabilities.

In this work, we propose an energy optimisation method called sleep-energy to improve the quality of hypnograms generated by automatic sleep staging procedures. The method evaluates the system's total energy based on conditional probabilities for each epoch's stage and employs an energy minimisation procedure. The energy comprises conditional probabilities from the ML model stage predictions, the prevalence of each sleep stage, and the stage transition probabilities. One can apply this method as a meta-optimisation layer over the sleep stage sequences generated by any ML classifier that outputs class probabilities, including state-of-the-art Deep Learning models. The highlights of our study are:

1) Our method can be used as a meta-optimisation layer to improve sleep stage sequence predictions from any ML-based model that generates prediction probabilities;
2) Our energy optimisation method reduces incoherent sleep stage transitions and improves the distribution of less frequent stages, such as N2, N3, and REM stages, while using low computational resources;
3) The proposed method outperforms the post-optimisation based on Hidden Markov Model - HMM, under a subject-independent evaluation paradigm, using Sleep-EDFx [24] and DRM-SUB [25] datasets.
4) To the best of our knowledge, this is the first use of an energy optimisation method on a sleep-scoring task or other EEG-based classification tasks.

II. PROPOSED METHOD

A. AUTOMATIC SLEEP STAGING

Sleep is composed of multiple cycles of sleep stages, which repeat during the night: Wake (W), Rapid Eye Movements (REM, denoted as R in this study), and Non-REM Stages 1 (N1), 2 (N2) and 3 (N3).[1] The final hypnogram should contain the sleep stage s_t at each 30s epoch t, where $t \in \{1, 2, \ldots, T\}$.

In typical setups, an ML model receives the EEG signal from an epoch t and returns the probabilities of these epochs belonging to each of the five possible sleep stages. When applied to the whole hypnogram, it generates a *predicted probability matrix* with shape $(5, T)$. With this predicted probability matrix, we apply a *argmax* function to extract the most likely sleep stage according to the model, generating a candidate hypnogram. We denote the predicted probability matrix as $P_{pred}(s_t)$ and use the notation s_t^{pred} to indicate the stage with the highest probability.

One can also extract other information from the predictions, such as the *confusion matrix* containing the probabilities of mispredictions. This matrix contains the actual sleep stage in the rows and the predicted state in the columns. By normalising the sum of values on each column, each matrix entry will contain the probability of the actual stage being s_t (row) given that the ML model predicted a state s_t^{pred} (column). We represent it as $P_{conf}(s_t|s_t^{pred})$.

During sleep, stage transitions have different probabilities. We construct a *transition probability matrix* by (i) measuring the frequencies of transitions among different stages; (ii) organising the source stages in the rows and the target stages in the columns; and (iii) normalising the sum of each row in the matrix. We use the notation $P_{trans}(s_{t-1} \rightarrow s_t)$ to denote the probability of transition from state s at time $t-1$ to state s at time t.

B. SLEEP-ENERGY OPTIMISATION METHOD FOR SLEEP STAGING

In this work, we propose sleep-energy, an energy optimisation method that uses the prediction probability, confusion, and transition probabilities matrices to optimise the energy of the generated hypnogram. The method evaluates the initial energy of the system and iteratively optimises the sleep stage sequence to reduce this overall energy, as shown in Figure 1.

Although one typically selects the stage with the highest probability, there can be multiple stages with similar probabilities, or the predictions can be wrong. We leverage the information from transition probabilities, predicted probabilities, and confusion matrices to improve the quality of the constructed hypnogram. These matrices will indicate the likelihood for each stage s on each epoch t, based on different criteria: (i) the sleep stage at the previous epoch $t-1$, (ii) the confidence of the ML model prediction, and (iii) the errors of the ML model. For instance, we could have $P_{pred}(s, t) = 0.8$, $P_{conf}(s|s_t^{pred}) = 0.5$, $P_{trans}(s' \rightarrow s, t) = 0.1$, and $P_{trans}(s \rightarrow s'', t+1) = 0.7$. We would have a different set of probabilities for each sleep stage $s \in \{W, N1, N2, N3, R\}$.

We define the whole hypnogram as the variable to be optimised. To optimise the sequence of sleep stages s_t, where $t \in \{1, 2, \ldots, T\}$, we define an energy function $E(s_t)$ for each epoch. We consider the total energy of the hypnogram as the sum of $E(s_t)$ for all t and use an optimisation technique to minimise its value. We define the energy function as:

$$E(s_t) = \alpha * [e_{trans}(s_{t-1} \rightarrow s_t) + e_{trans}(s_t \rightarrow s_{t+1})] + (1-\alpha) * [e_{conf}(s_t|s_t^{pred}) + e_{pred}(s_t)] \quad (1)$$

where $\alpha \in [0, 1]$ indicates the relative contribution of the transition probability matrix in the energy function compared to the confusion matrix and prediction probabilities. The

[1] As defined by the American Academy of Sleep Medicine (AASM).

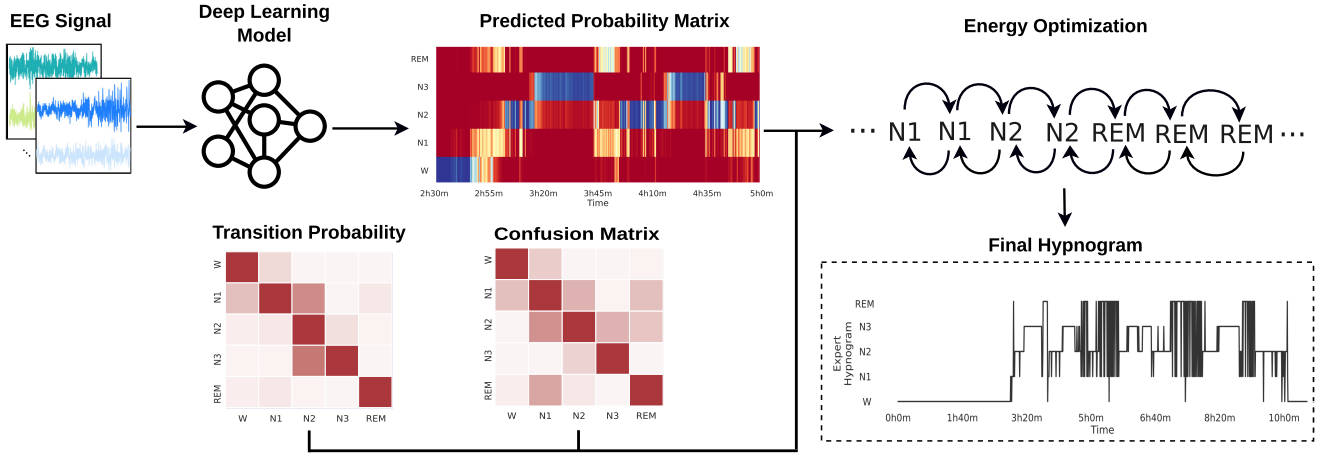


FIGURE 1. The general pipeline for the energy optimisation method. The ML model receives a sequence of 30*s* epochs and generates the predicted probability matrix. The energy optimiser uses this matrix to generate a candidate hypnogram and evaluates the system's energy using the confusion and transition probability matrices. It then iteratively optimises the sleep stage sequence to reduce this overall energy and generates the final hypnogram.

errors are given by:

$$e_{trans}(s_{t-1} \rightarrow s_t) = \frac{1-\epsilon_t}{P_{trans}(s_{t-1} \rightarrow s_t) - \epsilon_t} - 1 \quad (2)$$

$$e_{conf}(s_t | s_t^{pred}) = \frac{1-\epsilon_p}{P_{conf}(s_t | s_t^{pred}) - \epsilon_p} - 1 \quad (3)$$

$$e_{pred}(s_t) = \frac{1-\epsilon_p}{P_{pred}(s_t) - \epsilon_p} - 1 \quad (4)$$

These equations have the format $1/p-1$ so that the energy is zero when the respective probability p is one and increases when p goes to zero. To prevent divergence at $p \rightarrow 0$ and limit its value, we include the constants ϵ_t and ϵ_p. When $p = 0$, the error will be $1/\epsilon_t - 1$ and $1/\epsilon_p - 1$, respectively. These constants also define the shape of the probability curve, with steeper curves for smaller ϵ values.

Our next step is to employ an optimisation technique for discrete search spaces to reduce the system's energy. We employ a stochastic optimisation process to prevent the system from being stuck in local minima.

C. ENERGY OPTIMISATION USING SIMULATED ANNEALING

We use an optimization strategy often used to model physical systems called *simulated annealing*. This strategy simulates the controlled heating and cooling of materials to change their properties in metallurgy applications. Translating into an optimization algorithm, the idea is to start the process with a higher temperature, initially permitting sleep stage transitions to higher energy (lower probability) stages and gradually reducing the temperature so that transitions to lower probability stages become less likely.

The process is conducted over multiple steps k, and the global state of the system at each moment is given by $S^k = (s_1^k, s_2^k, \ldots, s_T^k)$. At each step of the simulated annealing process, the algorithm selects one candidate epoch t to update its value s_t^k. We update a single position at each step because it also changes the energy of its neighbours $t-1$ and $t+1$. The total energy of the hypnogram at update step k is:

$$E(S^k) = \sum_{t=1}^{T} E(s_t^k; S^k) \quad (5)$$

where $E(s_t^k, t; S^k)$ is the energy for stage s_t at the optimisation step k (Equation 1) when considering that the hypnogram is at global state S^k.

We select the position t to update at each step k using the Boltzmann distribution, which depends on the state energy of each position $E(s_t^k)$ and the current temperature $1/\beta_k$. This distribution is the same as the softmax function and is given by:

$$P_{update}(t) = 1/Z * e^{-\beta_k * E(s_t^k; S^k)} \quad (6)$$

where Z is the normalization factor, also called partition function, given by:

$$Z = \sum_{s'=1}^{T} e^{-\beta * E(s_{s'}^k; S^k)}$$

Finally, the algorithm selects the new state s_t^{k+1} for the selected position t using the same probability distribution:

$$P(s_t^{k+1}) = 1/Z * e^{-\beta * \Delta E(s_t^{k+1}; S^k)} \quad (7)$$

where s^{k+1} is one the 5 possible states $\{W, 1, 2, 3, R\}$, Z is the normalisation factor over these states, and $\Delta E(s_t^{k+1}; S^k)$ is the difference in the energy $E(S^k)$ of the hypnogram caused by the transition from state s_t^k to the new state s_t^{k+1}:

$$\Delta E(s_t^{k+1}; S^k) = \sum_{t'=t-1}^{t+1} E(s_{t'}^{k+1}; S^{k+1}) - E(s_{t'}^k; S^k) \quad (8)$$

where S^{k+1} denotes the global state of the system at step $k+1$ if the transition $s_t^k \rightarrow s_t^{k+1}$ occurs. We repeat the optimization process for K update steps, reducing the temperature $1/\beta_k$ using a given annealing schedule.

III. EXPERIMENTAL METHODOLOGY

A. DATASETS

We evaluated the energy optimisation method using two classical polysomnography datasets, the sleep EDFx expanded dataset, sleep cassette subset,[2] and DREAMS subjects database (DRM-SUB) [25].[3] The polysomnography exams contain two or three EEG electrodes (Fpz-Ca and Pz-Oz in the first dataset; FP1-A1, O1-A1, and Cz-A1 or C3-A1 depending of the subject in the second dataset), 1 EOG electrode, and 1 EMG electrode. We used only the EEG channels to perform the classification task. The EDFx dataset has 78 healthy subjects with ages between 25 and 101 and up to two sessions per subject, totalling 153 records [21], [24]. The DRM-SUB dataset contains 20 subjects with one whole night of polysomnography recording for each.

One specialist determined the corresponding sleep stage for each 30-second epoch in both datasets. Sleep stages were originally defined following the Rechtschaffen & Kales protocol and later converted to the AASM protocol, which has stages Wake, N1, N2, N3 and REM, which we denote $\{W, R, 1, 2, 3\}$. We re-sampled the EEG datasets to 100Hz. EEG series were normalised to a scale between $\pm 1\mu V$ and combined with a temporal bandpass filter $[0, 30]$ Hz to remove high-frequency noise and artefacts. We selected only the EEG channels to perform the classification task and discarded EOG and EMG data.

B. MODEL TRAINING AND EVALUATION

We used the state-of-the-art Chambon [19] and U-Sleep [20] neural networks for sleep stage classification. Chambon is a small and efficient network with 11 layers, ten for feature extraction and one for classification. U-Sleep consists of an adapted U-Net for high-frequency features [26], with 11 encoding and 11 decoding layers. Both architectures delivered solid results in several machine learning paradigms and datasets [27], [28], [29], [30], [31].

We trained both neural network models from scratch with Xavier weight initialisation [32] and AdamW optimiser [33], with $\beta_1 = 0.9$, $\beta_2 = 0.999$, learning rate 1^{-3}, weight decay 0.0005 (following [33] recommendation), and Cross-Entropy Loss. We used batch size 32 and 300 epochs, with early stopping regularisation and 30 epochs patience, following the methodology proposed by [31] for systematic comparisons.

We used subject-wise 5-fold cross-validation, with 20% of the subjects for testing, and split the remaining subjects into 80% for training and 20% for validation. We used the validation data for early stopping for the neural network training. For sleep-energy, we used the training data for generating the transition probabilities and the validation data for the confusion matrix.

We evaluated sleep-energy using the test subjects. We first generate the predicted probabilities for the sleep stage at each epoch t and then optimise the energy of the entire sleep sequence using the simulated annealing procedure. We used the parameters $\alpha = 0.5$, $\epsilon_t = 0.1$, and $\epsilon_p = 0.1$ unless otherwise noted. We used the accuracy, balanced accuracy and F1 Score metrics to compare the improvements obtained using sleep-energy. Finally, we used the paired Wilcoxon signed-rank test to assess statistical significance, with Holm-Bonferroni corrections for multiple comparisons. We used a significance level of 5% for the statistical tests ($p-$value < 0.05).

We trained the deep learning architecture using pytorch [34] and braindecode [35] on an Nvidia DGX with 4 A100 boards. We implemented the energy method in Python, using numpy [36], scikit-learn and scipy. The source code of the energy model and experimental results are available at https://github.com/rycamargo/sleep-energy.

IV. RESULTS

Using sleep-energy improved the accuracy of both neural network models (U-Sleep and Chambon) on both datasets (Sleep EDFx and DRM-SUB) with high statistical significance, as shown in Table 1. The most significant improvement was 4.0% for the U-Sleep with the EDFx dataset, and the smallest was 0.7% for Chambom in the DRM-SUB dataset. The smaller Chambon model had better accuracy on the small DRM-SUB dataset, with only 20 subjects, while the larger U-Sleep was better on the EDFx dataset, with 87 subjects. The energy model also improved the balanced accuracy in all cases (Sleep-EDFx was statistically significant), which indicates that the model does not simply enhance the accuracy of the most prevalent sleep stage classes to improve the overall accuracy. Moreover, the average F1 Score was also slightly improved (although not statistically significant), which also indicates that the model maintains an equilibrium between the precision and recall of the multiple sleep stage classes.

Looking at the predictions for each subject (Figure 2), we see a general improvement in the accuracy when using the sleep-energy, except for outlier subjects with very low initial accuracy. In this case, the model could not improve the accuracy since it depends on the quality of the predictions from the original models to generate the confusion matrices and prediction probabilities. Also, each epoch's neighbours must have correct predictions to adjust the system's energy correctly.

When evaluating the individual sleep stages (Figure 3), the improvements in accuracy and F1-score occurred primarily for the REM and N2 classes ($p < 0.01$). At the same time, for N3 there was a small increase in accuracy and decrease in F1-score. For W and N1 there was no statistical difference in using the energy optimisation.

The improvements in the accuracy for each sleep stage class are also evident when comparing the confusion matrices for the original U-Sleep predictions and the energy-optimised ones (first row in Figure 4 for one subject). The energy method improves the prediction for most stage classes (main diagonal), except for stage N3.

[2] https://www.physionet.org/content/sleep-edfx/1.0.0/
[3] https://zenodo.org/record/2650142#.ZB4fs9KZOV4

TABLE 1. Sleep stage prediction accuracy, balanced accuracy and average F1 score (in %) using 5-fold cross-validation (mean ± standard deviation, $^{*}p < 0.05$ and $^{**}p < 0.01$).

Dataset	Sleep-EDFx			SUB-DRM		
Model	Accuracy	Balanced Acc	Average F-1	Accuracy	Balanced Acc	Average F-1
U-Sleep	74.30±9.98	74.00±10.06	65.24±12.10	64.56±9.67	61.09±9.43	57.96±9.99
U-Sleep+Energy	78.31±9.94**	75.72±10.09*	68.68±12.21	67.40±10.92**	63.22±10.44	59.70±11.37
Chambon	72.16±10.91	68.94±9.76	61.28±12.06	76.78±6.69	73.36±8.76	70.89±8.39
Chambon+Energy	75.31±10.66**	70.16±10.60**	63.51±12.54	77.47±7.09**	74.14±8.96	71.54±9.07

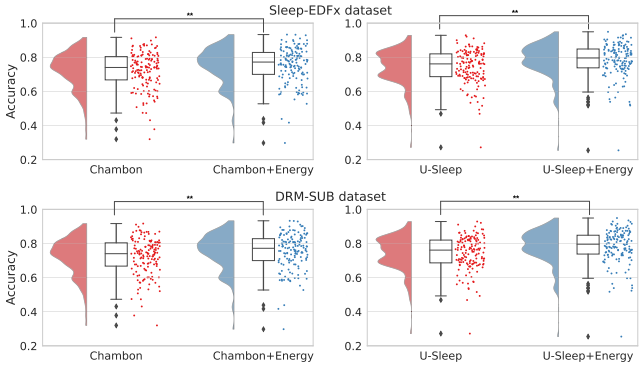


FIGURE 2. Boxplot with accuracy distribution per EEG recording before and after the energy optimisation ($^{**}p < 0.01$).

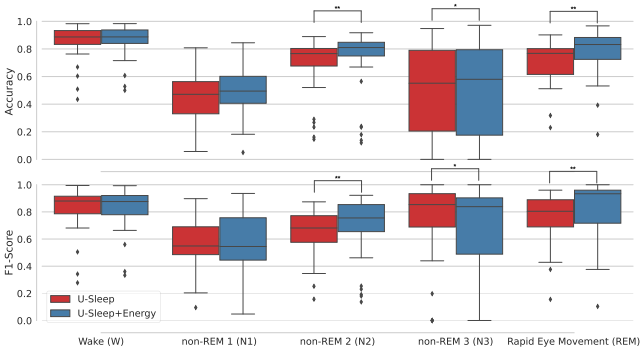


FIGURE 3. Distribution of two machine learning metrics for each sleep stage in the Sleep-EDFx dataset after optimisation of U-Sleep model predictions with sleep-energy ($^{*}p < 0.05$ and $^{**}p < 0.01$).

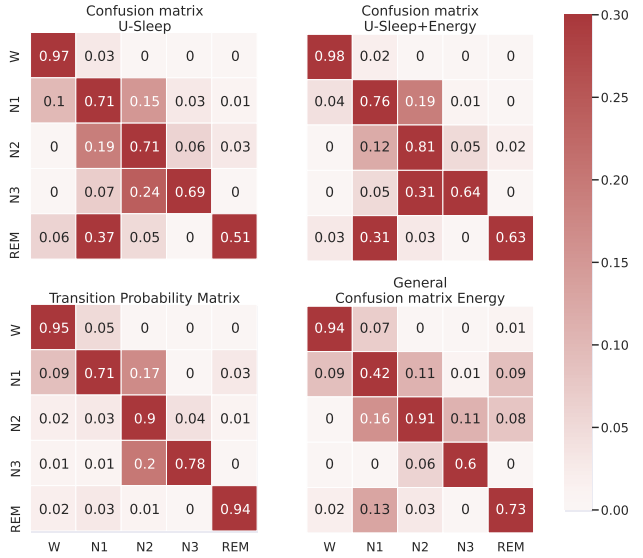


FIGURE 4. (Top): Confusion Matrix for the sleep stages predictions from the U-Sleep model for the EDFx dataset before and after sleep-energy application. Lines represent the target sleep stage, and columns the predicted stages. (Bottom): Input matrices to sleep-energy. Transition probability matrix for sleep stages on the EDFx dataset (left-hand side) and confusion matrix for the U-Sleep model using the validation set (right-hand side), where lines represent target stages and columns the predicted ones.

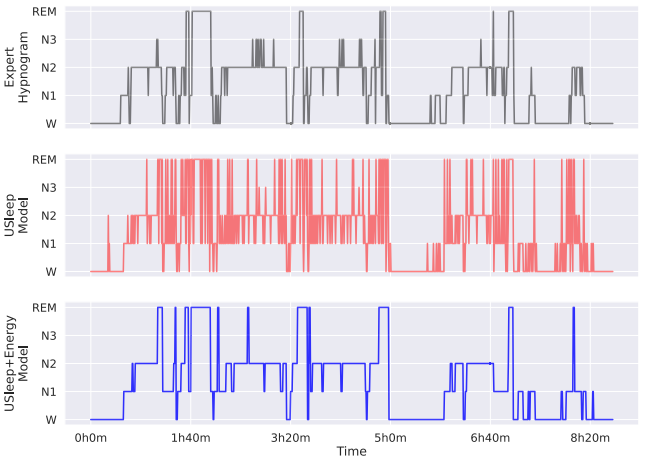


FIGURE 5. Example of a hypnogram from the EDFx dataset with the target defined by a clinical expert (top), followed by the U-Sleep predictions (middle), and the improved predictions using sleep-energy (bottom).

The origin of the accuracy improvements becomes clear when evaluating the hypnogram (Figure 5). The energy optimisation corrects the multiple mispredictions of individual epochs by adjusting them according to their neighbours and making the hypnogram less noisy and more similar to the expert hypnograms. The model depends on the initial quality of the predictions, as shown in the period before 1h40, where there were multiple predictions for REM stages. Although the sleep-energy removed most of these predictions, some remained.

The transition probability matrix used as input for sleep-energy shows that, in most cases, the sleep stage tends to stay the same in consecutive epochs (Figure 4, bottom left). There are also some transitions that never or rarely occur and cause a substantial increase in the model's energy if maintained. But it also depends on the confusion matrix (Figure 4, bottom right) and the probability for each prediction generated by the neural network model. We should

TABLE 2. Sleep stage accuracy, balanced accuracy and average F1 score (in %) using 5-fold cross-validation for different α values (mean ± standard deviation).

Model	α	Sleep-EDFx			SUB-DRM		
		Accuracy	Balanced Accuracy	Average F1 Score	Accuracy	Balanced Accuracy	Average F1 Score
U-Sleep	0.00	74.76 ± 9.81	74.12 ± 9.93	65.55 ± 12.00	63.84 ± 9.36	60.35 ± 8.75	57.14 ± 9.45
	0.25	77.57 ± 9.72	75.88 ± 9.95	68.04 ± 12.19	66.03 ± 10.39	62.51 ± 9.90	59.00 ± 10.48
	0.50	78.31 ± 9.94	75.72 ± 10.09	68.68 ± 12.21	67.40 ± 10.92	63.22 ± 10.44	59.70 ± 11.37
	0.75	79.30 ± 9.46	75.12 ± 9.49	69.00 ± 11.57	67.97 ± 10.54	63.71 ± 10.30	60.41 ± 10.96
	1.00	69.59 ± 9.40	62.91 ± 8.27	57.47 ± 9.75	64.91 ± 9.51	58.43 ± 9.57	55.73 ± 10.09
Chambon	0.00	72.08 ± 10.80	68.57 ± 9.68	61.06 ± 11.96	76.70 ± 6.61	73.10 ± 8.41	70.64 ± 7.89
	0.25	74.73 ± 10.61	70.50 ± 10.33	63.41 ± 12.41	77.12 ± 6.99	73.87 ± 9.03	71.21 ± 8.93
	0.50	75.31 ± 10.66	70.16 ± 10.60	63.51 ± 12.54	77.47 ± 7.09	74.14 ± 8.96	71.54 ± 9.07
	0.75	75.63 ± 10.17	68.57 ± 10.18	62.66 ± 12.02	77.27 ± 7.43	73.81 ± 8.78	71.24 ± 9.22
	1.00	67.42 ± 10.26	59.35 ± 9.48	53.61 ± 10.52	74.99 ± 5.96	67.82 ± 7.75	66.21 ± 8.25

note that the confusion matrix estimated from the validation set has some differences from the test set (Figure 4), possibly due to variability among individuals. A better estimation of the confusion matrix could improve the energy method optimisations.

We also evaluate the effect of using different values for α in Equation 1. When $\alpha = 0$, we are using only the prediction probability and confusion matrix information in the energy model, and with $\alpha = 1$, we use only the stage transition probability matrix. We used $\alpha = 0.5$ as the default value in all other experiments to provide a balanced contribution for the transition and prediction/confusion matrices. We can note that using α between 0.25 and 0.75 delivered similar results (Table 2, showing the method's robustness to changes in this parameter value. With $\alpha = 1$, when only information on transition probabilities is used, there was a clear decrease in all metrics, while with $\alpha = 0$, the result is similar to that without sleep-energy.

V. DISCUSSION

Other projects also explored using sleep stage transition probabilities to improve automatic sleep staging. For instance, Yang et al. [23] used a Hidden Markov Model (HMM) to improve the sleep stage sequences generated by a single electrode convolutional neural network model, using sleep stage transition probabilities for the HMM model. They obtained an improvement of 2.69% in accuracy and 6% in the mean F1 score in the EDFx dataset and an F1 score of 1.61% in the DRM-SUB dataset. Our energy optimisation method leveraged the information on prediction errors and probabilities from the neural networks to obtain 4.01% and 3.15% improvements in accuracy in the U-Sleep and Chambon models in the EDFx dataset. For the DRM-SUB dataset, the improvements were 2.84% and 0.68% for the U-Sleep and Chambon models. We obtained the most significant improvements when using the EDFx dataset, probably due to more data availability, as it allows a better estimation of the confusion and transition matrices.

A possible approach to use information from neighbouring is to provide the neural network classifier with the EEG signal from an epoch and its immediate neighbours. This provides a better context for the neural network to classify the epoch sleep stage. The most recent models are based on the Transformer architecture [37], [38], [39] and use the Attention mechanism to obtain context information from neighbouring epochs. It is still being determined if applying sleep-energy to the sleep stage sequences generated by these Transformer-based models could improve their results since they already use contextual information.

We did not fine-tune the model parameters, which could improve the results. For instance, we showed in Table 2 that depending on the neural network model, and dataset, different α values provided better results. Also, different values of of ϵ_t and ϵ_p in Equations 2, 3 and 4 could alter the results. These parameters could be optimised using the validation sets from the cross-validation. Nevertheless, the differences were small, as shown in Table 2 for *alpha* and reasonable values of ϵ_t and ϵ_p.

One possible improvement to the model is to use different sleep transition and confusion matrices, depending on the moment of the sleep period, the age and gender of the subject, and sleep abnormalities. For instance, the deep sleep N3 stage appears mostly at the beginning of sleep, while REM occurs mostly at the end. Also, time spent on deep sleep tends to decrease in older subjects, while the number of WASO (wake after sleep onset) events increases [40]. Classifying automatically if a subject has sleep abnormalities is more challenging, and multiple machine-learning techniques have been proposed [41]. Providing sleep transition and confusion matrices should generate substantial improvements to the sleep model optimisations as it will use probabilities that resemble the expected sleep stage patterns more closely.

ACKNOWLEDGMENT

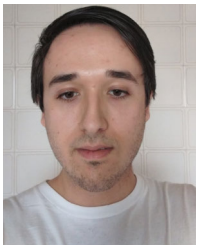
For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

REFERENCES

[1] M. A. Reimer and W. W. Flemons, "Quality of life in sleep disorders," *Sleep Med. Rev.*, vol. 7, no. 4, pp. 335–349, 2003.
[2] W. W. Flemons, "Obstructive sleep apnea," *New England J. Med.*, vol. 347, no. 7, pp. 498–504, 2002.
[3] B. M. Altevogt and H. R. Colten, "*Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. New York, NY, USA: Committee on Sleep Medicine and Research, 2006.
[4] S. Chokroverty, "Overview of sleep & sleep disorders," *Indian J. Med. Res.*, vol. 131, no. 2, pp. 126–140, 2010.
[5] K. Chin, "Overcoming sleep disordered breathing and ensuring sufficient good sleep time for a healthy life expectancy," *Proc. Jpn. Acad. B*, vol. 93, no. 8, pp. 609–629, 2017.

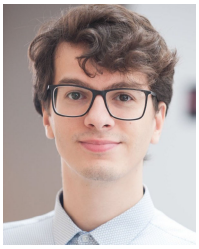
[6] J. W. Cho and J. F. Duffy, "Sleep, sleep disorders, and sexual dysfunction," *World J. Men's Health*, vol. 37, no. 3, pp. 261–275, 2019.
[7] B.-H. Huang, B. del Pozo Cruz, A. Teixeira-Pinto, P. A. Cistulli, and E. Stamatakis, "Influence of poor sleep on cardiovascular disease-free life expectancy: A multi-resource-based population cohort study," *BMC Med.*, vol. 21, no. 1, p. 75, Mar. 2023, doi: 10.1186/s12916-023-02732-x.
[8] S. Stranges, W. Tigbe, F. X. Gómez-Olivé, M. Thorogood, and N.-B. Kandala, "Sleep problems: An emerging global epidemic? Findings from the INDEPTH WHO-SAGE study among more than 40,000 older adults from 8 countries across Africa and Asia," *Sleep*, vol. 35, no. 8, pp. 1173–1181, Aug. 2012.
[9] H. Phan and K. Mikkelsen, "Automatic sleep staging of EEG signals: Recent development, challenges, and future directions," *Physiol. Meas.*, vol. 43, no. 4, Apr. 2022, Art. no. 04TR01.
[10] M. H. Silber, S. Ancoli-Israel, M. H. Bonnet, S. Chokroverty, M. M. Grigg-Damberger, M. Hirshkowitz, S. Kapen, S. A. Keenan, M. H. Kryger, T. Penzel, M. R. Pressman, and C. Iber, "The visual scoring of sleep in adults," *J. Clin. Sleep Med.*, vol. 3, no. 2, pp. 121–131, Mar. 2007.
[11] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *J. Neural Eng.*, vol. 16, no. 5, Oct. 2019, Art. no. 051001.
[12] O. Yildirim, U. Baloglu, and U. Acharya, "A deep learning model for automated sleep stages classification using PSG signals," *Int. J. Environ. Res. Public Health*, vol. 16, no. 4, p. 599, Feb. 2019.
[13] S. Motamedi-Fakhr, M. Moshrefi-Torbati, M. Hill, C. M. Hill, and P. R. White, "Signal processing techniques applied to human sleep EEG signals—A review," *Biomed. Signal Process. Control*, vol. 10, pp. 21–33, Mar. 2014.
[14] M. Hanaoka, M. Kobayashi, and H. Yamazaki, "Automated sleep stage scoring by decision tree learning," in *Proc. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2, Oct. 2001, pp. 1751–1754.
[15] L. G. Doroshenkov, V. A. Konyshev, and S. V. Selishchev, "Classification of human sleep stages based on EEG processing using hidden Markov models," *Biomed. Eng.*, vol. 41, no. 1, pp. 25–28, Jan. 2007.
[16] B. Koley and D. Dey, "An ensemble system for automatic sleep stage classification using single channel EEG signal," *Comput. Biol. Med.*, vol. 42, no. 12, pp. 1186–1195, 2012.
[17] A. M. Mora, C. M. Fernandes, L. J. Herrera, P. A. Castillo, J. J. Merelo, F. Rojas, and A. C. Rosa, "Sleeping with ants, SVMs, multilayer perceptrons and SOMs," in *Proc. 10th Int. Conf. Intell. Syst. Design Appl.*, Nov. 2010, pp. 126–131.
[18] S. Santaji and V. Desai, "Analysis of EEG signal to classify sleep stages using machine learning," *Sleep Vigilance*, vol. 4, no. 2, pp. 145–152, Dec. 2020.
[19] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 758–769, Apr. 2018.
[20] M. Perslev, S. Darkner, L. Kempfner, M. Nikolic, P. J. Jennum, and C. Igel, "U-sleep: Resilient high-frequency sleep staging," *NPJ Digit. Med.*, vol. 4, no. 1, pp. 1–12, Apr. 2021.
[21] A. L. Goldberger, L. A. N. 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*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.
[22] S. Mishra and R. Birok, "Literature review: Sleep stage classification based on EEG signals using artificial intelligence technique," in *Recent Trends in Communication and Electronics*. Boca Raton, FL, USA: CRC Press, 2021, pp. 241–244.
[23] B. Yang, X. Zhu, Y. Liu, and H. Liu, "A single-channel EEG based automatic sleep stage classification method leveraging deep one-dimensional convolutional neural network and hidden Markov model," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102581.
[24] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. C. Kamphuisen, and J. J. L. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 9, pp. 1185–1194, Sep. 2000.
[25] S. Devuyst, T. Dutoit, J. F. Didier, F. Meers, E. Stanus, P. Stenuit, and M. Kerkhofs, "Automatic sleep spindle detection in patients with sleep disorders," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2006, vol. 1, no. 1, pp. 3883–3886, doi: 10.1109/IEMBS.2006.259298.
[26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.
[27] H. Banville, O. Chehab, A. Hyvärinen, D.-A. Engemann, and A. Gramfort, "Uncovering the structure of clinical EEG signals with self-supervised learning," *J. Neural Eng.*, vol. 18, no. 4, 2021, Art. no. 046020.
[28] X. Wei et al., "2021 BEETL competition: Advancing transfer learning for subject independence and heterogenous EEG data sets," in *Proc. NeurIPS Competitions Demonstrations Track*, in Proceedings of Machine Learning Research, vol. 176, D. Kiela, M. Ciccone, and B. Caputo, Eds. PMLR, Dec. 2022, pp. 205–219. [Online]. Available: https://proceedings.mlr.press/v176/wei22a.html
[29] C. Rommel, T. Moreau, J. Paillard, and A. Gramfort, "CADDA: Class-wise automatic differentiable data augmentation for EEG signals," in *Proc. 10th Int. Conf. Learn. Represent. (ICLR)*. OpenReview.net, Apr. 2022.
[30] H. Banville, S. U. N. Wood, C. Aimone, D.-A. Engemann, and A. Gramfort, "Robust learning from corrupted EEG with dynamic spatial filtering," *NeuroImage*, vol. 251, May 2022, Art. no. 118994.
[31] C. Rommel, J. Paillard, T. Moreau, and A. Gramfort, "Data augmentation for learning predictive models on EEG: A systematic comparison," *J. Neural Eng.*, vol. 19, no. 6, p. 066020, 2022.
[32] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
[33] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–19.
[34] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, vol. 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2019, pp. 8024–8035.
[35] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Aug. 2017.
[36] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The NumPy array: A structure for efficient numerical computation," *Comput. Sci. Eng.*, vol. 13, no. 2, pp. 22–30, 2011.
[37] E. Eldele, Z. Chen, C. Liu, M. Wu, C. K. Kwoh, X. Li, and C. Guan, "An attention-based deep learning approach for sleep stage classification with single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 809–818, 2021.
[38] H. Phan, K. Mikkelsen, O. Y. Chen, P. Koch, A. Mertins, and M. De Vos, "SleepTransformer: Automatic sleep staging with interpretability and uncertainty quantification," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 8, pp. 2456–2467, Aug. 2022.
[39] J. Pradeepkumar, M. Anandakumar, V. Kugathasan, D. Suntharalingham, S. L. Kappel, A. C. De Silva, and C. U. S. Edussooriya, "Towards interpretable sleep stage classification using cross-modal transformers," 2022, *arXiv:2208.06991*.
[40] L. G. Doroshenkov, "Classification of human sleep stage," *Biomed. Eng.*, vol. 41, no. 1, pp. 24–28, 2007.
[41] S. Xu, O. Faust, S. Seoni, S. Chakraborty, P. D. Barua, H. W. Loh, H. Elphick, F. Molinari, and U. R. Acharya, "A review of automated sleep disorder detection," *Comput. Biol. Med.*, vol. 150, Nov. 2022, Art. no. 106100.

BRUNO ARISTIMUNHA is currently pursuing the Ph.D. degree in computer science with the Federal University of ABC. His research interests include brain–computer interfaces, brain decoding, electroencephalogram, machine learning, and deep learning.



ALEXANDRE JANONI BAYERLEIN is currently pursuing an undergraduate degree in computer science and neuroscience with the Federal University of the ABC. His research interests include machine learning, deep learning, electroencephalogram, and data classification.

M. JORGE CARDOSO is a Reader of artificial medical intelligence with King's College London, where he leads a research portfolio on big data analytics, quantitative radiology, and value-based healthcare. He is also the CTO of the new London Medical Imaging and AI Centre for Value-Based Healthcare. He was a Lecturer at UCL, the Technical Lead of the Quantitative Radiology Initiative with the National Hospital for Neurology and Neurosurgery (NHNN), and the Engineering Lead of the Neuro-oncology Flagship Programme, Institute of Healthcare Engineering, UCL. He has more than 12 years of expertise in advanced image analysis, big data, and artificial intelligence and co-leads the development of NiftyNet, a deep-learning platform for artificial intelligence in medical imaging. He is also the Founder of BrainMiner, a medtech startup aiming to bring quantitative biomarkers and predictive models to neurological care.



WALTER HUGO LOPEZ PINAYA received the Ph.D. degree in computer science and information engineering (neuroscience and cognition) from the Federal University of ABC, Brazil. He is a Research Fellow with the Artificial Medical Intelligence Group (AMIGO), King's College London. Recently, he has led the MONAI Generative Models Initiative to make applications involving generative models more accessible to the medical image community. His main research interests include developing generative models-based methods to improve triage and anomaly detection in medical images.

RAPHAEL YOKOINGAWA DE CAMARGO received the B.S. degree in molecular sciences from the University of Sao Paulo, where he received the master's degree from the Institute of Physics and the Ph.D. degree in computer science from the Institute of Mathematics and Statistics. He is an Associate Professor with the Center for Mathematics, Computing, and Cognition, Federal University of ABC (UFABC), where he is currently a permanent Advisor of the master's and doctoral programs in computer science as well as neuroscience and cognition. He has the experience and works in high-performance computing (GPGPU, heterogeneous computing, and scheduling), machine learning (high-performance computing, smart cities, and neuroscience applications), and computational neuroscience (computational models of memory and time perception, connectivity analysis, and decoding of neural signals).