

Article

Not peer-reviewed version

Statistical Complexity Analysis of Sleep Stages

[Cristina D. Duarte](#)*, [Mariana Pacheco](#), [Francisco R. Iaconis](#), [Osvaldo A. Rosso](#), [Gustavo Gasaneo](#), [Claudio A. Delrieux](#)

Posted Date: 26 November 2024

doi: 10.20944/preprints202411.1941.v1

Keywords: permutation entropy; statistical complexity; generalized weighted permutation entropy; sleep stages









Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Statistical Complexity Analysis of Sleep Stages

Cristina D. Duarte ^{1,*} , Marianela Pacheco ^{1,2,4} , Francisco R. Iaconis ¹ , Osvaldo A. Rosso ³ , Gustavo Gasaneo ¹  and Claudio A. Delrieux ⁴ 

¹ Departamento de Física, Instituto de Física del Sur, Universidad Nacional del Sur and CONICET, Bahía Blanca, Argentina

² Universidad Nacional de Córdoba

³ Instituto de Física, Universidade Federal de Alagoas UFAL, Maceió 57072-900, Brazil

⁴ Departamento de Ing. Eléctrica y Computadoras, Instituto de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur and CONICET, Bahía Blanca, Argentina

* Correspondence: cristina.duarte@uns.edu.ar

Abstract: Studying sleep stages is crucial for understanding sleep architecture, which can help identify various health conditions, including insomnia, sleep apnea, and neurodegenerative diseases, and allowing for better diagnosis and treatment interventions. In this paper we explore the effectiveness of generalized weighted permutation entropy (GWPE) in distinguishing between different sleep stages from EEG signals. We provide a comparison of the performance of various machine learning algorithms using features derived from both standard permutation entropy and GWPE, demonstrating that GWPE significantly improves classification accuracy, particularly in identifying the transition between N1 and REM sleep. The results highlight the potential of GWPE as a valuable tool for understanding sleep neurophysiology and improving the diagnosis of sleep disorders.

Keywords: permutation entropy; statistical complexity; generalized weighted permutation entropy; sleep stages

1. Sleep Stage Interpretation and Analysis

Sleep and waking states are periodic biobehavioral features that are characterized by changes in brain electrical activity, manifested as alterations in consciousness, reduced sensory responsiveness, decreased muscle tone, and relative inactivity [1]. This process has an important adaptive function in the homeostasis of mammals, guaranteeing their survival [2]. In humans, alterations in the quality, quantity, and patterns of sleep (*i.e.*, sleep disorders) negatively impact the individual's overall health and functionality [1,3–5]. Sleep disorders are extremely common in the general population [1,5].

Normal human sleep consists of two primary states: rapid eye movement (REM) and non-REM, each with distinct stages and characteristics. Furthermore, non-REM sleep includes several stages, ranging from light to deep sleep, while REM sleep is associated with vivid dreaming and brain activity patterns similar to wakefulness. These sleep stages undergo age-related changes and are influenced by factors such as sleep-wake history, circadian timing, and pharmacological agents, all of which impact the distribution and quality of REM and non-REM sleep [6]. Accurately assessing these sleep stages is essential for understanding various health conditions, including insomnia, sleep apnea, and neurodegenerative diseases such as Parkinson's and Alzheimer's [7].

For example, disruptions in deep sleep have been linked to cognitive decline [8,9], while abnormalities in REM sleep may indicate certain mood disorders [10,11] or post-traumatic stress disorder [12,13]. This analysis requires robust methods to discriminate sleep stages from physiological signals. The electrical signals of the brain recorded by EEG are the most used technique to understand the microstructure of sleep waves [14]. By analyzing these signals, the different stages of sleep can be identified—from light sleep (N1, N2) to deep sleep (N3) and REM stages—and assess their duration and how a person transitions between them. This allows for accurate identification of sleep disorders, which can lead to more effective diagnostic and treatment interventions [15].

The EEG signal reflects the electrical activity of the brain, which forms a chaotic, non-linear temporal signal whose features have been studied from the standpoint of statistical complexity, with efforts aimed at developing automatic and precise methods for classifying sleep stages. Among the earliest studies to apply statistical complexity analysis in sleep stages is presented in [16]. This paper

discusses how permutation entropy (PE) can be applied to analyze EEG sleep signals in artifact-free segments of 30 s, providing measures that help in identifying different stages and detecting anomalies, such as those related to neurological conditions. The study shows that PE is useful for characterizing sleep EEG and offers a computationally efficient tool to measure the complexity of brain activity during different sleep stages, contributing to a better understanding of sleep physiology and potential disorders. In [14], the authors explore the statistical complexity of EEG signals during sleep, particularly focusing on the Cyclic Alternating Pattern (CAP), which is a key feature of sleep microstructure. The study applied complexity based features like fractal dimension and sample entropy to evaluate these activations and compared them to non-activation periods in NREM sleep. Their findings suggest that different subtypes of CAP activations (A1, A2, and A3) have distinct levels of EEG complexity, with A3 showing the highest complexity across all sleep stages. The ability of ordinal patterns to detect minute features in EEG signals was exposed in [17]. A modified PE assessment is here understood as a measure of distance from white noise, with the support of a specific statistical hypotheses test in which critical significance values are provided. This paper presented the first evidence of automated analysis that goes beyond expert human assessment, since the method is able to detect delta waves patterns and graphic features which are only visible on a time scale of several seconds. At the same time, the "distance from white noise" measure serves as an indicator of sleep depth, in which the awake state is being treated as the nearest to white noise.

The use of machine learning methods in classifying sleep stages is also proposed in [18], specifically to score sleep stages using single-channel electrooculogram (EOG) signals. These signals are analyzed using several features (discrete wavelet transform, spectral entropy, moment-based measures, composite multiscale dispersion entropy, and autoregressive coefficients). The discriminative power of these features is assessed through ANOVA tests, and a feature reduction is applied to decrease model complexity and retain the most discriminative features. Finally, sleep stage classification is performed using Random Under-Sampling Boosting (RUSBoost), Random Forest (RF), and Support Vector Machines (SVM). On a similar setting, in [19] the authors combine entropy features (including fuzzy measure entropy, fuzzy entropy, and sample entropy), with a multi-class SVM to classify sleep stages. The method analyzes 30 s segments of Fpz-Cz and Pz-Oz EEG signals, together with horizontal EOG signals from the Sleep Telemetry Study dataset [20]. Given the widespread success of ordinal patterns and statistical complexity based signal analysis, there was a recent influx of alternative techniques that may outperform the more general methods. In particular, in [21] the authors propose ensemble improved permutation entropy and multiscale ensemble improved PE for taking into account both permutation relations and amplitude information. The proposed methods show better discriminating power and robustness, specifically in the analysis of EEG signals. In this work we adopt a similar aim to identify sleep stages in polysomnographs, by training a classifier using novel statistical complexity measures as features, specifically Generalized Weighted Permutation Entropy.

While deep learning methods excel at automating feature extraction, their black-box nature and high resource demands can make them less suitable for EEG analysis, where interpretability and performance with minimal data are often priorities. Our results outperform traditional analysis based on ordinal patterns, achieving accuracies comparable to those obtained with deep learning [22]. The advantage of using feature-engineered white-box models in this context (as opposed to data-driven deep learning models) is that handcrafted features are more interpretable in neurophysiological terms, providing clearer insights into the underlying neural mechanisms, which is crucial in fields like neuroscience and clinical diagnosis in which model explainability is required. Additionally, feature-engineered models are usually less computationally intensive and can work effectively with smaller datasets, and are much easier to re-train, making them a more practical choice in scenarios where labeled EEG data is varying or limited.

2. Materials and Methods

2.1. Database Description

The dataset used were obtained from the Bob Kemp's publicly available database on Physionet, linked to the Sleep Telemetry Study [20] already mentioned, which consists of a set of polysomnographic signal recordings. That study investigated the effects of temazepam on sleep in 22 Caucasian men and women, ages 18 to 79, 7 men (mean age 35.71) and 15 women (mean age 42.26). The subjects had mild difficulties falling asleep, were not taking any other medications, and were healthy subjects. This dataset is made up of a total of 44 recordings in the hospital, corresponding to two nights, that is, two recordings per subject. Polysomnographic recordings last nine hours, including the location of the Fpz-Cz and Pz-Oz electrodes, and were performed after temazepam ingestion or after placebo ingestion. For the present work, the corresponding records were included for subjects between 18 and 40 years old after taking placebo. The dataset was made up of 12 polysomnographic records of 12 subjects (total mean age for men and women 25.83, standard deviation: 6.55, sample rate 100 hz). More specifically, the study was composed of 8 women (mean age women 27.87, standard deviation: 7.20) and 4 men (mean age men 21.75, standard deviation: 1.71). We used only the Fpz-Cz channel of the EEG signal. These recordings included expert annotations of sleep stages, labeled in 30-second intervals which the experts termed *epochs*, each consisting of 3000 data points.

2.2. Generalized probabilities of ordinal patterns

Permutation entropy (PE) [23] is a robust, non-linear time series analysis method that quantifies the complexity and randomness of a system, gaining widespread adoption due to its simplicity, computational efficiency, and ability to capture dynamic patterns in diverse fields such as neuroscience, finance, and engineering. The speed and robustness that characterizes PE makes it particularly useful for handling large datasets without requiring extensive preprocessing or parameter tuning [17].

The computation of PE involves transforming the time series into a sequence of symbols known as *ordinal patterns* (OPs). These OPs are characterized by two key parameters: the length of the pattern, D , and the time delay τ , which denotes the interval between the samples in the signal used to define the patterns. Given a time series x_t where $t = 1, \dots, T$, overlapping segments s of length D are taken with a τ interval between consecutive samples. This results in a total of $T - (D - 1)$ segments. Within each segment, the D samples are labeled with ordinal numbers $1, 2, \dots, D$ according to their ordinal relationships with other values in the same segment. Then, each segment is labeled with an OP, which is one permutation π among the $D!$ possible permutations of D ordinal values. In a given signal, the relative frequency (or empirical probability) of each OP π_i is then

$$p(\pi_i) = \frac{\sum_{s, \pi_s = \pi_i}}{\sum_s}, \quad (1)$$

where s represent each of the sequences and π_s the corresponding OP of s . The probability distribution function (PDF) of the $D!$ OPs is then $P_\pi = \{p(\pi_1), \dots, p(\pi_{D!})\}$. From this PDF, the PE of the signal can be computed using Shannon entropy:

$$S = - \sum_i^{D!} p(\pi_i) \log(p(\pi_i)). \quad (2)$$

It can be shown that S takes its minimum value of 0 when there is total certainty about the series, *i.e.*, a given permutation has probability one. The maximum value corresponds to the case where uncertainty is maximized ($\log(D!)$), which corresponds to a uniform probability distribution $P_U = \{p_i = 1/M, i = 1, \dots, M\}$, where all states have the same occurrence probability.

Statistical complexity, in turn, is defined as the product of normalized Shannon entropy H and disequilibrium Q :

$$C = Q[P_\pi, P_u] \cdot H[P_\pi], \quad (3)$$

where normalized Shannon entropy H is defined as S/S_{max} and disequilibrium Q is defined in term of the Jensen-Shannon divergence $Q = Q_0 \cdot D[P_\pi, P_u]$, and the divergence D provides a measure of the similarity between two probability densities. Here, the PDF is compared with the uniform distribution.

$$D[P_\pi, P_u] = S\left(\frac{P_\pi + P_u}{2}\right) - \left(\frac{S[P_\pi] + S[P_u]}{2}\right) \quad (4)$$

The lowest complexity value occurs when the system is either entirely random or entirely ordered. When the system is in a state of maximal disorder, there is no structural information, the complexity is low. When the system is in a highly ordered state is fully predictable, meaning there is no uncertainty in the pattern distribution, which also results in low complexity. Entropy and complexity can be articulated together in the complexity-entropy causality plane. In [24] it is shown how to simultaneously quantify both the information content and structural complexity of a time series, distinguishing between stochastic noise and deterministic chaotic behavior, leading to many applications in analyzing and characterizing time series and signals from physiology, physics, and many other fields.

While PE has been effective in various applications, the resulting assessments may have limitations when the local amplitude of the time series is very uneven, since smaller fluctuations should exert less influence in the overall analysis. To this avail, in [25] the authors propose Weighted PE (WPE), which assigns more significance to OPs with higher amplitude fluctuations. This is accomplished incorporating the local variance of each OP as a weighing factor of its contribution to the overall empirical probabilities. This modification allows WPE to capture both the ordinal structure and amplitude information of a time series, making it more suitable for applications where amplitude plays a key role. Furthermore, in [26] the authors propose the Generalized Weighed PE (GWPE), which also considers an entropic index q that allows to discern between the effects of small or large fluctuations in the overall complexity of the signal. The associated probability distribution is then

$$p(\pi_i, q) = \frac{\sum_{s, \pi_s = \pi_i} w_s^{q/2}}{\sum_s w_s^{q/2}}. \quad (5)$$

GWPE coincides with WPE for $q = 2$, and with the original PE for $q = 0$, making the notation and parameters compatible with other nonextensive measures (*f.e.*, Tsallis entropy) or nonstationary fluctuation analysis (*f.e.*, multifractal spectrum). As with other nonextensive/nonstationary models, entropic indices are used to generalize the concept of entropy to systems with long-range interactions or memory effects, fractal structures, or complex correlations, providing a more accurate description of their statistical behavior and properties. In our particular context, the relevance of the use of GWPE lays in how different ordinal patterns are weighted, given that negative q values enhance the contribution of small fluctuations in the OP distribution, while positive values enhance the contribution of large fluctuations.

In this study we are focused on determining how entropic indices may enhance the ability of a PE-based polysomnographic signal analysis. As already mentioned, this approach enables a more flexible and detailed evaluation of the disorder and structure of the time series. A likely interpretation of Eq. 5 is that now PE is a function of the entropic index and thus signals can be assessed according to this function. In particular, instead of evaluating this PE function for signals, or characterizing them in the complexity-entropy causality plane as in [26], our idea is to use the generalized probabilities $p(\pi_i, q)$ of the OPs as features for classification, and study the ability of the resulting classifier, as compared to the standard PE, to distinguish apart the different sleep stages, as will be described in the next subsection.

2.3. Data Analysis

For each subject, the EEG signal was segmented according to the labeled epochs, with each epoch corresponding to a specific sleep stage. The number of epochs across all subjects was $N_1=885$, $N_2=5281$, $N_3=1995$, $R=2599$, and $W=803$, for a total of 11563 epochs. Given the uneven distribution of epochs, we downsampled N_2 class randomly dropping half of the epochs, and we performed a data augmentation procedure with N_1 and W (waking state), creating new epochs in the middle of each pair of consecutive epochs belonging to the same class (*i.e.*, starting at the midpoint of the first epoch and ending at the second epoch). Furthermore, for each sleep cycle, we removed the two boundary epochs marking the transition from one sleep phase to the next. The final number of epochs in each state was $N_1 = 978$, $N_2 = 2247$, $N_3 = 1643$, $R = 2444$, and $W = 1017$. The total number of epochs across all states is 8329.

In calculating the generalized probabilities $p(\pi_i, q)$ for each subject and epoch, we used a time delay of $\tau = 1$, a segment length $D=4$. These parameter values were chosen based on [27], which demonstrated their effectiveness in characterizing sleep stages. The index q was varied in integer values in the $[-10,10]$ interval, for a total of 21 q values. Using these values, the generalized probabilities of each of the $D!$ ordinal patterns in each epoch were computed, giving in all 504 features (24×21) for each epoch. Thus the shape of the dataset for training the classifier was 8329×505 , being the former the rows corresponding to each epoch, and the later the columns, including the 504 features plus the corresponding sleep stage label.

With this dataset, several classification models were trained, including random forest, support vector machines and XGBoost. In each case, the dataset was split into 85% for training and 15% for testing. The training data was further divided using 5-fold cross-validation, hyperparameter tuning procedures were performed, and also feature selection for evaluating the effectiveness of the different features in distinguishing sleep stages W , N_1 , N_2 , N_3 , and REM . The optimal hyperparameters for each model were determined using RandomizedSearch implemented through the scikit-learn library [28], with 75 random permutations of parameter values. The performance of the best hyperparameters was evaluated based on accuracy. The above procedures were applied to the full set of features (corresponding to the generalized probabilities) and with only the 24 features corresponding to standard PE ($q = 0$). In addition, the PE and complexity of the 21 q values was assessed.

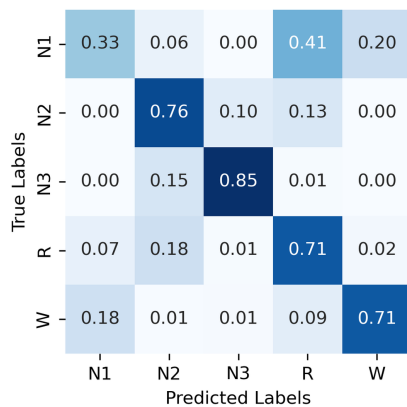
3. Results

3.1. Classifier and Feature Set Comparison for Sleep Stage Classification

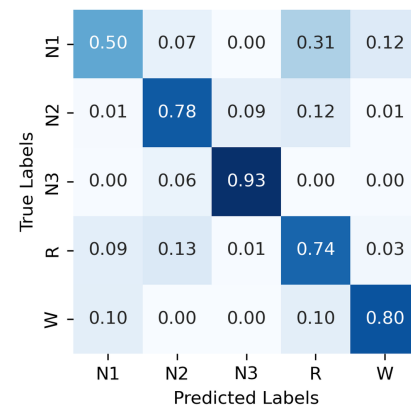
To illustrate the advantages of the generalized probabilities of the OPs in improving sleep stage classification, we trained three algorithms to automatically classify the five stages. We evaluated the algorithm performance with different feature sets to identify which features best separate the sleep stages. Each algorithm was computed in the following order:

- Permutation entropy (PE) and complexity (C)
- Generalized weighted permutation entropy (GWPE) and complexity (GWPEC)
- The probability distribution function of the ordinal patterns (PDF)
- The generalized weighted probability distribution of the ordinal patterns (GWPDF)

The normalized confusion matrices for the classification of sleep stages using XGBoost, trained with (PE) and (C), as well as (GWPE) and (GWPEC), are presented in Figure 1. Similarly, Figure 2 shows the confusion matrices for the classification using the probability distribution of the ordinal patterns (PDF) and generalized weighted probability distribution (GWPDF). All confusion matrices were calculated using the classifier optimized through cross-validation. Furthermore, the matrices were generated after performing feature importance analysis on both the generalized weighted permutation entropy and complexity, as well as on the generalized weighted probability distribution of ordinal patterns. For simplicity and clarity, only the confusion matrices for the XGBoost classifier are presented in the main text. Confusion matrices for the other classifiers have been included in Appendix A.1 for reference.

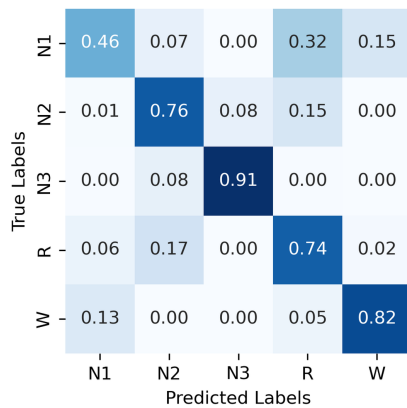


(a) XGBoost algorithm; PE and C as features

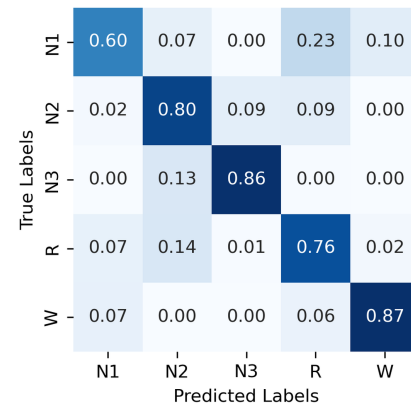


(b) XGBoost algorithm; GWPE and GWPEC as features

Figure 1. Normalized confusion matrix (%) for XGBoost algorithm classifier, with PE and C on the left and, GWPE and GWPEC on the right as features.



(a) XGBoost algorithm; PDF as features



(b) XGBoost algorithm; GWPDF as features

Figure 2. Normalized confusion matrix (%) for XGBoost algorithm classifier, with PDF on the left and, GWPDF on the right as features.

To compare the performance of all classifiers across the different feature sets, Table 1 summarizes the mean accuracy and standard deviation obtained from cross-validation. Feature importance analysis was conducted to identify the most relevant features that contribute to the model's predictive performance using XGBoost and random forest. The results for XGBoost, using GWPDF as features, as well as the results for GWPE and GWPEC as features, are presented in Table 2.

Table 1. Mean accuracy (Acc.) and standard deviation (SD) for the classifiers: random forest (RF), support vector machine (SVM) and XGBoost.

| Features | RF | | SVM | | XGBoost | |
|----------------|------|------|------|------|---------|------|
| | Acc | SD | Acc | SD | Acc | SD |
| PE and C | 0.71 | 0.01 | 0.72 | 0.01 | 0.71 | 0.01 |
| GWPE and GWPEC | 0.76 | 0.01 | 0.77 | 0.01 | 0.77 | 0.01 |
| PDF | 0.74 | 0.01 | 0.74 | 0.01 | 0.74 | 0.01 |
| GWPDF | 0.79 | 0.01 | 0.78 | 0.01 | 0.80 | 0.01 |

The importance of each feature was calculated based on the total accuracy gain of the feature when it is used in training. In Table 2 are presented, for both features sets, the features, the corresponding

index q used in the calculation, and their relative importance. Only the top 10 features are displayed, ordered based on their impact on the classification accuracy. Using OP, the top 10 features contributed to 35% of the cumulative importance, whereas for GWPE, these same top 10 features contribute to 64% of the cumulative importance.

Table 2. Top ten features ranked by relative importance for XGBoost classifiers. The classifiers were trained with GWPDF (left) and GWPE and GWPEC (right). For both sets, the first column lists the respective feature, the second column presents the entropic index q used to calculate the feature, and the third column shows the relative importance of each feature.

| OPs | | | GWPE and GWPEC | | |
|--------------|-----|---------------------|----------------|-----|---------------------|
| Features | q | Relative Importance | Features | q | Relative Importance |
| $P_{[3210]}$ | 0 | 0.0868 | H | 1 | 0.1655 |
| $P_{[0123]}$ | 1 | 0.0716 | H | 0 | 0.1028 |
| $P_{[0123]}$ | -1 | 0.0355 | C | 3 | 0.0874 |
| $P_{[1032]}$ | 2 | 0.0290 | C | 1 | 0.4366 |
| $P_{[1032]}$ | 3 | 0.0286 | H | -1 | 0.0595 |
| $P_{[3210]}$ | 1 | 0.0285 | C | 0 | 0.0416 |
| $P_{[0123]}$ | 0 | 0.0210 | C | -1 | 0.0300 |
| $P_{[3102]}$ | 2 | 0.0206 | H | 2 | 0.0277 |
| $P_{[0213]}$ | 2 | 0.0204 | C | 2 | 0.0214 |
| $P_{[0213]}$ | 3 | 0.0124 | C | 7 | 0.0208 |

3.2. Behavior of PE, C vs. q and POs vs. q

To illustrate how the entropic index q exerts influence in differentiating between sleep stages, we computed the median values of permutation entropy $\langle PE \rangle$ and complexity $\langle C \rangle$ for all subjects across each sleep stage, as a function of q parameter. The exploration of the q parameter is showed in Figure 3. This sheds light on how different values of q affect the separation of sleep stages based on generalized permutation entropy and complexity. In both cases, $\langle PE \rangle$ and $\langle C \rangle$, for $q < -1$, the medians across all stages remain closely clustered. Starting from $q = -1$, the median values begin to diverge. This behavior is consistent with the feature selection performed in the cases of RF and XGBoost classifiers. In these cases, it is shown that entropy and complexity with positive exponents have the highest relative importance.

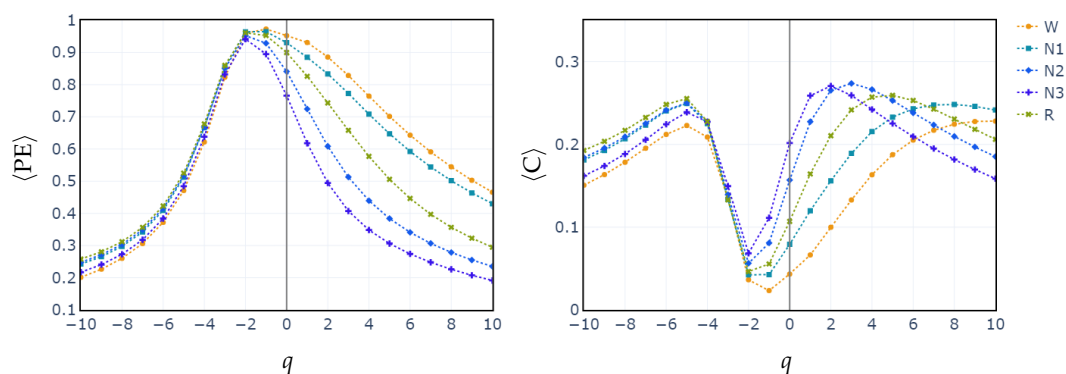


Figure 3. Median of values of generalized permutation entropy (left) and complexity (right) for q across all sleep stages.

Additionally, we plot the median values of the probability distributions for the patterns [0123] and [3210]. We choose this patterns because they exhibit the highest relative importance in the XGBoost classification. We observe once again that the separation between the stages begins at $q \geq -1$.

4. Discussion

Starting with a comparison of the performance of the three classifiers (RF, SVM, and XGBoost) across the different feature sets, we observe that all algorithms show an improvement in median accuracy when moving from standard PE and C to GWPE and GWPEC (Table 1). The inclusion of the entropic index q increases the performance consistently, indicating that it captures more fully the complexity of the EEG signal. The use of PDF as a feature set improves the classification as compared with the standard PE but does not fully capture the signal's complexity. This is likely due to the fact that local signal amplitudes may be relevant in the identification of sleep stages. In contrast, the GWPE distribution function provides the best median accuracy across all classifiers, achieving more than 80% accuracy with XGBoost. This demonstrates that this feature set is the most effective at distinguishing sleep stages. Moreover, an analysis of feature informativeness reveals a contrasting distribution in predictive power across the two feature sets. In plain PE, the 10 most informative features capture only 35% of the total predictive performance gain of the classifiers, indicating a relatively dispersed distribution of informativeness across features. In contrast, in the GWPE feature set, the top 10 features account for 65% of the total gain, suggesting a more concentrated distribution where fewer features contribute a much larger amount of information relevant to classification accuracy. This discrepancy suggests differing levels of feature redundancy, with potential implications for model complexity and feature selection efficiency.

As the individual progresses into deeper levels of sleep, cortical brain waves become increasingly synchronized. This synchronization increases the signal amplitude while reducing frequency, which significantly alters the probabilities of the ordinal patterns, particularly increasing the likelihood of monotonic patterns [0123] and [3210], common in deterministic and periodic signals. This phenomenon results in a reduction in entropy and an increase in signal complexity. This feature explains the higher accuracies achieved for the N3 stage (as reflected in the confusion matrices across all features sets) since this stage mainly contains low-frequency, high-amplitude delta waves, in which these monotonic patterns prevail.

For N2, classification accuracy was generally high across the three algorithms using PE and C as features. However, adding the entropic index q further enhanced performance in both GWPE and GWPDF features sets (see Figures 1, 2, A1 and A2). Stage N2 is characterized by distinctive K-complexes (brief, large-amplitude spikes) and sleep spindles (12-16 Hz oscillatory waves). Considering signal amplitude likely improves the classifiers' ability to detect these patterns by highlighting specific amplitude and frequency occurrences, thereby reducing the risk of misinterpretation as noise, which may occur when using only PE and C.

In contrast, classifying N1 sleep remains challenging, particularly distinguishing it from wakefulness and REM, due to overlapping EEG signal features. During wakefulness, the EEG generally shows lower complexity due to regular alpha or beta waves. As the signal transitions into stage N1, there is an intermediate increase in complexity as alpha waves decrease and theta waves emerge, leading to GWPE values that are distinct from W and REM. In REM sleep, the EEG signal is highly complex and disordered, with a mixture of theta and beta waves and increased signal variability, resulting in higher GWPE values (see Figure 3). Incorporating generalized probabilities of ordinal patterns (OPs), weighted by the entropic index q , significantly improves the classifiers' ability to differentiate these stages. The confusion matrices demonstrate a significant enhancement in N1 classification accuracy, increasing from 33% (Figure 1a) to 60% (Figure 2b) with the use of generalized probabilities. This improvement can be observed in two ways. First, as shown in Figure 3, N1 become more distinguishable from W and R at $q \geq -1$ in PE and C. Second, as shown in Figure 4, the probability of weighted patterns captures finer distinctions between N1 and the other sleep stages. This suggests that incorporating both the variance of the EEG signal and the occurrence probabilities of each pattern may allow for deeper capture of the microstructure within sleep stages. Specifically, at values of $q \geq -1$, there is a marked improvement in stage discrimination, suggesting that large fluctuations in EEG

signals play a critical role in enhancing classification accuracy. This can be seen in Figures 3 and 4, where distinctions among all stages are apparent.

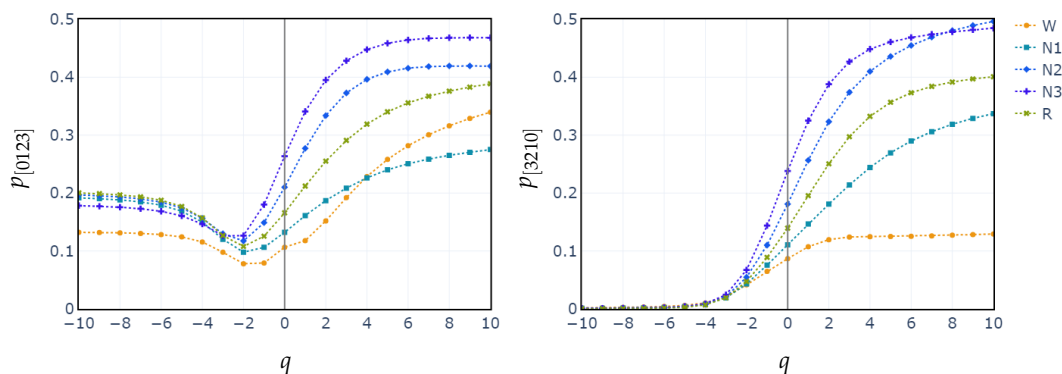


Figure 4. Median probabilities of the monotonic patterns [0123] and [3210] across all sleep stages for varying entropic index q .

This effect is also reflected in the feature importance analysis of the classifiers. As shown in Table 2, the entropic indices that achieved the highest relative importance in classifying sleep stages are those greater than or equal to -1. This applies to both the probability of ordinal patterns and the complexity and entropy measures (PE and C) derived from these probabilities. Thus, GWPE shows better performance in classifying sleep stages in EEG signals.

5. Conclusion

In this study, three classification algorithms (Random Forest, SVM, and XGBoost) were compared using different sets of features derived from labeled EEG signals. Our results demonstrate that incorporating weighted generalized permutation entropy (GWPE) and weighted generalized complexity (GWPEC) significantly improves the performance of classifiers, especially compared to standard approaches based on traditional permutation entropy and complexity. The use of the entropic index q was shown to be crucial for more effectively capturing the complexity of EEG signals, as it allows for more precise modulation of ordinal probabilities based on EEG fluctuations. On the other hand, classification of N1 and REM stages presents challenges due to similarities in brain activity in these stages. However, the use of generalized ordinal probabilities weighted by q contributed to improving the distinction between these phases, especially in values of $q \geq -1$, suggesting that wide fluctuations in the EEG signal are a key factor in effective classification.

Furthermore, feature importance analysis showed that the values of $q \leq -1$, have a prominent influence on classification, particularly on improving the accuracy at the N1 stage. Generalized weighed permutation entropy combined with machine learning pose a promising approach to improve automatic classification of sleep stages from the standpoint of statistical complexity. The ability to adjust ordinal probabilities to capture the nonlinear dynamics and complexity of the EEG has potential applications both in the clinical diagnosis of sleep disorders and in the study of other complex signals and time series. In future work, it would be necessary to explore in depth the effects of different values of the entropic index q in other populations and conditions such as in subjects with specific sleep disorders. On the other hand, it would be valuable to investigate whether the integration of other physiological signals, in addition to EEG, could further complement and improve the classification results, creating more robust and accurate diagnostic tools in the field of sleep disorders.

Author Contributions: Conceptualization, C.D.D., M.P., F.I., G.G. and C.A.D.; methodology, C.D.D. and C.A.D.; software, F.I. and C.D.D.; validation, C.A.D, G.G. and O.A.R ; formal analysis, O.A.R.; writing—original draft preparation, C.D.D and M.P.; writing—review and editing, C.A.D, G.G., F.I. and O.A.R.; supervision, C.A.D. and G.G.; funding acquisition, C.A.D. and G.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the doctoral and posdoctoral fellowships awarded by the National Research and Technology Council of Argentina (CONICET).

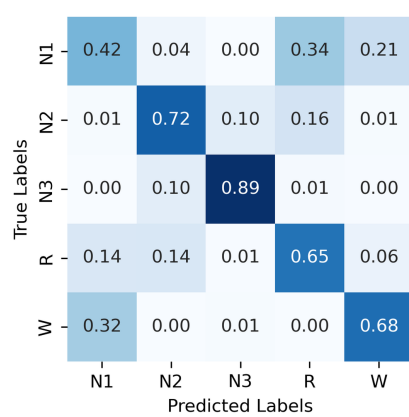
Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A.

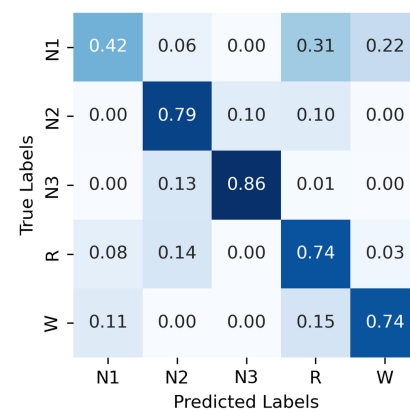
Appendix A.1. Confusion Matrix for Random Forest and Support Vector Machines.

This appendix presents the confusion matrices corresponding to the classification algorithms Random Forest (RF) and Support Vector Machine (SVM). The matrices are calculated using four different feature sets, allowing for the evaluation of each algorithm's performance in classifying sleep stages.

RF

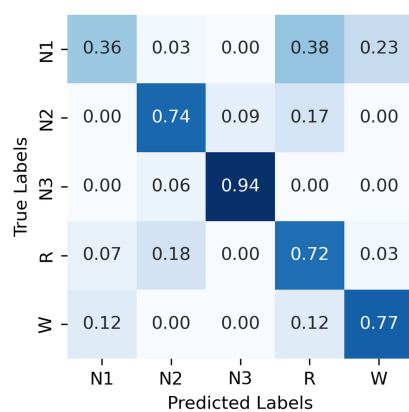


(a) RF algorithm; PE and C as features

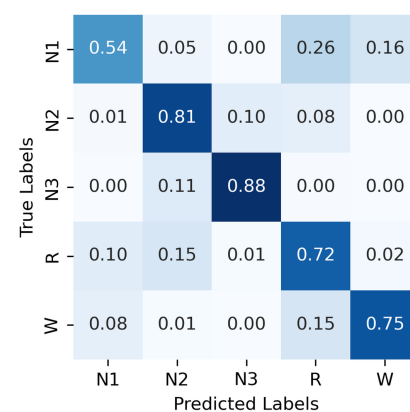


(b) RF algorithm; GWPE and GWPEC as features

SVM



(c) SVM algorithm; PE and C as features



(d) SVM algorithm; GWPE and GWPEC as features

Figure A1. Normalized confusion matrix (%) for RF and SVM algorithms classifier, with PE and C on the left and, GWPE and GWPEC on the right as features.

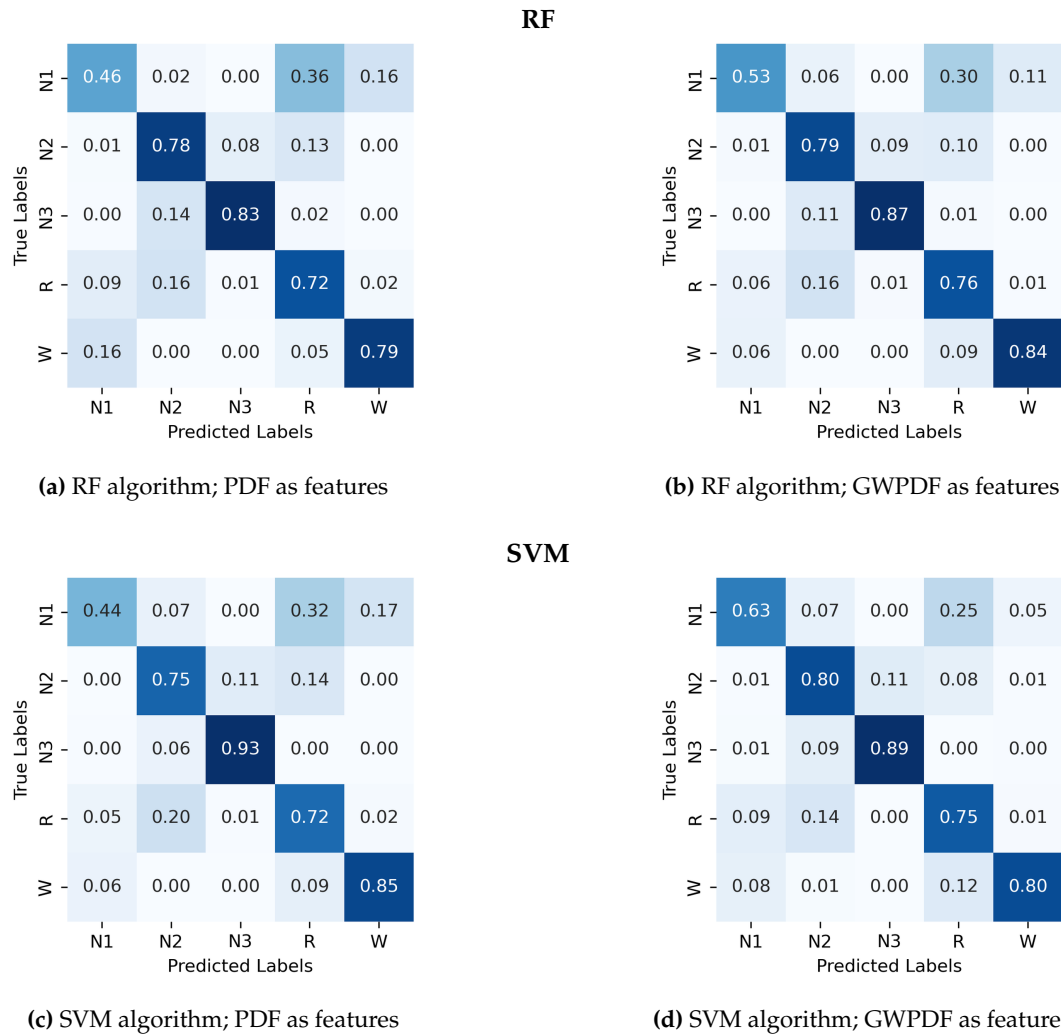


Figure A2. Normalized confusion matrix (%) for RF and SVM algorithms classifier, with PDF on the left and, GWPDF on the right as features.

References

- Gottesman, R.F.; Lutsey, P.L.; Benveniste, H.; Brown, D.L.; Full, K.M.; Lee, J.M.; Osorio, R.S.; Pase, M.P.; Redeker, N.S.; Redline, S.; Spira, A.P.; on behalf of the American Heart Association Stroke Council; Council on Cardiovascular, Nursing, S.; on Hypertension, C. Impact of Sleep Disorders and Disturbed Sleep on Brain Health: A Scientific Statement From the American Heart Association. *Stroke* **2024**, *55*, e61–e76, [<https://www.ahajournals.org/doi/pdf/10.1161/STR.0000000000000453>]. doi:10.1161/STR.0000000000000453.
- Rial, R.V.; Akaârîr, M.; Canellas, F.; Barceló, P.; Rubiño, J.A.; Martín-Reina, A.; Gamundí, A.; Nicolau, M.C. Mammalian NREM and REM sleep: Why, when and how. *Neuroscience & Biobehavioral Reviews* **2023**, *146*, 105041.
- Kumar, V.M. Sleep and sleep disorders. *Indian Journal of Chest Diseases and Allied Sciences* **2008**, *50*, 129.
- Palagini, L.; Geoffroy, P.A.; Miniati, M.; Perugi, G.; Biggio, G.; Marazziti, D.; Riemann, D. Insomnia, sleep loss, and circadian sleep disturbances in mood disorders: A pathway toward neurodegeneration and neuroprogression? A theoretical review. *CNS Spectrums* **2022**, *27*, 298–308. doi:10.1017/S1092852921000018.
- Panossian, L.A.; Avidan, A.Y. Review of Sleep Disorders. *Medical Clinics of North America* **2009**, *93*, 407–425. Common Neurologic Disorders, doi:<https://doi.org/10.1016/j.mcna.2008.09.001>.
- Patel, A.K.; Reddy, V.; Shumway, K.R.; Araujo, J.F. Physiology, sleep stages. In *StatPearls [Internet]*; StatPearls Publishing, 2024.

7. Malhotra, R.K. Neurodegenerative Disorders and Sleep. *Sleep Medicine Clinics* **2022**, *17*, 307–314. Commemorative Issue: 15 Years of the Sleep Medicine Clinics – PART 1: Sleep and Sleep Disorders, doi:<https://doi.org/10.1016/j.jsmc.2022.02.009>.
8. Gilley, R.R. The Role of Sleep in Cognitive Function: The Value of a Good Night's Rest. *Clinical EEG and Neuroscience* **2023**, *54*, 12–20. PMID: 35369784, doi:10.1177/15500594221090067.
9. Porter, V.R.; Buxton, W.G.; Avidan, A.Y. Sleep, Cognition and Dementia. *Current Psychiatry Reports* **2015**, *17*. doi:10.1007/s11920-015-0631-8.
10. Yan, T.; Qiu, Y.; Yu, X.; Yang, L. Glymphatic Dysfunction: A Bridge Between Sleep Disturbance and Mood Disorders. *Frontiers in Psychiatry* **2021**, *12*. doi:10.3389/fpsy.2021.658340.
11. Pearson, O.; Ugluk-Marucha, N.; Miskowiak, K.W.; Cairney, S.A.; Rosenzweig, I.; Young, A.H.; Stokes, P.R. The relationship between sleep disturbance and cognitive impairment in mood disorders: A systematic review. *Journal of Affective Disorders* **2023**, *327*, 207–216. doi:<https://doi.org/10.1016/j.jad.2023.01.114>.
12. de Boer, M.; Nijdam, M.J.; Jongedijk, R.A.; Bangel, K.A.; Olff, M.; Hofman, W.F.; Talamini, L.M. The spectral fingerprint of sleep problems in post-traumatic stress disorder. *Sleep* **2019**, *43*, zsz269, [<https://academic.oup.com/sleep/article-pdf/43/4/zsz269/33026528/zsz269.pdf>]. doi:10.1093/sleep/zsz269.
13. Barone, D.A. Dream enactment behavior—a real nightmare: A review of post-traumatic stress disorder, REM sleep behavior disorder, and trauma-associated sleep disorder. *Journal of Clinical Sleep Medicine* **2020**, *16*, 1943–1948.
14. Chouvarda, I.; Rosso, V.; Mendez, M.O.; Bianchi, A.M.; Parrino, L.; Grassi, A.; Terzano, M.; Cerutti, S. Assessment of the EEG complexity during activations from sleep. *Computer methods and programs in biomedicine* **2011**, *104*, e16–e28.
15. Kryger, M.H. *Atlas of Clinical Sleep Medicine: Expert Consult-Online*; Elsevier Health Sciences, 2022.
16. Nicolaou, N.; Georgiou, J. The Use of Permutation Entropy to Characterize Sleep Electroencephalograms. *Clinical EEG and Neuroscience* **2011**, *42*, 24 – 28. doi:10.1177/155005941104200107.
17. Bandt, C. A new kind of permutation entropy used to classify sleep stages from invisible EEG microstructure. *Entropy* **2017**, *19*, 197.
18. Rahman, M.M.; Bhuiyan, M.; Hassan, A. Sleep stage classification using single-channel EOG. *Computers in biology and medicine* **2018**, *102*, 211–220. doi:10.1016/j.combiomed.2018.08.022.
19. Zhang, Z.; Wei, S.; Zhu, G.; Liu, F.; Li, Y.; Dong, X.L.; Liu, C.; Liu, F. Efficient sleep classification based on entropy features and a support vector machine classifier. *Physiological Measurement* **2018**, *39*. doi:10.1088/1361-6579/aae943.
20. Kemp, B.; Zwinderman, A.; Tuk, B.; Kamphuisen, H.; Oberyé, J. Sleep-edf database expanded. *Physionet.org* **2018**.
21. Chen, Z.; Ma, X.; Fu, J.; Li, Y. Ensemble Improved Permutation Entropy: A New Approach for Time Series Analysis. *Entropy* **2023**, *25*. doi:10.3390/e25081175.
22. Loh, H.W.; Ooi, C.P.; Vicnesh, J.; Oh, S.L.; Faust, O.; Gertych, A.; Acharya, U.R. Automated detection of sleep stages using deep learning techniques: A systematic review of the last decade (2010–2020). *Applied Sciences* **2020**, *10*, 8963.
23. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Physical review letters* **2002**, *88*, 174102.
24. Rosso, O.A.; Larrondo, H.; Martin, M.T.; Plastino, A.; Fuentes, M.A. Distinguishing noise from chaos. *Physical review letters* **2007**, *99*, 154102.
25. Fadlallah, B.; Chen, B.; Keil, A.; Principe, J. Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information. *Physical review. E, Statistical, nonlinear, and soft matter physics* **2013**, *87*, 022911. doi:10.1103/PhysRevE.87.022911.
26. Stosic, D.; Stosic, D.; Stosic, T.; Stosic, B. Generalized weighted permutation entropy. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **2022**, *32*, 103105, [https://pubs.aip.org/aip/cha/article-pdf/doi/10.1063/5.0107427/19804706/103105_1_online.pdf]. doi:10.1063/5.0107427.
27. Mateos, D.M.; Gómez-Ramírez, J.; Rosso, O.A. Using time causal quantifiers to characterize sleep stages. *Chaos, Solitons & Fractals* **2021**, *146*, 110798.

28. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E.; Louppe, G. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2012**, *12*.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.