Metabolic index of the best preserved hemisphere of 18F-FDG PET imaging combination behavioral CRS-R scores predict recovery from disorders of consciousness

Kun Guo

1512649468@qq.com

Department of Nuclear Medicine, Xijing Hospital https://orcid.org/0000-0002-9728-2766

Guiyu Li

Xijing Hospital, Department of Nuclear Medicine

Zhiyong Quan

Xijing Hospital Department of Nuclear Medicine

Yirong Wang

Xijing Hospital, Department of Nuclear Medicine

Junling Wang

Xijing Hospital, Department of Nuclear Medicine

Fei Kang

Xijing Hospital, Department of Nuclear Medicine

Jing Wang

Xijing Hospital, Department of Nuclear Medicine

Research Article

Keywords: Consciousness Disorders, Positron Emission Tomography, Fluorodeoxyglucose F18, Neuroimaging, Deep learning

Posted Date: February 27th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3849572/v1

Abstract

Background The objective of this study was to develop a prognostic model for predicting one-year post-injury outcomes in chronic disorders of consciousness (DoC) by detecting relatively preserved brain metabolism through ^{18}F-fluorodeoxyglucose positron emission tomography (^{18}F-FDG PET). This approach allows the assessment of the level of consciousness and the prediction of the likelihood of wakefulness.

Methods Eighty-seven DoC patients newly diagnosed with behavioral Revised Coma Recovery Scale (CRS-R) and ^{18}F-FDG PET/CT studies were included. PET images were standardized by the metabolic index of the best-preserved hemisphere (MIBH) and the ratio SUV (SUVR), respectively. The training of image-based classification was conducted using the DenseNet121 network, while tabular-based deep learning was employed for training depth features extracted from imaging models and behavioral CRS-R scores. The performance of the models was assessed using the area under the curve (AUC).

Results Of the 87 DoC patients who received routine treatments, consciousness recovery was observed in 52 patients, while consciousness non-recovery was observed in 35 patients. The classification performance of the MIBH model was found to be superior to that of the SUVR model, with AUC values of 0.751 ± 0.093 and 0.412 ± 0.104 on the test sets, respectively. The MIBH + CT multimodal model was determined to perform better than the MIBH-only model, achieving an AUC of 0.784 ± 0.073 on the test sets. The combination of MIBH + CT depth features with behavioral CRS-R scores resulted in the best classification accuracy, with AUC values of 0.950 ± 0.027 and 0.933 ± 0.015 on the training and test sets, respectively.

Conclusions The prediction of recovery in DoCs was facilitated by a model based on a combination of multimodal imaging features and behavioral CRS-R scores.

Introduction

Severe brain injury causes chronic disorders of consciousness (DoC) [1], primarily including unresponsive awakening syndrome (UWS) and minimally conscious state (MCS). In China, there are over 500,000 patients with chronic DoC, witnessing an annual increase of 70,000-100,000 [2]. These patients endure bedridden conditions, lack functional communication, exhibit no purposeful behavior, and are entirely reliant on others for care. The medical community regularly encounters the challenge of providing prognostic guidance to families of DoC patients. Despite limited treatment options, recent research indicates potential benefits for certain chronic DoC cases through awakening interventions [3–5]. Moreover, determining a patient's suitability for recent pilot studies and ensuring accurate prognostication has become an essential step.

Historically, the prognostication for DoC patients has relied on meticulous and repeated behavioral assessments over a sufficient period. However, these assessments are unavoidably subjective and

susceptible to various personal biases [6]. Motor impairments, sensory deficits, cognitive damage, fluctuations in vigilance, and medical complications may contribute to misjudgments. There is an urgent need for accurate and objective biomarkers in clinical practice. Some clinical studies have highlighted the significance of etiology, incidence age, and duration of DoC as crucial indicators for prognosis, resulting in the development of prognostic models predicting conscious recovery [7]. Over the past decades, pilot prognostic models based on neurological examination features [8], electroencephalogram abnormalities [9–10], and anatomical and functional changes observed in neuroimaging have been also explored [11–13]. Despite considerable efforts, any singular method carries the risk of false predictions, making the identification of efficient biomarkers for outcome prediction still challenging.

^{18}F-fluorodeoxyglucose positron emission tomography (^{18}F-FDG PET) facilitates the visualization of brain glucose metabolism, where preserved metabolism aligns with specific behavioral or perceptual functions. The ratio standard uptake value (SUVR), typically standardized by whole brain or cerebellar metabolism, quantifies glucose metabolism in key networks of DOC patients. Stender et al. demonstrated that preserved metabolism in the frontoparietal networks predicts clinical outcomes one year later in 74% of patients [14]. However, SUVR may be overestimated in patients with extensive injuries. Stender et al. introduced the metabolic index of the best-preserved hemisphere (MIBH), considering evidence that conscious awareness can be maintained with only one hemisphere. This index scales metabolic activity by setting the mean activity of extracerebral regions. The metabolic criterion accurately predicted 88% of all known patient outcomes [15]. Nevertheless, existing studies primarily focus on classifying healthy and pathological groups, and to date, the metabolic network pattern allowing single patient prognosis prediction remains incompletely established.

In this study, our initial objective was to assess the predictive efficacy of two distinct PET image standardization methods, MIBH and SUVR. Additionally, given the progress in medical science, the prognostication of patients with chronic DoC has evolved towards a multidomain paradigm, the second aim was to develop an approach for forecasting the individual prognosis of DoC patients by incorporating superior standardization techniques for ^{18}F-FDG PET and behavioral scores.

Materials and methods

2.1 Subjects

We conducted a retrospective review of a dataset comprising 114 patients recently diagnosed with chronic DoC who underwent ^{18}F-FDG PET/CT scans between January 2013 and December 2021. The level of consciousness was determined through a comprehensive assessment involving the Coma Recovery Scale Revised (CRS-R) and ^{18}F-FDG PET imaging [14]. CRS-R assessments were administered by a trained neuropsychologist at least once a day over 5 days, and the best result was selected. The interpretation of ^{18}F-FDG PET scans involved visual examination by two Nuclear Medicine physicians to identify hypometabolic and preserved regions. Complete bilateral hypometabolism in the associative

frontoparietal cortex with preserved metabolism pointed toward a diagnosis of UWS, while incomplete hypometabolism and partial preservation of activity within these areas indicated a MCS. In cases of disagreement, a third doctor was consulted to make a final decision. If ambiguity persisted or there was disagreement within the team, reassessment of the patient was carried out until a consensus was reached. Inclusion criteria were: 1) DoC patients (age > 18 years) diagnosed with CRS-R and subclassified into UWS or MCS, and 2) more than 28 days post-brain injury. 3) During the PET examination, the patient remained still, and the image quality was satisfactory. Exclusion criteria were: 1) neuroimaging examination within 28 days from brain insult (10 cases excluded), 2) significant focal brain damage involving more than two-thirds of one hemisphere or severe deformation due to surgery or trauma, as confirmed by a certified neuroradiologist (9 cases excluded), 3) absence of follow-up data after 1 year (6 cases excluded), and 4) poor image quality (2 cases excluded), such as severe image artifacts caused by head movement. Ultimately, a total of 87 patients newly diagnosed with chronic DoC and receiving routine treatment were included in the subsequent analysis.

One year after the initial assessment, functional outcomes were evaluated using the Glasgow Outcome Scale-Extended (GOS-E), extracted from the patients' medical records in our institution. In cases of incomplete data, the attending physician or legal guardian was contacted. The outcomes were categorized into two groups: "consciousness recovery" (GOS-E > 2) and "consciousness non-recovery" (GOS ≤ 2). This study received approval from the Human Subject Research Ethics Committee of Xijing Hospital, Fourth Military Medical University, and informed consent was obtained as required.

2.2 ^{18}F-FDG PET imaging acquisition

All participants received an injection of 3.7 MBq/kg of ^{18}F-FDG with a radiochemical purity exceeding 95%, following a fasting period of more than 6 hours, in a subdued and quiet environment for a 40-minute uptake duration. PET images were acquired using a PET/CT scanner (United Imaging, uMI780) during a 15-minute bed position and 3D whole-head acquisition. A low-dose attenuation correction CT was performed, covering the entire skull with the following scan parameters: tube voltage of 120 kV and tube current of 200 mA. All images were reconstructed with a slice thickness and increment of 0.5 mm. Axial PET images, corrected for attenuation, were reconstructed using a 3D ordered-subset expectation maximization (OSEM) algorithm (8 iterations and 32 subsets, 3-mm cutoff).

2.3 Imaging preprocessing

The ^{18}F-FDG PET images underwent visual examination by two experienced board-certified Nuclear Medicine physicians, blinded to clinical information, adhering to current international terminology. Using the 3D Slicer tool (version 5.1.0, https://www.slicer.org), the original 3D CT and ^{18}F-FDG PET images were manually cropped to eliminate areas below the skull base and unused background parts. Subsequently, each subject's PET image underwent partial volume correction (PVC) using the Van Citert algorithm [16] and was coregistered with CT using the General BRAINS registration algorithm from the 3D Slicer tool. The CT and PET images were then skull-stripped using the HD-Brain Extraction Tool from 3D Slicer, and brain masks were obtained simultaneously. Finally, utilizing a deep learning-based segmentation model

[17], the PET images were automatically segmented into different brain regions. The cerebellar region was chosen, and the mean SUV (SUVmean) value of the cerebellar VOI served as a reference for subsequent SUVR calculations.

2.4 PET image standardization

For CT image standardization, the CT values underwent clipping within a range of 0 to 200 after CT preprocessing. Regarding PET image standardization, this study employed two methods, MIBH and SUVR. The MIBH method, derived from Johan Stender et al. [15], involved calculating individualized MIBH values using the subject's extracellular regions. The extracerebral VOI of each subject was determined as follows:

The extracerebral VOI = whole brain PET with skull - skull-stripped brain PET

Subsequently, the SUVmean of the extracerebral VOI served as a standardized reference. The MIBH value for each PET brain image was calculated using the following formula:

MIBH = (voxelwise SUV of sull-stripped PET image)/(SUVmean-of-extracerebral VOI)

The second standardized method was the conventional SUVR method, utilizing the SUVmean of the cerebellum as a reference to calculate the SUVR of the cerebellum in PET, expressed as:

SUVR = (voxelwise SUV of sull-stripped PET image)/(SUVmean-of-cerebellum VOI)

2.5 Image-based deep learning classification

Five classification tasks were examined based on different PET image standardization methods and different modality combinations, including MIBH + CT, SUVR + CT, MIBH only, SUVR only, and CT only. The deep learning model training was conducted using the Medical Open Network for Artistic Intelligence (MONAI) framework (version 1.0.1, https://monai.io) in a Windows 11 operating system, and Nvidia RTX A2000 GPU 8GB. The DenseNet121 network, provided by the MONAI framework, was adopted with default parameter configurations, including 64 filters in the first convolution layer, 32 filters added in each subsequent layer, (6, 12, 24, 16) layers in each pooling block, RELU activation, and batch normalization for feature normalization. The loss function employed was cross-entropy, and the optimization algorithm used was the adaptive moment estimation (Adam) optimizer with a learning rate of 1e-5 and a weight decay of 1e-4. Moreover, all input images were uniformly resized to (128, 128, 128), Min-Max normalized, and data-augmented by applying random flipping and rotating techniques before starting model training. The training epoch was set to 250 and the batch size to 4.

2.6 Tabular-based deep learning classification

In the DenseNet121-based deep learning classification, the multimodal MIBH + CT classifier demonstrated optimal classification performance. Consequently, 1024 deep features were extracted, representing the 1024 vector values output by the last fully connected layer in the DenseNet121 model.

After applying a common logarithmic transformation to the CRS-R scores of each subject, (lg10 (CRS-R) was concatenated with its 1024 deep features. Python Tabular (v1.0.2, https://pytorch-tabular.readthedocs.io/en/latest/) utilized the provided Tab Transformer Model for deep learning classification modeling based on tabular data. During training, the optimizer employed the Adam algorithm, and the loss function was cross-entropy. Epochs were set to 400, and the batch size was set to 128.

2.7 Statistical analysis

A total of 87 cases were randomly divided into training, validation, and test datasets in a 4:1:1 ratio. Each classifier underwent stratified 5-fold cross-validation with 100 repetitions to assess performance. The performance of each repetition was computed as the mean across the five folds, and the overall model performance was determined as the mean across all 100 repetitions. Binary classification model performance was evaluated based on sensitivity (Sen), specificity (Spe), accuracy (ACC), and the area under the ROC curve (AUC).

To visualize deep learning outputs, Grad CAM [18] was employed to extract the gradient class activation map generated by the last convolutionlayer in the final MIBH + CT DenseNet121 classification model. Then, a heatmap was generated by multiplying the gradient class activation maps with the original PET MIBH image, which produced a coarse localization map highlighting the important regions in the image for predicting the target. For tabular-based MIBH + CT deep features and CRS-R scores, the principal component analysis (PCA) algorithm was firstly utilized to reduce feature dimensionality to 10. Subsequently, t-SNE [19] was employed to generate a two-dimensional distribution map. The entire workflow is depicted in Fig. 1.

Results

3.1 Clinical Data

A total of 114 patients with consciousness disorders due to brain damage underwent ^{18}F-FDG PET/CT examinations at the Department of Nuclear Medicine, Xijing Hospital. Following stringent exclusion criteria, our study included 87 patients (58 (66.7%) males, mean age ± SD 64.4 ± 8.04 years). During PET scans, 32 (36.8%) patients were diagnosed with unresponsive wakefulness syndrome, while 55 (63.2%) were diagnosed with a minimally conscious state according to CRS-R. Among these, 45 (51.7%) patients had traumatic brain injuries, 33 (37.9%) had intracerebral hemorrhage and ischemic stroke, and 9 (10.3%) had toxic encephalopathy. The average time from the event occurrence to PET scan was 56.42 ± 42.5 days (ranging from 28 to 210 days). CRS-R scores at the PET scan were significantly higher in the consciousness recovery groups ($P < 0.001$). Patients with non-traumatic brain injuries exhibited worse consciousness recovery than those with traumatic brain injuries ($P = 0.026$). No significant differences were observed in age, sex, or time from event occurrence to PET scan between the consciousness recovery and non-recovery groups ($P > 0.05$). Detailed clinical characteristics are presented in Table 1.

Table 1
Detailed clinical characteristics in DOC patients

	Consciousness recovery (n = 52)	**Consciousness not recovery** (n = 35)	*P*
Gender (n,%)	35 (67.3%)	23(65.7%)	0.877
Male	17 (32.7%)	12(34.3%)	
Female			
Age (years, mean ± SD)	46.0 ± 17.2	50.74 ± 16.1	0.432
Diagnosis at PET scan (n,%)	5 (9.6%)	27 (77.1%)	0.001
UWS	47 (90.4%)	8 (22.9%)	
MCS			
CRS-R scores at PET scan	4.75 ± 0.94	2.00 ± 0.00	0.001
Time from event occur to PET scan (days, mean ± SD)	60.3 ± 53.1	54.6 ± 35.6	0.326
Etiology(n,%)	32 (61.5%)	13 (37.1%)	0.026
Traumatic	20 (38.5%)	22 (62.9%)	
Non-traumatic			

3.2 Performance of image-based deep learning classifier

We assessed the ability to predict consciousness recovery in all patients with five tasks: MIBH + CT, SUVR + CT, MIBH only, SUVR only, and CT only.

The MIBH-only task outperformed the SUVR-only task in predicting consciousness recovery, with AUCs of 0.764 ± 0.028 vs. 0.667 ± 0.039, 0.686 ± 0.170 vs. 0.670 ± 0.091, and 0.751 ± 0.093 vs. 0.412 ± 0.104 in the training, validation, and independent test sets, respectively. Accuracy in these datasets between MIBH and SUVR tasks was 0.670 ± 0.059 vs. 0.610 ± 0.027, 0.771 ± 0.123 vs. 0.605 ± 0.061, and 0.629 ± 0.070 vs. 0.500 ± 0.000. Sensitivity in the SUVR task was higher than that in the MIBH task in the training, validation, and independent test sets (0.852 ± 0.035 vs. 0.788 ± 0.100, 0.834 ± 0.034 vs. 0.796 ± 0.103, and 0.830 ± 0.040 vs. 0.707 ± 0.096, respectively). Specificity in the MIBH task was higher than that in the SUVR task in these datasets (0.629 ± 0.153 vs. 0.500 ± 0.006, 0.695 ± 0.133 vs. 0.522 ± 0.061, 0.796 ± 0.123 vs. 0.532 ± 0.052). Grad-CAM was utilized to visualize the MIBH heatmap for a single patient, as shown in Fig. 2.

The MIBH + CT task demonstrated superior performance in predicting consciousness recovery, achieving AUCs of 0.803 ± 0.024, 0.804 ± 0.059, and 0.784 ± 0.073 in the training, validation, and independent test sets, respectively. In these datasets, sensitivity was 0.813 ± 0.041, 0.845 ± 0.069, and 0.794 ± 0.055, while

specificity was 0.807 ± 0.043, 0.806 ± 0.076, and 0.807 ± 0.062, respectively. The SUVR + CT task exhibited poorer performance compared to MIBH + CT, with AUCs of 0.791 ± 0.036, 0.632 ± 0.067, and 0.612 ± 0.192 in the training, validation, and independent test sets, respectively. The MIBH + CT task outperformed SUVR + CT in both sensitivity and specificity. Figure 3 displays the average AUC comparisons for MIBH + CT and SUVR + CT classifiers.

The CT-only task yielded suboptimal results in predicting consciousness recovery, with AUCs of 0.632 ± 0.080, 0.563 ± 0.155, and 0.624 ± 0.160 in the training, validation, and independent test sets, respectively. In these datasets, sensitivity was 0.570 ± 0.140, 0.585 ± 0.153, and 0.580 ± 0.160, while specificity was 0.870 ± 0.260, 0.820 ± 0.186, and 0.820 ± 0.136, respectively. The average performance results are summarized in Table 2. Figure 4 illustrates the boxplots of accuracy, AUC, sensitivity, and specificity on the test dataset for all five tasks after 5-fold cross-validation.

Table 2
Multimodal classification performance

Modality	Dataset	Accuracy (mean ± SD)	AUC (mean ± SD)	Sensitivity (mean ± SD)	Specificity (mean ± SD)
MIBH + CT	training	0.816 ± 0.025	0.803 ± 0.024	0.813 ± 0.041	0.807 ± 0.043
	validation	0.808 ± 0.084	0.804 ± 0.059	0.845 ± 0.069	0.806 ± 0.076
	test	0.723 ± 0.054	0.784 ± 0.073	0.794 ± 0.055	0.807 ± 0.062
SUVR + CT	training	0.685 ± 0.020	0.785 ± 0.034	0.722 ± 0.124	0.752 ± 0.272
	validation	0.630 ± 0.024	0.629 ± 0.060	0.692 ± 0.090	0.733 ± 0.075
	test	0.636 ± 0.049	0.623 ± 0.173	0.687 ± 0.070	0.700 ± 0.288
MIBH	training	0.670 ± 0.059	0.764 ± 0.028	0.788 ± 0.100	0.629 ± 0.153
	validation	0.771 ± 0.123	0.686 ± 0.170	0.796 ± 0.103	0.695 ± 0.133
	test	0.629 ± 0.070	0.751 ± 0.093	0.707 ± 0.096	0.796 ± 0.123
SUVR	training	0.610 ± 0.027	0.667 ± 0.039	0.852 ± 0.035	0.500 ± 0.006
	validation	0.605 ± 0.061	0.670 ± 0.091	0.834 ± 0.034	0.522 ± 0.061
	test	0.500 ± 0.000	0.412 ± 0.104	0.830 ± 0.040	0.532 ± 0.052
CT	training	0.517 ± 0.034	0.632 ± 0.080	0.570 ± 0.140	0.870 ± 0.260
	validation	0.535 ± 0.055	0.563 ± 0.155	0.585 ± 0.153	0.820 ± 0.186
	test	0.500 ± 0.000	0.624 ± 0.160	0.580 ± 0.160	0.820 ± 0.136

3.3 Performance of Tabular based deep learning classifier based on MIBH + CT image depth features and CRS-R scores

Compared to image-based deep learning classifiers, the tabular-based classifier combined MIBH + CT deep features and clinical CRS-R scores achieved optimal classification performance, with an accuracy of 88.5% and 82.2%, AUC of 0.950 and 0.933, sensitivity of 0.93 and 0.8, and specificity of 0.83 and 0.85 on the training and test sets, respectively. Figure 5 displays the average AUC comparisons between the combination of CRS-R scores and deep features abular-based classifier and MIBH + CT image-based classifiers. Visualization of the Tabular-based classifier, combined with clinical CRS-R scores and PET/CT deep features, is presented in Fig. 6. The t-SNE plot resulted in a clear differentiation between MCS and UWS two clusters which reflected the feasibility and advantages of using a joint of image deep features and clinical behavioral scores for classification.

Discussion

Predicting which patients will recover consciousness is a challenging task. In this prospective cohort study of chronic DoC patients, image-based deep learning revealed that PET MIBH outperformed the PET SUVR task in predicting consciousness recovery. A prognostic model combining PET MIBH + CT with behavioral CRS-R scores successfully discriminated patients who would recover consciousness one year later from those who would not, achieving an AUC of 0.93 in independent test datasets.

The correlation between whole-brain energy metabolism and the individual level of consciousness have been substantiated [15]. SUVR standardizes regional ^{18}F-FDG activity to global brain ^{18}F-FDG activity; thus, widespread reductions in brain glucose metabolism may be underestimated. Nevertheless, researcher have demonstrated that normalization to the cerebellum does not impact whole-brain glucose metabolism [20]. MIBH enables accurate diagnosis and prediction of recovery from DoC because multiple studies confirm that a preserved glucose metabolic rate in at least one hemisphere is a crucial condition for regaining awareness after brain injury [15]. Bertrand et al. affirmed that quantitative ^{18}F-FDG PET using the MIBH quantification procedure is robust in diagnosing the state of consciousness, outperforming EEG-based automatic classification of the conscious state [21]. In our study, employing a deep learning framework, we demonstrated that the PET MIBH task surpasses the PET SUVR task in predicting consciousness recovery. Building on previous research, it is confirmed that PET MIBH standardization based on extracerebral mean SUV as a reference is more suitable for evaluating the level of consciousness in DoC patients. However, a crucial point in our research is that, although the specificity in the PET MIBH task surpassed that in the PET SUVR task, the classification threshold signifies the minimal energetic requirements for sustained awareness, highlighting the value of PET MIBH in predicting a good prognosis. The sensitivity in the SUVR task outperformed that in the PET MIBH task, indicating that PET SUVR standardization identifies more patients likely to regain consciousness. Predicting consciousness recovery with high sensitivity is crucial to avoid premature withdrawal of life-sustaining therapy.

Each imaging technology brings unique advantages, and the trend toward multimodal imaging for comprehensive evaluation aims to enhance diagnostic accuracy and develop prognostic models. Diego et al. demonstrated that combining EEG with PET based on cardiac-evoked responses yielded superior

diagnostic efficiency compared to behavioral studies, indicating better detection of implicit consciousness over explicit consciousness [22]. Carol et al. highlighted preserved default mode network anti-correlations and high metabolism in patients with EMCS and healthy controls, while UWS or MCS patients exhibited decreased metabolism and pathological between-network hyperconnectivity [23]. In our study, we observed that multimodal combinations, whether PET MIBH + CT or PET SUVR + CT, outperformed single modalities in predicting consciousness recovery. Despite the gradual replacement of brain CT by MRI, structural and density changes still contribute value to consciousness assessment [24]. Additionally, CT scans serve for attenuation correction of PET images, especially in the current scenario where PET/MR applications are not yet widespread.

Prior studies have separately compared models using only imaging features or only clinical characteristics, with findings suggesting that combining these predictors achieves greater accuracy [25–28]. Our results align with these previous studies, showing that combining whole-brain energy metabolism detected from ^{18}F-FDG PET with clinical CRS-R scores delivers the best performance in predicting whether a patient will recover from DoC or not. This underscores the necessity of accumulating evidence from multiple assessments, each possessing different sensitivity and specificity in detecting the capacity for consciousness recovery. Our model affirms that the combination of the metabolic index of the best-preserved hemisphere from ^{18}F-FDG PET imaging and behavioral CRS-R scores plays crucial roles in predicting consciousness recovery. The accuracy in classifying a patient as "consciousness recovery" or not one year after injury reached 82.2%, and the sensitivity of 80% across the testing datasets demonstrated a low false-negative rate, preventing the misprediction of nonrecovery in a patient capable of recovery. Simultaneously, the specificity across the testing datasets was 85.0%, indicating the model's precision in identifying patients with a high potential for consciousness recovery and reducing false positives in predicting low-potential patients.

In addition, this study also explored the intrinsic mechanisms of image-based deep learning classification models via feature visualization. The Grad-CAM [15] algorithm is a novel class-discriminative localization technique and is commonly used to visualize and localize suspicious lesion areas. Although the DoC patients may not present obvious lesion areas, the Grad-CAM heatmap can help reveal and visualize which regions in the image are important ones for predicting the target. As shown in Fig. 2, which compares the Grad-CAM MIBH heatmaps of one UWS patient and one MCS case the brain cortex demonstrates its importance in classification. Moreover, there exists an obvious difference in the highlighted cortex regions between UWS and MCS cases in terms of range and value, which is consistent with the generally different metabolic levels of UWS and MCS.

While this study combined MIBH of ^{18}F-FDG PET imaging and behavioral CRS-R scores for predicting whether a patient recovery from DoC, several limitations exist. First, the sample size used for constructing deep learning frameworks, although consisting of data from 87 DoC patients, is relatively small, requiring a larger cohort for further validation. Second, numerous additional factors, such as medical complications and nutrition, are associated with DoC patient outcomes. Future work should integrate

more information to construct a comprehensive prediction model. Third, the study focused on DoC patients from a single center, lacking external data for verification.

Conclusions

In conclusion, the proposed prognostic model, combining MIBH of ^{18}F-FDG PET imaging and behavioral CRS-R scores, predicts one-year outcomes for DoC patients at an individual level. The model achieved an AUC of 0.950 and 0.933 in the training and testing datasets for classifying a patient as "consciousness recovery" or not. Additionally, visualizing the results of image-based deep learning allows physicians and scientists to comprehend the potential mechanisms for consciousness recovery. These advantages offer an objective prognosis for DoC patients, optimizing their management and deepening our understanding of brain function during unconsciousness.

Abbreviations

DoC	**disorders of consciousness**
UWS	unresponsive awakening syndrome
MCS	minimally conscious state
^{18}F-FDG PET	**^{18}F-fluorodeoxyglucose positron emission tomography**
CRS-R	Revised Coma Recovery Scale
MIBH	metabolic index of the best-preserved hemisphere
SUVR	ratio SUV
GOS-E	Glasgow Outcome Scale-Extended
AUC	area under the curve

Declarations

Funding: This study was supported by the National Natural Science Foundation of China (grant number. 92259304, 81971646, 82122033, 82202208, 82171977] and Ministry of Science and Technology Foundation of China [grant number. 2022ZD0208000].

Conflicts of interest/Competing interests: The authors declare that they have no conflict of interest.

Availability of data and material: Data and materials are available from Jing Wang.

Research involving human participants and/or animals: The research involved human participants. This study was permitted by the Research Ethics Committee of the Fourth Military Medical University of Xijing hospital.

Consent to participate: All patients gave written informed consent prior to enrollment in this study.

References

1. Edlow BL, Claassen J, Schiff ND, Greer DM. Recovery from disorders of consciousness: mechanisms, prognosis and emerging therapies. Nat Rev Neurol. 2021;17:135–56.
2. Bao W, Li X, Luo B. A novel prognostic approach to predict recovery in patients with chronic disorders of consciousness. Neurosci Bull. 2019;35:953–4.
3. Thibaut A, Schiff N, Giacino J, Laureys S, Gosseries O. Therapeutic interventions in patients with prolonged disorders of consciousness. Lancet Neurol. 2019;18:600–14.
4. Saleem GT, Ewen JB, Crasta JE, Slomine BS, Cantarero GL, Suskauer SJ. Single-arm, open-label, dose escalation phase I study to evaluate the safety and feasibility of transcranial direct current stimulation with electroencephalography biomarkers in paediatric disorders of consciousness: a study protocol. BMJ Open. 2019;9:e029967.
5. Kundu B, Brock AA, Englot DJ, Butson CR, Rolston JD. Deep brain stimulation for the treatment of disorders of consciousness and cognition in traumatic brain injury patients: a review. Neurosurg Focus. 2018;45:E14.
6. Giacino JT, Schnakers C, Rodriguez-Moreno D, Kalmar K, Schiff N, Hirsch J. Behavioral assessment in patients with disorders of consciousness: gold standard or fool's gold? Prog Brain Res. 2009;177:33–48.
7. Estraneo A, Moretta P, Loreto V, Lanzillo B, Santoro L, Trojano L. Late recovery after traumatic, anoxic, or hemorrhagic long-lasting vegetative state. Neurology. 2010;75:239–45.
8. Kang J, Huang L, Tang Y, Chen G, Ye W, Wang J, et al. A dynamic model to predict long-term outcomes in patients with prolonged disorders of consciousness. Aging. 2022;14:789–99.
9. Steppacher I, Eickhoff S, Jordanov T, Kaps M, Witzke W, Kissler J. N400 predicts recovery from disorders of consciousness. Ann Neurol. 2013;73:594–602.
10. Scarpino M, Lolli F, Hakiki B, Atzori T, Lanzo G, Sterpu R, et al. Prognostic value of post-acute EEG in severe disorders of consciousness, using American Clinical Neurophysiology Society terminology. Neurophysiol Clin. 2019;49:317–27.
11. Maas AI, Steyerberg EW, Butcher I, Dammers R, Lu J, Marmarou A, et al. Prognostic value of computerized tomography scan characteristics in traumatic brain injury: results from the IMPACT study. J Neurotrauma. 2007;24:303–14.
12. Wu X, Zou Q, Hu J, Tang W, Mao Y, Gao L, et al. Intrinsic functional connectivity patterns predict consciousness level and recovery outcome in acquired brain injury. J Neurosci. 2015;35:12932–46.
13. Zhang J, Zhang H, Yan F, Zhang H, Zhang E, Wang X, et al. Investigating the mechanism and prognosis of patients with disorders of consciousness on the basis of brain networks between the thalamus and whole-brain. Front Neurol. 2022;13:990686.

14. Stender J, Gosseries O, Bruno MA, Charland-Verville V, Vanhaudenhuyse A, Demertzi A, et al. Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. Lancet. 2014;384:514–22.
15. Stender J, Mortensen KN, Thibaut A, Darkner S, Laureys S, Gjedde A, et al. The minimal energetic requirement of sustained awareness after brain injury. Curr Biol. 2016;26:1494–9.
16. ohka J, Reilhac A. Deconvolution-based partial volume correction in Raclopride-PET and Monte Carlo comparison to MR-based method. NeuroImage. 2008;39:1570–84.
17. Shiyam Sundar LK, Yu J, Muzik O, Kulterer OC, Fueger B, Kifjak D, et al. Fully automated, semantic segmentation of Whole-Body 18F-FDG PET/CT images based on data-centric artificial intelligence. J Nucl Med. 2022;63:1941–8.
18. Jiang M, Qiu Y, Zhang W, Zhang J, Wang Z, Ke W et al. Visualization deep learning model for automatic arrhythmias classification. Physiol Meas. 2022; 43.
19. Van der Maaten, Hinton GE. Visualizing high-dimensional data, using t-SNE. J Mach Learn Res. 2008;9:2579–605.
20. Madsen K, Hesby S, Poulsen I, Fuglsang S, Graff J, Larsen KB, et al. Comparison of analytical methods of brain [18F]FDG-PET after severe traumatic brain injury. J Neurosci Methods. 2017;291:176–81.
21. Hermann B, Stender J, Habert MO, Kas A, Denis-Valente M, Raimondo F, et al. Multimodal FDG-PET and EEG assessment improves diagnosis and prognostication of disorders of consciousness. Neuroimage Clin. 2021;30:102601.
22. Candia-Rivera D, Annen J, Gosseries O, Martial C, Thibaut A, Laureys S, et al. Neural responses to heartbeats detect residual signs of consciousness during resting state in postcomatose patients. J Neurosci. 2021;41:5251–262.
23. Di Perri C, Bahri MA, Amico E, Thibaut A, Heine L, Antonopoulos G, 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;15:830–42.
24. Skandsen T, Kvistad KA, Solheim O, Strand IH, Folvik M, Vik A. Prevalence and impact of diffuse axonal injury in patients with moderate and severe head injury: a cohort study of early magnetic resonance imaging findings and 1-year outcome. J Neurosurg. 2010;113:556–63.
25. Song M, Yang Y, He J, Yang Z, Yu S, Xie Q, et al. Prognostication of chronic disorders of consciousness using brain functional networks and clinical characteristics. Elife. 2018;7:e36173.
26. Steppacher I, Fuchs P, Kaps M, Nussbeck FW, Kissler J. A tree of life? Multivariate logistic outcome-prediction in disorders of consciousness. Brain Inj. 2020;34:399–406.
27. Scarpino M, Lolli F, Hakiki B, Lanzo G, Sterpu R, Atzori T, et al. EEG and Coma Recovery Scale-Revised prediction of neurological outcome in Disorder of Consciousness patients. Acta Neurol Scand. 2020;142:221–8.
28. Chennu S, Annen J, Wannez S, Thibaut A, Chatelle C, Cassol H, et al. Brain networks predict metabolism, diagnosis and prognosis at the bedside in disorders of consciousness. Brain.

2017;140:2120–32.

Figures

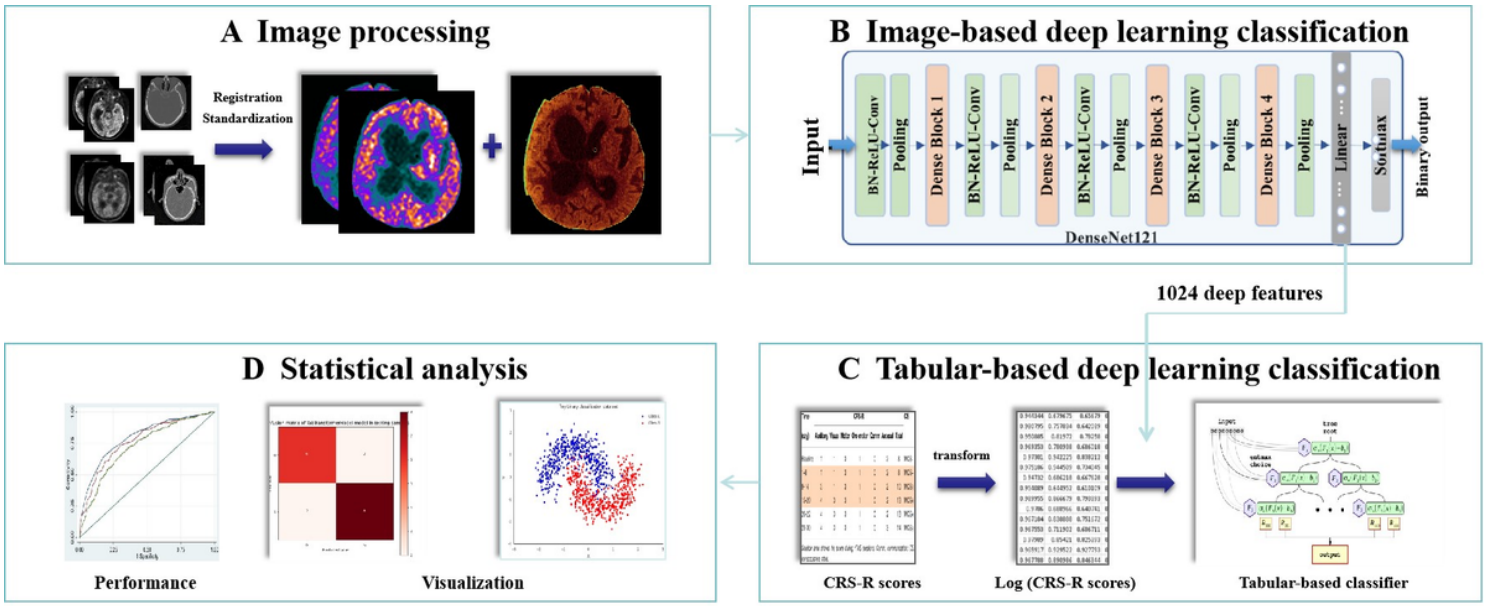

Figure 1

Whole procedure of the present study. (A) PET and CT images were preprocessed and standardized. (B) The DenseNet121, provided by the MONAI framework, was used for deep learning training, setting a binary classification of consciousness recovery or not. (C) CRS-R concatenated with 1024 deep features extracted from MIBH+CT in the DenseNet121 model. Python Tabular was used for deep learning classification modeling. (D) Sensitivity (Sen), specificity (Spec), accuracy (Acc), and the area under the ROC curve were evaluated to assess the performance of the models. Grad CAM was used for visualization of image-based deep-learning classification, and t-SNE was used to generate a two-dimensional distribution map of the original data.

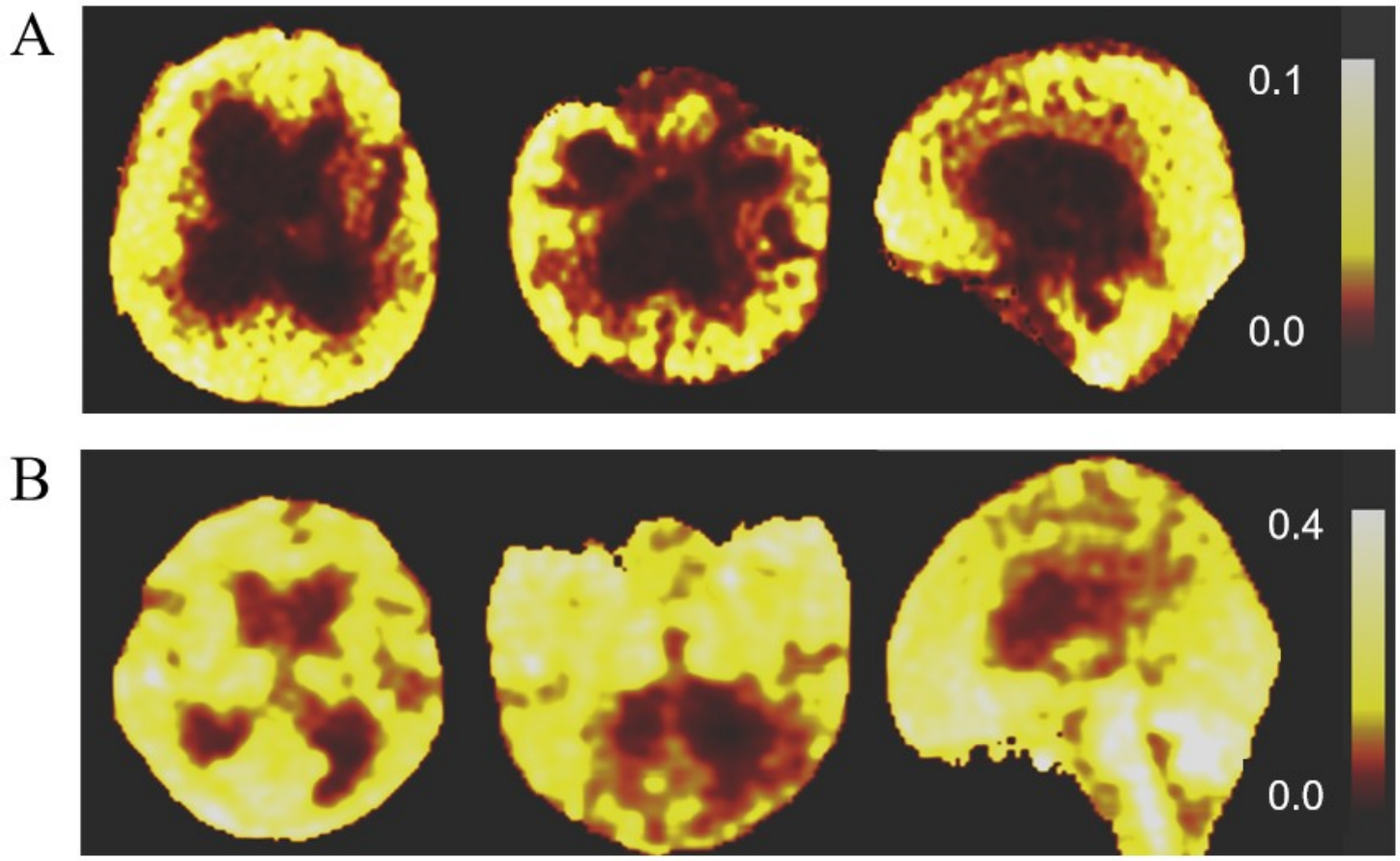

Figure 2

Grad-CAM MIBH heatmap visualization of image-based DL classifier. (A) Transverse, coronal, sagittal views of an MCS case. (B) Transverse, coronal, sagittal views of a UWS case.

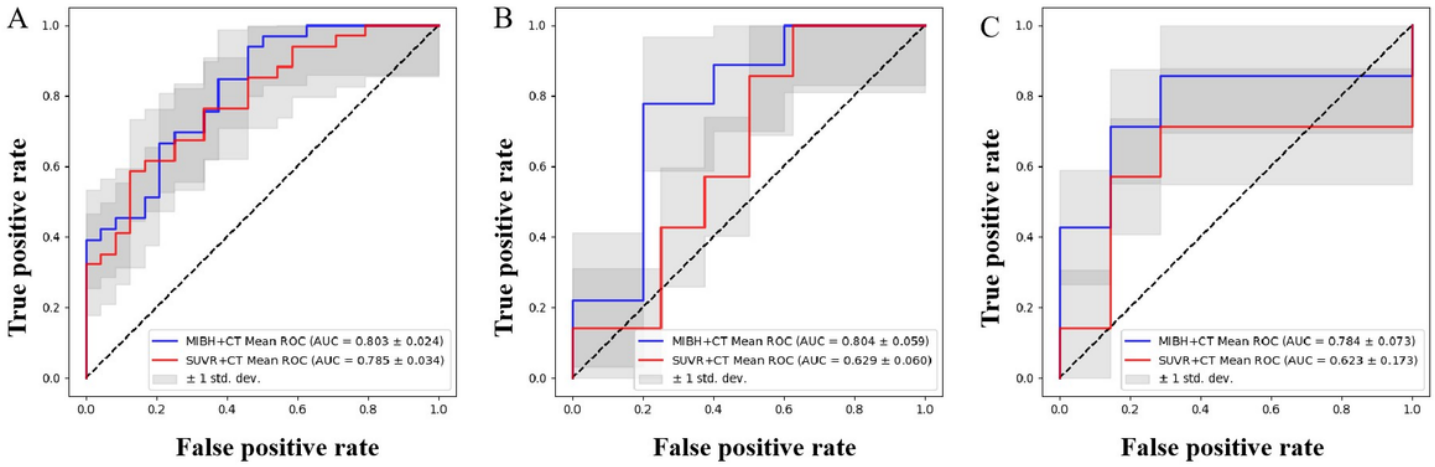

Figure 3

MIBH+CT and SUVR+CT classifiers average ROC comparison. The AUCs of the MIBH+CT model are higher in the training (A) validation (B) and independent test sets (C) compared to the SUVR+CT model.

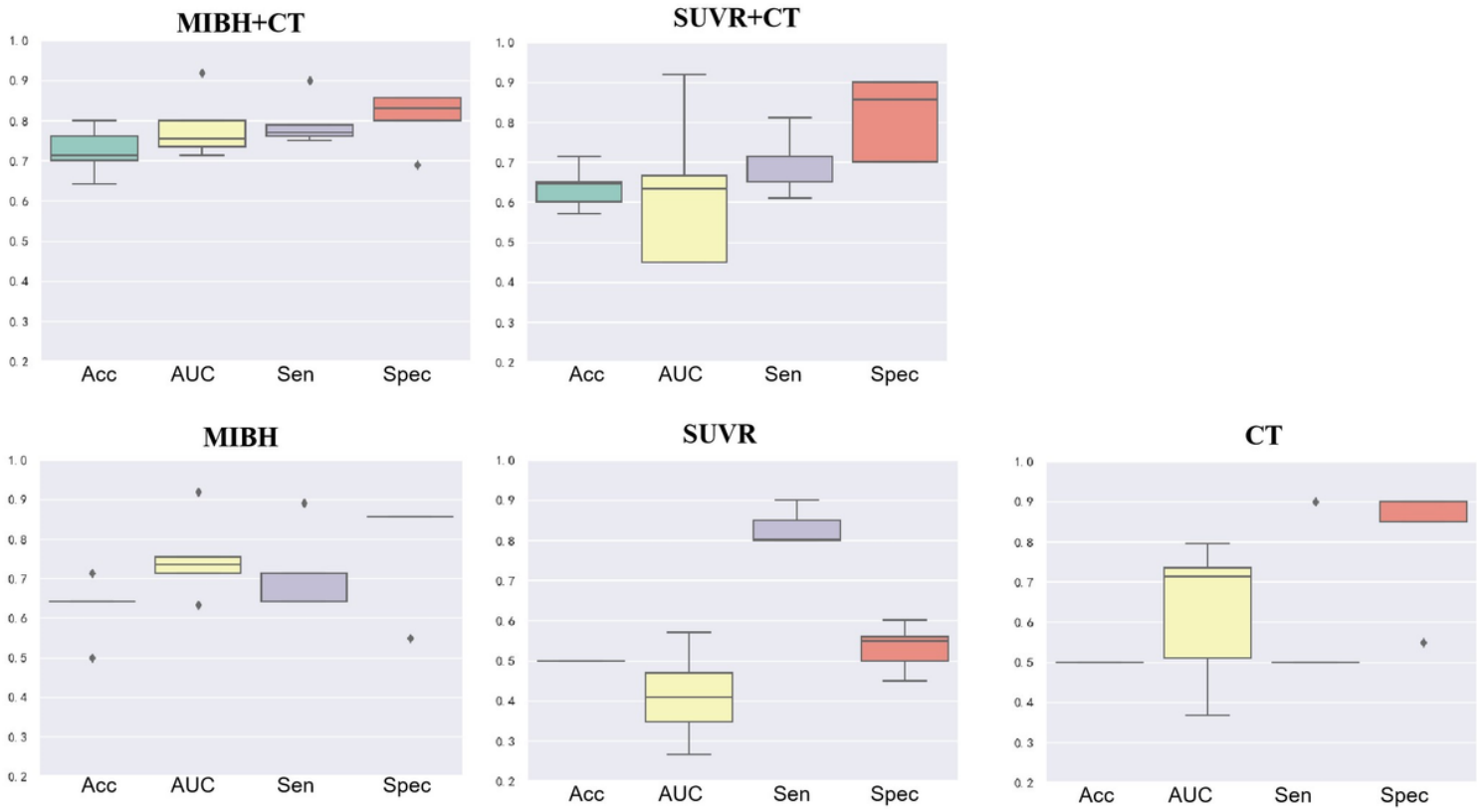

Figure 4

Acc, AUC, Sen and Spec boxplots on the test dataset for all five image tasks after 5-fold cross-validation. (A) MIBH+CT model (B) SUVR+CT model (C) MIBH only model (D) SUVR model (E) CT only.

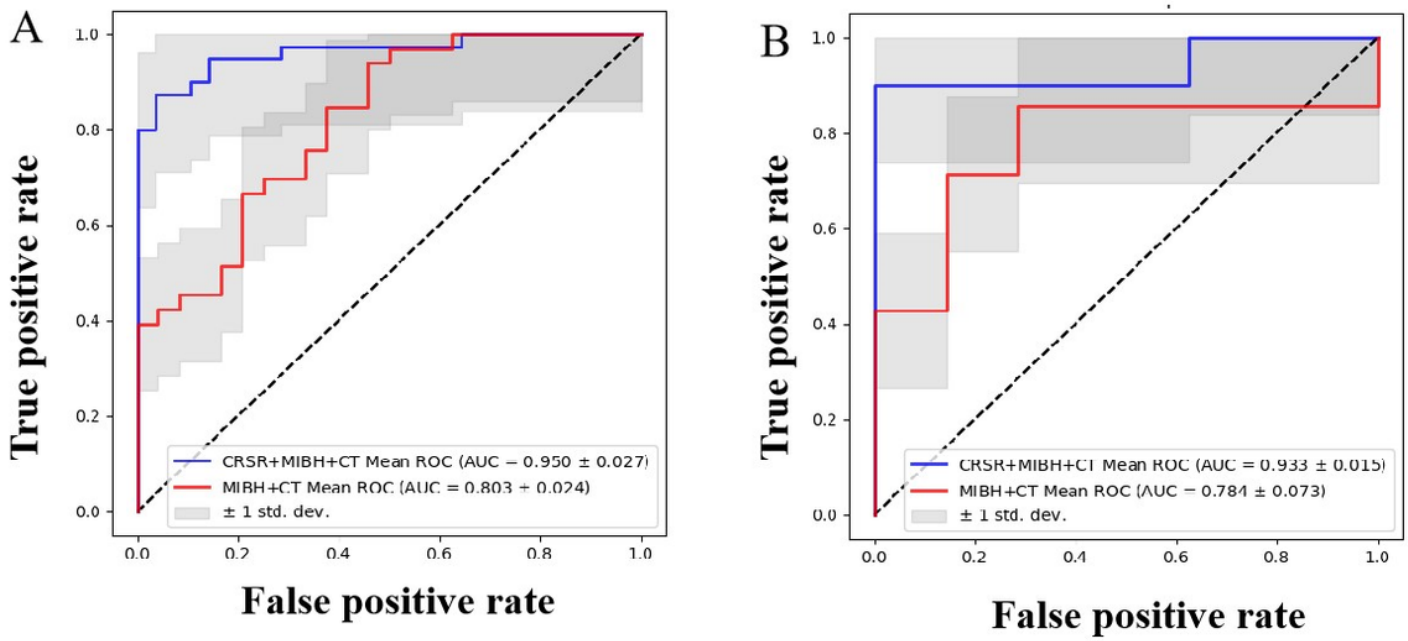

Figure 5

CRS-R combination MIBH+CT deep-feature classifers and MIBH+CT classifiers average ROC comparison. Compared to the MIBH+CT image-based classifier, CRS-R combination MIBH+CT deep-feature classifiers achieved better performance, with an accuracy of 88.5% and 82.2%, AUC of 0.950 and 0.933, Sen of 0.93 and 0.8, Spec of 0.83 and 0.85 on the training set (A) and test set (B), respectively.

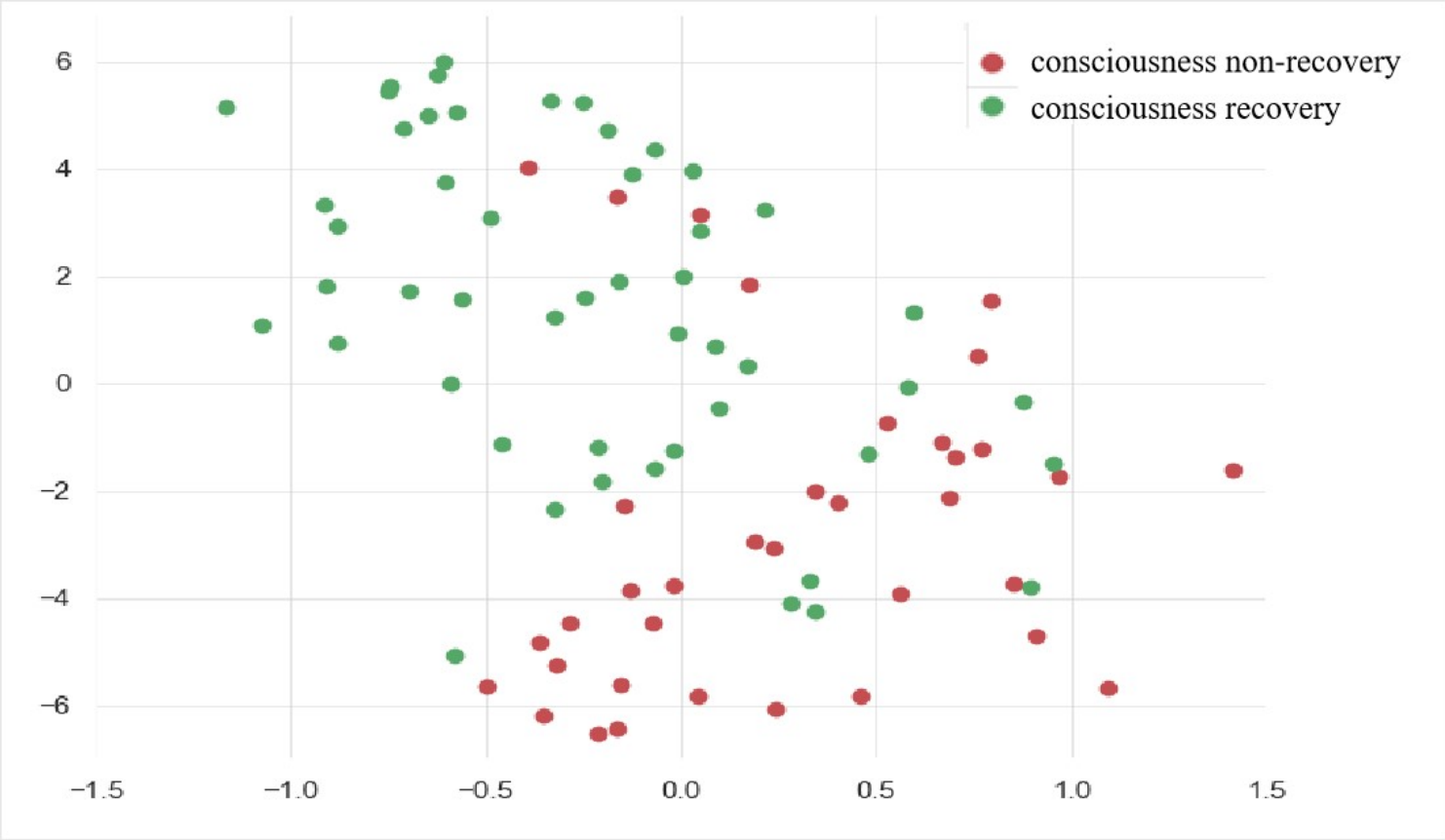

Figure 6

t-SNE visualization of tabular-based DL classifier. A classifier combining clinical CRS-R scores and MIBH+CT deep features accurately predicts patients' recovery from disorders of consciousness or not.