

MindRun: an Accessible Multi-Player BCI Game for Collaborative Motor Imagery Training

Alessio Palatella¹, Stefano Tortora^{1,2}, Emanuele Menegatti^{1,2}, Luca Tonin^{1,2}, Maryam Alimardani³

¹Department of Information Engineering, University of Padova, Padova, Italy

²Padova Neuroscience Center, University of Padova, Padova, Italy

³Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

E-mail: alessio.palatella@phd.unipd.it

ABSTRACT: Motor imagery (MI) brain-computer interfaces (BCIs) often suffer from limited engagement and reliance on high-cost electroencephalography (EEG) systems. We present MindRun, a continuous MI-BCI game supporting both single and multi-player interactions using a low-cost eight-channel EEG headset. The system was evaluated in a feasibility study with 10 healthy participants who engaged in both collaborative and solo game modes. EEG signals were decoded using a Riemannian geometry-based framework, yielding probabilistic outputs for continuous avatar control. In collaborative mode, the classifier combined individual outputs through a weighted probabilistic fusion to achieve stable joint control. Classification accuracy was significantly above chance level, and participants were able to achieve goal-directed steering in the game. These results demonstrate that continuous single and collaborative MI-BCI gaming is feasible with low-cost EEG hardware, highlighting the importance of interaction design and probabilistic control strategies for accessible real-time MI-BCI systems.

INTRODUCTION

Brain-computer interfaces (BCIs) have advanced considerably over the past decades, enabling applications in communication, control, and neurorehabilitation for both clinical and non-clinical populations [1]. Among the various paradigms, systems based on motor imagery (MI) are the most widely adopted because they allow users to voluntarily modulate their brain activity without external stimulation, which is relevant for applications such as prosthetic control and post-stroke neuro-rehabilitation. However, despite technological advances, user training remains one of the most critical bottlenecks limiting the practical deployment of motor imagery BCI systems [2]. The ability to reliably modulate neural activity is not innate, but requires sustained practice and cognitive effort [3]. Training is often mentally demanding, fatiguing, and frustrating, which can reduce motivation and increase dropout rates [4]. Enhancing user engagement during BCI training is therefore not only advantageous, but essential for long-term usability and widespread adoption. Gamification has emerged as a promising strategy to enhance motivation and improve user experience within

BCI training protocols [5]. Video games inherently incorporate structured goals, adaptive feedback, reward mechanisms, and immersive environments, transforming repetitive training tasks into meaningful interactive experiences [6]. Several studies have demonstrated that game-based BCI training can increase engagement and, in some cases, improve performance outcomes. For instance, Wang et al. [7] developed a virtual reality (VR) game that simulates real-life motor tasks to deliver adaptive MI-BCI neurofeedback, reporting higher event-related desynchronization (ERD) activation and improved classification accuracy. Similarly, Bermúdez i Badia et al. [8] introduced a VR-based MI-BCI system with personalized motor tasks and adaptive training, reporting higher levels of user enjoyment and acceptance. Mercado et al. [9] developed *FarmerKeeper*, a runner-style BCI video game designed for neurofeedback training in children with autism, demonstrating improved usability and user experience compared to a commercial BCI game while supporting sustained attention during sessions.

Despite these promising developments, most approaches focus primarily on single-user scenarios and individual training tasks, with limited consideration of social interaction during training. Social interaction is a powerful driver of motivation in learning and skill acquisition contexts, fostering emotional engagement, sustained attention, and intrinsic motivation [10]. Collaborative training introduces shared objectives, coordination, and interdependence, which can enhance both enjoyment and persistence. Therefore, integrating social collaboration into the BCI training protocols may represent an important step toward more sustainable and user-centered training approaches. However, empirical investigations and technological implementations in this area remain scarce. Bonnet et al. [11] were among the first to explore collaborative BCI paradigms by designing *BrainArena*, a two-player football-style video game fully controlled by MI-BCI. In this game, each participant moved a ball left or right by imagining the corresponding hand movements. The classifier outputs of both players were then combined to determine the ball's movement, supporting both collaborative and competitive modes. Their results indicated that multi-user BCI setups can be more engaging than tra-

ditional single-user protocols, with comparable levels of performance and motivation across interaction modes. Building on these early findings, Palatella et al. [12] investigated a collaborative training strategy based on a collaborative brain-machine interface (cBMI), in which two participants learned to operate a real robotic arm through cooperative control. Experiments involving three pairs of subjects over a five-day training period provided preliminary evidence that a collaborative framework may accelerate skill acquisition and help reducing frustration levels typically experienced by naive users.

Despite these explorations, accessibility remains a major barrier in socially engaging BCI research. The adoption of multi-user BCIs is particularly limited, in part because their implementation requires hyperscanning setups, often involving two research-grade electroencephalography (EEG) devices [13], which are typically costly and time-consuming to configure. To enable a larger adoption of collaborative BCI devices, reliable consumer-grade EEG systems and classifiers that can function effectively with a reduced number of electrodes are needed. In this work, we introduce *MindRun*¹, a video game that can be controlled using single or dual low-cost EEG headsets equipped with only eight channels. Through optimized signal processing and interaction design, we demonstrate that both single-user and collaborative BCI gameplay through MI can be achieved effectively despite the reduced channel density. The plug-and-play nature of the commercial headset, combined with our classifier architecture, substantially reduces setup time and technical complexity, facilitating faster deployment and improving accessibility beyond specialized research facilities.

The primary contribution of this study is a proof-of-concept continuous-control MI-BCI gaming architecture explicitly designed to support both individual and cooperative interaction using low-density EEG hardware. To achieve stable shared control in collaborative scenarios, we implemented a probabilistic multi-user fusion strategy based on softmax-weighted Riemannian decoding outputs, ensuring robust performance without degradation. By combining accessible hardware, cooperative gameplay, and task-driven assessment, this work advances BCI research toward scalable, ecologically valid, and user-centered real-world applications.

SYSTEM DEVELOPMENT

BCI Game Description: *MindRun* simulates a snowball rolling down a mountain valley. Along the path, rows of trees appear as obstacles, leaving a gap on either the right or left side of the screen. Participants must guide the snowball through the gap by performing the MI task. The game is organized into trials, each including one obstacle segment spanning a floor width of 60 *Uu* (Unity units). The snowball's maximum velocity is capped to ensure a minimum inter-trial interval of 10 s. The open-

ing between the trees corresponds to 20% of the floor width. Its outer boundaries never exceed the floor limits, while its inner boundaries never cross the center of the screen, ensuring that the opening remains entirely on either the left or the right side while maintaining a constant width. Within each trial, obstacle rows are presented in randomized order, with minor variations in scale, rotation, number of trees and position to enhance realism.

To prepare participants for each trial, three thin colored lines (red, green, and blue) are displayed on the ground. A fog wall positioned at 200 *Uu* restricts the visibility of subsequent segments. The red line, located 30 *Uu* after each obstacle row, signals the beginning of a new trial. After an additional 30 *Uu*, a green line appears and serves as a fixation cue to prepare participants for the MI task [14]. Following a further 60 *Uu*, the blue line marks the onset of the task (Figure 1). At this point, participants must steer the snowball toward the opening located 180 *Uu* ahead. The fog wall is positioned beyond the obstacle distance so that limited visibility becomes possible only a little before crossing the blue line (Figure 1). This configuration prevents anticipatory artifacts during the MI phase while still providing a preparatory cue before task initiation.

After each trial, if the *Continuous Feedback (CF) Reset* option is enabled in the main menu (Figure 1), the snowball automatically returns to the center of the screen. Under this configuration, lateral movement is disabled until the blue line has been crossed, otherwise the subjects can steer the snowball for all the run duration. All data acquisitions were performed with this option enabled.

Successful passage of the snowball through the opening triggers positive visual feedback (i.e., an animation of a purple explosion). Conversely, collisions with the trees slightly push the snowball backward, reducing its velocity and initiating a blinking animation that persists until the red line is crossed. During this phase, the snowball can pass through the trees without further interaction. This mechanism prevents the snowball from becoming stuck, however, if contact occurs while correctly passing through the opening, no backward displacement is applied and the snowball maintains its velocity.

During calibration runs, the snowball autonomously moves toward the opening, whereas during evaluation runs participants actively control its movement. In the evaluation mode, the snowball increases in size by 5% after each successful passage, up to a maximum of 140% of its original size, thereby progressively increasing task difficulty as performance improves. Conversely, following a collision, the snowball decreases in size by 5%, but never below the default diameter of 1.2 *Uu*.

EEG Acquisition: Brain signals were acquired using the 8-channel Unicorn Hybrid Black EEG headset (g.tec Medical Engineering, Schiedlberg, Austria) at a sampling rate of 250 Hz. Electrodes were positioned over the sensorimotor cortex according to the international 10–20 EEG system at locations *Fz*, *C3*, *Cz*, *C4*, *Pz*, *PO7*, *Oz*, and *PO8*. For the collaborative gameplay, two identical EEG

¹A playable version of *MindRun* is available at <https://github.com/Brain-Machine-Interface-VU/MindRun>.

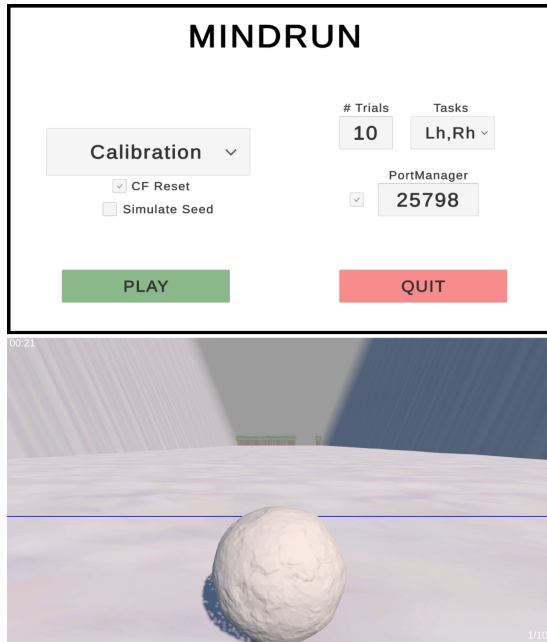


Figure 1: (Top) MindRun main menu. The interface allows users to select the run type, enable continuous feedback reset, recreate a game by specifying a seed in the PortManager, set the number of trials, choose the MI task, and configure the connection port for external scripts. If the connection manager is disabled, the game can be played using the arrow keys. (Bottom) Screenshot of a continuous feedback session. The top-left corner displays the elapsed time, while the bottom-right corner shows the number of completed trials out of the total.

headsets with the same configuration were used.

Riemannian BCI Classifier: A two-class MI paradigm was used, in which the participants performed kinesthetic imagination of the movements of the right and left hand. EEG signals were processed and classified using a Riemannian geometry-based pipeline designed for continuous control. To enhance local sensorimotor activity, EEG signals were first spatial filtered using a Laplacian filter. For each electrode, the spatial neighborhood was defined by the immediate adjacent electrodes along the horizontal and vertical axes of the grid. Because of the electrode layout, connections were restricted to directly adjacent electrodes and were not allowed to cross intermediate rows. Subsequently, the signals were bandpass filtered using a second-order Butterworth filter between 6 Hz and 24 Hz, with a stopband between 14 Hz and 18 Hz, to isolate the α and β frequency bands relevant for MI. The filtered signals were segmented using a sliding window of 1 s with a shift of 40 ms (update rate of 25 Hz).

For each window, spatial covariance matrices (SCM) were computed and regularized using Ledoit–Wolf normalization [15] to ensure that the resulting matrices were well-conditioned and remained symmetric positive definite (SPD). These SCMs were then used to train a Riemannian classifier based on the Fisher geodesic minimum distance to the mean (FgMDM) [16].

To enable continuous horizontal control of the snowball in the game, the classifier outputs were smoothed using

an exponential integration function [17], with smoothing parameter α ranging between 0.93 and 0.99. Due to the intrinsic nature of the classifier, which converts distances to class centroids into posterior probabilities, and the limitation of the exponential smoothing function, the output values rarely reach extreme probabilities close to 0 or 1. Consequently, to allow the snowball to reach the extreme horizontal positions, a hard decision rule was applied to the classifiers output, in which the class with the highest posterior probability was assigned probability 1 and the other class probability 0.

To prevent small fluctuations around the decision boundary from producing abrupt directional changes and jittery movements of the snowball, an additional thresholding rule was introduced. Specifically, probabilities below 0.51 were considered uncertain and reassigned to 0.50, thus directing the snowball toward the center of the screen and stabilizing its motion. With this approach, the value of $(1 - \alpha)$ in the integration function represents the contribution of each window to the movement of the snowball toward the right or left side.

In the collaborative condition, each subject retained an individual classifier. Their outputs were combined using a weighted mean before applying the same hard-decision remapping and integration procedures. The weighting strategy was designed to promote robustness and decoding stability rather than maximize ensemble accuracy. Subject-specific weights were derived from a composite performance score S_i computed on calibration data using probabilistic performance metrics including negative log-likelihood, Brier score, AUC, and expected calibration error, which are widely used measures for evaluating probabilistic classifiers [18, 19]. The collaborative weights were then obtained using a softmax normalization:

$$w_i = \frac{\exp(\beta S_i)}{\sum_{j=1}^M \exp(\beta S_j)},$$

where M is the number of users, S_i is the subject score and β controls the sharpness of the weighting. This strategy ensures stable and normalized weighting while preventing hard selection of a single participant.

VALIDATION STUDY

Participants: A total of 10 healthy participants (mean age: 23 ± 3 years, 4 males and 6 females) were recruited and randomly assigned to two groups. The first group included six players, randomly paired into three dyads for collaborative play. The second group consisted of four single players for solo play. None of the participants had previous experience with a BCI system. All participants completed two training sessions conducted on separate days with a 3-day interval between sessions. Written informed consent was obtained from participants after providing detailed information about the study goals, data management policy, and privacy protection measures. The study was approved by the Research Ethics Review Committee of the Faculty of Science, Vrije Uni-

versiteit Amsterdam.

Experimental Procedure: The experimental protocol comprised two sessions, separated by three days. Each session consisted of 10 runs of the proposed game 2, with each run including 10 trials, equally and randomly distributed between right- and left-hand MI. Two types of runs were defined: calibration runs, during which data were collected to train the decoder and feedback was systematically positive, and evaluation runs, in which the trained decoder was integrated into the proposed BCI framework, allowing participants to control the in-game avatar (i.e., snowball).

Session 1 was identical for all 10 participants who all played the game individually. Six calibration runs were first conducted to train the Riemannian BCI decoder, followed by four evaluation runs to assess its performance. In Session 2, participants were divided into two groups. Depending on the assigned group, they performed both the calibration and evaluation phases either individually or collaboratively. The session began with four additional calibration runs. The decoder was then retrained using the six calibration runs from Session 1 together with the four newly recorded runs to account for potential inter-session feature shifts. Participants subsequently completed six evaluation runs.

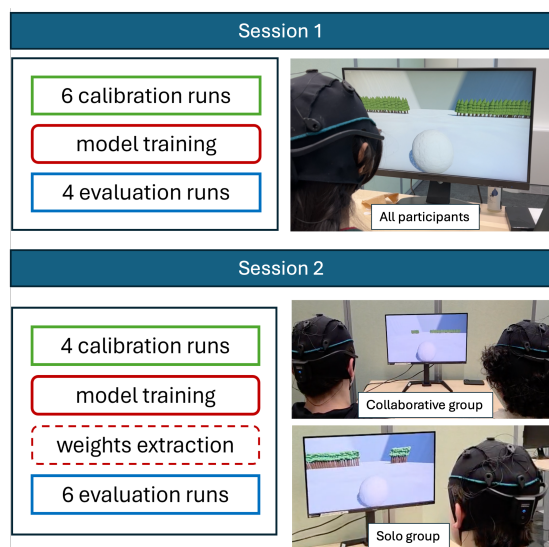


Figure 2: Experimental procedure illustrating distribution of calibration and evaluation runs across the two sessions. Weights extraction for BCI classification was applied only to the collaborative group. Videos of calibration and collaborative evaluation runs are available at <https://cloud.dei.unipd.it/index.php/s/gHiaG3JJjSaGkag>

Evaluation metrics: To evaluate the system and gameplay, two complementary sets of metrics were defined: in-game performance and BCI decoding accuracy. In-game performance was quantified as the lateral distance between the snowball and the center of the target gap at the moment of crossing, normalized by the total floor width. Since the gap corresponds to 20% of the floor width and is fully located on either the left or right side,

successful passage through the gap corresponds to normalized distances below 10%. The performance of the BCI decoder was assessed using three metrics. *Trial success rate* was defined as the percentage of trials in which the snowball successfully passed through the opening. This metric captures the effective control achieved by the participant within the game environment. *Sample accuracy* was defined as the proportion of covariance matrices correctly assigned to the intended MI class during the continuous feedback period over the threshold rule. Finally, *rejection rate* was calculated as the proportion of classifier outputs whose posterior probability fell below the uncertainty threshold of 0.51. Lower rejection rates, therefore, indicate more confident classifier outputs and stable control signals.

RESULTS

Figure 3 summarizes in-game performance and BCI decoding accuracies for collaborative and single players across the two sessions. No significant differences were observed between Session 1 and Session 2 in window accuracy (Wilcoxon signed-rank test: $p = 0.85$), trial success rate ($p = 0.32$), or distance-to-gap ($p = 0.70$), indicating stable performance for all participants.

In-Game Performances: Across both sessions, median distances in both solo- and collaborative-play conditions were generally close to the 10% gap threshold, and no statistically significant difference was observed between the two conditions (Mann–Whitney U tests: $p = 0.76$). This indicates that users were able to intentionally control the snowball’s trajectory and achieve goal-directed steering in both modes of the game.

BCI Decoding Accuracy: BCI classification accuracy remained consistently above the chance level (50%) across participants and sessions, with a mean of 0.61 ± 0.07 (Wilcoxon signed-rank test comparing accuracies vs. chance: $p = 0.002$, $r = 0.89$). All participants performed above chance level, indicating that the Riemannian geometry-based decoder reliably discriminated short-time MI states. In contrast, the overall trial success rate remained moderate, consistent with the gap distance results, particularly during Session 1, reflecting the stricter nature of trial metrics. Brief classification errors or transient uncertainty periods can invalidate an entire trial, making trial success rate a conservative measure of control performance. Rejection rates were moderate to high, effectively reducing rapid oscillations near the decision boundary. Crucially, elevated rejection rates did not impair task performance, as indicated by the distance-to-gap metric.

DISCUSSION

The purpose of this study was not to optimize MI classification accuracy, but to introduce and validate an integrated continuous MI-BCIs game framework supporting both single and collaborative interaction using a com-

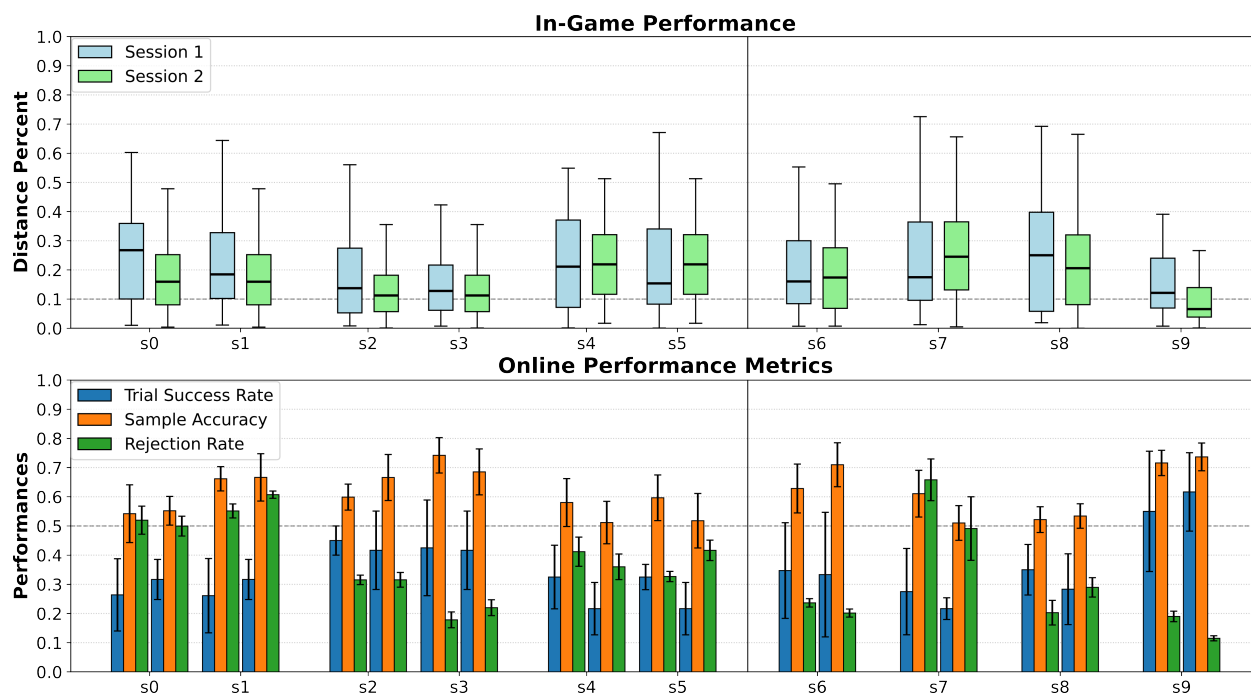


Figure 3: In-game performances (top) and BCI decoding accuracies (bottom) for collaborative dyads (s0–s1, s2–s3, s4–s5) and single players (s6, s7, s8, s9) across the two experimental sessions. In-game performance was quantified by the lateral distance between the snowball and the center of the target gap, expressed as a percentage of the total floor width (successful passage at 10%) (Bottom) BCI decoding performance was measured using three metrics: trial success rate, sample accuracy (chance level at 50%), and rejection rate.

mercially available EEG system. The observed above-chance sample accuracy demonstrates that a low-density consumer-grade EEG headset provides sufficient signal fidelity for Riemannian MI decoding. Game performance and BCI decoding results collectively show that MI control can be reliably translated into stable, continuous avatar steering in an interactive gaming environment. Furthermore, the collaborative fusion strategy enabled shared control without compromising system stability, confirming the feasibility of multiplayer MI-BCI interaction within a lightweight configuration. These findings confirm that functional continuous MI control is achievable with reduced channel density; however, direct comparisons with high-density systems are needed to objectively assess efficiency and potential performance trade-offs. While limited due to the different application scenarios and control paradigms, the results of current study can be compared to those of Palatella et al. [12] and Bonnet et al. [11], both introducing a collaborative MI-BCI using research fidelity EEG. When considering sample accuracy, the results obtained in this study are comparable to those reported by Bonnet et al. [11]. Specifically, they reported that overall sample accuracies across subjects ranged from 67% to 81%, with considerable variability between subjects. In contrast, Palatella et al. [12] reported trial accuracies that were highly dependent on the session, with a median value of approximately 45% and a range from 17% to 73%, which are also comparable to the results of the present study. However, it should be noted that our framework relies

on a more lightweight configuration, using a low-density EEG headset with fewer electrodes and a lower sampling rate. An additional analysis was performed by recomputing offline classification after removing the Fz channel, which is more susceptible to ocular contamination. As accuracies and rejection rates remained unchanged across participants (61.7% vs. 61.6% paired t-test: $t = 0.21$, $p = 0.84$ and 35.5% vs. 35.1%; paired t-test: $t = 0.35$, $p = 0.73$), the results suggest that decoding performance was not driven by EOG-related activity.

Importantly, decoding performance translated into measurable goal-directed in-game behavior. The results of the distance-to-gap metric confirm that participants were able to generate sustained and goal-directed steering rather than isolated classification events. This distinction is critical because successful BCI gaming requires temporal stability and controllability [5], not merely statistical separability of classes. Particularly in the collaborative gameplay, the probabilistic fusion of individual BCI outputs proved technically viable. The weighted softmax strategy maintained stable shared control without collapsing into dominance by a single user or introducing instability. This result supports the feasibility of multiplayer MI-BCI paradigms beyond competitive designs, extending toward cooperative interaction models.

The lower trial success rate compared to the sample accuracy should be interpreted within the task design context. The 20% gap width imposed strict spatial precision, making trial success an intentionally demanding outcome measure. Consequently, trial success rate functions as

a conservative measure of sustained control rather than instantaneous decoding capability. Similarly, higher rejection rates result from explicit design choices that prioritize smooth interaction over aggressive classification. In continuous control environments, stability and predictability are often more important than maximizing raw classifier output. The integration and thresholding mechanisms were therefore implemented as interaction design decisions aimed at improving control stability and usability. Nevertheless, the rejection mechanism does not resolve the intermediate control state. The rejection interval only filters uncertain outputs and is not a true “rest” class. Users must still maintain outputs near the decision boundary, representing a framework limitation.

By integrating social gameplay elements with an accessible hardware platform, this work contributes toward more scalable and ecologically valid BCI systems. The findings challenge the assumption that continuous BCI control necessarily requires high-density research-grade EEG setups and instead highlight the importance of interaction, classification robustness, and usability in real-world applications. Future research should extend this framework to larger cohorts and longitudinal protocols to evaluate learning effects, motivational impact, and the long-term benefits of collaborative training paradigms. To this end, we are currently collecting data from a larger sample size across 12 sessions, enabling the analysis of longitudinal learning dynamics between single-player and collaborative MI-BCI training conditions.

ACKNOWLEDGMENT

The work is partially funded by the Piano Nazionale di Ripresa e Resilienza (PNRR), Project PE8 “AGE-IT” Spoke 9 and Progetti di Rilevante Interesse Nazionale (PRIN 2022, 2022BCZ52A)-Ministero dell’Università e della Ricerca. This work reflects only the authors’ views and opinions, neither the European Union and European Commission can be considered responsible for them.

REFERENCES

- [1] Hramov A, Maksimenko V, Pisarchik A. Physical principles of brain–computer interfaces and their applications for rehabilitation, robotics and control of human brain states. *Physics Reports*. 2021.
- [2] Mladenović J. Standardization of protocol design for user training in eeg-based brain–computer interface. *Journal of Neural Engineering*. 2020;18.
- [3] Tortora S et al. Neural correlates of user learning during long-term bci training for the cybathlon competition. *Journal of NeuroEngineering and Rehabilitation*. 2022;19.
- [4] Roc A et al. A review of user training methods in brain computer interfaces based on mental tasks. *Journal of Neural Engineering*. 2020;18.
- [5] Atilla F, Postma M, Alimardani M. Gamification of motor imagery brain-computer interface training protocols: A systematic review. *Computers in Human Behavior Reports*. 2024;16:100508.
- [6] Vermeir JF, White M, Johnson D, Crombez G, Ryckeghem DVV. The effects of gamification on computerized cognitive training: Systematic review and meta-analysis. *JMIR Serious Games*. 2020;8.
- [7] Wang K et al. Adaptive neurofeedback training using a virtual reality game enhances motor imagery performance in brain–computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2025;33:2956–2966.
- [8] Badia SBI, Morgade AG, Samaha H, Verschure P. Using a hybrid brain computer interface and virtual reality system to monitor and promote cortical reorganization through motor activity and motor imagery training. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2013;21:174–181.
- [9] Mercado J, Espinosa-Curiel I, Escobedo L, Tentori M. Developing and evaluating a bci video game for neurofeedback training: The case of autism. *Multimedia Tools and Applications*. 2018;78:13675–13712.
- [10] Wang C, Huang L. A systematic review of serious games for collaborative learning: Theoretical framework, game mechanic and efficiency assessment. *Int. J. Emerg. Technol. Learn.*. 2021;16.
- [11] Bonnet L, Lotte F, Lécuyer A. Two brains, one game: Design and evaluation of a multiuser bci video game based on motor imagery. *IEEE Transactions on Computational Intelligence and AI in Games*. 2013;5(2):185–198.
- [12] Palatella A, Forin P, Tortora S, Menegatti E, Tonin L, et al. An alternative training protocol for a motor imagery bmi based on a collaborative approach. In: -. 2024.
- [13] Xue Z, Zhang Y, Li H, Chen H, Shen S, Du H. Instrumentation, measurement, and signal processing in electroencephalography-based brain-computer interfaces: Situations and prospects. *IEEE Transactions on Instrumentation and Measurement*. 2024;73:1–28.
- [14] Tonin L, Leeb R, Sobolewski A, Millán JdR. An online eeg bci based on covert visuospatial attention in absence of exogenous stimulation. *Journal of neural engineering*. 2013;10(5):056007.
- [15] Chen Y, Wiesel A, Eldar YC, Hero AO. Shrinkage algorithms for mmse covariance estimation. *IEEE transactions on signal processing*. 2010;58(10):5016–5029.
- [16] Barachant A, Bonnet S, Congedo M, Jutten C. Riemannian geometry applied to bci classification. In: *International conference on latent variable analysis and signal separation*. 2010, 629–636.
- [17] Tonin L, Bauer FC, Millán JdR. The role of the control framework for continuous teleoperation of a brain–machine interface-driven mobile robot. *IEEE Transactions on Robotics*. 2019;36(1):78–91.
- [18] Filho TS, Song H, Perello-Nieto M, Santos-Rodríguez R, Kull M, Flach PA. Classifier calibration: A survey on how to assess and improve predicted class probabilities. *Machine Learning*. 2021;112:3211–3260.
- [19] Waghmare KG, Ziegel J. Proper scoring rules for estimation and forecast evaluation. *Annual Review of Statistics and Its Application*. 2025.