Check for updates

Analysis of instantaneous brain interactions contribution to a motor imagery classification task

Jorge Humberto Cristancho Cuervo[1]*, Jaime F. Delgado Saa[2] and Lácides Antonio Ripoll Solano[3]

[1]Biomedical Signal Processing and Artificial Intelligence, Department of Electrical and Electronics Engineering, Universidad del Norte, Barranquilla, Colombia, [2]SciFork SARL, Geneva, Switzerland, [3]Grupo de Investigación en Telecomunicaciones y Señales, Department of Electrical and Electronics Engineering, Universidad del Norte, Barranquilla, Colombia

The purpose of this study is to analyze the contribution of the interactions between electrodes, measured either as correlation or as Jaccard distance, to the classification of two actions in a motor imagery paradigm, namely, left-hand movement and right-hand movement. The analysis is performed in two classifier models, namely, a static (linear discriminant analysis, LDA) model and a dynamic (hidden conditional random field, HCRF) model. The impact of using the sliding window technique (SWT) in the static and dynamic models is also analyzed. The study proved that their combination with temporal features provides significant information to improve the classification in a two-class motor imagery task for LDA (average accuracy: 0.7192 no additional features, 0.7617 by adding correlation, 0.7606 by adding Jaccard distance; $p < 0.001$) and HCRF (average accuracy: 0.7370 no additional features, 0.7764 by adding correlation, 0.7793 by adding Jaccard distance; $p < 0.001$). Also, we showed that adding interactions between electrodes improves significantly the performance of each classifier, regarding the nature of the interaction measure or the classifier itself.

KEYWORDS

brain interactions, correlation, Jaccard distance, classifier, static model, dynamic model, sliding window

Introduction

A brain-computer interface (BCI) is a communication and control scheme that sends messages and commands to the external world by interpreting brain waveforms (Wolpaw et al., 2002; Abiri et al., 2019). A regular BCI system has a brain monitoring system, a signal preprocessing stage, a stage for extracting features of the preprocessed signal, and a classification stage where features are decoded into commands or messages. Some manners of BCI activation are evoked potentials, brain rhythms, and motor imagery, among others (Scherer et al., 2007; Yao et al., 2022).

Different studies have suggested that combining frequency-temporal features with blinding source separation techniques improves the performance of classifiers in motor imagery paradigms significantly. One of the most outstanding techniques is the common spatial patterns (CSPs) that extract mutual features from a mixture of two populations by maximizing the different proportions of the variances in each population (Koles et al., 1990; Wang et al., 2021). In this way, a linear transformation (spatial filtering) is performed, preserving the number of sources but missing any interactions between electrodes. To compensate for the lack of information, a series of improvements have been introduced to CSPs, such as the filter bank CSP (FBCSP). In this study, the collected brain signal is processed by a set of different frequency band filters, and each filtered signal is processed by a CSP (Ang et al., 2008, 2012; Ferrero et al., 2021). Therefore, the contribution of a series of frequency bands is preserved while CSP extracts mutual features by the band. CSP and FBCSP can be applied over raw or preprocessed data, such as filtered data extracted from wavelet packet decomposition (Luo et al., 2019, 2020; Voinas et al., 2022), recombined data from CSP (Jalilpour Monesi and Hajipour Sardouie, 2019), or frequency data extracted from CSP (Oikonomou et al., 2020).

However, since CSP and FBCSP come from a linear transformation, the resultant operation only shows the projection of each electrode to the new space, such as its contribution to the total variance of the chosen population concerning the joint population (Bezdek and Pal, 1995; Wang et al., 2021). Furthermore, some recent studies have proposed to add interactions between electrodes—as correlation—to improve the CSP projection rather than being an input feature to a classifier by themselves (Zhang et al., 2013; Gubert et al., 2020; Ghanbar et al., 2021). Hence, the statistical contribution of these interactions is implicit as a CSP improvement and not as a feature. Other features, such as power bands, wavelet coefficients, and auto-regressive models, do not ever use any interactions between electrodes or brain zones (Aggarwal and Chugh, 2019; Mohdiwale et al., 2021).

The purpose of this study is to analyze statistically the contribution of the interactions between electrodes, measured here as correlation or Jaccard distance, as an additional input feature to the classification of two actions in a motor imagery paradigm, namely, right-hand movement or left-hand movement. The analysis uses two classifier models, namely, a static model and a dynamic model. The static model consists of a linear discriminant analysis classifier (LDA). The dynamic model consists of a linear conditional random fields (CRF) model, where features only interact with hidden variables rather than interacting with labels, named hidden CRF (HCRF). Also, the impact of using the sliding window technique (SWT) in the static and dynamic models is analyzed in this study.

Methods

Experiment and dataset description

The used dataset comes from the BCI Competition IV, dataset 2b (Leeb et al., 2007; Tangermann et al., 2012). Nine naïve volunteer subjects participated in the experiment. Each one was right-handed and had normal or corrected-to-normal vision. Each subject sat in an armchair and watched a flat screen placed 1 m away at eye level. Five sessions were performed for each subject: the first two without feedback and the last three with feedback. Each session consists of several runs preceded by 5 min of electrooculography (EOG) estimation at the beginning of each session.

The first two sessions used the following paradigm: a cue-based screening paradigm consisting of two classes, namely, left-hand movement and right-hand movement. Each session consisted of six runs, and each run had ten randomized trials by class, summarizing 120 repetitions per session. Each trial started with a fixation cross and a warning tone by 1 s approximately, followed by an arrow indicating either the left side or the right side for 1.25 s. Subsequently, the subject imagined the hand movement for 4 s. Next, there was a randomized pause for about 1.5 s to avoid adaptation.

The later three sessions made use of a smiley face for feedback, with four runs and 20 randomized trials by class and run, for a total of 160 repetitions per session. Each trial started with a gray smiley and a warning tone for ∼1 s, followed by a cue period of 3 s, where the smiley displaced to the left or right. In this time interval, depending on the hand movement, the smiley became red when the subject was wrong and green when the moved hand was correct. Moreover, the mouth of the smiley also changed to sad (corners of the mouth downward) or happy (corners of the mouth upward), with wrong or right movement, respectively. Next, there was a randomized pause between 1 and 2 s. We employed data only from feedback sessions.

Waveforms were recorded from three EEG bipolar electrodes, namely, C3, Cz, and C4, with a frequency sample of 250 Hz. Fz channel was used as EEG ground. Later, data were bandpass filtered between 0.5 and 100 Hz, followed by a 50 Hz notch filter. In addition, the EOG is available from three monopolar electrodes and a similar amplification configuration. Additional details of the experiment are available in Leeb et al. (2007).

Software implemented

Once the dataset was acquired, preprocessing, feature extraction, and classification stages were programmed in MATLAB® 2021. HCRF was implemented by the Hidden-state Conditional Random Field Library version 2.0b (HCRF Library, 2011).

Data preprocessing

During the imagery task, relevant information was collected between 3 and 6 s after the beginning of a trial. Then, the waveform was sub-segmented by sliding windows, varying the window size and the slide size along the experiments.

A single sub-segment can provide up to three types of features, namely, alpha/mu (8–12 Hz) and beta (15–25 Hz) power bands (Singh et al., 2011a,b, 2013; Pfurtscheller and McFarland, 2012) by each electrode, and a similarity (or distance) measure between a pair of electrode signals (C3–Cz, C4–Cz, or C3–C4). Then, power band data got scaled to an order of magnitude 10. Experiments used alpha and beta power bands as main features and similarity (or distance) measures as additional features. All features were obtained by using the SWT.

Sliding window technique

The sliding window technique is a set of instructions executed over a subset of k consecutive values of $\mathbf{X}$, being $\mathbf{X} = \{x_1, x_2, \ldots, x_{N-1}, x_N\}$, a discrete time series arrangement for N equally spaced time samples whose initial point is x_i: $\mathbf{X_{i,k}} = \{x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k}\}$. Once the set of instructions was performed, the position of the initial point displaces by a distance Δi and the algorithm takes another k points to set the new subset $\mathbf{X_{i+\Delta i,k}}$ $= \{x_{i+\Delta i}, x_{i+\Delta i+1}, \ldots, x_{i+\Delta i+k-1}, x_{i+\Delta i+k}\}$. The instructions process the new subset. The routine continues until the value of $\mathbf{X}$ corresponding to the final point x_N is reached. The size of the subset k and the displacement of initial point Δi could be predetermined (Bandettini et al., 1993) or dynamically adaptive (BenYahmed et al., 2015).

The sliding window technique is convenient for data to get simple representations (BenYahmed et al., 2015; Hota et al., 2017) or find dynamic patterns (Mokhtari et al., 2019; Vergara et al., 2019) in a time series set. In this way, we propose to apply the SWT to increase the number of features obtained by trial: alpha and beta power values and the similarity (or distance) measures.

Linear discriminant analysis

A discriminant classifier is a function that allocates an input vector x to one of K classes $\mathcal{C}_k$ (Bishop, 2013). If we assume that $f_k(\mathbf{x})$ is a multivariate Gaussian with a vector of mean μ_k and a covariance matrix Σ_k, we get the following discriminant function $\delta_k(\mathbf{x})$ (Hastie et al., 2009):

$$\delta_k(\mathbf{x}) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(\mathbf{x} - \mu_k)^T \Sigma_k^{-1}(\mathbf{x} - \mu_k) + \log\pi_k$$

If each class has its covariance matrix Σ_k, the discriminant function is quadratic by making the decision boundary between a pair of classes k and l as $\delta_k(\mathbf{x}) = \delta_l(\mathbf{x})$. However, if we suppose a shared covariance matrix Σ for all classes, the discriminant $\delta_k(\mathbf{x})$ becomes linear (Hastie et al., 2009; Bishop, 2013). Notice that LDA-based classifiers are generative since they mostly assume Gaussian distributions in the data (Martens et al., 2011) and that LDA is a static model since time is not a relevant parameter for the classifier.

Hidden conditional random fields

The conditional random field (CRF) classifier is a member of the probabilistic graphical models (PGMs) family. CRF represents complex distributions through products of local factors on small subsets of variables (Sutton and McCallum, 2011). Unlike other PGM models, CRF does not take the dependencies among entries but models directly the conditional distribution between input vectors x and output labels $\mathbf{y}$ as $p(\mathbf{y}|\mathbf{x})$.

To expand the CRF models, we could add a set of hidden variables $\mathbf{h} = \{h_1, h_2, \ldots, h_m\}$ not observed during the training stage, associated with a singular output label y (Quattoni et al., 2007). Restricting the model to have disjoint sets of hidden states related to a certain value y of output label $h_j \in \mathcal{H}_y$, a hidden conditional random field (HCRF) takes the following form:

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z}\sum_{h_j, h_{j'} \in \mathcal{H}_y}\prod_{t=1}^{T}\exp\left\{\sum_{k=1}^{K}\theta_k f_k\left(h_{j,t}, h_{j',t-1}, \mathbf{x}_t\right)\right\}$$

The partition function is a normalization function along all possible values of hidden variables along y. As CRF, a hidden variable $h_{j,t}$ depends only on its predecessor $h_{j',t-1}$ and the corresponding input variables.

Pearson correlation

Pearson correlation, defined for two zero-mean and real-valued random variables x, y, is the coefficient between the cross-correlation of the random variables $E[xy]$ and the product of the square root of their variances $\sigma_x\sigma_y$ (Benesty et al., 2008, 2009):

$$\rho(x, y) = \frac{E[xy]}{\sigma_x\sigma_y}$$

We remark that the purpose of the study is to analyze statistically the contribution of the interactions between electrodes. In this study, we use correlation values to quantify the interaction between electrodes as a similarity measure.

Jaccard distance

On the contrary, the Jaccard distance comes from its counterpart, the Jaccard Index J. The latter is a similarity measure for two sets A, B, defined as the coefficient between the size of the intersection $|A \cap B|$ and the size of its union $|A \cup B|$ (Fletcher and Isla, 2018). It can extend as the ratio between the measure of the intersection $\mu (A \cap B)$ and its union $\mu (A \cup B)$, with an arbitrary measure μ. If we define μ as the dot product between two multivariate variables A, B: $\mu (A \cap B) = A \bullet B$, and $\mu (A) = \|A\|^2$, being $\|A\|$ the Euclidean norm of A, and using the relationship between the intersection and the union of two sets, the Jaccard Index J becomes:

$$J (A, B) = \frac{A \bullet B}{\|A\|^2 + \|B\|^2 - A \bullet B} = \frac{A \bullet B}{\|A - B\|^2 + A \bullet B}$$

With this definition, Jaccard Index takes values between $-1/3$ (when $B = -A$) and 1 (for $B = A$). However, for obtaining a non-negative metric, the Jaccard distance $J_D = 1 - J$ is defined as follows (Cha, 2007):

$$J_D (A, B) = \frac{\|A - B\|^2}{\|A\|^2 + \|B\|^2 - A \bullet B} = \frac{\|A - B\|^2}{\|A - B\|^2 + A \bullet B}$$

With this definition, Jaccard distance takes values between 0 and 4/3, becoming a non-negative measure. In this study, we use Jaccard distance values to quantify the interaction between electrodes as a distance measure.

Performance metrics

A recurrent measure of performance for classification is *accuracy*. It estimates the closeness between measured or predicted values and their actual values (Clifford, 1985). For multiple classes, it is defined as the rate between the trace of a confusion matrix H and the total number of samples N_s (Schlögl et al., 2005):

$$p_0 = \frac{trace (H)}{N_s}$$

where trace (H) is the number of samples correctly classified. The accuracy varies from 0 to 1, where 1 denotes a perfect classification. The study used five rounds of 3-fold cross-validation, where training data tune LDA and HCRF model parameters by an intern 4-fold cross-validation.

Statistical analysis

A one-way randomized blocks ANOVA tested the statistical significance of differences between data with and without additional features. If the ANOVA test rejects the null hypothesis

TABLE 1 Results of accuracy in the LDA model, with the presence or absence of interactions between electrodes.

Subject	Type of interaction between electrodes			
	Correlation added		Jaccard distance added	
	Yes	No	Yes	No
1	0.757	0.714	0.743	0.704
2	0.569	0.527	0.565	0.526
3	0.623	0.580	0.604	0.565
4	0.975	0.933	0.976	0.937
5	0.798	0.756	0.821	0.782
6	0.758	0.715	0.764	0.725
7	0.712	0.669	0.707	0.668
8	0.902	0.859	0.898	0.859
9	0.762	0.720	0.767	0.728
Average	0.762	0.719	0.761	0.722
Standard deviation	0.125	0.125	0.130	0.130
P-value	<0.001	1	<0.001	1
Overall performance	0.74		0.741	

of statistical equality of averages, a Tukey-Kramer test performs a *post-hoc* comparison. The average value of each classifier and type of data is compared against the overall average accuracy, rejecting the null hypothesis if the average by class is greater than the overall average.

Also, a linear multiple-way randomized blocks ANOVA tested the statistical significance of differences in the data. The window size and the slide size were the tested parameters, and subjects were taken as randomized blocks in the model. If the ANOVA test rejects the null hypothesis of statistical equality of averages, a Tukey-Kramer test performs a *post-hoc* comparison.

Results

The results of this study refer to the average performance obtained by each classifier, measured with the accuracy metric. All measures were obtained from the testing dataset of each subject.

Results with the LDA model

Table 1 shows the accuracies obtained by comparing data with and without correlation features with the LDA classifier model. The 3-s trial was used without modifying other parameters. According to the ANOVA test [F: 34.922; degree of freedom (d.f.): 1; $p < 0.001$], the null hypothesis of average equality must get rejected, so we performed the *post-hoc* test. Their p-values are illustrated in Table 1 showing that only data

TABLE 2 Results of accuracy in the HCRF model, with the presence or absence of interactions between electrodes.

Subject	Type of interaction between electrodes			
	Correlation added		Jaccard distance added	
	Yes	No	Yes	No
1	0.752	0.712	0.753	0.711
2	0.596	0.557	0.600	0.558
3	0.568	0.529	0.571	0.529
4	0.991	0.951	0.992	0.950
5	0.790	0.751	0.809	0.767
6	0.778	0.738	0.788	0.747
7	0.763	0.724	0.750	0.708
8	0.919	0.880	0.919	0.878
9	0.831	0.792	0.832	0.790
Average	0.776	0.737	0.779	0.738
Standard deviation	0.135	0.135	0.135	0.135
P-value	<0.001	1	<0.001	1
Overall performance	0.757		0.758	

with additional features are statistically significant regarding the average accuracy.

Table 1 also shows the accuracies obtained by comparing data with and without Jaccard distance features with the LDA classifier model. As before, a one-way randomized blocks ANOVA tested the statistical significance of differences between data with and without additional features. According to the ANOVA test (F: 26.613; d.f.: 1; $p < 0.001$), the null hypothesis of average equality must get rejected, so we performed the *post-hoc* test. This indicates again that only data with additional features are statistically significant regarding the average accuracy.

The following step is to implement the sliding window algorithm in the data with power alpha and beta bands and additional features to obtain features dynamically. Hence, we used three sliding window sizes (0.5, 1, and 2 s) against the whole 3-s window and three slide sizes (0.125, 0.25, and 0.5 s) to compare the performance of the LDA classifier.

Table 2 shows the performance of the LDA model by window size and slide size by implementing the sliding window algorithm with correlation or Jaccard distance as additional features. According to the ANOVA test (F_{window_size}: 46.61; d.f.: 3; $p_{window\_size} < 0.001$; F_{slide_size}: 144.60; d.f.: 2; $p_{slide\_size} < 0.001$), the null hypothesis of average equality must get rejected, so we performed the *post-hoc* test. Results of Table 2 show that window sizes 2 and 3 s have the most outstanding performances. Meanwhile, slides 0.25 and 0.5 have the most statistically relevant values.

Considering the Jaccard Distance as an additional parameter, a similar procedure to correlation values was performed.

According to the ANOVA test (F_{window_size}: 30.28; 3 d. f.; $p_{window\_size} < 0.001$; F_{slide_size}: 137.73; 2 d. f.; $p_{slide\_size} < 0.001$), the null hypothesis of average equality must get rejected, so we implemented the *post-hoc* test. Results of Table 2 show that window sizes 2 and 3 s have the most significant performance. Meanwhile, slides 0.25 and 0.5 have the most statistically significant values. Meanwhile, although the classifier performance is the highest when the slide size is 0.125 s, the average result is slightly better than the average performance with whole data, as illustrated in Table 2.

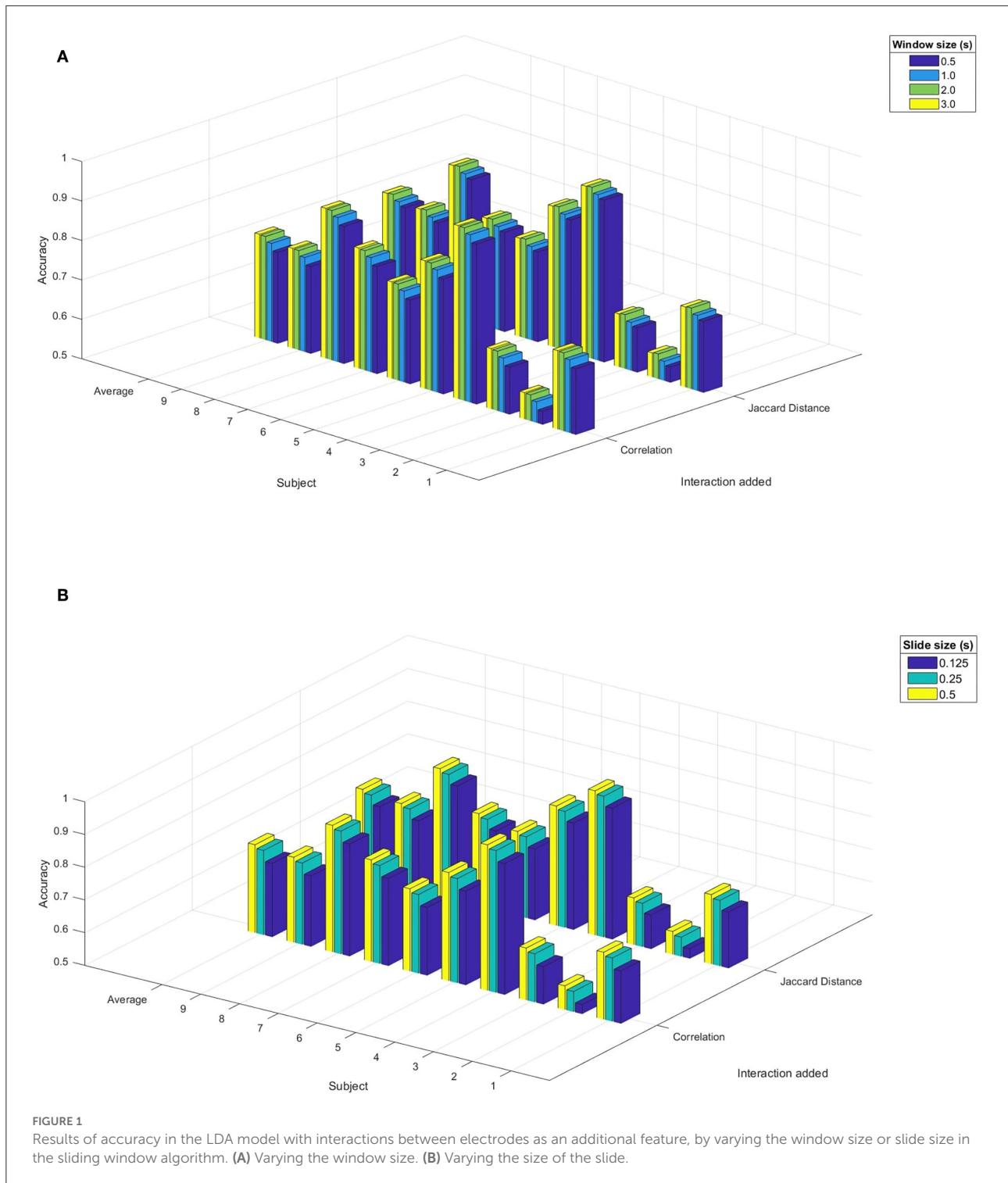## Results with the HCRF model

Since HCRF is a dynamic model regarding the LDA model, implementation of the sliding window algorithm is necessary to establish the corresponding timestamps of HCRF. To compare data with and without additional correlation features, we tested the model with two window sizes (0.5 and 2 s) and two slide sizes (0.125 and 0.5 s). Figure 1 shows the accuracies obtained by comparing data with and without correlation features implemented in the HCRF classifier. According to the ANOVA test (F: 117.232; d.f.: 1; $p < 0.001$), the null hypothesis of average equality must get rejected, so we performed the *post-hoc* test. Their p-values are illustrated in Figure 1, showing that only data with additional features are statistically significant regarding the average accuracy.

Figure 1 also shows the accuracies obtained by comparing data with and without Jaccard distance features with the HCRF classifier. As before, a one-way randomized blocks ANOVA tested the statistical significance of differences between data with and without additional features. According to the ANOVA test (F: 148.475; d.f.: 1; $p < 0.001$), the null hypothesis of average equality must get rejected, so we performed the *post-hoc* test. It indicates again that only data with additional features are statistically significant regarding the average accuracy.

The following step is to implement the sliding window algorithm in the data with power alpha and beta bands and additional features to get features dynamically. Hence, we used three sliding window sizes (0.5, 1, and 2 s) and three slide sizes (0.125, 0.25, and 0.5 s) to compare the performance of the HCRF classifier.
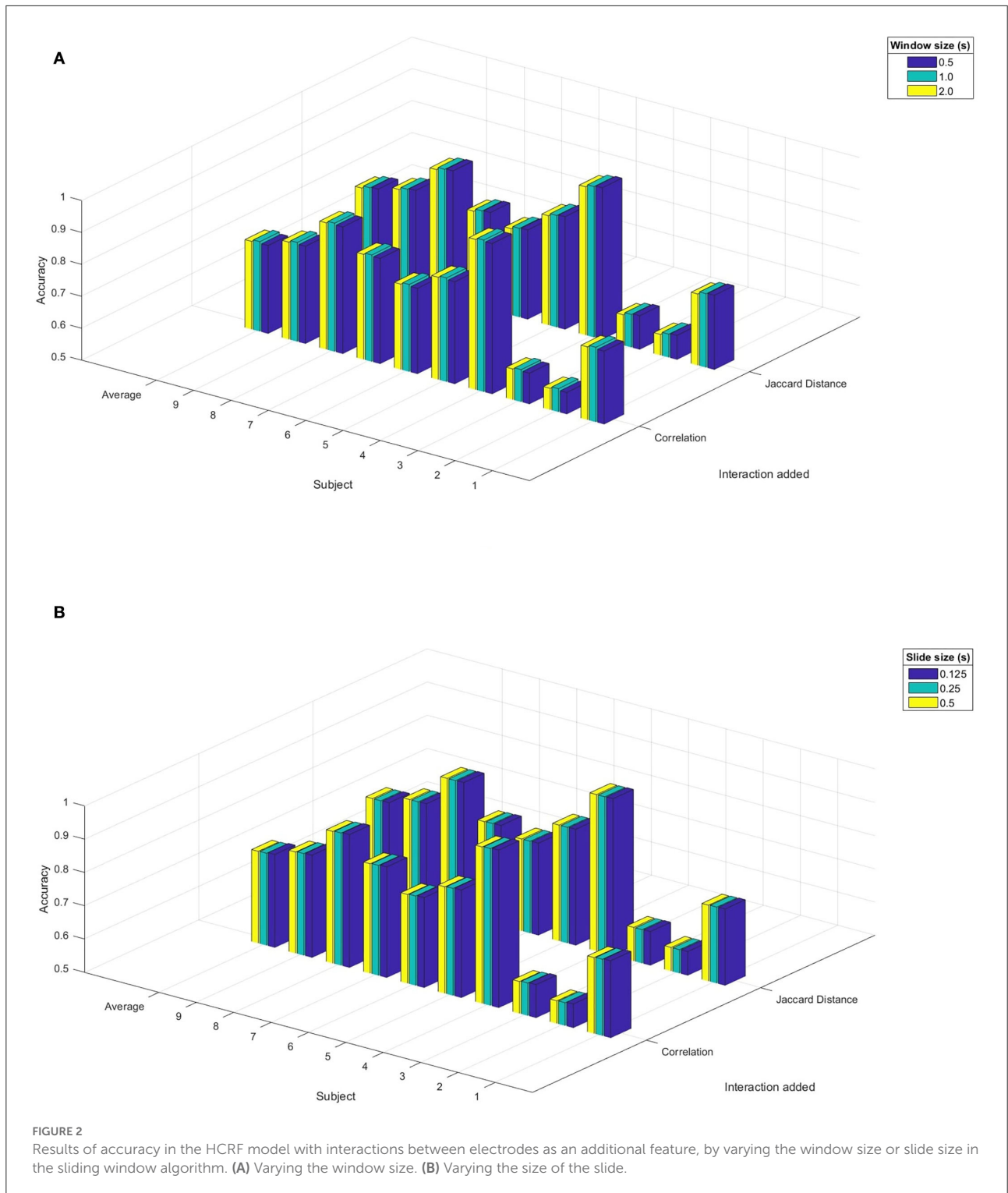
Figure 2 shows the performance of the HCRF model by window size and slide size by implementing the sliding window algorithm with correlation as additional features. According to the ANOVA test (F_{window_size}: 4.20; d.f.: 2; $p_{window\_size} = 0.016$; F_{slide_size}: 1.95; d.f.: 2; $p_{slide\_size} = 0.143$), the null hypothesis of average equality must get rejected only for the window size, so we performed the *post-hoc* test. Results of Figure 2 show that a window size of 1 s has the most significant performance.

Considering the Jaccard distance as an additional parameter, a similar procedure to correlation values was performed. According to the ANOVA test (F_{window_size}: 12.35; d.f.: 3;

FIGURE 1
Results of accuracy in the LDA model with interactions between electrodes as an additional feature, by varying the window size or slide size in the sliding window algorithm. **(A)** Varying the window size. **(B)** Varying the size of the slide.

$p_{\text{window\_size}} < 0.001$; $F_{\text{slide\_size}}$: 0.15; d.f.: 2; $p_{\text{slide\_size}} = 0.86$), the null hypothesis of average equality must get rejected only for window size, so we implemented the corresponding *post-hoc* test. Results of Figure 2 show that the window sizes of 0.5 s have

the most significant performance. Meanwhile, slides of 0.125 s have the most statistically significant performance, although it is not significant compared with the other slide sizes, as illustrated in Figure 2.

FIGURE 2
Results of accuracy in the HCRF model with interactions between electrodes as an additional feature, by varying the window size or slide size in the sliding window algorithm. **(A)** Varying the window size. **(B)** Varying the size of the slide.

Discussion

Results from Table 1 and Figure 1 suggest that adding correlation or Jaccard distance to the existing features improves significantly the performance of LDA and HCRF classifiers. It indicates that having available information on similarity or

distance relations between channels gives additional knowledge about the classes that carry to a more accurate classification. However, it is the only behavior that the models have in common when electrode interactions get added to the features. Also, it is crucial to remark although correlation and Jaccard distance are quantifications of interactions with distinct attributes, we get

a performance improvement for both measures. It means that adding interactions between electrodes significantly improves the performance of given classifiers, independent of the nature of the interaction measure.

In the LDA model, results from Table 2 show that only a few additional pieces of information provided by the SWT are enough to improve the classification. It is performed by adding brain interactions from 2 s sliding windows and displacements of 0.5 s, or even with the whole trial with no shifts. It is due to the nature of the LDA model, where a dimensionality reduction of data to 1 dimension is necessary to perform the discrimination analysis (Bishop, 2006). With more features added, the complexity of data gets enhanced because of their dimensionality. Since dimensionality reduction always implies loss and distortion of information, preserving most of them in a tractable processing core is mandatory (Gracia et al., 2014; Zenil et al., 2016). One mode is handling data with reduced dimensions before the dimensionality reduction step, which preserves information with less distortion and loss. It happens by using brain interactions from 2 s or higher sized windows and displacements of 0.5 s.

On the contrary, the HCRF model has different behavior. Results from Figure 2 show that the window size is the most relevant parameter to implement the SWT. Simultaneously, the change in the slide size had no or less effect on the classifying performance. It indicates that information from window size smaller than or equal to 1/3 trial size is more relevant than the one coming from longer windows. It also leads to the longest available window displacement usage without losing relevant information, reducing the dimensionality and, hence, the input size of the features to the model.

In summary, although electrode interactions do not contribute significantly to classification in a multiclass task by themselves (Miller et al., 2014), this study proved that their combination with temporal features provides significant information to improve the classification in a two-class task, such as motor imagery. Also, we showed that performance improvement is independent of the nature of the interaction measure. The future direction of the study point is to use the electrode interactions as additional features in motor imagery tasks with more than two classes, more than three electrodes, and dividing the frequency power bands into smaller sizes.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.bbci.de/competition/iv/.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Funding

Acknowledgments

Conflict of interest

JD was employed by SciFork SARL.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., and Zhao, X. (2019). A comprehensive review of EEG-based brain-computer interface paradigms. *J. Neural Engg.* 16, 011001. doi: 10.1088/1741-2552/aaf12e

Aggarwal, S., and Chugh, N. (2019). Signal processing techniques for motor imagery brain computer interface: a review. *Array* 1–2, 100003. doi: 10.1016/j.array.2019.100003

Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., and Zhang, H. (2012). Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Front. Neurosci.* 6, 1–9. doi: 10.3389/fnins.2012.00039

Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2008). "Filter bank common spatial pattern (FBCSP) in -computer interface," in *Proceedings of International* Joint *Conference* on *Neural Networks* (Hong Kong), 2390–2397.

Bandettini, P. A., Jesmanowicz, A., Wong, E. C., and Hyde, J. S. (1993). Processing strategies for time-course data sets in functional mri of the human brain. *Magn. Reson. Med.* 30, 161–173. doi: 10.1002/mrm.1910300204

Benesty, J., Chen, J., and Huang, Y. (2008). On the importance of the pearson correlation coefficient in noise reduction. *IEEE Trans. Audio Speech Lang. Process.* 16, 757–765. doi: 10.1109/TASL.2008.919072

Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). "Optimal filters in the time domain," in *Springer Topics in Signal Processing* ed H. Sawada (IEEE), 1–18.

BenYahmed, Y., Abu Bakar, A., RazakHamdan, A., Ahmed, A., and Abdullah, S. M. S. (2015). Adaptive sliding window algorithm for weather data segmentation. *J. Theor. Appl. Inf. Technol.* 80, 322–333.

Bezdek, J. C., and Pal, R. N. (1995). An index of topological preservation for feature extraction. *Pattern Recognit.* 28, 381–391. doi: 10.1016/0031-3203(94)00111-X

Bishop, C. M. (2006). *Pattern Recoginiton and Machine Learning* New York, NY: Springer.

Bishop, C. M. (2013). *Pattern Recognition and Machine Learning*. New York, NY: Springer-Verlag.

Cha, S. H. (2007). Comprehensive survey on distance/Similarity measures between probability density functions. *Int. J. Math*. Model Method Appl. Sci. 1, 300–307. Available online at: http://www.fisica.edu.uy/cris/teaching/Cha_pdf_distances_2007.pdf

Clifford, P. M. (1985). The international vocabulary of basic and general terms in metrology. *Measurement* 3, 72–76. doi: 10.1016/0263-2241(85)90006-5

Ferrero, L., Quiles, V., Ortiz, M., Ianez, E., and Azorin, J. M. (2021). "BCI based on lower-limb motor imagery and a state machine for walking on a treadmill," *International IEEE/EMBS Conference on Neural Engineering 2021-May*, 431–434.

Fletcher, S., and Isla, M. Z. (2018). Comparing sets of patterns with the Jaccard index. *Aust. J. Inf. Syst.* 22, 1–17. doi: 10.3127/ajis.v22i0.1538

Ghanbar, K. D., Rezaii, T. Y., Farzamnia, A., and Saad, I. (2021). Correlation-based common spatial pattern (CCSP): a novel extension of CSP for classification of motor imagery signal. *PLoS ONE* 16, e0248511. doi: 10.1371/journal.pone.0248511

Gracia, A., González, S., Robles, V., and Menasalvas, E. (2014). A methodology to compare dimensionality reduction algorithms in terms of loss of quality. *Inf. Sci.* 270, 1–27. doi: 10.1016/j.ins.2014.02.068

Gubert, P. H., Costa, M. H., Silva, C. D., and Trofino-Neto, A. (2020). The performance impact of data augmentation in CSP-based motor-imagery systems for BCI applications. *Biomed. Signal Process. Control* 62, 102152. doi: 10.1016/j.bspc.2020.102152

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. New York, NY: Springer-Verlag.

HCRF Library (2011). *HCRF Library (including CRF and LDCRF)*. SourceForge. net. Available at: https://sourceforge.net/projects/hcrf/ (accessed November 14, 2022).

Hota, H. S., Handa, R., and Shrivas, A. K. (2017). Time series data prediction using sliding window based RBF neural network. *Int. J. Comput. Intell. Res.* 13, 1145–1156.

Jalilpour Monesi, M., and Hajipour Sardouie, S. (2019). Extended common spatial and temporal pattern (ECSTP): a semi-blind approach to extract features in ERP detection. *Pattern Recognit.* 95, 128–135. doi: 10.1016/j.patcog.2019.05.039

Koles, Z. J., Lazar, M. S., and Zhou, S. Z. (1990). Spatial patterns underlying population differences in the background EEG. *Brain Topogr.* 2, 275–284. doi: 10.1007/BF01129656

Leeb, R., Lee, F., Keinrath, C., Scherer, R., Bischof, H., and Pfurtscheller, G. (2007). Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment. *IEEE Trans. Neural Syst. Rehabil. Eng.* 15, 473–482. doi: 10.1109/TNSRE.2007.906956

Luo, J., Gao, X., Zhu, X., Wang, B., Lu, N., and Wang, J. (2020). Motor imagery EEG classification based on ensemble support vector learning. *Comput. Methods Progr. Biomed.* 193, 105464. doi: 10.1016/j.cmpb.2020.105464

Luo, J., Wang, J., Xu, R., and Xu, K. (2019). Class discrepancy-guided sub-band filter-based common spatial pattern for motor imagery classification. *J. Neurosci. Methods* 323, 98–107. doi: 10.1016/j.jneumeth.2019.05.011

Martens, S. M. M., Mooij, J. M., Hill, N. J., Farquhar, J., and Schölkopf, B. (2011). A graphical model framework for decoding in the visual ERP-based BCI speller. *Neural Comput.* 23, 160–182. doi: 10.1162/NECO_a_00066

Miller, K. J., Ojemann, J. G., and Henderson, J. M. (2014). "Instantaneous interactions between brain sites can distinguish movement from rest but are relatively poor at resolving different movement types," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*. Chicago, IL.

Mohdiwale, S., Sahu, M., Sinha, G. R., and Nisar, H. (2021). Investigating feature ranking methods for sub-band and relative power features in motor imagery task classification. *J. Healthc. Eng.* 2021, 3928470. doi: 10.1155/2021/3928470

Mokhtari, F., Akhlaghi, M. I., Simpson, S. L., Wu, G., and Laurienti, P. J. (2019). Sliding window correlation analysis: Modulating window shape for dynamic brain connectivity in resting state. *Neuroimage* 189, 655–666. doi: 10.1016/j.neuroimage.2019.02.001

Oikonomou, V. P., Nikolopoulos, S., and Kompatsiaris, I. (2020). Robust motor imagery classification using sparse representations and grouping structures. *IEEE Access* 8, 98572–98583. doi: 10.1109/ACCESS.2020.2997116

Pfurtscheller, G., and McFarland, D. J. (2012). "BCIs that use sensorimotor rhythms," in *Brain–Computer InterfacesPrinciples and Practice* (New York, NY: Oxford University Press), 228–240.

Quattoni, A., Wang, S., Morency, L. P., Collins, M., and Darrell, T. (2007). Hidden conditional random fields. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 1848–1853. doi: 10.1109/TPAMI.2007.1124

Scherer, R., Schloegl, A., Lee, F., Bischof, H., Janša, J., and Pfurtscheller, G. (2007). The self-paced graz brain-computer interface: methods and applications. *Comput. Intell. Neurosci.* 2007:79826. doi: 10.1155/2007/79826

Schlögl, A., Lee, F., Bischof, H., and Pfurtscheller, G. (2005). Characterization of four-class motor imagery EEG data for the BCI-competition 2005. *J. Neural Eng.* 2, L14–L22. doi: 10.1088/1741-2560/2/4/L02

Singh, A., Kammermeier, S., Plate, A., Mehrkens, J. H., Ilmberger, J., and Bötzel, K. (2011a). Pattern of local field potential activity in the globus pallidus internum of dystonic patients during walking on a treadmill. *Exp. Neurol.* 232, 162–167. doi: 10.1016/j.expneurol.2011.08.019

Singh, A., Levin, J., Mehrkens, J. H., and Bötzel, K. (2011b). Alpha frequency modulation in the human basal ganglia is dependent on motor task. *Eur. J. Neurosci.* 33, 960–967. doi: 10.1111/j.1460-9568.2010.07577.x

Singh, A., Plate, A., Kammermeier, S., Mehrkens, J. H., Ilmberger, J., and Bötzel, K. (2013). Freezing of gait-related oscillatory activity in the human subthalamic nucleus. *Basal Ganglia* 3, 25–32. doi: 10.1016/j.baga.2012.10.002

Sutton, C., and McCallum, A. (2011). An introduction to conditional random fields. *Found. Trends Mach. Learn.* 4, 267–373. doi: 10.1561/2200000013

Tangermann, M., Müller, K. R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci.* 6, 1–31. doi: 10.3389/fnins.2012.00055

Vergara, V. M., Abrol, A., and Calhoun, V. D. (2019). An average sliding window correlation method for dynamic functional connectivity. *Hum. Brain Mapp.* 40, 2089–2103. doi: 10.1002/hbm.24509

Voinas, A. E., Das, R., Khan, M. A., Brunner, I., and Puthusserypady, S. (2022). "Motor imagery EEG signal classification for stroke survivors rehabilitation," in *2022 10th International Winter Conference on Brain-Computer Interface* (Gangwon-do), 1–5. doi: 10.1109/BCI53720.2022.9734837

Wang, B., Wong, C. M., Kang, Z., Liu, F., Shui, C., Wan, F., et al. (2021). Common spatial pattern reformulated for regularizations in brain-computer interfaces. *IEEE Trans. Cybern.* 51, 5008–5020. doi: 10.1109/TCYB.2020.2982901

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Yao, L., Jiang, N., Mrachacz-Kersting, N., Zhu, X., Farina, D., and Wang, Y. (2022). Performance variation of a somatosensory BCI based on imagined sensation: a large population study. *IEEE Trans. Neural Syst. Rehabil. Eng.* 30, 2486–2493. doi: 10.1109/TNSRE.2022.3198970

Zenil, H., Kiani, N. A., and Tegnér, J. (2016). Quantifying loss of information in network-based dimensionality reduction techniques. *J. Complex Netw.* 4, 342–362. doi: 10.1093/comnet/cnv025

Zhang, R., Xu, P., Liu, T., Zhang, Y., Guo, L., Li, P., et al. (2013). Local temporal correlation common spatial patterns for single trial EEG classification during motor imagery. *Comput. Math. Methods Med.* 2013, 591216. doi: 10.1155/2013/591216