Multimodal multicentre investigation of diagnostic and prognostic markers in disorders of consciousness

Dragana Manasova, Laouen Mayal Louan Belloli, Martin Justinus Rosenfelder, Lina Willacker, Emilia Fló Rama, Chiara Valota, Bertrand Hermann, Brigitte Charlotte Kaufmann, Alice Pirastru, Chiara Camilla Derchi, Theresa Raiser, Melanie Valente, Aude Sangare, Başak Türker, Nadya Pyatigorskaya, Benoît Béranger, Michele Colombo, Esteban Munoz-Musat, Anira Escrichs, Tiziana Atzori, Francesca Baglio, Constantin Lapa, Ansgar Berlis, Kristina Krüger, Tina Luther, Vincent Perlbarg, Gustavo Deco, Yonathan Sanz-Perl, Enzo Tagliazucchi, Louis Puybasset, Benjamin Rohaut, Lionel Naccache, Angela Comanducci, Anat Arzi, Mario Rosanova, Andreas Bender, Jacobo Diego Sitt

Angaben zur Veröffentlichung / Publication details:

Manasova, Dragana, Laouen Mayal Louan Belloli, Martin Justinus Rosenfelder, Lina Willacker, Emilia Fló Rama, Chiara Valota, Bertrand Hermann, et al. 2026. "Multimodal multicentre investigation of diagnostic and prognostic markers in disorders of consciousness." Brain, awaf412. https://doi.org/10.1093/brain/awaf412.

BRAIN
ORIGINAL ARTICLE

Multimodal multicentre investigation of diagnostic and prognostic markers in disorders of consciousness

Dragana Manasova,[1,2] Laouen Mayal Louan Belloli,[1,3,4] Martin Justinus Rosenfelder,[5,6,7] Lina Willacker,[5] Emilia Fló Rama,[1] Chiara Valota,[8,9] Bertrand Hermann,[10,11] Brigitte Charlotte Kaufmann,[1] Alice Pirastru,[9] Chiara Camilla Derchi,[9] Theresa Raiser,[5] Melanie Valente,[1] Aude Sangare,[1] Başak Türker,[1] Nadya Pyatigorskaya,[1] Benoît Béranger,[1] Michele Colombo,[8] Esteban Munoz-Musat,[1,12] Anira Escrichs,[13] Tiziana Atzori,[9] Francesca Baglio,[9] Constantin Lapa,[14] Ansgar Berlis,[15] Kristina Krüger,[15] Tina Luther,[5,6] Vincent Perlbarg,[16] Gustavo Deco,[13,17] Yonathan Sanz-Perl,[1,17] Enzo Tagliazucchi,[4,18] Louis Puybasset,[16,19] Benjamin Rohaut,[1,20] Lionel Naccache,[1,21] Angela Comanducci,[9] Anat Arzi,[1,22,23] Mario Rosanova,[8] Andreas Bender[5,6,24] and Jacobo Diego Sitt[1]

Severely brain-injured patients may enter a spectrum of conditions collectively known as disorders of consciousness. This spectrum includes clinical conditions such as unresponsive wakefulness syndrome or minimally conscious state, where the behavioural assessment of consciousness can often be deceptive.

To bridge this dissociation, neuroimaging techniques are employed to identify the residual brain functions. Each neuroimaging modality imperfectly captures distinct aspects of brain preservation—functional, anatomical, or both. In this study, we adopt a comprehensive approach by integrating the neurophysiology and neuroimaging modalities available from the standard and advanced clinical assessments through interpretable machine learning. The electrophysiological modalities included high-density EEG (resting state and task), whereas neuroimaging modalities included anatomical and resting-state functional MRI, diffusion MRI and ^{18}F-fluorodeoxyglucose PET.

Our investigation reveals that specific modalities, such as functional assessments, provide comprehensive insights into the currently evaluated state of consciousness, the diagnosis of the patients. Conversely, structural modalities offer valuable information about the patient's evolution within the consciousness spectrum. We validate the proposed analysis with data coming from other centres with different acquisition parameters. Importantly, we demonstrate that model performance improves with an increase in the number of modalities. We observe a higher inter-modality disagreement for minimally conscious state patients and those patients who improve. Lastly, we observe a difference in feature importances between diagnosis and prognosis, with an interaction between modality and anatomical structures: some subcortical markers tend to contribute more to prognosis, while other cortical markers are more informative for diagnosis.

This integrative multimodal and machine learning methodology presents a promising avenue for a more nuanced understanding of disorders of consciousness, contributing to enhanced diagnostic precision, prognostic capabilities and the personalization of rehabilitative strategies in clinical practice.

Received November 21, 2024. Revised July 22, 2025. Accepted September 03, 2025. Advance access publication January 7, 2026

1 Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, Sorbonne Université, Paris 75013, France
2 Université Paris Cité, Paris 75006, France
3 Laboratorio de Inteligencia Artificial Aplicada, Instituto de Ciencias de la Computación, Universidad de Buenos Aires, Buenos Aires C1053, Argentina
4 Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ministry of Science, Technology and Innovation, Buenos Aires C1053, Argentina
5 Department of Neurology, University Hospital of the Ludwig-Maximilians-Universität München, Munich 82152, Germany
6 Therapiezentrum Burgau, Hospital for Neurological Rehabilitation, Burgau 89331, Germany
7 Clinical and Biological Psychology, Institute of Psychology and Education, Ulm University, Ulm 89081, Germany
8 Department of Biomedical and Clinical Sciences, University of Milano, Milan 20157, Italy
9 IRCCS Fondazione Don Carlo Gnocchi ONLUS, Milan 20148, Italy
10 Inserm 1266, Institute of Psychiatry and Neurosciences of Paris, Université Paris Cité, Paris F-75014, France
11 Medical Intensive Care Unit, HEGP Hôpital, Assistance Publique - Hôpitaux de Paris-Centre (APHP-Centre), Paris 75014, France
12 Centre Mémoire de Ressources et de Recherche, Paris Nord/Université Paris-Cité, Paris 75006, France
13 Center for Brain and Cognition, Computational Neuroscience Group, Universitat Pompeu Fabra, Barcelona 08005, Spain
14 Nuclear Medicine, Faculty of Medicine, University of Augsburg, Augsburg 86156, Germany
15 Diagnostic and Interventional Neuroradiology, Faculty of Medicine, University of Augsburg, Augsburg 86156, Germany
16 BRAINTALE SAS, Paris 75013, France
17 Institució Catalana de la Recerca I Estudis Avançats (ICREA), Barcelona 08010, Spain
18 Latin American Brain Health Institute (BrainLat), Universidad Adolfo Ibáñez, Santiago 7941169, Chile
19 GRC 29, AP-HP, DMU DREAM, Department of Anaesthesiology and Critical Care Medicine, Pitié-Salpêtrière Hospital, Sorbonne University, Paris 75013, France
20 AP-HP, Hôpital de la Pitié Salpêtrière, Neuro ICU, DMU Neurosciences, Paris 75013, France
21 AP-HP, Hôpital Pitié - Salpêtrière, Service de Neurophysiologie Clinique, Paris 75013, France
22 Department of Medical Neurobiology, Institute for Medical Research Israel-Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem 9112102, Israel
23 Department of Cognitive and Brain Sciences, The Hebrew University of Jerusalem, Jerusalem 9112102, Israel
24 Department of Neurorehabilitation, Medical Faculty, University of Augsburg, Augsburg 86156, Germany

Correspondence to: Dragana Manasova
Paris Brain Institute - Institut du Cerveau
Hôpital Pitié, 47 Bd de l'Hôpital, Paris 75013, France
E-mail: dragana.manasova@gmail.com

Correspondence may also be addressed to: Jacobo Diego Sitt
E-mail: jacobo.sitt@inserm.fr

Keywords: disorders of consciousness; electrophysiology; neuroimaging; multimodal; machine learning

Introduction

Disorders of consciousness (DoC) encompass a spectrum of conditions resulting from various causes of brain injury. Patients with a DoC have a range of sensorimotor deficits that can, to varying extents, impair both their state of consciousness and their capacity to express it.[1] Owing to an unreliable ability to express consciousness through behavioural responses, there can be a dissociation between unresponsiveness (based on following commands with motor outputs) and unconsciousness.[2,3] Furthermore, the assessment of patients with DoC can be limited due to the presence of medical devices (such as mechanical ventilation or tracheostomy tubes), acute pain, or medications that affect arousal.[4] All of these aspects pose challenges to the correct assessment of the patient's consciousness.

The main clinical conditions on the DoC spectrum are unresponsive wakefulness syndrome (UWS) and the minimally conscious state (MCS). UWS patients are behaviourally diagnosed by eyes opening during arousal but with no signs of awareness,[5] whereas patients in an MCS show reproducible, though subtle, behavioural signs of consciousness (visual pursuit or a response to simple commands).[6] Although there is no clear consensus where the spectrum of DoC ends and healthy consciousness begins,[7] generally patients are said to be emergent from MCS (EMCS) when they regain some basic communication capacity or when they are capable of functional object use.[6] This recovery of consciousness can occur at any point in the patient's clinical evolution—from the acute to chronic stages.[7]

The current diagnostic gold standard in the field is the Coma Recovery Scale-Revised (CRS-R).[8] Although the systematic and repeated use of this scale decreases the rate of misdiagnosis,[8] the remaining uncertainty due to behavioural and neural disparities has yet to be addressed systematically. Thus, despite extensive standardization of the administration of behavioural scales,[9] current guidelines recommend the use of neuronal or physiological signals

across distinct modalities to increase the certainty of achieving a correct assessment of a patient's state.[1,4,10-13] Additionally, individual electrophysiology or neuroimaging modalities are limited because they occupy a narrow space in the temporal-spatial resolution plane and are designed to evaluate only specific neural structures or activities. A natural question that arises is to what extent each modality is informative in terms of diagnosis and prognosis.

Throughout the years, various studies have focused on single modalities and reported the potential of each in improving diagnostic power.[2,14-32] However, clinical teams, especially those in hospitals with established expertise in treating DoC patients, have access to and can combine information from various multimodal tests.[1,11,33] Recent evidence indicates that multimodal assessment enhances neuroprognostics in clinically unresponsive critical-care patients with brain injury.[33] This suggests that latent integration of information by clinicians contributes to improved decision-making outcomes, thereby making a strong argument for the exploration of automatic fusion approaches using machine learning. This is also emphasized by the international guidelines for the clinical approach to DoC, calling for multimodal assessments, especially due to their heterogeneous pathophysiology.[10,13,34] On the question of how we can systematically make use of the complementary information contained in the different neurophysiological signals, numerous studies have assessed the possibilities of an integrative neuroimaging approach.[1,11,35-42] These studies highlight the need to investigate various dimensions of brain preservation [e.g. anatomical MRI (aMRI), functional MRI (fMRI), electrophysiology or brain metabolism] to more accurately assess a patient's current state and progression. However, to date, there has been no large-scale multicentre study involving commonly used neuroimaging and electrophysiological modalities analysed under the same methodological umbrella to evaluate the differences and complementarity of these modalities in assessing a patient's current condition and evolution.

In this study, we took the multimodal integrative neuroimaging and electrophysiological approach one step further and adopted a multicentric, comprehensive approach by separately analysing and then combining six neuroimaging modalities through interpretable machine learning, to investigate the evaluated consciousness (diagnosis) of patients and future change (prognosis).

Materials and methods

This project is part of an EU-funded 4-year consortium (PerBrain) involving several institutions, including the Pitié Salpetrière Hospital, the University Hospital of the Ludwig-Maximilians-University of Munich, the Therapiezentrum Burgau (hospital for neurological rehabilitation), the University of Milan, Fondazione Don Carlo Gnocchi, and the Weizmann Institute of Science. The consortium's aims are explained in Willacker *et al.*[43] The included patients are from three centres and split into Dataset 1 (France), Dataset 2 (Germany) and Dataset 3 (Italy).

Ethics statements

This research was approved by the ethical committee of the Pitie-Salpetriere under the French label of 'routine care research' (Comité de Protection des Personnes no 2013-A01385-40, Ile de France 1, Paris, France under the code 'Recherche en soins courants', protocol numbers NEURO-DoC/HAO-006/20130409 and M-NEURO-DoC/NCT04534777); the ethics committee of the medical faculty of Ludwig-Maximilians-Universität München (protocol numbers 20-634 and 20-635); and the ethical committee section of the IRCCS Fondazione Don Carlo Gnocchi (ethics committee IRCCS Regione Lombardia, protocol number 32/2021/CE_FdG/FC/SA). Written informed consent from patients was obtained either through their legal guardian or, in the absence of one, from the closest relative. Table 1 provides an overview of the number of patients from all three datasets who have a particular modality per prediction category (diagnostic or prognostic). The patient inclusion and behavioural assessment (including the CRS-R[8] and the Glasgow Outcome Scale-Extended[44]) and methodological details can be found in the Supplementary material.

Modalities

In this study, the included modalities were high-density EEG-resting state (RS) and a two-level local-global (LG) auditory regularity task, aMRI, resting-state functional MRI (fMRI-RS), diffusion MRI (dMRI) and ^{18}F-fluorodeoxyglucose PET (FDG-PET) (Table 2). The acquisition protocols and parameters, as well as the preprocessing details and analyses of the markers, are provided in the Supplementary material. The markers we extracted from the modalities later used in the machine learning models[45] are given in Table 2.

The main methodological outline of the paper is given in Fig. 1. We included two prediction targets: (i) diagnosis; and (ii) prognosis of the patients. We ran models per modality (referred to as unimodal) and then combined the predictions, yielding a multimodal analysis. The acquisition details, preprocessing, marker extraction, machine learning methods and statistical analysis are described in detail in the Supplementary material.

Results

Neuroimaging modalities carry independent diagnostic and prognostic information

In the diagnostic classification (Fig. 2A), the highest balanced accuracy was observed for PET (0.73 ± 0.13), followed by dMRI (0.69 ± 0.08), EEG-LG (0.69 ± 0.06), EEG-RS (0.66 ± 0.09), fMRI-RS (0.63 ± 0.15), and the lowest for aMRI (0.51 ± 0.1). For the prognostic classification (Fig. 2B), the highest balanced accuracy was achieved by dMRI (0.74 ± 0.09), followed by fMRI-RS (0.63 ± 0.18), aMRI (0.58 ± 0.11), EEG-LG (0.55 ± 0.06), PET (0.5 ± 0.16), and the lowest for EEG-RS (0.49 ± 0.1). For both diagnostic and prognostic classification, we calculated the quality of the difference between the model distributions and the surrogate distributions (Supplementary Fig. 2).

EEG recordings, whether RS or LG, exhibited a classification accuracy close to 0.7 for diagnosis but dropped near the chance level for prognosis. Conversely, aMRI showed an increase from chance level for diagnosis to 0.58 for prognosis, while PET exhibited the opposite trend, becoming non-informative in more of the cross-validated splits for prognosis (in other words, one part of the distribution of the PET prognostic results was at chance level). dMRI and fMRI-RS remained relevant for both diagnosis and prognosis, with dMRI gaining 0.05 points for prognosis.

Using two independent datasets (Dataset 2 from Germany and Dataset 3 from Italy), we examined the generalization of diagnostic prediction in four modalities. The balanced accuracy aligned with the training set for fMRI-RS (0.7 ± 0.12) and EEG-LG (0.62 ± 0.03) in Dataset 2 and for EEG-LG (0.75 ± 0.05), EEG-RS (0.64 ± 0.04) and aMRI (0.58 ± 0.12) in Dataset 3 (Italy). The aMRI models for diagnosis for Dataset 1 were at chance level, which implied that we could not

Table 1 Overview of the number of patients from all three datasets having a particular neuroimaging modality per prediction category (diagnostic or prognostic)

Modality	All patients	Diagnostic categories (Current state of the patients)		Prognostic categories (Evolution of the patients)	
		UWS	MCS	Not-improved	Improved
Dataset 1 (France)					
EEG-RS	120	63	57	50	39
EEG-LG	290	138	152	117	90
fMRI-RS	44	21	23	20	10
FDG-PET	53	21	32	35	9
dMRI	151	79	72	49	46
aMRI	101	54	47	40	26
Dataset 2 (Germany)					
EEG-RS	50	30	20	–	–
EEG-LG	42	24	18	–	–
aMRI	12	10	2	–	–
fMRI-RS	7	5	2	–	–
Dataset 3 (Italy)					
EEG-RS	25	12	13	–	–
EEG-LG	17	6	11	–	–
aMRI	12	3	9	–	–
fMRI-RS	12	3	9		–

The table presents data obtained through various neuroimaging modalities. In Dataset 1 (France), there were a total of 326 patients for diagnosis, and 232 patients with prognostic data; in Dataset 2 (Germany), there were 54 patients; and in Dataset 3 (Italy), 30 patients. Modalities included EEG-resting state (EEG-RS), EEG Local-Global paradigm (EEG-LG), resting-state functional MRI (fMRI-RS), anatomical MRI (aMRI), diffusion MRI (dMRI) and ^{18}F-fluorodeoxyglucose PET (FDG-PET). The patient population was further categorized based on diagnostic information, distinguishing between those in a vegetative state/unresponsive wakefulness state (UWS) and those in a minimally conscious state (MCS). Prognostic insights into patient evolution are provided, indicating the number of patients showing improvement and those not showing improvement across different modalities. The counts in each cell represent the corresponding number of patients within the specified modality and diagnostic/prognostic category. Not all the patients have prognostic information; thus, the sum of the prognostic categories is not equal to the counts in the 'All patients' column. Information on the patients' sex, aetiology (traumatic brain injury, anoxic damage, stroke and other causes), age, and whether they were acute or chronic at the time of the tests is given in Supplementary Tables 1–4.

Table 2 Overview of multimodal neuroimaging markers used as features

Modality type	Modality	Paradigm	Metrics
Dynamic, functional	EEG	Resting state	Spectral, information theory, connectivity markers
Dynamic, functional	EEG	Local-global auditory task	Spectral, information theory, connectivity and evoked markers
Dynamic, functional	fMRI	Resting state	Cortical and subcortical functional connectivity
Static, functional	FDG-PET	Resting state	Cortical & subcortical metabolic activity
Static, anatomical	dMRI	Resting state	White matter tract fractional anisotropy and mean diffusivity
Static, anatomical	aMRI	Resting state	Cortical thickness, subcortical volume

This table provides a comprehensive overview of the neuroimaging and electrophysiology modalities used in this work, encompassing both dynamic or static and anatomical or functional properties. Each modality is detailed with its associated paradigm and metric. Dynamic modalities, which have multiple time points, include EEG in the resting state and task local-global paradigms, as well as fMRI in the resting state. Static modalities include FDG-PET, dMRI, and aMRI. The second type refers to whether the modality captures anatomical properties (aMRI and dMRI) or functional ones (EEG, fMRI, FDG-PET). Each modality is associated with specific metrics such as cortical and subcortical functional connectivity, metabolic activity, white matter tract properties, and structural measures like cortical thickness and subcortical volume. aMRI = anatomical MRI; dMRI = diffusion MRI; fMRI = functional MRI; FDG-PET = ^{18}F-fluorodeoxyglucose PET.

test for their generalization. The models that did not generalize above chance level are fMRI-RS (0.44 ± 0.08 median balanced accuracy) in Dataset 3 (Italy) and EEG RS (0.49 ± 0.02 median balanced accuracy) in Dataset 2.

Although it was not the focus of this study, we ran additional analyses splitting the unimodal results from Dataset 1 (Paris) per MCS subgroups (MCS– and MCS+) (Supplementary Fig. 3). We observed that the results were stable, with some variation, whereby dMRI and EEG-LG showed better balanced accuracy when comparing UWS and MCS+, and EEG-RS and aMRI exhibited higher performance when contrasting UWS and MCS–.

We ran similar splits for three aetiologies: traumatic brain injury, anoxia and other (where other was a combination of stroke and other combinations or occurrences of aetiologies). The diagnostic per-aetiology splits are given in Supplementary Fig. 4 and the prognostic ones in Supplementary Fig. 5. Owing to the low number of patients in each subcategory per modality being variable and, in some cases, too low, we refrained from further interpretations but provide the supplementary analysis in case it is helpful to future studies.

Pairwise disagreements of unimodal models are higher in MCS and improved patients

When examining unimodal predictions per patient, we observed a discernible difference in the extent of disagreements across patient

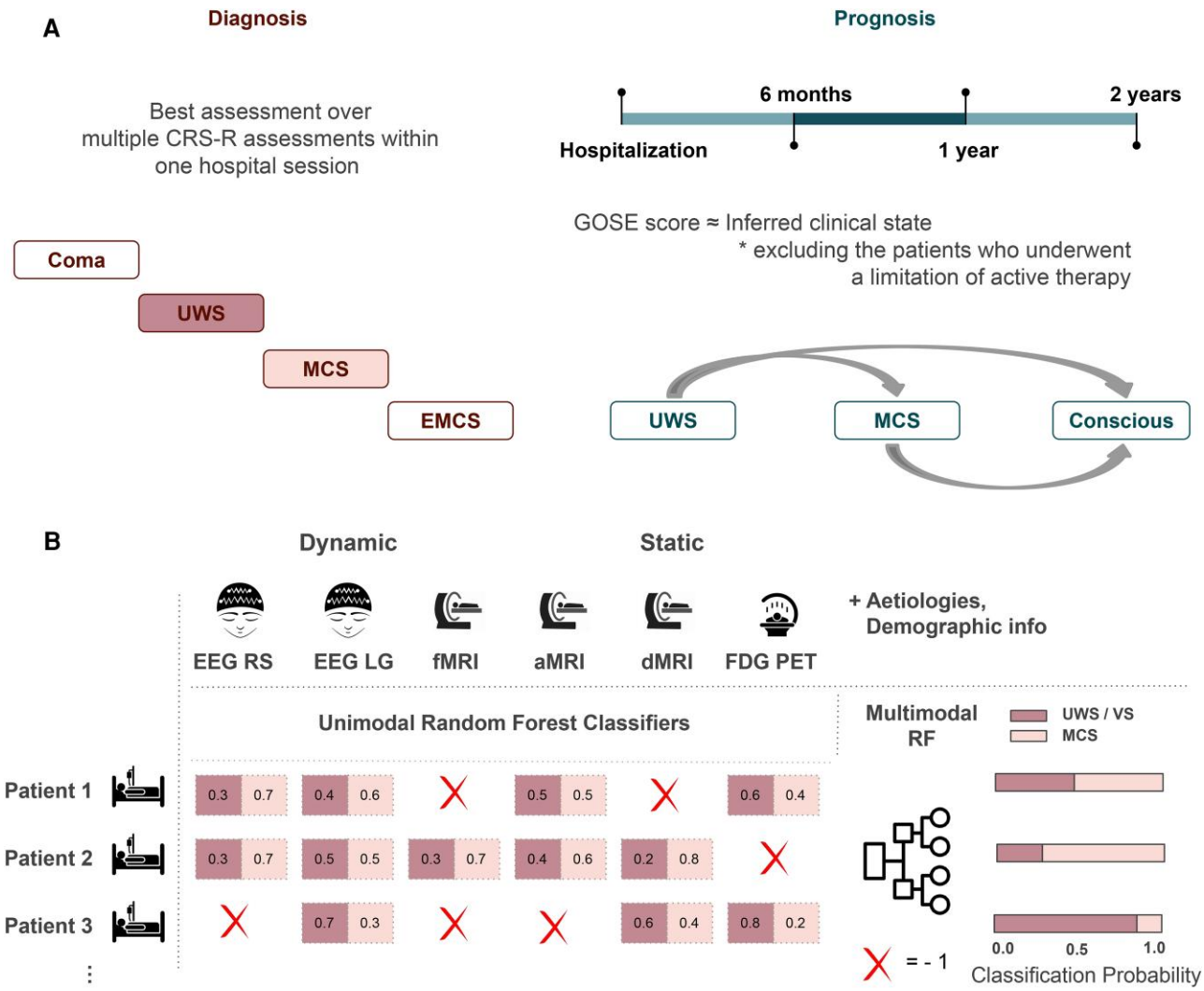

Figure 1 Multimodal assessment methods of diagnostic and prognostic categories of patients with DoC. (**A**) Patients underwent multiple CRS-R assessments during their hospital stay and the best assessment in that range of a week (typically) was taken as the gold-standard clinical diagnosis. The diagnostic clinical categories of patients included in the prediction are UWS and MCS. The prognostic categories are improved or not improved (explained in the 'Materials and methods' section). (**B**) For both the diagnostic and prognostic classification, we ran unimodal RFC to obtain probabilistic estimates of each patient belonging to one or another category. The probabilistic outputs are then combined using a second-level RFC either alone or in combination with the aetiologies and demographic information. Missing values are substituted with −1 (a data imputation approach). The final output is a probability of belonging to either a diagnostic or prognostic category. aMRI = anatomical MRI; CRS-R = Coma Recovery Scale-Revised; dMRI = diffusion MRI; DoC = disorders of consciousness; EMCS = emergent minimally conscious state; FDG-PET = ^{18}F-fluorodeoxyglucose PET; fMRI = functional MRI; GOSE = Glasgow Outcome Scale-Extended; LG = Local Global paradigm; MCS = minimally conscious state; RFC = Random Forest classifiers; RS = resting state; UWS = unresponsive wakefulness syndrome.

groups. Notably, pairwise disagreements were markedly higher in patients classified as being in an MCS compared to those in a UWS (Fig. 3B). The trend in diagnosis was primarily driven by combinations involving aMRI, PET, fMRI and EEG-LG, leaving EEG-RS to be more similar to the other modalities. A similar trend, although less pronounced, was evident when comparing patients who showed improvement versus those who did not (Fig. 3C), where the modalities contributing significantly to these disagreements were dMRI and the EEG paradigms.

Feature importance differs in diagnostic and prognostic models

The fluctuations in diagnostic and prognostic accuracy across individual modalities required exploration of the most influential features for prediction, examining whether these varied when assessing the patient's current state or their outcome. To investigate this, we ranked the features per modality based on their average importance scores across all model iterations. Figure 4A, D and F depict the feature importance scores for fMRI, PET and aMRI, showcasing cortical and subcortical regions; whereas Fig. 4B, C and E show the mean feature importance combined into groups for the EEG-RS and -LG and dMRI scans. The diagnostic aMRI models, together with the prognostic EEG-RS and PET models, were all at chance level; thus, we did not analyse their feature importance scores.

Examining spectral subcategories of low bands (delta, theta, alpha) and high bands (beta and gamma), we found that, in both paradigms, for diagnosis, low-frequency-based features were most relevant (Fig. 4B and E and Supplementary Fig. 16G and I). In

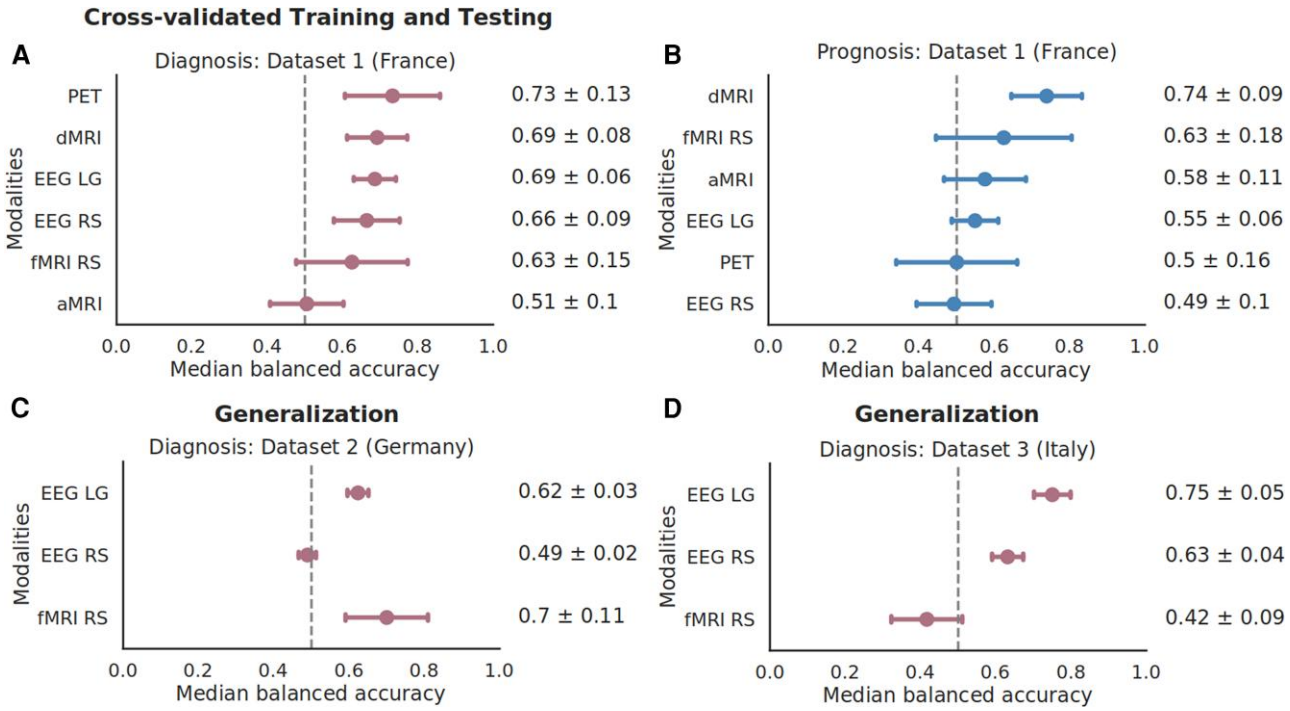

Figure 2 Multivariate unimodal models' accuracy differs depending on the classification target that can be diagnostic or prognostic. (**A**) Unimodal Random Forest classifiers for the six modalities give a diagnostic classification accuracy which is given next to the distributions (median ± standard deviation). The diagnostic classification patient categories are patients in UWS or MCS. (**B**) The same as in **A** but for the prognostic categories (improved and not improved). (**C** and **D**) The same as **A** but for the diagnostic classification trained on Dataset 1: France and tested on the available modalities from Dataset 2: Germany and Dataset 3: Italy. aMRI = anatomical MRI; dMRI = diffusion MRI; fMRI = functional MRI; LG = Local Global paradigm; MCS = minimally conscious state; RS = resting state; UWS = unresponsive wakefulness syndrome.

diagnostic EEG-LG, connectivity-derived features were also highly relevant, while high-frequency-based features were less informative. Conversely, for EEG-LG prognostic prediction (Fig. 4E and Supplementary Fig. 16J), high-frequency features became the most informative. However, the evoked features remained the least important for both diagnosis and prognosis. In the aMRI prognostic models, the most relevant features were the subcortical volume features, followed by salience and visual network cortical thicknesses, with the features from the default mode network (DMN) being the lowest scoring. In the fMRI-RS functional connectivity analysis, subcortical regions to cortical network connectivity features were most informative for diagnostic classification, followed by within-subcortical and cortical-to-cortical region of interest connectivities (Fig. 4A). Conversely, for prognosis (Fig. 4A), the somatomotor cortical network gained importance, accompanied by an increased relevance of the limbic and frontoparietal networks. The visual and salience networks and subcortical to cortical functional connectivity decreased in significance. In diagnostic FDG-PET analysis (Fig. 4D), mean metabolic activity per left or right hemisphere emerged as the most informative feature, followed by cortical networks like the somatomotor and visual networks, while subcortical areas were less informative. In dMRI, the most important feature distinguishing UWS from MCS patients was the combined fractional anisotropy (FA global) (Fig. 4C). This was followed by the right superior fronto-occipital fasciculus, right posterior limb of the internal capsule, and left and right corona radiata. When grouping tracts, projection fibres and brainstem fibres were most informative for diagnosis based on FA (Fig. 4C). For prognosis, the mean diffusivity of commissural fibres rose in importance (Fig. 4C), and the brainstem tracts measured by FA remained among the most informative. Furthermore, there appeared to be an interaction between modality and cortical and subcortical markers, with some subcortical markers being more informative for prognosis and cortical markers contributing more to diagnosis (Fig. 4G).

In addition, we calculated the univariate area under the curve (AUC) values per feature (Supplementary Figs 8–13), both for diagnosis and prognosis. We found a non-linear relationship between the mean feature importance scores and the feature AUC values (Supplementary Fig. 19).

Multimodal integration improves predictive accuracy

In this section, we address two questions: (i) whether the model accuracy improves with an increase in the number of modalities; and (ii) whether extended models perform better than basic ones.

In the case of diagnostic prediction (Fig. 5A), we observed an increase in balanced accuracy for both basic and extended models. The basic model started at chance level and progressively improved to achieve an accuracy above 0.83. For prognosis (Fig. 5B), there was an upward trend in accuracy, with a notable deviation when patients had four modalities, leading to a drop in accuracy, particularly for the basic model. In most cases, the extended model demonstrated superior performance, indicating non-redundant information derived from demographic details and aetiological divisions.

When looking into the trends for the models using a Spearman correlation test, for the diagnostic models, we saw an increase in balanced accuracy for the basic model $r(2294) = 0.49$, $P < 0.0001$, 95% confidence interval (CI): 0.46–0.52; and for the extended model $r(2294) = 0.399$, $P < 0.0001$, 95% CI: 0.36–0.43. For the prognosis, the

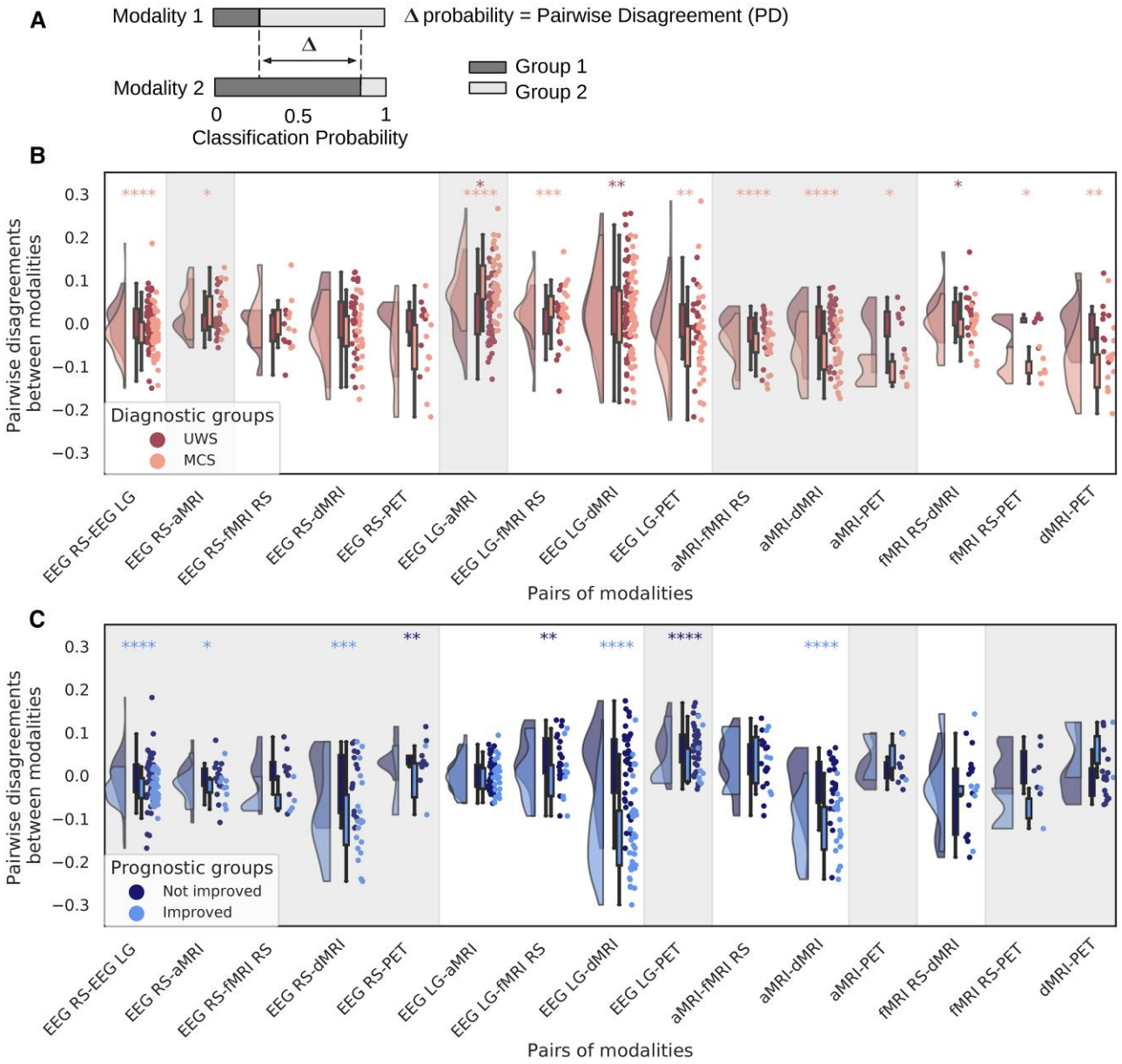

Figure 3 Pairwise disagreements of the classification probabilities are higher for patients in the minimally conscious state and patients who show an improvement in Dataset 1 (France). (**A**) The pairwise disagreement is calculated per patient per pair of modalities as the absolute difference in the classification probabilities (probability of being in Group 1 versus Group 2, either for the diagnostic or the prognostic groups) between two modalities. (**B**) Pairwise disagreements of the classification probabilities are more common in MCS patients. Results are displayed separately for the two diagnostic groups (UWS, dark red; MCS, light red). (**C**) Pairwise disagreements are higher in improved patients than in not improved patients with a difference statistically less strong than the one of the diagnostic groups. Pairwise disagreements of the classification probabilities are more common in improved patients. Results are displayed separately for the two diagnostic groups (Not improved, dark blue; Improved, light blue). The distributions of the pairwise disagreements are tested using a Wilcoxon signed-rank test to see if two paired samples are from the same distribution. The stars above the distributions denote the significance in the colour related to the diagnostic group (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$). The pairwise disagreements that have a grey background are those that include a model that was at chance level in Fig. 2A or B. aMRI = anatomical MRI; dMRI = diffusion MRI; fMRI = functional MRI; LG = Local Global paradigm; RS = resting state; MCS = minimally conscious state; UWS = unresponsive wakefulness syndrome.

positive correlation of the balanced accuracy with the number of modalities was less strong $r(2037) = 0.105$, $P < 0.0001$, 95% CI: 0.06–0.15, and increased for the extended model to $r(2037) = 0.335$, $P < 0.0001$, 95% CI: 0.3–0.37.

The statistical difference between the basic and extended model in diagnosis was for $n = 1$ modalities [$U(500,500) = 55562$, $P < 0.0001$], and $n = 4$ [$U(495,495) = 98743$, $P < 0.0001$], whereas in prognosis was for $n = 2$ [$U(500,500) = 107557$, $P = 0.0007$], $n = 4$ [$U(465,465) = 68766$, $P < 0.0001$] and $n = 5$ [$U(74,74) = 1748.5$, $P = 0.0002$] modalities. For the diagnosis for $n = 2$ [$U(500,500) = 124698$, $P = 1$], $n = 3$ [$U(500,500) = 130672$, $P = 1$] and $n = 5$ [$U(301,301) = 42645$, $P = 0.92$], the basic and extended model results were not statistically different; the same was true for $n = 1$ [$U(500,500) = 116245$, $P = 0.28$] and $n = 3$ [$U(500,500) = 124105$, $P = 1$] for prognosis.

It is worth noting that the number of patients with five or more modalities was low. When removing the patients with five

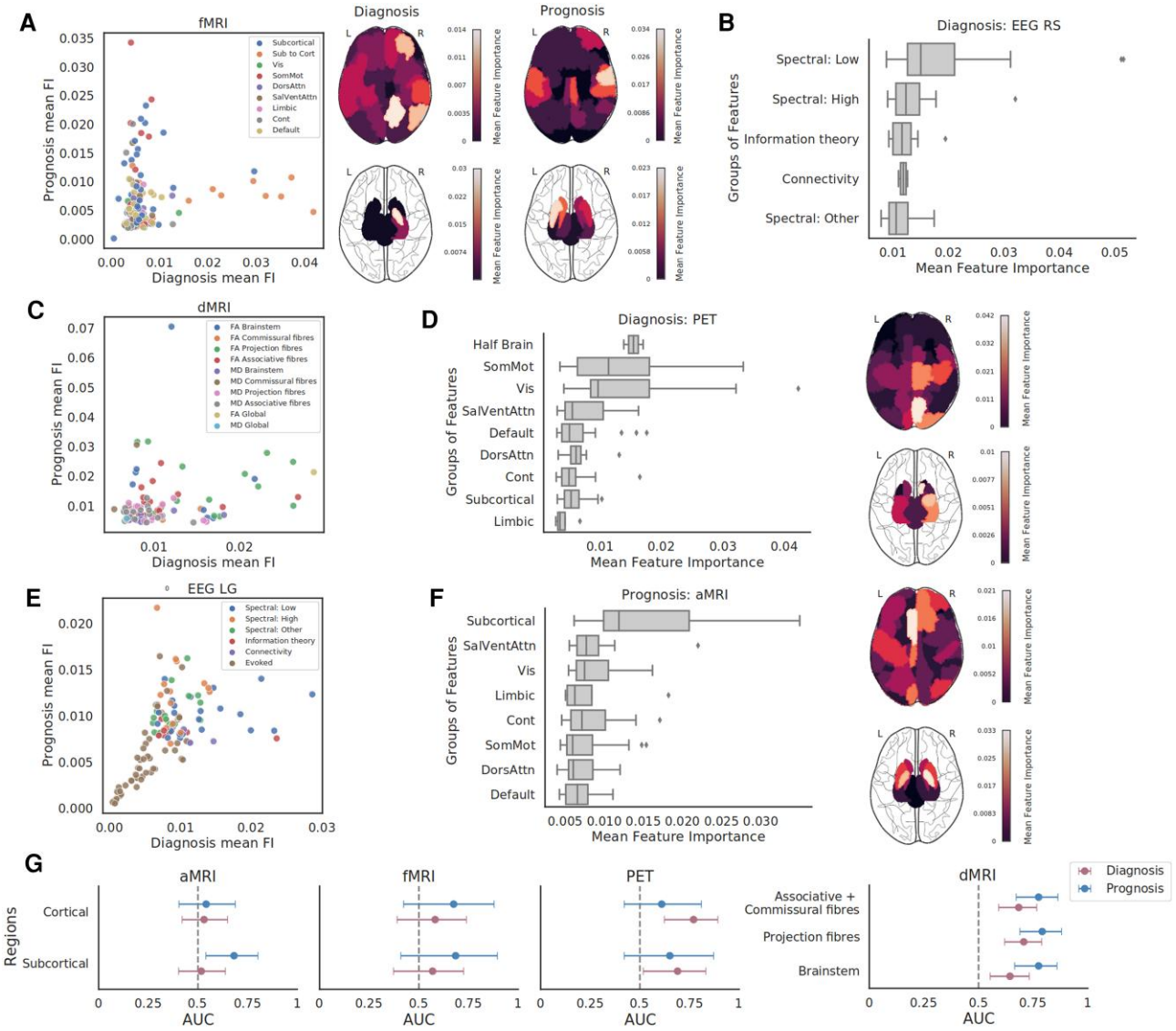

Figure 4 Reordering of the feature importance (FI) distributions per group of features for diagnosis and prognosis in Dataset 1 (France). (**A**) Feature importance of the diagnostic and prognostic prediction using the fMRI-RS scans. The swarm plot shows the feature importance split into within subcortical functional connectivity, within cortical connectivity subdivided into the seven cortical networks, and subcortical to cortical functional connectivity. The brain plots show the feature importance of the within cortical and within subcortical functional connectivity per region of interest. (**B**) Feature importance of the diagnostic prediction using the EEG-RS recordings. The bar plot shows the feature importance split into conceptual families: connectivity (wSMI), information theory (Kolmogorov complexity and permutation entropy), low spectral (delta, theta and alpha frequency bands), high spectral (beta and gamma bands), and other spectral marker summaries. (**C**) Feature importance of the diagnostic and prognostic prediction using the dMRI scan. The swarm plot shows the feature importance split into two measures, fractional anisotropy (FA) and mean diffusivity (MD), which are subdivided into global brain-wide measures and families of tracts: projection fibres, brainstem, commissural fibres and associative fibres. (**D**) Feature importance of the diagnostic prediction using the FDG-PET scan. On the brain plots, the feature importance of the metabolic activity per cortical or subcortical region of interest are shown. The bar plot shows the feature importance split into cortical networks and the subcortical regions, as well as the importance of the half-brain (left or right hemisphere) metabolic activity. (**E**) Feature importance of the diagnostic and prognostic prediction using the LG task-based EEG recordings. The bar plot shows the feature importance split into the same conceptual families as the EEG-RS with the addition of the evoked markers coming from the task-based paradigm. (**F**) Feature importance of the prognostic prediction using the aMRI scan. (**G**) Average AUC value per cortical or subcortical regions of interest (first three plots) or fibre tracts (fourth plot) for the neuroimaging modalities, split per diagnosis and prognosis. The positive classes are MCS and Improved patients, if the AUC value is from 0.5 to 1, the given marker is higher in MCS or Improved patients. If the AUC value is from 0 to 0.5, the given marker is higher in the UWS or not-Improved group of patients. The feature importance of cortical thickness and subcortical volume are shown on the brain plots. All of the bar plots, including the brain plots from four different views, are given in Supplementary Figs 10–12. AUC = area under the curve; aMRI = anatomical MRI; dMRI = diffusion MRI; FDG-PET = ^{18}F-fluorodeoxyglucose PET; fMRI = functional MRI; LG = Local Global paradigm; RS = resting state; MCS = minimally conscious state; UWS = unresponsive wakefulness syndrome; wSMI = weighted symbolic mutual information.

modalities, the increasing trend remained; however, the correlation was less strong [diagnostic basic model $r(1993) = 0.451$, $P < 0.0001$, 95% CI: 0.42–0.49; diagnostic extended model $r(1993) = 0.327$, $P < 0.0001$, 95% CI: 0.29–0.37; prognostic basic model $r(1963) = 0.09$, $P = 0.0001$, 95% CI: 0.05–0.13; prognostic extended model $r(1963) = 0.302$, $P < 0.0001$, 95% CI: 0.26–0.34].

Discussion

Interpretable modelling approach for sparse multimodal neuroimaging datasets

The assessment of DoC presents significant clinical challenges due to the potential dissociation between behavioural responsiveness and consciousness, as well as the limitations of behavioural scales in the presence of confounding factors. While individual neuroimaging and electrophysiological modalities have shown potential in improving diagnostic and prognostic accuracy, they are inherently limited in scope. Multimodal approaches, which integrate information across various modalities, hold promise for addressing these limitations, but there has been a lack of large-scale, multicentric studies systematically evaluating their complementarity and effectiveness. In this study, we adopted a comprehensive, multicentric approach to overcome these challenges. We integrated six neuroimaging modalities using interpretable machine learning methods. To handle the inherent challenges of sparse and heterogeneous multimodal datasets, we proposed a two-model stacking approach. This methodology effectively addressed issues such as the low patient-to-feature ratio, modality-specific data sparsity, and the heterogeneity of modality combinations, enabling a robust analysis of both diagnostic and prognostic dimensions of DoC.

Differentiating unimodal classification scores for a patient's current state and its evolution

Our analysis of neuroimaging and electrophysiological modalities for patients with DoC revealed intriguing modality differences in diagnostic and prognostic accuracy. Modalities that capture the structural preservation of the brain and its networks (aMRI and dMRI) become more relevant in the evolution of a patient's state, and hence for prognosis. Conversely, modalities that capture electrical activity (EEG) and metabolic activity (PET) are mostly relevant to diagnosis. This shift in accuracy rankings for modalities between diagnostic and prognostic classifications underscores their complementarity.

PET displayed the highest accuracy in discriminating between UWS and MCS patients but carried little prognostic information in half of the cross-validated splits (Fig. 2A and B). The diagnostic results aligned with previous studies showing that metabolic data distinguish MCS from UWS patients,[35,41,46,47] with one study reporting that the gradient increase continues to EMCS patients and healthy controls.[36] Furthermore, FDG-PET demonstrated better suitability in discriminating DoC diagnoses compared to MRI-derived measures, including active fMRI (where PET had higher sensitivity for identifying MCS patients),[35] aMRI and fMRI-RS.[47] In our work, we observed a drop in the prognostic accuracy of PET compared to diagnostic; however, the distribution was large (Fig. 2A and B), indicating that certain cross-validated splits are better predicted whereas other splits show opposing trends to their training subset. This large variation in accuracy could arise due to the large imbalance in patients who improve ($n = 9$) versus those who do not ($n = 35$) (Table 1). One study from the literature reported a drop in the patient recovery value of PET compared to the diagnostic, albeit still as high as 74%,[35] whereas another study, in a sample of 20 patients, did not show a prognostic value.[48] Comparing the PET with the EEG models, one study showed a higher sensitivity of EEG models compared to FDG-PET, although the AUC of the diagnostic prediction did not differ significantly.[41]

In our study, we used the same RS and LG paradigm EEG markers as those reported in two previous studies.[18,27] There is a partial overlap in the data with previous studies, and the results are in accordance with our study. Additionally, after collecting the markers into large groups and contrasting their diagnostic and prognostic accuracies, we observed a drop in the accuracy of both EEG paradigms for the evolution of patient states (Fig. 2A and B) (in contrast to higher accuracy in the diagnostic models), as previously shown in other studies.[24,49] On the contrary, other studies were able to show both the diagnostic and prognostic potential of EEG.[23,39,40,50]

In our work, we observed a diagnostic accuracy of aMRI close to chance levels (Fig. 2A). One study found similar diagnostic results (balanced accuracy ranging from 0.45 to 0.63).[31] However, Annen *et al.*[32] showed a high diagnostic area under the receiver operating curve of 96% using grey matter and white matter volume, a prediction comparable to that from FDG-PET. Importantly, we used a different implementation of the cortical thickness and subcortical volume estimates that was created specifically for clinical data with various resolutions and originating from different neuroimaging centres.[51-54] The differences in our findings compared with those reported by Annen *et al.*[32] may have arisen due to the different methodologies and diverse cohorts, leading to questions that should be answered in future work, comparing both FreeSurfer implementations across different aetiologies (for example, traumatic versus anoxic). In comparison to the diagnostic models, we saw an increase in accuracy when looking into prognosis (Fig. 2A and B). When there was physical damage to tissue that could be quantified with neuroimaging, regeneration was slow, and this could dictate patient prognosis. A study by van der Vliet *et al.*[55] showed that even patients with severe initial deficits can reach favourable outcomes but require longer time constants for recovery, consistent with the importance of anatomical preservation for prognosis. In future studies, it can be tested whether anatomical preservation scales with the time for improvement or recovery in a linear way. The differential relevance of some subcortical markers for prognosis and other cortical markers for diagnosis (Fig. 4G) may reflect underlying differences in neuroplasticity mechanisms and their timescales. Importantly, cortical plasticity has been more extensively studied,[56,57] in contrast to subcortical plasticity.[56] A few studies suggest that subcortical structures are not passive relays but play a central role in cortical plasticity and cross-modal functional reorganization.[57-59] Thus, anatomical preservation, especially subcortical, could be more indicative of improvement due to its influence on cortical plasticity processes. A follow-up study would be to test the functional preservation at different time points post-injury of patients that show higher subcortical anatomical preservation.

For fMRI-RS, we saw a similar median balanced accuracy for diagnosis and prognosis but a larger distribution for prognostic prediction, similar to the results for PET, suggesting an influence depending on the random cross-validated splits. However, the ordering of the modalities' performances was different, making fMRI-RS one of the most informative modalities for patient state evolution. Previous work has shown the potential of fMRI-RS to discriminate between patients and controls[17] and between UWS and MCS patients.[21,29,39] The higher accuracy reported in the literature compared to our results could be attributed to methodological differences between seed-based and atlas-based parcellation of RS networks, or the fact that one of the studies[21] used data only from patients for whom the clinical diagnosis based on CRS-R was congruent with PET scans. Two studies have tested the outcome prediction of DoC patients at 3 months with an accuracy range of 0.69–0.78,[40] and 0.81,[22] but no predictive value at 12 months was observed.[40] Given that we looked at a different

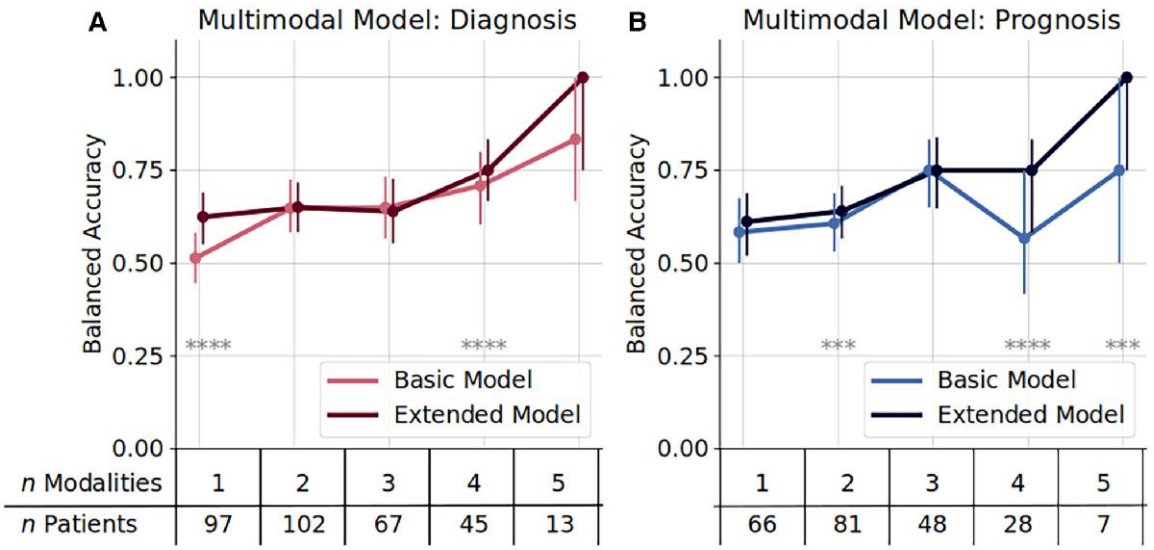

Figure 5 Increasing trends in the multimodal model balanced accuracy for the basic models (only neural modalities) and the extended models (neural modalities plus information on the patient aetiologies and demographics) of Dataset 1 (France). (**A**) Balanced accuracy for the diagnostic models (basic and extended) for patients with 1–5 different neuroimaging modalities (*x*-axis). (**B**) Same as **A**, apart from the prognosis. The error bars represent the first (Q1) and third (Q3) quartile intervals of the distributions. On the *x*-axis, the *top row* represents the number of modalities across the function, and the *second row* represents the number of patients with the given number of modalities. The stars represent significance following Mann–Whitney U-tests (Bonferroni corrected) between the basic and extended models ($^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$). *n* = number of elements in the given distribution.

prognostic metrics, the results are not directly comparable, but there is consistent evidence that fMRI activity does contain prognostically relevant information for DoC patients.

Previous diagnostic studies using dMRI have reported accuracies as high as 0.95[60] and in the range 0.81–0.84 using a multivariate searchlight analysis of whole-brain thalamo-cortical tracts.[61] In prognostic studies of cardiac arrest patients, FA values were shown to reach values of 0.95 sensitivity and 1 specificity,[62] 0.98 AUC in a larger follow-up study,[28] and 0.93 AUC 1-year prognostic value of global deep white matter metrics in TBI patients.[63] The consistency of our results with previously reported findings emphasizes the importance and potential of using dMRI to aid the diagnostic and prognostic assessment of patients with DoC.

Generalization tests across independent datasets demonstrated varying performance (Fig. 2C and D). The modalities that could not be generalized include fMRI-RS from Dataset 3. This discrepancy may be attributed to the heterogeneity in the acquisition parameters compared to the training set (see Supplementary material, 'Methods' section). Task EEG (EEG-LG) outperformed EEG-RS across all centres, with a notably reduced effect in Dataset 2 from Germany. The higher performance of EEG-LG-based models highlights the critical role of active paradigms[1] in assessing patients with DoC, as these paradigms likely enhance and regulate patients' attentional states. In contrast, resting-state paradigms may be less robust to cross-centre variability due to their dependence on intrinsic brain activity, which is more susceptible to external and patient-specific factors. These findings, taken together, underscore the importance of accounting for modality-specific and centre-related acquisition parameters to improve model generalizability across centres.

Pairwise disagreements between modalities

Studying pairwise disagreements across modalities is important, as it can point to cases for whom a dissociation can elucidate the potential for recovery. We observed more pairwise disagreements for MCS patients and for those that improved (Fig. 3), suggesting that some signals may capture a more positive clinical picture, while others do not. The sources of these disagreements can be neural or non-neural. Neural examples include the case of a functional hemispherectomy, when a patient showed almost no metabolic activity in the left hemisphere with preserved white matter tracts,[64] or islands of preserved cortical activity that are posited to exist in this group of patients.[65] Furthermore, UWS patients with unfavourable EEG features have shown an increase in fMRI between-network connectivity and a decrease in DMN within-network connectivity (but not significant).[39] Although EEG and PET have been shown to be highly correlated, EEG connectivity patterns differed in PET-negative and PET-positive patients.[24] Another study found a diagnostic difference (between healthy controls and DoC patients) in metabolic activity and mixed results in positive and negative DMN connectivity, but no significant results in grey matter volume.[36] A disagreement between metabolic activity and grey matter has been found in the left-sided language network of MCS– and MCS+ patients.[37] The first exhibited lower metabolic values in the left middle temporal cortex and a metabolic functional disconnection between the left angular gyrus and the left prefrontal cortex. The authors concluded that brain function and not grey matter structure supports the clinical signs of language processing.[37] In some patients, only the dMRI images showed a consistent loss of white matter compared to the seemingly unchanged appearance of structural images.[66] Furthermore, the reliability of discriminating between MCS and UWS patients is often compromised due to the limited sensitivity of scalp EEG, as demonstrated by instances where pathological brain activity masks normal neuronal patterns in awake individuals, suggesting the potential for complex dissociations in severe cases of brain injury.[10] The proposal to rename the MCS to a cortically mediated state[67] underlines the fact that the MCS encompasses a broad and heterogeneous range of

conditions. This spectrum includes unconscious patients who exhibit residual cortical activity leading to observable behaviour, and conscious patients who, despite possibly being self-aware, are hindered by executive deficits that prevent them from effectively using a communication code or responding functionally to commands.[67,68] This may explain why there are more pairwise disagreements in the MCS compared to the UWS, and that some specific modalities might not capture the heterogeneity. A complementary perspective is that the differences in prediction can also stem from aetiology specificities, such as EEG alpha power, which has been shown to be suppressed in severely post-anoxic patients and does not differ between patient groups with other aetiologies.[30,69] All these examples corroborate the fact that pairwise disagreements from a neural origin are common, and their hierarchical importance in diagnosis and prognosis needs to be further studied.

One non-neural source of disagreement was data quality, which, even with stringent exclusion criteria, was lower in the DoC patient group, possibly affecting analyses down the line. Furthermore, the state of the patients fluctuates across various time scales, whereas in our case, we worked with one diagnosis per patient, which could be the source of disagreement between the diagnostic models. This would not be the case for prognostic prediction. A way to surpass this would be to look at the variability of multiple CRS-R tests and check the disagreement in light of the patient's clinical fluctuations.

Differing importance of feature groups within modalities, for diagnostic and prognostic prediction

EEG

Previous results have shown that the most informative EEG features to differentiate between UWS and MCS patients are absolute alpha power, permutation entropy, Kolmogorov complexity, and a connectivity measure in the theta band (weighted symbolic mutual information, wSMI).[18,27] In our work, we found that in EEG, the spectral feature groups were the most informative (Fig. 4B and E and Supplementary Fig. 16G, I and J), with low frequencies being important for diagnosis and high frequencies for prognosis. This finding can be related to the mesocircuit hypothesis that provides a framework for understanding the recovery of consciousness after severe brain injuries by focusing on the interconnected roles of cortical and subcortical structures.[7,70] Specifically, the 'ABCD' model of neuronal recovery proposes that sequential changes in EEG power spectra can be categorized into four broad stages, each reflecting varying degrees of thalamocortical deafferentation severity.[7,70] The importance of high beta and gamma frequencies (AUC values in Supplementary Figs 8 and 9) in the DoC patients could further help in distinguishing the MCS patients that belong to group C of the 'ABCD' model of corticothalamic dynamics. However, for prognosis, the AUC of the various high-frequency features was higher for patients who did not improve (Supplementary Figs 8 and 9).

In prognostic EEG studies of prolonged DoC, the presence of dominant delta frequencies and reduced EEG amplitudes was related to worse outcomes, whereas the dominance of alpha frequencies, preserved EEG reactivity, and an increase in the dominant frequency were associated with improvement.[10,23] On the contrary, in one study, higher delta power was associated with improved outcomes for patients transitioning from UWS to MCS,[48] which we also found in the EEG-LG results (Supplementary Fig. 9). These discrepancies can come from aetiology-dependent differences such as the slowing of EEG being relevant for prognosis in toxic encephalopathies, a transient increase in slow waves or suppression of sensory stimuli in patients with a traumatic brain injury, contrary to the increase in gamma and alpha frequencies in acute patients with a subarachnoid haemorrhage.[10]

We observed that connectivity metrics from LG had a stronger weight compared to RS (Fig. 4B and E and Supplementary Figs 8 and 9). Previous work has demonstrated the importance of network metrics over frequency power,[24] the relevance of coherence across various regions and frequency bands for improving UWS patients,[23] and a stronger delta network connectivity in patients with negative outcomes.[24] On the contrary, one study found no network features related to outcome at 3 or 6 months post-injury,[49] with only relative alpha power improving prediction accuracy at 3 months in contrast to prediction using only clinical features. This is consistent with our findings, where univariate AUC values of the connectivity features were at chance level for prognosis, both for EEG-RS and -LG.

When using oddball auditory perception paradigms, event-related potentials such as mismatch negativity and P300 have been reported to have low sensitivity in MCS patients.[10,71] Accordingly, in our work, most of the evoked features were the least informative ones (Fig. 4E).

Neuroimaging

Multiple studies using PET or fMRI-RS have reported a brain-wide network difference between UWS and MCS patients,[21,41,46] with some reporting that the left hemisphere is more impaired in UWS.[31,47] In our work, in most neuroimaging modalities, we observed a distributed brain-wide feature importance (Fig. 4A, D and F, brain plots), with the exception of fMRI-RS prognostic feature importance.

However, the DMN has mostly been shown to differ among diagnostic groups, to have lower activity in UWS compared to MCS,[15,17,22,47] both at enrollment and at discharge from the intensive care unit,[39] and to relate to recovery outcome.[20,22,25,26,47] Structural information, such as grey matter volume[32] and structural integrity,[31] has also been reported to be highest in DMN regions. In our prognostic results, the DMN features were moderately informative compared to the other networks in the aMRI and fMRI scans (Fig. 4A and F and Supplementary Fig. 16B and D). In the fMRI scans, where both diagnostic and prognostic models were comparable, DMN features were more informative for patient prognosis.

A few studies have reported the strongest metabolic activity reduction in frontoparietal areas[46,47,72] or specifically in the medial prefrontal cortex (part of DMN) or lateral parietal cortex.[22] The primary and associative somatosensory areas have been associated with diagnosis,[46] with one FDG-PET study observing this in the best-preserved hemisphere.[68] A few studies have reported specific diagnostic differences only for the auditory network in fMRI-RS[17,21] and glucose metabolism to be higher in MCS than in UWS.[48] In PET and fMRI-RS, we also observed the higher relevance of somatomotor and visual networks. In contrast, one study reported that higher-order networks (DMN, salience, dorsal attention network, left and right fronto-parietal network and temporal network) have better diagnostic accuracy than low-order networks (sensorimotor, auditory and visual networks) as derived by their structural integrity,[31] but we could not compare these findings to our results because the aMRI model was at chance level.

We found that the subcortical areas are more important for prognosis than diagnosis; however, this difference has not been the focus of neuroimaging investigations. Various subcortical areas have been reported to differ between diagnostic groups, such as

lower metabolic activity in the brainstem,[32,46] thalamus,[32,46] and the caudate and para-hippocampal areas.[32] Thalamic white matter integrity is also affected in patients,[60] along with the pathway linking the posterior cingulate cortex/precuneus with the thalamus, as evidenced by their mean FA values[73]; whereas brainstem white matter tract preservation has only been observed in ischaemic-hypoxic patients,[66] and no mean diffusivity brainstem differences have been identified between MCS and UWS.[60] Our results suggest that the preservation of subcortical structures (potentially reflecting the integrity of large-scale arousal and integration networks, as well as their aforementioned neuroplasticity) may be a key factor supporting long-term recovery, highlighting its relevance to both outcome prediction and potential therapeutic targets (as previously investigated).[74,75] It is important to note that the signal-to-noise ratio when imaging subcortical areas can be lower than that in cortical regions.

Multimodal combinations

An increase in the balanced accuracy for patients imaged using multiple modalities was expected, according to the literature, for the combinations of EEG and FDG-PET,[41] EEG and fMRI-RS,[39,40] and EEG and dMRI,[42] aligning with our results. Similar studies on cardiac arrest patients revealed that a model using only three FA features outperformed models incorporating either only the FA global scores, clinical data or grey matter apparent diffusion coefficient,[62] or an enhanced AUC with the integration of scores and metrics from multiple modalities (EEG, aMRI, dMRI).[28] A case study by Comanducci *et al.*[11] illustrated how a longitudinal multimodal analysis can reveal covert signs of consciousness in an unresponsive patient. Additionally, Rohaut *et al.*[33] demonstrated that integrating multimodal observations enhances neuro-prognostication performance. While features across modalities may correlate due to shared neural sources (e.g. slow EEG rhythms and reduced metabolism), each modality also captures distinct noise. This uncorrelated noise can lead to synergy in multimodal models, even when the signals overlap, by improving the overall signal-to-noise ratio. Thus, synergy may reflect both complementary information and the statistical benefits of combining modalities.

Benefits, caveats and the future of integrative multimodal neuroimaging for DoC

The current electrophysiology and neuroimaging modalities capture some aspect of brain preservation—either the underlying structure or dynamics, and they contain non-redundant information. The importance of having a multidimensional perspective of this clinical group has been increasingly emphasized.[4,13,33,39-41] Overall, numerous reviews have placed focus on the advantage of having multimodal acquisitions[4,7,12,13,76]; however, the practicalities, given the limitations already exposed in this paper, make implementation challenging. Furthermore, machine learning-based approaches, trained on behavioural labels to differentiate between UWS and MCS patients, may overlook conscious but unresponsive individuals, posing a circularity problem; however, there is some robustness to mislabelling if classifiers are trained with a sufficient amount of data.[27] Importantly, random forest classifiers are non-linear models, thus the relationship between feature importance and the models (measured here through the balanced accuracies) is non-linear. While data collection procedures for Dataset 1 (France) were standardized across time for EEG (LG and RS) using the same system, one change occurred in the neuroimaging protocols, due to the installation of a new hospital scanner. Although this may have introduced variability, all acquisition parameters were carefully documented as described in the 'Materials and methods' section and Supplementary material. For Datasets 2 and 3, collected across multiple centres, some variability in EEG hardware and MRI acquisition was unavoidable. However, we applied harmonized preprocessing pipelines and used feature extraction methods robust to differences in EEG channel count and MRI preprocessing to reduce site-related bias. A potential shortcoming of our study is that long-term outcome was assessed using a single phone-guided CRS-R evaluation, which, although validated,[33,77] may be subject to inter-rater variability and reduced precision. The absence of repeated assessments limited our ability to account for potential fluctuations in clinical state over time. Furthermore, although feature importance can be very informative in understanding how decisions are made, it has important limitations, and further analysis of redundancy and synergy can paint a clearer image of their relationships. Another limitation of the study is the lack of healthy controls, which would render the results more reliable. The reliability of the unimodal model results can be confirmed if the models classify the healthy controls as being in an MCS category and not in UWS. Lastly, future investigations should focus on the distinction between different aetiologies of patients with DoC using combined multimodal approaches (due to interactions of neural signals with aetiology[30,69]).

Conclusions

In this study, we developed an explainable machine learning approach for the classification of DoC patients from a large dataset of multimodal neuroimaging and electrophysiology recordings. The observed distinctions in accuracy, feature importance and pairwise disagreements underscore the need for tailored strategies in leveraging different modalities for enhanced clinical decision-making. Comparing the current states of patients and their evolution across modalities and features (regions or other signal summaries) may pave the way for more thorough investigations into aetiologies or integrative neuroimaging studies with narrower hypotheses.

Data availability

The data are not publicly available. Codes used in the analyses will be made publicly available upon publication at https://github.com/DraganaMana/multimod_doc.

Acknowledgements

We thank all the participants who took part in the studies. We would like to thank the work and support of the clinicians at the Neuro ICU, DMU Neurosciences, APHP Sorbonne Université, Hôpital de la Pitié Salpêtrière, Paris, France; University Hospital of the Ludwig-Maximilians-University of Munich; Therapiezentrum Burgau; the University of Milan; Fondazione Don Carlo Gnocchi Santa Maria Nascente; the Weizmann Institute of Science, and the patient families whose consent and understanding are essential to the progress of the field.

Funding

This work was supported by the Ecole Doctorale Frontières de l'Innovation en Recherche et Education–Fondation Bettencourt

Schueller (to D.M.). This project is part of the multicentric application for the EU ERAPerMed Joint Translational Call for Proposals for 'Personalised Medicine: Multidisciplinary research towards implementation' (ERA PerMed JTC2019). It is funded by local funding agencies of the participating countries (for France it is the Agence Nationale de la Recherche ANR, funding code: ANR-19-PERM-0002, for Germany the Federal Ministry of Education and Research BMBF, funding code: 01KU2003, for Italy, the Fondazione Regionale per la Ricerca Biomedica, funding code: GA 77982), also supported and funded by the Italian Ministry of Health—Ricerca Corrente 2025–2027 (A.C., C.D., T.A.). This project is supported by the MODELDxConsciousness Consortium (Flag-ERA JTC 2023); as well as the ECOS-SUD A20M02 (Argentina/France), and by Paris Brain Institute America's project on Consciousness mapping. ERC-2022-SYG Grant number 101071900 neurological mechanisms of injury and sleep-like cellular dynamics (NEMESIS) (to M.R.)

Competing interests

J.D.S. and L.N. are scientific co-founders of NeuroMeters (have scientific advisory activity but no executive or management activity). M.R. is shareholder and scientific advisor of Intrinsic Powers, a spin-off of the University of Milan. This affiliation in no way affects the content of this article.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Bodien YG, Allanson J, Cardone P, et al. Cognitive motor dissociation in disorders of consciousness. *N Engl J Med*. 2024;391:598-608.
2. Owen AM, Coleman MR, Boly M, Davis MH, Laureys S, Pickard JD. Detecting awareness in the vegetative state. *Science*. 2006;313:1402.
3. Bruno MA, Vanhaudenhuyse A, Thibaut A, Moonen G, Laureys S. From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes: Recent advances in our understanding of disorders of consciousness. *J Neurol*. 2011;258:1373-1384.
4. Gallucci A, Varoli E, Del Mauro L, et al. Multimodal approaches supporting the diagnosis, prognosis and investigation of neural correlates of disorders of consciousness: A systematic review. *Eur J Neurosci*. 2024;59:874-933.
5. Jennett B, Plum F. Persistent vegetative state after brain damage: A syndrome in search of a name. *Lancet*. 1972;299:734-737.
6. Giacino JT, Ashwal S, Childs N, et al. The minimally conscious state: Definition and diagnostic criteria. *Neurology*. 2002;58:349-353.
7. Edlow BL, Claassen J, Schiff ND, Greer DM. Recovery from disorders of consciousness: Mechanisms, prognosis and emerging therapies. *Nat Rev Neurol*. 2021;17:135-156.
8. Giacino JT, Kalmar K, Whyte J. The JFK coma recovery scale-revised: Measurement characteristics and diagnostic utility. *Arch Phys Med Rehabil*. 2004;85:2020-2029.
9. Wannez S, Heine L, Thonnard M, Gosseries O, Laureys S. The repetition of behavioral assessments in diagnosis of disorders of consciousness. *Ann Neurol*. 2017;81:883-889.
10. Comanducci A, Boly M, Claassen J, et al. Clinical and advanced neurophysiology in the prognostic and diagnostic evaluation of disorders of consciousness: Review of an IFCN-endorsed expert group. *Clin Neurophysiol*. 2020;131:2736-2765.
11. Comanducci A, Casarotto S, Rosanova M, et al. Unconsciousness or unresponsiveness in akinetic mutism? Insights from a multimodal longitudinal exploration. *Eur J Neurosci*. 2024;59:860-873.
12. Edlow BL, Fecchio M, Bodien YG, et al. Measuring consciousness in the intensive care unit. *Neurocrit Care*. 2023;38:584-590.
13. Kondziella D, Bender A, Diserens K, et al. European academy of neurology guideline on the diagnosis of coma and other disorders of consciousness. *Eur J Neurol*. 2020;27:741-756.
14. Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L. Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci U S A*. 2009;106:1672-1677.
15. Vanhaudenhuyse A, Noirhomme Q, Tshibanda LJF, et al. Default network connectivity reflects the level of consciousness in non-communicative brain-damaged patients. *Brain*. 2010;133:161-171.
16. Faugeras F, Rohaut B, Weiss N, et al. Event related potentials elicited by violations of auditory regularities in patients with impaired consciousness. *Neuropsychologia*. 2012;50:403-418.
17. Demertzi A, Gómez F, Crone JS, et al. Multiple fMRI system-level baseline connectivity is disrupted in patients with consciousness alterations. *Cortex*. 2014;52:35-46.
18. Sitt JD, King JR, El Karoui I, et al. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain*. 2014;137:2258-2270.
19. King JR, Sitt JD, Faugeras F, et al. Information sharing in the brain indexes consciousness in noncommunicative patients. *Curr Biol*. 2013;23:1914-1919.
20. Silva S, De Pasquale F, Vuillaume C, et al. Disruption of posteromedial large-scale neural communication predicts recovery from coma. *Neurology*. 2015;85:2036-2044.
21. Demertzi A, Antonopoulos G, Heine L, et al. Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients. *Brain*. 2015;138:2619-2631.
22. Wu X, Zou Q, Hu J, et al. Intrinsic functional connectivity patterns predict consciousness level and recovery outcome in acquired brain injury. *J Neurosci*. 2015;35:12932-12946.
23. Schorr B, Schlee W, Arndt M, Bender A. Coherence in resting-state EEG as a predictor for the recovery from unresponsive wakefulness syndrome. *J Neurol*. 2016;263:937-953.
24. Chennu S, Annen J, Wannez S, et al. Brain networks predict metabolism, diagnosis and prognosis at the bedside in disorders of consciousness. *Brain*. 2017;140:2120-2132.
25. Kondziella D, Fisher PM, Larsen VA, et al. Functional MRI for assessment of the default mode network in acute brain injury. *Neurocrit Care*. 2017;27:401-406.
26. Threlkeld ZD, Bodien YG, Rosenthal ES, et al. Functional networks reemerge during recovery of consciousness after acute severe traumatic brain injury. *Cortex*. 2018;106:299-308.
27. Engemann DA, Raimondo F, King JR, et al. Robust EEG-based cross-site and cross-protocol classification of states of consciousness. *Brain*. 2018;141:3179-3192.
28. Velly L, Perlbarg V, Boulier T, et al. Use of brain diffusion tensor imaging for the prediction of long-term neurological outcomes in patients after cardiac arrest: A multicentre, international, prospective, observational, cohort study. *Lancet Neurol*. 2018;17:317-326.
29. Demertzi A, Tagliazucchi E, Dehaene S, et al. Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Sci Adv*. 2019;5:eaat7603.
30. Colombo MA, Comanducci A, Casarotto S, et al. Beyond alpha power: EEG spatial and spectral gradients robustly stratify disorders of consciousness. *Cereb Cortex*. 2023;33:7193-7210.

31. Medina Carrion JP, Stanziano M, D'Incerti L, et al. Disorder of consciousness: Structural integrity of brain networks for the clinical assessment. *Ann Clin Transl Neurol*. 2023;10:384-396.
32. Annen J, Frasso G, Crone JS, et al. Regional brain volumetry and brain function in severely brain-injured patients. *Ann Neurol*. 2018;83:842-853.
33. Rohaut B, Calligaris C, Hermann B, et al. Multimodal assessment improves neuroprognosis performance in clinically unresponsive critical-care patients with brain injury. *Nat Med*. 2024;30:2349-2355.
34. Bender A, Eifert B, Rubi-Fessen I, Jox RJ, Maurer-Karattup P, Müller F. The neurological rehabilitation of adults with coma and disorders of consciousness. *Dtsch Arzteblatt Int*. 2023;120:605-612.
35. Stender J, Gosseries O, Bruno MA, et al. Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: A clinical validation study. *Lancet*. 2014;384:514-522.
36. Di Perri C, Bahri MA, Amico E, et al. Neural correlates of consciousness in patients who have emerged from a minimally conscious state: A cross-sectional multimodal imaging study. *Lancet Neurol*. 2016;15:830-842.
37. Aubinet C, Cassol H, Gosseries O, et al. Brain metabolism but not gray matter volume underlies the presence of language function in the minimally conscious state (MCS): MCS+ versus MCS– neuroimaging differences. *Neurorehabil Neural Repair*. 2020;34:172-184.
38. Candia-Rivera D, Annen J, Gosseries O, et al. Neural responses to heartbeats detect residual signs of consciousness during resting state in postcomatose patients. *J Neurosci*. 2021;41:5251-5262.
39. Amiri M, Fisher PM, Raimondo F, et al. Multimodal prediction of residual consciousness in the intensive care unit: The CONNECT-ME study. *Brain*. 2022;146:50-64.
40. Amiri M, Raimondo F, Fisher PM, et al. Multimodal prediction of 3- and 12-month outcomes in ICU patients with acute disorders of consciousness. *Neurocrit Care*. 2023;40:718-733.
41. Hermann B, Stender J, Habert MO, et al. Multimodal FDG-PET and EEG assessment improves diagnosis and prognostication of disorders of consciousness. *Neuroimage Clin*. 2021;30:102601.
42. Altmayer V, Sangare A, Calligaris C, et al. Functional and structural brain connectivity in disorders of consciousness. *Brain Struct Funct*. 2024;229:2285-2298.
43. Willacker L, Raiser TM, Bassi M, et al. PerBrain: A multimodal approach to personalized tracking of evolving state-of-consciousness in brain-injured patients: Protocol of an international, multicentric, observational study. *BMC Neurol*. 2022;22:468.
44. Wilson JTL, Pettigrew LEL, Teasdale GM. Structured interviews for the Glasgow outcome scale and the extended Glasgow outcome scale: Guidelines for their use. *J Neurotrauma*. 1998;15:573-580.
45. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in python. *J Mach Learn Res*. 2011;12:2825-2830.
46. Stender J, Kupers R, Rodell A, et al. Quantitative rates of brain glucose metabolism distinguish minimally conscious from vegetative state patients. *J Cereb Blood Flow Metab*. 2015;35:58-65.
47. Rosazza C, Andronache A, Sattin D, et al. Multimodal study of default-mode network integrity in disorders of consciousness. *Ann Neurol*. 2016;79:841-853.
48. Golkowski D, Kiel T, Schorr B, et al. Simultaneous EEG-PET-fMRI measurements in disorders of consciousness: An exploratory study on diagnosis and prognosis. *J Neurol*. 2017;264:1986-1995.
49. O'Donnell A, Pauli R, Banellis L, et al. The prognostic value of resting-state EEG in acute post-traumatic unresponsive states. *Brain Commun*. 2021;3:fcab017.
50. Stefan S, Schorr B, Lopez-Rolon A, et al. Consciousness indexing and outcome prediction with resting-state EEG in severe disorders of consciousness. *Brain Topogr*. 2018;31:848-862.
51. Gopinath K, Greve DN, Das S, Arnold S, Magdamo C, Iglesias JE. Cortical analysis of heterogeneous clinical brain MRI scans for large-scale neuroimaging studies. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. MICCAI 2023. Lecture Notes in Computer Science*. Vol 14227. Springer;2023:35-45.
52. Billot B, Greve DN, Puonti O, et al. SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. *Med Image Anal*. 2023;86:102789.
53. Billot B, Magdamo C, Cheng Y, Arnold ID SE, Das SI, Eugenio Iglesias J. Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain MRI datasets. *Proc Natl Acad Sci U S A*. 2023;120:e2216399120.
54. Iglesias JE, Billot B, Balbastre Y, et al. SynthSR: A public AI tool to turn heterogeneous clinical brain scans into high-resolution T1-weighted images for 3D morphometry. *Sci Adv*. 2023;9:eadd3607.
55. van der Vliet R, Selles RW, Andrinopoulou ER, et al. Predicting upper limb motor impairment recovery after stroke: A mixture model. *Ann Neurol*. 2020;87:383-393.
56. Duffau H. Does post-lesional subcortical plasticity exist in the human brain? *Neurosci Res*. 2009;65:131-135.
57. Nudo RJ. Recovery after brain injury: Mechanisms and principles. *Front Hum Neurosci*. 2013;7:887.
58. Jones EG. Cortical and subcortical contributions to activity-dependent plasticity in primate somatosensory Cortex. *Annu Rev Neurosci*. 2000;23:1-37.
59. Ewall G, Parkins S, Lin A, Jaoui Y, Lee HK. Cortical and subcortical circuits for cross-modal plasticity induced by loss of vision. *Front Neural Circuits*. 2021;15:665009.
60. Fernández-Espejo D, Bekinschtein T, Monti MM, et al. Diffusion weighted imaging distinguishes the vegetative state from the minimally conscious state. *NeuroImage*. 2011;54:103-112.
61. Zheng ZS, Reggente N, Lutkenhoff E, Owen AM, Monti MM. Disentangling disorders of consciousness: Insights from diffusion tensor imaging and machine learning. *Hum Brain Mapp*. 2017;38:431-443.
62. Luyt CE, Galanaud D, Perlbarg V, et al. Diffusion tensor imaging to predict long-term outcome after cardiac arrest: A bicentric pilot study. *Anesthesiology*. 2012;117:1311-1321.
63. Puybasset L, Perlbarg V, Unrug J, et al. Prognostic value of global deep white matter DTI metrics for 1-year outcome prediction in ICU traumatic brain injury patients: An MRI-COMA and CENTER-TBI combined study. *Intensive Care Med*. 2022;48:201-212.
64. Bruno MA, Fernández-Espejo D, Lehembre R, et al. Multimodal neuroimaging in patients with disorders of consciousness showing "functional hemispherectomy". In: *Progress in brain research*. Vol 193. Elsevier B.V.; 2011:323-333.
65. Bayne T, Seth AK, Massimini M. Are there islands of awareness? *Trends Neurosci*. 2020;43:6-16.
66. Newcombe VFJ, Williams GB, Scoffings D, et al. Aetiological differences in neuroanatomy of the vegetative state: Insights from diffusion tensor imaging and functional implications. *J Neurol Neurosurg Psychiatry*. 2010;81:552-561.
67. Naccache L. Minimally conscious state or cortically mediated state? *Brain*. 2018;141:949-960.
68. Hermann B, Sangaré A, Munoz-Musat E, et al. Importance, limits and caveats of the use of "disorders of consciousness" to theorize consciousness. *Neurosci Conscious*. 2021;2021:niab048.
69. Maschke C, Belloli L, Manasova D, Sitt JD, Blain-Moraes S. The role of etiology in the identification of clinical markers of consciousness: Comparing EEG alpha power, complexity, and spectral exponent. *Cereb Cortex*. 2025;35:bhaf254.
70. Schiff ND. Mesocircuit mechanisms underlying recovery of consciousness following severe brain injuries: Model and predictions. In: *Brain function and responsiveness in disorders of consciousness*. Springer International Publishing; 2016:195-204.

71. Daltrozzo J, Wioland N, Mutschler V, Kotchoubey B. Predicting coma and other low responsive patients outcome using event-related brain potentials: A meta-analysis. *Clin Neurophysiol*. 2007;118:606-614.
72. Bodart O, Gosseries O, Wannez S, et al. Measures of metabolism and complexity in the brain of patients with disorders of consciousness. *Neuroimage Clin*. 2017;14:354-362.
73. Fernandez-Espejo D, Soddu A, Cruse D, et al. A role for the default mode network in the bases of disorders of consciousness. *Ann Neurol*. 2012;72:335-343.
74. Tasserie J, Uhrig L, Sitt JD, et al. Deep brain stimulation of the thalamus restores signatures of consciousness in a nonhuman primate model. *Sci Adv*. 2022;8:eable5547.
75. Schiff ND, Giacino JT, Butson CR, et al. Thalamic deep brain stimulation in traumatic brain injury: A phase 1, randomized feasibility study. *Nat Med*. 2023;29:3162-3174.
76. Coleman MR, Bekinschtein T, Monti MM, Owen AM, Pickard JD. A multimodal approach to the assessment of patients with disorders of consciousness. *Prog Brain Res*. 2009;177(C):231-248.
77. Sterling A, Bodien Y, Bergin M, et al. Validity of the telephone-administered coma recovery scale-revised and confusion assessment protocol for standardized remote assessment of persons with disorders of consciousness. *Arch Phys Med Rehabil*. 2024;105:e23-e24.