Decoding Sensorimotor Rhythms:
Offline Classification of Motor Imagery for EEG-Based Brain–Computer Interfaces

Micah Baldonado
Carnegie Mellon University
Professor Bin He
November 3, 2025

Introduction

Brain–computer interfaces (BCIs) have become increasingly prominent over the last several decades as advances in neural engineering have made it possible to detect and interpret human intent directly from brain activity. These systems enable individuals with severe motor impairments, such as limb amputations or paralysis, to control external devices through neural signals alone. One widely used approach relies on sensorimotor rhythms (SMRs), recorded from electroencephalography (EEG) during motor imagery, the mental rehearsal of movement without overt execution. Motor imagery produces well-characterized changes in the sensorimotor cortex, including contralateral event-related desynchronization (ERD) in the alpha (8–12 Hz) and beta (13–30 Hz) bands. These rhythm modulations can be decoded and transformed into control signals for tasks such as moving cursors, selecting letters, or manipulating robotic devices.

Motor-imagery decoding can be performed through either online or offline classification. Online classification refers to real-time decoding of neural activity, where the system must update continuously as new data arrives. Offline classification, in contrast, is performed retrospectively on segmented trials and often achieves higher accuracy because it uses full-trial information and can leverage more sophisticated models. In this project, we worked with both approaches. In our experimental setup, the participant controlled a small on-screen ball, which functioned as the cursor during the task. The online classifier, provided as part of the experimental setup, used spatial filtering and linear classification—specifically common spatial patterns (CSP) and linear discriminant analysis (LDA)—to estimate cursor velocity by comparing alpha-band power between electrodes C3 and C4. This enabled real-time left–right or up–down cursor control driven purely by imagined movement.

In addition to evaluating the performance of the online system, we developed offline classifiers designed to decode SMR activity with greater accuracy. These offline models, which ranged from CSP-based linear methods to deep-learning architectures, consistently outperformed the online decoder. In the following sections, I describe the relevant background, experimental methodology, offline analysis pipeline, and the resulting performance comparisons.

Background and Review of State of Art

Over the years, Brain Computer Interfaces (BCIs) have evolved to encompass many forms, but most of them can be categorized by either involving direct stimulation to the central nervous system (CNS) or providing a novel output to the CNS that is not neuromuscular in nature (He, 2025). Common examples include deep brain stimulation (DBS), transcranial electrical stimulation (TES), transcranial magnetic stimulation (TMS), and transcranial focused ultrasound (tFUS). This literature review will focus on BCIs which provide a new output to the nervous system. Another important classification in BCI systems is that they can be either invasive or non-invasive. For example, while there is scalp EEG, there is also intracranial EEG (iEEG). One key example of invasive BCI is electrocorticography (ECoG). In 2004, Leuthardt et al. validated ECoG as a viable method for controlling a computer cursor. While invasive BCI modalities inherently collect signals which are closer to their source, it is ultimately much more dangerous for the user. In fact, Leuthardt et al. (2021) later published further work detailing the risks that come along with invasive BCI. Of the clinical risks, Leuthardt et al. point out

surgery-related hazards and long-term device risks (e.g., rejection, inflammation, hardware failure). Nevertheless, the classification between invasive and non-invasive technology is not entirely black and white as Leuthardt et al. (2021) emphasize that there is a difference between “invasive” and “minimally invasive” BCIs. Regardless, the subsequent literature review will focus on non-invasive BCI modalities. Non-invasive BCIs, which do not require surgery, are much more accessible and less expensive. The experimental setups are also well-established and widely replicated, leading to their broad use in BCI research. While this review focuses on non-invasive BCIs that generate new outputs to the nervous system, these systems cannot be considered in isolation.Together with invasive approaches, all BCI modalities are in conversation with each other, contributing collectively to unveiling the brain’s complex electrophysiological mechanisms.

Historically, the field lacked a unified framework, and early BCI systems appeared limited in scope. While BCIs have been around for at least half a century, the field became formally structured after a landmark review by Wolpaw et al. (2002), which outlined the basic BCI framework. With the basic principles for non-muscular communication publicized, this was one of the early papers which set the stage for future BCI technology, giving hope that it could be used for persons with severe motor impairments. Just a year later, Mason, S. G., and Birch, G. E. (2003) published *A General Framework for Brain–Computer Interface Design*, which proposed a systematic approach for constructing BCI systems. It broke down elements of the BCI pipeline including signal acquisition, feature extraction, and translation components, the latter referring to the algorithms that convert neural features into actionable system commands. Since then, BCI research has been steadily increasing in prevalence with the field garnering 1,431 publications between 2004 and 2024 (Zhang et al., 2024).

A key technology within BCI research is the electroencephalogram (EEG). Introduced in 1924, EEG has steadily advanced to become one of the most used and accessible methods for exploring brain activity and function (Britton et al., 2016). The widespread adoption of EEG stems from its practicality, cost-effectiveness, and non-invasive design. In fact, “Electroencephalogram recordings were the first method developed for direct and noninvasive measurements of brain activity from human subjects” (Woodman, 2010). Another major advancement in BCI-related technology was the discovery of the event-related desynchronization (ERD) by Pfurtscheller, G. and Aranibar, A. in 1977. Rather than relying on averaged brainwave data across participants, it became possible to identify event-related patterns shared among individuals exposed to the same stimulus and to develop a clearer representation of how brain responses evolve in real time. This work helped establish the broader class of event-related potentials (ERPs), which capture systematic neural responses time-locked to sensory or motor events and laid the foundation for SSVEP-based BCIs and motor-imagery BCIs (Abiri et al., 2019). Sensorimotor rhythm (SMR) BCIs are especially significant because they link brainwave activity to voluntary movement and laid the groundwork for prosthetic devices directly controlled by the brain. For persons with a recently amputated limb, BCIs offered hope for reestablishing communication between the brain and assistive devices by decoding signals from regions that previously controlled the missing limb. A major breakthrough occurred when researchers discovered that actual movement was not required for detecting movement-related brain activity. The advent of sensorimotor imagery BCI was a critical turning point, as now, even patients who have long been amputees are capable of external control without relying on the nerves their brain uses for limbs, specifically. For example, SMR-based BCIs were demonstrated by Wolpaw et al. (1991) were among the first to demonstrate how an EEG-based BCI could be used to control a computer cursor. Showing that people could learn to voluntarily control a BCI-system to move a cursor in real-time showed promise for the future of BCI technologies, setting the stage for even more complex control, beyond a simple cursor. At the time, even one-dimensional cursor control represented a remarkable technological milestone, generating considerable excitement in the emerging BCI community. Roughly a decade later, Wolpaw and McFarland (2004) later expanded their work in a subsequent study demonstrating control of a two-dimensional cursor using EEG sensorimotor rhythms. SMR has also been used with other modalities of BCI. Since then,

sensorimotor EEG-BCI technology has advanced substantially, offering renewed promise for individuals with limb loss.

Within the field of EEG BCI, exciting methodological changes have been made since the mid-90s which have shaped critical components of the BCI pipeline, including signal processing, feature extraction, and translation components. Key advancements include the Common Spatial Pattern (CSP) algorithm introduced by Ramoser et al. (2000) and the Independent Component Analysis (ICA) algorithm by Bell and Sejnowski (1995). The CSP algorithm was a key advancement in signal processing because it allowed spatial features to automatically differentiate between two different mental states through maximizing variance between them. With this advancement, it was much easier for researchers to produce high-performing online BCI technology which could decode neural intent. ICA was particularly important in signal processing because of its ability to remove artifacts such as blinking, which are notoriously disruptive to EEG signals in the field of BCI. Another important methodological tool in BCIs is Linear Discriminant Analysis (LDA), which, while first introduced in 1936, was not primarily used in BCIs by Müller-Gerking, J., Pfurtscheller, G., and Flyvbjerg, H. in 1999. What made LDA particularly useful was the ability to cleanly separate neural intent into different classes in linear fashion, which was extremely robust in online-settings. Combining CSP and LDA became a strong methodological strategy for creating robust and high-performing BCI systems. Unlike a simple linear classifier, which struggles with uncertainty arising from data mixing and denoising prior to intent classification, CSP and LDA together more effectively manage the inherently noisy nature of EEG data. Finally, while more computationally expensive, Riemannian geometry methods are another advancement for EEG-based BCIs specifically, as they enabled even more robust feature extraction (Congedo et al., 2017). However, because geometric methods are much more computationally expensive compared to CSP and LDA, even if they tend to outperform CSP and LDA, CSP and LDA remain to be much more widely used for online classification in most pipelines. Finally, with the explosion of deep learning in the 2000s, EEGNet was born, a “A compact convolutional neural network for EEG-based brain–computer interfaces” (Lawhern, V.J., et al., 2018). Rather than manual feature extraction from raw EEG signals, EEGNet was capable of automatically learning key temporal and spatial features for tasks, especially motor imagery. Indeed, EEGNet was a breakthrough that enabled people to have strong analysis of EEG data with minimal feature engineering.

In spite of the vast advancements and promising future of the field of BCI, the field still suffers some technical challenges which are often a bottleneck to our ability to create BCI systems with advanced capabilities. For example, the inverse problem remains a huge technical barrier: because BCIs rely on mapping features to an intended action, the ground truth signals often remain elusive, contributing to non-stationarity, the idea that users have to recalibrate frequently due to small shifts in equipment because even small shifts in equipment can distort the learned mapping between recorded features and neural intent (Chandrasekaran et al., 2021). In general, S. Saha et al., 2021 also provide a comprehensive view of obstacles in the field of BCI citing non-stationary brain signals, user fatigue, calibration, and learning burden. The authors write that adaptive machine learning could be the future, rather than static models which don’t bear as much of the burden of calibration and user fatigue. Furthermore, Yuan, Y., & He, B. (2014) have pointed out that user training, inter-subject variability, signal quality, and translation to real-world devices still remain huge barriers in the field of SMR-based BCIs. Nevertheless, even if there are many unsolved problems in the field of BCI, the field continues to grow and offers promise to those who are in dire need of it. Just recently, Choi et al. (2020) demonstrated that incorporating virtual reality (VR) into motor-imagery training can improve BCI performance and make the system more engaging for users. Edelman et al. (2019) used a hybrid BCI to improve continuous robotic arm control, which offers significant promise to those who need neural-controlled prosthetic arms. The field of BCIs remains in its infancy, yet its future holds immense promise for those who stand to benefit most from its advancement.

Experimental Protocol

The experiment followed a standard sensorimotor-rhythm (SMR) motor-imagery paradigm in which participants controlled a one-dimensional on-screen cursor (a small ball) by imagining left–right or up–down movements. Two task types were used: imagining left or right hand movement to move the cursor horizontally, and imagining both hands versus rest to move it vertically.

Hardware and Electrode Configuration

EEG was recorded using a Neuroscan wet-electrode cap with gel-based Ag/AgCl sensors connected to a Neuroscan amplifier and BCI2000 acquisition software. Only 27 gelled electrodes were used, primarily spanning the sensorimotor strip (Figure 1), and impedance was maintained below 5 kΩ (with all channels under 10 kΩ before starting). Although 27 channels were recorded, the online decoder used a core subset of motor-area electrodes, while the remaining channels were included to support richer offline analysis. Data were sampled at 1000 Hz and saved in MATLAB (.mat) format.

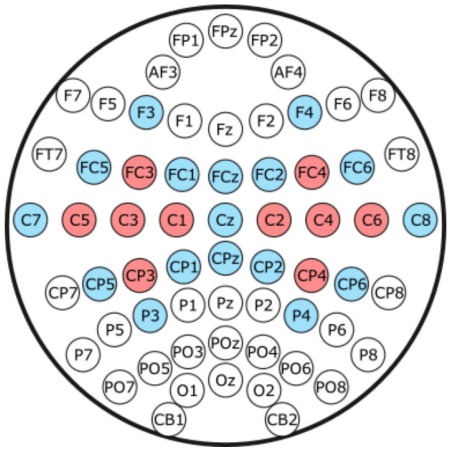

Figure 1. Electrode layout used in the SMR BCI experiment. Red electrodes indicate channels used by the online decoder, and blue electrodes represent additional gelled electrodes available for offline analysis. Adapted from the 42-633 BCI Group Project Protocol.

Session Structure

Each participant completed 14 runs per session: 7 left–right runs and 7 up–down runs, each containing 25 trials (350 trials total per session). After capping, the full experimental session lasted roughly 80 minutes. At the start of each trial, a target appeared on the screen, cuing the participant to imagine movement in the appropriate direction. Two seconds later, the cursor/ball appeared, and the online classifier began updating its velocity estimate. Trials ended with a hit, miss, or abort.

Online and Offline Decoding

The online classifier implemented a standard SMR pipeline: mu- and beta-band (8–30 Hz) features were extracted using common spatial patterns (CSP) and classified with linear discriminant analysis (LDA) to update cursor velocity in real time. The timing in the BCI2000 logs marks ball onset, not target onset, as trialStart. Because meaningful motor imagery begins immediately after cue onset, our offline analysis incorporated pre-ball activity.

For offline decoding, we worked with the runData structures stored by BCI2000, which included continuous EEG, trial start/end indices, target labels, and trial outcomes (hit, miss, abort). Although two full seconds of pre-ball activity were available, we used the final 1.5 seconds before ball onset, capturing stable motor imagery while avoiding early cue-processing activity. All aborted trials were excluded from offline classifier training.

Offline Analysis Pipeline

Our team followed a broadly similar preprocessing strategy rooted in standard SMR motor-imagery practice. We restricted the data to the 27 gelled electrodes positioned over the sensorimotor cortex, applied a band-pass filter to isolate the SMR frequency range, downsampled the recordings to reduce computational load, and aligned each trial to a consistent temporal reference by extracting a window spanning activity shortly before the onset of the ball and into the feedback period. In the offline analysis, we also excluded all trials classified as aborted by the online decoder, since these trials provide no usable motor-imagery signal and are excluded when computing Percent Valid Correct (PVC), the performance metric used throughout this project. Beyond these shared steps, the remainder of the pipeline diverged across team members.

Preprocessing

Although the overall structure of our group's preprocessing was similar, my implementation introduced several additional steps and constraints that were important for offline decoding. I restricted the EEG to the 27 gelled electrodes over the sensorimotor cortex and applied an 8–30 Hz band-pass filter to isolate the mu (8–12 Hz) and beta (13–30 Hz) rhythms most relevant to motor-imagery activity. I then applied common-average referencing to stabilize baseline activity across channels.

A key part of my preprocessing was defining a fixed 3000-sample window centered around the moment the ball appeared on screen. Each trial was required to contain at least 1500 samples from the feedback period to ensure sufficient temporal structure for extracting stable SMR features using CSP and deep learning models. In addition to the feedback segment, I included the 1500 samples immediately preceding ball onset. This pre-onset interval carries meaningful information because the online decoder is already estimating user intent during this time, and the participant has not yet reacted to the ball's movement (as decided by the online decoder). I deliberately did not include the full two-second interval before ball appearance, as the participant is unlikely to have processed the target location that early. By capturing the short interval after target presentation together with the feedback period, each trial contained both the initial rise of motor-imagery-related rhythms and the sustained SMR activity that drives cursor/ball motion. Finally, all extracted windows were downsampled by a factor of five to reduce data dimensionality, and aborted trials were excluded entirely because they do not contain strong, interpretable MI-related signals. These preprocessing decisions produced consistent, information-rich epochs that formed the foundation for the feature extraction and classification steps that followed.

Feature Extraction

For feature extraction, I used the Common Spatial Pattern (CSP) method, which generates spatial filters that maximize the variance difference between the two motor-imagery classes. I used six CSP pairs, giving twelve log-variance features per trial. Rather than fitting CSP once on the full dataset, I recalculated the CSP filters inside each fold of cross-validation so that no information from the test split leaked into the training process. This yielded more stable filters across folds and made the comparison between models scientifically consistent. To better understand what the CSP filters were learning, I also computed CSP spatial patterns and plotted them as topographic maps. These maps allowed me to verify that the discriminative patterns projected onto sensorimotor regions, which provided an important physiological check on model behavior.

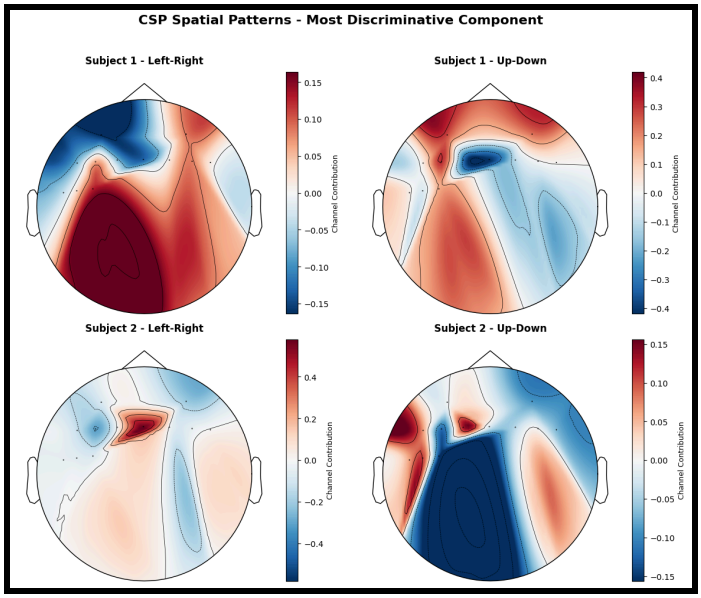

Figure 2. Topographic maps of the most discriminative CSP component for each subject and task. Warmer colors indicate channels contributing positively to class separation and cooler colors indicate negative contributions. Patterns localize to expected sensorimotor areas, confirming that CSP captured meaningful SMR features.

Although extended versions of CSP, such as FBCSP, were available, I chose not to rely on them for the main analysis. My experiments showed that FBCSP features did not improve performance with the dataset size and trial lengths we were working with, and EEGNet already incorporates its own learned temporal filtering within the architecture. For this reason, my final offline models primarily drew on CSP features or directly learned representations from the raw EEG.

Classifier Design

My offline analysis centered on testing how well different model classes could generalize across all motor-imagery conditions. Unlike other team members who trained separate models for left–right and up–down runs, I trained each classifier on the pooled dataset so that the models learned patterns that generalized across MI tasks. This choice reflects a more realistic BCI scenario, where a decoder must handle changes in direction, timing, and user strategy without retraining between tasks.

I implemented several models with increasing levels of representational complexity. As structured baselines, I trained CSP+LDA and CSP+MLP models to evaluate how well linear and shallow nonlinear classifiers performed with my CSP features. I also trained tree-based models, including Random Forest and XGBoost, which capture nonlinear interactions between CSP features and performed surprisingly well across tasks. Finally, I implemented EEGNet, a compact deep-learning architecture designed for EEG decoding. My EEGNet reproduced the temporal and depthwise convolutions from the original 2018 paper and was trained on raw EEG epochs for 800 epochs using a learning rate of 2e-4 and weight decay of 1e-3 to reduce overfitting (Lawhern et al., 2018). These classifiers, taken together, allowed me to compare a wide range of modeling approaches and analyze their strengths and weaknesses relative to the online decoder.

Cross Validation and Evaluation Strategy

To evaluate each model fairly, I used stratified five-fold cross-validation on the pooled dataset. This ensured that both motor-imagery classes were proportionally represented in every fold and that performance estimates reflected the variability across different subsets of trials. For models using CSP, the spatial filters were always computed only on the training portion of each fold to prevent information leakage. I reported the mean

accuracy and standard deviation across folds for each model and compared these values to the Percent Valid Correct (PVC) of the online decoder. In addition to absolute accuracy, I examined the improvement over the online PVC and the variance of each model's performance. This allowed me to assess not only which models achieved higher accuracy but also which ones consistently generalized across subjects and tasks.

Summary

Overall, my offline analysis combined structured feature extraction, several classifier families, and a rigorous evaluation framework to assess how well different decoding strategies improved upon the online system. While the team shared many of the same initial preprocessing steps, my exact preprocessing pipeline, feature extraction steps, and classifier implementations distinguish this pipeline from the others. These choices allowed me to investigate both linear and nonlinear relationships in the EEG, evaluate subject and task variability, and determine which modeling approaches transferred best across MI conditions. The results of this analysis are presented in the next section, where I compare the performance of each classifier against the online decoder and examine the strengths and limitations of each approach.

Analysis Results

Online Classifier Results

The online decoder served as the real-time baseline for all offline comparisons. During each six-second trial, the system attempted to move the ball/cursor toward the correct target based on the participant's SMR modulation, producing one of three outcomes: hit, miss, or abort. Performance followed expected SMR patterns, with left–right trials outperforming up–down trials because contralateral desynchronization is more pronounced during unilateral imagery.

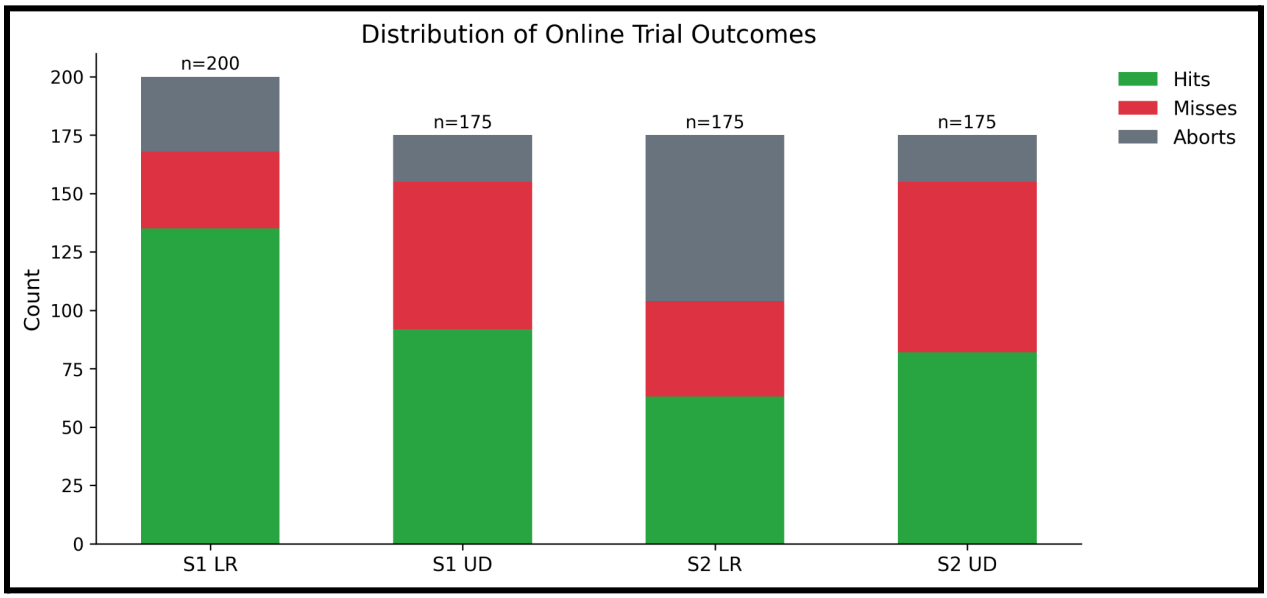

Figure 3. Distribution of hits, misses, and aborts across subjects and tasks.

Across both subjects, the decoder produced more hits than misses, indicating accuracy above chance, and Subject 1 consistently outperformed Subject 2, suggesting stronger SMR expression. To quantify performance, I computed Percent Valid Correct (PVC), which measures accuracy on trials where the decoder made a directional decision, and Percent Total Correct (PTC), which additionally penalizes aborted trials. PVC therefore reflects classification accuracy, while PTC reflects overall reliability.

Subject	Session Type	PTC	PVC
Subject 1	LR	0.675000	0.803571
Subject 1	UD	0.525714	0.593548
Subject 2	LR	0.360000	0.605769
Subject 2	UD	0.468571	0.529032

Table 1. Online PVC and PTC across subjects and tasks.

To integrate these metrics into the analysis consistently, we represent their definitions visually below. This figure appears here because the formulas underpin all subsequent performance comparisons and provide a clear distinction between intent classification and full-trial success.

$$PVC = \frac{Hits}{Hits+Misses} \qquad PTC = \frac{Hits}{Hits+Misses+Aborts}$$

Figure 4. PVC and PTC Equations.

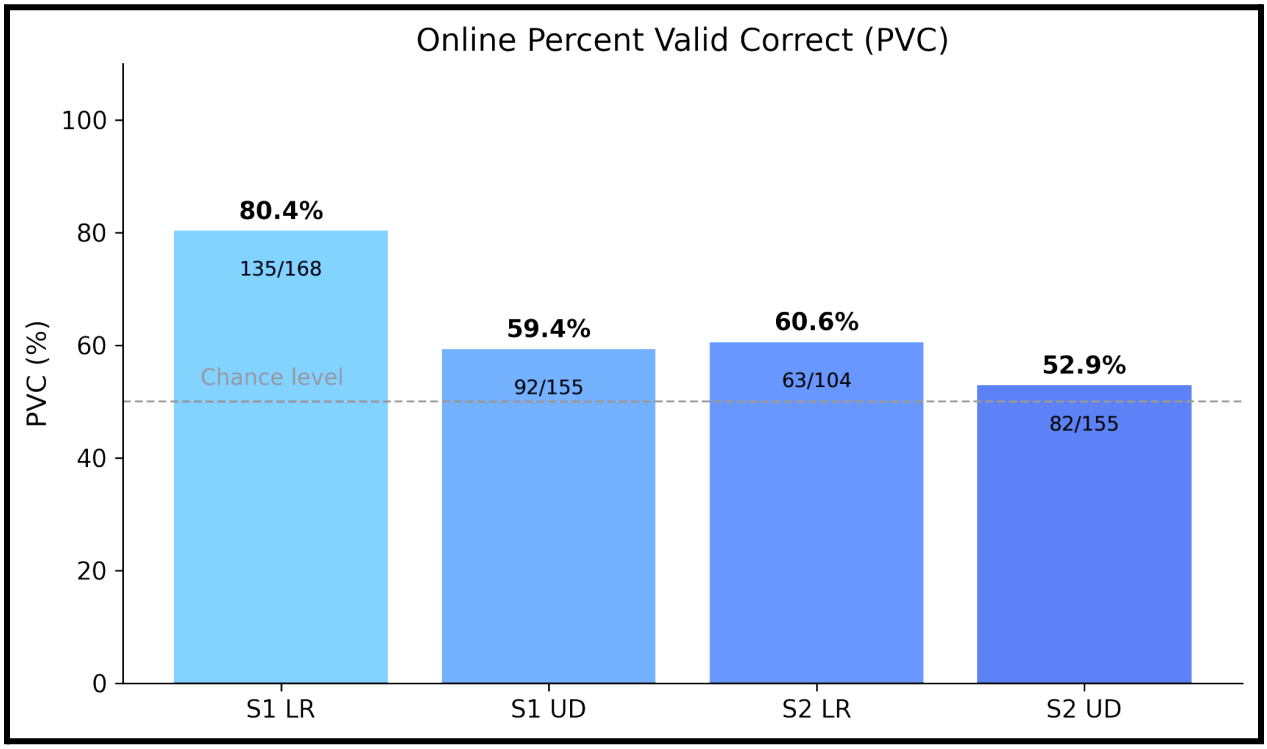

Figure 5. Online PVC values for each subject and task, relative to the 50 percent chance level.

Together, these results establish the baseline that all offline classifiers must exceed to demonstrate meaningful improvement.

Choice of Performance Metric (PVC)

Because the offline classifiers operate strictly as binary decision-makers, they cannot produce an “abort” output. The online decoder, however, generates hits, misses, and aborts, which makes Percent Total Correct

(PTC) unfair for comparison because it penalizes offline models for a category they cannot predict. Percent Valid Correct (PVC) is therefore the only appropriate metric for evaluating offline performance against the online baseline.

PVC reflects pure classification accuracy, while PTC measures real-time control reliability, something offline models are never tasked with. Using PTC would artificially deflate offline accuracy and misrepresent model capability. For consistency and fairness, PVC is adopted throughout this analysis. After removing aborted trials, 329 valid samples remained for cross-validation.

CSP + LDA Baseline Results

As a baseline offline model, I trained a linear discriminant analysis (LDA) classifier on features extracted using the Common Spatial Pattern (CSP) method. Each trial was transformed into twelve log-variance features from six CSP filter pairs, capturing the spatial components that maximized variance differences between the two motor-imagery classes. To avoid information leakage, CSP filters were recomputed within each fold of cross-validation so that spatial patterns were learned strictly from training data. Once preprocessing was finalized, this baseline model reliably outperformed the online PVC across all subjects and tasks, confirming that the data pipeline was sound and establishing a foundation from which more complex models could build.

Across all subjects and tasks, the CSP + LDA model exceeded the online PVC, confirming that the preprocessing pipeline and CSP extraction produced meaningful, physiologically grounded features. This alignment with the online baselines was essential, as it demonstrated that even a simple linear model could outperform the real-time decoder when trained offline on cleanly preprocessed data.

Subject	Session Type	PTC	PVC	CSP+LDA Mean	CSP+LDA Std
Subject 1	LR	0.675000	0.803571	0.83	0.05
Subject 1	UD	0.525714	0.593548	0.68	0.15
Subject 2	LR	0.360000	0.605769	0.68	0.16
Subject 2	UD	0.468571	0.529032	0.53	0.12

Table 2. CSP + LDA Cross-Validation Results.

To assess whether the CSP filters captured meaningful spatial structure, I visualized the most discriminative spatial patterns for each subject and task (Figure 2). Although the exact topographies varied across subjects and paradigms, the dominant components consistently reflected organized activity over central and sensorimotor areas rather than random or diffuse patterns. This indicates that CSP extracted physiologically relevant structure from the EEG, even if the small dataset and inter-subject variability limited the emergence of clearly lateralized SMR patterns.

Additional Offline Classifiers

4.1 MLP Trained on CSP Features

To test whether nonlinear decision boundaries could improve upon linear decoding, I trained a multilayer perceptron (MLP) on the same twelve CSP log-variance features used by the LDA baseline. Because CSP already concentrates the most discriminative spatial information into a low-dimensional feature set, the MLP's role was not to discover new representations but to refine the class boundary beyond what a linear model can express. Across subjects and tasks, the MLP matched or exceeded CSP + LDA performance, with the largest gains appearing in Subject 2's left–right condition. These results indicate that modest nonlinear modeling can

extract additional structure from CSP features without requiring the complexity or data demands of deeper architectures.
(See Table 3.)

4.2 EEGNet Trained on Raw EEG

I implemented EEGNet, a compact convolutional architecture designed for EEG decoding (Lawhern et al., 2018). Unlike the CSP-based models, which rely on handcrafted spatial filters, EEGNet learns both temporal and spatial feature representations directly from the raw 27-channel EEG. This allows the model to capture patterns that may not be fully expressed in variance-based features, particularly subtle temporal dynamics that characterize SMR activity. EEGNet performed strongly across subjects, with especially notable improvements in the up–down task where temporal cues appear more informative than purely spatial ones. Its competitive performance demonstrates that, even with limited data, learned spatiotemporal representations can rival or exceed the effectiveness of manually engineered features.
(Refer to Figure 6 and Table 3.)

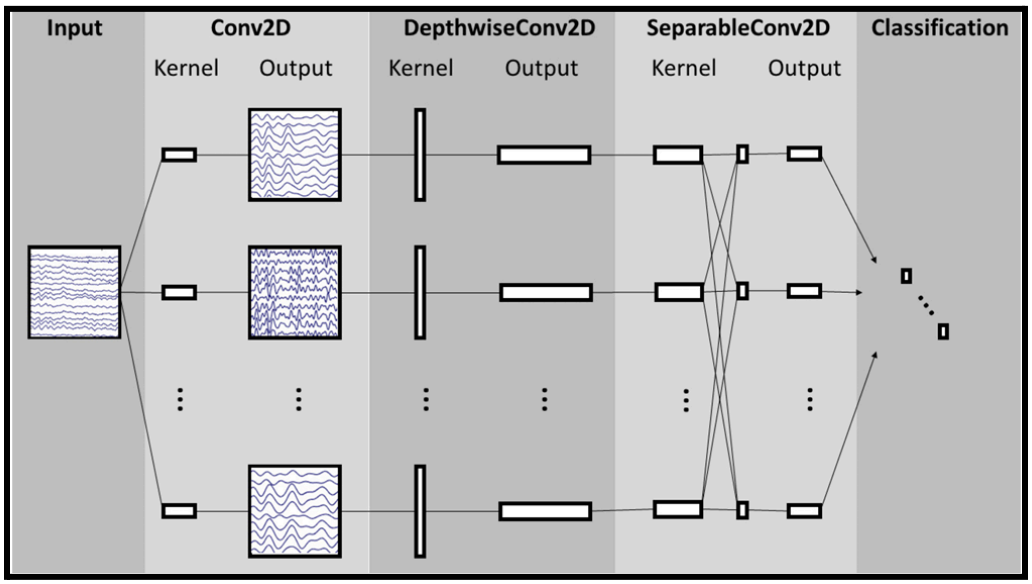

Figure 6. Overall visualization of the EEGNet architecture. Reproduced from Lawhern et al. (2018).

4.3 Random Forest and XGBoost

To broaden the comparison beyond linear and neural network models, I also evaluated two ensemble tree-based classifiers: Random Forest and XGBoost. Both operate on the same CSP-derived feature set as the baseline models, offering a useful contrast in how feature interactions are handled. Random Forest provided steady improvements over the online PVC but did not surpass the performance of the nonlinear MLP or EEGNet.

XGBoost, however, delivered the strongest overall performance across all subject–task combinations, despite receiving far less tuning attention than the deep learning models. Its ability to model nonlinear feature interactions and automatically prioritize informative feature subsets likely contributed to its advantage in this moderate-dimensional, structured feature space. The Summary Statistics Table and the Model Ranking Table reflect this trend clearly, showing XGBoost achieving the highest mean improvement relative to the online classifier.
(Refer to Table 3 and Table 4.)

Combined Model Comparison and Interpretation

To compare performance across all offline classifiers, I evaluated CSP+LDA, CSP+MLP, EEGNet, Random Forest, and XGBoost against the online PVC baseline for each subject and task. Despite their architectural differences and varying levels of tuning, every offline model surpassed the online decoder, demonstrating that offline learning on preprocessed EEG provides a substantial advantage over real-time classification.

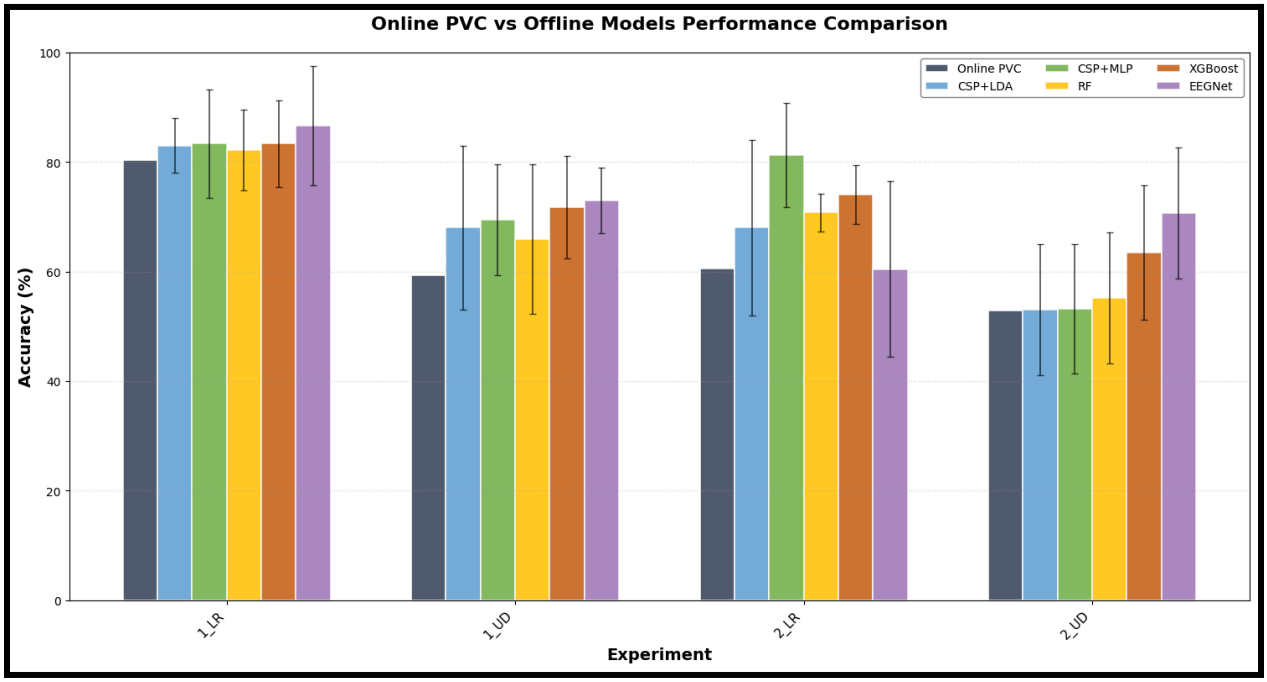

Figure 7. *Comparison of online PVC and offline model performance across all subjects and tasks. Error bars denote cross-validation variability.*

CSP+LDA established a strong linear baseline, and CSP+MLP improved upon it by capturing modest nonlinear structure in the CSP feature space. EEGNet performed competitively with both CSP-based models, particularly in the up–down conditions where temporal patterns in the raw EEG appear more informative than purely spatial filters.

Although Random Forest demonstrated consistent improvements over the online classifier, XGBoost ultimately achieved the highest overall accuracy. This aligns with XGBoost's strength in modeling nonlinear interactions and leveraging structured, moderate-dimensional feature sets—an advantageous match to CSP-derived inputs.

Subject	Task	PTC	PVC	CSP+LDA	CSP+MLP	EEGNet	RF	XGBoost
Subject 1	LR	0.6750	0.8036	0.83	0.8333	0.8667	0.8222	0.8333
Subject 1	UD	0.5257	0.5935	0.68	0.6941	0.7294	0.6588	0.7176
Subject 2	LR	0.3600	0.6058	0.68	0.8121	0.6046	0.7076	0.7409
Subject 2	UD	0.4686	0.5290	0.53	0.5316	0.7063	0.5516	0.6342

Table 3. Summary of PVC across all offline classifiers.

To quantify these differences more explicitly, I computed the mean improvement over online PVC for each model across all subject–task combinations. The ranking is shown below.

Rank	Model	Mean Improvement (%)	Std Improvement (%)
1	XGBoost	9.853	4.115
2	EEGNet	9.378	6.840
3	CSP+MLP	8.480	7.875
4	Random Forest	5.208	3.405
5	CSP+LDA	4.703	3.479

Table 4. Model ranking based on mean improvement over online PVC.

The ranking highlights that no single model dominates universally, but several achieve strong performance when paired with robust preprocessing and CSP-based feature extraction. The consistent improvement across all offline methods reinforces a central pattern in SMR BCI decoding: model choice matters less than the quality of the preprocessing pipeline and the representation of the neural signal.

Limitations of Offline Analysis

Our offline results are strong, but several limitations should be noted. First, the dataset was small—only 329 valid trials—which increases variability and makes complex models more susceptible to overfitting. Second, our cross-validation split was random rather than chronological. Because EEG drifts over time, mixing early and late trials can introduce information leakage and artificially inflate performance. A more realistic evaluation would train on earlier runs and test on later ones to better reflect real BCI generalization. Differences between subjects, such as the performance gap between S1 and S2, also highlight the inherent variability in SMR expression and its influence on model accuracy.

Another important consideration is that all classifiers were trained on the combined set of left–right and up–down trials. These paradigms rely on different spatial patterns—lateralized desynchronization for LR versus more midline involvement for UD—so training a single model across both increases task variability and likely suppresses maximum achievable accuracy. Models trained separately on LR-only or UD-only data would almost certainly perform even higher, meaning the results reported here should be interpreted as a conservative estimate of model capability.

Finally, offline decoding benefits from conditions absent in real-time BCI systems. Models operate on clean, pre-segmented data and face no latency or continuous-control constraints. As a result, offline accuracy should be viewed as an upper bound rather than a predictor of online performance.

Overall, the offline analyses demonstrate that robust preprocessing and well-constructed feature representations enable a wide range of classifiers to outperform the online decoder. Although the models differ in complexity, all captured sufficient SMR structure to decode motor intent with meaningful accuracy, validating both the experimental design and the strength of the offline analysis pipeline.

Summary

Beating an online classifier for SMR motor-imagery EEG was far from straightforward. The project made it clear that strong decoding hinges less on model complexity and far more on disciplined preprocessing and well-chosen features. EEG is noisy, SMR rhythms are subtle, and the brain rarely offers patterns that align

cleanly with a classifier's expectations. Even so, researchers have continued to push the field forward, creating systems that give people with profound physical limitations a way to communicate and act when no other channel is available. Despite being a relatively young technology, SMR-based BCIs have already shown extraordinary potential, hinting at a future where neural interfaces may become an ordinary part of how humans interact with the world.

Access to that future is still limited, but our team was fortunate to reproduce a full SMR motor-imagery experiment and evaluate it with our own offline decoders. We ran two complete EEG sessions, compared online results with models we implemented ourselves, and incorporated state-of-the-art methods across preprocessing, feature extraction, and classification. Each offline pipeline we built outperformed the online decoder, reaffirming a central insight in BCI research: motor intent can be reliably inferred even when no physical movement occurs. With a small team and tight time constraints, we were able to replicate a core phenomenon of the field and demonstrate measurable decoding of intent from EEG alone, underscoring how much can be achieved with careful engineering and a clear analytical framework.

Author Contributions

All team members contributed meaningfully to the project, especially during data collection. Everyone participated in the experimental setup, including EEG cap preparation and electrode gelling. Alexis and I also served as BCI subjects.

For offline analysis, we divided the work by model class. Joseph focused on CSP + LDA and general classifiers. I worked on deep learning approaches, including an MLP that drew from CSP features and EEGNet that drew from raw EEG data. Dheemant implemented a separate CNN-based method that drew from CSP features. JD, Defne, and Alexis concentrated on Random Forest models (CSP/FBCSP-based). Every member contributed to the final presentation: JD covered preprocessing, Defne handled the introduction, I discussed deep learning architectures and their relevance to BCI, Dheemant presented his CNN approach, Alexis and Defne reviewed the Random Forest results, and Joseph discussed feature extraction. Throughout data preparation, Joseph was particularly helpful in organizing and explaining the loading and preprocessing steps. During the Q&A, I contributed significantly to discussion regarding data leakage and the EEGNet implementation.

GenAI Usage

I used ChatGPT throughout the project, but throughout the draft phase of the project, never in a way that substituted for my own understanding. For small tasks, like writing quick plotting snippets, it was simply a convenience. I already know how to do that work by hand and didn't want to spend time on boilerplate. I also used it to polish parts of my data loading and preprocessing code, although I stayed hands-on there because understanding the structure of the data mattered more than any speed boost. For the modeling phase, ChatGPT helped me prototype a wide range of architectures so I could explore the landscape faster than if I had written every version manually. Once I identified the models worth keeping for the final report, I tuned them myself and made sure I understood every component and hyperparameter that ended up in the final pipeline. After the report draft, I used ChatGPT more heavily to tie up loose ends in my report, doing a more thorough analysis: this involved using generative AI to implement FBCSP, topographic CSP maps, an RF and XGBoost model, as well as overall visualizations of performance between models. However, none of the models I included were unfamiliar to me, and I could reproduce them on my own if needed. For writing the report draft, I barely used ChatGPT. Mainly, I used it to tighten up some sections, but I wrote almost the entire thing myself. After I drafted up the report, I relied more heavily on ChatGPT during revision to refine the clarity and scientific tone of the report.

References:

1. Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., & Zhao, X. (2019). *A comprehensive review of EEG-based brain–computer interface paradigms. Journal of Neural Engineering, 16*(1), 011001. https://doi.org/10.1088/1741-2552/aaf12e
2. Bell, A. J., & Sejnowski, T. J. (1995). *An information-maximization approach to blind separation and blind deconvolution. Neural Computation, 7*(6), 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129
3. Britton, J. W., Frey, L. C., Hopp, J. L., St. Louis, E. K., & Frey, L. C. (Eds.). (2016). *Appendix 6. A brief history of EEG.* In *Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants.* American Epilepsy Society. https://www.ncbi.nlm.nih.gov/books/NBK390348/
4. Chandrasekaran, S., Nanivadekar, A., McNamara, C., & Lebedev, M. A. (2021). *Historical perspectives, challenges, and future directions of brain–machine interfaces. Bioelectronic Medicine, 7*(9). https://doi.org/10.1186/s42234-021-00076-6
5. Choi, J. W., Kim, Y. J., Jeong, J. H., & Kim, H. T. (2020). *Observing actions through immersive virtual reality enhances motor imagery training. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28*(7), 1614–1622. https://doi.org/10.1109/TNSRE.2020.2994855
6. Congedo, M., Barachant, A., & Bhatia, R. (2017). *Riemannian geometry for EEG-based brain–computer interfaces: A primer and a review. Brain–Computer Interfaces, 4*(3), 155–174. https://doi.org/10.1080/2326263X.2017.1297192
7. Edelman, B. J., Meng, J., Suma, E. A., Zurn, C., Nagarajan, S. S., & He, B. (2019). *Noninvasive neuroimaging enhances continuous neural tracking for robotic device control. Science Robotics, 4*(31), eaaw6844. https://doi.org/10.1126/scirobotics.aaw6844
8. He, B. (2025). *2 BCI definition and structure* [Lecture slides]. *Introduction to Neural Engineering.* Carnegie Mellon University.
9. Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). *EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. Journal of Neural Engineering, 15*(5), 056013. https://doi.org/10.1088/1741-2552/aace8c
10. Leuthardt, E. C., Decuir, J., Farahmand, S., Garabedian, L., Sharma, M., Moran, D. W., & Alomar, S. (2021). *Defining surgical terminology and risk for brain–computer interfaces. Journal of Neural Engineering, 18*(6), 066006. https://doi.org/10.1088/1741-2552/ac35d0
11. Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G., & Moran, D. W. (2004). *A brain–computer interface using electrocorticographic signals in humans. Journal of Neural Engineering, 1*(2), 63–71. https://doi.org/10.1088/1741-2560/1/2/001
12. Mason, S. G., & Birch, G. E. (2003). *A general framework for brain–computer interface design. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11*(1), 70–85. https://doi.org/10.1109/TNSRE.2003.810426
13. Müller-Gerking, J., Pfurtscheller, G., & Flyvbjerg, H. (1999). *Designing optimal spatial filters for single-trial EEG classification in a movement task. Clinical Neurophysiology, 110*(5), 787–798. https://doi.org/10.1016/S1388-2457(98)00038-8
14. Pfurtscheller, G., & Aranibar, A. (1977). *Event-related cortical desynchronization detected by power measurements of scalp EEG. Electroencephalography and Clinical Neurophysiology, 42*(6), 817–826. https://doi.org/10.1016/0013-4694(77)90235-8
15. Ramoser, H., Müller-Gerking, J., & Pfurtscheller, G. (2000). *Optimal spatial filtering of single-trial EEG during imagined hand movement. IEEE Transactions on Rehabilitation Engineering, 8*(4), 441–446. https://doi.org/10.1109/86.895946

16. Saha, S., Mamun, K. A., Ahmed, K., Mostafa, R., Naik, G. R., Darvishi, S., Khandoker, A. H., & Baumert, M. (2021). *Progress in brain–computer interface: Challenges and opportunities. Frontiers in Systems Neuroscience, 15*, 578875. https://doi.org/10.3389/fnsys.2021.578875
17. Wolpaw, J. R., & McFarland, D. J. (2004). *Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans. Proceedings of the National Academy of Sciences of the United States of America, 101*(51), 17849–17854. https://doi.org/10.1073/pnas.0403504101
18. Wolpaw, J. R., McFarland, D. J., Neat, G. W., & Forneris, C. A. (1991). *An EEG-based brain–computer interface for cursor control. Electroencephalography and Clinical Neurophysiology, 78*(3), 252–259. https://doi.org/10.1016/0013-4694(91)90061-5
19. Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). *Brain–computer interfaces for communication and control. Clinical Neurophysiology, 113*(6), 767–791. https://doi.org/10.1016/S1388-2457(02)00057-3
20. Woodman, G. F. (2010). *A brief introduction to the use of event-related potentials in studies of perception and attention. Attention, Perception, & Psychophysics, 72*(8), 2031–2046. https://doi.org/10.3758/APP.72.8.2031
21. Yuan, Y., & He, B. (2014). *Brain–computer interfaces using sensorimotor rhythms: Current state and future perspectives. Frontiers in Neuroengineering, 7*, 27. https://doi.org/10.3389/fneng.2014.00027
22. Zhang, Z., Li, W., Han, L., & Zhang, X. (2024). *Bibliometric and visual analysis of brain–computer interface research in rehabilitation (2004–2024). Frontiers in Human Neuroscience, 18*, 1486167. https://doi.org/10.3389/fnhum.2024.1486167