WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

EEG-Based Brain–Computer Interface for Motor Imagery Classification: Challenges, Applications, and Future Directions

Simran Kumari[1], Neelam Batham[2]

[1] School of Computer Applications Babu Banarasi Das University, Lucknow rai420171@gmail.com

[2] (Assistant Professor) School of Computer Applications Babu Banarasi Das University, Lucknow neelam.mca01@bbdu.ac.in

DOI : https://doi.org/10.5281/zenodo.23829732

ABSTRACT—

A brain–computer interface (BCI) allows a person to operate a computer or assistive device through brain activity rather than muscle movement. Electroencephalography (EEG) is widely used for non-invasive BCI work because it records electrical activity at the scalp with fairly portable equipment. In this paper we review EEG-based motor imagery (MI) classification: how signals are acquired, cleaned, converted into features and classified, where the technology can be applied, and why it remains difficult. Imagining a left- or right-hand movement can produce sensorimotor rhythms that a system can distinguish, but doing so reliably is hard. Noise, artifacts, differences between users, small labelled datasets, drift between sessions and limited computing power all get in the way. We present a conceptual pipeline that links preprocessing, feature representation and machine-learning or deep-learning classification. The paper rests on the literature alone. The pipeline has not been tested experimentally, so we do not claim any classification accuracy. Future work should implement it on a suitable public EEG dataset, using reproducible protocols and standard evaluation metrics.

***Index Terms*—**EEG, brain–computer interface, motor imagery, signal processing, feature extraction, machine learning, deep learning.

I. INTRODUCTION

A brain–computer interface gives a person a channel to an outside system by translating measurable brain activity into a command. Rather than depending only on muscles, a BCI can use neural patterns to select an item, move a cursor, drive a simulated device or support assistive technology [2]. EEG is a common non-invasive choice. Electrodes on the scalp pick up tiny changes in electrical potential, and they follow those changes with very good timing. The equipment is usually cheaper and easier to carry than most other brain-imaging tools. The price of this convenience is that the signals are weak and spatially blurred.

Motor imagery means mentally rehearsing a movement without carrying it out. A participant might imagine moving the left or right hand while staying perfectly still. Doing so changes sensorimotor rhythms, particularly activity in the mu and beta bands. These changes differ from person to person, and even in one person they shift between trials and between sessions. A classification system has to find the task-related pattern and turn it into an output that can be trusted.

In practice the classifier is only one part of the story. Electrode placement, signal quality, cue timing, preprocessing, feature representation, training strategy and feedback design all shape what comes out at the end. Eye blinks, eye movements, muscle activity, mains interference, electrode shifts and fatigue can bury the activity we are looking for. A model trained on one person also tends to perform worse on another unless it is calibrated or adapted.

This paper offers a literature-based overview and a conceptual implementation workflow for EEG-based MI classification. It has four aims: to describe the main stages of an MI-BCI pipeline, to summarize common feature-extraction and classification approaches, to discuss applications and practical limits, and to outline a reproducible evaluation plan for later implementation. The contribution is organizational and methodological. We make no claim of new experimental performance.

II. RELATED WORK

This section reviews earlier papers one by one: who wrote them, what they set out to do, and what they found. Papers are grouped as foundations of motor imagery BCIs, classical decoding methods, deep-learning work, and recent work on running BCIs on small devices. Where a paper is a review, we say so, because a review reports no new accuracy of its own. For papers where only the abstract and the survey summary below were available to us, we describe only what they state.

Foundations (Wolpaw et al., 2002) [1]. Jonathan Wolpaw and colleagues wrote one of the best-known early reviews of brain–computer interfaces for communication and control, published in Clinical Neurophysiology. It explains what a BCI is, which brain signals can drive one, how a system is built and how its performance should be judged, and it lists the obstacles that kept the technology from wider use. It is a review, so it reports no experiment and no accuracy figure. Its value for us is that it sets the vocabulary and the overall framework that later motor imagery work builds on.

Brain rhythms (McFarland et al., 2000) [5]. McFarland, Miner, Vaughan and Wolpaw compared the mu and beta rhythm topographies recorded over the scalp during actual and imagined movement, in Brain Topography. The study supports the physiological idea behind motor imagery BCIs: imagined movement changes rhythms over sensorimotor areas in a way that can be measured on the scalp. It is a physiology study and not a classifier paper, so it gives no classification accuracy.

Common Spatial Patterns (Ramoser et al., 2000) [3]. Ramoser, Müller-Gerking and Pfurtscheller applied spatial filtering to multichannel EEG recorded while people imagined left- and right-hand movement. The filters are learned from training trials so that the two classes differ as much as possible in variance, and this made the classes easier to tell apart than raw channels. The paper is the usual reference for CSP in MI-BCI. Later work, including [4], points out that CSP is sensitive to noise, small training sets and distribution shift.

Filter Bank CSP (Ang et al., 2008) [6]. Kai Keng Ang, Zheng Yang Chin, Haihong Zhang and Cuntai Guan addressed a weakness of plain CSP, namely that the best frequency band differs from person to person. Their Filter Bank CSP splits the EEG into several sub-bands, learns spatial filters in each, and then selects the most useful features automatically. It became a common reference method, and newer deep models are often compared against it.

A decade of classifiers (Lotte et al., 2018) [4]. Fabien Lotte and co-authors updated an earlier review of classification algorithms for EEG-based BCIs, in the Journal of Neural Engineering, covering the ten years since the first edition. They grouped newer methods into adaptive classifiers, matrix and tensor classifiers, transfer learning and deep learning, and gave guidance on which to try for which situation. As a review it has no single accuracy figure. Its main message for our work is that reported results only make sense together with the dataset and the validation protocol.

Deep ConvNets (Schirrmeister et al., 2017) [7]. Robin Tibor Schirrmeister and colleagues, in Human Brain Mapping, designed deep and shallow convolutional networks that learn directly from raw EEG. They studied how design choices such as batch normalization and cropped training affect decoding, and used visualization to see what the networks relied on. On movement-related decoding, their networks reached a mean accuracy of 84.0% against 82.1% for FBCSP, so the deep model matched or slightly beat the classical one, but only by a small margin.

EEGNet (Lawhern et al., 2018) [8]. Vernon Lawhern and colleagues proposed EEGNet in the Journal of Neural Engineering, a compact convolutional network that uses depthwise and separable convolutions to keep the number of parameters very small. They tested it on four kinds of EEG data (P300 responses, error-related negativity, movement-related cortical potentials and sensorimotor rhythms) and reported performance comparable to established methods, also when little training data was available. Its small size is the reason many later embedded BCI studies start from it.

Two reviews of deep learning for EEG (Roy et al., 2019; Craik et al., 2019) [9], [10]. Yannick Roy and colleagues reviewed 156 papers that used deep learning for EEG. They reported a median accuracy gain of about 5.4% over traditional baselines, and also noted that the amount of detail given about data and methods varied a lot, which makes studies hard to reproduce. Craik, He and Contreras-Vidal reviewed deep learning for EEG classification across tasks such as motor imagery, and found that convolutional, recurrent and deep belief networks generally did better than stacked autoencoders and multilayer perceptrons. Both reviews support the same cautious reading: deep models help, but how much depends on the data and the evaluation.

Edge AI for BCIs (Nguyen et al., 2025) [11]. The most relevant recent paper for this work is a survey by Manh-Dat Nguyen, Thomas Do, Xuan-The Tran, Quoc-Toan Nguyen and Chin-Teng Lin of the University of Technology Sydney, published in IEEE Transactions on Neural Systems and Rehabilitation Engineering (vol. 33, 2025, pp. 4051–4066). The authors ask how a BCI can run its signal processing and AI model on the wearable itself instead of on a PC or in the cloud. They first treat the recording chain: the number of channels, the sampling rate (at least 250 samples per second for research), the ADC resolution (16 or 24 bits), and the rule that recording, cleaning and the AI decision for one trial must finish within the trial time. They then discuss electrode contact monitoring and the choice between active and passive electrodes, compare feature-based methods with end-to-end deep learning, and compare five tools that shrink models for small chips (ESP-DL, STM32.AI, TFLite Micro, Edge Impulse and CMSIS-NN). A table of earlier on-device studies is the survey's main evidence. It lists, for example, driver drowsiness detection at 96.11% accuracy and 25.44 ms, motor imagery at 90.88% and 21.5 ms, person identification from EEG at 99.74% and 4.58 ms, 13-class imagined speech at 42.8% and 40.9 ms, and eye-blink robot control at 99.3% and 200 ms. The authors conclude that the whole chain, from electrode contact to the final decision, has to be designed together. The survey also names applications in robotics, emotion recognition, medicine, consumer devices and industry. It gathers reported results and does not run a new experiment, and the studies it lists use different tasks and data, so their numbers cannot be ranked against each other.

Studies on small devices [12]–[15]. Four papers from the survey's reference list are closest to motor imagery on small hardware. Schneider, Wang, Hersche, Cavigelli and Benini built Q-EEGNet, an 8-bit quantized and parallel version of EEGNet for edge motor-imagery interfaces [12]. Mei and

colleagues describe a wearable system whose lightweight CNN, MI-BMInet, is summarized in [11] as reaching 90.88% accuracy with about 21.5 ms latency and 0.45 mJ per inference on a GAP9 microcontroller [13]. Cao and colleagues designed a dedicated EEGNet processor for low-power, real-time wearable classification [14], and Bian and colleagues studied on-device learning of an EEGNet-style network for a wearable motor-imagery BCI [15]. We know these four papers through their titles and the survey's summary, so we do not repeat figures beyond those the survey reports.

Research gap. Read together, these papers leave a practical gap. Accuracy-focused work on CSP and deep networks [6]–[10] says little about what a model costs to run, while the embedded studies [11]–[15] report latency and energy but use different boards, datasets and validation schemes, which makes them hard to compare. Preprocessing, representation, classification, adaptation, efficiency and safety are rarely treated as one linked design problem. This paper lays those decisions side by side. We do not propose a new algorithm or claim an experimentally proven improvement.

TABLE I

SELECTED RELATED STUDIES. FIGURES ARE THOSE REPORTED BY THE CITED AUTHORS, NOT RESULTS OF THIS PAPER.

Ref.	Study	Focus	Main takeaway
[5]	McFarland et al.	Mu/beta rhythms in imagined and real movement	Physiological basis for MI decoding
[6]	Ang et al.	Filter Bank CSP	Automatic, subject-specific band and feature selection
[7]	Schirrmeister et al.	Deep ConvNets on raw EEG	Comparable to FBCSP (84.0% vs 82.1%)
[8]	Lawhern et al.	EEGNet	Very small CNN that works across paradigms
[9]	Roy et al.	Review of 156 DL-EEG studies	Median gain of about 5.4% over baselines; uneven reporting
[10]	Craik et al.	Review of DL for EEG classification	CNN, RNN and DBN tend to outperform autoencoders and MLP
[11]	Nguyen et al.	Edge AI-BCI survey	Acquisition, model and latency need co-design
[12]	Schneider et al.	Q-EEGNet	8-bit quantized EEGNet for edge MI-BCI
[13]	Mei et al.	Wearable BMI with lightweight CNN	90.88% accuracy, about 21.5 ms (as summarized in [11])
[14]	Cao et al.	EEGNet processor	Low-power, real-time wearable classification
[15]	Bian et al.	On-device learning, EEGNet-style network	Wearable MI-BCI adaptation on the device

III. PROPOSED METHODOLOGY

This study is a literature-based review. Instead of reporting a finished experiment, it proposes a conceptual workflow. The stages below are meant to guide a future implementation and to make the assumptions of the pipeline explicit (Fig. 1: EEG acquisition → preprocessing → feature extraction → classification → decision/command → feedback/adaptation; a conceptual architecture, not an experimentally validated one).

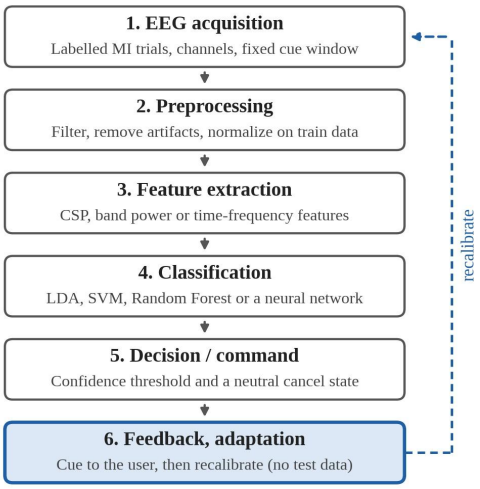


Fig. 1. Conceptual MI-BCI pipeline (not an experimentally validated system). Stages 1 to 3 turn raw trials into features, stage 4 assigns a class, stage 5 decides whether to act on it, and stage 6 returns feedback and recalibration to the start. Any tuning inside this loop must use training data only.

A. EEG Acquisition and Trial Definition

The first step is to pick a labelled motor imagery dataset that documents its channels, sampling frequency, class labels, trial boundaries and recording conditions. Participants are usually cued to imagine left- or right-hand movement, with a rest period between trials. The cue period and analysis window must be defined identically throughout. If a public dataset is used, its version and any licensing or usage terms should be noted. Channels should be chosen on the basis of the task and the available montage, and not after looking at test results.

B. Preprocessing and Artifact Handling

Preprocessing can include band-pass filtering, notch filtering where needed, trial segmentation, baseline correction and normalization. Filtering removes irrelevant frequencies but also alters the signal, so filter limits and parameters should always be reported. Eye blinks and facial muscle activity produce large artifacts. They can be handled by automatic detection, careful rejection or established artifact-reduction methods. Whichever route is taken, it should not discard trials in a way that unbalances the classes. Normalization parameters must be estimated from training data only and then applied to held-out data. Otherwise information leaks in from the test set.

C. Feature Representation

Features should preserve the information that separates the imagery classes. CSP builds spatial features from the covariance structure of the EEG and is often paired with LDA or SVM. Band-power features summarize energy in chosen frequency ranges and are easy to inspect. Time-frequency representations show how activity changes over time and frequency, but they can inflate dimensionality considerably. Neural networks can learn representations directly from the input, though they usually need careful regularization and enough training examples. Any feature selection or parameter tuning has to stay inside the training and validation process. *D. Classification and Decision Output*

Candidate classifiers include LDA, SVM, Random Forest and neural networks. CSP with LDA is a convenient baseline because it is simple and gives a reference point for judging more complex methods. Predicted classes can then be mapped to a simulated cursor movement, an interface selection or an assistive command. A system intended for real use should not act on every uncertain prediction immediately. Confidence thresholds, a neutral or no-control state and a safe way to cancel all help reduce unintended actions.

E. Evaluation Design

Before any model is trained, a future evaluation should decide what kind of generalization it wants to measure. Within-session evaluation estimates performance on held-out trials from a similar recording. Cross-session evaluation tests how well a model copes with change over time. Cross-subject

evaluation checks transfer to a different person. These protocols answer different questions and should not be treated as interchangeable. Hyperparameters should be chosen on training and validation data only, and the test set should be used once, for the final report. Class balance, random seeds, preprocessing settings, software versions and any exclusions should be documented.

F. Practical Considerations for Data Quality

Data quality should be checked before a classifier is trained. Useful checks look for missing channels, unusually flat channels, clipping, inconsistent sampling rates, duplicated trials and unexpected class counts. They do not replace signal processing, but they catch recording and file-handling mistakes that would otherwise look like a machine-learning problem. The quality-control rules should be fixed before test results are seen and applied consistently.

Trial segmentation is another important decision. A window that begins too early may capture visual processing of the cue instead of the motor imagery itself, while one that is too short may miss useful changes in sensorimotor rhythms. The best interval depends on the dataset protocol and the research question, so it is better to record the chosen interval and compare alternatives within the training procedure than to assume one window suits every dataset.

Class balance and differences between participants also deserve a look. If one class has far more trials than the other, accuracy alone gives an incomplete picture. Stratified splits help preserve class proportions in within-subject experiments, whereas subject-wise evaluation may be needed when the goal is to generalize to new people. In every case the splitting strategy should match the setting in which the system will eventually be used.

G. Deployment Constraints for On-Device Inference

If the classifier is meant to run on a wearable board instead of a PC, the earlier choices have to be revisited. Following the system view in [11], acquisition, preprocessing and inference must together finish within the length of one trial:

$$T + T + T \leq T \qquad (1)\ acq\ pre\ inf\ trial$$

The amount of data a trial produces grows with channel count, sampling rate and ADC resolution, so shorter trials, more channels or faster sampling all squeeze the time left for inference. The same survey recommends at least 16-bit conversion and a sampling rate of no less than 250 Hz for research-grade EEG [11]. Since mu and beta activity lies below roughly 30 Hz, that is not far from what MI work needs, which is one reason low-channel, low-rate setups suit embedded MI-BCIs.

Model size matters too. Compact networks and 8-bit quantization [8], [12] are the usual ways to cut memory use and multiply-accumulate load, and any accuracy drop they cause should be measured instead of assumed. For a future study we therefore suggest reporting, next to accuracy, the parameter count, model memory, per-trial latency on the target hardware and, where possible, energy per inference.

IV. EXPERIMENTAL RESULTS

No experiment was carried out for this paper. The proposed workflow is conceptual and has not been run on newly collected or public EEG data, so no accuracy, precision, recall, F1-score or latency is measured here. Table II and Figs. 2 and 3 only restate values that the reviewed papers reported themselves, mainly the survey by Nguyen et al. [11], plus [7] and [9]. These studies used different tasks, data and testing methods, so their numbers must not be compared directly with each other.

TABLE II RESULTS REPORTED IN THE REVIEWED PAPERS (NOT PRODUCED BY THIS PAPER)

Task / comparison	Source	Accuracy	Response
Raw-EEG movement decoding: Deep ConvNet vs FBCSP (mean)	[7]	84.0% vs 82.1%	–
156 deep-learning EEG studies: median gain over baselines	[9]	+5.4%	–
Driver drowsiness, four-sensor headband	[11]	96.11%	25.44 ms
Drowsiness, brain + motion sensors	[11]	83%	Not reported
Motor imagery	[11], [13]	90.88%	21.5 ms

Person ID from EEG, smart glasses	[11]	99.74%	4.58 ms
Imagined speech, 13 classes	[11]	42.8%	40.9 ms
Eye-blink robot control	[11]	99.3%	200 ms

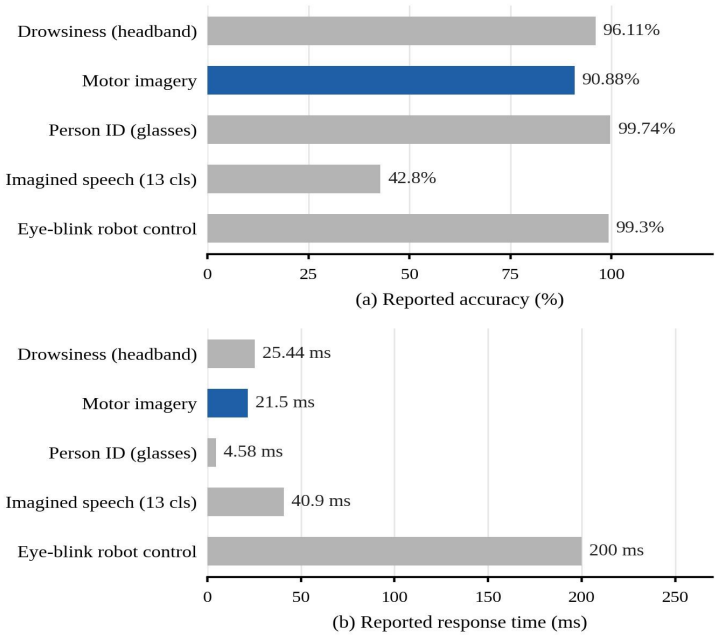


Fig. 2. **Accuracy (a) and response time (b) reported in the survey [11] for five on-device BCI tasks. Values are quoted, not re-measured; motor imagery is highlighted.**

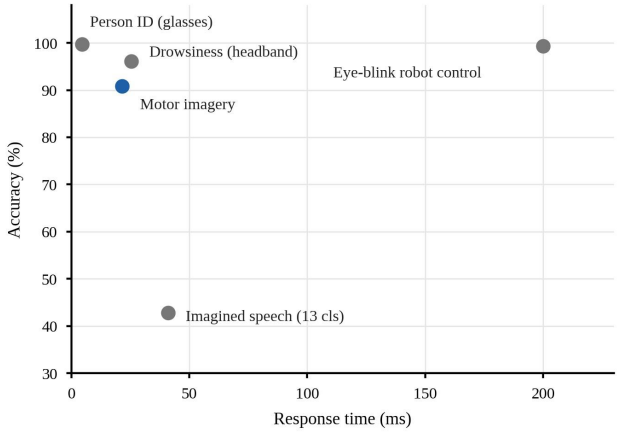


Fig. 3. **Reported accuracy versus response time. Different tasks and datasets; shown for context only.**

Taken together, the quoted studies suggest that several useful BCI tasks can already run on small, low-power wearable devices with response times of roughly 4 to 41 ms, while a hard task such as 13-class imagined speech remains much less accurate. The eye-blink system is slower (200 ms) but accurate, which shows that latency and accuracy depend on the task and hardware. Among classical and deep decoders, the reviewed work shows only small average gains for deep models over FBCSP [7] and a median gain of about 5.4% across many studies [9], so the benefit is real but modest and depends on the data.

A future experiment should compare at least one classical baseline with one alternative pipeline on the same data split. Useful metrics include accuracy, precision, recall, F1-score, a class-wise confusion matrix and inference time. When classes are imbalanced, macro-averaged metrics and balanced accuracy add information. Results should also show how much performance varies across folds, participants or sessions wherever the design allows.

A. Metrics for a Future Empirical Study

Accuracy is the share of trials classified correctly, but it should not be used alone. Precision tells us how often a predicted class is actually right, and recall tells us how many trials of a class were found. The F1-score combines the two through their harmonic mean. A confusion matrix shows which classes get mixed up, and inference latency shows how quickly a prediction can be produced. For a system that issues commands, the cost of false activations can matter as much as the overall classification rate.

Metrics should fit the experimental design. In binary leftversus right-hand imagery, class-wise precision and recall can expose an uneven error pattern. Macro-averaging gives each class equal weight, which helps when the class distribution is uneven. With several participants, a single pooled score can hide users for whom the system works poorly, so the mean, the spread and per-participant results give a fuller picture. Confidence intervals or repeated splits can show uncertainty where the data allow it.

Latency should cover every processing step, not just the classifier's prediction call. Filtering, feature extraction, data transfer and any smoothing or thresholding logic all add delay between the end of a trial and the output. Offline estimates are fine early in development, but real-time performance should eventually be checked on the target hardware. These are proposed evaluation requirements, not results from this paper.

V. DISCUSSION

The workflow highlights that robust MI classification is a system-level problem. Even a strong classifier cannot recover information lost to poor electrode contact or heavy artifacts. Likewise, an elaborate feature representation may add complexity without improving generalization when data are limited. The order and consistency of the processing stages matter as much as the final algorithm.

Classical approaches remain useful when training data are modest and a transparent baseline is needed. CSP with LDA or SVM can be fast, and the feature-generation stage can be inspected. However, CSP fits best in carefully defined classification settings and can become unstable when covariance estimates are noisy. Regularization, channel selection and validation-based parameter choices may help, but they should be justified by experiments rather than assumed to guarantee improvement.

Deep learning offers a different trade-off. By learning features from examples, it may capture interactions that are hard to specify by hand. Its performance, though, can depend heavily on how much labelled data is available and how varied it is. It may also need more computing resources and make it harder to explain why a particular trial received a particular label. A fair comparison should therefore consider data requirements, inference speed, stability and interpretability, and not only predictive scores.

User variability and session drift matter most for real-world use. Signals can change because of electrode placement, skin contact, fatigue, attention, learning, or natural differences in how people imagine movement. Calibration and adaptive learning may improve usability, but adaptation has to be tested carefully so that it neither uses future test information nor destabilizes the model. An assistive system should give understandable feedback and a safe way to stop or ignore uncertain commands. Fig. 4 summarizes these interconnected challenges: noise and artifacts, latency and compute, subject variability, session drift, interpretability and limited labels all bear on reliable EEG-MI classification.

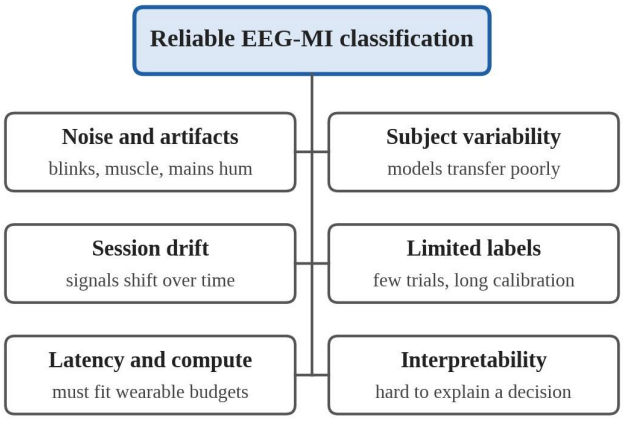


Fig. 4. Interconnected challenges affecting robust EEG motor imagery classification. No single challenge can be fixed alone: noisy signals make small datasets worse, drift erodes calibration, and a heavier model that copes better may no longer fit the latency budget of a wearable.

A. Applications and Deployment Context

Assistive communication is one possible application. A user could select letters, icons or ready-made messages through repeated mental imagery trials, which might offer another way to access devices for people who find conventional input difficult. The real benefit still depends on the individual's ability, fatigue, training demands and how reliable the system is. The design should let the user cancel a selection and should never treat a low-confidence prediction as a definite command.

Rehabilitation research may use BCI feedback as part of a wider training programme. Here the timing and relevance of feedback can matter as much as the predicted class. A system might show a visual cue or move a simulated limb in response to a classification decision. Even so, any clinical conclusion needs proper validation and cannot rest on an offline classification score alone.

MI-BCI ideas have also been explored for cursor control, smart-environment interaction, and robotic or prosthetic control. These settings carry different safety needs. A lab cursor task can tolerate occasional mistakes that would be unacceptable for a physical robot or a mobility device. Command confirmation, conservative thresholds, emergency-stop mechanisms and supervised testing therefore become important when moving from offline research to interactive hardware.

B. Reproducibility, Ethics, and Responsible Reporting

Reproducibility takes more than naming the dataset and the classifier. Researchers should describe the trial window, channel set, filtering parameters, artifact policy, normalization method, feature configuration, hyperparameter search and exact split logic. Software versions should be recorded. These details make it easier to separate real algorithmic differences from differences in preprocessing or evaluation.

EEG is personal physiological data, so data governance matters even in non-medical research. Researchers should respect dataset access conditions, consent restrictions, institutional requirements and applicable privacy rules. When collecting new data, the protocol should state what is recorded, how it is stored, who can access it and whether it will be shared. This is even more important when the system is intended for a vulnerable or clinical population.

Responsible reporting also means keeping measured results separate from expectations. A conceptual diagram can explain the planned architecture and a qualitative comparison can clarify design trade-offs, but neither is evidence that a model reached a certain accuracy. This is why the paper labels its diagrams as conceptual and reports no invented numbers. Once the implementation is complete, the charts should be replaced or supplemented with results produced by the stated evaluation protocol.

VI. LIMITATIONS AND FUTURE WORK

The main limitation is that this is a literature-based conceptual review and not a completed empirical study. The sources discussed are selected foundational and review works, not an exhaustive systematic review. No new EEG data were collected, no classifier was trained, and no numerical performance result is available. The workflow should be read as an implementation plan, not a validated system.

Future work should implement the workflow on a suitable public motor imagery dataset and publish the exact preprocessing and evaluation settings. A good first experiment would compare CSP-LDA and CSP-SVM with a clearly specified deep-learning baseline, using subject-wise or session-wise splits where possible. Ablation studies could show whether filtering, channel selection, feature choice or adaptation drives performance. Confidence intervals and per-subject results would be more informative than a single average score.

Other directions include transfer learning between users, lightweight models for portable devices, artifact-resistant feature learning, interpretable predictions, and adaptive calibration that keeps the burden on the user low. Real-world studies should also examine comfort, training time, accessibility, the cost of false commands and user trust. In assistive applications, a slightly lower offline score may be perfectly acceptable if the system is more stable, responsive and safe.

On-device deployment deserves its own line of future work.

A natural next step is to run the same CSP-LDA, CSP-SVM and EEGNet-style baselines on a microcontroller-class board and report latency and memory next to accuracy, in the spirit of the embedded studies surveyed in [11]. Quantization-aware training [12] and on-device adaptation [15] are two options worth testing against session drift.

VII. IMPLEMENTATION

A modular implementation can be divided into data loading, preprocessing, feature extraction, model training, evaluation and command simulation. First, load the labelled EEG trials and inspect channel names, sampling rate, class distribution and trial lengths. Next, apply filtering and artifact handling with a documented configuration. Then generate features using CSP or spectral representations, or prepare the data for a neural model. Fit the classifier on the training partition only, tune settings on validation data, and keep the test set for the final evaluation.
A reproducible implementation should record the dataset identifier and version, software libraries, filter settings, channel list, trial window, random seed, classifier hyperparameters and splitting method. The pipeline should also check for leakage, such as normalizing all trials before splitting or selecting features on the whole dataset. When comparing algorithms, use identical partitions and report the same metrics for each. The last stage can map a predicted label to a simulated interface command, well before any physical device is considered. The intended sequence is:

1. Load labelled EEG trials and metadata.
2. Split the data into training, validation and test sets.
3. Fit preprocessing parameters on training data.
4. Transform trials and extract CSP or spectral features.
5. Tune candidate classifiers on validation data.
6. Evaluate once on the held-out test set.
7. Report metrics, confusion matrix, latency and limitations.

A. Suggested Experiment Sequence

A first implementation can deliberately stay small. The initial experiment should use one public dataset and a binary motor imagery task with clearly documented class labels. A classical baseline such as CSP-LDA can be evaluated with a subject-specific split. Once the data pipeline is verified, a second model such as CSP-SVM can be tested on exactly the same partitions. A deep-learning model should be added only when its input format, training needs and computational cost can be supported by the available data and hardware.

The experiment should proceed in stages: verify the raw data, define the task and splits, fit preprocessing and feature parameters on training data, tune models on validation data, freeze the configuration, and evaluate on held-out trials. Error analysis comes after the final evaluation and should include the confusion matrix and any variation between participants or sessions. If the test results suggest changes, those changes should be checked on a new hold-out set or through fresh nested validation, not by repeatedly optimizing against the same test data.

The final report should say plainly what was tested and what remains unknown. If only within-session results exist, conclusions should stay within that setting and not claim cross-user generalization. If only simulated commands were used, the paper should not claim real-world device control. This staged approach gives a practical route from the present conceptual framework to a defensible experimental study.

VIII. CONCLUSION

EEG-based brain–computer interfaces offer a non-invasive way to turn brain activity into commands, and motor imagery is an important paradigm for doing so. Reliable classification remains difficult because EEG signals are noisy, vary across users and sessions, and often have to be processed with little labelled data and low latency. This paper has set out a conceptual workflow linking acquisition, preprocessing, feature extraction, classification and command generation, and has discussed the trade-offs between classical machine learning and deep learning.

The framework has not yet been validated experimentally, so we claim no performance advantage or numerical accuracy. The next step is a reproducible implementation on a documented EEG dataset, with transparent data splits and metrics that capture both predictive quality and practical efficiency. For MI-BCI systems to be useful outside controlled experiments, robustness, safety, interpretability and user experience must be weighed alongside classification scores.

IX. REFERENCES

[1] J. R. Wolpaw et al., “Brain-computer interfaces for communication andcontrol,” Clinical Neurophysiology, vol. 113, no. 6, pp. 767–791, 2002.

[2] B. Blankertz et al., “The Berlin Brain–Computer Interface: Non-medicaluses of BCI technology,” Frontiers in Neuroscience, vol. 4, art. 198, 2010. (Please verify against the original source before submission.)

[3] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, “Optimal spatialfiltering of single trial EEG during imagined hand movement,” IEEE Transactions on Rehabilitation Engineering, vol. 8, no. 4, pp. 441–446, 2000.

[4] F. Lotte et al., “A review of classification algorithms for EEG-basedbrain-computer interfaces: A 10 year update,” Journal of Neural Engineering, vol. 15, no. 3, 2018.

[5] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, “Muand beta rhythm topographies during motor imagery and actual movements,” Brain Topography, vol. 12, no. 3, pp. 177–186, 2000.

[6] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank commonspatial pattern (FBCSP) in brain-computer interface," in Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN), 2008.

[7] R. T. Schirrmeister et al., "Deep learning with convolutional neuralnetworks for EEG decoding and visualization," Human Brain Mapping, vol. 38, no. 11, pp. 5391–5420, 2017.

[8] V. J. Lawhern et al., "EEGNet: A compact convolutional neuralnetwork for EEG-based brain–computer interfaces," Journal of Neural Engineering, vol. 15, no. 5, art. 056013, 2018.

[9] Y. Roy et al., "Deep learning-based electroencephalography analysis: asystematic review," Journal of Neural Engineering, vol. 16, no. 5, art. 051001, 2019.

[10] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning forelectroencephalogram (EEG) classification tasks: a review," Journal of Neural Engineering, vol. 16, no. 3, 2019.

[11] M.-D. Nguyen, T. Do, X.-T. Tran, Q.-T. Nguyen, and C.-T. Lin, "EdgeAI-brain–computer interfaces system: A survey," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 33, pp. 4051–4066, 2025.

[12] T. Schneider, X. Wang, M. Hersche, L. Cavigelli, and L. Benini,"Q-EEGNet: An energy-efficient 8-bit quantized parallel EEGNet implementation for edge motor-imagery brain– machine interfaces," in Proc. IEEE Int. Conf. Smart Computing (SMARTCOMP), 2020, pp. 284–289.

[13] L. Mei et al., "An ultra-low power wearable BMI system withcontinual learning capabilities," IEEE Trans. Biomed. Circuits Syst., vol. 19, no. 3, 2025. (Details as listed in [11]; verify before submission.)

[14] J. Cao et al., "An optimized EEGNet processor for low-power andreal-time EEG classification in wearable brain–computer interfaces," Microelectronics Journal, vol. 145, art. 106134, 2024.

[15] S. Bian et al., "On-device learning of EEGNet-based network forwearable motor imagery brain–computer interface," in Proc. ACM Int. Symp. Wearable Computers, 2024, pp. 9–16.