Article:
du Bois, N., Korik, A., Hodge, S. et al. (2026) Advancing EEG-based assessment of consciousness and cognition in prolonged disorders of consciousness. Communications Medicine. ISSN: 2730-664X

https://doi.org/10.1038/s43856-026-01574-x

Communications Medicine

https://doi.org/10.1038/s43856-026-01574-x

Article in Press

Advancing EEG-based assessment of consciousness and cognition in prolonged disorders of consciousness

Received: 9 July 2025

Accepted: 23 March 2026

Cite this article as: Bois, N., Korik, A., Hodge, S. *et al.* Advancing EEG-based assessment of consciousness and cognition in prolonged disorders of consciousness. *Commun Med* (2026). https://doi.org/10.1038/s43856-026-01574-x

Naomi du Bois, Attila Korik, Stephanie Hodge, Leah Hudson, Ainjila S. Elahi, Alain Bigirimana, Natalie Dayan, Jose M. Sanchez-Bornot, Alison McCann, Kudret Yelden, Lloyd Bradley, Krishnan P. S. Nair, Simon Judge, Damon Hoad, Emma Vines, Venu Harilal, Sheryl Parke, Paul Johnson, Jacqueline Pogue, Emma Dodds, Abayomi Salawu, Raymond Carson, Karl McCreadie, Jacqueline Stow, Jacinta McElligott, Aine Carroll & Damien Coyle

Title: Advancing EEG-Based Assessment of Consciousness and Cognition in Prolonged Disorders of Consciousness

Authors: Naomi du Bois[1], Attila Korik[1], Stephanie Hodge[2], Leah Hudson[2], Ainjila S. Elahi[2], Alain Bigirimana[3], Natalie Dayan[2], Jose M. Sanchez-Bornot[2], Alison McCann[4], Kudret Yelden[5], Lloyd Bradley[6], Krishnan P. S. Nair [7], Simon Judge[8], Damon Hoad[9], Emma Vines[9], Venu Harilal[10], Sheryl Parke[10], Paul Johnson[11], Jacqueline Pogue[12], Emma Dodds[13], Abayomi Salawu[14], Raymond Carson[4], Karl McCreadie[2], Jacqueline Stow[4], Jacinta McElligott[4], Aine Carroll[4, 15], and Damien Coyle[1, 2 *]

Affiliations

[1] Bath Institute for the Augmented Human (IAH), University of Bath, Bath, BA2 7AY, UK
[2] Intelligent Systems Research Centre, Ulster University, BT48 7JL, UK
[3] Queens University Belfast, Belfast, BT7 1NN, UK
[4] National Rehabilitation Hospital, Dublin, A96 RPN4, IRE
[5] Kings College Hospital, London, SE5 9RS, UK
[6] Royal Hospital for Neuro-disability, Putney, London, SW15 3SW, UK
[7] Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, S10 2JF, UK
[8] Barnsley Hospital NHS Foundation Trust, Barnsley, S75 2EP, UK
[9] South Warwickshire University NHS Foundation Trust, Rugby, CV21 3SR, UK
[10] Norfolk Community Health and Care NHS Trust, Norwich, NR2 3TU, UK
[11] Western Health and Social Care Trust, Derry, BT47 6SB, UK
[12] Northern Health and Social Care Trust, Antrim, BT41 2RL, UK
[13] Oxford University Hospitals NHS Foundation Trust, Oxfordshire, OX3 9DU, UK
[14] Hull University Teaching Hospitals NHS Trust, Hull, HU3 2JZ, UK
[15] Health Sciences Centre, University College Dublin, Dublin 4, IRE

* Prof Damien Coyle (corresponding author)
Affiliation: Bath Institute for the Augmented Human
Address: Bath Institute for the Augmented Human (IAH), University of Bath, Claverton Down, Bath, BA2 7AY, UK
Email: dhc30@bath.ac.uk
Phone: +441225386896

Abstract

Background: Accurate assessment of residual awareness in patients with Prolonged Disorders of Consciousness (PDoC) remains a major clinical challenge, as conventional behavioural tools can underestimate covert cognition. This study evaluates whether a structured, multi-phase motor imagery Brain–Computer Interface (MI-BCI) protocol provides objective electroencephalography (EEG)-based indicators of awareness that complement behavioural assessments.

Methods: Forty-four participants (N = 44) completed repeated imagined-movement tasks using wearable EEG (PDoC: Unresponsive Wakefulness Syndrome (UWS, n = 14), Minimally Conscious State (MCS, n = 17), Locked-In Syndrome (LIS, n = 11); two able-bodied participants as benchmarks; ClinicalTrials.gov: NCT03827187; 30-01-2019). The protocol assessed sensorimotor rhythm modulation, training with and without neurofeedback, and binary question answering across phases. Standard behavioural assessments (CRS-R and WHIM) were administered at each session.

Results: Significant MI-BCI decoding accuracy (DA) is achieved by 73.8% of patients, of whom 90% progress to Q&A testing and frequently exceed the 70% usability threshold, revealing marked inter-individual heterogeneity. For significant MI-BCI runs, LIS outperform MCS (p = 0.007) and UWS (p = 0.048), while UWS exceed MCS during Q&A (p = 0.049), driven by familiar-voice stimuli. Using leave-one-subject-out cross-validation, combining predictions from DA and behavioural assessments improves balanced diagnostic accuracy to 62% (from 55%), increasing sensitivity to MCS (39% to 69%), with a modest reduction in LIS sensitivity (78% to 67%). Task-related activity over sensorimotor and parietal cortices differentiate diagnostic groups.

Conclusions: The structured MI-BCI protocol demonstrates potential as a movement-independent, EEG-based tool for distinguishing UWS, MCS and LIS. Integrating DA and spatial patterns yields diagnostic information that may augment behavioural assessment and advance objective tools for evaluating awareness in PDoC.

Plain language summary

Some people with severe brain injuries seem unresponsive, yet their brains may still show signs of awareness. This study tested a brain-computer system that detects these signs by recording tiny electrical signals from the scalp when a person imagines moving.

We worked with people in different states of consciousness and those with locked-in syndrome, a rare condition marked by complete paralysis, but intact consciousness – across hospitals, care homes, and private homes. The results showed that many could change their brain signals, suggesting intentional thinking, especially those with locked-in syndrome.

These findings show that this kind of brain-computer system could help doctors detect hidden awareness in people who cannot move or speak, and may eventually assist with diagnosis and basic communication.

Introduction

Disorders of Consciousness (DoC) refer to a spectrum of altered conscious states resulting from an acquired brain injury. When a DoC persists beyond four weeks, it is termed a Prolonged Disorder of Consciousness (PDoC) [1]. Consciousness comprises two components: wakefulness, driven by subcortical arousal systems (primarily the brainstem and thalamus), and awareness, which depends on a distributed frontoparietal network supporting higher-order cognitive processing and perception of self and environment [2–4]. Clinical presentations are on a continuum; coma (no wakefulness or awareness, typically transient); Unresponsive Wakefulness Syndrome (UWS), wakefulness is present but awareness is absent; and the Minimally Conscious State (MCS), wakeful with minimal or fluctuating awareness [1]. In the UK alone, it is estimated that there are between 4,000 and 16,000 patients diagnosed with UWS, with up to three times as many living in MCS [5]. Although not part of the DoC spectrum, Locked-In Syndrome (LIS) can be misdiagnosed as a DoC using behavioural assessment tools, as the condition results in near-total paralysis that can mask preserved awareness [6–8]. People with LIS retain consciousness awareness and can typically communicate via eye movements or blinking – although this ability is lost in Complete Locked-In Syndrome (CLIS) due to additional oculomotor impairment [7,8].

Several diagnostic tools are available to assess consciousness, including the Coma Recovery Scale-Revised (CRS-R), the Wessex Head Injury Matrix (WHIM), and the Sensory Modality Assessment and Rehabilitation Technique (SMART) [9]. However, a fundamental limitation shared by all behavioural assessment tools is their reliance on the patient's ability to produce overt motor or verbal responses. For patients with a PDoC, this presents a significant challenge – fluctuations in arousal, sedative side effects of medication, the presence of a tracheotomy, or severe motor impairments may prevent consistent behavioural output. Assessment quality is also influenced by clinician expertise and experience [10]. Consequently, up to 40% of MCS patients are misdiagnosed as being in UWS [10,11]. Accurate diagnosis is critical for appropriate prognostication, to maximise quality of life and to guide decisions around longer term continuation of treatment [10–13]. Misdiagnosis is not limited to PDoC. A survey of forty-four LIS patients in France reported an average diagnostic delay of 78.1 days, with some cases exceeding 1,460 days (four years) – attributed to misinterpretation of symptoms [6]. These findings highlight the urgent need for objective diagnostic tools that do not rely on motor behaviour or vocalisation alone to assess consciousness.

Alternative approaches to behavioural assessment have emerged following advances with neuroimaging techniques. A landmark study by Owen et al. (2006) demonstrated covert awareness in a patient with an apparent UWS by measuring neural responses to motor imagery tasks using functional magnetic resonance imaging (fMRI) [14]. Monti et al. (2010) extended this fMRI-based protocol to achieve yes/no communication [14,15]. However, fMRI is costly and not always accessible, prompting the development of alternative methods such as electroencephalography (EEG). Cruse et al. (2011) [16] demonstrated that EEG, a non-invasive, bedside-compatible method, could detect covert awareness in DoC. Sensorimotor rhythms (SMRs), typically in the 8–12 Hz (μ) and 18–26 Hz (β) ranges, are modulated during both movement and kinaesthetic motor imagery (MI) [17]. Cruse et al. (2011) demonstrated intentional SMR modulation in three UWS patients, indicating misdiagnosis or state transition [16]. Coyle et al. (2015) [18] moved beyond one-off detection to closed-loop training for MCS patients, using a motor imagery brain-computer interface (MI-BCI), marking a shift toward BCI as a trainable skill. Participants received both auditory and visual real-time feedback across multiple sessions, with auditory feedback (particularly music) shown to improve motor imagery Decoding Accuracy (DA). This study demonstrated the feasibility of MI-BCI training in PDoC populations –

catalysing a transition in EEG-based BCI research from passive testing to interactive, real-time systems for PDoC populations [19].

Despite growing support for BCIs in supplementing the CRS-R, the current gold-standard behavioural assessment tool [9,20,21], the Royal College of Physicians (RCP) guidelines continues to recommend behavioural assessments (favouring the CSR-R) [22]. The RCP argues that there is not enough evidence for neuroimaging-based methodologies to enter routine clinical practice alone [22]. In contrast, the American Academy of Neurology (AAN) and the European Academy of Neurology (EAN) recognise the additive potential of neuroimaging and electrophysiological tools under specific conditions [23,24]. All three organisations share concerns, however, regarding standardisation, cost-effectiveness, and real-world feasibility of these technologies [23]. Therefore, ongoing research continues to address these concerns by developing practical, scalable, and clinically integrated solutions.

Recent efforts have aimed to broaden the diagnostic capabilities of EEG-based BCIs in (P)DoC. In conjunction with, and subsequent to the Coyle et al. (2015) MI-BCI study, hybrid EEG-based BCI paradigms have been developed that combine visually evoked responses, i.e., the P300 event-related potential (ERP) and the Steady-State Visually Evoked Potential (SSVEP) – to detect command following and preserved reasoning in this population [25,26]. This multi-input approach has improved BCI sensitivity and reliability, particularly when supported by advanced classification algorithms [18,27–31]. Kim et al. (2022) [32] demonstrated the value of a multi-EEG marker approach to the assessment of cognitive function in a brain-injured paediatric cohort. By combining the EEG correlates of motor command-following and auditory evoked responses, they enabled cognitive profiling in this cohort, reinforcing the diagnostic/prognostic potential for this approach in clinical populations – highlighting the diagnostic potential of integrating multiple neural signals.

However, despite these advances, EEG-based MI-BCI systems remain blind to the integrity of the neural networks supporting consciousness. Functional connectivity (FC) analysis offers a systems-level view of the network dynamics that support awareness – including the Default Mode Network (DMN) and the Fronto-Parietal Network (FPN). Research findings show that disruptions in these intrinsic networks are consistently associated with loss of awareness [33–38], while residual or reorganised connectivity in key hubs such as the medial Prefrontal Cortex (mPFC), Posterior Cingulate Cortex (PCC), and dorsolateral Prefrontal Cortex (dlPFC) can distinguish between DoC states [39,40]. Furthermore, cortico-thalamic integrity, particularly within the ascending arousal system, is an established neural marker for poor outcomes in acute DoC patients [41,42]. Combining FC with EEG-based MI-BCI could move diagnostic tools beyond binary classifiers toward network-informed diagnostics.

Building on this foundational work (e.g., Owen *et al.*, 2006 [14]; Cruse *et al.*, 2011 [16]; Coyle et al., 2015 [18]), the present study evaluates a multi-phase, longitudinal EEG-based MI-BCI protocol for patients with PDoC and LIS. The study protocol incorporates command-following assessment, real-time neurofeedback, and structured yes/no communication – rolled out in three phases: assessment (sessions 1–2), training and feedback (sessions 3–6), and binary question-answering (sessions 7–10/13). Phase I identifies participants capable of SMR modulation. Phase II provides guided MI practice with auditory feedback to support learning. Phase III investigates participants' binary responses to biographical, numerical, logical, and situational questions using the MI-BCI. The primary focus is to determine whether cognition can be profiled by analysing the aggregate decoding performance within each question category – thereby using the BCI as a scientific instrument to probe domain-specific awareness. Importantly, the outcomes of this phase also provide an indication of the system's potential for future development as a communication mechanism.

In addition to EEG-signal classification, used for group-level comparisons of DA – the study incorporates further EEG analyses to probe underlying neural mechanisms of awareness. Specifically, a between-group comparison of the features selected for classification is included, to provide insight into differential neural activation across diagnostic categories – as well as a between-group analysis of functional connectivity, targeting intrinsic networks implicated in awareness, to explore systems-level differences in network integrity between PDoC and LIS participants.

This study represents one of the largest EEG-based MI-BCI investigations to date involving patients with UWS, MCS, and LIS/CLIS (*n* = 42). Participants were tested in real-world environments, such as hospitals, care homes, and private residences – enhancing the system's ecological validity and clinical relevance. Importantly, this work moves beyond one-off detection paradigms by integrating training, communication, and behavioural comparison (CRS-R, WHIM) in a longitudinal framework. The primary objective is to determine whether the structured BCI protocol provides an objective, scalable complement to existing diagnostic frameworks for patients with PDoC. Specifically, we assess whether DA from a two-class imagined-movement task differentiates diagnostic groups (UWS, MCS, and LIS) and reveals consistent neural markers of intentional modulation. In line with the study's basic-science and diagnostic focus, the final phase involving a structured closed Q&A assesses the feasibility of extending this framework to future applications in understanding cognition in unresponsive patients, and explores whether responses may provide insights into patients' cognitive state or needs.

We demonstrate that most participants achieve significantly above chance level MI-BCI decoding accuracy, with LIS generally achieving higher accuracies than UWS and MCS across paradigms. Group-level modelling identifies DA as a key predictor distinguishing LIS from UWS, while the inclusion of traditional behavioural measures further improves classification. Despite paradigm-dependent variation in DA, overall DA, motor-associative cortical activation patterns, and connectivity features supplement existing behavioural indicators, suggesting that DA offers complementary diagnostic information beyond traditional behavioural assessment. Exploratory analyses using domain-specific question stratification (i.e., grouping yes/no items by cognitive category) revealed promising initial evidence that supports further development of refined BCI-based assessment tools capable of probing distinct cognitive domains and, potentially, enabling more effective communication within this framework.

Methods

Study sample

This study was registered on ClinicalTrials.gov (NCT03827187; 30th January, 2019), and included 42 patients (aged 17–73 years; 13 females, 29 males). Patients were diagnosed as follows: 14 with unresponsive wakefulness syndrome (UWS), 17 with minimally conscious state (MCS), 10 with locked-in syndrome (LIS), and 1 with complete locked-in syndrome (total LIS participants, *n* = 11). Two able-bodied males (aged 20 and 23) were included as benchmarks (AB, *n* = 2). Previous publications involving subsets of this cohort have reported early findings [18,43,44], with a recent preprint reporting on the findings of a separate evaluation [45]. All participants were naïve to the task, with the exception of one MCS participant who had previously taken part in an earlier study by Coyle *et al.* (2015)[46] that investigated the ability of MCS patients to learn to modulate sensorimotor rhythms through MI-BCI training.

Ethical approval was granted by the NHS Health Research Authority (Integrated Research Application System (IRAS) UK and Scotland, Project IDs: 136640 and 247815; Research Ethics Committee (REC)

reference: 18/WA/0186; approved 07/10/2020). Ethical approval to conduct this research study in the Republic of Ireland, was granted by the National Rehabilitation University Hospital (NRH) REC. Informed consent was obtained from participants or via proxy consent where necessary. The approved protocol involved experimental procedures that did not include any therapeutic or clinical intervention; accordingly, the research was conducted as a non-interventional basic science study. All procedures complied with the Declaration of Helsinki.

Clinical consultants screened patients for eligibility and collected consent. For participants lacking capacity, consent was provided by a consultee, such as the participant's next of kin or legally authorised carer. Inclusion criteria were a diagnosis of prolonged disorder of consciousness (PDOC), low awareness states, or (C)LIS. Exclusion criteria included progressive neurological diseases, uncontrolled epilepsy or pain, medications impairing cognition or causing excessive fatigue, and skull deformities preventing electrode contact. Sex and gender were not variables under investigation, however, participants' sex is reported in the population description. The study population did not include socially relevant subgroupings; therefore, race and ethnicity were not reported.

Experimental paradigm

Movement imagery tasks were selected for each patient with input from family or clinicians, avoiding injured brain regions and favouring arm movements when possible. Patients imagined one movement combination throughout the study: left vs. right arm, right arm vs. feet, or left arm vs. feet. Before each session, patients were instructed to keep their eyes open, remain still, and avoid teeth grinding. With consent from a family member or caregiver, a gentle massage between runs was permitted to rouse the patient if they appeared to fall asleep.

Table 1. Structure of the study phases and associated runs.

	Run>>>	1	2	3	4
Phase I (Assessment paradigm)	Session 1	A1	F(PN)	-	-
	Session 2	A2	T	F(PN)	-
Phase II (Training and Feedback paradigms)	Sessions 3 - 4	T	T	F(PN)	F(M)
	Sessions 5 - 6	T	F(M)	F(M)	F(M)
Phase III (Q&A paradigm)	Sessions 7 - 8	T	F(M)	Q&A	Q&A
	Sessions 9 – 10	F(M)	F(M)	Q&A	Q&A
	Sessions 11-12	F(M)	Q&A	Q&A	Q&A

A = Assessment, F(PN) = Feedback (Pink Noise), T = Training, F(M) = Feedback (Music), Q&A = Question and Answer session. NOTE: For some participants, training, feedback and Q&A sequencing was adjusted to accommodate individual progression and responses.

The study consisted of three phases (see Table 1).

Phase I (Sessions 1–2): *Assessment* of awareness through imagery tasks following cues, plus real-time Feedback runs. Data were analysed offline (see Offline single-run analysis for BCI calibration for details). Participants advanced to the next phase when decoding accuracy exceeded 70% or when task-related DAs were statistically greater than baseline (one-tailed paired *t*-test, $\alpha = 0.05$). The 70% threshold corresponds to performance well above the 99% upper bound of the binomial chance level for 90 trials, providing a conservative operational benchmark for reliable BCI control.

Phase II (Sessions 3–6): Additional Training and Feedback runs.

Phase III (Sessions 7–10): Training, Feedback, and Q&A runs.

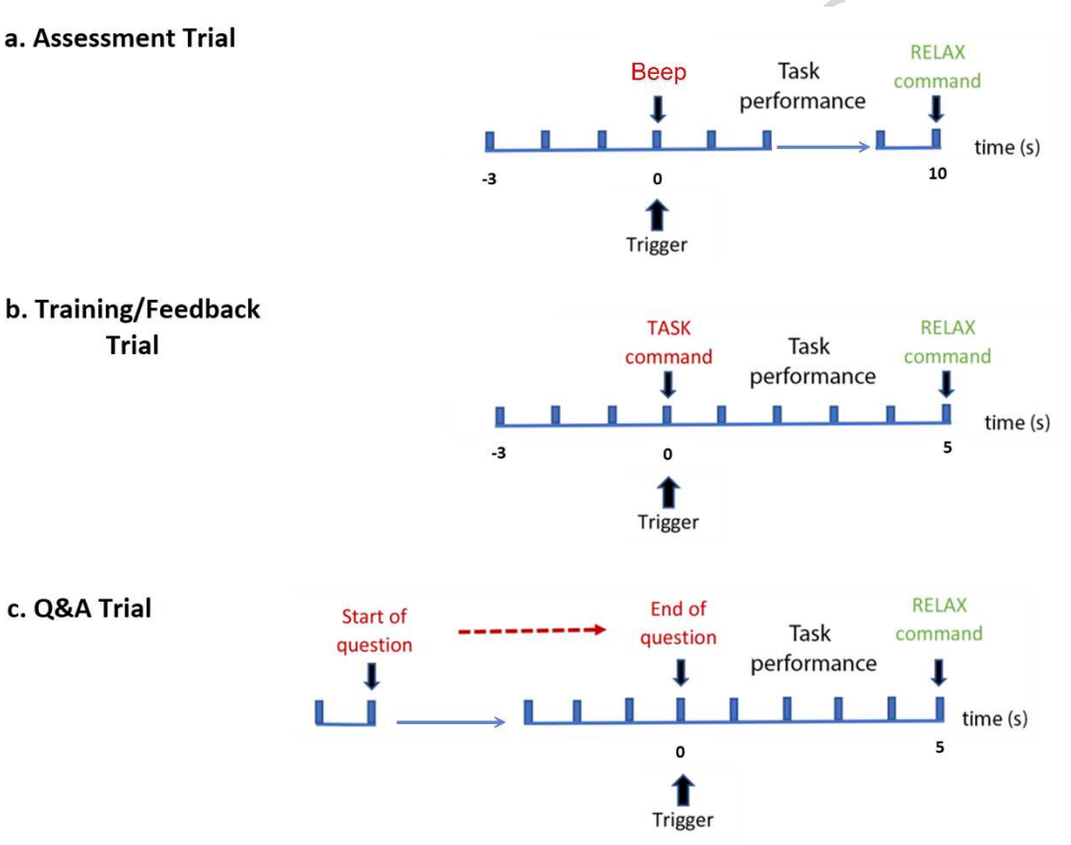

Figure 1. Timing of the trials. a: Assessment. b: Training, and Feedback. c: Q&A. Note: the Q&A task began with a question. Therefore, the trigger for the start of the task performance period was sent at the end of the question. Paradigm specific trial instructions are provided in Supporting Information, Supplementary Note 1.

Trial timings are presented in Figure 1. Participants who advanced completed 4–13 sessions, each comprising 1-5 runs, although the number varied depending on health and scheduling. Tasks were performed without distractions or visual aids once sessions began.

Experimental procedure

Each daily session (1–2 hours) included 1–5 runs selected from Assessment, Training, Feedback, and Q&A paradigms. EEG setup was performed at the start of each session. The verbal instructions given (via headphones) at the start of each task, are provided in Supporting Information: Methods, Supplementary Note 1. Paradigm Specific Instructions.

Assessment runs: Patients were verbally and visually instructed to imagine specified movements (e.g., lifting a weight with the right arm). Each run followed a block design: six blocks of 15 trials, alternating between two imagined movements (three blocks per movement). Beeps (0.5s) cued imagery every 8s, with pre-recorded instructions at block start. Each Assessment run lasted ~17 minutes, including rest periods. The protocol was adapted from Cruse et al. 2011 [16].

Training runs: Conducted at the start of each Phase II/III session, Training runs compressed the Assessment format: 60 randomly ordered trials (30 per movement class) without block structure. Commands ("Left"/"Right") were given via pre-recorded voice, followed by a 5s imagery period and 3s relax cue. Each Training run lasted ~8 minutes.

Feedback runs: Similar to Training runs, but included real-time auditory feedback. Sound (broadband noise or music) shifted azimuthally between ±90° based on correct decoding of imagined movements. Misclassification led to erratic or misdirected auditory feedback. Runs lasted ~8 minutes

Q&A runs: Each Q&A run (8–11 minutes) presented 48 yes/no questions from two combined categories (basic logic, numbers/letters, situational, biographical) using familiar caregiver-recorded audio to produce self-relevant stimuli which were expected to maximise engagement [47,48]. Patients responded by imagining movements associated with "yes" or "no." Task periods lasted 5s, followed by relax cues. Categories rotated to balance linguistic and numerical content. Questions were adapted from the Montreal Cognitive Assessment (MOCA) [49] and in accordance with the National Clinical Guidelines for PDoC [50]. A full list of the questions for each category is provided in Supporting Information: Methods, Supplementary Note 2. Questions were personalised by inserting the, correct (for a “yes” response) or incorrect (for a “no” response), question completion information where there are blanks in the text. For example, the biographical question set contains the question “Is your name _____ ?”. This question is asked twice in one biographical questions run – once with the participant’s correct name as a “yes” response question, and once with the participant’s incorrect name as a "no" response. Questions were presented in random order.

Clinical Measures: Routine CRS-R and WHIM assessments [51–53], were conducted daily to compare behavioural scores with EEG-based BCI performance (i.e., DA). Sessions alternated between morning and afternoon to account for diurnal variation in patient alertness [54].

Data Acquisition: EEG was recorded from 16 channels (Figure 2) with active electrodes using a g.Nautilus Wireless Research EEG headset [55]. The reference electrode was fixed on the right earlobe and the ground electrode was positioned over the AFz electrode location according to the international 10/20 EEG standard. The EEG was filtered (Butterworth, 0.5-100Hz, eighth order), and sampled (sampling rate: 250Hz, down-sampled to 125Hz).

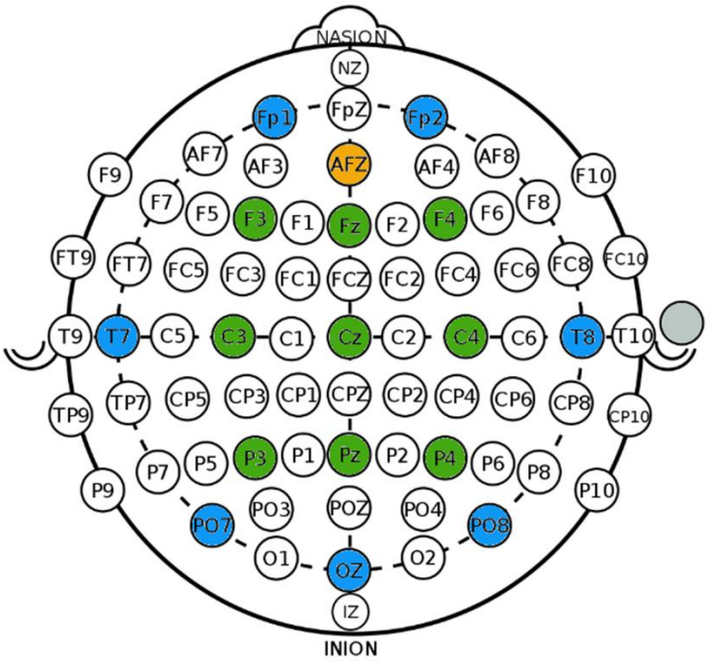

Figure 2. EEG montage. The positions of EEG (blue and green) and ground (orange) electrodes are presented in this figure. Nine EEG channels covering motor and imagined movement-related cortical areas are indicated in green. The reference clip was attached to the right ear (as indicated by the circle in grey).

A user datagram protocol (UDP) managed communication between a MATLAB Simulink [56] module (that was used for EEG data acquisition and online signal processing), and the experimental protocol controller application presented in the Unity 3D Game Engine [57].

Offline signal processing

Offline analysis and BCI calibration were performed using a filter-bank common spatial patterns (FBCSP) and mutual information (MI)-based feature selection framework [58,59] following a structure similar to Korik *et al.* [60]. This calibrated model was used to provide auditory feedback in online runs. FBCSP is a well-established EEG classification method for distinguishing between imagined movements [61,62], as shown in Figure 3.

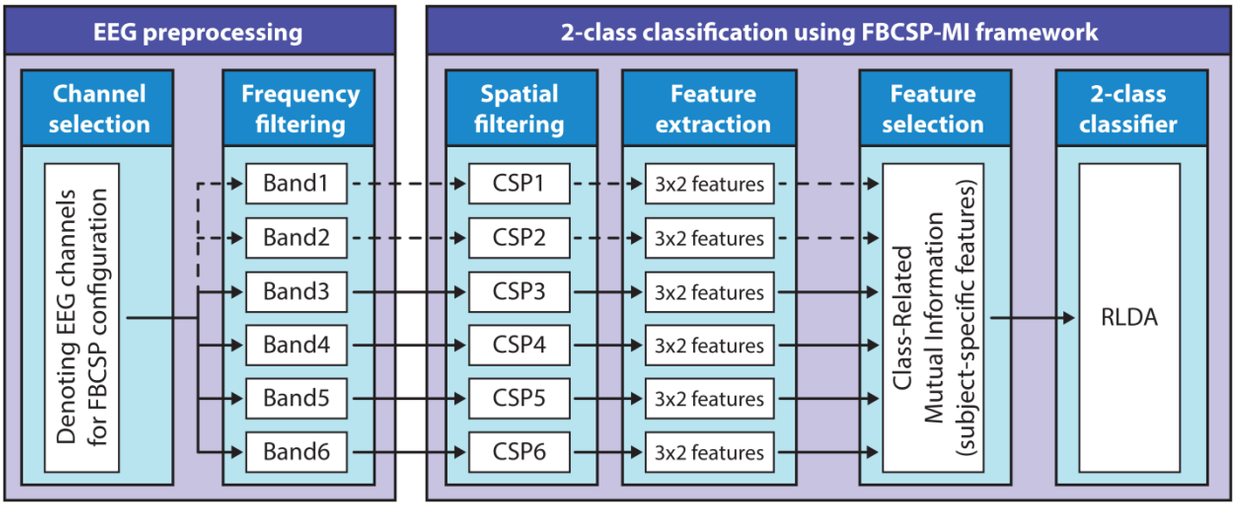

Figure 3. Structure of the applied FBCSP-MI based 2-class classification method. The block diagram illustrates the signal processing framework, including filter bank common spatial patterns and mutual information (FBCSP-MI) based features selection modules, and a regularized linear discriminant analysis (RLDA) 2-class classifier. The analysis performed using three different sets of frequency bands, involving four adjacent bands from the six pre-processed bands (i.e., delta, theta, mu, low-beta, high-beta, low-gamma bands). The illustrated setup represents the highest four-band option.

EEG signals were filtered into six standard frequency bands [0.5–4 Hz (delta), 4–8 Hz (theta), 8–12 Hz (mu), 12–18 Hz (low beta), 18–28 Hz (high beta), 28–40 Hz (low gamma)] using Simulink FIR filters [56] (band-pass attenuation: 0 dB; band-stop: 60 dB). For each participant and run, FBCSP-MI calibration involved testing combinations of:

1. Two EEG channel sets [16-channel full coverage; 9-channel motor cortex-focused]
2. Three frequency band sets ([delta–low beta], [theta–high beta], [mu–low gamma], each comprising four adjacent bands from the six preprocessed bands, see Figure 3**Error! Reference source not found.**)
3. Classification windows of 1 s and 2 s
4. Between 4–10 of the highest ranked MI-selected features

Feature extraction and classification

*Epoching***:** For each run, task-relevant EEG segments were extracted from -3s to +5s relative to the onset of the motor imagery cue. These epochs were extracted from the frequency-filtered dataset and processed separately for each EEG channel configuration.

Spatial Filtering (CSP): A common spatial patterns (CSP) method was applied to maximise discriminability between two motor imagery classes. CSP generates a set of spatial filters that maximise the variance of EEG signals from one class while minimising it for the other [63,64]. Each filter corresponds to a row in a linear transformation matrix applied to the band-pass filtered EEG signals. For each frequency band, three CSP filter pairs (six filters total) were retained. These were selected based on their ability to produce the highest variance separation between classes.

Feature Extraction: CSP-filtered signals were transformed into feature vectors using a sliding window of either 1 s or 2 s (depending on the FBCSP-MI setup), with a 40 ms step size between windows. For each window, the feature vector was calculated as the natural logarithm of the signal variance:

$$\bar{\boldsymbol{\omega}} = \boldsymbol{log}(\boldsymbol{var}(\boldsymbol{E})) \quad (1)$$

where, $\bar{\boldsymbol{\omega}}$ is the feature vector and $\boldsymbol{E}$ is the CSP-transformed EEG signal.

Feature Selection (Mutual Information): Features were ranked and selected using a "best individual feature" algorithm based on mutual information (MI) [59]. The MI between each feature and the class label was computed using a fixed quantisation level of 3. For each configuration, 4 to 10 features with the highest MI values were retained for classification (evaluated across multiple configurations).

*Classification (RLDA)***:** Selected features were classified using regularised linear discriminant analysis (RLDA), implemented via the RCSP toolbox [64]. RLDA constructs a linear decision boundary (hyperplane)

that separates the two classes. Classification output is determined by the sign of the feature vector's projection onto the weight vector:

- Positive output → Class 1
- Negative output → Class 2

The classifier's decision confidence is proportional to the distance from the hyperplane [65].

MI-BCI Performance Evaluation: Decoding accuracy (DA) values were computed for each setup (1 s vs. 2 s window, channel configuration, frequency band set, and number of features). These were compared across configurations to 1. determine the optimal BCI calibration per participant and 2. to maximise sensitivity at the individual level to detect evidence of intentional sensorimotor modulation within a BCI framework.

Offline single-run analysis for BCI calibration

Each combination of FBCSP-MI setup parameters (channel set, frequency band set, classification window length, and number of features) was evaluated independently for every participant and run. Classification performance was assessed using six-fold cross-validation (CV). The configuration yielding the highest peak DA in the task period was selected as optimal for that participant and run (see Feature extraction and classification).

Time-Varying Decoding Accuracy Calculation: For each fold, DA was calculated at each time point, generating a time-resolved accuracy curve. The baseline (reference) interval was defined from -1000 ms to 0 ms before the task cue. Although motor imagery tasks were cued at 0 ms and nominally expected between 0 to 5000 ms, the task interval was defined from 400 ms to 7000 ms post-cue. This adjusted window accounts for potential stimulus-related activity in the first 400 ms (which could confound motor imagery signals), and known latency in PDoC patient responses, which may extend beyond 5s post-cue [66]. A 200 ms moving average smoothing was applied to each DA time-series to reduce short-term variance.

Peak DA Extraction: For both reference and task intervals, the smoothed DA curve was scanned to identify the maximum DA value (smoothed peak). A final peak DA was then selected from the non-smoothed DA curve as the local maximum within a ±150 ms (300 ms total) window centred on the smoothed peak. This two-step process reduced the likelihood of selecting spurious DA spikes in otherwise low-accuracy regions.

Statistical Comparison of Task vs. Baseline DAs: Task and baseline DAs were compared using a one-tailed paired *t*-test across CV folds, to test the hypothesis that DA during the task interval was higher than baseline. A statistically significant result at $\alpha=0.05$ indicated that classifier performance during motor imagery exceeded baseline variability. Runs without significant task-related increases indicated non-engaged runs. Following initial group analyses that focused specifically on differences in DA, subsequent analyses were conducted only on runs with significant DA. This procedure ensures that significant results reflect intentional modulation rather than random variability.

Evaluation Against Chance: Although the theoretical chance level for two-class classification is 50%, actual chance levels can be influenced by trial count, label distribution, and temporal correlations [67]. To assess whether observed DA exceeded chance performance, a permutation test was conducted:

- 100 permutations of the trial labels (class 1 vs. class 2) were generated.

- For each permutation, the full six-fold CV procedure was repeated, producing 100 time-resolved random DA curves.
- At each time point, the empirical *p*-value was calculated as:

$$\boldsymbol{p} = \frac{\left|\left\{\boldsymbol{D}' \in \widehat{\boldsymbol{D}} : \boldsymbol{ac}(\boldsymbol{D}') \geq \boldsymbol{ac}(\boldsymbol{D})\right\}\right| + \mathbf{1}}{\boldsymbol{n} + \mathbf{1}} \quad (2)$$

where, $\widehat{D}$ is a set of n-randomised versions D' of the original data D, and $ac(D)$ is the accuracy achieved with the non-randomised data [68]. The null hypothesis (that observed DA could be obtained by chance) was rejected for α = 0.05. Time-resolved DA curves from both the original data and the permutation baseline were plotted together for visual comparison.

Paradigm-Wise Performance Tracking: Peak DA values from each experimental paradigm (Assessment, Training, Feedback, Q&A) were extracted for each participant and session. These were visualised in separate plots to monitor consistency across paradigms.

Frequency and topographical contribution analysis

To explore which frequency bands and EEG channels contributed most to class separation, the CSP filter outputs and associated mutual information (MI) weights were analysed post-calibration.

Band-wise Contribution (Heatmap Basis): At each time point *t*, the contribution of frequency band *b* to classification was computed as the average MI weight of all features n_b within that band:

$$\boldsymbol{M_{b,t}} = \frac{\sum_{\boldsymbol{n_b}} \boldsymbol{M_{b,t}^{n_b}}}{\boldsymbol{N_b}} \quad (3)$$

where $M_{b,t}^{n_b}$ is the MI weight of feature n_b at time *t*, and N_b is the number of features from band *b*.

Topographical Contribution per Band: To generate spatial maps of DA-relevant EEG activity, CSP filter outputs were weighted by their MI scores:

$$\boldsymbol{F_{b,n_b,i,t}} = \boldsymbol{C_{i,t}^{b,n_b}} \boldsymbol{M_t^{b,n_b}} \quad (4)$$

where $F_{b,i,t}$ is the cumulative contribution of EEG channel *i* within frequency band *b* at time *t*. This is computed by summing over all features n_b from band *b*, where each feature's contribution $F_{b,i,t}^{n_b}$ is the product of the CSP transformation value $C_{i,t}^{b,n_b}$ for channel *i*, and its mutual information weight M_t^{b,n_b}. Thus, Equation 1 aggregates these weighted contributions:

$$\boldsymbol{F_{b,i,t}} = \sum_{\boldsymbol{n_b}} \boldsymbol{F_{b,i,t}^{n_b}} \quad (5)$$

Multi-band Channel Contribution: To integrate across all frequency bands, a composite score for each channel was computed as:

$$\boldsymbol{F_{i,t}} = \frac{\sum_{\boldsymbol{b}} \boldsymbol{F_{i,t}^{b}}}{\boldsymbol{N}} \quad (6)$$

where *N* is the number of bands.

Task vs. Baseline Topographical Difference: Finally, to isolate spatial patterns specifically related to motor imagery (vs. baseline activity), the difference between multi-band contributions at peak task and baseline times was computed:

$$\boldsymbol{F_{i,t_1-t_0}} = \boldsymbol{F_{i,t_1}} - \boldsymbol{F_{i,t_0}} \quad (7)$$

where, F_{i,t_1-t_0} represents the difference in multi-band contributions for EEG channel *i* between the task period (F_{i,t_1}) and the reference period (F_{i,t_0}).

The topographical maps using mutual information weighted CSP filters are calculated at the sensor level using MATLAB [69] and the location of source activity are estimated and plotted using the sLORETA software package [70].

Online BCI configuration

In online runs, real-time auditory feedback was generated using the optimised FBCSP-MI configuration selected during offline calibration (see previous sections). For each time point t within a trial n, the classifier produced a time-varying signed distance (TSD) – a scalar value representing the projection of the feature vector onto the RLDA decision boundary [71,72]. The TSD value was computed as described in 8:

$$\boldsymbol{TSD_t^n} = \boldsymbol{w^T \bar{\omega}^n} - \boldsymbol{a_0} \quad (8)$$

where, w^T is the transposed weight vector (slope of the RLDA hyperplane), $\bar{\omega}^n$ is the feature vector at time *t* in trial *n*, and a_0 is the bias term (intercept of the RLDA hyperplane).

***I**nterpretation of TSD*: The sign of the TSD determines the predicted class (e.g., left vs. right imagery), and the magnitude of the TSD reflects the classifier confidence, which was mapped to the extent of auditory feedback movement (e.g., how far the sound panned toward the left or right ear).

Bias correction: To reduce systematic classification bias and improve feedback stability, the TSD was de-biased online by subtracting a rolling average of TSD values over the preceding 35 seconds. This dynamic correction helped stabilise the feedback direction over time, particularly in cases of class imbalance or drift.

Offline analysis of Q&A category results

To explore the feasibility of the SMR-BCI system as a communication tool for patients with PDoC, offline analysis examined whether DA varied as a function of participant group, question category, and experimental paradigm. The groups included patients diagnosed with UWS, MCS, LIS, and AB controls. The four question categories consisted of biographical, numerical, logical, and situational content. Experimental paradigms included assessment, training, feedback, and Q&A runs.

The analysis replicated the methods described in the Offline Single-Run Analysis for BCI Calibration, applying the same procedures for feature extraction, classification, cross-validation, and peak DA calculation. For Q&A runs specifically, trials were grouped by question category, and DA was computed separately for each participant and group. These results were then compared across groups and categories to identify patterns in BCI communication performance. Additionally, all Q&A trials across the four categories were pooled into a general 'question category' to assess overall decoding performance within each participant group. An exploratory analysis was performed to evaluate the question-level accuracies. For each question category (Biographical, Situational, Basic Logic, Numbers/Letters) and participant group (UWS, MCS, LIS, AB), we computed confusion matrices comparing target to decoded class, using smoothed peak DA as the decision metric in the task and reference periods separately. From these matrices we derived overall question-level accuracy, per-class bias indices, and per-class false-negative and miss rates.

Statistics and Reproducibility

All statistical analyses were performed using R-studio (version version 4.4.2) [73]. Two-tailed tests were used unless otherwise specified, with a 95% confidence interval (CI) applied throughout. Statistical results are reported using exact *p*-values where available; adjusted post-hoc comparisons are labelled and presented using standard threshold notation. The Shapiro–Wilk test is used to assess normality, and Levene's test evaluates the homogeneity of variances across groups. Prior to hypothesis testing, data are screened for outliers using the interquartile range (IQR) method. Outliers are retained unless they clearly influence the variance or distribution of the data, as such cases may reflect clinically meaningful variability rather than artefactual noise, given the heterogeneity of PDoC populations. To account for potential influence from these values, robust statistical approaches are employed. Specifically, for analyses of variance Welch's ANOVA is applied to mean DA data when assumptions of normality are met, as it is robust to unequal variances and sample sizes. The Kruskal–Wallis test is applied to median DA data, as it is appropriate for non-normally distributed data that may include meaningful extreme values. To control for Type I error across pairwise comparisons, where omnibus effects are significant, Games-Howell post-hoc tests are conducted for Welch's ANOVA, and Dunn's post-hoc tests with Holm correction are performed for the Kruskal-Wallis analyses. In addition, planned Welch contrasts are included to test a priori hypotheses comparing (1) UWS against the combined mean of MCS and LIS, and (2) MCS against LIS. In each case, sample size (including *n* per group), means and standard deviations are reported descriptively in the Supporting Information, Supplementary Table 1 (a) and (b)). These tables include the AB group as a benchmark. While not included in the inferential analyses, descriptive data for the AB group are provided to illustrate the performance of all diagnostic groups relative to healthy individuals. For further visualisation, the AB group is included in Figure 5-Figure 6, and Figure 9. For all analyses of DA data, unless specified, only DA values for significant runs were used, i.e., motor imagery (MI) -BCI runs where peak DA during the task period was significantly higher than the corresponding peak baseline DA (one-tailed *t*-test, $\alpha = 0.05$; consistent with our directional hypothesis, see *Statistical Comparison of Task vs. Baseline DAs* in Methods Section, Offline single-run analysis for BCI calibration). In between-group analyses and correlations, each participant constitutes one independent observation. In linear mixed-effects model (LMM) analyses, repeated measures of DA across paradigms are treated as within-subject replicates (statistically modelled via random intercepts).

Bivariate associations were evaluated using Pearson's correlation. Partial correlations were computed to control for group membership, which was entered as a categorical variable using dummy coding. To examine DA across task paradigms within each group, a LMM was fitted with *task paradigm* as a fixed effect and participant as a random intercept to account for repeated measures. Model

assumptions of normality, homoscedasticity, and linearity were assessed using residuals versus fitted values and Q–Q plots. Where significant effects of paradigm were identified, estimated marginal means (EMMs) were computed, and pairwise comparisons between paradigms were performed using Holm-adjusted p-values to control for Type I error. Bivariate associations were evaluated using Pearson's correlation. Partial correlations were computed to control for group membership, which was entered as a categorical variable using dummy coding.

To predict diagnostic category (UWS, MCS, LIS), a multinomial logistic regression was performed using DA, CRS-R, and WHIM scores as predictors, with injury type (Traumatic = 1; Non-Traumatic = 0) and time since injury (in months) included as covariates. Model performance was evaluated using leave-one-subject-out (LOSO) cross-validation, in which the model was trained on all participants except one, and then used to predict the diagnosis of the held-out participant. This procedure was repeated for each participant to provide subject-level out-of-sample predictions. Overall, accuracy, balanced accuracy, and macro-F1 scores were calculated from the aggregated LOSO predictions, with uncertainty quantified via exact and bootstrap confidence intervals. To explore whether integrating DA and behavioural measures improves diagnostic prediction, we performed a post-hoc ensemble analysis that combined three single-measure models – DA, CRS-R, and WHIM, each fitted with the same covariates as the main multinomial model. Using leave-one-subject-out (LOSO), we obtained held-out predictions and class probabilities for every participant from each model. For the ensemble, the final predicted class for each participant was determined by majority vote across the three models; in cases of a three-way tie, we applied probability-based tie-breaking by averaging per-class probabilities and choosing the maximum. Ensemble performance (accuracy, balanced accuracy, macro-F1, and class-wise recall with confidence intervals) was computed from these LOSO-derived predictions to allow direct comparison with the individual predictor measures and the combined multinomial model.

Topographical analysis methods

Topographical contributions of cortical regions were assessed using data from runs with statistically significant DA. For each task period, common spatial pattern (CSP) weights were multiplied by their corresponding mutual information (MI) scores to determine the contribution of each electrode. Following the interpretive framework proposed by Haufe et al. (2014)[74], these MI-weighted CSP values are treated as activation patterns (i.e., estimates of each sensor's relative contribution to classification performance), rather than as direct representations of cortical source activity. To aid qualitative visualisation, these sensor-level activation patterns were optionally projected into source space using sLORETA; this step was exploratory and not intended as a quantitative inverse solution. Accordingly, scalp-level MI-weighted CSP maps remain the primary analytical result. Electrodes were grouped by cortical region: frontal (F3, Fz, F4), left-temporal (T7, C3), right-temporal (C4, T8), sensorimotor (C3, Cz, C4), parietal (P3, Pz, P4), and occipital (PO7, Oz, PO8)[75].

For each paradigm (assessment, training, feedback, Q&A), MI-weighted CSP values are averaged across electrodes within each cortical region and then across participants, such that each participant contributes one independent observation per region to the group-level comparisons.

Functional Connectivity analysis methods

Functional connectivity (FC) differences between diagnostic groups were evaluated during motor imagery (MI) task performance, using source-level, frequency-specific connectivity measures. The analysis aimed to identify how task-evoked network dynamics varied across diagnostic categories.

The networks of interest included the Default Mode Network (DMN), Fronto-Parietal Network (FPN), Auditory Network, and Thalamocortical Network, due to their respective roles in endogenous awareness [39,76], executive control [39,77], exogenous awareness and auditory processing [33,39,78], and thalamocortical information integration [39,79,80]. Nodes used for each network were defined as follows – DMN: medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC); FPN: left and right dorsolateral prefrontal cortex (l-dlPFC, r-dlPFC), and left and right inferior parietal lobule (l-IPL, r-IPL); Thalamocortical: left and right thalamus (l-Thal, r-Thal); and Auditory: left and right superior temporal gyrus (l-STG, r-STG). The Montreal Neurological Institute (MNI) coordinates for the locations of the nodes are presented in Supporting Information: Additional Exploratory Results, Supplementary Table 6), replicated from the Aubinet et al. (2018) study examining resting-state FC in a cohort of MCS patients [39].

Connectivity was computed within six frequency bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), low-beta (12–18 Hz), high-beta (18–28 Hz), and low-gamma (28–40 Hz). Data were epoched into 1-second windows centred on the peak DA time point within each significant MI trial. Only runs that showed statistically significant DA were included.

EEG source localisation was performed using the New York head model [81], which incorporates standard electrode positions and a forward model aligned to the Montreal Neurological Institute (MNI) template [82]. In this template, the forward model is constrained to a priori selected dipoles with perpendicular orientation to the cortical surface. The dipole locations are represented by the network nodes of interest listed in Supplementary Table 6. To solve the source localisation problem, eLORETA was applied, as it provides zero localisation error under ideal conditions [83,84] and an estimated localisation error of ~1 cm for general scenarios [85–87]. The inverse solution equations followed from the formulation described in equations Eq. 32 and Eq. 37 by Pascual-Marqui et al. (2007) [88], using a regularisation parameter $\alpha = 0.01$.

FC was estimated between the predefined cortical nodes using imaginary coherence (iCOH). iCOH quantifies the strength of phase coupling between two source signals, while minimising spurious effects from volume conduction and common source artifacts [89].

The iCOH measure between signals $x(t)$ and $y(t)$ at frequency f was computed as:

$$\mathrm{iCOH}_{xy}(f) = \frac{\mathrm{Im}[S_{xy}(f)]}{\sqrt{S_{xx}(f)\, S_{yy}(f)}} \tag{9}$$

where, $S_{xy}(f)$ is the cross-spectrum between x and y, $S_{xx}(f)$ and $S_{yy}(f)$ are their respective power spectrum values at f , and Im[·] denotes the imaginary component.

Group-level FC differences were assessed using Spectral Functional Connectivity Difference (SFCD) matrices, derived from the iCOH signed values computed between all the 66 paired interactions of twelve predefined cortical nodes (regions) across 6 frequency bands. Firstly, epoched regional activity was concatenated for all the participants in each diagnostic group, then the FC metric was estimated for each group and SFCD matrices were obtained from the group differences. In parallel, a non-parametric max-statistics approach was implemented using a Monte Carlo permutation test (1000 replications). Here, the condition labels were shuffled prior to the calculation of a surrogate SFCD, finally resulting in a single value corresponding to the maximum absolute difference value of SFCD elements, for each replication. Under the null hypothesis of non-differences between the conditions, this non-parametric max-statistic test, provides a distribution for the evaluation of significant

differences between the conditions while controlling for multiple comparisons. Significant differences were tested with a threshold of $\alpha = 0.01$.

Manuscript editing

To improve clarity and conciseness, manuscript editing was aided by ChatGPT (OpenAI, San Francisco, CA). All content was reviewed and verified by the authors for accuracy and appropriateness.

Results

Decoding accuracy examples and metrics

AB participants served as a benchmark representing optimal MI-BCI performance, providing a reference framework for interpreting diagnostic group outcomes. Following Phase I (assessment), 8 of 14 UWS, 13 of 17 MCS, 10 of 11 LIS patients, and both AB participants progressed to Phase II (training and feedback; *N* = 33, 73.8% of patients passed assessment). Having completed Phase II, 6 UWS, 12 MCS, 10 LIS, and both AB participants advanced to Phase III (Q&A; *N* = 30, 90% of patients progressed to Q&A). The number of sessions completed by each patient varied between 4 and 13 – however, all patients who progressed to Phase III (Q&Q) completed a minimum of 9 sessions. The two AB participants completed 4 and 5 sessions, respectively – one completed 20 runs (including 4 Q&A), and the other completed 15 runs (including 2 Q&A). Decoding Accuracy (DA) refers to the percentage of correctly classified imagined movements – and throughout the analyses, 'significant runs' refer to MI BCI runs where DA during the task period was significantly higher than during the corresponding baseline (one-tailed t-test, $\alpha = 0.05$).

Examples of decoding accuracy dynamics and associated spectral contributions for individual participants from each diagnostic group are provided in the Supporting Information: Methods (Supplementary Figure 6). As shown there, significant task-related modulation of brain activity is reflected by DA peaks exceeding baseline and permutation-derived chance levels, with corresponding time-frequency maps highlighting sensorimotor rhythm contributions, typically within mu and beta bands. These examples serve to illustrate the physiological basis of decoding accuracy and validate the feature-selection and classification approach used in the main analyses.

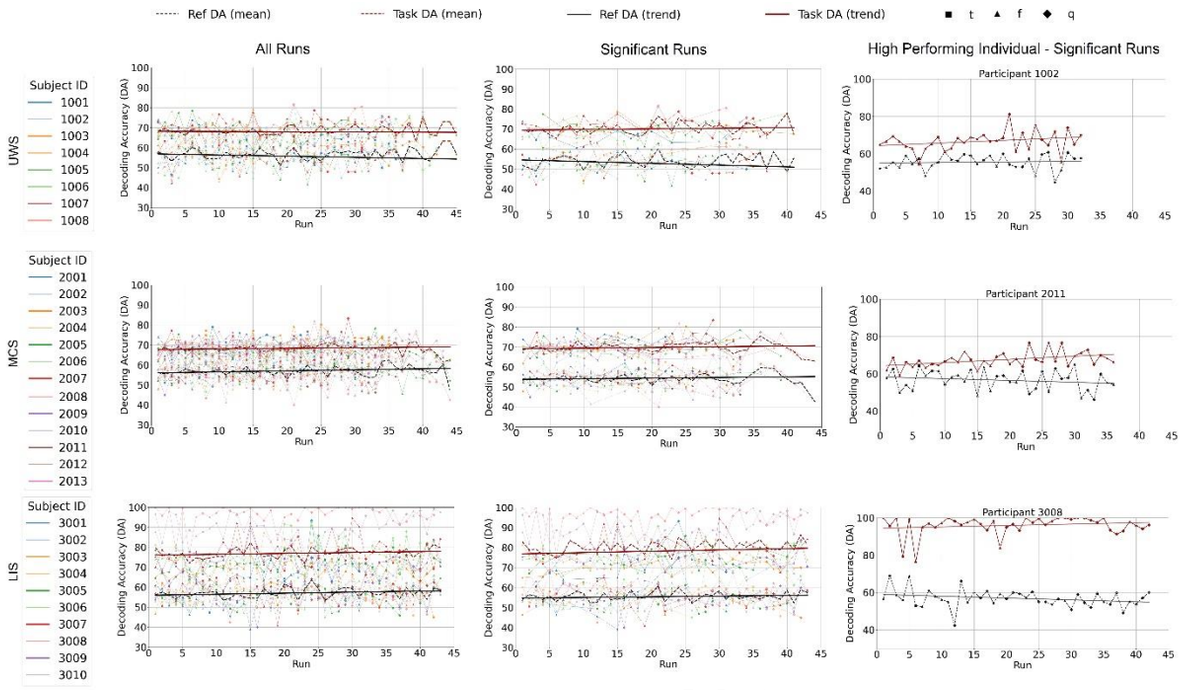

Figure 4. Decoding accuracy (DA) across runs for all participants. Graphs are displayed according to diagnostic category (top row – UWS; middle row – MCS; bottom row – LIS), shown separately for all runs (left-hand column) and runs reaching statistical significance (middle column); group-level trajectories for Ref DA (black) and Task DA (red), including participant traces (see subject ID for each group), mean trajectories (dashed), and participant-averaged linear trends (solid). Representative high-performing individuals, illustrating stable high performance and task-reference separation, are presented in the right-hand column.

Inspection of participant- and run-specific DA, visualised in Figure 4, reveal substantial variability in both initial performance and longitudinal patterns across runs and sessions. Some LIS participants demonstrated consistently high task decoding accuracy throughout the protocol, whereas others exhibited gradual improvements over time (see trendlines in Figure 4). Strong task-related DA was not uniformly accompanied by corresponding changes in reference/baseline decoding accuracy, particularly for PDoC participants, indicating that task-specific modulation and baseline performance can vary independently across these individuals. Together, these patterns illustrate overall heterogeneity in MI-BCI performance expression across participants, regardless of diagnostic category.

Group analysis

Decoding accuracy (DA) was calculated for each participant and paradigm (assessment, training, feedback, Q&A) using the procedure described in the *Offline Single-Run Analysis* section. Subsequent group comparisons were based on these calculated DA values and conducted in two ways: (1) using all runs (Figure 5a), and (2) using only runs where task-related DA significantly exceeded baseline DA (one-tailed t-test, $\alpha = 0.05$; Figure 5b).

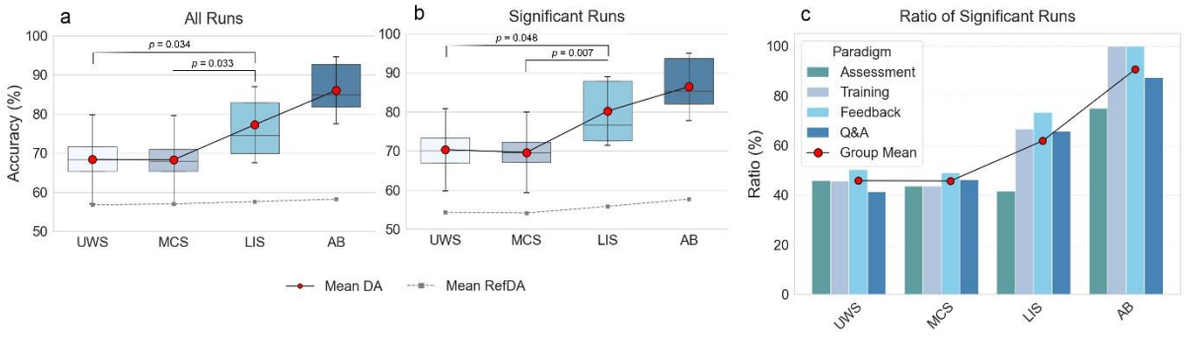

Figure 5. Illustration of the comparison of DA values obtained for each group (UWS, MCS, LIS, AB). DA values obtained from (a) all runs (*n* = 1233 run/33 participants; UWS = 293/8, MCS = 472/13, LIS = 433/10, AB = 35/2) and (b) runs showing a significant task-related peak (one-tailed *t*-test comparing task vs. reference interval, α = 0.05; *n* = 681 run/31 participants; UWS = 138/8, MCS = 211/13, LIS = 299/10, AB = 33/2). Boxplots display the data distribution for each group, showing the interquartile range (25th–75th percentile) with the median as a horizontal line. Individual data points represent DA values from multiple runs per participant, illustrating within-group variability. Red circles mark group mean DA, with vertical error bars showing ±1 *SD*. Black circles and solid lines connect mean task-related DA across groups, while grey squares and dashed lines indicate mean DA during the reference interval. Panel (c) illustrates the proportion of significant runs per paradigm (assessment, training, feedback, and Q&A), by group. Bars represent the mean percentage of significant runs for each group and paradigm. Individual data points represent proportions from each participant's runs, illustrating within-group variability. Red points and connecting lines indicate group-level averages across paradigms. Exact P values for diagnostic group comparisons are displayed in the figure.

A Kruskal-Wallis test on data from all runs revealed a significant group effect ($H_{(2)} = 12.047$, $p = 0.0012$, $\varepsilon^2 = 0.36$). Dunn post hoc comparisons with Holm correction showed the LIS group achieved significantly higher DA than both MCS ($p = 0.033$) and UWS ($p = 0.034$) groups, with no significant difference between MCS and UWS. Restricting the analysis to significant runs (Figure 5b) confirmed group differences (Kruskal-Wallis: ($H_{(2)} = 10.026$, $p = 0.007$, $\varepsilon^2 = 0.29$). Dunn post hoc tests with Holm correction again showed higher DA in the LIS group compared to MCS ($p = 0.007$), and UWS ($p = 0.048$). Figure 5c shows the proportion of significant runs per paradigm and group. A Kruskal-Wallis test was used to evaluate diagnostic group variance in the proportion of significant runs across paradigms and did not find significant group differences ($H_{(2)} = 4.65$, $p = 0.098$, $\varepsilon^2 = 0.12$).

Descriptive statistics for both tests indicated that the AB group's mean and median were higher than those of the diagnostic groups, providing a benchmark for interpreting patients' MI results (Supporting Information, Supplementary Table 1(a)).

Q&A question category analysis, across diagnostic groups

In the Q&A phase, the MI-BCI was used primarily as a cognitive assessment tool to examine whether participants with PDoC could intentionally modulate neural activity in response to different categories of questions (biographical, logical, numerical, and situational). While the diagnostic potential of the framework is the focus of the earlier phases, this phase was designed to probe domain-specific awareness and intentional modulation, with the results also providing initial indications of future potential as a communication aid for patients with PDoC and CLIS.

Welch ANOVAs revealed significant group differences (all runs; $F_{(2,\ 13.69)} = 4.64$, $p = 0.029$, $\omega^2_{adj} = 0.3$, significant runs; ($F_{(2,\ 13.52)} = 7.94$, $p = 0.005$, $\omega^2_{adj} = 0.46$). Games-Howell post-hoc tests revealed that, with all runs included, LIS had higher DA than MCS ($p = 0.028$) and UWS ($p = 0.042$). When analyses

included significant runs only, DA was higher for the UWS group compared to the MCS group (p = 0.049) and not compared with the LIS group. A significant difference was also observed between the LIS and MCS groups (p = 0.008).

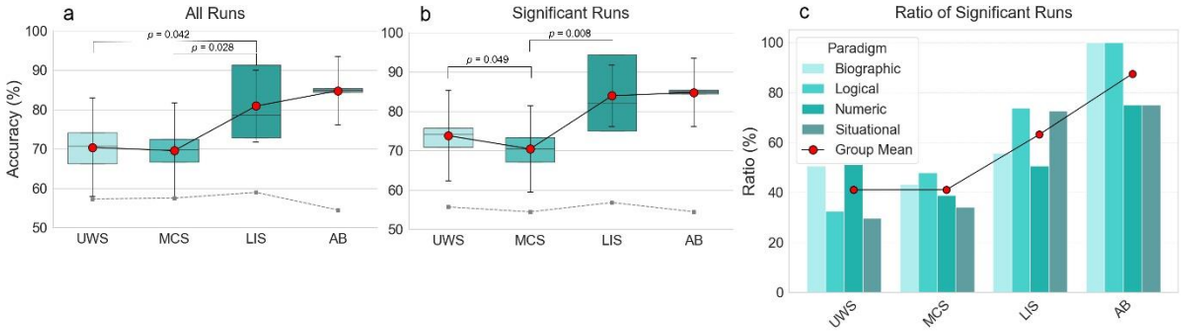

Figure 6. Comparison of decoding accuracy (DA) values obtained for Q&A categories, across groups (UWS, MCS, LIS, AB). DA values obtained from (a) all runs (n = 297 run/30 participants; UWS = 57/6, MCS = 112/12, LIS = 124/10, AB = 4/2) and (b) runs showing a significant task-related peak (one-tailed t-test comparing task vs. reference interval, $\alpha = 0.05$; n = 169 run/29 participants; UWS = 26/5, MCS = 51/12, LIS = 88/10, AB = 4/2). Boxplots display the data distribution for each group, showing the interquartile range (25th–75th percentile) with the median as a horizontal line. Individual data points represent DA values from multiple runs per participant, illustrating within-group variability. Red circles mark group mean DA, with vertical error bars showing ±1 *SD*. Black circles and solid lines connect mean Q&A task-related DA across groups, while grey squares and dashed lines indicate mean DA during the reference interval. Panel (c) illustrates the proportion of significant runs per Q&A category (Biographical, Basic Logic (Logic), Numbers and Letters (Numeric), and Situational), by group. Bars represent the mean percentage of significant runs for each group and question category. Individual data points represent proportions from each participant's runs, illustrating within-group variability. Red points and connecting lines indicate group-level averages across paradigms. Exact P values for diagnostic group comparisons are displayed in the figure.

The proportion of significant runs by question category is illustrated in (Figure 6c). Again a Kruskal-Wallis test evaluating diagnostic group variance in the proportion of significant runs across question categories and did not find significant group differences ($H_{(2)} = 4.98$, $p = 0.083$, $\varepsilon^2 = 0.12$). Descriptive statistics for Q&A session performance, based on both all runs and significant runs only, show higher mean and median values for the AB group, again providing a normative reference for evaluating patient performance (Supporting Information, Supplementary Table 1(b)). Exploratory analyses of question-level accuracies computed using confusion matrices, across cognitive categories are provided in Supporting Information: Additional Exploratory Results. Supplementary Tables 4 (a and b) summarise the accuracy of decode responses, with interpretive guidance provided in Supplementary Note 3. Supplementary Table 5, together with Supplementary Note 4 and 5, summarises question-level accuracy rankings by group. UWS and MCS participants achieved mean accuracies of 76-78%, and LIS participants exceeded 80%, whereas accuracies derived from baseline intervals were substantially lower 54-65.5%).

Paradigm-specific differences in MI-BCI Decoding Accuracy across diagnostic groups

A linear mixed-effects model (LMM) was fitted with task paradigm as a fixed effect and participant as a random intercept, to test whether MI-BCI performance varied across task paradigms. The model included significant MI-BCI runs only. The assessment paradigm was excluded because it employed a

block design, which introduces temporal correlations that may inflate classification performance by capturing slow signal drift rather than stimulus-specific neural responses [90]. As this repeated-measures analysis focuses on within-subject variation, including a structurally different paradigm would confound the comparison. Consequently, the training paradigm (no feedback) was used as the reference condition.

For the UWS group, adding random intercepts did not significantly improve model fit (p = 0.13), whereas including task paradigm as a fixed effect significantly improved fit compared to the null model (p = 1.76 x10^{-6}, see green band in Figure 7). Pairwise comparisons showed higher DA during Q&A compared to training (p = 0.002) and feedback (p = 1.22 x10^{-6}), with no difference between training and feedback (p = 0.12). The Intraclass Correlation Coefficient (ICC) indicated a low between-participant variance (ICC = 0.076).

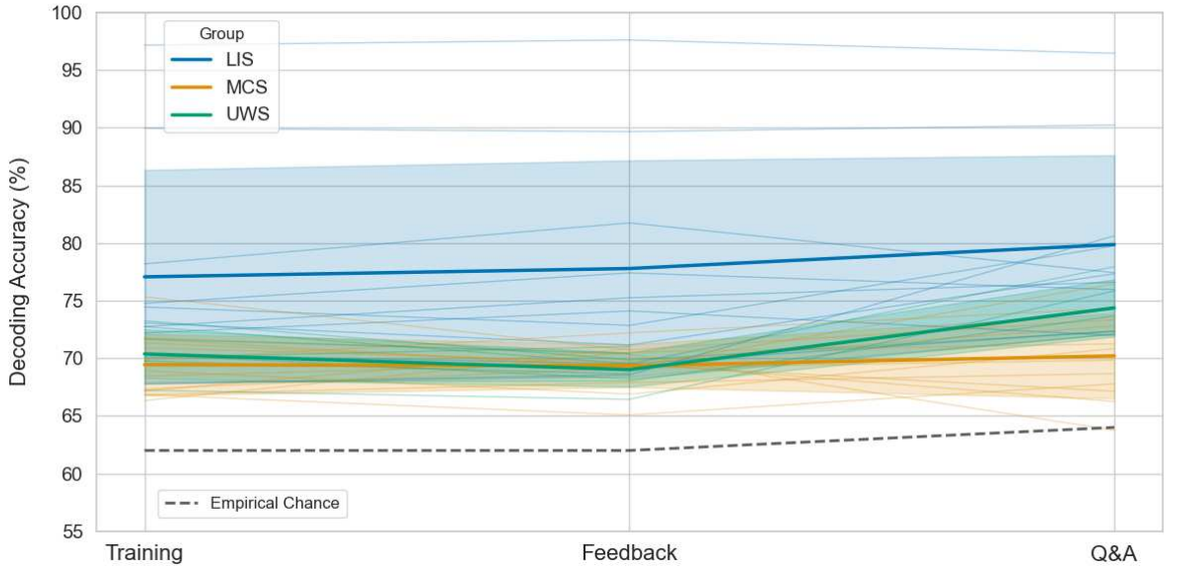

Figure 7. Mean DA across task paradigms (Training, Feedback, and Q&A) for significant runs only, shown separately for diagnostic groups (UWS, *n* = 8; MCS, *n* = 13; LIS, *n* = 10;). Thin lines represent individual participants, illustrating within-group variability. Bold lines and shaded bands denote group-level means ±1 SD. The dashed line indicates empirical chance levels for each paradigm, computed using the "better-than-random" binomial rule [91]. Group colours: UWS = green, MCS = orange, LIS = blue.

In contrast, random intercepts significantly improved model fit for both the MCS (p = 0.009) and the LIS (p = 2.76 x10^{-77}) group, while paradigm had no significant effect on DA in either group (α = 0.05). Pairwise comparisons were not significant (α = 0.05). However, between-participant variance was low in the MCS group (ICC = 0.103), suggesting that most variability occurred within individuals across trials (as opposed to paradigms); while in the LIS group, ICC was high (0.731), indicating that inter-individual differences were the dominant source of variation in DA.

Between-participant variance was low for the MCS group (ICC = 0.103) but high for the LIS group (ICC = 0.731), indicating that variability in DA for LIS was driven mainly by inter-individual differences. In contrast, MCS variability arose primarily within participants across trials rather than across paradigms.

These results suggest that only the UWS group was sensitive to paradigm differences, whereas inter-individual variability among LIS patients accounted for the majority of variance in DA (see blue band in Figure 7).

Analysis of Decoding Accuracy and CRS-R and WHIM assessment scores

CRS-R and WHIM scores were strongly correlated (Pearson's $r(28) = 0.66$, $p = 6.94 \times 10^{-5}$, 95% CI [0.40, 0.82]), indicating substantial agreement between both behavioural assessment measures. Partial correlations between participants peak DA and CRS-R and WHIM scores, controlling for diagnostic group, revealed DA was significantly associated with WHIM (partial Pearson's $r(28) = 0.43$, $p = 0.02$), but not with CRS-R ($p = 0.89$) measures. Group differences in both CRS-R and WHIM scores were significant, indicating both measures significantly differentiate between the diagnostic groups. For CRS-R: Welch ANOVA; $F_{(2, 17.55)} = 42.61$, $p = 1.84 \times 10^{-7}$, $\omega^2_{adj} = 0.8$ (Games-Howell comparisons; UWS MCS $p = 0.002$; UWS vs LIS $p = 5.6 \times 10^{-7}$; MCS vs LIS $p = 0.012$). For WHIM: Welch ANOVA; $F_{(2, 15.55)} = 11.42$, $p = 8.87 \times 10^{-4}$, $\omega^2_{adj} = 0.53$ (Games-Howell comparisons; UWS vs MCS $p = 0.012$; UWS vs LIS $p = 0.004$; MCS vs LIS $p = 0.018$). However, given the CRS-R and/or WHIM (and/or similar) measures were employed to form the diagnosis, these results are not surprising. At issue is whether diagnoses formed using assessment measures that require overt behavioural responses are accurate.

Therefore, to evaluate the predictive value of DA, CRS-R, and WHIM for diagnostic classification, we fitted a multinomial logistic regression model with diagnosis (UWS, MCS, LIS) as the outcome. UWS was set as the reference category. The model included DA, CRS-R, and WHIM as predictors, with injury type (TBI vs. non-TBI) and time since injury (IOT) included as covariates. To provide subject-level external validation, predictive performance was estimated using leave-one-subject-out (LOSO) cross-validation. This approach ensured that each participant's diagnosis was predicted from a model trained on all other participants, avoiding within-sample bias. A balanced accuracy of 55.4% (CI: 38-73%) was observed, indicating moderate classification performance across groups – correctly identifying 4/8 UWS, 5/13 MCS, and 7/9 LIS cases. The model performed best for LIS classification (recall = 77.78%, CI: 45-94%), and was weaker for MCS (recall = 38.46%, CI: 18-64%). While predictive metrics were modest and uncertainty wide, the LIS group shows the clearest separation from UWS (see Figure 8). A complimentary full-sample association analysis found MI-BCI DA has the strongest association with LIS relative to UWS, followed by CRS-R however, as association analyses are not cross-validated, this finding cannot be taken as evidence of out-of-sample predictive performance.

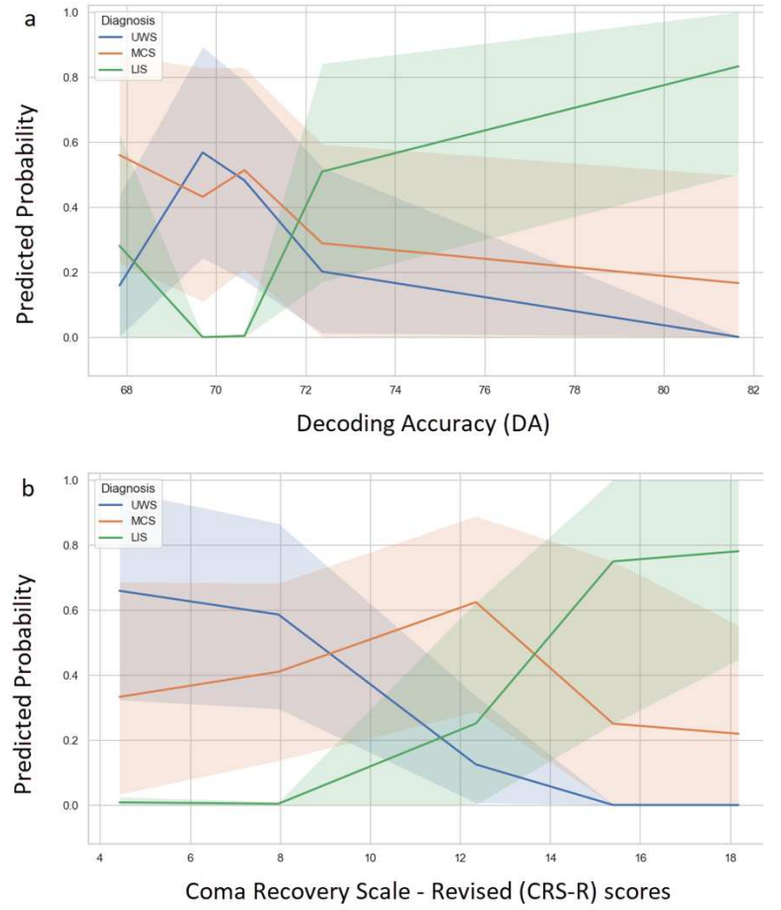

Figure 8. Predicted probabilities of clinical diagnosis (UWS, MCS, LIS) derived from a multinomial logistic regression model as a function of (a) average decoding accuracy (DA) and (b) CRS-R score. The model was fitted using participant-level averages within each diagnostic group. Lines show model-estimated probabilities with 95% bootstrapped confidence intervals. Colours denote diagnostic groups (UWS = blue, MCS = red, LIS = green).

As an exploratory post hoc analysis, we examined whether combining all three single-measure models – DA, CRS-R, and WHIM (each fitted with the same covariates as the main multinomial model) – improves classification relative to any single predictor measure. Using LOSO held-out predictions, single-modality balanced accuracies were 57.3% for DA, 49.4% for CRS-R, and 59.4% for WHIM. A three-model majority-vote ensemble with probability-based tie-breaking achieved a balanced accuracy of 61.81% ((95% CI: 43-80.4), macro-F1 = 0.633, and class-wise recalls of 50.0% (UWS), 69.2% (MCS), and 66.7% (LIS).

A ~5% improvement in diagnostic accuracy is achieved when integrating MI-BCI DA as a metric alongside behavioural (CRS-R, WHIM) predictor measures; most notably improving recall for the intermediate MCS group – while acknowledging the wide uncertainty around the estimates. However, the reduced predictive performance for LIS in the integrated ensemble model suggests that MI-BCI DA alone remains the most robust predictor of LIS.

Brain topography analysis

Topographical maps were visually inspected to identify brain regions contributing most to MI task classification. These maps represent sensor-level activation patterns derived from MI-weighted CSP

values, indicating relative discriminative contributions of each electrode rather direct neural source estimates [74]. sLORETA projections provide a qualitative visualisation to assess whether scalp-level contributions corresponded to expected sensorimotor regions. Task-related activity was calculated as the difference between common spatial patterns-mutual information (CSP-MI) weights during task and baseline periods, for each group. In the AB (benchmark) group, activity was consistently localised to sensorimotor and somatosensory regions across paradigms (see Supplementary Figure 1(d)). In contrast, the PDoC and LIS groups (particularly PDoC) showed more spatially diverse and less stereotyped patterns (Supplementary Figure 1(a–c)).

Group differences in regional CSP-MI activity were evaluated through an analysis of task-related CSP-MI weights across six cortical regions; frontal, left temporal, right temporal, sensorimotor, parietal, and occipital – within each paradigm (full details in Methods; significant results in Supporting Information, Supplementary Data 1 (F) and descriptive statistics in Supplementary Data 1 (G).)

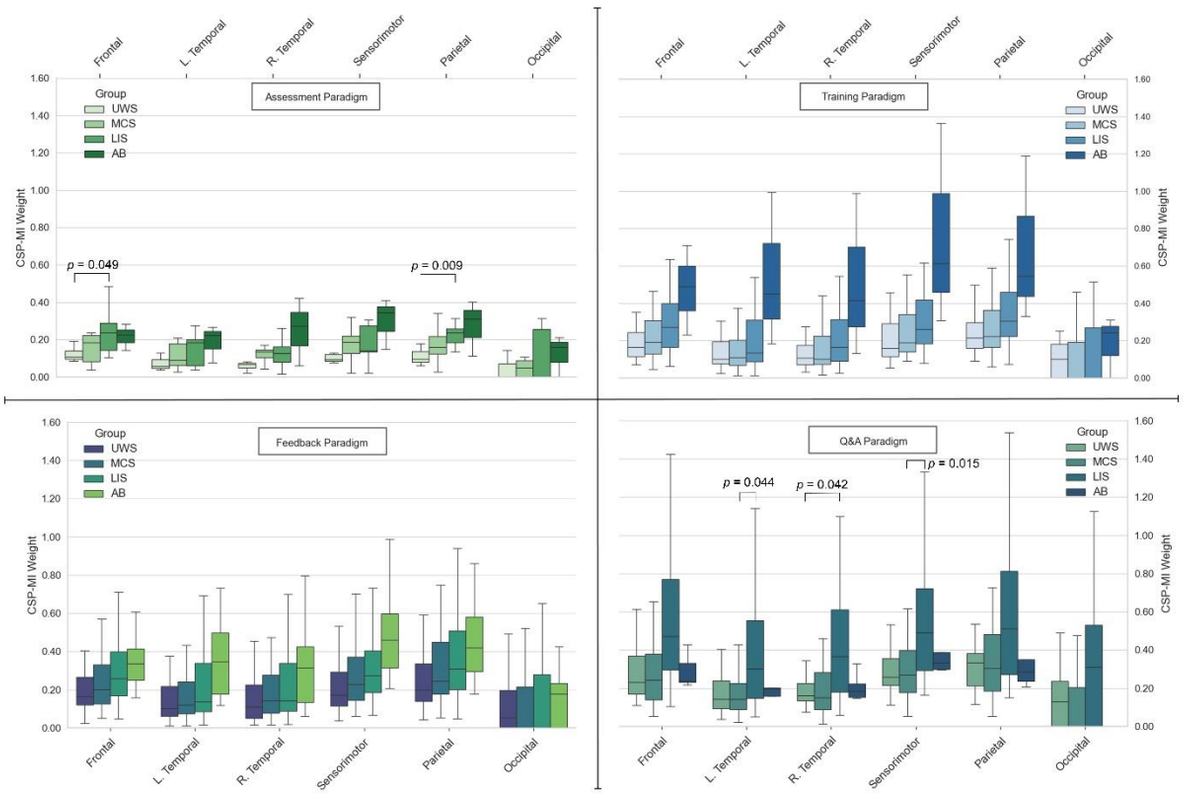

Figure 9. Clustered boxplots illustrating CSP–MI weights across brain regions for each diagnostic group (UWS, MCS, LIS, AB) and paradigm: Assessment (top left), Training (top right), Feedback (bottom left), and Q&A (bottom right). Each box represents the distribution of CSP-MI weights across individual runs (multiple runs per participant) within each group and brain region, with boxes showing the interquartile range (25th-75th percentile) and horizontal lines marking the median. Whiskers extend to 1.5× the interquartile range, and outliers are not displayed. Colours denote diagnostic groups as indicated in the legend. Exact P values for diagnostic group comparisons are displayed in the figure. See Supplementary Data 1 (H) for full details on the graphed units of observation.

The results show group differences in MI-weighted CSP values emerged only in brain regions contributing to MI task classification during the Assessment and Q&A paradigms, with higher MI-weighted CSP values distinguishing LIS from PDoC groups (as illustrated in Figure 9). Notably, during

the Assessment paradigm the difference in MI-weighted CSP values between LIS and UWS in the parietal region was significant (Welch ANOVA and Games-Howell post-hoc, $F_{(2,\ 11.8)} = 7.16$, $p = 0.009$, $\omega^2_{adj} = 0.45$, LIS>UWS, $p = 0.009$) – and during the Q&A paradigm, the difference between LIS and MCS in the sensorimotor region was significant (Kruskal-Wallis and Dunn post-hoc tests with Holm correction; $H_{(2)} = 8.801$, $p = 0.012$, $\varepsilon^2 = 0.28$, LIS>MCS, $p = 0.015$). These results suggest that LIS participants produce stronger and more discriminable motor imagery-related activation patterns compared with PDoC patients, particularly early on in the protocol, as well as in the Q&A task which requires a yes/no response to a question as opposed to simple command following.

Functional connectivity

Exploratory source-level FC analyses suggested group differences largely consistent with known thalamo-frontal and DMN–FPN disruptions in UWS compared with MCS and LIS. These patterns were most evident in the alpha band, extending into the theta band for UWS–LIS comparisons during the Q&A task (see Supplementary Figure 7). Given the low-density (16-channel) montage, these source-space results are presented as exploratory only; full details are reported in Supporting Information: Additional Exploratory Results, comprising Supplementary Tables 6 and 7, Supplementary Note 6, and Supplementary Figure 7. The main conclusions are instead supported by sensor-level and aggregate source-space summaries.

Discussion

The current study evaluated a multi-phase MI-BCI protocol across four groups: patients with Prolonged Disorders of Consciousness (PDoC) (unresponsive wakefulness syndrome (UWS, *n*=14) and minimally conscious state (MCS, *n*=17), Locked-In Syndrome (LIS, *n* = 11) and able-bodied controls (AB, *n* = 2). The AB participants were included as descriptive benchmarks of the upper bound of decoding performance under limited session exposure, rather than as part of the formal inferential tests, since they completed fewer sessions than the patients. As expected, descriptive statistics show they achieved the highest decoding accuracies (DAs); however, peak performances in some diagnostic groups occasionally approached these benchmark levels [92,93]. All primary statistical analyses and conclusions remain focused on the diagnostic groups (UWS, MCS, LIS).

The three-phase protocol comprised: Phase I, assessment of initial motor imagery (MI) modulation; Phase II, MI training with neurofeedback; and Phase III, binary question–response testing using closed questions from specific categories, designed to evaluate the potential for cognitive profiling and domain-specific awareness with the MI-BCI framework. Overall, 8/11 UWS, 13/17 MCS, and 10/11 LIS patients (73.8% of recruited patients) achieved significant DA during Phase I, with 6/8 UWS, 12/13 MCS and 10/10 LIS (90% of the patients who passed the Assessment phase) completing phase III – suggesting volitional modulation of brain activity was preserved in a high proportion of patients within this study cohort. An evaluation of diagnostic group differences in the proportion of significant runs, across (a) paradigms and (b) question categories, revealed no significant effects, suggesting comparable (a) motor imagery capacity and (b) potential ability to communicate yes-no answers, across diagnostic groups (see Figure 5(c) and Figure 6(c)). The latter interpretation remains theoretical and requires further empirical testing to assess whether patients can reliably use motor imagery to communicate yes-no responses on a single trial basis however the results demonstrate the average overall response to repetitions of questions may enable, albeit slow, communication.

As illustrated in Figure 4, group averages can obscure inter-individual variability – masking individuals whose MI-BCI performance is substantially better than is suggested by their diagnostic group mean. The marked heterogeneity in MI-BCI performance across PDoC groups reinforces the growing

consensus that preserved covert awareness may be underestimated in patients with PDoC, when assessment relies on tools that require overt behavioural responses [96–98]. Furthermore, the presence of exceptionally high-performing LIS participants highlights the potential for some individuals with LIS to approach maximal DA, in a short time-frame. Such cases motivate the development of responder-focused, individually tailored MI-BCI protocols.

In binary motor imagery BCIs, classification accuracies of approximately 70% are commonly adopted as a criterion for reliable control, reflecting statistically significant performance above chance and practical usability, while higher accuracies (≥75–80%) are typically required for sustained online or clinical applications [67,94,95] . Here, we report within-run cross-validation accuracies, as the primary focus is group-level comparative analysis. Nevertheless, the individual results shown in Figure 4 are indicative of performance that exceeds the commonly adopted criterion level for BCI utilisation. It should be noted, however, that inter-run and inter-session variability, which is more pronounced in these participant groups due to factors such as fluctuating awareness, fatigue, and engagement, makes it more difficult to assume reliable cross-run DA.

The multi-stage protocol adopted in this study, incorporating real-time continuous feedback, is intended as a stepping stone toward training participants to produce reliable single-trial responses. The gradual performance improvements observed in a subset of participants (Figure 4) suggest that feedback-driven learning and sustained engagement can enhance MI control. These findings indicate that performance may further improve with extended exposure, real-time adaptive feedback, and motivational engagement paradigms.

When statistical analyses of group variance in DA only included significant runs, the difference in overall DA between the LIS and UWS groups was significant (p = 0.048), whereas their DA during the Q&A sessions did not differ (p = 0.13). In contrast, the difference in DA between the LIS and MCS groups remained consistently significant. While there was no overall DA difference between the MCS and UWS groups (p = 0.87), a significant difference emerged between these PDoC groups during the Q&A sessions (p = 0.049). This pattern suggests that, when analyses were performed using only significant runs, UWS participants performed marginally better than the MCS group during Q&A sessions, thereby narrowing the apparent MI performance gap with the LIS group. Further support for this finding was provided by the significant variation in Q&A MI performance compared to Training and Feedback, in the UWS group (see Figure 7). Notably, the Q&A task presented pre-recorded questions spoken by a familiar caregiver or relative. The suggestion is that emotionally salient auditory input potentially enhanced cognitive engagement in UWS patients, consistent with prior evidence showing that meaningful stimuli can activate residual cortical networks in UWS despite widespread disconnection [99]. That this paradigm-specific effect emerged only in UWS indicates a potential moderating effect of emotionally salient stimuli on BCI engagement in lower awareness states. Additionally – or alternatively – the UWS participants may collectively exhibit more awareness than previously assumed, which was not evident through standard behavioural diagnostic tests but became observable over time, via movement-independent responses using the MI-BCI.

Grouping questions into distinct cognitive categories enabled assessment of category-specific decoding accuracy as a means of probing differential conscious awareness, rather than validating the MI-BCI as a communication channel per se. The randomised presentation and balanced phrasing of Yes/No questions minimised the likelihood of auditory or linguistic confounds in significant runs. In addition, all auditory cues were temporally and procedurally separated from the MI task window, identical across conditions, and too brief to account for sustained decoding differences; occasional tactile re-alerting occurred only between runs. These controls make it unlikely that the observed MI-specific effects were driven by auditory or arousal-related factors, although we acknowledge this as a

potential limitation. These preliminary findings indicate that MI-BCI responses may offer a window into preserved cognitive processing in participants with PDoC. Future studies will implement counterbalanced Yes/No mappings, auditory-control conditions, and formal multiplicity corrections to further substantiate and refine this approach.

A multinomial modelling framework was used to examine the relative contributions of MI-BCI DA and behavioural measures to diagnostic classification. Under leave-one-subject-out (LOSO) cross-validation, overall classification accuracy was modest (Balanced Accuracy; 55.41%), with DA and CRS-R contributing independently to differentiation from UWS. Whereas CRS-R scores showed lower sensitivity for LIS, DA achieved higher recall for LIS cases (77.8%) and provided clearer separation of LIS from UWS and MCS in predicted probabilities compared to CRS-R (see Figure 8). A post-hoc ensemble combining DA, CRS-R and WHIM predictions improved balanced accuracy to 63.21% (95% CI: 44–80%), with MCS predictive accuracy increasing from 38.46% in the multinomial model to 69.23% in the ensemble. Previous MI-BCI studies in PDoC populations have primarily used the CRS-R as the comparative behavioural measure, as this measure is considered the clinical gold standard for assessing responsiveness [100–102]. However, by including the WHIM, our findings reveal that DA correlates significantly with WHIM scores (p = 0.02) but not with CRS-R scores (p = 0.89). The WHIM contributed the least to diagnostic classification in the multinomial logistic regression, whereas both DA and CRS-R showed significant predictive power. Nonetheless, combining all three measures modestly improved model performance. Given that the CRS-R aggregates behaviour into broad categorical levels, while the WHIM provides a finer-grained, context-sensitive measure of spontaneous and goal-directed responses, these findings suggest that WHIM may more closely reflect the neural substrates underlying covert command-following detected by MI-BCI. At the same time, each measure appears to capture distinct aspects of residual function, together improving diagnostic sensitivity. Together, these results suggest that MI-BCI can complement existing clinical standards for the assessment of patients with PDoC, improving diagnostic accuracy.

Regarding analyses at the sensor and source space, analysis of spatial features derived from task-related activity over sensorimotor and parietal cortices further supports distinctions between diagnostic groups. CSP-MI weights applied for feature selection over these motor-associative areas of cortex differentiated LIS and PDoC groups, specifically distinguishing LIS from UWS during the Assessment paradigm (parietal; p = 0.009) and LIS from MCS during the Q&A paradigm (sensorimotor; p = 0.015). These findings implicate the sensorimotor-parietal network in the preserved capacity for motor imagery and volitional modulation in LIS – and more broadly confirm the role of these networks in MI-BCI control; demonstrating that BCI performance reflects physiologically meaningful activation, rather than random variance. Exploratory FC analyses suggested network-level alterations in UWS relative to MCS and LIS, most prominently involving thalamo-frontal and DMN–FPN interactions (see Supporting Information: Additional Exploratory Results, Supplementary Tables 6 and 7, Supplementary Note 6, and Supplementary Figure 7). These patterns were most evident during cognitively demanding tasks such as the Q&A paradigm – and align with known mechanisms of impaired awareness in DoC [38,41,42]. Although limited by the 16-channel montage, these exploratory findings support the idea that task-based FC can reveal context-specific neural disruptions relevant to residual awareness, complementing the main sensor-level and behavioural results.

Taken together, the convergence of decoding, spatial, and preliminary functional connectivity findings underscores the diagnostic potential of this multi-phase MI-BCI protocol as a clinically viable tool to augment existing behavioural based protocols. However, several limitations should be acknowledged. While this study represents the most comprehensive MI-BCI protocol trialled in a PDoC population to date, involving fifteen NHS hospitals in the UK and the NRH in Ireland in patient recruitment – unequal

group sizes constrained some statistical comparisons, though the analysis of all MI-BCI runs achieved 91% power [103], and large effect sizes were found for the analyses of group variance based on MI-BCI DA values. Although improvements were made to BCI parameter updating and paradigm sequencing compared to previous work [18], future performance may benefit from fully adaptive BCI systems capable of real-time adjustment to a participant's mental state [104].

EEG instability due to non-stationary brain states, linked to fatigue, arousal, medication, or learning – remains a challenge, often resulting in covariate shift [105,106]. Changes were observed in the evolution of frequency responses and feature importance, sometimes introducing features that degraded performance. DA could likely be improved through more fine-grained, participant-specific band optimisation, with adaptive classifiers [107], feature adaptation [108], or data space adaptation [109,110] also representing promising strategies to mitigate these effects. The current protocol included per-session recalibration; however, if a prior session's BCI setup outperformed the new calibration, the previous model was reused. While this maximised neurofeedback quality, it may have compromised patient learning, as classifier features could vary across sessions. Future work should develop globally optimised training data selection and automation for classifier updates.

Furthermore, although concurrent EMG was not recorded, visual inspection did not reveal sustained artefacts consistent with muscle activity. Given the severe motor impairment of participants, systematic EMG contamination is unlikely. Nevertheless, the integration of EMG and eye-tracking would benefit future research, to further rule out residual movement influences.

Additionally, while the FC analysis revealed interactions involving thalamo-frontal regions, these findings are exploratory and should be interpreted cautiously given the 16-channel EEG configuration and limited spatial resolution. Future work employing high-density EEG or beamformer-based inverse solutions (e.g., Atlantis Source Connectivity Toolbox (ASCT)) [111–113], will be required to substantiate these observations. Additionally, while imaginary coherence (iCOH) is robust to volume conduction and well-suited for non-parametric testing, it has interpretive limitations. iCOH detects non-zero phase-lagged coupling but cannot reliably infer temporal directionality – signals with vastly different delays (e.g., 10 ms vs 90 ms) can yield identical values [114]. Although faster communication is often assumed in low-frequency bands, this assumption is not always valid. Techniques like Phase Slope Index or Granger Causality could offer richer insights but are themselves limited (e.g., GC's vulnerability to volume conduction [115]).

Although anecdotal, feedback from families suggests a possible therapeutic benefit from engaging with the MI-BCI system. Some patients appeared to become aware of their influence on the system, potentially stimulating cognitive activity. This aligns with the observed shift in FC profiles between UWS and MCS patients during the Q&A task – becoming more similar to UWS-LIS differences, suggesting possible adaptive effects that merit further investigation (Supplementary Figure 7).

Regarding the the support for the integration of DA and the CRS-R to improve diagnostic prediction, these findings should be interpreted cautiously given the broad confidence intervals, and the potential instability of parameter estimates in multinomial models with small group counts. Although LOSO cross-validation mitigates within-sample bias, it remains an internal validation procedure; external replication in larger, independent cohorts will be essential to confirm the robustness and generalisability of these classification patterns.

A key limitation of the study design concerns the inclusion threshold applied during Phase I: participants had to achieve a mean DA > 70% and/or a significant DA peak (one-tailed *t*-test, $\alpha = 0.05$) during the task versus baseline. Given that participants were initially classified using behavioural

assessments – known to produce misdiagnoses, especially in UWS [10,11] – this introduces diagnostic circularity. To truly evaluate the MI-BCI's potential, future protocols should use updated clinical diagnoses, blind researchers to these diagnoses during testing, and remove assessment thresholds to avoid sampling bias. A second related important limitation which warrants acknowledgement, is that the paired-samples *t*-test used to assess whether DA during the task exceeded baseline DA at the fold level does not assume independence of CV folds. This test was retained for descriptive consistency, and comparison with alternative statistical methods indicated that our conclusions are robust to the choice of inference procedure (see Supporting Information: Methods, Supplementary Table 3).

While the exploratory analysis of the question-level accuracy attained for the illustrative case is of considerable interest in view of the potential for communication-relevant applications – the findings should be interpreted cautiously, as the analysis is based on a snapshot sample of one per group, and the confusion matrices were computed from offline results. Future work will be needed to test the reliability and practicality of MI-BCI use for communication in larger and more systematic designs.

Furthermore, practical constraints also introduced variability. While improving ecological validity, conducting sessions across hospitals, care homes, and private homes also created inconsistencies in intersession intervals and testing environments. The inherent instability of PDoC patients further complicates standardisation. Ideally, recordings would occur during each patient's considered peak arousal window, but this is often impractical due to logistical constraints. While our protocol already included multiple recording sessions per participant, collecting data across additional sessions could further increase the likelihood of assessing each patient during their optimal state of awareness. Runs identified as significant—or sessions yielding the highest number of significant runs—may indicate periods of peak arousal or cognitive responsiveness, suggesting that denser longitudinal sampling could enhance sensitivity to intra-individual fluctuations in conscious state.

In conclusion, despite some limitations, the study findings highlight the transformative potential of this multi-phase motor imagery BCI protocol in reshaping diagnostic practice for patients with severe motor and cognitive impairments. By uncovering covert cognitive capacities often missed by standard behavioural assessments, the protocol enables a more nuanced differentiation across clinical categories. Therefore, integrating DA as a metric, with CRS-R and WHIM behavioural indices, offers a more comprehensive assessment framework, improving diagnostic precision and sensitivity to residual awareness, particularly for LIS classifications. Furthermore, analysis of MI-weighted CSP values over motor-associative cortical regions indicates that motor imagery-related activation is preserved to a greater extent in LIS than in PDoC, consistent with decoding results showing that LIS participants consistently achieved the highest DA – a generally significant group effect confirming preserved volitional control despite total motor paralysis. In contrast, paradigm-specific neural signatures proved more informative for distinguishing patients with PDoC. Notably, a significant boost in DA among UWS patients during the familiar-voice Q&A paradigm suggests that emotionally salient stimuli can transiently unlock hidden awareness. Although the functional connectivity (FC) analysis was exploratory, distinct thalamo–frontal and DMN–FPN network alterations were revealed in UWS, and to a lesser extent, in MCS, relative to LIS. These patterns are consistent with established markers of impaired awareness in disorders of consciousness[38,41,42]. Together, DA, CSP-MI features, and connectivity metrics capture complementary facets of brain function, indicating that an interactive, personalised BCI approach holds promise as a scalable, movement-independent diagnostic tool – that may eventually support clinical decision-making and give voice to those otherwise unable to communicate.

Data availability

The data that support the findings of this study are available here: https://doi.org/10.15125/BATH-01632[116]. This anonymised dataset includes electroencephalography (EEG) recordings, descriptive variables, and session-level Coma Recovery Scale-Revised (CRS-R) and Wessex Head Injury Matrix (WHIM) scores. All data contained within the repository are provided under controlled access. Access will be granted upon reasonable request, which should include a brief description of the proposed research use, confirmation of relevant ethical approval where applicable, and agreement to data use conditions that prohibit data redistribution or use beyond the approved scope. Requests should be directed to the corresponding author (Prof D. Coyle, dhc30@bath.ac.uk) and will be reviewed on a case-by-case basis. A response will typically be provided within two weeks of receipt of a complete request. The source data graphed in Figures 4-9 and Supplementary Figure 7 are provided in Supplementary Data 1 (worksheets Supplementary Data A-E, H and I, respectively).

Code availability

The offline signal processing code used during the current study to analyse a single participant/run is available here: https://doi.org/10.15125/BATH-01632[116]. The code contained in the repository provides an analysis framework developed for the EEG data files also in this repository. Therefore, this code is provided under controlled access to ensure appropriate use within the approved methodological context. Access will be granted upon reasonable request, which should include a brief description of the proposed research use, confirmation of relevant ethical approval where applicable, and agreement to data use conditions that prohibit data redistribution or use beyond the approved scope. Requests should be directed to the corresponding author (Prof D. Coyle, dhc30@bath.ac.uk) and will be reviewed on a case-by-case basis. A response will typically be provided within two weeks of receipt of a complete request. Statistical analyses were performed using R-studio (version 4.4.2) [73].

References

1. Royal College of Physicians. *Prolonged disorders of consciousness following sudden onset brain injury: National clinical guidelines*. (RCP, 2020).

2. Cavanna, A. E., Shah, S., Eddy, C. M., Williams, A. & Rickards, H. Consciousness: A neurological perspective. *Behav. Neurol.* **24**, 107–116 (2011).

3. Jöhr, J., Pignat, J. M. & Diserens, K. Neurobehavioural evaluation of disorders of consciousness. *Schweizer Arch. fur Neurol. und Psychiatr.* **166**, 163–169 (2015).

4. Laureys, S. The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends Cogn. Sci.* **9**, 556–559 (2005).

5. Parliamentary Office of Science and Technology. *Disorders of consciousness*. https://researchbriefings.files.parliament.uk/documents/POST-PN-0674/POST-PN-0674.pdf (2022).

6. León-Carrión, J., Van Eeckhout, P. & Domínguez-Morales, M. D. R. Review of subject: The locked-in syndrome: A syndrome looking for a therapy. *Brain Inj.* **16**, 555–569 (2002).

7. Heilinger, A. *et al.* Performance Differences Using a Vibro-Tactile P300 BCI in LIS-Patients Diagnosed With Stroke and ALS. *Front. Neurosci.* **12**, 1–10 (2018).

8. Lugo, Z. R. *et al.* Mental imagery for brain-computer interface control and communication in non-responsive individuals. *Ann. Phys. Rehabil. Med.* **63**, 21–27 (2020).

9. Seel, R. T. *et al.* Assessment scales for disorders of consciousness: Evidence-based recommendations for clinical practice and research. *Arch. Phys. Med. Rehabil.* **91**, 1795–1813 (2010).

10. Gill-Thwaites, H. Lotteries, loopholes and luck: Misdiagnosis in the vegetative state patient. *Brain Inj.* **20**, 1321–1328 (2006).

11. Erp, V. W. S., Lavrijsen, J. C. M. & Vos, P. E. The Vegetative State: Prevalence, Misdiagnosis, and Treatment Limitations. *JMDA* **16**, e9-85.e14 (2015).

12. Wang, F. *et al.* Detecting Brain Activity Following a Verbal Command in Patients With Disorders of Consciousness. *Front. Neurosci.* **13**, 1–9 (2019).

13. Sprung, C. L. *et al.* End-of-Life Practices in European Intensive Care Units: The Ethicus Study. *Jama* **290**, 790–797 (2003).

14. Owen, A. M. *et al.* Detecting awareness in the vegetative state. *Sci. AAAS* **313**, 1402 (2006).

15. Monti, M. M. *et al.* Willful Modulation of Brain Activity in Disorders of Consciousness. *N. Engl. J. Med.* **362**, 579–589 (2010).

16. Cruse, D. *et al.* Bedside detection of awareness in the vegetative state: A cohort study. *Lancet* **378**, 2088–2094 (2011).

17. Vaughan, T. M. *et al.* The Wadsworth BCI Research and Development Program: At Home With BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.* **14**, 229–233 (2006).

18. Coyle, D., Stow, J., McCreadie, K., McElligott, J. & Carroll, Á. Sensorimotor modulation assessment and brain-computer interface training in disorders of consciousness. *Arch. Phys. Med. Rehabil.* **96**, S62–S70 (2015).

19. Gibson, R. M., Owen, A. M. & Cruse, D. Brain–computer interfaces for patients with disorders of consciousness. *Prog. Brain Res.* **228**, 241–291 (2016).

20. Xiao, J. *et al.* An Auditory BCI System for Assisting CRS-R Behavioral Assessment in Patients with Disorders of Consciousness. *Sci. Rep.* **6**, 1–13 (2016).

21. Wang, F. *et al.* Enhancing clinical communication assessments using an audiovisual BCI for patients with disorders of consciousness. *J. Neural Eng.* **14**, 1–9 (2017).

22. Royal College of Physicians. *Prolonged disorders of consciousness following sudden onset brain injury: National clinical guidelines*. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Royal+College+of+Physicians.+Prolonged+disorders+of+consciousness+following+sudden+onset+brain+injury%3A+National+clinical+guidelines&btnG= (2020).

23. Magnani, F. G., Barbadoro, F., Cacciatore, M. & Leonardi, M. The importance of instrumental assessment in disorders of consciousness: a comparison between American, European, and UK International recommendations. *Crit. Care* **26**, 1–9 (2022).

24. Giacino, J. T. *et al.* Practice guideline update recommendations summary: Disorders of consciousness. *Neurology* **91**, 450–460 (2018).

25. Pan, J. *et al.* Detecting awareness in patients with disorders of consciousness using a hybrid brain-computer interface. *J. Neural Eng.* **11**, (2014).

26. Li, Y. *et al.* Detecting number processing and mental calculation in patients with disorders of consciousness using a hybrid brain-computer interface system. *BMC Neurol.* **15**, 1–14 (2015).

27. Annen, J. *et al.* Auditory and somatosensory p3 are complementary for the assessment of patients with disorders of consciousness. *Brain Sci.* **10**, 1–14 (2020).

28. Huang, J. *et al.* Hybrid asynchronous brain-computer interface for yes/no communication in patients with disorders of consciousness. *J. Neural Eng.* **18**, (2021).

29. Li, Y. *et al.* Multimodal BCIs: Target Detection, Multidimensional Control, and Awareness Evaluation in Patients with Disorder of Consciousness. *Proc. IEEE* **104**, 332–352 (2016).

30. Han, C. H. *et al.* Electroencephalography-based endogenous brain-computer interface for online communication with a completely locked-in patient. *J. Neuroeng. Rehabil.* **16**, 1–13 (2019).

31. Formaggio, E. *et al.* EEG to Identify Attempted Movement in Unresponsive Wakefulness Syndrome. *Clin. EEG Neurosci.* **51**, 339–347 (2020).

32. Kim, N. *et al.* Cognitive-Motor Dissociation Following Pediatric Brain Injury: What About the Children? *Neurol. Clin. Pract.* **12**, 248–257 (2022).

33. Demertzi, A. *et al.* Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients. *Brain* **138**, 2619–2631 (2015).

34. Threlkeld, Z. D. *et al.* Functional networks reemerge during recovery of consciousness after acute severe traumatic brain injury. *Cortex* **106**, 299–308 (2018).

35. Carrière, M. *et al.* Auditory localization should be considered as a sign of minimally conscious state based on multimodal findings. *Brain Commun.* **2**, 1–15 (2020).

36. Shah, S. A. *et al.* Executive attention deficits after traumatic brain injury reflect impaired recruitment of resources. *NeuroImage Clin.* **14**, 233–241 (2017).

37. Li, H. *et al.* Functional networks in prolonged disorders of consciousness. *Front. Neurosci.* **17**, (2023).

38. Cacciola, A. *et al.* Functional brain network topology discriminates between patients with minimally conscious state and unresponsivewakefulness syndrome. *J. Clin. Med.* **8**, (2019).

39. Aubinet, C. *et al.* Clinical subcategorization of minimally conscious state according to resting functional connectivity. *Hum. Brain Mapp.* **39**, 4519–4532 (2018).

40. Chen, H. *et al.* Disturbed functional connectivity and topological properties of the frontal lobe in minimally conscious state based on resting-state fNIRS. *Front. Neurosci.* **17**, (2023).

41. Enciso-Olivera, C. O. *et al.* Structural and functional connectivity of the ascending arousal network for prediction of outcome in patients with acute disorders of consciousness. *Sci. Rep.* **11**, 1–12 (2021).

42. Edlow, B. L., Claassen, J., Schiff, N. D. & Greer, D. M. Recovery from disorders of consciousness: mechanisms, prognosis and emerging therapies. *Nat. Rev. Neurol.* **17**, 135–156 (2021).

43. Dayan, N. *et al.* Towards Answering Questions in Disorders of Consciousness and Locked-In Syndrome with a SMR-BCI. *Proc. 8th Graz Brain-Computer Interface Conf.* (2019) doi:10.3217/978-3-85125-682-6-65.

44. Coyle, D. *et al.* Towards electroencephalography-based consciousness assessment and cognitive function profiling in prolonged disorders of consciousness. *Res. Sq.* (2022) doi:10.21203/rs.3.rs-2349135/v1.

45. du Bois, A. N. *et al.* An evaluation of combined objective neurophysiologic markers to aid assessment of prolonged disorders of consciousness (PDoC). *medRxiv* (2024) doi:10.1101/2024.10.09.24315104.

46. Coyle, D., Stow, J., McCreadie, K., McElligott, J. & Carroll, Á. Sensorimotor modulation assessment and brain-computer interface training in disorders of consciousness. *Arch. Phys. Med. Rehabil.* **96**, S62–S70 (2015).

47. Perrin, F., Castro, M., Tillmann, B. & Luauté, J. Promoting the use of personally relevant stimuli for investigating patients with disorders of consciousness. *Front. Psychol.* **6**, 1–9 (2015).

48. Kempny, A. M. *et al.* Patients with a severe prolonged Disorder of Consciousness can show classical EEG responses to their own name compared with others' names. *NeuroImage Clin.* **19**, 311–319 (2018).

49. De Guise, E. *et al.* The montreal cognitive assessment in persons with traumatic brain injury. *Appl. Neuropsychol.* **21**, 128–135 (2014).

50. Turner-Stokes, L. F., Wade, D. & Playford, D. Prolonged Disorders of Consciousness: National Clinical Guidelines. (2013).

51. Giacino, J. T., Kalmar, K. & Whyte, J. The JFK Coma Recovery Scale-Revised: Measurement characteristics and diagnostic utility. *Arch. Phys. Med. Rehabil.* **85**, 2020–2029 (2004).

52. Bodien, Y. G., Carlowicz, C. A., Chatelle, C. & Giacino, J. T. Sensitivity and Specificity of the Coma Recovery Scale-Revised Total Score in Detection of Conscious Awareness. *Arch Phys Med Rehabil* **97**, 490–492 (2016).

53. Shiel, A. *et al.* The wessex head injury matrix (WHIM) main scale: A preliminary report on a

scale to assess and monitor patient recovery after severe head injury. *Clin. Rehabil.* **14**, 408–416 (2000).

54. Blume, C., del Giudice, R., Wislowska, M., Lechinger, J. & Schabus, M. Across the consciousness continuum—from unresponsive wakefulness to sleep. *Front. Hum. Neurosci.* **9**, 1–14 (2015).

55. g.tec medical engineering. g.NAUTILUS RESEARCH | Wearable EEG Headset. https://www.gtec.at/product/gnautilus-pro/ (2020).

56. Simulink for Matlab (The MathWorks, Inc.). (2020).

57. Unity Technologies. Unity Real-Time Development Platform | 3D, 2D VR & AR Visualizations. *Unity Technologies* https://unity.com/ (2020).

58. Ang, K. K., Chin, Z. Y., Zhang, H. & Guan, C. Filter Bank Common Spatial Pattern (FBCSP) in brain-computer interface. *Proc. Int. Jt. Conf. Neural Networks* 2390–2397 (2008) doi:10.1109/IJCNN.2008.4634130.

59. Pohjalainen, J., Räsänen, O. & Kadioglu, S. Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits. *Comput. Speech Lang.* **29**, 145–171 (2015).

60. Korik, A. *et al.* Competing at the Cybathlon championship for people with disabilities: long-term motor imagery brain–computer interface training of a cybathlete who has tetraplegia. *J. Neuroeng. Rehabil.* (2022) doi:10.1186/s12984-022-01073-9.

61. Ang, K. K., Chin, Z. Y., Wang, C., Guan, C. & Zhang, H. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Front. Neurosci.* **6**, 1–9 (2012).

62. Korik, A., Sosnik, R., Siddique, N. & Coyle, D. Decoding Imagined 3D Arm Movement Trajectories from EEG to Control Two Virtual Arms - A Pilot Study. *Front. Neurorobot.* **13**, 1–22 (2019).

63. Lotte, F. & Guan, C. Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Trans. Biomed. Eng.* **58**, 355–362 (2011).

64. Lotte, F. & Guan, C. Regulized Common Spatial Patterns (RCSP) toolbox. (2010).

65. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F. & Arnaldi, B. A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.* **4**, (2007).

66. Coyle, D., Dayan, N., Stow, J., McElligott, J. & Carroll, A. Answering questions in Prolonged disorders of consciousness with a brain-computer interface. in *7th International BCI Meeting* 1 (2018).

67. Mueller-Putz, G., Scherer, R., Brunner, C., Leeb, R. & Pfurtscheller, G. Better than random: A closer look on BCI results. *Int. J. Bioelectromagn.* **10**, 52–55 (2008).

68. Ojala, M. & Garriga, G. C. Permutation tests for studying classifier performance. *J. Mach. Learn. Res.* **11**, 1833–1863 (2010).

69. Matlab (The MathWorks, Inc.). (2020).

70. Pascual-Marqui, R. D. sLORETA, low resolution brain electromagnetic tomography. (2018).

71. Pfurtscheller, G. *et al.* Current trends in Graz Brain-Computer Interface (BCI) research. *IEEE Trans. Rehabil. Eng.* **8**, 216–219 (2000).

72. Ramlackhansingh, A. F. *et al.* Inflammation after trauma: Microglial activation and traumatic brain injury. *Ann. Neurol.* **70**, 374–383 (2011).

73. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. (2024).

74. Haufe, S. *et al.* On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* **87**, 96–110 (2014).

75. Karrasch, M., Laine, M., Rapinoja, P. & Krause, C. M. Effects of normal aging on event-related desynchronization/synchronization during a memory task in humans. *Neurosci. Lett.* **366**, 18–23 (2004).

76. Stawarczyk, D., Majerus, S., Maquet, P. & D'Argembeau, A. Neural correlates of ongoing conscious experience: Both task-unrelatedness and stimulus-independence are related to default network activity. *PLoS One* **6**, (2011).

77. Wu, W. *et al.* Is frontoparietal electroencephalogram activity related to the level of functional disability in patients emerging from a minimally conscious state? A preliminary study. *Front. Hum. Neurosci.* **16**, 1–10 (2022).

78. Qu, S. *et al.* Analyzing brain-activation responses to auditory stimuli improves the diagnosis of a disorder of consciousness by non-linear dynamic analysis of the EEG. *Sci. Rep.* **14**, 1–13 (2024).

79. Zheng, Z. S., Reggente, N., Lutkenhoff, E., Owen, A. M. & Monti, M. M. Disentangling disorders of consciousness: Insights from diffusion tensor imaging and machine learning. *Hum. Brain Mapp.* **38**, 431–443 (2017).

80. Di Perri, C. *et al.* Neural correlates of consciousness in patients who have emerged from a minimally conscious state: A cross-sectional multimodal imaging study. *Lancet Neurol.* **15**, 830–842 (2016).

81. Haufe, S. & Ewald, A. A Simulation Framework for Benchmarking EEG-Based Brain Connectivity Estimation Methodologies. *Brain Topogr.* **32**, 625–642 (2019).

82. Montreal Neurological Institute. The McConnell Brain Imagining Centre. https://www.mcgill.ca/bic/software/tools-data-analysis/anatomical-mri/atlases.

83. Pascual-Marqui, R. D. Lehmann, D. *et al.* Assessing interactions in the brain with exact low resolution electromagnetic tomography. *Zurich Open Repository and Archive* vol. 369 1–22 (2011).

84. Dattola, S. *et al.* Effect of Sensor Density on eLORETA Source Localization Accuracy. in *Smart Innovation, Systems and Technologies - Neural Approaches to Dynamics of Signal Exchanges,* (eds. Esposito, A., Faundez-Zanuy, M., Morabito, F. C. & Pasero, E.) vol. 151 403–414 (Springer, 2020).

85. Akalin Acar, Z. & Makeig, S. Effects of forward model errors on EEG source localization. *Brain Topogr.* **26**, 378–396 (2013).

86. Brodbeck, V. *et al.* Electroencephalographic source imaging: A prospective study of 152 operated epileptic patients. *Brain* **134**, 2887–2897 (2011).

87. Michel, C. M. & Brunet, D. EEG source imaging: A practical review of the analysis steps. *Front. Neurol.* **10**, (2019).

88. Pascual-Marqui, R. D. Discrete, 3D distributed, linear imaging methods of electric neuronal

activity. Part 1: exact, zero error localization. *arXiv:0710.3341 [math-ph]* http://arxiv.org/abs/0710.3341 (2007).

89. Sanchez-Bornot, J. M. *et al.* High-dimensional brain-wide functional connectivity mapping in magnetoencephalography. *J. Neurosci. Methods* **348**, (2021).

90. Li, R. *et al.* The Perils and Pitfalls of Block Design for EEG Classification Experiments. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 316–333 (2021).

91. Mueller-Putz, G., Scherer, R., Brunner, C., Leeb, R. & Pfurtscheller, G. Better than random: A closer look on BCI results. *Int. J. Bioelectromagn.* **10**, 52–55 (2008).

92. Ahn, M., Cho, H., Ahn, S. & Jun, S. C. High theta and low alpha powers may be indicative of BCI-illiteracy in motor imagery. *PLoS One* **8**, (2013).

93. Ahn, M. & Jun, S. C. Performance variation in motor imagery brain-computer interface: A brief review. *J. Neurosci. Methods* **243**, 103–110 (2015).

94. Blankertz, B. *et al.* The Berlin Brain-Computer Interface: Non-Medical Uses of BCI Technology. *Front. Neurosci.* **4**, 198 (2010).

95. Combrisson, E. & Jerbi, K. Exceeding chance level by chance : The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods* **250**, 126–136 (2015).

96. Di, H., Boly, M., Weng, X., Ledoux, D. & Laureys, S. Neuroimaging activation studies in the vegetative state: Predictors of recovery? *Clin. Med. J. R. Coll. Physicians London* **8**, 502–507 (2008).

97. Giacino, J. T., Fins, J. J., Laureys, S. & Schiff, N. D. Disorders of consciousness after acquired brain injury: The state of the science. *Nat. Rev. Neurol.* **10**, 99–114 (2014).

98. Schiff, N. D. *et al.* Brain–Computer Interfaces for Communication in Patients with Disorders of Consciousness: A Gap Analysis and Scientific Roadmap. *Neurocrit. Care* **41**, 129–145 (2024).

99. Laureys, S., Owen, A. M. & Schiff, N. D. Brain function in coma, vegetative state, and related disorders. *Lancet Neurol.* **3**, 537–546 (2004).

100. Chennu, S. *et al.* Brain networks predict metabolism, diagnosis and prognosis at the bedside in disorders of consciousness. *Brain* **140**, 2120–2132 (2017).

101. Wang, Q., Huang, Y., Meng, X., Feng, Z. & Bai, Y. Neuroimaging and electrophysiology techniques unveiling the mystery of disorders of consciousness : a narrative review. 86–104 (2024) doi:10.4103/ATN.ATN-D-24-00006.

102. Ballanti, S. *et al.* EEG-based methods for recovery prognosis of patients with disorders of consciousness: A systematic review. *Clin. Neurophysiol.* **144**, 98–114 (2022).

103. Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).

104. Myrden, A. & Chau, T. Effects of user mental state on EEG-BCI performance. *Front. Hum. Neurosci.* **9**, 1–11 (2015).

105. Sugiyama, M., Krauledat, M. & Muller, K.-R. Covariate Shift Adaptation by Importance Weighted Cross Validation. *J. Mach. Learn. Res.* **8**, 985–1005 (2007).

106. Mohammadi, R., Mahloojifar, A. & Coyle, D. Unsupervised Short-term Covariate Shift

Minimization for Self-paced BCI. in *IEEE Sympsosium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain* 101–106 (2013).

107. Vidaurre, C., Schlöogl, A., Cabeza, R., Scherer, R. & Pfurtscheller, G. A Fully On-Line Adaptive BCI. *IEEE Trans Biomed Eng.* **53**, 1214–9 (2006).

108. Satti, A., Guan, C., Coyle, D. & Prasad, G. A Covariate Shift Minimisation Method to Alleviate Non-stationarity Effects for an Adaptive Brain-Computer Interface. in *2010 20th International Conference on Pattern Recognition* 105–108 (Ieee, 2010). doi:10.1109/ICPR.2010.34.

109. Arvaneh, M., Guan, C., Ang, K. K. & Quek, C. EEG Data Space Adaptation to Reduce Inter-session Non-stationarity in Brain- Computer Interface. *Neural Comput.* **25**, 2146–71 (2013).

110. Samek, W., Member, S. & Meinecke, F. C. Transferring Subspaces Between Subjects in Brain-Computer Interfacing. *Biomed. Eng. (NY).* **60**, 1–10 (2013).

111. ASCT: Atlantis Source Connectivity Toolbox. https://atlantis.psychologia.uj.edu.pl (2025).

112. Spadone, S., Wyczesany, M., Della Penna, S., Corbetta, M. & Capotosto, P. Directed Flow of Beta Band Communication during Reorienting of Attention within the Dorsal Attention Network. *Brain Connect.* **11**, 717–724 (2021).

113. Adamczyk, A. K. & Wyczesany, M. Theta-band Connectivity within Cognitive Control Brain Networks Suggests Common Neural Mechanisms for Cognitive and Implicit Emotional Control. *J. Cogn. Neurosci.* **35**, 1656–1669 (2023).

114. Nolte, G. *et al.* Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin. Neurophysiol.* **115**, 2292–2307 (2004).

115. Bastos, A. M. & Schoffelen, J. M. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.* **9**, 1–23 (2016).

116. Coyle, D., du Bois, N. & Korik, A. *Motor-imagery brain-computer interface electroencephalography and behavioural assessment datasets in prolonged disorders of consciousness*. https://doi.org/10.15125/BATH-01632%0A doi:10.15125/BATH-01632.

Author contributions

The authors confirm their contribution to the paper as follows: D. C., K. M., A. M., K. Y., J. S., J. M., and A. C., study conception and design; A. M., K. Y., J. M., A. C., R. C., L. B., S. N., S. J., D. H., E. V., V. H., S. P., P. J., J. P., E. D., J. S., and A. S., participant recruitment; D. C., S. H., L. H., N. D., A. S. E., and N. B., data collection; D. C., A. K., A. D. B., and K. M., development of the Brain-Computer Interface software and feedback systems; J. M. S., development of the functional connectivity (FC) pipeline; D.C. and A.K., development of the motor-imagery task related EEG data analysis framework and analysis to extract decoding accuracies (DAs), plus FC analysis; D.C and N. B., statistical analysis to determine group differences. D. C., A. K., and N. B., interpretation of results; D. C., A. K., J. M. S., and N. B., draft manuscript preparation. A. K., preparation of figures 3, 4, 9, and Supplementary Figures 1-4; N. B., preparation of remaining figures and tables; All authors reviewed the results and approved the final version of the manuscript.

Acknowledgements

We would like to thank the clinical teams at the participating UK NHS hospitals and the National Rehabilitation Hospital (NRH) of Ireland for their dedication to patient recruitment. We are also grateful to the care teams and staff at the hospitals and care homes where participants resided, whose support was essential to the success of the trials. To the participants and their families, we extend our deepest thanks for their generous participation, trust, and commitment to this research. This work was supported by internal funds from the University of Bath and Ulster University; by access to the Tier 2 High Performance Computing resources provided by the Northern Ireland High Performance Computing (NI-HPC) facility, funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant number EP/T022175; and the UK Research and Innovation (UKRI) Turing AI Fellowship 2021-2025 funded by the EPSRC under Grant number EP/V025724/1. Author K. P. S. N. was supported by the NIHR Sheffield Biomedical Research Centre/NIHR Sheffield Clinical Research Facility. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the NRH or the Department of Health and Social Care.

Competing Interests

Prof Damien Coyle, Chief Investigator on this project, is the founder, Chief Executive Officer, and a shareholder of NeuroCONCISE Ltd., a company involved in the development of neurotechnology and wearable electroencephalography (EEG) systems.

Figure Captions and Main Tables

Figure Captions

Figure 1. Timing of the trials. a: Assessment. b: Training, and Feedback. c: Q&A. Note: the Q&A task began with a question. Therefore, the trigger for the start of the task performance period was sent at the end of the question. Paradigm specific trial instructions are provided in Supporting Information, Supplementary Note 1.

Figure 2. EEG montage. The positions of EEG (blue and green) and ground (orange) electrodes are presented in this figure. Nine EEG channels covering motor and imagined movement-related cortical areas are indicated in green. The reference clip was attached to the right ear (as indicated by the circle in grey).

Figure 3. Structure of the applied FBCSP-MI based 2-class classification method. The block diagram illustrates the signal processing framework, including filter bank common spatial patterns and mutual information (FBCSP-MI) based features selection modules, and a regularized linear discriminant analysis (RLDA) 2-class classifier. The analysis performed using three different sets of frequency bands, involving four adjacent bands from the six pre-processed bands (i.e., delta, theta, mu, low-beta, high-beta, low-gamma bands). The illustrated setup represents the highest four-band option.

Figure 4. Decoding accuracy (DA) across runs for all participants. Graphs are displayed according to diagnostic category (top row – UWS; middle row – MCS; bottom row – LIS), shown separately for all runs (left-hand column) and runs reaching statistical significance (middle column), with representative high-performing individuals highlighted (right-hand column). Left and middle columns show group-level trajectories for Ref DA (black) and Task DA (red), including participant traces (see subject ID for each group), mean trajectories (dashed), and participant-averaged linear trends (solid). Right column shows single-participant examples illustrating stable high performance and task-reference separation.

Figure 5. Illustration of the comparison of DA values obtained for each group (UWS, MCS, LIS, AB). DA values obtained from (a) all runs (n = 1233 run/33 participants; UWS = 293/8, MCS = 472/13, LIS = 433/10, AB = 35/2) and (b) runs showing a significant task-related peak (one-tailed t-test comparing task vs. reference interval, $\alpha = 0.05$; n = 681 run/31 participants; UWS = 138/8, MCS = 211/13, LIS = 299/10, AB = 33/2). Boxplots display the data distribution for each group, showing the interquartile range (25th–75th percentile) with the median as a horizontal line. Individual data points represent DA values from multiple runs per participant, illustrating within-group variability. Red circles mark group mean DA, with vertical error bars showing ±1 SD. Black circles and solid lines connect mean task-related DA across groups, while grey squares and dashed lines indicate mean DA during the reference interval. Panel (c) illustrates the proportion of significant runs per paradigm (assessment, training, feedback, and Q&A), by group. Bars represent the mean percentage of significant runs for each group and paradigm. Individual data points represent proportions from each participant's runs, illustrating within-group variability. Red points and connecting lines indicate group-level averages across paradigms. Statistical significance of diagnostic group differences is indicated by asterisks:*$p < 0.05$, **$p < 0.01$.

Figure 6. Comparison of decoding accuracy (DA) values obtained for Q&A categories, across groups (UWS, MCS, LIS, AB). DA values obtained from (a) all runs (n = 297 run/30 participants; UWS = 57/6, MCS = 112/12, LIS = 124/10, AB = 4/2) and (b) runs showing a significant task-related peak (one-

tailed t-test comparing task vs. reference interval, $\alpha = 0.05$; n = 169 run/29 participants; UWS = 26/5, MCS = 51/12, LIS = 88/10, AB = 4/2). Boxplots display the data distribution for each group, showing the interquartile range (25th–75th percentile) with the median as a horizontal line. Individual data points represent DA values from multiple runs per participant, illustrating within-group variability. Red circles mark group mean DA, with vertical error bars showing ±1 SD. Black circles and solid lines connect mean Q&A task-related DA across groups, while grey squares and dashed lines indicate mean DA during the reference interval. Panel (c) illustrates the proportion of significant runs per Q&A category (Biographical, Basic Logic (Logic), Numbers and Letters (Numeric), and Situational), by group. Bars represent the mean percentage of significant runs for each group and question category. Individual data points represent proportions from each participant's runs, illustrating within-group variability. Red points and connecting lines indicate group-level averages across paradigms. Statistical significance of diagnostic group differences is indicated by asterisks:*$p < 0.05$, **$p < 0.01$.

Figure 7. Mean DA across task paradigms (Training, Feedback, and Q&A) for significant runs only, shown separately for diagnostic groups (UWS, MCS, LIS). Thin lines represent individual participants, illustrating within-group variability. Bold lines and shaded bands denote group-level means ±1 SD. The dashed line indicates empirical chance levels for each paradigm, computed using the "better-than-random" binomial rule 91. Group colours: UWS = green, MCS = orange, LIS = blue.

Figure 8. Predicted probabilities of clinical diagnosis (UWS, MCS, LIS) derived from a multinomial logistic regression model as a function of (a) average decoding accuracy (DA) and (b) CRS-R score. The model was fitted using participant-level averages within each diagnostic group. Lines show model-estimated probabilities with 95% bootstrapped confidence intervals. Colours denote diagnostic groups (UWS = blue, MCS = red, LIS = green).

Figure 9. Clustered boxplots illustrating CSP–MI weights across brain regions for each diagnostic group (UWS, MCS, LIS, AB) and paradigm: Assessment (top left), Training (top right), Feedback (bottom left), and Q&A (bottom right). Each box represents the distribution of CSP-MI weights across individual runs (multiple runs per participant) within each group and brain region, with boxes showing the interquartile range (25th-75th percentile) and horizontal lines marking the median. Whiskers extend to 1.5× the interquartile range, and outliers are not displayed. Colours denote diagnostic groups as indicated in the legend. Statistical significance of diagnostic group differences is indicated by asterisks:*$p < 0.05$, **$p < 0.01$.

Main Tables

Table 1. Structure of the study phases and associated runs.

Editorial Summary: Du Bois et al. use a structured, multiphase brain–computer interface to test how people with severe brain injuries can modulate brain activity through imagined movements. The results show that decoding these signals provides objective evidence of awareness and diagnostic information beyond traditional behavioural assessments.

Peer review information: *Communications Medicine* thanks Masaya Ueda, Seitaro Iwama and Reinhold Scherer for their contribution to the peer review of this work.