applied sciences

MDPI

Article

Efficient Sleep–Wake Cycle Staging via Phase–Amplitude Coupling Pattern Classification

Vinícius Rosa Cota [1,*], Simone Del Corso [2], Gianluca Federici [2], Gabriele Arnulfo [2] and Michela Chiappalone [1,2,*]

[1] Rehab Technologies Lab, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy
[2] Department of Informatics, Bioengineering, Robotics, System Engineering (DIBRIS), University of Genova, Via all'Opera Pia 13, 16145 Genova, Italy; gabriele.arnulfo@unige.it (G.A.)
* Correspondence: vinicius.rosacota@iit.it (V.R.C.); michela.chiappalone@unige.it (M.C.); Tel.: +39-327-212-4371 (V.R.C.); +39-380-357-7986 (M.C.)

Abstract: The objective and automatic detection of the sleep–wake cycle (SWC) stages is essential for the investigation of its physiology and dysfunction. Here, we propose a machine learning model for the classification of SWC stages based on the measurement of synchronization between neural oscillations of different frequencies. Publicly available electrophysiological recordings of mice were analyzed for the computation of phase–amplitude couplings, which were then supplied to a multilayer perceptron (MLP). Firstly, we assessed the performance of several architectures, varying among different input choices and numbers of neurons in the hidden layer. The top performing architecture was then tested using distinct extrapolation strategies that would simulate applications in a real lab setting. Although all the different choices of input data displayed high AUC values (>0.85) for all the stages, the ones using larger input datasets performed significantly better. The top performing architecture displayed high AUC values (>0.95) for all the extrapolation strategies, even in the worst-case scenario in which the training with a single day and single animal was used to classify the rest of the data. Overall, the results using multiple performance metrics indicate that the usage of a basic MLP fed with highly descriptive features such as neural synchronization is enough to efficiently classify SWC stages.

Keywords: slow-wave sleep; REM sleep; wakefulness; cross-frequency coupling; neural synchronizations; modulation index; multilayer perceptron; signal processing

Citation: Cota, V.R.; Del Corso, S.; Federici, G.; Arnulfo, G.; Chiappalone, M. Efficient Sleep–Wake Cycle Staging via Phase–Amplitude Coupling Pattern Classification. *Appl. Sci.* **2024**, *14*, 5816. https://doi.org/10.3390/app14135816

Academic Editor: Hariton-Nicolae Costin

Received: 14 May 2024
Revised: 25 June 2024
Accepted: 2 July 2024
Published: 3 July 2024

1. Introduction

Historically considered merely a transient shutdown of brain activity, sleep is today better understood in not only its intricate generation mechanisms [1] but also in its crucial roles in supporting several different neural functions and homeostasis in general [2]. In fact, the alternation of states that characterize the sleep–wake cycle (SWC) in a circadian rhythmic pattern is the result of complex interactions of brain structures in multiple levels of organization, from cellular and molecular to widespread neural circuits [3]. By its turn, such a precise coordination of processes seems to be of uttermost importance to neurochemical cleansing [4], to the maturation and development of neural tissue in newborns and infants [5], and for memory consolidation in adults [6]. On the other side of the same coin, sleep may be disrupted in neurological scenarios [7] and the deterioration of sleep can induce neural dysfunction [8,9].

Each stage of the SWC displays distinctive overt behavior and neural activity signatures, usually assessed by electromyography (EMG) and electroencephalography (EEG), respectively. In a simplistic way, the SWC can be divided into three major stages. First, wakefulness (WK) is characterized by high-power EMG with desynchronized low-amplitude EEG. By its turn, slow-wave sleep (SWS) features low EMG with synchronized high-amplitude low-frequency EEG patterns. Finally, rapid eye movement sleep (REM) displays

virtually flat EMG and low-amplitude EEG tracings, intercalated with synchronization episodes in the theta frequency range. Importantly here, each of these different stages are inherently correlated with distinct neurobiological processes, such as the differential roles of SWS and REM in the consolidation of recently acquired memory traces [10–12].

For a proper assessment of the SWC and its relationship with brain function and dysfunction, it is thus of major importance to be able to characterize the rhythmic alternance of its stages [13,14]. This is performed by detecting each of its stages in an objective and reproducible manner. Even though there are clear guidelines for their recognition by expert visual inspection of electrophysiological recordings [15], the process is very laborious, tedious, and prone to errors due to subjectivity, fatigue, and bias. For this reason, different automatic signal processing methods have been put forward by scholars in the field. One pioneering work proposed and tested a method based on the automatic clustering of points in a state map built from spectral content ratios, applied to the multi-channel local field potential (LFP) recording of animals [16]. More recently, the application of machine learning methods has become a major trend for sleep stage classification [17]. By also using the spectral content of mice LFP and EMG power assessed by the mean-covariance restricted Boltzmann machine approach, Katsageorgiou and collaborators were able to detect SWC stages, including NREM substages, while observing important differences across different genetic backgrounds [18]. Huang and collaborators employed a support vector machine over a collection of signal features pre-screened from a set of 62 measurements extracted from time and frequency domains and of a non-linear nature [19]. Finally, convolutional neural networks have been applied to EMG and LFP signals, previously z-scored according to a novel proposal, in the work of Barger et al., 2019 [20]. The authors observed good agreement between automatic and expert classification while also incorporating the algorithm into an open-source toolbox named Accusleep.

Although representing valid contributions to automatic SWC staging, complex topologies of artificial neural networks (ANNs) yield no more than marginal improvements while requiring considerably higher computational resources and impairing neurobiological interpretability [21]. A more rational approach to this problem could be the usage of simpler machine learning models applied to more descriptive electrographic features. Among them, neural synchronization has been extensively described as a direct correlate of brain states and processes [22–24], including sleep [25] and its disorders [26]. For instance, an assessment of cross-frequency coupling (CFC), which is a measurement of the driving force between neural circuits or substrates oscillating at different frequencies [27–29], has proven to be very insightful in investigations of neuroplasticity [30] and visual binding [31] and in neurological disorders [32–34].

Based on this rationale, in this work, we assessed the possibility and the convenience of an SWC stage detection strategy by using a multilayer perceptron (MLP) trained and tested with a CFC measurement, the phase–amplitude coupling values (PAC). For this, the performance of different MLP architectures were objectively assessed by means of the receiver operating characteristic (ROC). Then, the top performing model was further assessed with multiple performance metrics in its application in the extrapolation of classifications, which simulates scenarios in an electrophysiology lab. We found that employing a basic MLP model supplied with detailed features, such as neural synchronization, is sufficient for accurately categorizing SWC stages in experimental conditions. We feel that our proposed approach opens new avenues of neurobiological investigation considering the importance of cross-frequency coupling for brain function. Particularly, it can enable the insightful interpretation of probable disturbances to be observed in neurological disorders.

2. Material and Methods

2.1. Signal Database, Surgery, and Electrophysiological Recordings

This study employed data obtained from in vivo experiments conducted in mice and made publicly available, as documented in the research by [20]. Briefly, 10 adult C57BL/6 mice (10–20 weeks old) were housed on a 12 h dark/12 h light cycle. For the EEG and EMG recordings, a reference screw was surgically implanted into the cranium, specifically positioned

on the top of the right cerebellum. The EEG signals were then captured using two screws positioned on the left and right cortex, located at coordinates anteroposterior –3.5 mm and medio-lateral +/−3 mm, respectively. For each mouse, 4 h long recordings were collected on 5 different consecutive days. The signals were acquired using a flexible recording cable that was connected to a pin header through a mini-connector. These signals were then amplified using an Intan RHD-2132 headstage amplifier (bandpass filter, 1–500 Hz; sampling rate, 1000 Hz; Intan Technologies, LLC—Los Angeles, CA, USA). All the recordings from this database were annotated with expert classification of the SWC stages, which was used here to evaluate our approach. All the experiments with animals were approved by the Animal Care and Use Committee at the University of California, Berkeley.

2.2. PAC Computation by Modulation Index and Comodulograms

The phase–amplitude coupling (PAC) was quantified using the modulation index (MI), which essentially measures the amount in which the amplitude of a higher-frequency oscillation is modulated by the phase of a slower oscillation, following a methodology detailed elsewhere [29]. Briefly, the LFPs were subjected to finite impulse response (FIR) bandpass filtering in both the forward and reverse directions using the "eegfilt" function in Matlab® 2023b to separate the LFPs into the low- and high-frequency bands of interest while also ensuring the elimination of phase distortion. Subsequently, the instantaneous phase of the low-frequency component was obtained, and the amplitude of the high-frequency oscillation was computed using the Hilbert transform. Afterward, a histogram of amplitudes of the high-frequency oscillation falling within the phase bins (N = 18, each spanning 20°) of the low-frequency oscillation was built. The resulting histogram/distribution P was then compared against a uniform distribution U, by means of applying the Kullback–Leibler distance (D_{KL}), defined in general form by:

$$D_{kL}(P,Q) = \sum_{j=1}^{N} P(j) log\left[\frac{P(j)}{Q\left(j\right)}\right]$$

where Q is any distribution. The distance was then normalized to obtain MI by:

$$MI = \frac{D_{k_L}(P,U)}{log(N)}$$

Multiple values of the MI using many combinations of low- and high-frequency pairs were then computed to create the comodulograms. Twenty frequency bands for the phase (1–20 Hz, 1 Hz increments, and 1 Hz bandwidth) and forty frequency bands for the amplitude (5–200 Hz, 5 Hz increments, and 10 Hz bandwidth), amounting to 800 values, were considered for the different forms of input data construction (Section 2.4). These discrete frequency vectors were deliberately adopted to ascertain an optimal balance between capturing potential PAC patterns and mitigating unnecessary computational burden. By focusing on this targeted frequency range, we were able to discern the salient features pertaining to the PAC while minimizing the impact of irrelevant oscillatory components that did not add significant informational value. A comodulogram was computed for each one of the 480 available consecutive 30 s segments (a total of 240 min), without temporal overlap, of the LFP spanning the entire 4 h long recording period.

2.3. Multilayer Perceptron Artificial Neural Network

A feedforward multiclass neural network known as the multilayer perceptron (MLP) was used in this work to classify the sleep stages, namely, WK, SWS, and REM, based on the information extracted from the set of MI values (Figure 1B). The MLP architecture was shallow and comprised three layers: input, hidden, and output. The input layer receives features extracted from the set of MI values (or PAC patterns), while the hidden layer performs computations and transformations to learn complex relationships within the data. The output layer yields the final predictions as binary classifiers of each stage based on the learned

representations from the preceding layers. Thus, the number of neurons in the output layer was always three, one for each of the stages of interest, while the number of input and hidden neurons varied in search of the best architecture. The number of input neurons depended on the choice of the input dataset (see Section 2.4.1). To determine the number of neurons in the hidden layer, empirically established guidelines were considered. The widely employed rule of thumb for this case suggests that the optimal size of the hidden layer typically lies between the dimensions of the input and output layers (Figure 1B, but see Section 2.4 for more details).

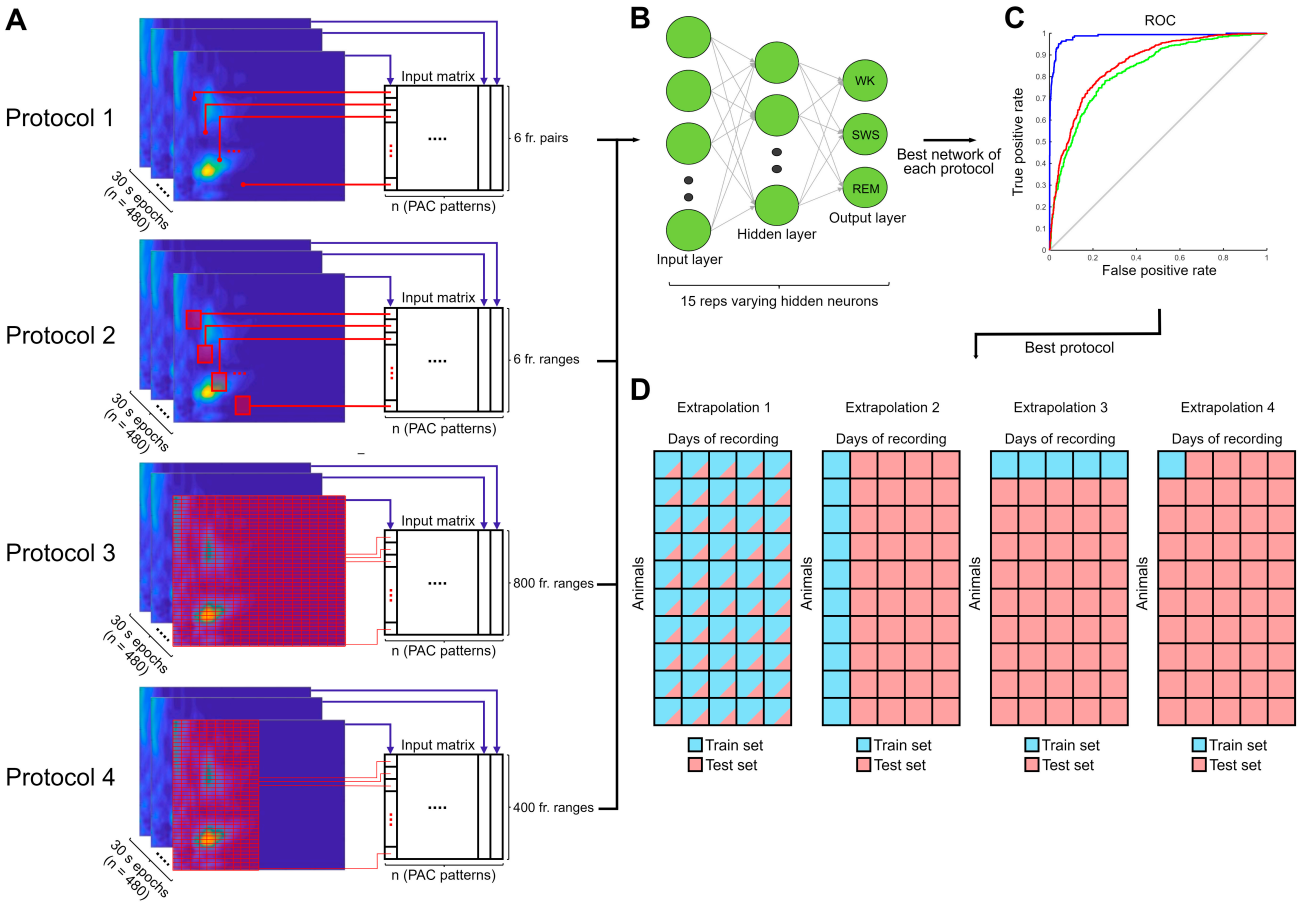

Figure 1. The experimental design: an assessment of the network architectures was performed in phase 1 of the experiments, which is depicted in panels A to C, and simulation of the application of the network with novel data with decreasing ratios of the train to test dataset extrapolations were performed in phase 2, depicted in panel D. (**A**) The input data: four different protocols were used for the construction of the input data, frequency points, frequency ranges, full comodulogram, and first (1–10 Hz) half of the comodulogram (left panels); the values were computed for each 30 s time window for all the animals and all the days and, thus, the input data were N (*MI* values) × 480 (time windows) × 10 (animals) × 5 days. (**B**) The multilayer perceptron (MLP) architecture consisted of an input layer (number of neurons equals to number of *MI* values), a hidden layer (number of neurons varied), and an output layer with three neurons, one for each sleep stage. (**C**) The ROC curve for all possible combinations of inputs and networks were computed and the AUC values obtained to determine the best performing choice. (**D**) Using the best performing network architecture, four different extrapolation strategies were tested in order to simulate usage in a lab scenario, from left to right: train with 70% and test 30% of the dataset; train with the first day of all the animals to test on all the other days; train with all the days of the first animal and test with all the other animals; and train with one day of one animal to test on the rest of the dataset.

For training the network, information was processed using the logistic sigmoid activation function in the hidden-layer neurons and the softmax activation function was employed in

the neurons of the outer layer to generate the final output. As a shallow network, issues like gradient vanishing are unlikely to occur and thus more sophisticated (and complex) algorithms such as ReLU (Rectified Linear Unit) are not needed. The error obtained at each parsing of the data was then used to update the network's weights using the scaled conjugate gradient (SCG) algorithm. This algorithm, a second-order optimization method, was chosen due to its efficiency in handling large-scale machine learning problems [35–37]. Here, the SCG optimized the Binary Cross-Entropy loss function, which is highly applicable to multi-label classification problems such as this. It is also the default when using Matlab® 2023b libraries. A total of 1000 epochs were employed in the training process to ensure adequate learning while preventing overfitting. To keep the method very straightforward, we also did not implement hyperparameter regularization procedures. The MLP was trained in batches, meaning that the neural network is presented with the entire training dataset at once and its weights are updated after each epoch. Batch training can be a beneficial approach, when the pertinent information necessary for making accurate predictions is consistently distributed across the entire dataset and there are sufficient computational resources available to process it simultaneously. Finally, in our dataset, the class distribution was roughly 6:3:1 (SWS:WK:REM). While this indicates some degree of imbalance, it is not severe enough to warrant specific algorithmic adjustments, such as resampling or cost-sensitive learning. In any case, we were able to assess the impact of this imbalance by computing multiple different metrics of network performance (see Section 2.4).

2.4. Experimental Design

The experiments performed with the MLP can be divided into two phases: (1) testing different combinations of network architecture, i.e., the number of neurons in the hidden layer and protocols for construction of the input dataset, in order to define the best performing method (Figure 1A–C) and (2) using the best method to test the classification of the SWC stages with new data in a simulation of the usage in real lab settings (Figure 1D).

2.4.1. Phase 1: Search for the Best MLP Architecture

The four different protocols used for the construction of the input data (Figure 1A left) were as follows:

- Protocol 1 or frequency points: The input layer comprised six neurons, each corresponding to *MI* values of specific frequency pairs that are of general neurobiological relevance, sleep function included (Table 1). These pairs of frequency values were defined on the basis of past literature, including a previous study of the group.
- Protocol 2 or frequency ranges: Analogous to protocol 1, but instead of using single *MI* values for each frequency pair, narrow bands centered around the values of Table 1 were used. For this protocol, all the *MI* values inside each band were averaged, thus also resulting in 6 data points.
- Protocol 3 or full comodulogram: The whole comodulogram was incorporated as the data input, encompassing not only the previously mentioned frequency pairs but the entire spectrum of calculated cross-frequency couplings.
- Protocol 4 or half comodulogram: The same as protocol 3 but, instead of the whole comodulogram, only its first (left) half (0–10 Hz in the horizontal axis) was used.

The set of *MI* values (PAC patterns) representing each 30 s signal window was concatenated into a single column so that the input consisted of a matrix in which each line was the *MI* values and each column represented a PAC pattern of a 30 s time window (Figure 1A right). This resulted in 6 × 480 matrices for protocols 1 and 2, 800 × 480 matrices for protocol 3, and 400 × 480 matrices for protocol 4. The PAC patterns of the entire recordings (all mice and all days) were used. In this phase, the data were always standardized (z-scored) and, subsequently, they were randomly divided into training (70%) and testing (30%) sets.

For each one of the four protocols, each possible network architecture, i.e., with the number of neurons in the hidden layer varying from 3 (number of output layer neurons) to the maximum number considered (equals the number of neurons in the input layer, 6,

400, or 800) was tested in multiple proofs (10 simulations). In protocols 3 and 4, to improve the computation time, we used different step sizes for different ranges of the number of neurons in the hidden layer (see Figure 2). In this first experimental phase, the performance of each MLP architecture and input protocol was assessed using the area under the ROC curve (AUC) calculated for each output (Figure 1C) of all possible architectures. Here, the reported results refer only to the testing part of the dataset. The MLP with the highest average (across the distinct sleep stages) AUC value among all the proofs and stages was chosen as the best network for the second phase of the experiments.

Table 1. Pairs of frequencies assessed by PAC.

Identifier	Frequency Used for Phase (Low)	Frequency Used for Amplitude (High)
A	Theta (7.5 Hz)	High gamma (140 Hz)
B	Theta (7.5 Hz)	Low gamma (70 Hz)
C	Slow wave (1 Hz)	Theta (11 Hz)
D	Theta (11 Hz)	Low gamma (70 Hz)
E	Delta (2.5 Hz)	Theta (11 Hz)
F	Slow wave (1 Hz)	Higher gamma (200 Hz)

2.4.2. Phase 2: Testing the Best Architecture in a Simulation of a Lab Scenario

After the best combination of the number of hidden layer neurons and the input construction protocol was detected (best architecture), the second phase of experiments was performed. Here, the optimal neural network was evaluated in its capability of correctly classifying novel data in scenarios that mimic a real lab setting (Figure 1D). Four different strategies were employed, simulating different dimensions and increasing levels of extrapolation, in which the ratio of the train dataset to the test dataset was progressively lower. Of particular importance here, the network was tested for extrapolation across the dimensions of multiple days of recording, multiple animals, or both. The extrapolation strategies were thus (Figure 1D from left to right):

- Strategy 1: train on 70% of all the data; test on the remaining 30%.
- Strategy 2: train on all the animals of day 1; test on all the remaining days.
- Strategy 3: train on all the days of a single animal; test on all the other animals.
- Strategy 4: train on a single animal in a single day; test on all the other animals on all the other days.

The first strategy is the same as the one used in the experiments of phase 1 but detailed for the best architecture, with the values per animal and per day, and with additional metrics (see below). Again, the reported results refer to the testing data only. Conversely, the performance values for strategies 2, 3, and 4 refer to the full dataset. Not only does this better simulate an application of the model in real situations from the lab but also the train part of the dataset in these cases is small (20%, 10%, and 2%, respectively). Finally, in the last strategy, we varied the single animal used for training to assess the possible variability across this dimension.

To fully assess model performance and its generalization capacity, in phase 2, we measured not only the AUC but also several metrics from the confusion matrices, including the precision, sensitivity, specificity, accuracy, and F1-score. The ground truth used here was always the sleep classification information annotated by experts in the original database study [20]. We used Matlab® 2023b machine learning toolboxes for creating, training, and testing the model. The code is freely available on the internet at https://github.com/vrcota/PAC-MLP (accessed on 8 May 2024).

2.4.3. Statistical Analysis

The AUC results of both phase 1 and phase 2 were assessed for their normality by using the Kolmorogov–Smirnov test and for its statistical significance by one-way analysis of variance followed by the Tukey post hoc test for pairwise comparison. In all cases, $p < 0.05$ was adopted for the significance level.

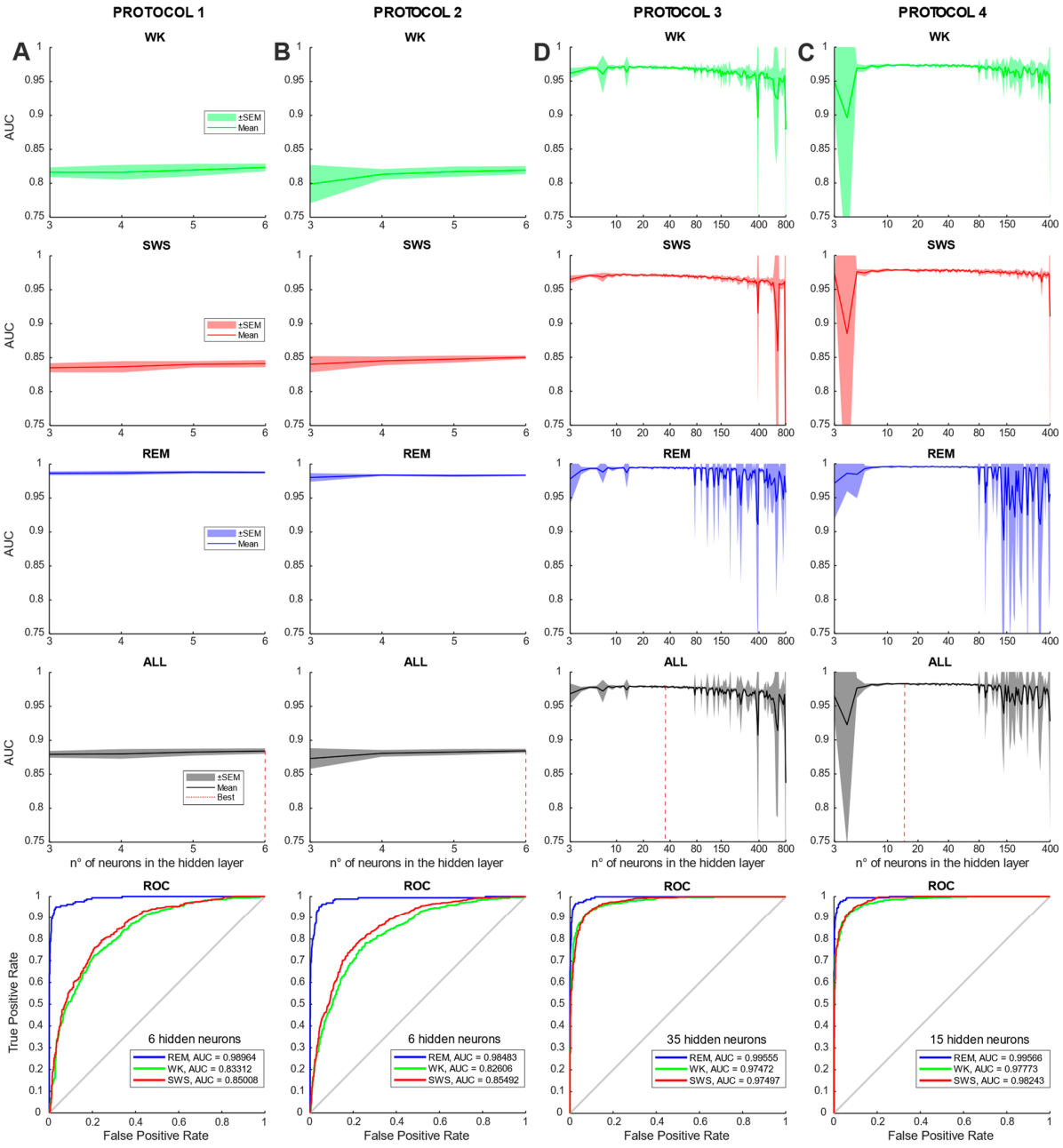

Figure 2. An assessment of the different network architectures. Each column represents one of the four different protocols for the construction of the input dataset: (**A**) Protocol 1, frequency points; (**B**) protocol 2, frequency ranges; (**D**) protocol 3, full comodulogram, and (**C**) protocol 4, first half of the comodulogram. The curves of the top four rows depict the variation in the AUC values with the number of neurons in the hidden layer, with the average across the proofs in dark color and the dispersion (s.e.m.) across the many proofs being depicted in the background shadow of the same color in a lighter hue. From top to bottom: WK in green, SWS in red, REM in blue, and ALL stages in black (this color convention is used for all the figures). For protocol 3, the ticks of the horizontal axis are not uniformly spaced given the different step sizes: step equal to 1 from 3 up to 49, step equal to 2 from 50 up to 98, step equal to 5 from 100 up to 195, step equal to 10 from 200 up to 490, and step equal to 25 from 500 up to 800. For protocol 4, we followed the same procedure as for protocol 3 but set the axis range up to 400. The bottom row depicts the ROC curves for the binary classification of each stage for the optimal number of neurons in the hidden layer: 6, 6, 35, and 15 (indicated as a vertical red dotted line in the fourth-row panels), for protocols 1 to 4, respectively.

3. Results

3.1. Top Performing MLP Architectures Are Obtained by Using Full or Half of the Comodulogram

The performance obtained by all the architectures in the first phase of the experiments are depicted in Figure 2. The AUC values are highly constant (with the variability almost always lower than 2%) in protocols 1 and 2 but vary considerably in the extremities in protocols 3 and 4, especially for the REM stage. Moreover, while the first two show an apparently uniform increase in the AUC values with the increase in the number of neurons, protocols 3 and 4 display a bell-shaped curve in this relationship. The maximum AUC values were thus obtained using 6, 6, 35, and 15 neurons in the hidden layer for protocols 1, 2, 3, and 4, respectively. For protocols 3 and 4, the maximum AUC values were extracted within the stable region of the AUC curves. After detecting the best network of the best architecture for each protocol (with the above-mentioned number of neurons in the hidden layer), the ROC curves and final AUC values were computed (Figure 2, bottom row). All of them are clearly above chance (AUC values > 0.85 always).

The AUC values encompassing all the stages were 0.884, 0.885, 0.979, and 0.982 for protocols 1 to 4, respectively. The architectures using larger portions of the comodulogram (full or half), and not only pairs of frequency of known neurobiological relevance, performed significantly better for all the stages (Figure 3). Interestingly, while the AUC values for REM were the same (no statistical difference)—and always high—across the different protocols, those for SWS and WK were considerably higher when using more of the comodulogram. Furthermore, for both the SWS and ALL cases, the performances obtained by using the half comodulogram were even significantly superior to the full comodulogram, even with a small difference (SWS: 0.971 versus 0.978; ALL: 0.979 versus 0.982). Finally, the AUC values for the REM stage were always considerably higher when compared to those of both SWS and WK, as it can be seen by the blue curves always occupying the top left-most portion of the ROC graphs (Figure 2).

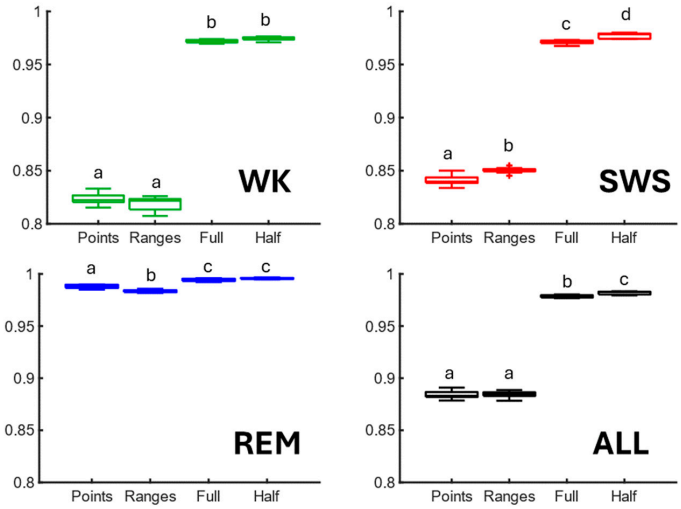

Figure 3. A comparison of classification performance between the four different protocols of the input data construction for WK (green; **top left**), SWS (red; **top right**), REM (blue; **bottom left**), and ALL (black; **bottom right**). Protocols using more of the comodulogram performed significantly better than those using only well-established frequency pairs described in the literature. Curiously, using only half the comodulogram performed even better than using its entirety, in the cases of SWS and ALL, even though only slightly. Using the one-way ANOVA and post hoc Tukey, the resulting statistics were as follows (mean ± s.e.m.): WK) points: 0.823 ± 0.001; ranges: 0.819 ± 0.001; full: 0.972 ± 0.001; and half: 0.974 ± 0.001 (difference a–b: $p < 0.0001$). SWS) points: 0.841 ± 0.001; ranges: 0.851 ± 0.001; full: 0.971 ± 0.001; and half: 0.978 ± 0.001 (differences a–b, a–c, a–d, b–c, and b–d: $p < 0.0001$; difference c–d: $p < 0.001$). REM) points: 0.988 ± 0.001; ranges: 0.984 ± 0.001; full: 0.994 ± 0.001; and half: 0.996 ± 0.001 (differences a–b, a–c, and b–c: $p < 0.0001$). ALL) points: 0.884 ± 0.001; ranges: 0.885 ± 0.001; full: 0.979 ± 0.001; and half: 0.982 ± 0.001 (differences a–b and a–c: $p < 0.0001$; b: $p < 0.0001$; and difference b–c: $p < 0.05$). The boxplots display the median and 25th and 75th percentiles in the main box; the whiskers depict the extremities of the data.

3.2. Top Performing Network of Protocol 4 Extrapolates Well across Animals and/or Days of Experiment

For the second phase of experiments, we chose the architecture using protocol 4 (first (left) half of the comodulogram; 1–10 Hz) for the input dataset with 15 neurons in the hidden layer, as it was the best performing one. The results obtained for the four extrapolation strategies are depicted in Figures 4–8, in the same sequence as shown in Figure 1D. In every case, the performance of the chosen architecture across all the testing scenarios was always above 0.9. In the first extrapolation strategy, with a random division of 70%/30% train and test data, the AUC values were always above 0.95, with REM showing the highest values, and without noticeable variation across the different days (Figure 4). The global accuracy in this first assessment was 0.924 (Table 2).

Table 2. Global accuracy for all extrapolation strategies.

	Accuracy
Strategy 1	0.924
Strategy 2	0.920
Strategy 3	0.898
Strategy 4—best	0.879
Strategy 4—worst	0.864

The same goes for extrapolating the training of one day (day 1), all animals, to classify the other days: the AUC remained high (>0.95) with a slight decrease as the days passed (Figure 5). The global accuracy here was 0.920 (Table 2).

The AUC values are also high when there is an extrapolation from one animal (animal 1), all days, to the others (Figure 6). Although the variability in the results is higher, the average AUC values were always higher than 0.94. The global accuracy was 0.898 (Table 2).

Even in the most demanding scenario, in which the network trained on a single day of a single animal is used to classify the rest of the dataset, the AUC values remained over 0.9 and without noticeable variation across the days (Figure 7), particularly for the case of REM. This applies to the extrapolations using both the best and worst animal in the dataset. The accuracies in strategy 4 were 0.879 and 0.864 for the best and worst cases, respectively (Table 2).

When all the strategies are compared against each other, it becomes clear that the performance decreases from the least (strategy 1) to the most demanding scenario (strategy 4) in a statistically significant way (Figure 8). Moreover, in strategy 4, the extrapolation using the worst animal (10; lowest prediction average in panel A) performs statistically worse than using the best one (7; highest prediction average in panel A). Here, again, the REM stage seems to be the least affected by the strategy choice.

Finally, Tables 3–5 below show, respectively, the results for the precision, recall, and F1-score for all the extrapolation strategies. Overall, all these metrics were also high, typically over 0.9 for SWS or 0.8 for WK, while REM shows both high and a couple of moderate values in the most demanding scenarios. The global average of all the values is 0.88.

The confusion matrices and the complete set of its derived metrics for all the strategies can be found in Figures S1–S5 and Tables S1–S5 of the Supplementary Materials. Graph versions of Tables 3–5 can be found in Figures S6–S8 of the Supplementary Materials.

Table 3. Precision values for all extrapolation strategies.

	WK	**SWS**	**REM**
Strategy 1	0.876	0.936	0.965
Strategy 2	0.912	0.928	0.897
Strategy 3	0.910	0.914	0.793
Strategy 4—best	0.827	0.901	0.923
Strategy 4—worst	0.871	0.920	0.649

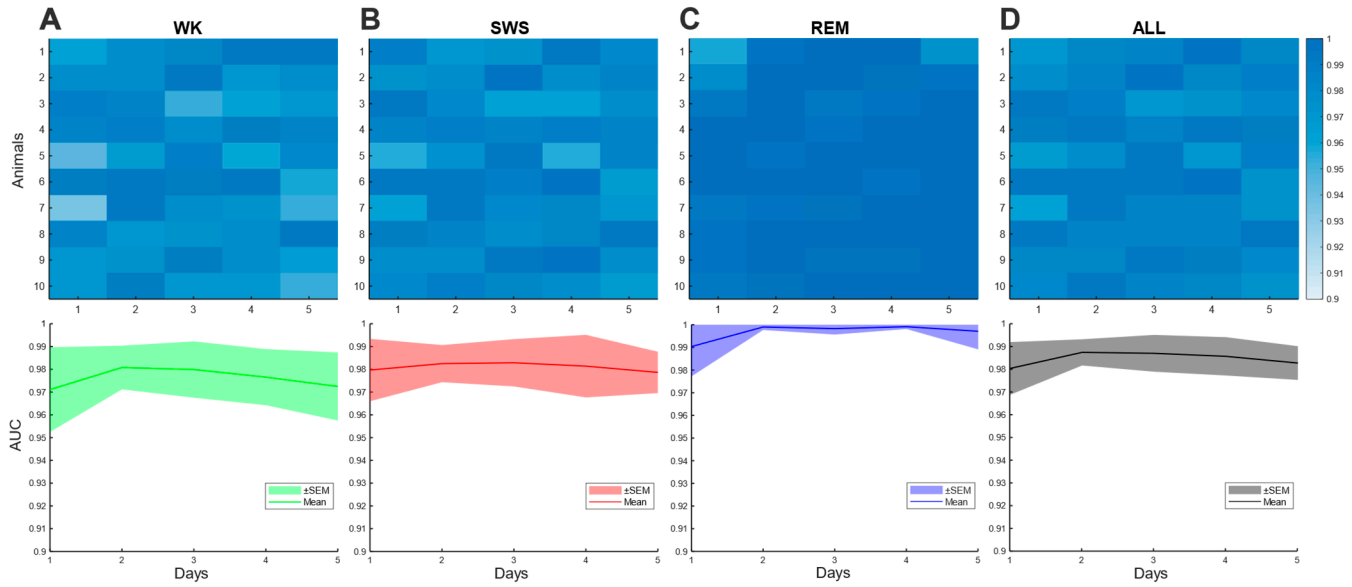

Figure 4. Extrapolation strategy 1. Training data on 70% of all the data to test on the remaining 30% resulted in high AUC values (average always >0.95) of all the animals, all days, with small variability (<5%) for all the stages: (**A**) WK; (**B**) SWS; (**C**) REM, and (**D**) ALL. The top row depicts each AUC value for each animal and each day as color-coded panels (see the lateral bar for scale). The bottom row shows the average of all the animals for each of the five days of the dataset with the variability (s.e.m.) being depicted in the background shadow of the same color in a lighter hue. The classification of the REM stage was shown to be particularly optimal considering both the high AUC values and low variability.

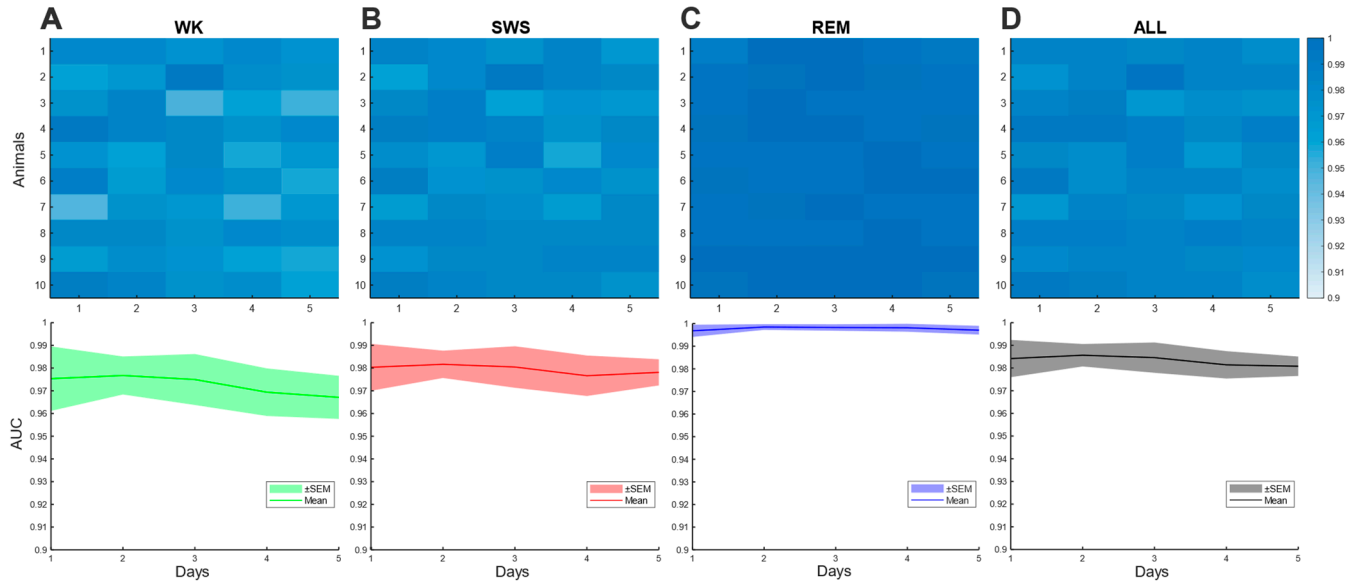

Figure 5. Extrapolation strategy 2. Training with the first day of all the animals also resulted in high AUC values (>0.95) for the classification of stages in the other days, with small variability: (**A**) WK; (**B**) SWS; (**C**) REM, and (**D**) ALL. There is a slight decrease in the values as the days pass. Analogously to Figure 4, the top row depicts each AUC value for each animal and each day, and the bottom row shows the average for all the animals across the five days of the dataset. Again, the classification of the REM stage was shown to be particularly optimal considering both the high AUC values and low variability.

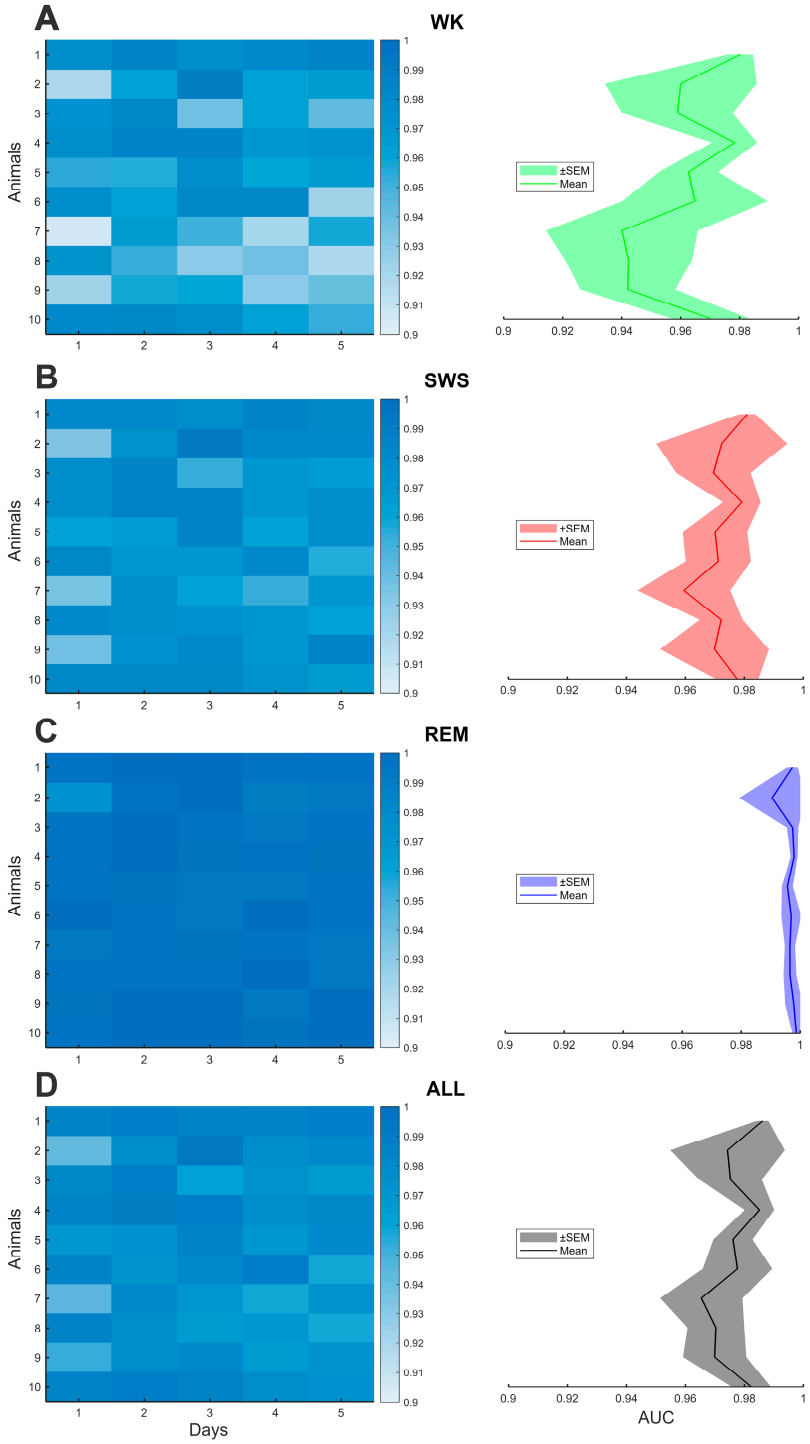

Figure 6. Extrapolation strategy 3. Training with all the days of the first animal resulted in high AUC values for the classification of the stages in the other animals, still with considerably small variability: (**A**) WK; (**B**) SWS; (**C**) REM, and (**D**) ALL. Considering that here our interest is to see the extrapolation across the dimension of the different animals, the AUC values were averaged for each animal across all days. Thus, in this figure, each row represents a sleep stage. The left column panels depict each AUC value for each animal and each day as a color-coded scheme (see the lateral bar for scale). The right column shows the average across all the days for each one of the ten animals in the dataset with the variability (s.e.m.) being depicted in the background shadow of the same color in a lighter hue. Once more, the classification of the REM stage was shown to be particularly optimal in terms of the AUC value.

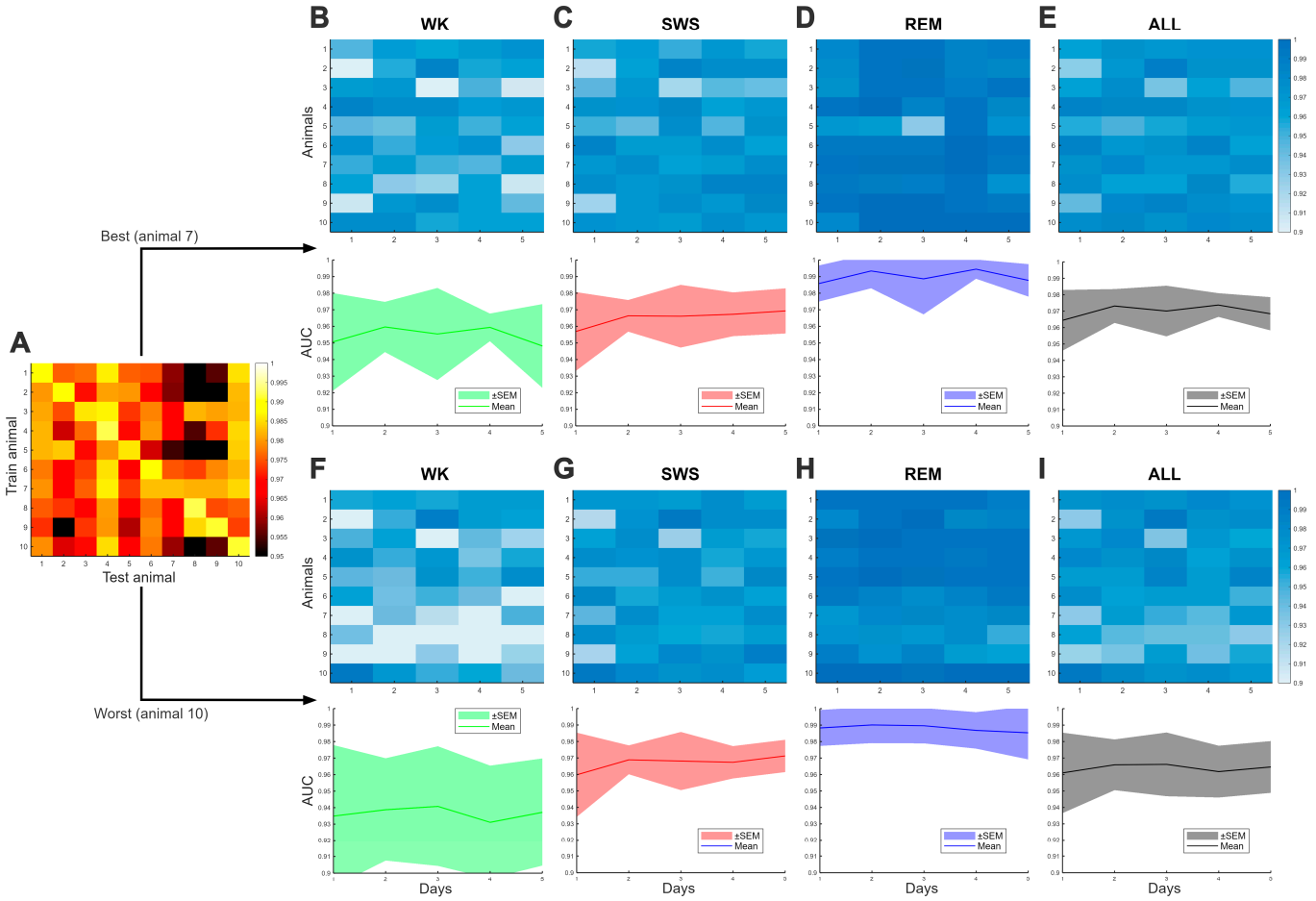

Figure 7. Extrapolation strategy 4. Training with a single animal and single day and testing on the rest of the dataset, the most demanding scenario, also resulted in high average AUC values across the days. (**A**) The color-coded panel on the left depicts the overall classification performances using all possible pairs of animals used for training and for testing. Each row is one animal used for training and each column is an animal used for testing. From this panel, the best performing animal (animal 7: the one that better predicted the others with a higher AUC on average; highest row-wise mean) and the worst performing animal (animal 10: the one that predicted the others with a lower AUC on average; lowest row-wise mean) were extracted. (**B**–**E**) The classification results using the best animal for each stage, including the individualized AUC (top row) and AUC averages for all the animals for each of the five days (bottom row). (**F**–**I**) Analogously to panels **B**–**E**, the classifications using the worst animal. Equal to the previous results, the REM stage displayed the highest performances and smaller variability.

Table 4. Sensitivity (recall) values for all extrapolation strategies.

	WK	SWS	REM
Strategy 1	0.878	0.950	0.890
Strategy 2	0.856	0.951	0.935
Strategy 3	0.789	0.951	0.924
Strategy 4—best	0.867	0.925	0.668
Strategy 4—worst	0.713	0.929	0.952

Table 5. F1-score values for all extrapolation strategies.

	WK	SWS	REM
Strategy 1	0.877	0.943	0.926
Strategy 2	0.883	0.940	0.916
Strategy 3	0.845	0.932	0.853
Strategy 4—best	0.846	0.913	0.775
Strategy 4—worst	0.784	0.924	0.772

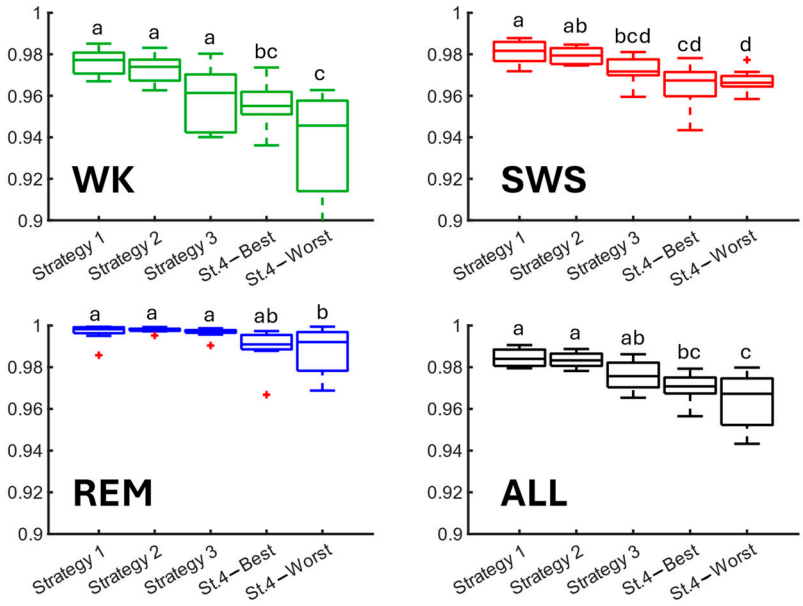

Figure 8. The classification performance decreases significantly as the extrapolation strategy becomes more demanding, in which the training data represent less of the total dataset. Using the one-way ANOVA and post hoc Tukey, the resulting statistics were as follows (mean ± s.e.m.): WK) strategy 1: 0.976 ± 0.005; strategy 2: 0.973 ± 0.005; strategy 3: 0.960 ± 0.005; strategy 4: 0.955 ± 0.005; and strategy 5: 0.937 ± 0.005 (differences a–bc and a–c: $p < 0.05$). SWS) strategy 1: 0.981 ± 0.002; strategy 2: 0.979 ± 0.002; strategy 3: 0.972 ± 0.002; strategy 4: 0.965 ± 0.002; and strategy 5: 0.967 ± 0.002 (differences a–bcd: $p < 0.05$; differences a–cd, a–d, and ab–cd: $p < 0.0001$; and differences ab–d: $p < 0.001$). REM) strategy 1: 0.997 ± 0.002; strategy 2: 0.998 ± 0.002; strategy 3: 0.997 ± 0.002; strategy 4: 0.990 ± 0.002; and strategy 5: 0.988 ± 0.002 (difference a–b: 0.05). ALL) strategy 1: 0.985 ± 0.003; strategy 2: 0.983 ± 0.003; strategy 3: 0.976 ± 0.003; strategy 4: 0.970 ± 0.003; and strategy 5: 0.964 ± 0.003 (differences: a–bc and ab–c: 0.01; and difference a–c: 0.0001). The boxplots display the median and 25th and 75th percentiles in the main box; the whiskers depict the extremities of the data; and the outliers are displayed as red + symbols.

4. Discussion

The objective and automatic detection of sleep stages is of paramount importance in its neuroscientific investigation and several different algorithms exist for this purpose. Instead of using computationally heavy and complex machine learning models, here we implemented and tested a simple and straightforward MLP network topology, which was trained and tested using conventional algorithms and with a very meaningful electrographic biomarker, the phase–amplitude coupling (PAC) between oscillations of different frequencies. The results of phase 1, designed to test different protocols of input data usage and to find the optimal number of hidden neurons, showed that such an approach can yield good to excellent overall performances, as assessed by the AUC metric (Figure 2). Protocols that use more data from the comodulogram (protocols 3 and 4, full or half comodulograms, respectively) displayed significantly improved performances, with AUC values around 0.98. Moreover, the results from multiple metrics of phase 2, which was designed to assess

the applicability of the network in a simulation of usage in the electrophysiology lab, indicate that the network architecture with the highest AUC values for the present data (half comodulogram input and 15 neurons in the hidden layer) can also perform well in extrapolations across different days (Figure 5), different animals (Figure 6), or even in the most demanding scenario in which data from a single animal in a single day can be used to classify the rest of the dataset (Figure 7). Yet, some caution is advised in these situations (see more on this further down the text).

Regarding the different choices of input (protocols 1 to 4), the usage of only the PAC values described as highly important for sleep phenomena in the literature (Table 1; protocols 1 and 2) were enough to classify all the stages reasonably well, at least as assessed by the AUC values only (Figure 2). On the other hand, it was somehow expected that the network architectures would perform significantly better when fed with more input values, i.e., whole or half comodulograms instead of frequency points or ranges, which happened (Figure 3). This suggests that, although the PAC between such frequency pairs does have a profound relationship with SWC oscillatory phenomena, they do not describe it completely, being more important to the REM stage, specifically. Conversely, the values of the PAC synchronization across other frequency pairs seem to be needed for optimal classification. Curiously, the usage of just the first half of the comodulogram yielded significantly better classification than the usage of the whole comodulogram, although only slightly. In our understanding, this is a compelling result in support of smart choices of input datasets and that brute force options are not always the solution. Not only do bigger datasets imply computationally heavier models, but they do not necessarily yield improved performance. Indirectly, this also supports the concept of "garbage in, garbage out" in signal processing and machine learning, which can be translated here as no matter how powerful one's network is, no good classification will be attained if the data are poor in their relevance to the brain process under investigation.

In the same vein, the present findings showed that the number of neurons in the hidden layer must be chosen with care in these applications. While simpler input protocols performed better with the maximum number, this was not the case for the more complex choices. Not only was the average performance lower with the maximum values but also performance was unstable as shown by the high variability across distinct runs in this extremity. The same happened in the lower extremity. In fact, for these protocols, there seems to be a narrower middle range in which the model is stable. Importantly, such range also yields the highest AUC values. For protocol 3, the range is between 15 and 60 neurons, and for protocol 4, it is between 6 and 70 neurons. Curiously, REM seems to be the most susceptible stage, showing greater variations in the extremities and the narrowest range of stable AUC values (in comparison with SWS for instance). The reasons for this behavior are unclear but may be related to the fact that REM is the class with the smaller number of samples in the dataset.

Overall, the results from phase 2 showed that, by using protocol 4 with 15 neurons in the hidden layer, experimenters can train the network on a small subset of the data to confidently classify the rest of it. In fact, all the extrapolation strategies tested here with this architecture, which are relatable to applications in the lab setting, typically yielded good performances that did not vary substantially in any of the dimensions tested, either across the days of experiments (Figures 4, 5 and 7) nor the different animals (Figure 6). As expected, the values of all the metrics dropped as less of the original dataset was used for training the network (Figure 8 and Tables 3–5). Yet, even in the worst-case scenario, in which the network is trained on data from a single day of the worst predicting animal, the average performance is still an over 0.95 AUC value. On the other hand, the F1-scores in both cases (best and worst animals) of strategy 4 were not as high, around 0.77 (Table 5). Interestingly, the results suggest that the generalizability of the training from a single animal and single day can be impaired in different ways, in which the precision or recall can be lower (see Tables 3 and 4). Of course, the acceptable level and type of error must be previously defined by potential users, so one can properly choose the strategy to be adopted. Naturally, if

very high overall performance is required, initial labor-intensive work of annotating larger portions of the data by expert inspection will be needed, impairing automation aspects. Even more, it is recommendable that the approach is validated once again on an initial annotated dataset for each new application, including searching for the optimal architecture (steps of phase 1). This will certainly maximize the performance metrics. Conversely, if users are more interested in detecting only majority classes (e.g., SWS), or if specificity is much more important than any other criteria, future users can safely adopt strategies requiring training on a small subset of the data.

It is important to highlight that the electrographic data used in this study are very homogeneous, once they were collected using the very same recording parameters, with the same electrophysiology setup and experimental protocol. Although not tested here, it is very plausible that changes in these factors between the training and testing datasets may strongly impact performance, yielding misclassifications. Particularly, it is expected that the type and anatomical position of electrodes will have a major influence in the PAC values captured, as they will be able to or not properly capture important oscillatory content related to the SWC. For instance, detection of the REM stage in rats by the usage of spectral state maps is heavily influenced by the presence or not of electrodes in the hippocampus [16] due to the important hippocampal theta oscillation. In the same vein, another aspect of major importance may be animal species and, hence, brain functional anatomy. Thus, in our understanding, a network trained in one experiment should never be used to classify data from a different one. Furthermore, and given the small differences between using a half or full comodulogram, new applications of this approach may consider using the full spectrum to avoid the risk of not capturing important cross-frequency couplings that did not matter here. Finally, the performance of the network may also depend on the quality and consistency of the recordings across animals and days. For instance, if the signal quality decays or varies too much as the days go by, as usually happens in very long experiments, it is expected that the classification will fail. On the other hand, it is notable that the high values of the performance metrics in this study were obtained from single-channel electrocorticogram recordings only (we did not use the EMG signals available in the database). This represents a simplified experimental protocol for the surgical implant of electrodes and recordings. Thus, if problems are expected to happen, experimenters can overcome them by adopting more robust recording strategies, which include EMG and/or multiple channels of electroencephalographic recordings. This is evidence that the approach may work well for practical purposes in neuroscientific investigation.

In our understanding, the good overall performance of the model corroborates our central argument that a straightforward network can efficiently classify SWC stages if the input data contain relevant neurobiological information. The fact that a shallow network, as the one implemented here, is enough has intrinsic technical and scientific value by itself. Yet, it is important to recognize that the model proposed here could be improved even without the need for deep learning architectures nor incurring considerable additional computational costs. Particularly, it seems that the network performance was impacted to some extent by not properly dealing with imbalanced data. In fact, the F1-scores were lower for the minority class REM stage, especially in the most demanding extrapolation strategy (Table 5; F1 around 0.77). Curiously, while training with the best animal (with the highest AUC values) showed good precision (0.923) and moderate sensitivity (0.668), the worst animal was the other way around, with moderate precision (0.649) but good recall (0.952). From the point of view of the F1-score, they were equivalent (0.775 and 0.772). Thus, it is plausible that the adoption of additional steps aimed at mitigating data imbalances (e.g., resampling or cost-sensitive learning) might yield improved performance. From the point of view of sleep neurobiology though, REM is usually considered the most "elusive" of the stages and different SWC staging methods can fail in properly detecting it. For instance, the REM cluster in spectral state spaces will certainly overlap the much more frequent WK cluster if the electrographic recording fails to capture enough hippocampal theta, which can happen when only intracortical microwire electrodes are used [16]. The resemblance

between WK and REM electrographical signatures is one of the reasons that the latter is many times called paradoxical sleep. Furthermore, we did not employ regularization techniques (e.g., dropout), as the performance metrics (F1-score, precision, and recall) indicated that overfitting was not a significant issue in our shallow network configuration. However, we acknowledge the potential benefits of these techniques and suggest their exploration in future work. Regarding signal processing, one limitation of the approach presented here is that the calculation of the *MI* and of comodulograms in general may be computationally intensive. This is largely due to the chosen algorithm, which was selected for its significance in the neuroscientific literature of memory and plasticity. However, alternative methods for computing PAC could be considered, such as those that do not require the creation of phase–amplitude distributions and calculations of distances to uniform patterns [38].

Overall, all this rationale, properly supported by the current findings, makes the case that the usage of neurobiologically relevant features can represent game changers when applying machine learning tools to analyze neural data in lab settings. In general, the network is expected to perform well when tested with brand new data, given that some consistency is granted. In the same vein, it is exactly the usage of highly interpretable features that can make approaches such as this one useful in the context of perturbations of the SWC neurodynamic. What would it mean if, in the case of a certain type of neural dysfunction animal model, different methods of SWC staging stopped agreeing? For instance, what would it mean if a Boltzmann machine strategy identifies WK, but a PAC-based method identifies SWS? That could be interpreted as an electrophysiologically dysfunctional brain state in which the animal is awake but with a brain synchronization typical of slow-wave sleep. This kind of valuable finding has been reported in the past (Dzirasa, 2006) [14] but largely depended on the visual inspection of an expert.

Finally, there has been some fruitful debate about the possibility that artificial intelligence and machine learning techniques would substitute signal processing [39]. There are no doubts of the wonders achieved by the new technologies, such as convolutional neural networks, deep learning, and the very hyped Large Language Models, in many areas of applied sciences and engineering. Yet, in the opinion of the authors, supported also by the present results, in-depth knowledge of the mechanisms underlying the processes of interest—the sleep–wake cycle in our case—helps with devising simple models with not only great performance but also yielding improved interpretability with reasonable computational cost. Again, the fact that the architecture that uses only the first half of the comodulogram is even better than the one using all of it is a strong argument in this sense. In basic sciences, in which understanding the phenomenon is the most important aspect, such due balance between well-informed machine learning approaches and signal processing is indispensable.

5. Conclusions

In this study, we showed that a shallow and straightforward neural network model can efficiently classify the stages of the sleep–wake cycle of animals from electrocorticograms. Instead of adopting computationally heavy deep models, the network was fed with an electrographic biomarker of improved neurobiological relevance, with phase–amplitude coupling between oscillations of different frequencies. Obtaining high-performance values comparable to other different strategies from the literature has intrinsic methodological and scientific value. Furthermore, it corroborates the central argument here that the usage of biologically relevant features in biomedical applications of machine learning tools is at least as important as network topologies and learning algorithms. By doing this, we believe that this manuscript contributes to the important debate of signal processing and artificial intelligence for neuroscience.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app14135816/s1, Figure S1: Confusion matrix for extrapolation strategy 1; Figure S2: Confusion matrix for extrapolation strategy 2; Figure S3: Confusion matrix for extrapolation strategy 3; Figure S4: Confusion matrix for extrapolation strategy 4, best animal; Figure S5: Confusion matrix for extrapolation strategy 4, worst animal; Figure S6: Precision values for all strategies according to stage; Figure S7: Recall values for all strategies according to stage; Figure S8: F1-score values for all strategies according to stage; Table S1: Complete list of performance metrics for extrapolation strategy 1; Table S2: Complete list of performance metrics for extrapolation strategy 2; Table S3: Complete list of performance metrics for extrapolation strategy 3; Table S4: Complete list of performance metrics for extrapolation strategy 4, best animal; Table S5: Complete list of performance metrics for extrapolation strategy 4, worst animal.

Author Contributions: Conceptualization, V.R.C., G.F. and M.C.; methodology, V.R.C., G.F. and S.D.C.; software, G.F. and S.D.C.; formal analysis, V.R.C. and S.D.C.; investigation, V.R.C., G.F. and S.D.C.; figures: V.R.C. and S.D.C.; resources, V.R.C. and M.C.; writing—original draft preparation, V.R.C. and G.F.; writing—review and editing, V.R.C., G.A. and M.C.; supervision, V.R.C. and M.C.; project administration, V.R.C. and M.C.; funding acquisition, V.R.C. and M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006)—A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022). V.R.C. was supported by the Marie Skłodowska-Curie Individual Fellowship MoRPHEUS, Grant Agreement no. 101032054, funded by the European Union under the framework programme H2020-EU.1.3.-EXCELLENT SCIENCE.

Data Availability Statement: The data presented in this study are openly available in https://doi.org/10.17605/OSF.IO/PY5EB, reference number [20].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Brown, R.E.; Basheer, R.; McKenna, J.T.; Strecker, R.E.; McCarley, R.W. Control of Sleep and Wakefulness. *Physiol. Rev.* **2012**, *92*, 1087–1187. [CrossRef] [PubMed]
2. Assefa, S.Z.; Diaz-Abad, M.; Wickwire, E.M.; Scharf, S.M.; Assefa, S.Z.; Diaz-Abad, M.; Wickwire, E.M.; Scharf, S.M. The Functions of Sleep. *Aims Neurosci.* **2015**, *2*, 155–171. [CrossRef]
3. Arrigoni, E.; Fuller, P.M. Chapter 5—The Circuit, Cellular, and Synaptic Bases of Sleep-Wake Regulation. In *Handbook of Behavioral Neuroscience*; Dringenberg, H.C., Ed.; Handbook of Sleep Research; Elsevier: Amsterdam, The Netherlands, 2019; Volume 30, pp. 65–88.
4. Eugene, A.R.; Masiak, J. The Neuroprotective Aspects of Sleep. *MEDtube Sci.* **2015**, *3*, 35–40. [PubMed]
5. Ringli, M.; Huber, R. Chapter 5—Developmental Aspects of Sleep Slow Waves: Linking Sleep, Brain Maturation and Behavior. In *Progress in Brain Research*; Van Someren, E.J.W., Van Der Werf, Y.D., Roelfsema, P.R., Mansvelder, H.D., Lopes Da Silva, F.H., Eds.; Slow Brain Oscillations of Sleep, Resting State and Vigilance; Elsevier: Amsterdam, The Netherlands, 2011; Volume 193, pp. 63–82.
6. Frankland, P.W.; Bontempi, B. The Organization of Recent and Remote Memories. *Nat. Rev. Neurosci.* **2005**, *6*, 119–130. [CrossRef] [PubMed]
7. Dyken, M.E.; Afifi, A.K.; Lin-Dyken, D.C. Sleep-Related Problems in Neurologic Diseases. *Chest* **2012**, *141*, 528–544. [CrossRef]
8. Bishir, M.; Bhat, A.; Essa, M.M.; Ekpo, O.; Ihunwo, A.O.; Veeraraghavan, V.P.; Mohan, S.K.; Mahalakshmi, A.M.; Ray, B.; Tuladhar, S.; et al. Sleep Deprivation and Neurological Disorders. *Biomed. Res. Int.* **2020**, *2020*, 5764017. [CrossRef]
9. Zamore, Z.; Veasey, S.C. Neural Consequences of Chronic Sleep Disruption. *Trends Neurosci.* **2022**, *45*, 678–691. [CrossRef] [PubMed]
10. Almeida-Filho, D.G.; Queiroz, C.M.; Ribeiro, S. Memory Corticalization Triggered by REM Sleep: Mechanisms of Cellular and Systems Consolidation. *Cell. Mol. Life Sci.* **2018**, *75*, 3715–3740. [CrossRef]
11. Ribeiro, S.; Mello, C.V.; Velho, T.; Gardner, T.J.; Jarvis, E.D.; Pavlides, C. Induction of Hippocampal Long-Term Potentiation during Waking Leads to Increased Extrahippocampal Zif-268 Expression during Ensuing Rapid-Eye-Movement Sleep. *J. Neurosci.* **2002**, *22*, 10914–10923. [CrossRef] [PubMed]
12. Ribeiro, S.; Gervasoni, D.; Soares, E.S.; Zhou, Y.; Lin, S.-C.; Pantoja, J.; Lavine, M.; Nicolelis, M.A.L. Long-Lasting Novelty-Induced Neuronal Reverberation during Slow-Wave Sleep in Multiple Forebrain Areas. *PLoS Biol.* **2004**, *2*, e24. [CrossRef]
13. Carver, K.; Saltoun, K.; Christensen, E.; Abosch, A.; Zylberberg, J.; Thompson, J.A. Towards Automated Sleep-Stage Classification for Adaptive Deep Brain Stimulation Targeting Sleep in Patients with Parkinson's Disease. *Commun. Eng.* **2023**, *2*, 95. [CrossRef]

14. Dzirasa, K.; Ribeiro, S.; Costa, R.; Santos, L.M.; Lin, S.-C.; Grosmark, A.; Sotnikova, T.D.; Gainetdinov, R.R.; Caron, M.G.; Nicolelis, M.A.L. Dopaminergic Control of Sleep–Wake States. *J. Neurosci.* **2006**, *26*, 10577–10589. [CrossRef] [PubMed]
15. Rechtschaffen, A.; Kales, A. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*; Public Health Service, US Government Printing Office: Washington, DC, USA, 1968.
16. Gervasoni, D.; Lin, S.-C.; Ribeiro, S.; Soares, E.S.; Pantoja, J.; Nicolelis, M.A.L. Global Forebrain Dynamics Predict Rat Behavioral States and Their Transitions. *J. Neurosci.* **2004**, *24*, 11137–11147. [CrossRef] [PubMed]
17. Hosseini, M.-P.; Hosseini, A.; Ahi, K. A Review on Machine Learning for EEG Signal Processing in Bioengineering. *IEEE Rev. Biomed. Eng.* **2021**, *14*, 204–218. [CrossRef] [PubMed]
18. Katsageorgiou, V.-M.; Sona, D.; Zanotto, M.; Lassi, G.; Garcia-Garcia, C.; Tucci, V.; Murino, V. A Novel Unsupervised Analysis of Electrophysiological Signals Reveals New Sleep Substages in Mice. *PLoS Biol.* **2018**, *16*, e2003663. [CrossRef] [PubMed]
19. Huang, W.; Guo, B.; Shen, Y.; Tang, X.; Zhang, T.; Li, D.; Jiang, Z. Sleep Staging Algorithm Based on Multichannel Data Adding and Multifeature Screening. *Comput. Methods Programs Biomed.* **2020**, *187*, 105253. [CrossRef] [PubMed]
20. Barger, Z.; Frye, C.G.; Liu, D.; Dan, Y.; Bouchard, K.E. Robust, Automated Sleep Scoring by a Compact Neural Network with Distributional Shift Correction. *PLoS ONE* **2019**, *14*, e0224642. [CrossRef] [PubMed]
21. Van Der Donckt, J.; Van Der Donckt, J.; Deprost, E.; Vandenbussche, N.; Rademaker, M.; Vandewiele, G.; Van Hoecke, S. Do Not Sleep on Traditional Machine Learning: Simple and Interpretable Techniques Are Competitive to Deep Learning for Sleep Scoring. *Biomed. Signal Process. Control* **2023**, *81*, 104429. [CrossRef]
22. Arnulfo, G.; Hirvonen, J.; Nobili, L.; Palva, S.; Palva, J.M. Phase and Amplitude Correlations in Resting-State Activity in Human Stereotactical EEG Recordings. *Neuroimage* **2015**, *112*, 114–127. [CrossRef]
23. Arnulfo, G.; Wang, S.H.; Myrov, V.; Toselli, B.; Hirvonen, J.; Fato, M.M.; Nobili, L.; Cardinale, F.; Rubino, A.; Zhigalov, A.; et al. Long-Range Phase Synchronization of High-Frequency Oscillations in Human Cortex. *Nat. Commun.* **2020**, *11*, 5363. [CrossRef]
24. Womelsdorf, T.; Schoffelen, J.-M.; Oostenveld, R.; Singer, W.; Desimone, R.; Engel, A.K.; Fries, P. Modulation of Neuronal Interactions Through Neuronal Synchronization. *Science* **2007**, *316*, 1609–1612. [CrossRef]
25. Steriade, M. The Corticothalamic System in Sleep. *Front. Biosci.* **2003**, *8*, d878–d899. [CrossRef]
26. Roascio, M.; Canessa, A.; Trò, R.; Mattioli, P.; Famà, F.; Giorgetti, L.; Girtler, N.; Orso, B.; Morbelli, S.; Nobili, F.; et al. Phase and Amplitude Electroencephalography Correlations Change with Disease Progression in People with Idiopathic Rapid Eye-Movement Sleep Behavior Disorder. *Sleep* **2022**, *45*, zsab232. [CrossRef]
27. Canolty, R.T.; Knight, R.T. The Functional Role of Cross-Frequency Coupling. *Trends Cogn. Sci.* **2010**, *14*, 506–515. [CrossRef] [PubMed]
28. Siebenhühner, F.; Wang, S.H.; Arnulfo, G.; Lampinen, A.; Nobili, L.; Palva, J.M.; Palva, S. Genuine Cross-Frequency Coupling Networks in Human Resting-State Electrophysiological Recordings. *PLoS Biol.* **2020**, *18*, e3000685. [CrossRef] [PubMed]
29. Tort, A.B.L.; Komorowski, R.; Eichenbaum, H.; Kopell, N. Measuring Phase-Amplitude Coupling Between Neuronal Oscillations of Different Frequencies. *J. Neurophysiol.* **2010**, *104*, 1195–1210. [CrossRef] [PubMed]
30. Fell, J.; Axmacher, N. The Role of Phase Synchronization in Memory Processes. *Nat. Rev. Neurosci.* **2011**, *12*, 105–118. [CrossRef] [PubMed]
31. Köster, M.; Finger, H.; Graetz, S.; Kater, M.; Gruber, T. Theta-Gamma Coupling Binds Visual Perceptual Features in an Associative Memory Task. *Sci. Rep.* **2018**, *8*, 17688. [CrossRef]
32. de Hemptinne, C.; Ryapolova-Webb, E.S.; Air, E.L.; Garcia, P.A.; Miller, K.J.; Ojemann, J.G.; Ostrem, J.L.; Galifianakis, N.B.; Starr, P.A. Exaggerated Phase–Amplitude Coupling in the Primary Motor Cortex in Parkinson Disease. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 4780–4785. [CrossRef]
33. Edakawa, K.; Yanagisawa, T.; Kishima, H.; Fukuma, R.; Oshino, S.; Khoo, H.M.; Kobayashi, M.; Tanaka, M.; Yoshimine, T. Detection of Epileptic Seizures Using Phase-Amplitude Coupling in Intracranial Electroencephalography. *Sci. Rep.* **2016**, *6*, 25422. [CrossRef] [PubMed]
34. Hirano, S.; Nakhnikian, A.; Hirano, Y.; Oribe, N.; Kanba, S.; Onitsuka, T.; Levin, M.; Spencer, K.M. Phase-Amplitude Coupling of the Electroencephalogram in the Auditory Cortex in Schizophrenia. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **2018**, *3*, 69–76. [CrossRef] [PubMed]
35. Møller, M.F. A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning. *Neural Netw.* **1993**, *6*, 525–533. [CrossRef]
36. Muñoz-Zavala, A.E.; Macías-Díaz, J.E.; Alba-Cuéllar, D.; Guerrero-Díaz-de-León, J.A. A Literature Review on Some Trends in Artificial Neural Networks for Modeling and Simulation with Time Series. *Algorithms* **2024**, *17*, 76. [CrossRef]
37. Sbrollini, A.; Leoni, C.; Morettini, M.; Rivolta, M.W.; Swenne, C.A.; Mainardi, L.; Burattini, L.; Sassi, R. Identification of Electrocardiographic Patterns Related to Mortality with COVID-19. *Appl. Sci.* **2024**, *14*, 817. [CrossRef]
38. Hülsemann, M.J.; Naumann, E.; Rasch, B. Quantification of Phase-Amplitude Coupling in Neuronal Oscillations: Comparison of Phase-Locking Value, Mean Vector Length, Modulation Index, and Generalized-Linear-Modeling-Cross-Frequency-Coupling. *Front. Neurosci.* **2019**, *13*, 573. [CrossRef]
39. Kwasinski, A.; Wu, M. What Is the Future of Signal Processing?: Views Across Our Community [Community Voices]. *IEEE Signal Process. Mag.* **2017**, *34*, 14–25. [CrossRef]