Improved ERD Detection of EEG Sensorimotor Rhythms through Wavelet Transform

Alejandro Quiroga[1[0000-0002-3657-1142]], Diana Vértiz del Valle[1[0000-0002-7332-8248]], Katherine Tschopp[1[0009-0000-8098-2590]], Leonardo Rufiner [2,3 [0000-0002-5967-9172]] and Rubén Acevedo[1[0000-0002-2057-2925]]

[1] Center of Rehabilitation Engineering and Neuromuscular Research, Faculty of Engineering, National University of Entre Ríos, Entre Ríos, Argentina
[2] Institute for Signals, Systems and Computational Intelligence, *sinc(i),* UNL-CONICET, Ciudad Universitaria UNL, 4th floor, FICH, Santa Fe, Argentina.
[3] Cybernetics Laboratory, Faculty of Engineering, National University of Entre Ríos, Entre Ríos, Argentina
alejandro.quiroga@uner.edu.ar

Abstract. Brain-computer interfaces are a novel tool to implement neurorehabilitation therapies in people with motor disabilities. One of the most used paradigms in neurorehabilitation is the one based on the electroencephalogram. During the execution or attempted execution of a movement, a decrease in sensorimotor rhythms occurs in the contralateral hemisphere known as event-related desynchronization (ERD). Power spectral density is widely used in the literature to detect ERD, under the assumption that SMRs are rhythmically sustained oscillations. A recent theory suggests that neural oscillations can be represented as rhythmically sustained oscillations with dynamic amplitude or also as bursts without underlying rhythmicity. This allows the use of the wavelet transform, in particular the discrete dyadic wavelet transform (DDWT), which has a representation through compact support functions that allows highlighting localized frequency characteristics of a signal. In this work, the performance of different DDWT-based feature extraction strategies and denoising techniques were compared in order to improve the performance of ERD detection of SMR. The DDWT with the bior2.8 wavelet and a polynomial SVM classifier yielded the best performance, achieving a high true positive rate. However, the overall accuracy did not match the favorable results. To address this limitation, future research incorporating data augmentation techniques and feature selection algorithms are proposed to reduce the dimensionality of the data.

Keywords: Wavelet transform, DDWT, BCI, ERD.

1 Introduction

Brain Computer Interfaces (BCI) are a novel tool to implement neurorehabilitation therapies in people with motor disabilities. By decoding brain activity, BCIs can interpret user intentions and generate corresponding outputs [1]. One of the most used

paradigms in neurorehabilitation is the one based on electroencephalogram (EEG), whose recording is done in the central area located on the sensorimotor cerebral cortex. During the execution or attempted execution of a movement, there is a decrease in sensorimotor rhythms (SMR) in the contralateral hemisphere known as event related desynchronization (ERD).

Power Spectral Density (PSD) is widely used in the literature for detecting ERD of SMRs [2], this assumes that the SMRs are rhythmically sustained oscillations. However, a recent theory has emerged suggesting that neural oscillations, which include SMRs, can also be represented as rhythmically sustained oscillation with amplitude dynamics and as burst-events with no underlying rhythmicity, see Fig. 1. The "bursting" interpretation comes with far-reaching implications, but its importance depends on its being an accurate reflection of physiology measures [3].

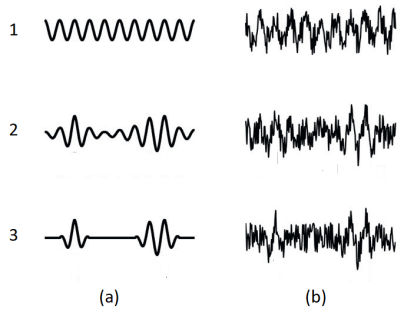

Fig. 1. Types of neural oscillations: (1) Rhythmically sustained oscillation without amplitude dynamics, (2) Rhythmically sustained oscillation with amplitude dynamics, (3) Burst-events with no underlying rhythmicity. (a) without noise and (b) with noise. Adapted from [3].

This makes us suppose that a representation of the signal using elements of short duration, and with a defined temporal location, would allow a better representation of the signal. Wavelet representation has a compact support that allows highlighting localized characteristics of a signal, such as those shown in Fig. 1 (2.a) and (3.a) in the time-frequency plane. This uses windows of different sizes, so that high frequencies are evaluated in the shorter window and low frequencies in the longer window. Therefore, it provides a flexible framework, from which it is possible to compactly represent different characteristics of the signal. [4]. In particular, the discrete dyadic wavelet transform (DDWT) is one of the most commonly used methods to generate orthogonal bases from the wavelet transform, due to its simple and inexpensive computational implementation.

The aim of this work was to analyze and compare extracting features strategies based on DDWT, using different wavelet mother functions and denoising techniques, in order to improve the performance of ERD detection of SMRs.

2 Methods

2.1 EEG dataset

The dataset used in this work was obtained from the Center for Neuromuscular and Sensory Rehabilitation and Research Engineering (CIRINS) at the Faculty of Engineering of the National University of Entre Ríos. The dataset comprised EEG signals from six volunteers without neurological or cognitive sequelae. The signals were recorded using the IM-tention software with a sampling frequency of 250 Hz and five recording channels in a monopolar configuration [5]. The electrodes were located at C3, Pz, C4, Fz and Cz; the ground and reference electrodes were placed at A1 and A2 respectively. For EEG signals preprocessing, a 2nd order bandpass Butterworth filter (1-40 Hz) and a notch filter to reject power line frequency of 50 Hz were used. To emphasize localized activity on the Cz electrode, a Laplacian spatial filter was used[1].

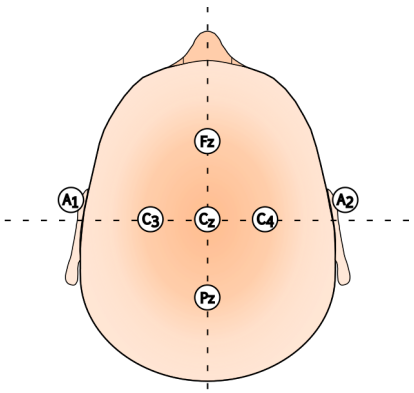

Fig. 2. Electrodes used in the EEG dataset.

Considering the stages needed in order to use a BCI, records were obtained in the calibration stage (calibration recordings) and in the closed-loop stage (online recordings). In the calibration recordings, visual cues (arrows presented on a monitor) were used to indicate which foot the volunteer should move (right or left) as well as when it should be at rest (pause sign). These visual instructions were randomly repeated 10 times for each foot during each series of recordings. Three series of EEG recordings were conducted for each volunteer. Then, temporal patterns were formed by segmenting the EEG signals using temporal marks that identified the appearance time of the visual cue. This process defined intervals corresponding to movement and rest, as illustrated in Fig. 3. The 500 msec following the visual cue were discarded, and the subsequent 2 s were considered as the interval during which the subject performed the movement. Similarly, the 500 msec preceding the cue were discarded and the 2 s prior were considered as the rest interval.

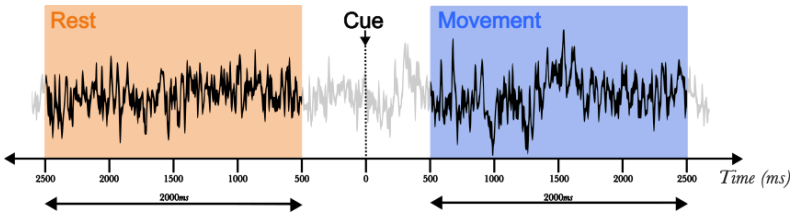

Fig. 3. EEG signal segmentation.

In the case of online recordings, three series were conducted, each consisting of 10 movements of the dominant foot and 10 rest periods. It is important to note that only actual foot movements were performed, with no attempted movements, in order to ensure the manifestation of the ERD, as the objective of this work is to evaluate the ERD detection.

2.2 Features extraction strategies

This section describes the feature extraction strategies evaluated in this work.

2.2.1 Power Spectral Density

Since ERD is a power decrease of SMR, the power spectral density (PSD) of temporal patterns was computed. There are different approaches to estimate the PSD and in this work, Welch's method for PSD estimation was employed. This approach divides the signal into overlapping windows, estimates the periodogram for each window, and averages them to obtain the PSD [6].

In the calibration stage, the PSD of the temporal patterns was calculated using the Welch method with 1Hz resolution and 3 Hamming windows of 1 sec (50% overlapping). This process resulted in two sets of 23 features (referred to as feature vectors), including only the frequencies in the 8-30 Hz range which correspond to SMR. During the Closed-loop stage, a single feature vector is extracted only from the movement interval.

2.2.2 Dyadic discrete wavelet transform

The wavelet transform is an important tool for signal processing, as it allows the representation of signals in the time-frequency plane and provides detailed analysis at both high and low frequencies (multiresolution analysis) as well as good response when dealing with nonstationary signals [7]. The wavelet transform is achieved by calculating the inner product between the signal of interest and the wavelet function (ϕ) at a scale and translation, determined by the scale function (ψ). This process yields coefficients corresponding to an orthogonal base which represents the original signal into different resolution levels.

In this work, the dyadic discrete wavelet transform (DDWT) was used, with a scaling factor of 2, resulting in a more efficient transform compared to the continuous-time wavelet transform. This is because the DDWT produces fewer

coefficients and reduces redundant information. The state of the art analysis brings a number of wavelet functions used in common EEG feature extraction problems [8]. Considering the similarity between the morphology of the wavelet functions and the bursts mentioned earlier, the following families of wavelet functions were chosen: Daubechies (db4, db6, db10, db13, db14 and db15), Biorthogonal (bior2.4, bior2.8, bior3.1, bior5.5 and bior6.8), Coiflet (coif5) and Symlet (sym5). Fig. 4 shows an example for each of the selected families.

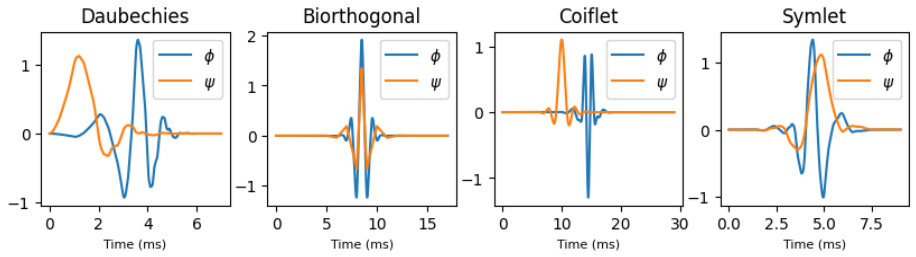

Fig. 4. Examples of Daubechies, Biorthogonal, Coiflet and Symlet wavelet families.

The DDWT algorithm implementation involves a tree decomposition (see Fig. 5) using a filter bank approach. At each decomposition level, a low-pass filter and a high-pass filter are applied to extract a set of coefficients known as approximation (A) and detail (D), respectively. Dyadic scaling allows to reduce the number of the coefficients of the previous level by half and enables the representation of specific frequencies using any selected coefficients.

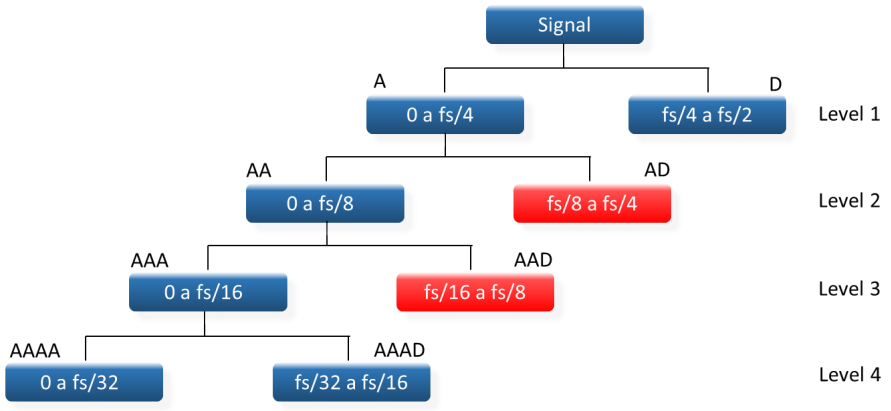

Fig. 5. DDWT tree decomposition.

As previously mentioned, the sample frequency was fs=250 Hz, resulting in a maximum signal frequency of 125 Hz. Based on this, at level 1, the decomposition consists of A and D, representing frequencies from 0 Hz to 62.5 Hz and 62.5 Hz to 125 Hz respectively. The level 2 consists of AA and AD, representing frequencies from 0 Hz to 31.25 Hz and AD 31.25 Hz to 62.5 Hz respectively. Continuing this pattern, AAA and AAD represent frequencies from 0 Hz to 15.625 Hz and 15.625 Hz to 31.25 Hz respectively at level 3. Finally, on level 4 AAAA represents frequencies from 0 Hz to 7.81 Hz and AAAD represents frequencies from 7.81 Hz to 15.625 Hz.

To focus on the frequency range of interest for SMRs (8-30 Hz), a denoising scheme was applied. Only the coefficients corresponding to AAAD and AAD (7.81 Hz to 31.2 Hz) were used. These coefficients are then concatenated to form the feature patterns, as shown in blue in Fig. 5.

2.3 Classifiers

According to [9], Fisher's linear discriminant analysis (LDA) and support vector machine (SVM) are suitable classifiers for studying the ERD phenomenon. Therefore, in this work, both LDA and SVM classifiers were implemented and compared.

2.3.1 Linear discriminant analysis

Fisher's linear discriminant is a linear classifier with easy implementation and low computational cost. It assumes that the classes are normally distributed with identical covariance (homoscedasticity assumption). Though the LDA classifier imposes very strong assumptions on the distribution of the data, the computation of the discriminative function is very efficient, that's why it has been popular in the BCI field [10].

The LDA, like any binary linear classifier, can be characterized by the Eq. (1):

$$g(z) = w^T.z + b \quad (1)$$

where $w = [w_1, ..., w_K]$ is the projection vector, $z \in R^K$ represents the input vector and b is the bias term. The classification function assigns the class label C_i to each pattern z depending on the sign of the function g(z). It is assumed that the probability distributions of each class follow a Gaussian distribution.

2.3.1 Support vector machine

One of the widely used classifiers in BCI for various applications is the Support Vector Machine (SVM), a kernel-based classifier [11]. For the specific problem addressed in this work, which involves 2 classes, SVM finds a hyperplane that separates the classes. This process involves projecting the data into a high-dimensional space, where the classes could be linearly separable. In cases where linear separability cannot be achieved, the use of appropriate kernel functions becomes necessary. Although different kernel functions were evaluated, this work presents the results obtained using the linear and polynomial kernels (Eq. 2), as they demonstrated the highest performance.

$$(< x, y > + c)^d, \; c \in \Re, \; d \in \aleph \quad (2)$$

In the SVM training process, grid-search and cross-validation techniques were employed to optimize the classifier's performance. Grid-search involved varying the values of key parameters, such as C, gamma (γ), and the polynomial degree. For the linear kernel, the parameter C determines the trade-off between misclassification and maximizing the margin. In the case of the polynomial kernel, γ controls the influence of individual training samples and the polynomial degree sets the degree of the polynomial kernel function.

Cross-validation was used to evaluate the performance of the SVM with a linear kernel across different C values and the SVM with a polynomial kernel using different parameter combinations. The selection of the optimal parameter set was based on the accuracy metric. In this work, the chosen parameters ranges were as follows: C values ranged from 10^{-2} to 10^{10}, γ values ranged from 10^{-9} to 10^{3} and the polynomial degree values were set to 2 and 3. These ranges were selected based on prior knowledge [12] and experimentation. Using grid-search and evaluating performance through cross-validation, the SVM classifier was fine-tuned to achieve the highest accuracy result.

2.4 Performance metrics

In the calibration stage, the performance of the classifier was evaluated using the Accuracy (Acc_{Cal}) metric, the feature vectors obtained in this stage were used. The objective of this metric is to estimate the effectiveness of the calibration process. In the closed-loop stage, the Accuracy (*Acc)* and True Positives Rate (*TPR*) were employed. These metrics can be calculated using the equations (3) and (4), where *TP* represents true positives, *TN* represents true negatives, *FP* represents false positives, and *FN* represents false negatives.

$$Acc\ [\%] = \frac{TP+TN}{TP+TN+FP+FN}.100 \tag{3}$$

$$TPR\ [\%] = \frac{TP}{TP+FN}.100 \tag{4}$$

In the context of neurorehabilitation, reporting the TPR is crucial because the BCI is active only during the execution or attempted execution of a movement. The rest of the time, the BCI remains inactive, which means that only the class related to the movement is available.

3. Results

To select the best wavelet functions within each family, the TPR at the close-loop was used instead of *Acc* due to the minimal variability between wavelet functions. Therefore, the TPR was analyzed using the three classifiers: LDA, SVM with linear kernel and SVM with polynomial kernel. In the case of the Daubechies family, the db6 wavelet function achieved the highest rate, as can be seen in Table 1; for the

Biorthogonal family, the bior2.8 function demonstrated the highest rate, as shown in table 2. This selection was not necessary for Coiflet and Symlet families, since only one function per family was considered.

Table 1. Daubechies family TRP results

Wavelet function	*Median TPR*		
	LDA	**kernel = Linear**	**kernel = Poly**
db4	50.00	50.50	61.11
db6	63.33	55.00	85.00
db10	55.00	51.11	57.77
db13	60.55	54.17	50.00
db14	55.55	49.16	66.11
db15	55.55	50.00	65.55

Table 2. Biorthogonal family TRP results

Wavelet function	*Median TPR*		
	LDA	**kernel = Linear**	**kernel = Poly**
bior2.4	61.66	50.55	91.11
bior2.8	66.70	58.30	96.70
bior3.1	53.00	57.78	92.22
bior5.5	57.22	53.33	75.55
bior6.8	56.66	56.67	79.45

The Fig. 6 to 8 shows the results comparison between the best wavelets functions and the PSD, using the latter as a reference.

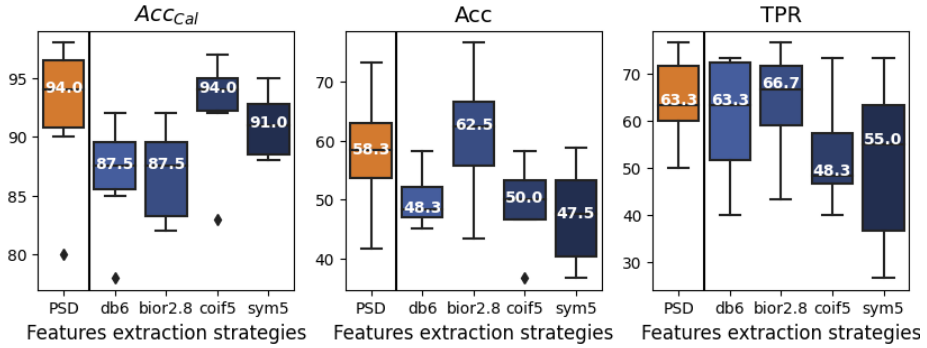

Fig. 6. Performance metrics using LDA.

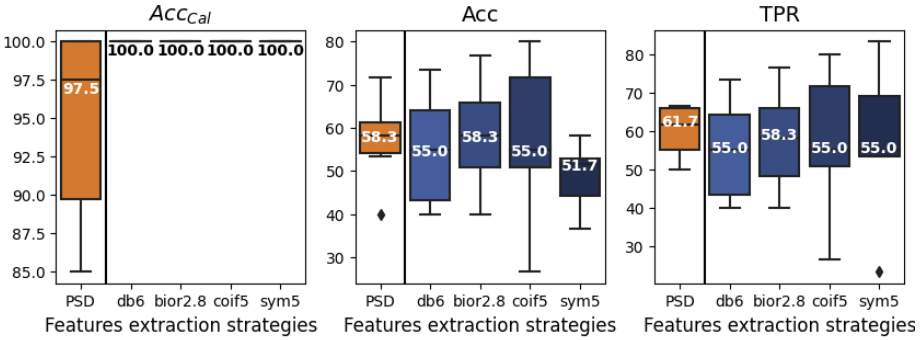

Fig. 7. Performance metrics using linear SVM.

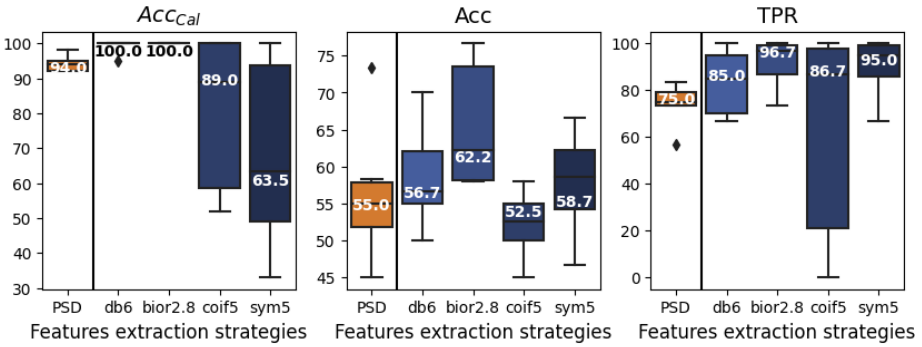

Fig. 8. Performance metrics using SVM with polynomial kernel.

By analyzing the previous figures, it was found that higher *TPR* is obtained by using bior2.8 wavelet function, the best case being the use in combination with the SVM with polynomial kernel. Fig. 9 shows a comparison with the reference (PSD).

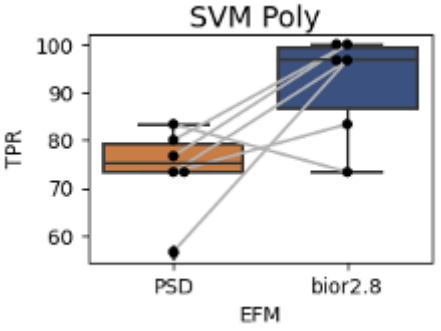

Fig. 9. Statistical analysis for PSD and bior2.8

A repeated measures ANOVA test (Anova-RM) was used to assess differences in TPR between bior2.8-SVM polynomial and the PSD-LDA combinations , the results are shown in table 3

Table 3. Anova-RM analysis for PSD and bior2.8

	F Value	**Num DF**	**Den DF**	**Pr > F**
Features extraction strategies	6.8088	1.0000	5.0000	0.0477

3 Conclusions

This paper presented a comparison of feature extraction strategies based of DDWT to detect ERD of sensorimotor rhythms for potential use in a BCI for neurorehabilitation.

According to the results obtained, the combination of bior2.8 wavelet function and SVM polynomial outperformed the other alternatives in terms of the *TPR* (96,7 %). Statistical analysis revealed a significant difference between bior2.8-SVM polynomial and the PSD-LDA combination, as indicated by a p-value below 0.05. However, it is important to keep in mind that *TPR* cannot be the only metric to draw a definitive conclusion, as high *TPR* may bias the results. This is clear, since the *Acc* for this combination is 62.2%, this implies that in this case the classifier has high sensitivity and low specificity. Future work should aim to improve the performance of the classifier trying to increase the *Acc* and consequently the specificity.

On the other hand, using wavelet functions as a feature extraction strategy can improve classification metrics due to the relationship between the burst events theory and the morphology of wavelet functions. This is evidenced through their inner

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (sinc.unl.edu.ar)
A. Quiroga, D. Vértiz, K. Tschopp, H. L. Rufiner & R. C. Acevedo; "Improved ERD Detection of EEG Sensorimotor Rhythms through Wavelet Transform"
Anales del XXVI Congreso Argentino de Bioingeniería, SABI, 2023.

product, as demonstrated by the similarity between the signal in Fig. 1 (2.a) and (3.a) and the wavelets functions in Fig. 4.

Another aspect to consider when analyzing the results is the limited amount of data available to train the classifiers (30 samples per class). This may have contributed to the challenges in achieving a good generalization by the strategies presented in this paper. To address this limitation in future works, it is proposed to employ data augmentation techniques as well as feature selection algorithms to reduce the dimensionality of the data.

Acknowledgments

This work was funded by the PID-UNER #6235 *"Application of machine learning to biomedical problems in the context of sparse data"* and CAID-UNL 50620190100151LI *"Algoritmos inteligentes profundos para análisis y clasificación de bioseñales"*.

References

1. Wolpaw JR, Wolpaw EW, editors. Brain-computer interfaces: principles and practice. Oxford ; New York: Oxford University Press; 2012. 400 p.
2. Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, et al. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. J Neural Eng. 2018 Jun 1;15(3):031005.
3. van Ede F, Quinn AJ, Woolrich MW, Nobre AC. Neural Oscillations: Sustained Rhythms or Transient Burst-Events? Trends Neurosci. 2018 Jul 1;41(7):415–7.
4. Acevedo R, Atum Y, Gareis I, Biurrun Manresa J, Medina Bañuelos V, Rufiner L. A comparison of feature extraction strategies using wavelet dictionaries and feature selection methods for single trial P300-based BCI. Med Biol Eng Comput. 2019 Mar;57(3):589–600.
5. Vértiz del Valle D, Carrere C, Acevedo R, Tabernig C. IM-tention: a software for brain computer interfaces with motor recovery purposes. In: IFMBE Proceedings book series. Ed. Springer; 2023.
6. Gursel Ozmen N, Gumusel L, Yang Y. A Biologically Inspired Approach to Frequency Domain Feature Extraction for EEG Classification. Comput Math Methods Med. 2018 Jan 23;2018:e9890132.
7. Rufiner HL, Goddard C J. A method of wavelet selection in phoneme recognition. In: Proceedings of 40th Midwest Symposium on Circuits and Systems Dedicated to the Memory of Professor Mac Van Valkenburg. 1997. p. 889–91 vol.2.
8. Ganorkar S, Raut V. Comparative Analysis Of Mother Wavelet Selection For EEG Signal Application To Motor Imagery Based Brain-Computer Interface. 2019;8(12).
9. Quiroga A, Vértiz del Valle D, Pilz M, Acevedo R. Performance comparison of

different classifiers to detect motor intention in EEG-based BCI. In Florianópolis - SC, Brazil: Springer Books; 2022.
10. Bashashati H, Ward RK, Birch GE, Bashashati A. Comparing Different Classifiers in Sensory Motor Brain Computer Interfaces. PLOS ONE. 2015 Jun 19;10(6):e0129435.
11. Rasheed S. A Review of the Role of Machine Learning Techniques towards Brain–Computer Interface Applications. Mach Learn Knowl Extr. 2021 Dec;3(4):835–62.
12. Quitadamo LR, Cavrini F, Sbernini L, Riillo F, Bianchi L, Seri S, et al. Support vector machines to detect physiological patterns for EEG and EMG-based human-computer interaction: a review. J Neural Eng. 2017 Feb;14(1):011001.