EEG-Based Brain-Computer Interfaces (2011-2025): A Systematic Mapping of Paradigms, Devices, and Reproducibility Practices

Sevgi Koyuncu Tunç, Ahmet Burak Albayrak, Lütfiye Dila Çalık, İlke Durmaz, Utku Eyerci, Şemsettin Ekin Işık, Ayberk Özel, İbrahim Sefa Sağlam, Murat Tut, Ayşenur Elif Abanoz

Department of Software Engineering, Faculty of Engineering, Çankaya University, Ankara
Yukarıyurtçu Mahallesi, Mimar Sinan Caddesi No:4, 06815 Etimesgut, Ankara, Türkiye
sevgik@cankaya.edu.tr,
{burak_albayrak0, ekinisik4}@icloud.com,
{dilaacalik, ilkedurmaz4, utkueyercii, ayberksu60, saglamibrahimsefa, tutmurattut, ayselifabanoz}@gmail.com

Ece Çağlayan

Department of Health Informatics, Graduate School of Informatics, Middle East Technical University (METU), Ankara, Türkiye
Üniversiteler Mahallesi, Dumlupınar Bulvarı No:1, 06800 Çankaya, Ankara, Türkiye
cece@metu.edu.tr

Abstract. *Reproducibility deficits, hardware-paradigm misalignment, and fragmented benchmarking constrain EEG-based brain-computer interface research. We systematically mapped 265 publications from Scopus and citation snowballing (2011-2025), coded across six dimensions by five reviewers (Cohen's κ = 0.83). Speller systems (38.3%) and command recognition (33.1%) dominated, while motor imagery accounted for 8.1%. Emotiv devices appeared in 41 of 52 P300 speller studies. Deep learning rose sharply after 2017, but transfer learning appeared in only 1.5%. Author-deposited data occurred in 5.3%, benchmark reuse in 8.7%, and verified code in 5.7%; 85.7% shared neither code nor data, and only 3.4% acknowledged reproducibility as a limitation.*

Keywords. EEG, brain-computer interface, systematic mapping, PRISMA, P300, SSVEP, motor imagery, open science, reproducibility, dataset availability

1 Introduction

EEG-based brain-computer interfaces decode neural activity recorded from the scalp into communication or control signals without muscular involvement (Farwell & Donchin, 1988; Wolpaw et al., 2002). Their appeal lies in non-invasiveness, millisecond temporal resolution, and falling hardware costs. Since Farwell and Donchin's P300 speller paradigm in 1988, the field has diversified to include steady-state visual evoked potential (SSVEP) systems, motor imagery (MI) interfaces, and mental command recognition, with growing interest in rehabilitation and affective computing. Affordable consumer headsets and off-the shelf machine learning libraries have substantially lowered the barrier to entry, driving a publication surge visible in our corpus: 9 papers in 2011 versus 33 in 2023.

Despite this growth, high within-subject accuracy in offline settings has not translated into deployable, generalisable systems. This gap is not unique to BCI research: applied intelligent systems more broadly face the same compounding problems of studies that cannot be replicated because neither data nor code are shared, design choices driven by hardware availability rather than task requirements, and progress that is hard to measure because each group benchmarks against private datasets. EEG-BCI is a particularly clear example, old enough to have produced several hundred papers and diverse enough to span paradigms and hardware grades, yet not previously examined for all three problems at once. We present a systematic mapping of the 2011-2025 EEG-BCI literature to characterise what the community has studied, how, and with what resources, and to identify the structural barriers that constrain progress. Five research questions guide the analysis:

RQ1: Which application domains and paradigms dominate the literature?

RQ2: Which signal processing and classification methods are most prevalent, and how has this changed over time?

RQ3: What EEG hardware is used, and how does device grade relate to paradigm choice?

RQ4: What is the state of open-science compliance in terms of dataset and code sharing?

RQ5: What methodological limitations recur, and how do they constrain the field structurally?

Prior EEG-BCI reviews have each addressed a slice of these questions. Nicolas-Alonso and Gomez-Gil

(2012) surveyed paradigms and signal processing methods up to 2012. Lotte et al. (2018) focused on classification algorithms over the following decade and noted that transfer learning remained underused despite its theoretical promise. Rashid et al. (2020) provided a broader landscape review but did not systematically examine hardware selection or open science practices. Each of these is a valuable contribution, but none treats hardware, methodology, and reproducibility jointly, and none attempts to measure whether the outputs of the field are actually reusable. More recent work has examined individual facets of these questions in isolation. Chevallier et al. (2024) built a large open benchmark (MOABB) that re implements and compares machine-learning pipelines across public EEG datasets, addressing reproducibility at the level of algorithm evaluation. Saibene et al. (2024) systematically reviewed deep learning for motor-imagery decoding across 394 papers, documenting the post-2017 shift toward deep architectures and the reproducibility practices used within that subfield. On the hardware side, Zhang et al. (2023) surveyed the proliferation of wearable EEG BCI devices for medical applications, and Lee et al. (2024) empirically evaluated consumer-grade wireless headsets against a research-grade reference and reported measurable signal-quality gaps. Each of these advances one dimension (benchmark reproducibility, deep-learning methodology, or hardware capability), but none maps paradigms, devices, methods, and open science practice jointly across the full corpus, which is the structural question this study addresses.

This study makes three contributions. First, it combines paradigm distribution, hardware selection, signal processing trends, and open-science compliance in a single coded dataset, so their interactions become visible. Second, it separates studies that deposited original data to a public archive from those that merely reused pre-existing benchmarks such as BCI Competition or PhysioNet: conflating the two overstates data sharing, since 23 papers reuse benchmarks but only 14 deposit new data, an author deposit rate of 5.3%. Third, it reads these findings as an instance of the broader reproducibility problem in intelligent systems research rather than a peculiarity of BCI.

2 Methodology

2.1 Search Strategy and Screening

The literature search was conducted in October 2025 using the Scopus database. The Boolean query combined EEG and BCI terminology with signal processing and hardware keywords: TITLE-ABS KEY(("electroencephalography" OR "EEG") AND ("brain computer interface" OR "BCI" OR "brain machine interface" OR "BMI") AND ("communication" OR "speller" OR "mental command" OR "command recognition" OR "assistive communication") AND ("signal processing" OR "feature extraction" OR "classification" OR "machine learning" OR "deep learning" OR "canonical correlation analysis" OR "CCA" OR "SSVEP" OR "P300" OR "motor imagery") AND ("headset" OR "OpenBCI" OR "Emotiv" OR "Muse" OR "wearable" OR "portable")) AND PUBYEAR > 2010 AND PUBYEAR < 2026 AND (LIMIT-TO (SRCTYPE,"p") OR LIMIT-TO (SRCTYPE,"j")) AND (LIMIT-TO (DOCTYPE,"cp") OR LIMIT-TO (DOCTYPE,"ar")) AND (LIMIT-TO (LANGUAGE,"English") OR LIMIT-TO (LANGUAGE,"Turkish"))

Table 1. Screening summary (PRISMA 2020)

Stage	Action	Records
Database search	Scopus Boolean query (October 2025)	254
Supplementary	Citation snowballing + open repository search	20
Total identified	Candidate pool before screening	274
Deduplication	Exact duplicates removed	-2
Exclusion	Non-EEG modality studies	-2
Exclusion	Conceptual or theoretical, no empirical validation	-2
Exclusion	Editorials and review-only publications	-3
Final corpus	Full-text analysis	265

Scopus was chosen as the primary database because it provides broad coverage of both engineering and computer science venues relevant to this mapping. We acknowledge that this choice has consequences for the composition of the corpus. Web of Science also indexes relevant interdisciplinary journals that may not be fully represented in Scopus. IEEE Xplore indexes a substantial body of biomedical engineering and signal processing work that does not always appear in Scopus, and PubMed covers clinical and rehabilitation-oriented BCI research that is similarly underrepresented. Studies in these venues tend to focus on motor imagery, high-density EEG, and clinical populations. The relative scarcity of motor imagery papers in our results (8.1% of empirical studies) should therefore be interpreted with this search boundary in mind: it may partly reflect a genuine paradigm distribution within Scopus-indexed venues rather than the field as a whole. The same caveat applies to the hardware and open science distributions reported in the Results section.

To partially compensate for this limitation, we supplemented the database search with citation snowballing from the retrieved papers and manual searches of four open data repositories: PhysioNet, OpenNeuro, BCI Competition, and Kaggle. This added 20 further studies, giving 274 candidates in total.

Screening followed PRISMA 2020 guidelines (Page et al., 2021). Two papers were removed as exact duplicates. Six further papers were excluded on the following grounds: two used a non-EEG primary modality, two were conceptual or theoretical with no

empirical validation, and two were editorials or review-only publications. The final corpus of 265 papers was retained for full-text analysis. Table 1 summarises each screening stage.

The corpus contains 248 empirical studies and 17 secondary-literature items, including review and survey papers. The 17 secondary items were retained to support background synthesis but are excluded from all quantitative counts in the Results section. Unless otherwise stated, percentages in the Results are computed over the empirical subset of 248 papers. Open-science and hardware counts use the full corpus of 265 papers, since every publication regardless of type makes an implicit reproducibility decision. Inclusion required: (i) EEG as the primary acquisition modality; (ii) description of a signal processing or classification pipeline; and (iii) empirical results in a BCI communication or control context.

2.2 Data Extraction and Inter-Rater Agreement

Five coders independently extracted data from 52-54 papers each. Each paper was coded across six dimensions: (1) primary application domain and paradigm; (2) EEG hardware and electrode type; (3) preprocessing, feature extraction, and classifier; (4) dataset availability, scored positive only when a verifiable URL, DOI, or named repository was cited (generic "available upon request" statements were not counted); (5) source code availability, scored positive only when a live, version-controlled repository link was provided; and (6) self-reported limitations, extracted verbatim and grouped into thematic categories. A shared codebook specifying decision rules and worked examples was used throughout. To estimate inter-rater reliability, a stratified random sample of 30 papers (11.3% of the corpus) was coded independently by a second coder. Disagreements were resolved by discussion. Cohen's kappa across all six dimensions on this overlap set reached 0.83 (substantial agreement, following the convention of Landis and Koch as cited in (Petersen et al., 2015)), with per-dimension values ranging from 0.78 (application domain) to 0.91 (code availability). Of the 248 empirical papers, 155 (62.5%) reported a quantitative accuracy metric (e.g., classification accuracy, AUC, or F1-score). The remaining 93 papers reported only qualitative outcomes or no classification accuracy metric and were excluded from that analysis. For consistency and comparability across studies, only classification accuracy values were included in the descriptive analysis presented in Section 4.4.

3 Results

3.1 Publication Trend

The 265 papers span 2011 to 2025. Annual volume grew from 9 papers in 2011 to a peak of 33 in 2023, with 29 papers in 2024 and 12 in 2025. The 2025 count is a lower bound due to indexing lag at the time of search and should not be interpreted as a downturn in publication activity. A sustained increase from 2018 onward accounts for 64.5% of the corpus (171 papers), coinciding with the appearance of deep learning methods in the EEG-BCI literature (Section 3.3) and the proliferation of low-cost consumer headsets. Figure 1 illustrates the annual publication trend across the corpus.

3.2 Application Domains and Paradigms (RQ1)

Table 2 shows paradigm frequencies for the 248 empirical papers. Categories are multi-label: a hybrid P300/SSVEP study is counted in both sub-rows, so percentages sum to more than 100%. Speller-type systems (95 papers, 38.3%) and command recognition (82 papers, 33.1%) together account for roughly 70% of empirical output. Motor imagery, despite its theoretical centrality, is the primary focus of only 20 papers (8.1%). Rehabilitation and neurofeedback account for 15 papers (6.0%), emotion recognition for 7 (2.8%), and drone or UAV control for 6 (2.4%), confirming that clinically and ecologically motivated applications remain marginal.

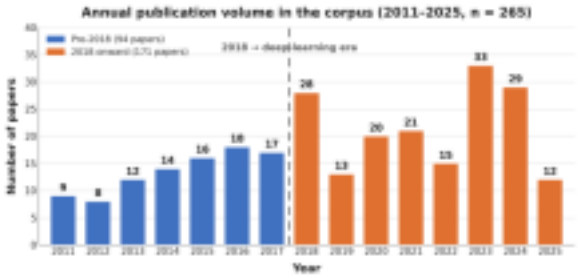


Figure 1. Annual publication volume in the corpus (2011-2025, n = 265). The dashed line marks 2018, after which deep learning methods appear in the corpus.

Table 2. Paradigm and application distribution (empirical papers, n = 248; multi-label; % of n = 248)

Application / Paradigm	n	%	Primary Signal Feature
Speller systems (total)	95	38.3%	ERP / frequency-tagged response
P300 speller	52	21.0%	N200/P300 complex at Pz
SSVEP speller	5	2.0%	Frequency-tagged occipital response
Command recognition	82	33.1%	Band power, spectral features
Binary control (yes/no, left/right)	47	19.0%	ERP or SSVEP frequency
Object / cursor control	43	17.3%	MI event-related desynchronisation
Motor imagery (primary focus)	20	8.1%	mu/beta ERD/ERS at C3-Cz-C4

Application / Paradigm	n	%	Primary Signal Feature
Wheelchair / mobility	18	7.3%	SSVEP or MI commands
Hybrid paradigms	12	4.8%	Multi-modal feature fusion
Rehabilitation / neurofeedback	15	6.0%	MI + closed-loop feedback
Emotion recognition	7	2.8%	Frontal alpha asymmetry, gamma
Drone / UAV control	6	2.4%	SSVEP / mental commands

3.3 Signal Processing and Classification (RQ2)

ERP-based P300 pipelines appear in 86 papers (32.5%), consistent with speller dominance. SSVEP with canonical correlation analysis (CCA) appears in 56 papers (21.1%). SVM (47, 17.7%) and LDA (44, 16.6%) remain the dominant classifiers (Lotte et al., 2018). CNN and other deep learning architectures, including EEGNet (Lawhern et al., 2018), appear in 51 papers (19.2%). Of these, 2 predate 2018 (2016 and 2017), while 49 are from 2018 onward, confirming a sharp adoption boundary at that year. Transfer learning or domain adaptation appears in only 4 papers (1.5%), well below the 4.1% identified by Lotte et al. in 2018 (Lotte et al., 2018). CSP-based spatial filtering, a well-established MI technique (Ang et al., 2012), appears in 14 papers (5.3%). Table 3 summarises the signal processing and classification methods identified in the corpus.

Table 3. Signal processing and classification methods (n = 265; multi-label; % of n = 265)

Method / Classifier	n	%	Paradigm Association
ERP / P300 pipeline	86	32.5%	P300 speller, ERP-based BCI
SSVEP + CCA pipeline	56	21.1%	SSVEP frequency decoding
SVM (linear, RBF, polynomial)	47	17.7%	MI, binary, command recognition
LDA / Fisher discriminant	44	16.6%	MI, P300, general classification
CNN / deep learning (EEGNet, LSTM)	51	19.2%	Cross-paradigm (MI, emotion), post-2017 primarily
FFT / power spectral density	29	10.9%	Preprocessing, SSVEP band power
CSP (common spatial patterns)	14	5.3%	Motor imagery spatial filtering
Wavelet transform (DWT/MODWT)	14	5.3%	Feature extraction, denoising
Transfer learning / domain adapt.	4	1.5%	Cross-subject generalisation

3.4 Hardware Utilisation (RQ3)

Table 4 presents device frequency across all 265 papers. Emotiv devices appear in 92 papers (34.7%), making them the dominant hardware platform by a wide margin. Including OpenBCI (23), NeuroSky (9), and Muse (8), consumer-grade devices appear in 132 studies (49.8%). Research-grade systems, namely g.tec (24), Neuroscan (13), and BrainProducts (4), collectively account for 41 papers (15.4%). The remainder use custom amplifiers, unspecified laboratory equipment, or benchmark datasets without recording hardware specification.

The hardware distribution has a direct methodological consequence. Table 5 presents a device-by-paradigm cross-tabulation derived from the coded dataset. Emotiv devices appear in 41 of the 52 coded P300 speller studies and in 29 of 82 command recognition studies. Section 4.2 examines why this concentration represents a mismatch between device capability and paradigm requirements rather than a design choice supported by the literature.

Table 4. EEG hardware platforms identified in the corpus (n = 265)

Device	n	Grade	Specifications (channels, sampling rate, electrode)	Suitable For
Emotiv EPOC / EPOC+ / Insight	92	Consumer	5-14 ch, 128-256 Hz, wet (saline felt)	Command, binary, SSVEP (speller: limited)
OpenBCI Cyton + Ultracortex	23	Consumer	8-16 ch, 250 Hz, wet/dry	SSVEP, motor imagery, P300 speller
g.tec (Nautilus, Unicorn, etc.)	24	Research	8-64 ch, 250-500 Hz, wet/hybrid	All paradigms; clinical and research grade
Neuroscan SynAmps	13	Research	32-64 ch, 1000 Hz, wet Ag/AgCl	Clinical ERP, source imaging
NeuroSky MindWave	9	Consumer	1 ch, 512 Hz, dry	Attention, binary (single-channel)
Muse / Muse 2	8	Consumer	4 ch, 256 Hz, dry	Introductory / attention paradigms
BrainProducts actiCHamp	4	Research	32-160 ch, 1000 Hz, active-wet	High-density ERP, source analysis
BCI2000 (software platform)	5	Platform	Device-agnostic framework (underlying device varies)	Framework; underlying device varies

Table 5. Device-by-paradigm cross-tabulation (columns cover the seven most frequent paradigm categories)

Device	P300 Speller	SSVEP	Command Recog.	Binary	Object Control	Motor Imagery	Rehab / NFB
Emotiv	41	4	29	19	12	6	7
OpenBCI	6	6	10	2	3	4	6
g.tec	14	2	6	1	4	0	1
Neuroscan	7	5	3	2	2	2	1
NeuroSky	2	1	3	3	1	0	1
Muse	0	0	1	0	0	0	0
BrainProducts	1	1	1	1	1	0	0
BCI2000	3	0	1	1	0	1	0

Each cell is the number of studies using that device for that paradigm. Because coding is multi-label, a single study may appear under more than one device or paradigm, so cell counts are not mutually exclusive. Shaded cells highlight the concentration of Emotiv consumer headsets in P300 speller and command-recognition studies.

3.5 Open-Science Compliance (RQ4)

Table 6 presents the open-science audit. Code was counted as available only when a live, version-controlled repository URL was verifiable at time of search; 15 papers (5.7%) meet this criterion. Dataset availability was divided into two categories: author deposited data (original data deposited by the paper's authors to GitHub, Zenodo, Figshare, OpenNeuro, OSF, IEEE Dataport, or Kaggle; n = 14, 5.3%) and benchmark reuse (papers that used a pre-existing public dataset such as BCI Competition II/III/IV or PhysioNet eegmmidb without depositing new data; n = 23, 8.7%). Only 14 papers (5.3%) share both verifiable code and some form of public data. The remaining 227 papers (85.7%) share neither.

Table 6. Open-science compliance (n = 265; % of n = 265)

Metric	n	%	Notes
Code: verified repository URL	15	5.7%	GitHub / GitLab / OSF link confirmed live at search date
Data: author deposited original data	14	5.3%	Deposited to Zenodo, Figshare, OpenNeuro, OSF, IEEE Dataport, or Kaggle by the authors
Data: benchmark reuse only	23	8.7%	Paper used BCI Competition II/III/IV, PhysioNet eegmmidb, or similar pre-existing repository
Both code and some public data	14	5.3%	Minimum artefacts for independent replication present
Neither code nor any public data	227	85.7 %	Independent replication not possible from published materials

The gap between objective non-sharing prevalence (85.7%) and self-reported acknowledgement is notable. Only 9 papers (3.4%) mention reproducibility or open-science as a limitation, indicating that the community treats non-sharing as a norm rather than a methodological deficit.

3.6 Self-Reported Limitations (RQ5)

Table 7 summarises limitation themes coded from the 265 papers. Signal quality issues (70 papers, 26.4%) are the most frequently cited limitation, followed by accuracy or performance ceiling (53, 20.0%) and real time or latency constraints (53, 20.0%). Cross-subject generalisation failure is cited in 31 papers (11.7%), and small or non-diverse participant cohorts in 35 (13.2%). Notably, 21 papers (7.9%) report no limitations at all, which is implausible for any empirical EEG study. Reproducibility or open-science deficit is cited in only 9 papers (3.4%), contrasting sharply with the 85.7% non-sharing rate reported in Section 3.5.

Table 7. Self-reported limitation themes (multi label; n = 265; % of n = 265)

Limitation Theme	n	%	Implication
Signal noise / artefact contamination	70	26.4 %	Exacerbated by dry consumer electrodes and limited channel counts
Accuracy / performance ceiling	53	20.0 %	Lab-only gains not demonstrated to hold in deployment conditions
Real-time / latency constraints	53	20.0 %	Most studies are offline; closed-loop validation is rare
Small / non-diverse participant cohorts	35	13.2 %	Healthy young adults only; clinical populations underrepresented
Lack of cross subject generalisation	31	11.7 %	Transfer learning underused (4 papers, 1.5%)
No limitations stated	21	7.9%	Implausible for empirical EEG studies
Reproducibility / open-science deficit	9	3.4%	Critically under-reported relative to actual non-sharing rate (85.7%)

4 Discussion

4.1 Paradigm Concentration and Application Gap

The paradigm distribution has remained stable over fourteen years. P300 and SSVEP speller systems together account for roughly 40% of empirical output, consistent with Nicolas-Alonso and Gomez-Gil (2012). This stability reflects path dependency rather than deliberate prioritisation: established paradigms attract further work by lowering the cost of entry through shared benchmarks and familiar pipelines. Rehabilitation, hybrid closed-loop systems, and emotion recognition collectively account for fewer than 15% of empirical papers despite having clearer clinical rationale. Motor imagery at 8.1% deserves separate attention given its theoretical centrality. Two explanations are plausible. MI research may concentrate in IEEE TNSRE and PubMed-indexed clinical venues not captured by our Scopus search. Alternatively, the dominance of Emotiv devices in this corpus creates a selection pressure against MI, which requires stable electrode placement at C3, Cz, and C4 that consumer headsets cannot reliably provide. Table 5 is consistent with the second explanation: MI appears in only 6 of 92 Emotiv studies. The 8.1% figure should be treated as a lower bound rather than an accurate representation of MI's role in the broader field. Transfer learning presents a related puzzle. Deep learning adoption grew sharply after 2017, yet transfer learning appears in only 1.5% of studies, lower than the 4.1% Lotte et al. identified in 2018. This is most plausibly a downstream consequence of the reproducibility problem: transfer learning requires multi-subject data collected under comparable conditions, and in a field where 85.7% of studies use private datasets, those preconditions rarely exist.

4.2 Hardware-Paradigm Misalignment

Emotiv devices appear in 41 of 52 coded P300 speller studies. This matters because P300 detection depends on a low-amplitude ERP component at parieto-occipital electrodes, and Emotiv's limited channel count and fixed placement provide restricted coverage of precisely that region. A practical alignment rule emerges from the data: SSVEP and command recognition paradigms are tolerant of consumer-grade hardware because their signal features do not require dense spatial coverage, whereas P300 spellers and motor imagery tasks have specific electrode placement requirements that consumer devices often cannot meet (Arpaia et al., 2022). Hardware availability rather than paradigm suitability appears to drive device selection across a substantial portion of the corpus, introducing avoidable performance constraints that are rarely disclosed.

4.3 The Structural Reproducibility Problem

The reproducibility picture from Section 3.5 is the paper's central structural finding. With most studies sharing neither code nor data, and benchmark reuse counted separately from genuine author deposits, the field generates 30 or more papers per year that are nearly all built on private datasets compared only against other private datasets within the same study. The result is a literature where reported accuracy figures cannot be aggregated and models cannot be validated across laboratories.

That only 9 papers cite reproducibility as a self-reported limitation despite an 85.7% non-sharing rate indicates that non-sharing is treated as a norm. These findings suggest that improving reproducibility requires structural interventions beyond individual researcher practices. Such interventions may include journal data availability policies, funder mandates for persistent archiving, and community agreement on a maintained shared repository for new EEG-BCI datasets. The infrastructure to implement these changes already exists.

4.4 Accuracy Reporting and Ecological Validity

Across the 155 empirical papers reporting a quantitative accuracy, the per-paper median accuracy was 90.9% and the mean is 87.6%. The large majority of these use offline, within-session, within-subject evaluation, the most favourable possible conditions for classification performance. Twenty-six papers report accuracy at or above 99%, nearly all from single subject offline evaluations. Cross-session or cross subject validation is rarely reported. High median accuracy therefore coexists with minimal cross-subject validation and no shared benchmark that would allow direct comparison across groups. Whether reported gains accumulate into field-level progress is a question that current reporting norms do not allow us to answer.

5 Conclusion

This systematic mapping of 265 EEG-BCI papers from 2011 to 2025 produces five findings mapped onto the research questions. The paradigm landscape (RQ1) is concentrated around P300 and SSVEP spellers and has not substantially diversified over fourteen years, leaving rehabilitation, hybrid closed-loop systems, and affective computing underdeveloped. Deep learning (RQ2) arrived sharply after 2017 and is now the fastest growing methodological strand, though classical SVM and LDA pipelines remain prevalent on small private datasets. Hardware selection (RQ3) is frequently misaligned with paradigm requirements, most visibly in the concentration of Emotiv consumer headsets in P300 speller studies. Open-science compliance (RQ4) is low: most studies share neither code nor data, and author-deposited datasets are rarer still. Finally, the

recurring self-reported limitations (RQ5), namely signal quality, accuracy ceiling, and real-time constraints, reflect structural features of the field rather than isolated study weaknesses, while reproducibility itself is almost never acknowledged as a limitation.

The most tractable interventions are not algorithmic. Researchers should: (i) select hardware based on paradigm requirements rather than convenience; (ii) deposit code to a persistent repository such as GitHub with a Zenodo DOI, and deposit EEG data to PhysioNet or OpenNeuro, as a default practice; (iii) report cross-subject or cross-session validation alongside within-subject offline results; and (iv) evaluate against established public benchmarks before collecting new private data. None of these steps is algorithmic; they are changes in editorial policy and community practice that would make open, requirement-driven research the default rather than the exception.

Study limitations. The search was limited to the Scopus database and did not include Web of Science, IEEE Xplore, or PubMed. Consequently, some relevant studies may not have been captured. Open science compliance was operationalised as the binary presence of a verifiable URL at search date, which is a conservative lower bound: repositories may have become unavailable after publication. Method classification relied on textual descriptions in papers rather than independent code verification. The 2025 publication count (n = 12) underestimates the true annual volume due to indexing lag.

References

Ang, K. K., Chin, Z. Y., Zhang, H., & Guan, C. (2012). Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. Frontiers in Neuroscience, 6, 39.

Arpaia, P., Esposito, A., Natalizio, A., & Parvis, M. (2022). How to successfully classify EEG in motor imagery BCI: A metrological analysis of the state of the art. Journal of Neural Engineering, 19(3), 031002. https://doi.org/10.1088/1741-2552/ac74e0

Chevallier, S., Carrara, I., Aristimunha, B., Guetschel, P., Sedlar, S., Lopes, B., Velut, S., Khazem, S., & Moreau, T. (2024). The largest EEG-based BCI reproducibility study for open science: the MOABB benchmark. arXiv preprint arXiv:2404.15319. https://doi.org/10.48550/arXiv.2404.15319

Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. Electroencephalography and Clinical Neurophysiology, 70(6), 510-523.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174.

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: A compact convolutional network for EEG-based brain-computer interfaces. Journal of Neural Engineering, 15(5), 056013.

Lee, S., Kim, M., & Ahn, M. (2024). Evaluation of consumer-grade wireless EEG systems for brain-computer interface applications. Biomedical Engineering Letters, 14, 1433-1443. https://doi.org/10.1007/s13534-024-00416-w

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for EEG-based BCIs: A 10-year update. Journal of Neural Engineering, 15(3), 031005.

Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012). Brain computer interfaces: A review. Sensors, 12(2), 1211-1279.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71.

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. Information and Software Technology, 64, 1-18. https://doi.org/10.1016/j.infsof.2015.03.007

Rashid, M., Sulaiman, N., Abdul Majeed, A. P. P., Musa, R. M., Ab Nasir, A. F., Bari, B. S., & Khatun, S. (2020). Current status, challenges, and possible solutions of EEG-based brain-computer interface: A comprehensive review. Frontiers in Neurorobotics, 14, 25.

Saibene, A., Ghaemi, H., & Dagdevir, E. (2024). Deep learning in motor imagery EEG signal decoding: A systematic review. Neurocomputing, 610, 128577. https://doi.org/10.1016/j.neucom.2024.128577

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.

Zhang, J., Li, J., Huang, Z., Huang, D., Yu, H., & Li, Z. (2023). Recent progress in wearable brain-computer interface (BCI) devices based on electroencephalogram (EEG) for medical applications: A review. Health Data Science, 3, 0096. https://doi.org/10.34133/hds.0096

Appendix: Open-Science Resource Catalogue

This appendix catalogues the studies in the reviewed corpus that provide publicly accessible datasets (Table A1) or verified source-code repositories (Table A2). URLs were verified at time of corpus search (October 2025). Studies are listed alphabetically by first author. In Table A1, entries marked * cite a pre-existing benchmark repository and are not original author

deposits; entries without * deposited new data. Brophy et al. (2021) appears only in Table A2: the study used existing public datasets but did not deposit new data.

Table A1. Studies providing publicly accessible datasets (n = 37 total; 14 author-deposited, 23 benchmark reuse). Entries marked * are benchmark reuse only.

Citation	Dataset / Repository	URL / DOI	Paradigm
Cariello et al., 2023	Project data and resources page	dees.unict.it/mbucolo/index.php/resources	Rehabilitation
Clayton et al., 2020	FEIS: 14-ch imagined speech (21 subjects)	doi.org/10.5281/zenodo.3369179	Imagined speech
Fouad et al., 2021	P300 EEG dataset (Kaggle)	kaggle.com/islamfouad/p300dataset	P300 speller
Ganesan et al., 2025	*PhysioNet eegmmidb (BCI2000, 109 subjects)	physionet.org/content/eegmmidb/1.0.0/	Motor imagery
Kim et al., 2019	Nakanishi 2015 SSVEP (12-target, 10 subjects)	github.com/mnakanishi/12JFPM_SSVEP	SSVEP
Lawhern et al., 2018	*BCI Comp IV-2a; Kaggle ERN; RSVP P300	bbci.de/competition/iv/	P300, ERN, MI, SMR
Li et al., 2024	Semantic EEG dataset (26 days)	github.com/cfli20/EEG	Semantic decoding
Liu et al., 2024	MI EEG, 50 stroke patients, BIDS (Figshare)	doi.org/10.1038/s41597-023-02787-8	Motor imagery / rehab
Mattioli et al., 2021	*PhysioNet eegmmidb (BCI2000)	physionet.org/content/eegmmidb/1.0.0/	Motor imagery
Mezzina et al., 2021	*BCI Competition III dataset	bbci.de/competition/iii/	Motor imagery
Molina-Cantero et al., 2019	Blinking experiments dataset (ResearchGate)	researchgate.net/publication/331320020	EOG / blink
Moro et al., 2017	*BCI Competition II Dataset IIb (Wadsworth)	bbci.de/competition/ii/	ERP / P300
Novais et al., 2024	*BCI Competition II and III (referenced)	bbci.de/competition/ii/	Motor imagery (review)
Parui et al., 2022	*BCI Comp IV DS1 + BCI Comp III DSIVa	bbci.de/competition/iv/	Motor imagery
Peterson et al., 2020	MI-OpenBCI dataset (GitHub)	github.com/vpeterson/MI-OpenBCI	Motor imagery
Pires et al., 2011	*BCI Competition 2003 dataset	bbci.de/competition/ii/	P300 speller
Rajashekhar et al., 2021	*MIT-BIH Noise Stress Test DB	physionet.org/content/nstdb/	Artefact benchmarking
Raji et al., 2024	OpenNeuro ds004956 (Haupt et al.)	openneuro.org/datasets/ds004956/versions/1.0.1	ERP / cognitive
Ramele et al., 2018	*BNCI-Horizon 008-2014; BCI Comp II IIb	bnci-horizon-2020.eu/database/data-sets	P300 speller
Sahoo et al., 2024	Odia BCI dataset (IEEE Dataport)	dx.doi.org/10.21227/wg17-6068	P300 / language BCI
Saibene et al., 2023	*BCI Comp + PhysioNet eegmmidb (review)	physionet.org/content/eegmmidb/1.0.0/	Motor imagery (review)
Sampanna et al., 2018	*BCI Competition II and III	bbci.de/competition/iii/	Motor imagery
Schneider et al., 2020	*BCI Competition IV-2a	bnci-horizon-2020.eu/database/data-sets	Motor imagery
Sharma et al., 2024	Bonn Univ epilepsy; DEAP; UCI alcoholism (reviewed)	Multiple; see paper	Review of datasets
Stock et al., 2016	*BCI Competition II Dataset 3	bbci.de/competition/ii/	SSVEP
Sun et al., 2023	*BCI Competition IV-2a (survey reference)	bbci.de/competition/iv/	SSVEP / drone (survey)
Suri et al., 2023	*BCI Comp III + IV; PhysioNet eegmmidb (review)	bbci.de/competition/	Motor imagery (review)
Tahmasebzadeh et al., 2013	*BCI Competition 2003 IIb	bbci.de/competition/ii/	Motor imagery
Tiwari et al., 2022	*BCI Competition III Dataset V	bbci.de/competition/iii/	Motor imagery
Tiwari et al., 2023	Mind Big Dataset (MBD)	mindbigdata.com/opendb/	Mental imagery / digits
Yan et al., 2023	*BCI Comp IV-2a; High-Gamma Dataset	bbci.de/competition/iv/	Motor imagery

Citation	Dataset / Repository	URL / DOI	Paradigm
Zao et al., 2014	BCI Web Portal + RESTful API	bci.pet.cs.nctu.edu.tw/databank	Multi-modal / game BCI
Zhao et al., 2020	*IFMBE MEDICON 2019; BCI Comp III DS-II	medicon2019.org/scientific-challenge/	ERP / ASD BCI
Zhu et al., 2021	SSVEP 102-subject, 12-target dataset (Figshare)	doi.org/10.6084/m9.figshare.13560281	SSVEP
Pavan et al., 2025	MATLAB scripts + analysis data (OSF)	osf.io/d9eha/	Neurofeedback / ERP
Molina-Cantero et al., 2019	IR-OG and EOG hardware + software (GitHub)	github.com/aljemoca/BlinkingIROG	EOG / blink
Peterson et al., 2020 (b)	Acquisition + processing pipeline (GitHub)	github.com/vpeterson/MI-OpenBCI	Motor imagery

Table A2. Studies providing verified public source-code repositories (n = 15). All URLs confirmed live at time of corpus search (October 2025).

Citation	Description	Repository URL	Paradigm
Brophy et al., 2021	DenoiseEEG-GAN: GAN-based EEG denoising	github.com/Brophy-E/DenoiseEEG-GAN	Artefact removal
Cardona-Alvarez et al., 2023	BCI-Framework: open-source BCI development platform	github.com/UN-GCPDS/BCI-framework	General framework
Ingolfsson et al., 2024/2025	VowelNet: lightweight CNN for EEG vowel decoding	github.com/pulp-bio/VowelNet	Imagined speech
Larsen et al., 2024	Eye-tracking + Unity LSL integration	gitlab.stud.idi.ntnu.no/group-92/eye-tracking-unity-lsl	Game / eye-tracking
Lawhern et al., 2018	EEGNet: compact CNN for EEG-BCI (Python/Keras)	github.com/vlawhern/arl-eegmodels	Cross-paradigm
Legeay et al., 2022	NeuXus: modular BCI pipeline framework (Python/PyPI)	github.com/LaSEEB/NeuXus	General pipeline
Li et al., 2024	Semantic EEG analysis code (MATLAB)	github.com/cfli20/EEG	Semantic decoding
Mattioli et al., 2021	MI-EEG-ID-CNN: subject identification CNN	github.com/Kubasinska/MI-EEG-ID-CNN	Motor imagery
Molina-Cantero et al., 2019	IR-OG and EOG hardware + signal processing software	github.com/aljemoca/BlinkingIROG	EOG / blink
Parbez et al., 2020	Blink-Fruity: blink-controlled BCI game (Python)	github.com/ranadepto/blinkfruity	Binary / blink
Pavan et al., 2025	MATLAB analysis scripts + data (OSF)	osf.io/d9eha/	Neurofeedback / ERP
Peterson et al., 2020	MI-OpenBCI: acquisition + processing pipeline	github.com/vpeterson/MI-OpenBCI	Motor imagery
Reyes et al., 2021	BMI eyes blinking: blink detection BCI (Python)	github.com/Andresreyesf/brain_machine_interface_eyes_blinking	Binary / blink
Saini et al., 2022	tinymlgen: ML model export for microcontrollers	github.com/eloquentarduino/tinymlgen	Embedded / IoT BCI
Schneider et al., 2020	Q-EEGNet: quantised EEGNet (PULP platform)	github.com/pulp-platform/q-eegnet	Motor imagery