Journal of NeuroEngineering and Rehabilitation

RESEARCH

Open Access

Characterization of error-related potentials during the command of a lower-limb exoskeleton based on deep learning

Paula Soriano-Segura[1,2], Mario Ortiz[1,2*], Cristina Polo-Hortigüela[1,2], Eduardo Iáñez[1,2] and José M. Azorín[1,2,3]

Abstract

Background Brain-Machine Interfaces (BMI) based on motor imagery (MI) are promising assistive neurotechnology tools for gait rehabilitation that allow users to control exoskeletons by imagining motor actions. Literature has proven the influence of BMIs over neuroplasticity mechanisms. However, the accuracy of MI-BMIs is often limited by the weak brain signals associated with lower-limb movements. To enhance system reliability and safety, Error-Related Potentials (ErrP), which are exogenous potentials evoked by erroneous system actions, can be integrated to correct commands. This study characterizes and detects ErrP during the use of a lower-limb exoskeleton, making use of a deep learning approach to improve accuracy and robustness over traditional classifiers.

Methods ErrP detection is performed using the EEG-Inception neural network, a convolutional deep learning model, and applying data augmentation techniques to the imbalanced dataset. The methodology is tested first for the characterization of ErrP during the start of gait with static data and, after confirming its improvement, regarding previous developments, it is also applied to motion data during the stop of the exoskeleton. With this objective, an experimental protocol is designed to evoke ErrP and NoErrP during motion, using tactile stimuli. ErrP is elicited when the exoskeleton stops erroneously in a gait region, while NoErrP is generated when it stops correctly in a stop region.

Results The proposed approach achieves a True Positive Rate (TPR) of approximately 95% and a False Positive Rate (FPR) below 20% in both static and motion conditions, significantly outperforming traditional ensemble classifiers. In terms of MI-BMI performance, these results indicate that most erroneous commands are successfully canceled, while only a small number of correct commands are wrongly canceled. In addition, statistical analysis revealed no significant differences between the detection of ErrP in static and motion scenarios, nor between sessions or subjects just in static. However, significant differences are observed between subjects in motion and also the outcomes of ErrP and NoErrP classes in both scenarios.

Conclusion The EEG-Inception neural network provides a robust and accurate method for ErrP detection. Future research will focus on integrating ErrP detection with MI classifiers and validating the system with SCI patients for improved gait rehabilitation therapies.

*Correspondence:
Mario Ortiz
mortiz@umh.es

Full list of author information is available at the end of the article

Keywords Electroencephalography(EEG), Brain machine interface (BMI), Error related potential (ErrP), Exoskeleton, Neuro-rehabilitation, Deep learning

Background

A Brain-Machine Interface (BMI) is a technology that involves recording brain signals and decoding them into commands to control a device, such as an exoskeleton [1, 2]. These signals are typically recorded non-invasively using electroencephalography (EEG), although invasive techniques, such as electrocorticography (ECoG), are sometimes employed. However, ECoG carries higher risks for patients as they need a surgical intervention.

BMI systems, which control lower-limb exoskeletons, represent a promising technology for gait rehabilitation in patients with spinal cord injuries (SCI) or stroke, as they actively involve the patient in the recovery process [3]. Complementary approaches to exoskeleton control, such as multimodal cooperative strategies [4] and gait planning based on central pattern generators [5], have also been explored in recent literature, although they focus on biomechanical coordination rather than neurophysiological feedback. Specifically, BMIs based on the kinesthetic motor imagery (MI-BMI) paradigm offer a highly intuitive therapeutic approach [6–8]. In this approach, the patient imagines the sensation of movement in their muscles and the positions of their joints to activate a robotic device that executes the corresponding movement, thereby aiding in neuroplasticity improvements.

However, in MI-BMIs for gait rehabilitation, brain signals are weaker because the motor areas responsible for leg movement are located in the interhemispheric region of the cortex [9]. Consequently, the system's accuracy is often limited, impeding its clinical applicability [10–12]. To address this challenge, a recent approach involves using Error-Related Potentials (ErrP) to correct erroneous commands in MI-BMI systems [13–19]. An ErrP is an exogenous potential generated when the system behaves unexpectedly or incorrectly, and the user perceives this error. A relevant example within the context of this research is a subject walking with an exoskeleton that suddenly stops, despite the user's intention to continue walking. This discrepancy in system behavior triggers the user's detection of an error, eliciting an ErrP.

In the literature, ErrP is typically described by four distinct peaks: two positive and two negative. The potential first shows a slight decrement at around 200 ms after the erroneous action, followed by a pronounced positive peak at 300 ms. It then decreases to its most negative point at approximately 500 ms and gradually rising to a positive peak around 700 ms [15]. In addition, a key advantage of ErrP is its stability over long periods, even years [13, 15]. However, the shape and amplitude of these peaks may vary depending on the feedback modality that evokes the potential. While some studies elicit ErrP visually during monitoring tasks [13, 20, 21], others use tactile stimuli [22–25]. In a previous study [26], ErrP was evoked in an MI-BMI setting for initiating gait with a lower-limb exoskeleton, using three types of feedback: tactile, visual, and visuotactile. Among these, tactile feedback produced the best results. Similar findings have been reported in other studies comparing different feedback modalities [22, 24], as tactile feedback is processed more directly through the somatosensory system without causing cognitive overload [27].

Previously, [26] analyzed ErrP in static conditions within an MI-BMI designed to control gait initiation with a lower-limb exoskeleton. The present study extends this analysis to ErrP during motion, focusing on a scenario where the exoskeleton must stop. In order to achieve this objective, an experiment was designed with two predefined regions: a gait region, where the subject should continue walking without stopping, and a stop region, where the subject should attempt to stop the exoskeleton. ErrP is evoked when the subject unintentionally stops in the gait region, while NoErrP is elicited when the subject successfully stops in the stop region. These potentials are triggered using a tactile stimulus, delivered through two vibrating wristbands, that activate just before the exoskeleton stops. Six subjects participated in the experiments, each performing three sessions.

The current research proposes a new methodology for the characterization of ErrP during the command of a lower-limb exoskeleton based on deep learning using the EEG-Inception neural network [28]. First, the new approach is tested under static conditions for the characterization of ErrP during the start of the exoskeleton compared to [26]. Then, the methodology is subsequently extended to address results in motion data for the characterization of ErrP during the stop of the exoskeleton. Therefore, the main hypothesis is that it is possible to detect Error-Related Potentials, under static and motion conditions, when controlling a lower-limb exoskeleton by means of deep learning neural networks. In particular, the EEG-Inception neural network is expected to outperform traditional machine-learning classifiers, such as the ensemble approach, in terms of robustness, even under motion conditions affected by movement artifacts.

Methods

This section explains the experimental protocol designed to evoke the error potential and the success potential by means of a tactile feedback, while a subject is walking

using a lower-limb exoskeleton, intending either to continue walking or stop within a defined region. Furthermore, it also describes the materials employed and the participants involved in the experiment. Subsequently, the steps applied to the recorded dataset during the experiments are detailed, from EEG signals processing to the classification process for detecting ErrP.

Subjects

Throughout this research, 9 subjects have participated during the two experiments under analysis: 4 females and 5 males, with ages ranging from 20 to 32 years old (25.88 ± 3.34). All participants were physically healthy and were not taking any pharmaceuticals. Furthermore, they received a detailed explanation of the experiments and they provided written informed consent in accordance with the Helsinki Declaration. The study was approved by the Responsible Research Office of Miguel Hernández University of Elche (Spain) (DIS.JAP.09.21). The following table (Table 1) shows the nomenclature of the subjects for the two experiments under study: S1-S6 correspond to subjects under static conditions and M1-M6 to subjects under motion conditions. Notably, three of the subjects participated in both experiments.

Equipment

In the experiments, the electroencephalographic (EEG) signals are acquired using the g.Nautilus cap (g.tec medical engineering GmbH, Austria). The 28 wet non-invasive electrodes register the signal at a sampling rate of 250 Hz following a international 10-10 system distribution: AF3, F3, FZ, FC3, FC1, FCZ, C5, C3, C1, CZ, CP3, CP1, CPZ, P3, PZ, PO3, AF4, F4, FC2, FC4, C2, C4, C6, CP2, CP4, P4, POZ, PO4. The ground electrode is placed at AFZ and the reference electrode is positioned on the right ear lobe. Additionally, the electrooculographic (EOG) signals are recorded, to remove eye movements and blinks from the EEG signals, employing 4 electrodes: VU (Vertical Up), VD (Vertical Down), HL (Horizontal Left), and HR (Horizontal Right).

Table 1 Participation of subjects in the different experiments and their corresponding nomenclature

Subject	Static Experiment	Motion Experiment
1	S1	M1
2	S2	M2
3	S3	-
4	S4	-
5	S5	-
6	S6	M3
7	-	M4
8	-	M5
9	-	M6

Subjects S1 to S6 participated in the static experiment, while subjects M1 to M6 participated in the motion experiment. Note that three subjects participated in both experiments

During the trials, participants are assisted by a lower-limb exoskeleton, H3 exoskeleton (Technaid, Spain), for their gait movements. The robotic device is controlled through a state-machine system, that alternates between static and motion states, via commands received from a computer through a Bluetooth connection. The H3 exoskeleton was designed for research purposes and includes predefined gait patterns, as a result of an extensive study of human gait kinematics, where a controller adjusts the motor torque according to the gait phase. This mode ensures a symmetric motion pattern on both legs. In addition, the mechanical structure is fully adjustable to accommodate different user anthropometries, with configurable segments between the ankle, knee, and hip, as well as adjustable back supports. All attachment points are covered with foam padding to improve comfort. During trials, subjects require the use of crutches and a person positioned in the back assists and balances the subject to allow the leg to swing naturally and maintain stability.

Furthermore, subjects are equipped with a pair of vibratory wristbands, one on each wrist, that contain vibration motors inside to provide the tactile feedback. The behavior of this device is programmed on an Arduino Uno R3 (Arduino Org, Italy).

Experimental protocol

The main objective of the study is to detect ErrP in a MI-BMI for its posterior use as a corrective layer in the command control of a lower-limb exoskeleton. For this purpose, data were recorded in two different experiments, depending on the command intention of the subject: start or stop. The first database includes static information for the characterization of ErrP during the start of the gait with different feedbacks. This database was previously analyzed with an ensembled method and it is used in this research as a benchmark comparison for the proposed improved methodology. This is the reason why only the tactile feedback is used here, as it was the one with former higher results. A detailed explanation of the experimental protocol for the start experiment can be seen in [26].

The second experimental data were recorded specifically for this research in order to characterize the ErrP during the stop of the gait. Experimental protocol is conceived in a similar way than the static one, but taking into consideration the motion of the exoskeleton and the space required for the stopping. Hence, it is designed to evoke two potentials: a success potential (NoErrP), elicited when the exoskeleton correctly stops when the subject pretended to, and an error potential (ErrP), triggered

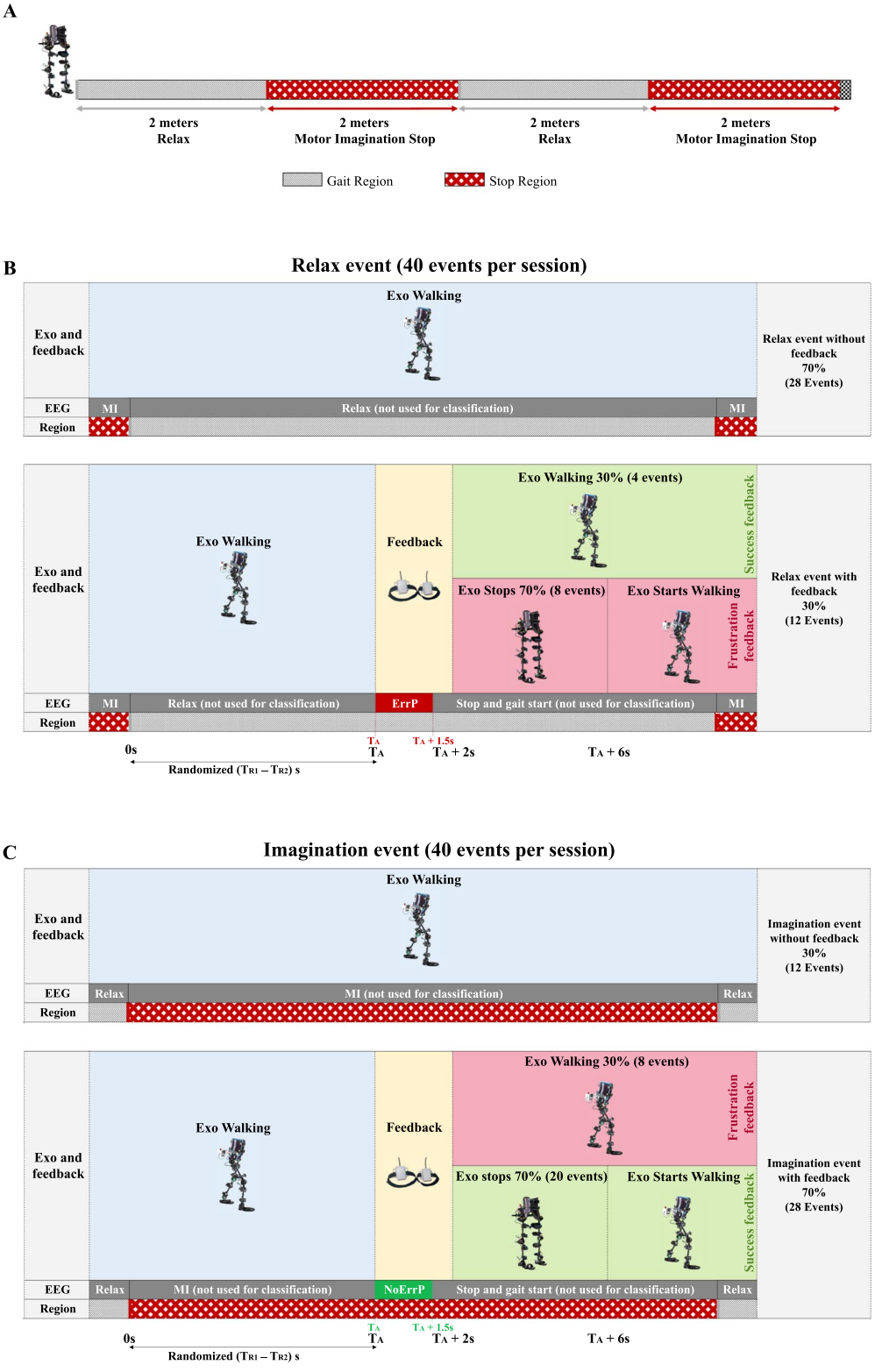

Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Experimental protocol to evoke ErrP in motion in a BMI for controlling gait stop using a lower-limb exoskeleton. **A** represents the space that the subject must cover with four regions: two gait regions (gray pattern), where the subject must keep walking, and two stop regions (red pattern), where they must attempt to stop by mean of MI. The mental tasks events are shown in **B** for Relax and **C** for Imagination. EEG periods, highlighted in blue during the exoskeleton's activity, correspond to the mental task executed by the subject, starting when the subject enters in the corresponding region. For reality likeliness purpose, error appearance is at a 30% rate. This means that the feedback does not activate, during the 70% of Relax events (correct relax) and 30% of MI periods (erroneous MI), which causes that the exoskeleton crosses the red region without stopping. Otherwise, in erroneous Relax (30%) and correct MI (70%), the feedback activates (yellow) for 2 s at a random time point (T_A) to evoke the potentials, which can be an ErrP (dark red) in Relax event or a NoErrP (dark green) in an Imagination event. Only the EEG signals during the vibration are extracted for classification in a window of 1.5s after the feedback activation. In this particular cases, the mental task (blue) duration (T_A) is randomized between T_{R1} and T_{R2}, which vary depending on the subject ability to walk with the exoskeleton. After feedback deactivates, the exoskeleton stops for 4 s in 70% of the cases (light red in Relax events and light green in Imagination events) and starts walking again automatically. In the remaining 30%, the exoskeleton keeps walking (light green in Relax events and light red in Imagination events) to simulate the command corrections and provide frustration and success feedback to the participant

when the exoskeleton stops unexpectedly, contrary to the user's intention.

The stopping protocol takes place in a straight line, shown in Fig. 1A, where the participant walks along a path divided into four distinct regions, each measuring two meters in length and marked on the floor with colored tape. These regions alternate between gait regions (gray pattern), where the subject must maintain the gait, and stop regions (red pattern), where the subject is required to stop the exoskeleton. Furthermore, the spatial layout of the trajectory is pretty similar to the regions used for some MI-BMI experiments, such as [12], to ensure consistency with the final application, in which ErrP detection will be combined with MI detection.

In order to achieve the goal of each region, the participant must perform two different mental tasks: idle state (Relax), where the subject is instructed to maintain calm and continue walking as relaxed as possible, and kinesthetic motor imagery (MI) of stopping the exoskeleton. Concretely, this kind of imagination consists of imagining how a concrete movement feels in the muscles and joints involved, but without performing the action. Hence, the participant must remain relaxed within the gait regions (gray pattern) to reach the next region without stopping. Once in the stop region (red pattern), they must perform MI of stopping to attempt to turn off the exoskeleton.

During each event, whether Relax or MI, the system can behave either as expected or unexpectedly wrong, to recreate both evoked potentials. Based on the literature [14], where a high error rate can lead the subject to habituation and a diminished or even disappearance of the potential, an error rate of 30% is set for each type of event to prevent this issue. Thus, Fig. 1B and C illustrate the four possible scenarios that can occur during the experiment.

On the one hand, in the Relax events shown in Fig. 1B, most of the times the subject walks continuously through the entire gait region without stopping (upper). However, in 30% of the cases, the exoskeleton unexpectedly stops (lower). On the other hand, in the MI events shown in Fig. 1C, the exoskeleton mostly stops within the designated stop region (lower), but in 30% of the cases, it passes through the region without stopping (upper). The event distribution during the session is predefined without any knowledge by the subject to create the needed evoked potentials. This way, the exoskeleton is commanded by the BMI in an opened-loop control.

In cases where the exoskeleton stops, whether in the stop region or in the gait region, the subject walks for a time interval between T_{R1} and T_{R2}, which is set before the trial and varies according to the subject's walking ability with the exoskeleton, as some cover the regions faster than others. The stop moment is randomly chosen within that range to ensure it occurs inside the designated region. Before the stop, tactile feedback (yellow) is activated (T_A) for two seconds ($T_A + 2$), evoking a potential that may be ErrP (red in EEG of Fig. 1B) if the vibration occurs within the gait region, or NoErrP (green in EEG of Fig. 1C) if the stop occurs within the stop region.

Following the vibration, the exoskeleton can either stop as expected or, in 30% of the cases, continue walking without stopping, simulating a scenario where some commands are modify by an ErrP classifier. This results in frustration feedback for MI events and success feedback for Relax events. When the exoskeleton stops, it remains stationary for 4 s, allowing the subject to reposition themselves before automatically start walking again without requiring any mental task or physical movement. Once the exoskeleton resumes walking or goes on after the feedback without stopping, it keeps walking to the next region or the end of the path, stopping at most once per area.

First, a familiarization session is conducted to allow participants to learn how to operate the exoskeleton and adapt to its movements. Then, each subject participates in three-day sessions, with 20 trials per session. The whole protocol lasts approximately 45 min plus the time required for instrumentation. Therefore, each subject provides a dataset of 84 NoErrP samples and 36 ErrP samples, i.e., 28 NoErrP and 12 ErrP samples per session.

Signal processing

The collected EEG signals undergo some preprocessing steps to prepare them for ErrP detection. Initially, the

signals are filtered using several techniques to eliminate noise and unwanted artifacts. The first filter applied is a 50 Hz Notch filter and a 1 Hz hardware high pass filter to remove powerline interference and low-frequency noise. This step is essential, as such noise can interfere with the effectiveness of subsequent filter. After that, the H∞ filter mitigates the artifacts from eye movements and blinks [29].

The original approach entailed the application of a band-pass filter between 1 and 10 Hz [26], as is commonly done in the literature [20, 22, 23], to focus on the frequency bands where the potential is most pronounced. However, when employing the neural network, this methodology resulted in a loss of valuable information for the model, leading to an increase in the false positive rate (FPR). Consequently, the decision was to avoid the use of a band-pass filter and to keep the valuable information of high frequencies. A similar issue was observed with the CAR spatial filter, which was also tested and discarded, due to its tendency to remove a significant amount of relevant information.

Once the signals are filtered, the feedback activation point is identified, and a 1.5 s window is extracted from the moment the feedback is triggered, as it can be seen in EEG section of Fig. 1B (red) and C (green). These signals are then divided into two classes: ErrP class, when the wristbands activate during a gait phase, and NoErrP class, when the vibration correctly occurs during a stopping phase.

ErrP detection

The main goal of this research is to characterize and detect the ErrP during the commanding of a lower-limb exoskeleton. In a previous study [26], the detection of ErrP in a static context, specifically at the beginning of the gait, was carried out by an ensembled of three Machine Learning classifiers. However, the system exhibited a high False Positive Rate (FPR), evidencing the need of further research to enhance the detection system and reduce erroneous detection. To address this issue, the BMI is improved through data augmentation and the employment of a novel detection method based on a Deep Learning neural network, known as EEG-Inception [28], to detect the ErrP in both static (Start ErrP) and motion scenarios (Stop ErrP).

Table 2 Hyperparameters used in the training of the EEG-Inception model for ErrP detection

Hyperparameter	Value
Sampling rate	250 Hz
Window size	1500 ms
Input size	(375, 28)
Number of electrodes	28 electrodes
Epochs	600 epochs
Batch size	128 samples
Activation function	ReLU
Optimizer	Adam
Loss function	Binary Cross Entropy
Learning rate type	Cyclical Cosine Annealing
Learning rate range	[0.005, 0.05]
Learning rate cycle period	300 epochs
Dropout rate	0.6
Filters per branch	14
Scales time	(750, 500, 250)

EEG-inception

The neural network employed to train the models for ErrP detection is EEG-Inception [28], a deep learning architecture based on a convolutional neural network (CNN). This classifier consists in a series of specialized blocks that apply both temporal and spatial filters to the EEG signals, extracting valuable features related to the potentials. As a result, the network effectively learns to distinguish between ErrP and NoErrP classes.

Training this classifier requires defining several hyperparameters to ensure an optimal learning process and accurate class discrimination. These hyperparameters are summarized in Table 2.

The parameters related to the input data are the number electrodes, which provides spatial information, and the window size of 1500 ms, which defines the temporal duration of the input signals. Given the sampling rate of 250 Hz, the resulting input matrix contains 375 time points across 28 electrodes.

The architecture of the neural network is defined by several parameters. One of the most important is the number of filters applied per branch, which is set to 14 filters. These filters are responsible for extracting features at different scales and allowing the network to capture information at multiple resolutions. The temporal scales at which these filters operate are set to 750, 500 and 250 time points. In addition, the dropout rate randomly deactivates neurons during training, reducing overfitting and enhancing the model's ability to generalize to unseen data. In this case, the dropout rate is set to 0.6, since the model's future application is in a closed-loop system with real-time data and varying conditions, where a robust generalization is crucial.

Other parameters also influence the neural network's performance. The activation function used is *ReLU*, which activates the neurons only for positive values, otherwise the negative values are set to zero. Furthermore, the loss function employed is *Binary Cross Entropy*, as the model is trained to differentiate between two classes. Furthermore, the optimizer chosen to enhance the convergence and efficiently update the weights is *Adam*. This optimizer depends on the learning rate, which controls

how much the weights adjust during each optimization step.

In this particular case, a *Cyclical Cosine Annealing* learning rate schedule is employed, where the rate decreases from 0.05 to 0.005 and each descending cycle lasts 300 epochs. Thus, diverse rates are explored during convergence and the cycle reset prevents from getting stuck in local minima. In order to perform at least two full cycles of the learning rate schedule, the training process is established in 600 epochs, with a batch size of 128 samples, which are the number of samples processed at the same time before updating the network's weights.

The model training is implemented using TensorFlow library in Python, and is executed by a GeForce RTX 2060 GPU via CUDA for faster computation. The training process for each model, given the size of the dataset, takes less than ten minutes. In addition, the average acquisition, processing and inference time of all algorithms is below 200 ms, which ensures low computational load and confirms the feasibility of the system for real-time BMI application.

Cross validation

The training of a model with the EEG-Inception neural network is based on a generic training, using data not only from the target subject, but also from other subjects. The main motivation for this approach is the necessity for neural networks to have a large and diverse dataset to train effectively, ensuring that the model can learn complex patterns and generalize well to unseen data. In addition, a larger dataset helps to reduce overfitting and enhances the robustness of the network during training.

Thus, Fig. 2 shows how the dataset is divided in two groups: a training dataset and a test dataset. The training dataset (light blue background) includes data from all subjects (blue, orange, green, pink, purple and red contours) across their three sessions, as well as data from two sessions of the subject is being evaluated. Then, the remaining session is used as test dataset (light yellow background). To ensure unbiased learning, both classes (ErrP and NoErrP) in the dataset are balanced across the training and test datasets.

The subject evaluation is performed applying a three-fold cross-validation, where each session is sequentially used as part of the test dataset in one fold, ensuring robust performance assessment.

Data augmentation

As it was highlighted during the explanation of the experimental protocol (Sect. 2.3), one of the key challenges of

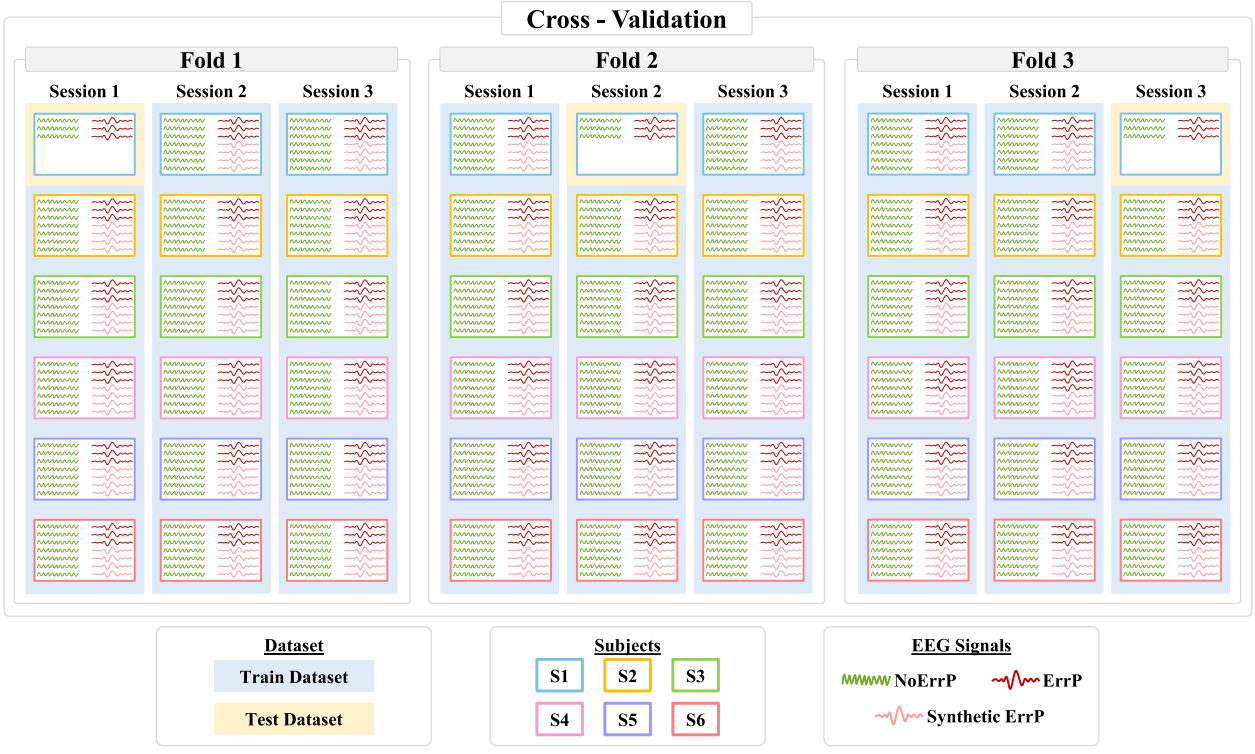

Fig. 2 Cross-validation process for a single subject (subject S1) across three folds (Fold 1, Fold 2, Fold 3), one for each session (Session 1, Session 2, Session 3). The data from S1 (blue framework) in the corresponding session is used as the test dataset (light yellow background), while data from the remaining sessions of the subject and from the rest of the subjects (S2: yellow framework, S3: green framework, S4: pink framework, S5: purple framework, S6: coral framework) are used to train the model (light blue background). The EEG signals of each subject include NoErrP (green), ErrP (red), and Synthetic ErrP (light red), each represented with distinct line styles

this study is the significant class imbalance. Since a 30% of error rate is necessary to prevent subject from getting used to errors, the resulting dataset is imbalanced. It contains a low number of ErrP samples (red EEG signals in Fig. 2) compared to NoErrP samples (green EEG signals Fig. 2). To tackle this problem, data augmentation is applied to the ErrP class in the training dataset (light red EEG signals in Fig. 2), improving the balance of the dataset for training without discarding any NoErrP samples. However, this method must not be applied to the test dataset to avoid introducing synthetic data that could bias the evaluation of the model's performance. In this case, the dataset is balanced using an undersampling method, which randomly removes a portion of the NoErrP samples to equalize both classes, as it can be seen inside the light yellow backgrounds in Fig. 2.

The data augmentation process, represented in Fig. 3, involves generating random combinations of three ErrP (light blue, light magenta and light yellow) samples within the dataset. The selection of the three random signals used belong to the same subject and session, ensuring that only signals from the same subject and session are mixed. Subsequently, the mean (red) of each set of three signals is computed, resulting in a new ErrP signal that is similar to the original potentials, but not identical [30]. This approach not only increases the data size of ErrP class, thereby balancing the two classes, but also introduces additional variability, providing the neural network with a richer dataset to enhance learning and generalization.

Metrics

This study addresses a binary classification problem to differentiate between ErrP and NoErrP classes. In these type of problems, a confusion matrix categorizes the predictions into four possible outcomes. On the one hand, when the real class is ErrP, the classifier can either correctly identify it as ErrP (*TP: True Positive*) or undetect it (*FN: False Negative*). On the other hand, when the real class is NoErrP, the classifier can either correctly predict it as NoErrP (*TN: True Negative*) or fail, predicting it as ErrP (*FP: False Positive*).

When evaluating the model trained with the neural network, it is important to get information not only about the overall performance of the model, but also about its ability to distinguish between classes. For this purpose, a set of metrics (*Accuracy*, *F1Score*, *True Positive Ratio (TPR)*, *False Positive Ratio (FPR)*) is employed, each providing valuable insights into different aspects of the model's behavior.

The most commonly used metric is *Accuracy*, which estimates the proportion of correctly classified samples out of the total number of samples. However, *Accuracy* alone does not provide a complete understanding of the model's performance. For this reason, *F1Score* is also employed as it combines two metrics into a single one by means of an harmonic mean: *Precision*, which indicates the proportion of *TP* among all ErrP predictions, and *Recall*, which calculates the ratio of *TP* among all actual ErrP samples. Thus, *F1Score* represents the classifier's effectiveness, providing a balance measure between *FP* and *FN*.

Additionally, since the detection of ErrP is crucial to enhance the performance of a BMI-MI, the values of *TPR*

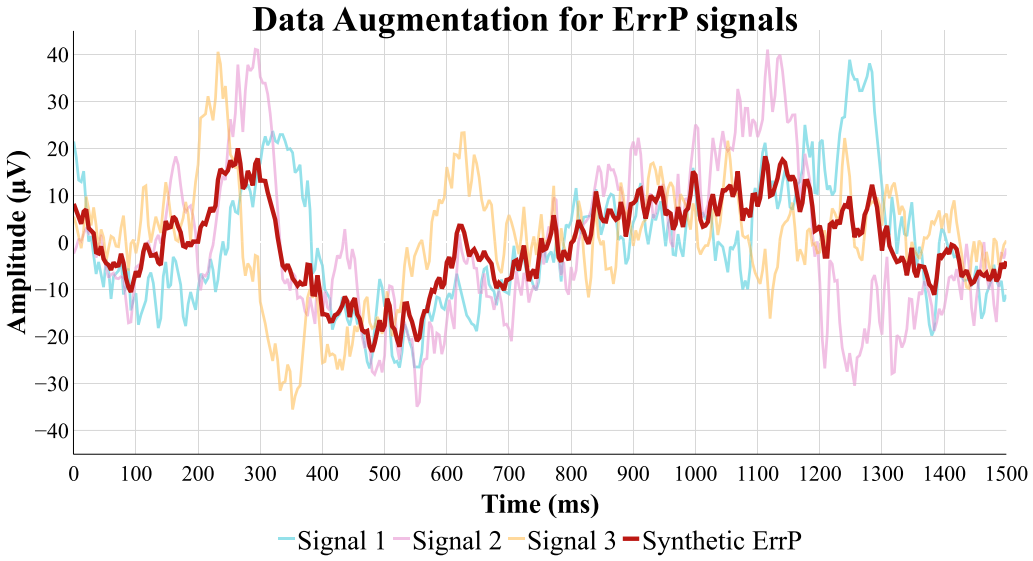

Fig. 3 Data augmentation process for ErrP signals. A combination of three signals (Signal 1: light blue, Signal 2: light magenta and Signal 3: light yellow) are randomly selected. The synthetic ErrP signal (red) is the average of the three signals

and *FPR* are essential for determining the system's overall effectiveness. Hence, the ability of the classifier to reliably differentiate between ErrP and NoErrP classes is reflected in a higher *TPR* and a lower *FPR*. Therefore, a high *TPR* value indicates that most erroneous MI detections are successfully identified and canceled, thus improving overall system performance. Conversely, a high *FPR* value means that correct MI detections are wrongly canceled, thereby reducing the overall performance.

Statistical analysis

After performing the data classification with the different models (static for start the gait and motion for stop), a statistical analysis is conducted to evaluate the significance of the results. This analysis helps to identify variables on which the system depends and whether there are statistically significant differences between the outcomes.

This statistical evaluation is carried out using IBM SPSS Statistics (IBM Corp., USA), where three dataset are created: one for static outcomes, another for motion outcomes, and a third combining both. These datasets share the same structure, where each row is an event predicted by the classifier and each column corresponds to each of the analyzed variables, along with the prediction outcome. Specifically, the columns include Method (Static/Motion), Gender (Female/Male), Subject (1-6), Session (1-3), Class (ErrP/NoErrP), and Detection (0 for incorrect predictions and 1 for correct predictions). For proper analysis, the categorical variables are codified as numerical values.

Thus, the variables individually evaluated for each ErrP type (Start or Stop), involve gender, subjects, sessions, and classes. Additionally, the combined data from both ErrP types are also examined to determine whether a significant difference exists between the two models.

Hence, the analysis starts with a Shapiro-Wilk test for normality to determine the appropriate statistical tests for the data. The test reveals that none of the variables follows a normal distribution, with p-values lower than 0.05. Then, the non-parametric Pearson's Chi-Square test of independence is performed for each variable to analyze whether there is a dependence between the variable and the outcome. Additionally, to determine whether there are differences in the predictions based on the value of each variable, the Kruskal-Wallis Test for k independent variables is performed for the variables 'subjects' and 'sessions', and the Mann–Whitney Test for two independent variables is applied to the variables gender, classes, and methods. For those variables with significant differences, a Dunn's Test for Multiple Comparisons is conducted to determine where the specific differences between values lie.

Results

This section presents the results of classifying ErrP with EEG-Inception neural network described in Sect. 2.5.1. First, the results obtained for the static error potential (Start ErrP) are shown, in order to check if this new classification method improves the results obtained with the assembled system of [26]. Then, the classification results of the motion ErrP (Stop ErrP) for its different sessions are shown. Finally, the results of the statistical analysis for the results obtained with the static and motion classifiers are presented, allowing to solve different hypotheses.

Start ErrP

First, the new classification method is applied to detect ErrP in the static dataset. The goal is to determine whether this approach employing a neural network achieves superior results compared to the initial classification method with the ensemble system. The previous method, documented in [26], was employed for the same static dataset to characterize the ErrP during the start of the gait.

In this particular case, when referring to the sessions, participants only performed one-day session, as the other two original sessions were done with a different feedback and are not comparable. In order to do a similar comparison, session is split into three. Here, session 1 is corresponding to the data at the beginning of the session, session 2 are the data in the middle of the session, and session 3 are the data by the ending of the session. This adjustment is made for nomenclature purposes and simplify the graphical representation. As a consequence of this reduction of data per session, it is decided to use a dropout of 0.35 for the classification of static data, instead of 0.6, so that the network could learn and mitigate the lack of data.

Figure 4 represents the results obtained for each subject in each of the sessions: Session 1 with blue borderline, Session 2 in yellow borderline and Session 3 in magenta. For each session, two bars are represented, one green bar showing the True Positives Rate (TPR) and, another red bar indicating the False Positives Rate (FPR). Additionally, the overall accuracy of each session is represented by a diamond in the corresponding color.

On the one hand, the plot shows that the TPR tends to be high, close to 100%, for most subjects and sessions. However, some subjects, such as S3 in Session 3 (66.67%), S4 in session 3 (66.67%), and S6 in session 1 (80.00%), exhibit a lower TPR. On the other hand, the FPR percentage is relatively low, below 20% in most sessions and even zero in many cases. Nonetheless, subject S5 shows a consistently elevated FPR across all sessions, reaching 33.33% in Sessions 1 and 3 and up to 50.00% in Session 2. In addition, accuracy is generally high across most sessions and subjects too, although occasionally accuracy

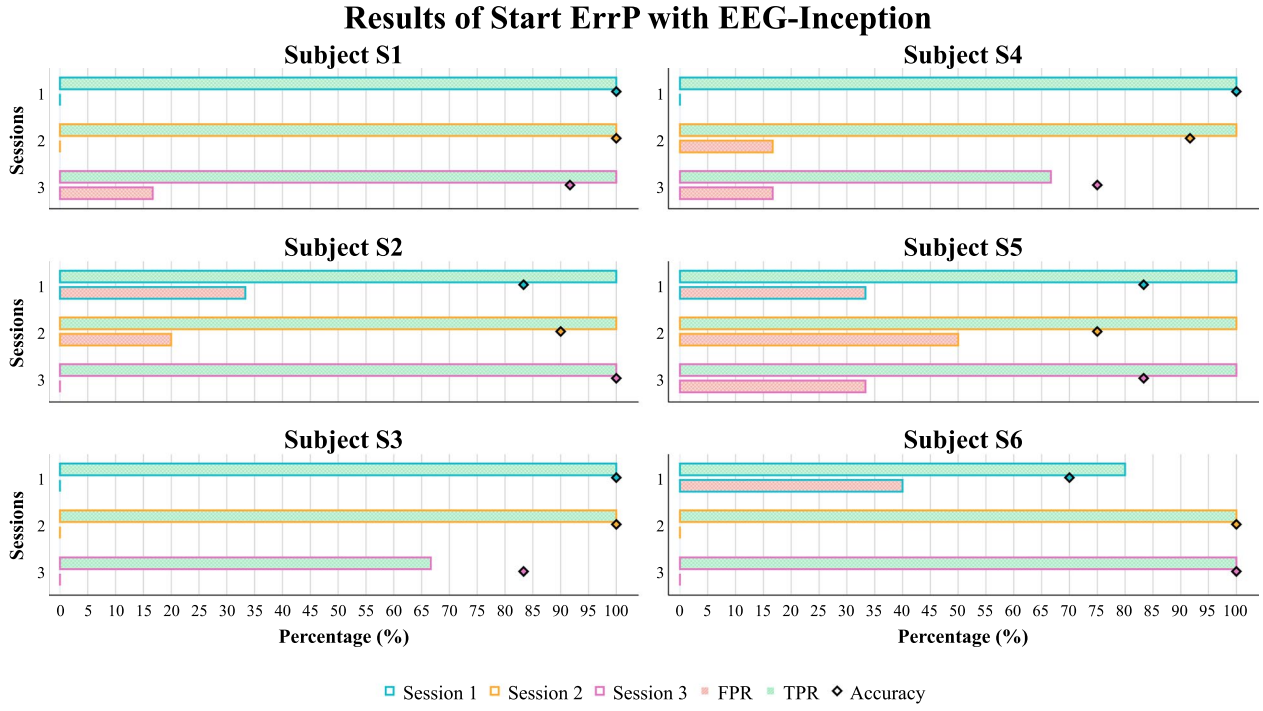

Fig. 4 Results of the Start ErrP detection with EEG-Inception for each subject (S1, S2, S3, S4, S5, S6) in each session: Session 1 with blue borderline, Session 2 in yellow borderline and Session 3 in magenta. For each session, two bars are displayed: a green bar showing the True Positives Rate (TPR), with higher values indicating better system performance, and a red bar representing the False Positives Rate (FPR), with higher values indicating worse performance. The overall accuracy of each session is represented by a diamond in the corresponding color

Table 3 Comparison of cross-validation results for Start ErrP detection using the Ensembled system [26] and EEG-Inception for each subject. The mean and standard deviation of the metrics for the six subjects appear in bolded text

Cross-validation Start ErrP								
Method	**Metric**	**Subjects**						**MEAN±STD**
		S1	**S2**	**S3**	**S4**	**S5**	**S6**	
Ensembled [26]	Accuracy(%)	64.18	76.78	66.18	60.35	59.28	68.35	**65.85±6.36**
	F1Score(%)	61.95	79.26	69.7	65.16	62.22	67.33	**67.60±6.44**
	TPR(%)	60.00	88.80	78.10	74.90	67.90	65.90	**72.60±10.23**
	FPR(%)	31.70	35.20	45.80	54.20	49.40	29.20	**40.92±10.27**
EEG-Inception	Accuracy(%)	97.22	91.11	94.44	88.89	80.55	90.00	**90.37±9.87**
	F1Score(%)	97.44	92.21	93.33	88.35	83.81	90.91	**91.01±9.19**
	TPR(%)	100.00	100.00	88.89	88.89	100.00	93.33	**95.19±7.72**
	FPR(%)	5.56	17.78	0.00	11.11	38.89	13.33	**14.44±16.09**

drops align with the observed decreases in TPR and increases in FPR.

Therefore, in general terms, the figure highlights the consistency of certain subjects, such as S1, S2, and S5, who achieve similar results across all three sessions. In contrast, other participants, like S3, S4, and S6, exhibit noticeable variations between sessions.

Table 3 presents the relevant classification metrics with their cross-validation results, i.e. the average of the three sessions for each subject.

The table compares the neural network results with the best results of the ensembled system [26]. The former results of the ensembled system present an average accuracy of 65.85% ± 6.36 with a TPR of 72.60% ± 10.23 and a high FPR of 40.92% ± 10.27. The accuracy across subjects varies considerably, with most subjects achieving around 65%, except for subject S3, who reaches the highest accuracy at 75.00%, and S5, who presents the lowest accuracy with 55.56%. In terms of TPR, subjects S1 and S3 show high values of 83.33% and 88.89%, respectively, while subject S5 achieves the lowest TPR at 50.00%. Furthermore, the FPR is notably high for most subjects, with subject S5 showing the highest FPR at 55.56%, while subject S3 achieves the lowest FPR at 27.78%. In contrast, the EEG-Inception results indicate an average accuracy of 90.37% ± 9.87 with a high TPR of 95.19% ± 7.72 and a low

FPR of 14.44% ± 16.09. The accuracy for most subjects is around 90%, except for subject S5, which is 80.55%, while S1 achieves the highest accuracy at 97.22%. Regarding TPR, three subjects, S1, S2, and S5, reach a perfect TPR, meaning that all error potentials are correctly detected, and the remaining subjects achieve around 90% precision. Additionally, the FPR is generally low, except for subject S5, who shows a FPR of 38.89%, while S3 achieves an FPR of 0.00%, indicating that no NoErrP instances are mistakenly classified as ErrP. Hence, the neural network method improves the overall performance by providing higher accuracy, increased TPR, and a lower FPR across subjects, indicating that the system effectively cancels erroneous commands, keeping the subject standing, while preserving correct ones for gait initiation.

Stop ErrP

The previous section demonstrates that the new methodology achieves higher results than the ones obtained with the ensemble system from [26] as it can be seen in Table 3. Therefore, this section presents the results of ErrP classification in motion conditions using the EEG-Inception neural network across the three sessions conducted for each subject.

Figure 5, similar to the previous figure, displays the TPR (green bars), FPR (red bars), and overall accuracy (diamonds) by subject and session, with color-coded borders for each session: blue for Session 1, yellow for Session 2, and magenta for Session 3.

On the one hand, the figure shows that TPR values remain high, close to 100%, across most subjects and sessions. However, some subjects, such as M3 (75.00%) and M6 (83.33%) both in Session 3, exhibit a lower TPR compared to the others. On the other hand, the FPR is low in most sessions, although unlike ErrP in static conditions, there are fewer sessions with a zero FPR. Furthermore, both subjects M3 and M4 show a high FPR in their first session, with 50.00% and 75.00%, respectively. Regarding accuracy, it remains high in most cases, except when the TPR is not as elevated or the FPR is not as low. For instance, M3 maintains an accuracy around 75% across all sessions. Therefore, it can be observed that this type of ErrP demonstrates less consistency across sessions than the start one, with only subjects M1 and M2 achieving similar results in all three sessions.

Table 4 presents the cross-validation results for the classification of ErrP in motion. It shows the averaged metrics obtained for each subject across the three sessions.

The table indicates an average accuracy of 87.96% ± 10.86, with a high TPR of 94.45% ± 7.45 and a low FPR of 18.52% ± 18.26. Generally, subject accuracy is around 80%, except for subjects M1 (100.00%) and M2 (95.83%), which are significantly higher, and M3 (76.39%), which is

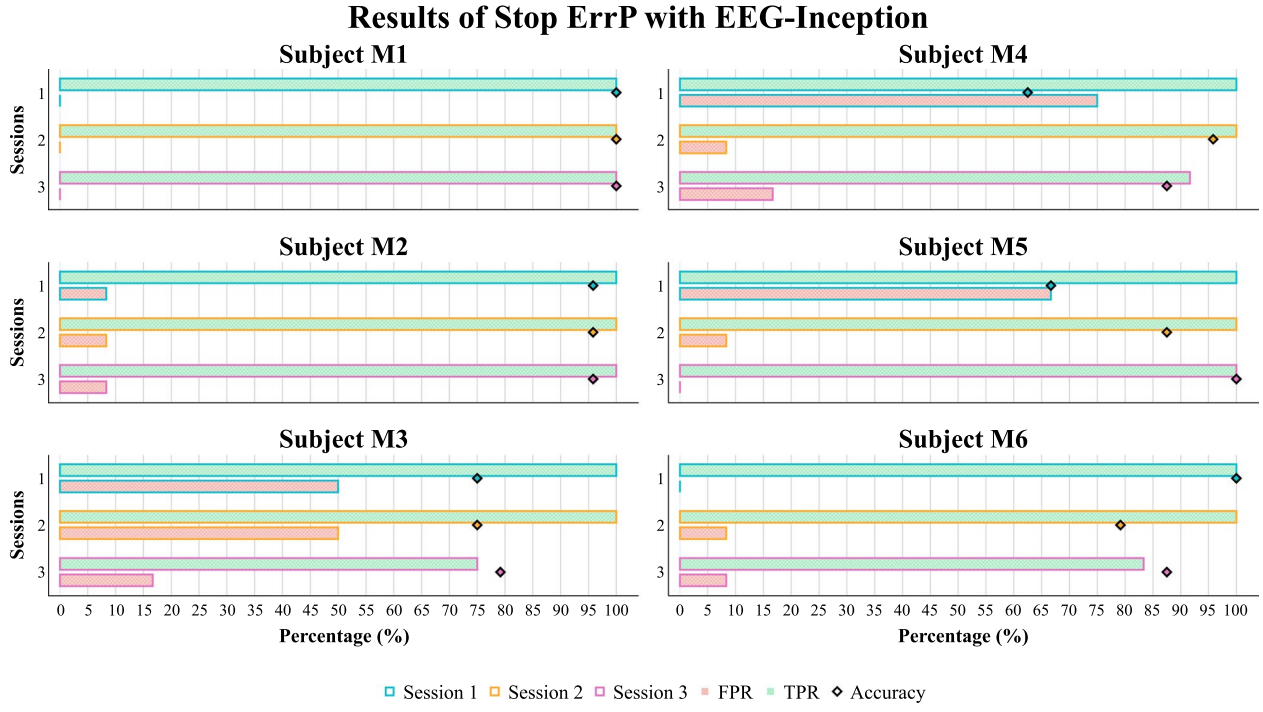

Fig. 5 Results of the Stop ErrP detection with EEG-Inception for each subject (M1, M2, M3, M4, M5, M6) in each session: Session 1 with blue borderline, Session 2 in yellow borderline and Session 3 in magenta. For each session, two bars are represented, one green bar showing the True Positives Rate (TPR), with higher values indicating better system performance, and a red bar representing the False Positives Rate (FPR), with higher values indicating worse performance. The overall accuracy of each session is represented by a diamond in the corresponding color

Table 4 Cross-validation results of the Stop ErrP detection with EEG-Inception for each subject. The mean and standard deviation of the metrics for the six subjects appear in bolded text

Cross-validation Stop ErrP				
Subject	**Accuracy (%)**	**F1Score (%)**	**TPR (%)**	**FPR (%)**
M1	100.00	100.00	100.00	0.00
M2	95.83	96.00	100.00	8.33
M3	76.39	79.42	91.67	38.89
M4	81.94	85.58	97.22	33.33
M5	84.72	87.32	94.44	25.00
M6	88.89	87.72	83.33	5.55
MEAN ± STD	**87.96±10.86**	**89.34±9.46**	**94.45±7.45**	**18.52±18.26**

notably lower. The TPR rate also tends to be high, with M1 (100.00%) and M2 (100.00%) correctly identifying all ErrP events, whereas M6 (83.33%) has a considerably lower rate than the others. In addition, the FPR varies widely among subjects. While M1 (0.00%), M2 (8.33%), and M6 (5.55%) reach near-zero FPRs, barely misclassifying NoErrP as ErrP, subjects M3 (38.89%), M4 (33.33%), and M5 (25.00%) exhibit greater difficulty in distinguishing NoErrP from ErrP. Therefore, these results indicate that the system effectively detects erroneous stop commands, allowing the subject to continue walking, while preserving correct ones to stop the gait when intended.

Statistical analysis

After predicting the classification outcomes for each subject across all sessions, a statistical analysis is conducted to confirm or reject several hypotheses previously suggested by the results presented.

The hypotheses formulated focus on determining whether the classification outcomes for each type of ErrP, both static and motion, depend on any of the following variables: gender, subject, session, or class. Furthermore, it also evaluates if there are statistically significant differences between the results of the two ErrP types, Start and Stop.

First, a Shapiro-Wilk normality test was performed, revealing that none of the variables followed a normal distribution, with p-values below 5%. Subsequently, various statistical analyses, as detailed in Sect. 2.6, were conducted to determine whether there was dependency on any variables or statistically significant differences. However, the outcomes for each variable were nearly identical across both tests. This similarity is due to the simple structure of the analysis table, containing binary categorical variables. This means that when comparing only two groups, the Chi-square test, Kruskal-Wallis test, and Mann–Whitney test tend to function in a similar manner, as they all assess whether the distributions between groups are statistically equal or different. As these are non-parametric tests designed to evaluate distributional differences, they yield comparable results when applied to binary or categorical data, since they essentially test the same hypothesis of equality across the categories. For this reason, Table 5 presents the resulting p-values from both statistical analyses for the variables Gender, Subjects, Session, and Classes under both static and motion conditions.

Table 5 Statistical p-values for ErrP types (Start/Stop) and their relationship with Gender, Subjects, Sessions, Classes

Statistical analysis *p*-values				
ErrP Type	**Gender**	**Subjects**	**Sessions**	**Classes**
Static	0.951	0.230	0.714	0.031
Motion	0.182	7.136e−5	0.097	3.565e−5

These *p*-values also indicate the statistically significant differences within those factors

On the one hand, the statistical analysis of the ErrP in static conditions indicates that classification outcomes do not depend on the subject's gender (p-value = 0.951), the subject (p-value = 0.230), or the session (p-value = 0.714). However, they do depend on the class being classified, with a p-value of 0.031, which is below the 0.05 significance level. This also implies that there are statistically significant differences between the results for the ErrP and NoErrP classes. In addition, the Dunn post-hoc test yields a p-value of 0.030552, confirming this difference between the two classes and indicating that the classifier responds differently when detecting each class.

On the other hand, the statistical analysis of ErrP under motion conditions indicates that classification outcomes depend on both the subject (p-value = 7.136e−5) and the class (p-value = 3.565e−5), with p-values below the 0.05 significance level. Nevertheless, these outcomes do not depend on gender (p-value = 0.182) or session (p-value = 0.097). Consequently, this indicates statistically significant differences between classes, as also observed in the static condition, as well as significant differences across subjects. Thus, when applying Dunn's test for subjects, several significant differences emerge: between subjects M1 and M3 with a p-value of 1.97e−4, between subjects M1 and M4 with a p-value of 0.012, and between subjects M2 and M3 with a p-value of 0.005. However, comparisons among the remaining subjects, all show p-values above the 0.05 significance level, which means that there are no statistically significant differences in classification performance between these other subjects.

Additionally, the non-parametric Mann–Whitney U test is conducted again to determine if significant differences exist between the ErrP results in static and motion conditions. The test yields a p-value of 0.318, which is greater than the 0.05 significance level, indicating that there is not statistically significant difference between detections during static and motion conditions.

Discussion

In previous studies, the detection of ErrP under static conditions was analyzed using an ensemble classification system [26]. However, this method obtained a TPR of 72.60% ± 10.23 and a FPR of 40.92% ± 10.27, indicating that improvements were needed to enhance the user's confidence in the system, since most of the correct detections of the MI classifier were incorrectly canceled. Therefore, before applying the ensemble classifier to the ErrP motion data, efforts were made to improve the ErrP detection results in static conditions using the EEG-Inception neural network.

Table 3 demonstrates how this new methodology improves previous results in [26], achieving a TPR of 95.19% ± 7.72 and an FPR of 14.44% ± 16.09. Subsequently, this neural network classification method was applied to ErrP motion data, with the results shown in Table 4, yielding a TPR of 94.45% ± 7.45 and an FPR of 18.52% ± 18.26. These results reflect a high capacity to accurately detect ErrP and avoid the misclassification of NoErrP as ErrP. Thus, the enhancement in results with this new method is confirmed.

Regarding the True Positive Rate (TPR), results are high under both static and motion conditions. For the Start ErrP (Table 3), subjects S1, S2, and S5 achieve a 100.00% detection rate, while the remaining subjects exhibit percentages close to 90%. Similarly, the TPR for the stop (Table 4) also remains high, with two subjects (M1 and M2) reaching 100% accuracy, and most of the other subjects attaining elevated percentages, except for M6, whose TPR slightly drops to 83.33%. These results reflect the models' high capability to detect ErrP in both conditions, although performance during the stop shows a slight decrease in some cases.

In terms of the False Positive Rate (FPR), the percentage in static conditions is below 15% for all subjects, except for S5, who shows a notably higher value of 38.89%. In motion conditions, while the average FPR is low, there is greater variability, with subjects M1, M2, and M6 displaying an almost null FPR, while subjects M3, M4, and M5 exceed 25%. This disparity is confirmed by statistical analysis, which reveals statistically significant differences among subjects only in the motion condition, particularly between the subject pairs M1 and M3, M1 and M4, and M2 and M3. This suggests that, although the model is effective in minimizing false ErrP detections under both conditions, some subjects experience greater variability in the motion condition.

These results indicate that the average accuracy of the system for the start detection is 90.37% ± 9.87, with most subjects achieving around 90% accuracy, except for subject S5, who shows an accuracy of 80.55%. In motion, the accuracy of stop detection is 87.96% ± 10.86, with some subjects reaching high percentages close to 100.00%, such as S1 and S2, while others, such as M3, exhibit lower values around 76.39%. Therefore, these results demonstrate that the reduction in system accuracy is directly related to an increase in the FPR.

Concerning the consistency of sessions, Fig. 4 shows that the start results are generally consistent across the three sessions for most subjects, except for the third session of S3 and S4, and the first session of S6. For the stop, Fig. 5 reveals less consistency between sessions; only M1 and M2 achieve similar results across all three sessions, while other subjects, such as M3, M4, and M5, in the first session and M3 also in the second session exhibit an elevated FPR. This initial variability may be related to a lack of familiarity with the system, especially in the early sessions, given that M1 (S1) and M2 (S2) had also previously participated in static experiments, whereas M3 (S6), despite also having participated in both experiments, fails to maintain consistency between them, showing an increased FPR in the motion experiments. However, statistical analysis indicates that there are no statistically significant differences between sessions in either condition (p-value>0.05)

Additionally, statistical analysis reveals significant differences between the ErrP and NoErrP classes in both static and motion scenarios, exposing that the model responds differently to each class. This difference may be produced by the fact that NoErrP class is just composed of original samples, whereas most of the data from ErrP class employed for training is generated by data augmentation, introducing greater variability in the dataset. This variability helps prevent the model from overfitting to the ErrP class, enabling it to generalize better to new samples. Furthermore, although no statistically significant differences were observed between static and motion conditions (p-value>0.05), the performance in motion is slightly lower due to the influence of movement artifacts. However, the larger amount of data available in motion, with three sessions compared to one in static, has helped to partially compensate for this effect.

Although no previous studies in the literature evoke the ErrP while walking, as in the present work, it is possible to compare this study with others sharing some common elements. In [13], ErrP is employed to enhance the accuracy of a BCI, achieving 85% precision in detecting ErrP and a mean false positive rate (FPR) below 21%. In [24], ErrP is evoked using tactile feedback, like in the current study, achieving a TPR of 73% and an FPR of 12%. In the study [31], the proposed neural network model achieved a TPR of 63% and an FPR around 12%, showing a moderate detection rate of ErrP with a relatively low FPR. Thus, in all these studies, the TPR and FPR percentages are comparable to or lower than those achieved in this study. Therefore, this research has achieved a good accuracy in

ErrP detection but there is still room for improvement in false ErrP detections.

Strengths and limitations

Throughout this investigation, strengths and limitations have been identified, shaping the development of the study.

The study represents a novel contribution to the ErrP characterization during the control of a lower-limb exoskeleton field. There are very few articles in the literature where ErrP during the start and stop of the gait is analyzed in combination with an exoskeleton.

Additionally, new datasets of evoked potentials during the stop of the exoskeleton have been recorded for this specific research with the help of a tactile vibration feedback from wristbands. The designed protocol is highly realistic, recreating real-life needs in the control of a exoskeleton thanks to cues related to position instead of acoustic signals. Stop experiments were conducted by six subjects, each performing three sessions. Although the number of subjects may seem limited, it aligns with the average sample size in studies of this type [32], considering the complexity of the experimental setup and the difficulty in recruiting participants available for multiple sessions. Therefore, the dataset generated from these experiments provides a valuable resource for further research on ErrP during motion tasks, as very few datasets of this kind exist.

One of the limitations of the protocol is its duration, lasting approximately 50 min plus instrumentation time, which can contribute to the subject fatigue. Besides, maintaining a 30% error ratio to prevent the subject from becoming used to the error and, thereby, avoiding the disappearance of the potential, results in a pretty low number of ErrP samples per session and an imbalanced dataset. This leads to a limited amount of relevant information for training the neural networks. A further limitation is that, in motion experiments, in which the subject walks using the exoskeleton, the signals may be affected by motion artifacts. Nevertheless, they are present in the same way in the ErrP and NoErrP classes in the motion trials, so they are not a differentiation factor although it can mask the brain patterns to decode ErrP. However, one differentiation point between static and motion models is the level of mental load that the subject faces. Walking with an exoskeleton while using crutches is a highly attention-demanding task, which may reduce the subject's focus on ErrP perception due to the inherent coordination required to move with the device during a motor imagery task. Some subjects even reported difficulties to perceive wrist vibration. This, combined with the overlapping of motor activation with the mental tasks to decode, makes it harder than in the static ErrP perception.

Regarding the exoskeleton, specific physical and mechanical limitations must be taken into consideration. The system is restricted to users between 160 cm and 185 cm in height and under 100 kg in weight, due to the torque capacity of the actuators. Furthermore, the gait pattern is limited to the sagittal plane, excluding rotational or lateral movements. Additionally, a battery, situated in the rear, modifies the user's center of gravity. For this reason, to compensate for these limitations, a person positioned behind the user assists and balances to the opposite side of the swinging leg, facilitating a more natural gait and preventing potential falls due to lack of stability.

Despite the limitations found during the experimental phase, the detection of ErrP stands out as one of the strengths of this research. First, to address the issue of limited data, data augmentation is proposed only for the ErrP class to increase the number of samples and balance the dataset. In addition, balanced data from other subjects are incorporated into the training process to reduce model over-fitting to individual subject data, increasing the diversity of the dataset.

The new approach developed with the EEG-Inception neural network [28] significantly improves the initial results obtained with the ensembled system [26]. This convolutional neural network is based on deep learning, enabling it to identify complex patterns in the data that traditional machine learning classifiers cannot detect. Another advantage of this new method is the reduced calibration time for closed-loop applications, as it eliminates the need for the extensive parameter selection process, such as electrodes, features, and classifiers. Additionally, the training time for the network is under 10 min, and if further reduction is desired, the number of epochs could be decreased while still allowing the model to learn effectively.

The results achieved with the neural network demonstrate high accuracy and robustness in detecting ErrP, showing comparable performance in both static and motion conditions. This indicates that, although motion artifacts may arise during movement, the neural network remains capable of detecting ErrP as effectively as under static conditions, where such artifacts are absent because the subject is stationary. Furthermore, the neural network reaches a high percentage of correct ErrP detections (TPR) of approximately 95% and a low rate of misclassifications confusing NoErrP with ErrP (FPR) below 20%.

These outcomes address another potential limitation of the system: the subject's confidence in using it. For instance, when the subject is walking and the exoskeleton attempts to stop against their intention, the erroneous command will be corrected in most cases, allowing the subject to continue walking, thereby reducing frustration

and increasing trust in the system. Conversely, whenever the subject intends to stop walking, the system will confirm the correct command in most of the cases and stop the exoskeleton, enhancing the safety of the system.

To sum up, the outcomes obtained in detecting Error-Related Potentials, both in static and motion scenarios, highlight the robustness of the new classification method using the EEG-Inception neural network for ErrP detection in a gait MI-BMI system designed for rehabilitation applications. This approach effectively maintains a balance between high TPR and a low FPR, which is crucial for ensuring both the safety and user experience of the system.

Future work

This study has presented the characterization of ErrP during the start and stop of a lower-limb exoskeleton. The future objective is to employ this as an additional layer of control of the exoskeleton in the BMI design. Consequently, the next steps in this research will focus on integrating ErrP detection with MI classifiers for a unified system capable of recognizing both start and stop intentions, while correcting the erroneous commands to enhance the overall system accuracy. This would need the design of a new protocol to mix both paradigms and evaluate the functionality of the system.

This integration of ErrP detection into MI-BMIs addresses critical limitations, such as enhancing user trust and system safety, by effectively correcting erroneous commands and accurately confirming the correct ones. For this reason, the balance between high precision and low false detection rates is crucial for the feasibility of the system.

Furthermore, for its clinical application, a shortening of the experimental protocols must be achieved. It is necessary to reduce the number of electrodes without compromising accuracy and to validate the use of dry electrodes in order to reduce instrumentation times. Ultimately, the goal is to validate the MI-BMI system, self-tuned with ErrP detection, by conducting long-term experiments with non-able-bodied subjects to assess its clinical utility for more efficient gait rehabilitation therapies.

Conclusions

This is an early-stage research for developing a neurorehabilitation technique based on a Brain-Machine Interfaces (BMI) that uses Motor Imagery (MI) paradigm for gait control with a lower-limb exoskeleton. Specifically, the present work focuses on Error-Related Potential (ErrP) detection for self-tuning commands within the MI-BMI system.

This study successfully demonstrates the capability for detecting Error-Related Potentials (ErrP) in both static and motion scenarios. The proposed deep learning approach employing the EEG-Inception neural network [28] achieves significant improvements over traditional classifiers, with a TPR of approximately 95% and an FPR below 20%, ensuring both accuracy and robustness. These results indicate that most ErrP are detected, and consequently erroneous commands are successfully canceled, while only a small number of correct commands are wrongly canceled.

The outcomes confirm that the deep learning model can maintain high performance even under motion conditions, where movement artifacts are present and the motor activity overlaps with the mental tasks, reflecting its adaptability to different scenarios. Notably, these results are achieved despite the challenge of maintaining a 30% error ratio, which resulted in a significantly unbalanced dataset with a limited number of ErrP samples. This highlights the robustness of the proposed approach and the positive contribution of ErrP detection makes in enhancing the safety and reliability of the MI-BMI system.

Future work will focus on integrating motor imagery (MI) detection with ErrP self-correction to create a BMI system to control the gait. This integration will be clinically validated in patients with spinal cord injury to assess its feasibility and effectiveness as a rehabilitation therapy.

Author contributions

Conceptualization, M.O., E.I., and J.M.A; Methodology, P.S. M.O., and C.P; Software, P.S and C.P; Validation, P.S., M.O., and E.I.; Formal Analysis, P.S., Investigation, P.S., M.O., C.P., E.I., and J.M.A; Data Curation, P.S., M.O. and C.P.; Writing - Original Draft, P.S.; Writing - Review and Editing, P.S., M.O., E.I., and J.M.A.; Visualization, P.S.; Supervision, M.O., E.I., J.M.A.; Project Administration, J.M.A.,, E.I.; Funding Acquisition, J.M.A., E.I.

Funding

This publication is part of grant PID2021-124111OB-C31, funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU. This research has also been supported by grant PRE2022-103336 funded by MICIU/AEI/10.13039/501100011033, and by the Valencian Graduate School and Research Network of Artificial Intelligence (ValgrAI), Generalitat Valenciana and European Union.

Data availability

The datasets supporting the conclusions of this article are available in the Zenodo repositories, (https://doi.org/10.5281/zenodo.10828804) for Error-Rel ated Potentials evoked by means of a tactile stimulus when starting gait and (https://doi.org/10.5281/zenodo.14190392) when stopping the gait, with a lower-limb exoskeleton.

Declarations

Conflict of interest

The authors declare no Conflict of interest.

Ethical approval

The study was approved by the Responsible Research Office of Miguel Hernández University of Elche (Spain) (DIS.JAP.09.21). All participants received a detailed explanation of the experiments and they provided written informed consent in accordance with the Helsinki Declaration.

Author details

[1]Brain-Machine Interface Systems Lab, Miguel Hernández University of Elche, Avda. de la Universidad, s/n, 03202 Elche, Alicante, Spain

[2]Engineering Research Institute of Elche – I3E, Miguel Hernández University of Elche, Avda. de la Universidad, s/n, Building Valverde, 03202 Elche, Alicante, Spain
[3]Valencian Graduated School and Research Network of Artificial Intelligence - ValGRAI, Camí de Vera, s/n, Building 3Q, 46022 Valencia, Spain

Received: 21 January 2025 / Accepted: 28 November 2025
Published online: 06 December 2025

References

1. McFarland DJ, Wolpaw JR. Brain-computer interfaces for communication and control. Commun ACM. 2011;5(54):60–6. https://doi.org/10.1145/1941487.1941506.
2. Ortiz M, Nathan K, Azorín JM, Contreras-Vidal JL. Brain-machine interfaces for neurorobotics. Springer Nature Singapore; 2023.
3. Ang KK, Guan C. Brain-computer interface in stroke rehabilitation. J Comput Sci Eng. 2013;6(7):139–46. https://doi.org/10.5626/JCSE.2013.7.2.139.
4. Zhan H, Kou J, Guo Q, Wang C, Chen Z, Shi Y, et al. Multilevel control strategy of human-exoskeleton cooperative motion with multimodal wearable training evaluation. IEEE Trans Control Syst Technol. 2025;3(33):434–48. https://doi.org/10.1109/TCST.2024.3477299.
5. Kou J, Wang Y, Chen Z, Shi Y, Guo Q. Gait planning and multimodal human-exoskeleton cooperative control based on central pattern generator. IEEE/ASME Trans Mechatron. 2025;8(30):2598–608. https://doi.org/10.1109/TMECH.2024.3453037.
6. Ortiz M, láñez E, Gaxiola J, Kilicarslan A, Contreras-Vidal JL, Azorín JM. Assessment of motor imagery in gamma band using a lower limb exoskeleton. In: 2019 IEEE International conference on systems, man and cybernetics (SMC); 2019;2773–2778.
7. Rodríguez-Ugarte M, láñez E, Ortiz M, Azorín JM. Improving real-time lower limb motor imagery detection using tDCS and an exoskeleton. Front Neurosci. 2018;10:12. https://doi.org/10.3389/fnins.2018.00757.
8. Colucci A, Vermehren M, Cavallo A, Angerhöfer C, Peekhaus N, Zollo L, et al. Brain-computer interface-controlled exoskeletons in clinical neurorehabilitation: Ready or not? Neurorehabil Neural Repair. 2022;12(36):747–56. https://doi.org/10.1177/15459683221138751.
9. Bakker M, Lange FPD, Helmich RC, Scheeringa R, Bloem BR, Toni I. Cerebral correlates of motor imagery of normal and precision gait. Neuroimage. 2008;7(41):998–1010. https://doi.org/10.1016/j.neuroimage.2008.03.020.
10. Ferrero L, Quiles V, Ortiz M, láñez E, Azorín JM. A BMI based on motor imagery and attention for commanding a lower-limb robotic exoskeleton: a case study. Appl Sci. 2021;4(11):4106. https://doi.org/10.3390/app11094106.
11. Ferrero L, láñez E, Quiles V, Azorín JM, Ortiz M. Adapting EEG based MI-BMI depending on alertness level for controlling a lower-limb exoskeleton. In: 2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE); 2022;399–403.
12. Ferrero L, Soriano-Segura P, Navarro J, Jones O, Ortiz M, láñez E, et al. Brain–machine interface based on deep learning to control asynchronously a lower-limb robotic exoskeleton: a case-of-study. J Neuroeng Rehabil. 2024;4(21):48. https://doi.org/10.1186/s12984-024-01342-9.
13. Ferrez PW, Millan JDR. Error-related EEG potentials generated during simulated brain-computer interaction. IEEE Trans Biomed Eng. 2008;3(55):923–9. https://doi.org/10.1109/TBME.2007.908083.
14. Chavarriaga R, Millan JDR. Learning from EEG error-related potentials in noninvasive brain-computer interfaces. IEEE Trans Neural Syst Rehab Eng. 2010;8(18):381–8. https://doi.org/10.1109/TNSRE.2010.2053387.
15. Chavarriaga R, Sobolewski A, Millán JDR. Errare machinale est: the use of error-related potentials in brain-machine interfaces. Frontiers Neurosci. 2014;7:8. https://doi.org/10.3389/fnins.2014.00208.
16. Cruz A, Pires G, Nunes UJ. Double errp detection for automatic error correction in an ERP-based BCI speller. IEEE Trans Neural Syst Rehabil Eng. 2018;1(26):26–36. https://doi.org/10.1109/TNSRE.2017.2755018.
17. Mousavi M, Krol LR, de Sa VR. Hybrid brain-computer interface with motor imagery and error-related brain activity. J Neural Eng. 2020;10(17):056041. https://doi.org/10.1088/1741-2552/abaa9d.
18. Fidêncio AX, Klaes C, Iossifidis I. Error-related potentials in reinforcement learning-based brain-machine interfaces. Front Hum Neurosci. 2022;6:16. https://doi.org/10.3389/fnhum.2022.806517.
19. Tao T, Jia Y, Xu G, Liang R, Zhang Q, Chen L, et al. Enhancement of motor imagery training efficiency by an online adaptive training paradigm integrated with error related potential. J Neural Eng. 2023;2(20):016029. https://doi.org/10.1088/1741-2552/acb102.
20. Omedes J, Iturrate I, Chavarriaga R, Montesano L. Asynchronous decoding of error potentials during the monitoring of a reaching task. In: 2015 IEEE International conference on systems, man, and cybernetics; 2015;3116–3121.
21. Lopes-Dias C, Sburlea AI, Müller-Putz GR. Online asynchronous decoding of error-related potentials during the continuous control of a robot. Sci Rep. 2019;11(9):17596. https://doi.org/10.1038/s41598-019-54109-x.
22. Tessadori J, Schiatti L, Barresi G, Mattos LS. Does tactile feedback enhance single-trial detection of error-related eeg potentials? In: 2017 IEEE International conference on systems, man, and cybernetics (SMC); 2017;1417–1422.
23. Schiatti L, Barresi G, Tessadori J, King LC, Mattos LS. The effect of vibrotactile feedback on ErrP-based adaptive classification of motor imagery. In: 2019 41st Annual international conference of the IEEE engineering in medicine and biology society (EMBC); 2019;6750–6753.
24. Ahkami B, Ghassemi F. Adding tactile feedback and changing ISI to improve BCI systems' robustness: an error-related potential study. Brain Topogr. 2021;7(34):467–77. https://doi.org/10.1007/s10548-021-00840-6.
25. Kim SK, Kirchner EA. Detection of tactile-based error-related potentials (ErrPs) in human-robot interaction. Front Neurorobot. 2023;12:17. https://doi.org/10.3389/fnbot.2023.1297990.
26. Soriano-Segura P, Ortiz M, láñez E, Azorín JM. Design of a brain-machine interface for reducing false activations of a lower-limb exoskeleton based on error related potential. Comput Methods Programs Biomed. 2024;10:255. https://doi.org/10.1016/j.cmpb.2024.108332.
27. Yao L, Mrachacz-Kersting N, Sheng X, Zhu X, Farina D, Jiang N. A multi-class BCI based on somatosensory imagery. IEEE Trans Neural Syst Rehabil Eng. 2018;8(26):1508–15. https://doi.org/10.1109/TNSRE.2018.2848883.
28. Santamaria-Vazquez E, Martinez-Cagigal V, Vaquerizo-Villar F, Hornero R. EEG-inception: a novel deep convolutional neural network for assistive ERP-based brain-computer interfaces. IEEE Trans Neural Syst Rehabil Eng. 2020;12(28):2773–82. https://doi.org/10.1109/TNSRE.2020.3048106.
29. Kilicarslan A, Grossman RG, Contreras-Vidal JL. A robust adaptive denoising framework for real-time artifact removal in scalp EEG measurements. J Neural Eng. 2016;4(13):026013. https://doi.org/10.1088/1741-2560/13/2/026013.
30. George O, Smith R, Madiraju P, Yahyasoltani N, Ahamed SI. Data augmentation strategies for EEG-based motor imagery decoding. Heliyon. 2022;8(8):e10240. https://doi.org/10.1016/j.heliyon.2022.e10240.
31. Gao Y, Tao T, Jia Y, et al. Error Related Potential Classification Using a 2-D Convolutional Neural Network. In: Liu H, Yin Z, Liu L, Jiang L, Gu G, Wu X, et al., editors. Intelligent robotics and applications. Cham: Springer International Publishing; 2022;711–21.
32. Wierzgala P, Zapala D, Wojcik GM, Masiak J. Most popular signal processing methods in motor-imagery BCI: a review and meta-analysis. Frontiers Neuroinf. 2018;11:12. https://doi.org/10.3389/fninf.2018.00078.

Publisher's Note