Puybasset *et al. Critical Care* (2026) 30:284
https://doi.org/10.1186/s13054-026-05909-x

Critical Care

RESEARCH Open Access

Deep white matter MRI predicts outcomes in coma of various etiologies: a cohort study

Louis Puybasset[1,2,15*], Pierre Siméone[2,3,4†], Martin Grange[2†], Didier Cassereau,[2], Damien Galanaud[2,5], Rémy Bernard[1], Lionel Velly[3,4], Vincent Perlbarg[6], Clément Capdeville[1], Valentine Battisti[1], Lamine Abdennour[1], Adrien Bouglé[7], Jean-Michel Constantin[8], Antoine Monsel[8,9], Juliette Chommeloux[10], Julien Mayaux[11,12], Benjamin Rohaut[13,14], Lionel Naccache[13,14], Alice Jacquens[1], Vincent Degos[1], Mélanie Pélégrini-Issac[2] on behalf of the MRI-COMA participants and investigators and CENTER-TBI principal investigators

Abstract

Background A reliable outcome prognostication tool for patients in coma of various etiologies would facilitate ICU treatment by providing objective information to caregivers and patients' relatives. This study aimed to predict outcome based on supervised machine learning and magnetic resonance diffusion tensor imaging (DTI) metrics.

Methods In this multicenter international study, a training set of 531 patients not responding to simple orders at day 5 after coma onset underwent diffusion-weighted MRI between day 5 and 45. A classifier was developed using DTI metrics, patient age, and delay between admission and MRI as features. Unfavorable outcome (UFO) was defined as GOSE 1–4 at one year. Three prognosis areas were defined: a "red" zone (specificity for UFO above 95%), a "green" zone (specificity for favorable outcome, FO, above 90%), and a "no determination zone" (NDZ) for patients classified in neither the red or green zone. The classifier was validated on an external test set of 211 patients.

Results The training set included 531 patients (age 48 ± 16 years; MRI at 19 ± 8 days post-injury), with 75.9% GOSE 1–4 and 24.1% GOSE 5–8 at one year. Normalized DTI metrics were FA 0.82 ± 0.12 and MD 1.10 ± 0.13. The external test set (n = 211; age 47 ± 16 years; MRI at 21 ± 12 days) showed similar outcome distribution (75.4% GOSE 1–4, 24.6% GOSE 5–8) and DTI values (FA 0.83 ± 0.09, MD 1.07 ± 0.12). Both sets were comparable in age, sex, initial GCS, and outcome ratios. In the external test set, ROC AUC was 0.89. For UFO classification, specificity was 98.1%, PPV 99.1%, and sensitivity 68.6%. For FO classification, specificity was 95.0%, PPV 77.8%, and NPV 86.3% whereas 30.8% of the patients were in the NDZ. After excluding patients for whom life sustaining therapies were withdrawn (n = 104), specificity was 96.6% and 82.4% for UFO and FO classification, respectively.

Conclusion This classifier demonstrates a high specificity to predict coma outcome while patients are still in the ICU, irrespective of coma etiology. These results may assist practitioners in making informed decisions.

Keywords Coma, Outcome, Prognosis, Diffusion tensor imaging, Deep white matter

†Pierre Siméone and Martin Grange contributed equally to this work.

*Correspondence:
Louis Puybasset
louis.puybasset@aphp.fr

Full list of author information is available at the end of the article

Background

Coma has multiple etiologies among which the most frequent, after exclusion of toxic or infectious causes, are cardiac arrest (CA), traumatic brain injury (TBI), aneurysmal subarachnoid hemorrhage (SAH), intraparenchymal hemorrhage (IPH) and hypoglycemia (HYP) [1]. Coma recovery may be complete, result in neurological deficits or leave the patient with a permanent disorder of consciousness. Disorders of consciousness encompass minimally conscious state[2] as well as vegetative state (also coined unresponsive wakefulness syndrome) [3, 4]. Regardless of cultural or cost issues, improved prognostication will always allow better titration of care, resource allocation and communication with relatives [5, 6].

The key challenge in a day-to-day practice is to develop tools based on personalized, reliable, and reproducible data, for predicting outcome with maximum specificity. The related goal is on the one hand, to avoid therapy withdrawal decisions in patients for whom there exists a potential for favorable recovery and, on the other hand, to allow dispassionate discussion and decisions in patients without any potential of recovery [5]. To date biological, electrophysiological, and imaging tools are available to clinicians to predict prognosis in comatose patients from various etiologies in the intensive care unit (ICU). Some of these biomarkers have been studied for prognostication across different etiologies of brain injury like in the present study. [5, 7] However, these tools and algorithms, although valuable, identifies, with a specificity sufficient to be clinically relevant, only a small proportion of patient eventually ending up with an unfavorable outcome [4, 8, 9] thereby highlighting the need for a more reliable approach to prognostication. During the last decade, diffusion tensor imaging (DTI) has emerged as a promising prognostic tool that outperforms conventional structural magnetic resonance imaging (MRI). [10, 11] This MRI technique quantifies the diffusion of water molecules in brain tissues and provides unique insights into the pathophysiology of white matter [12, 13].

According to the so-called “global neuronal workspace theory” (GNWT), deep white matter is a key anatomical substrate of consciousness [14]. GNWT emphasizes the role of long-distance connectivity and the integration of information across different brain regions. Given that all pathophysiological processes of coma from “structural” etiologies lead to deep white matter alteration [11, 15, 16], we thereby hypothesized that a global measurement of deep white matter integrity using DTI would be correlated with coma prognosis independently of etiology.

The goal of the present study was thus to develop a robust outcome classifier based on supervised machine learning using DTI metrics as well as basic clinical information as features, allowing accurate prognostication while the patient is still in the ICU.

Methods

Participants

All studies were approved by the local ethics committees in each participating country. The study adhered to the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cohort studies and Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)-AI guidelines for machine learning studies and was conducted in accordance with the Declaration of Helsinki (see Supplementary Appendix, Cohorts).

Training set

Patients from MRI-COMA and CENTER-TBI cohorts (see Supplementary Appendix, Cohorts and eTable 1) were included if they did not respond to simple orders, unresponsiveness not explained by sedation managed according to the international guidelines, between day 5 and day 45 after CA, TBI, SAH, IPH or HYP. Coma of toxic/inflammatory/infectious origins were excluded (Supplementary Appendix, eTable 2). SAH, IPH and HYP patients were pooled with already published patients with CA[7] and TBI[8] (Supplementary Appendix, Cohorts and eTable 1).

Independent external test set

Patients from COMABASE cohort (see Supplementary Appendix, Cohorts and eTable 1) enrolled within the 7 ICUs of la Pitié-Salpêtrière Hospital (Paris, France) from 2015 to 2023 were included in the test dataset with the same inclusion criteria as the patients of the training set.

Patient outcome

The primary study endpoint was 1-year neurological outcome assessed by the Glasgow Outcome Scale extended (GOSE). GOSE 1–4 was defined as an unfavorable outcome (UFO) and, accordingly, GOSE 5–8 as a favorable outcome (FO).

Healthy volunteers

At least five healthy volunteers were recruited in each center for calibration purposes (Supplementary Appendix, eTable 3). In a given center, the healthy volunteers underwent the same imaging protocol (coil, sequence parameters) as the patients.

Data

Data acquisition

Demographic and clinical information including initial severity data (Glasgow coma scale, GCS) were collected. MRI acquisitions were performed between day 5 and day 45 after the injury on 1.5 T or 3 T scanners from three manufacturers: GE Medical Systems (Milwaukee, WI), Siemens Medical Solutions (Erlangen, Germany), and

Philips Medical Systems (Eindhoven, The Netherlands) (Supplementary Appendix, eTable 3). All the clinical inclusion and non-inclusion criteria for MRI acquisition are summarized in Supplementary Appendix, eTable 4. In addition to a conventional three-dimensional structural T_1-weighted sequence, a diffusion-weighted imaging (DWI) sequence with a minimum of 29 diffusion gradient directions was acquired. For MRI-COMA and COMABASE cohorts, anonymized MR raw images were transferred to the coordinating center. CENTER-TBI data were collected through the Quesgen e-CRF (Quesgen Systems Inc, USA) hosted on the INCF platform and extracted via the INCF Neurobot tool (INCF, Stockholm, Sweden). Image data collection was facilitated by and hosted on the Icometrix platform (Icometrix, Leuven, Belgium). Version Center Core 2.1 of CENTER-TBI dataset was used in this study.

Diffusion-weighted data analysis

All DWI data were processed by using the brainQuant™ module of brainTale-care™ medical device (version 2.2.1) to extract calibrated fractional anisotropy (FA) and mean diffusivity (MD) parameters averaged within 19 regions of interest (ROIs) of the John Hopkins University deep white matter atlas (see Supplementary Appendix, MR diffusion-weighted data analysis) [17]. Briefly, DWI data first underwent a systematic sanity and quality check. Then, they were corrected from between-volume motion by performing linear registration of each acquired volume to the reference b_0 volume (i.e., with no diffusion gradient applied) [18]. After voxel-wise estimation of the diffusion tensor, resulting FA and MD maps were co-registered onto the Montreal Neurological Institute (MNI) standard space and metrics values were averaged across all deep white-matter ROIs. These average values were finally normalized with respect to similar measures obtained from the healthy volunteers to reduce between-center variability [19]. The data of the already published patients with CA [10] and TBI [11] were entirely reprocessed using this procedure, which led to the exclusion of some of these patients due to quality issues.

Descriptive statistics

Means and standard deviations were used for continuous variables (age, delay from injury to the MRI exam, length of stay in ICU, FA and MD); medians and interquartile ranges were used for GCS. Group differences between training and test datasets were assessed with unpaired two-sample T-tests, Fisher's exact test, Wilcoxon rank sum test, and MANOVA (Pillai's trace statistic) where appropriate ($p < 0.05$ considered as statistically significant). All analyses were performed using R version 4.3.3 [20] and RStudio [21].

Machine learning prognostication framework

The patients were characterized by four equally weighted input features: age (in years), time from injury to the MRI exam (in days), FA and MD (normalized average values). The prognostication procedure relied on a supervised machine learning approach involving a support vector classifier in a repeated nested cross-validation (rnCV) framework, implemented using *scikit-learn* v1.0.2 [22] for Python 3.9 (see details in Supplementary appendix, rnCV framework; eFigure 1–3). Depending on their resulting prediction score, the patients were prognosticated either to have a UFO with more than 95% specificity ("red" zone), or to have a FO with more than 90% specificity ("green" zone) or remained in a "no determination zone" (NDZ) if they could not be classified in the "red" or the "green" zone. A lower specificity was chosen for FO than for UFO considering a higher false positive rate more acceptable, from the clinical point of view, for the former category than for the latter. The whole rnCV procedure eventually yielded a total of 500 prediction models, which were subsequently tested on the independent external test dataset. Shapley additive explanations (SHAP) values were calculated to interpret the predictions made by the rnCV model and plotted according to global feature importance for the training and the external test sets, respectively. [23] Based on game theory, SHAP is a unified approach that aims to explain the output of any machine learning model by measuring the contribution of each feature to the prediction. [24] Some features have a large impact on the prediction, yielding large absolute SHAP values, whereas others have a small impact, yielding small absolute SHAP values. Feature impact is defined as the change in the expected value of the model's output when a feature is observed versus unknown. The results are often displayed in a summary plot as a beeswarm representation that combines feature importance with feature effects.

Regarding the external test data set, a final decision was made to assign a given patient to the "red" zone (respectively, the "green" zone) if 99.7% of the 500 models (i.e., empirically within 3 standard deviations of the mean prediction score over all models) classified the patient as having a UFO (or respectively, a FO) and to assign the patient to the NDZ, otherwise.

Results

Overall, 1,041 patients met the inclusion criteria during the enrolment period. Of these, 262 were excluded from the analysis following MRI data sanity and quality check or due to MRI preprocessing issues (N = 201 and 61, respectively), and 37 were lost to follow-up (Fig. 1). Finally, 742 patients were included in the analysis, of whom 502 (67.7%) were scored GOSE 1–3; 60 (8.1%) had GOSE 4; and 180 (24.2%) had GOSE 5–8; (Table 1). A

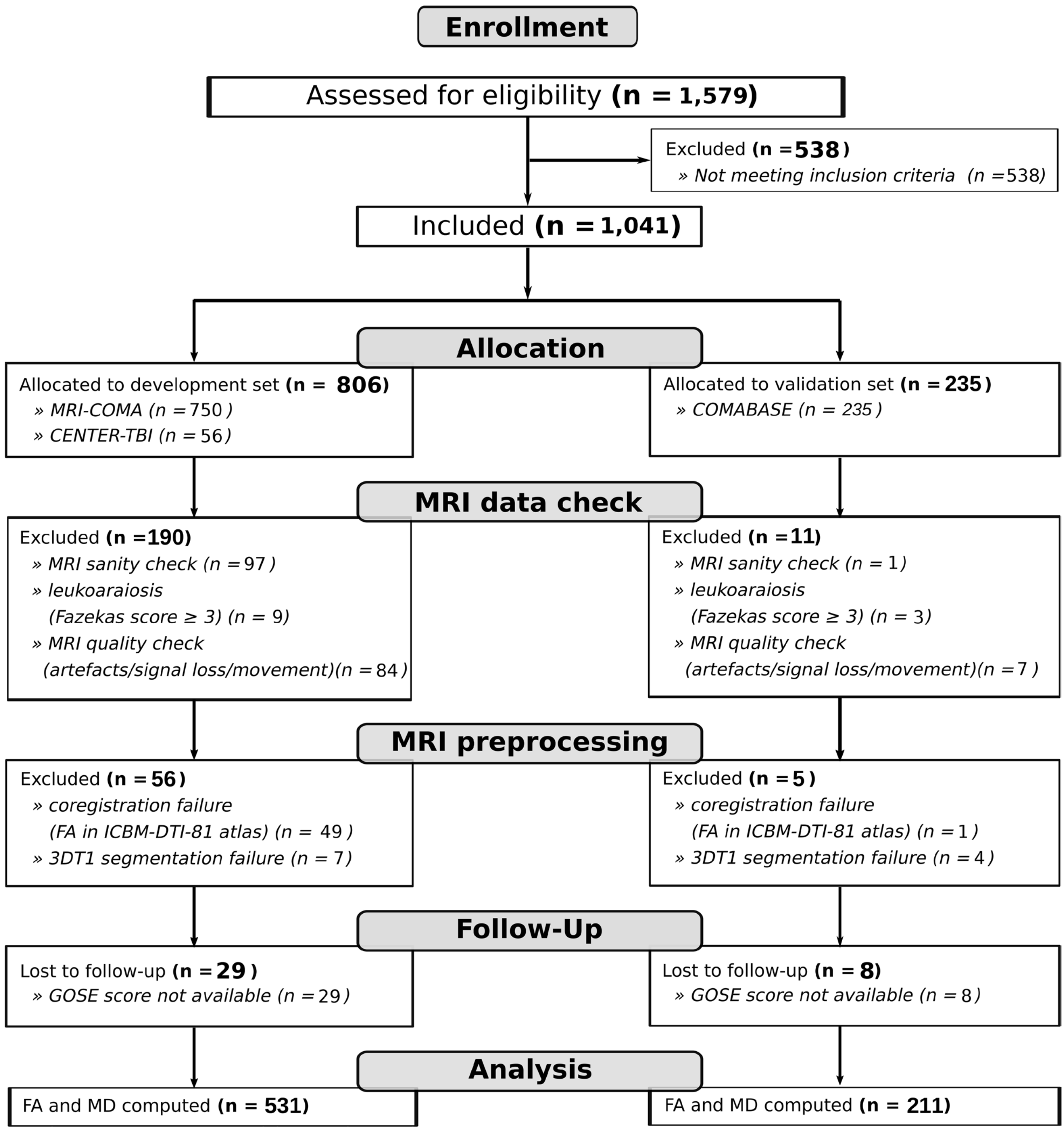

Fig. 1 Flow chart of the study

total of 215 healthy controls were included for MRI calibration (eTable 3).

Training set

A total of 531 patients with a mean age of 48 ± 16 years were included in the training set (Table 1). MRI was performed 19 ± 8 days post-injury in 170 (32.0%) CA, 245 (46.1%) TBI, 55 (10.4%) SAH, 51 (9.6%) IPH and 10 (1.9%) HYP patients. Of these patients, 403 were GOSE 1–4 (75.9%) and 128 patients were GOSE 5–8 at one year (24.1%). Overall, DTI metrics revealed a normalized FA of 0.82 ± 0.12 and a normalized MD of 1.10 ± 0.13. Univariate logistic regression results obtained to classify patients with UFO or FO are given in Supplementary Appendix, eTable 5. The rnCV procedure showed a high prognostic performance to classify UFO patients (GOSE 1–4) with an area under the receiver operating characteristic curve (AUC_{ROC}) of 0.930 (95% confidence interval CI, 0.927 to 0.933). In the "red" zone, sensitivity was 73.0% (95% CI, 72.5 to 73.6), with a specificity set at 95%

Table 1 Main clinical characteristics of training set and external test set patients

Training set	All	GOSE 1–3	GOSE 4	GOSE 5–6	GOSE 7–8
N	531	355	48	81	47
Sex (M/F)	365/166	238/117	34/14	61/20	32/15
Mean age (SD)[a]	47.6 (16.0)	50.8 (15.4)	45.6 (14.8)	38.8 (13.8)	40.6 (17.2)
Mean MRI delay (SD)[b]	19 (8)	19 (8)	21 (8)	18 (8)	17 (6)
Mean LOS in ICU (SD)[c]	36 (29)	37 (30)	48 (28)	32 (23)	18 (17)
Median GCS (IQR)[d]	4 (4)	3 (3)	6 (4)	4 (6)	8 (11)
Mean FA (SD)	0.819 (0.117)	0.773 (0.111)	0.875 (0.056)	0.916 (0.049)	0.943 (0.056)
Mean MD (SD)	1.100 (0.132)	1.116 (0.155)	1.068 (0.067)	1.070 (0.048)	1.053 (0.048)
External test set	**All**	**GOSE 1–3**	**GOSE 4**	**GOSE 5–6**	**GOSE 7–8**
N	211	147	12	33	19
Sex (M/F)	145/66	98/49	8/4	24/9	15/4
Mean age (SD)[a]	47.1 (16.5)	50.7 (15.7)	49.0 (16.1)	38.5 (15.0)	33.8 (13.2)
Mean MRI delay (SD)[b]	21 (12)	19 (11)	31 (8)	21 (11)	30 (13)
Mean LOS in ICU (SD)[e]	31 (22)	30 (24)	46 (25)	31 (15)	30 (14)
Median GCS (IQR)[f]	4 (4)	3 (3)	4 (3)	7 (7)	7 (6)
Mean FA (SD)	0.826 (0.088)	0.800 (0.089)	0.853 (0.049)	0.884 (0.047)	0.913 (0.038)
Mean MD (SD)	1.071 (0.121)	1.070 (0.142)	1.101 (0.060)	1.076 (0.043)	1.055 (0.049)

N: number; M: male; F: female; GOSE: Glasgow Outcome Scale extended; IQR: interquartile range; SD: standard deviation; LOS: length of stay in ICU (days); ICU: intensive care unit; GCS: Glasgow Coma scale

[a] *Age is at the time of injury. Age is expressed in years*

[b] *Delay is expressed in days post-injury*

[c] *Data were unavailable for 43 patients*

[d] *Data were unavailable for 10 patients*

[e] *Data were unavailable for 2 patients*

[f] *Data were unavailable for 11 patients*

for UFO prediction. In the "green" zone, sensitivity was 77.8% (95% CI, 76.8 to 78.8) and specificity was 89.5% (95% CI, 89.0 to 89.9) for FO prediction (see Supplementary Appendix, eTable 6). SHAP values are shown in Fig. 2a.

External test set

The 211 patients had a mean age of 47 ± 16 years and underwent an MRI session 21 ± 12 days post-injury. There were 91 (43.1%) CA, 91 (43.1%) TBI, 10 (4.7%) SAH, 17 (8.1%) IPH and 2 (1.0%) HYP patients. Overall, DTI metrics revealed a normalized FA of 0.83 ± 0.09 and a normalized MD of 1.07 ± 0.12. Of these patients, 159 were GOSE 1–4 (75.4%) and 52 patients were GOSE 5–8 (24.6%) at one year (Table 1). The test set did not significantly differ from the training test in terms of age (two-sample T-test, $t = -0.32$, $p = 0.75$), sex ratio (Fisher's exact test, $p = 1$), initial GCS (Wilcoxon's $W = 49.376$, $p = 0.25$), and UFO/FO ratio (Fisher's exact test, $p = 0.92$). The mean delay between injury and MRI session was significantly higher in the external test set than in the training set (two-sample T-test, $t = 2.14$, $p = 0.03$), as well as the mean length of stay in ICU (two-sample T-test, $t = 2.19$, $p = 0.03$). Normalized FA and MD were negatively correlated with one another in both sets ($p < 0.001$); MANOVA showed a significant multivariate effect (Pillai's statistic = 0.01, $p = 0.02$) and univariate post-hoc test showed that MD was significantly lower in the external test set than in the training set (Welch's one-way ANOVA statistic = 7.54, $p = 0.006$).

Patients were classified as follows when considering 99.7% of the 500 trained models: "red" zone (n = 110, 52.1%), NDZ (n = 65, 30.8%), "green" zone (n = 36, 17.1%) (Fig. 3). UFO prediction had a positive predictive value (PPV) of 99.1% with a specificity of 98.1% and a sensitivity of 68.6%. FO prediction had a PPV of 77.8% and a negative predictive value (NPV) of 86.3% with a specificity of 95.0% (see Supplementary Appendix, eTable 7 and 8). SHAP values are shown in Fig. 2b.

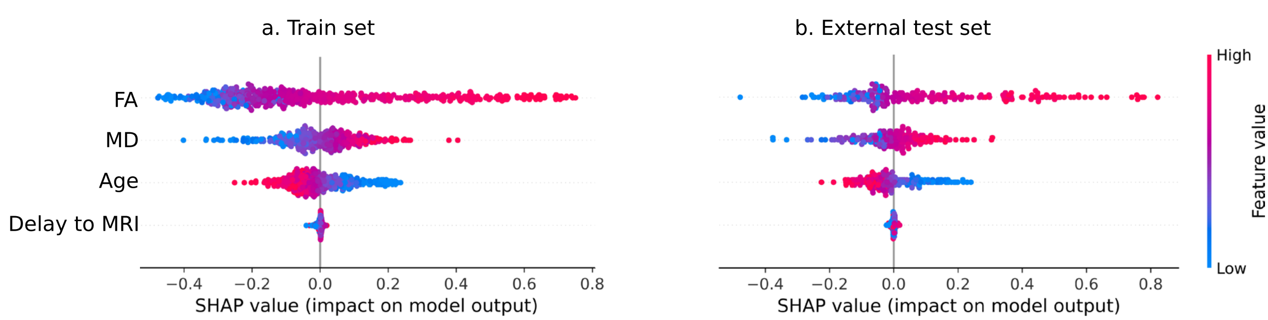

Fig. 2 Beeswarm summary plots of Shapley additive explanations (SHAP) values for the training (**a**) and external test (**b**) sets. The SHAP explanation is represented by a single dot on each feature row for each patient. The SHAP value of each feature determines the position of the dot on the X-axis. The lower the SHAP value in the negative X direction, the higher the impact of the feature on UFO prediction. The higher the SHAP value in the positive X direction, the higher the impact of the feature on FO prediction. The colorscale displays the original value of the feature. For instance, high FA values have high impact on FO prediction; high age values have more impact on UFO prediction than low age values

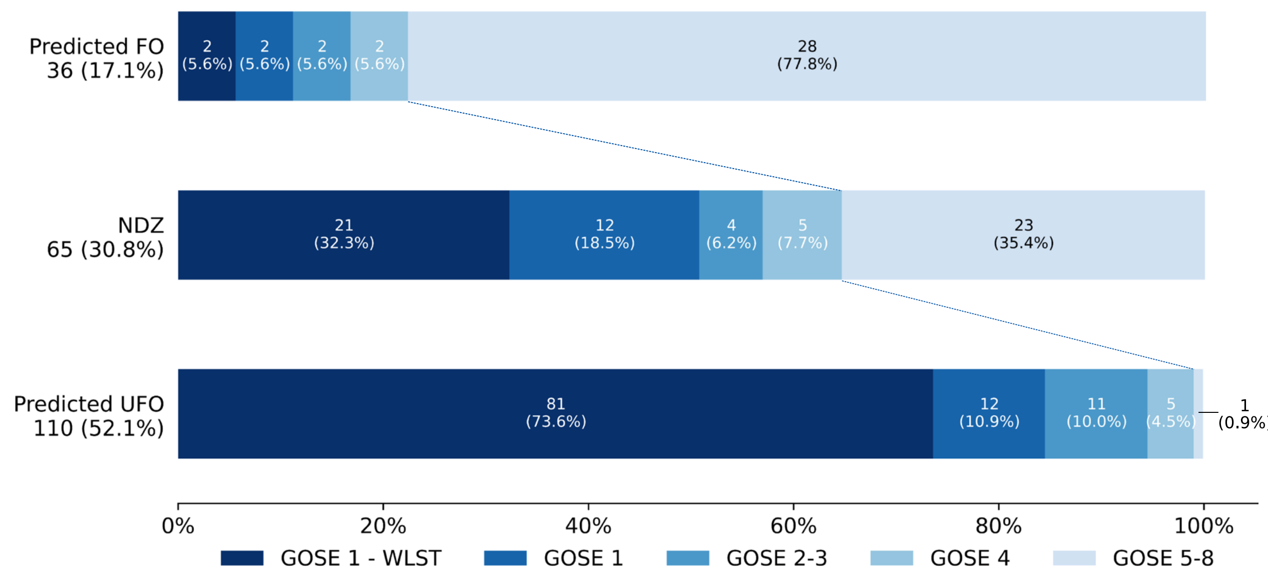

Fig. 3 Distribution of Glasgow Outcome Scale Extended (GOSE) scores for patients predicted by the classifier with favorable outcome (FO), with unfavorable outcome (UFO), and patients in the "no determination zone" (NDZ). Bar percentages were rounded to 1 decimal place and may not add up to 100%

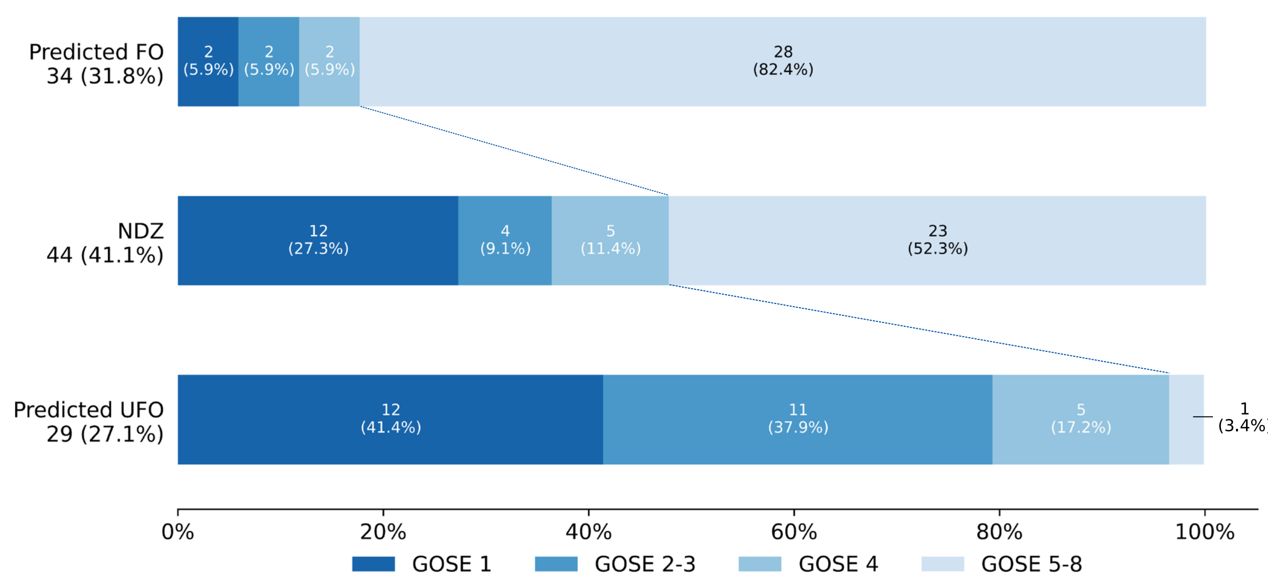

Fig. 4 Distribution of Glasgow Outcome Scale Extended (GOSE) scores for 107 patients predicted by the classifier with favorable outcome (FO), with unfavorable outcome (UFO), and patients in the "no determination zone" (NDZ), after exclusion of the patients concerned by withdrawal of life support therapy (WLST). Bar percentages were rounded to 1 decimal place and may not add up to 100%

One hundred and four patients died in the ICU from withdrawal of life-sustaining therapy (WLST, data available in the Supplementary Appendix, eTable 9). From the 107 remaining patients, 28 of 29 patients (96.6%) predicted in the "red" zone ultimately had UFO among whom 12 patients died in the year following ICU discharge; 28 of 34 patients (82.4%) predicted in the "green" zone ultimately had FO. In the NDZ, 23 of 44 patients (52.3%) had FO and 21 of 44 patients (47.7%) had UFO (Fig. 4).

Discussion

The present study focused on comatose patients in the ICU who did not regain consciousness by day 5 after injury. Our procedure classified 70% of the patients of the independent external test set in one outcome category, with a specificity above 95% for UFO (GOSE 1–4) and FO (GOSE 5–8). We believe that such a prediction system based on minimal data derived from DTI and basic clinical data adds highly relevant information to the individual decision-making process irrespective of coma etiology. This is especially true in those patients for whom neuroprognostication is challenging, i.e., without typical favorable or unfavorable evolution during the first days after onset, with a high dependency on care.

The classifier that we used for classification is a non-linear combination of two diffusion metrics quantifying global microstructural changes in deep white matter, FA and MD, enriched with patient's age and MRI delay between injury and MRI session. DTI metrics reflect deep white matter integrity and, thereby, potential damage to brain networks. This work can thus be embedded within the GNWT framework, which posits that conscious awareness arises when information being processed in specialized, unconscious modules is broadcast to a wider network of neurons, known as the "global neuronal workspace" [25]. GNWT has already been applied in various clinical settings [26, 27], including anesthesia [28]. Although this theory offers a useful framework for interpreting these findings, other theories and mechanisms may also play significant roles in consciousness and coma recovery [29]. This work provides a basis for a better understanding and management of disorders of consciousness and promotes the use of DTI metrics to assess white matter integrity within large bundles in such patients.

Regarding the clinical value of DTI metrics, it has been shown that a decrease in FA can be induced by different pathophysiological processes, including myelin damage that occurs after traumatic brain injury, and axonal degeneration observed after ischemic anoxic brain injury. In contrast, FA can be normal in the presence of cytotoxic brain swelling, as observed two weeks after cardiac arrest, and associated with a decrease in MD [13]. The classifier takes these various physiological changes into account, yielding overall predictive performances for 1-year UFO prediction and 1-year FO prediction that are compatible with reliable and personalized clinical use.

The rationale for setting a higher specificity threshold for UFO (≥ 95%) than for FO (≥ 90%) reflects a deliberate clinical trade-off. In coma prognostication, falsely predicting UFO might result in premature WLST decisions. The specificity threshold set at 95% therefore prioritizes minimizing false positives for UFO, ensuring that patients with no realistic chance of recovery are reliably identified, while accepting a slightly higher error rate for FO predictions. This approach aligns with the ethical imperative to avoid inappropriate limitation of care in patients who may still recover.

The decision to rely on minimal clinical inputs (specifically age, injury-to-MRI delay, and DTI metrics) represents a deliberate balance between feasibility and prognostic accuracy. By limiting features to robust, objective measures, we ensure reproducibility and minimize bias introduced by subjective or inconsistently documented clinical variables. However, this parsimonious approach may overlook potentially informative prognostic factors, such as detailed neurological assessment, biomarkers, or electrophysiological data, which could refine individual predictions. While this trade-off ensures broad applicability and reduces confounding, it also underscores the need for future integration of multimodal data to capture the full complexity of coma recovery trajectories particularly for patients in NDZ. The current model thus serves as a standardized foundation, but its clinical utility could be further strengthened by complementary assessments in challenging cases.

It is noteworthy that at the time when the external test patients were included, DTI metrics and outcome prediction for CA and TBI patients were available to clinicians through the brainTale-care™ interface. To avoid the effect of a possible self-fulfilling prophecy, we analyzed the performance of the prediction score without patients who died following WLST and observed conserved specificity in the predetermined targets, at the expense of an NDZ increasing from 30.8% up to 41.1% of the patients. While this strategy may reduce negative bias in patients predicted to have UFO, we cannot exclude that access to prognostic scores might have biased results toward extended care in patients predicted to have FO.

The prognosis tool that we propose gives reliable predictive scores based on robust MRI biomarkers. However, several drawbacks may slow adoption of the technique in the clinical context. First, only patients with a coma due to one of the five etiologies listed can be tested. This prognostic system is thus not usable in coma from infectious, toxic, encephalitis or fat embolism origin. The two main reasons for this choice were (1) the low incidence of coma complicating these pathologies and (2) the difficulties in defining a precise date for disease onset. Second, the periodic need for upgrade of MRI machines, together with the difficulties associated with scanning healthy subjects using clinical equipment, makes the procedure challenging insofar as calibration is required. A third issue is the possible existence of MRI contraindications and factors that can affect the reliability of DTI metrics. Another limitation stems from the exclusion of 262 patients (25.2%) due to MRI data sanity or preprocessing failures, primarily related to motion artifacts, protocol deviations, or incidental neurological abnormalities. It is therefore important to select patients in order to avoid significant contraindications for DTI interpretation, but it is also important to note that, as with any technical implementation, exclusions due to technical errors or patient movement decrease over time as teams gain experience. These factors include a history of neurological disorders, leukoaraiosis, hydrocephaly, and acquisition artifacts, which are known to impact accuracy of DTI measurements. Fourth, around 30% of patients in the external test set were unable to be prognosticated at the time of MRI acquisition due to their scores falling within the so-called "no determination zone". There are several possible approaches to address the uncertainty regarding the outcome of these patients. On the one hand, the additive predictive value of a rescan one or two weeks after the initial MRI session may provide information regarding dynamic changes in diffusion metrics. On the other hand, as for every patient, a broader approach consists in integrating quantified DTI metrics within a multimodal assessment [30] combining clinical markers, advanced quantified electrophysiology [31, 32], and if available, brain metabolic MRI [33, 34] or positron emission tomography [35].

While the proposed pipeline relies on standardized preprocessing and normalization using healthy volunteer scans, this approach ensures robust, reproducible, and high-quality DTI metrics across centers. This rigorous calibration enhances the reliability and generalizability of prognostic predictions, ultimately supporting more accurate and personalized clinical decision-making. The logistical challenges of data transfer, while notable, can be addressed through secure, compliant platforms—highlighting the importance of infrastructure investment to integrate advanced neuroimaging into routine.

Conclusions

We believe that the proposed classifier may serve as a valuable aid within a multimodal approach to outcome prediction in coma patients. It offers a unified tool to help evaluate coma from various etiologies and could contribute to better assess prognostic and determine therapeutic strategies. Further validation and clinical studies are necessary to fully establish its efficacy and reliability in diverse clinical settings.

Abbreviations

MRI	Magnetic resonance imaging
DTI	Diffusion tensor imaging
GOSE	Glasgow Outcome Scale extended
UFO	Unfavorable outcome
FO	Favorable outcome
FA	Fractional anisotropy
MD	Mean diffusivity
GCS	Glasgow Coma Scale
ICU	Intensive care unit
NDZ	No determination zone
WLST	Withdrawal of life-sustaining therapies
PPV	Positive predictive value
NPV	Negative predictive value
ROC AUC	Receiver operating characteristic area under the curve
SHAP	Shapley additive explanations
CA	Cardiac arrest
TBI	Traumatic brain injury
SAH	Subarachnoid hemorrhage
IPH	Intraparenchymal hemorrhage
HYP	Hypoglycemia
GNWT	Global neuronal workspace theory
rnCV	Repeated nested cross-validation
ROI	Region of interest

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13054-026-05909-x.

Supplementary Material 1: Supplementary Appendix

Supplementary Material 2: Institutional groups

Acknowledgements

We would like to thank patients, relatives, and clinical and research staff in all trial sites.

Author contributions

LP and MPI contributed equally to the conception and design of the study. MG and MPI developed the statistical method used in this study. LP, PS, MG, DC, DG, LV, VP, CC, VB, LA, AB, JMC, AM, JC, JM, BR, LN, AJ, and MPI contributed to data acquisition, analysis, and interpretation. LP, PS, and MPI drafted the manuscript. All authors critically revised the manuscript and approved the final version.

Funding

Funding The MRI-COMA trial was funded by independent research grants from non-profit or governmental agencies: French Ministry of Health, Paris, France (Programme Hospitalier de Recherche Clinique 2005 #051061), and the French National Agency for Research (ANR) for the program "Investissements d'avenir" ANR-10-IAIHU-06 (to the Brain and Spine Institute); Italian Ministry of health and Regione Lombardia (Ricerca Finalizzata 2010—RF-2010–2319503). PS received an academic mobility grant from "Bourse à la mobilité de SFAR—2024" and "Bourse à la mobilité de Phoceo—2024". The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data availability

AP-HP, as manager, is the owner of the data and no use or disclosure to third parties may be made without its prior consent. The data that support the findings of this study are not publicly available due to privacy or ethical restrictions. They may be obtained upon request to the corresponding author, subject to approval from Assistance Publique – Hôpitaux de Paris. The Python code of the rnCV procedure is available for reviewers in Zenodo.org at https://tinyurl.com/3a234swt.

Declarations

Ethics approval and consent to participate

This study was approved by the local ethics committees and adhered to the Reporting of Observational Studies in Epidemiology (STROBE) guidelines and the Declaration of Helsinki. Written informed consent was obtained from all participants or their legal representatives.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]APHP, Hôpital Pitié-Salpêtrière, Neurosurgical Intensive Care Unit, Sorbonne Université, Paris, France
[2]CNRS, INSERM, Laboratoire d'Imagerie Biomédicale (LIB), Sorbonne Université, Paris, France
[3]Department of Anesthesiology and Intensive Care, APHM, Aix-Marseille Université, CHU Timone, Marseille, France
[4]CNRS, Inst Neurosci Timone, Aix Marseille Université, Marseille, France
[5]Department of Neuroradiology, APHP, Sorbonne Université, Hôpital Pitié-Salpêtrière, Paris, France
[6]BrainTale SAS, Strasbourg, France
[7]Department of Cardiovascular and Thoracic Surgery, APHP, Sorbonne Université, Hôpital Pitié-Salpêtrière, Institute of Cardiology, Paris, France

[8]Department of Anaesthesiology and Critical Care, APHP, Sorbonne Université, Hôpital Pitié-Salpêtrière, GRC 29, DMU DREAM, Paris, France
[9]APHP, INSERM, Immunology-Immunopathology-Immunotherapy, UMR-S 959, Hôpital Pitié-Salpêtrière, Paris, France
[10]APHP, Hôpital Pitié-Salpêtrière, Institut de Cardiologie, Medical Intensive Care Unit, Sorbonne Université, Paris, France
[11]APHP, Hôpital Pitié-Salpêtrière, Médecine Intensive Réanimation (Département "R3S"), Sorbonne Université, Paris, France
[12]INSERM, UMR-S1158, Neurophysiologie Respiratoire Expérimentale Et Clinique, Sorbonne Université, Paris, France
[13]INSERM, CNRS, Institut du Cerveau–Paris Brain Institute–ICM, Sorbonne Université, Paris, France
[14]APHP, DMU Neurosciences, Hôpital Pitié-Salpêtrière, Paris, France
[15]Department of Anesthesiology and Intensive Care, Groupe Hospitalier Pitié–Salpêtrière, 47-83 Boulevard de l'Hôpital, Paris 75013, France

Received: 9 December 2025 / Accepted: 14 February 2026
Published online: 04 March 2026

References

1. Edlow JA, Rabinstein A, Traub SJ, Wijdicks EFM. Diagnosis of reversible causes of coma. Lancet. 2014;384:2064–76. https://doi.org/10.1016/S0140-6736(13)62184-4.
2. Giacino JT, Ashwal S, Childs N, et al. The minimally conscious state: definition and diagnostic criteria. Neurology. 2002;58:349–53.
3. Laureys S, Celesia GG, Cohadon F, et al. Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome. BMC Med. 2010;8:1–4.
4. Sandroni C, Cariou A, Cavallaro F, et al. Prognostication in comatose survivors of cardiac arrest: an advisory statement from the European Resuscitation Council and the European Society of Intensive Care Medicine. Intensive Care Med. 2014;40:1816–31. https://doi.org/10.1007/s00134-014-3470-x.
5. Fischer D, Edlow BL. Coma prognostication after acute brain injury: a review. JAMA Neurol. 2024. https://doi.org/10.1001/jamaneurol.2023.5634.
6. Herridge MS, Azoulay É. Outcomes after critical illness. N Engl J Med. 2023;388:913–24.
7. Bodien YG, Busl KM, Chang CWJ, et al. Disorders of consciousness diagnosis, interventions, and prognostication for the intensivist: report of the 2025 ISICEM roundtable. Intensive Care Med. 2025. https://doi.org/10.1007/s00134-025-08224-1.
8. Sondag L, Ruijter BJ, Tjepkema-Cloostermans MC, et al. Early EEG for outcome prediction of postanoxic coma: prospective cohort study with cost-minimization analysis. Crit Care. 2017;21:1–8.
9. Rossetti AO, Oddo M, Logroscino G, Kaplan PW. Prognostication after cardiac arrest and hypothermia: a prospective study. Ann Neurol. 2010;67:301–7.
10. 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;17:317–26. https://doi.org/10.1016/S1474-4422(18)30027-9.
11. 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;48:201–12. https://doi.org/10.1007/s00134-021-06583-z.
12. Alexander AL, Lee JE, Lazar M, Field AS. Diffusion tensor imaging of the brain. Neurotherapeutics. 2007;4:316–29. https://doi.org/10.1016/j.nurt.2007.05.011.
13. Sotak CH. The role of diffusion tensor imaging in the evaluation of ischemic brain injury–a review. NMR Biomed Int J Devoted Dev Appl Magn Reson Vivo. 2002;15:561–9.
14. Dehaene S, Kerszberg M, Changeux J-P. A neuronal model of a global workspace in effortful cognitive tasks. Proc Natl Acad Sci. 1998;95:14529–34. https://doi.org/10.1073/pnas.95.24.14529.
15. Baars BJ. The global workspace theory of consciousness. Blackwell Companion Conscious. 2007 236–246
16. Simeone P, Auzias G, Lefevre J, et al. Long-term follow-up of neurodegenerative phenomenon in severe traumatic brain injury using MRI. Ann Phys Rehabil Med. 2021;65:101599.
17. Van Der Eerden AW, Khalilzadeh O, Perlbarg V, et al. White matter changes in comatose survivors of anoxic ischemic encephalopathy and traumatic brain injury: comparative diffusion-tensor imaging study. Radiology. 2014;270:506–16.
18. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Med Image Anal. 2001;5:143–56. https://doi.org/10.1016/s1361-8415(01)00036-6.
19. Cai LY, Yang Q, Kanakaraj P, et al. MASiVar: multisite, multiscanner, and multisubject acquisitions for studying variability in diffusion weighted MRI. Magn Reson Med. 2021;86:3304–20.
20. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Httpwww R-Proj Org. 2016
21. Team P. RStudio: integrated development environment for R. Posit Software, PBC, Boston, MA. 2023
22. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825–30.
23. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017 30
24. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng. 2018;2:749–60.
25. Sergent C, Dehaene S. Neural processes underlying conscious perception: experimental findings and a global neuronal workspace framework. J Physiol-Paris. 2004;98:374–84.
26. Dehaene S, Changeux J-P, Naccache L. The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications. In: Dehaene S, Christen Y, editors. Characterizing Consciousness: From Cognition to the Clinic? Berlin Heidelberg, Berlin, Heidelberg: Springer; 2011. p. 55–84.
27. Bartolomei F, Naccache L. The global workspace (GW) theory of consciousness and epilepsy. Behav Neurol. 2011;24:67–74. https://doi.org/10.3233/BEN-2011-0313.
28. Mashour GA, Roelfsema P, Changeux J-P, Dehaene S. Conscious processing and the global neuronal workspace hypothesis. Neuron. 2020;105:776–98. https://doi.org/10.1016/j.neuron.2020.01.026.
29. Fischer D, Edlow BL, Giacino JT, Greer DM. Neuroprognostication: a conceptual framework. Nat Rev Neurol. 2022;18:419–27. https://doi.org/10.1038/s41582-022-00644-7.
30. 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. https://doi.org/10.1038/s41591-024-03019-1.
31. Sitt JD, King J-R, El Karoui I, et al. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. Brain. 2014;137:2258–70.
32. Claassen J, Doyle K, Matory A, et al. Detection of brain activation in unresponsive patients with acute brain injury. N Engl J Med. 2019;380:2497–505. https://doi.org/10.1056/NEJMoa1812757.
33. Tollard E, Galanaud D, Perlbarg V, et al. Experience of diffusion tensor imaging and ^{1}H spectroscopy for outcome prediction in severe traumatic brain injury: preliminary results. Crit Care Med. 2009;37:1448–55. https://doi.org/10.1097/CCM.0b013e31819cf050.
34. Quintard H, Velly L, Boussen S, et al. Value of assessment of multivoxel proton chemical shift imaging to predict long term outcome in patients after out-of-hospital cardiac arrest: a preliminary prospective observational study. Resuscitation. 2019;134:136–44. https://doi.org/10.1016/j.resuscitation.2018.09.007.
35. Stender J, Gosseries O, Bruno M-A, et al. Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study.

Publisher's Note