













# Quantifying arousal and awareness in altered states of consciousness using interpretable deep learning

Minji Lee <sup>1</sup>, Leandro R. D. Sanz <sup>2,3</sup>, Alice Barra<sup>2,3</sup>, Audrey Wolff<sup>2,3</sup>, Jaakko O. Nieminen <sup>4,5</sup>, Melanie Boly <sup>4,6</sup>, Mario Rosanova <sup>7,8</sup>, Silvia Casarotto <sup>7,9</sup>, Olivier Bodart<sup>2</sup>, Jitka Annen <sup>2,3</sup>, Aurore Thibaut<sup>2,3</sup>, Rajanikant Panda <sup>2,3</sup>, Vincent Bonhomme <sup>10,11,12</sup>, Marcello Massimini<sup>7,9</sup>, Giulio Tononi <sup>4</sup>, Steven Laureys<sup>2,3</sup>, Olivia Gosseries <sup>2,3,4,13,15</sup> ✉ & Seong-Whan Lee <sup>14,15</sup> ✉

Consciousness can be defined by two components: arousal (wakefulness) and awareness (subjective experience). However, neurophysiological consciousness metrics able to disentangle between these components have not been reported. Here, we propose an explainable consciousness indicator (ECI) using deep learning to disentangle the components of consciousness. We employ electroencephalographic (EEG) responses to transcranial magnetic stimulation under various conditions, including sleep ( $n = 6$ ), general anesthesia ( $n = 16$ ), and severe brain injury ( $n = 34$ ). We also test our framework using resting-state EEG under general anesthesia ( $n = 15$ ) and severe brain injury ( $n = 34$ ). ECI simultaneously quantifies arousal and awareness under physiological, pharmacological, and pathological conditions. Particularly, ketamine-induced anesthesia and rapid eye movement sleep with low arousal and high awareness are clearly distinguished from other states. In addition, parietal regions appear most relevant for quantifying arousal and awareness. This indicator provides insights into the neural correlates of altered states of consciousness.

<sup>1</sup>Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea. <sup>2</sup>Coma Science Group, GIGA-Consciousness, GIGA Research Center, University of Liège, Liège, Belgium. <sup>3</sup>Centre du Cerveau<sup>2</sup>, University Hospital of Liège, Liège, Belgium. <sup>4</sup>Wisconsin Institute for Sleep and Consciousness, Department of Psychiatry, University of Wisconsin, Madison, USA. <sup>5</sup>Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Espoo, Finland. <sup>6</sup>Department of Neurology, University of Wisconsin, Madison, WI, USA. <sup>7</sup>Department of Biomedical and Clinical Sciences “L. Sacco”, University of Milan, Milan, Italy. <sup>8</sup>Fondazione Europea di Ricerca Biomedica, FERB Onlus, Milan, Italy. <sup>9</sup>IRCCS Fondazione Don Carlo Gnocchi ONLUS, Milan, Italy. <sup>10</sup>Department of Anesthesia and Intensive Care Medicine, University Hospital of Liège, Liège, Belgium. <sup>11</sup>University Department of Anesthesia and Intensive Care Medicine, CHR Citadelle, Liège, Belgium. <sup>12</sup>Anesthesia and Intensive Care Laboratory, GIGA-Consciousness, GIGA Research Center, University of Liège, Liège, Belgium. <sup>13</sup>Department of Psychology, University of Wisconsin, Madison, WI, USA. <sup>14</sup>Department of Artificial Intelligence, Korea University, Seoul, Republic of Korea. <sup>15</sup>These authors contributed equally: Olivia Gosseries, Seong-Whan Lee. ✉email: [ogosseries@uliege.be](mailto:ogosseries@uliege.be); [sw.lee@korea.ac.kr](mailto:sw.lee@korea.ac.kr)

Responsiveness is often thought to reflect consciousness, and for a long time, unresponsiveness was considered as a surrogate of unconsciousness. However, consciousness and responsiveness are two different concepts<sup>1</sup>. Consciousness is considered to be absent during sleep or anesthesia, but in certain instances, subjective experience can still occur (e.g., dreaming)<sup>2,3</sup>. Similarly, consciousness has been described as a result of both arousal and awareness components<sup>4</sup>. Arousal refers to the overall state of alertness (or wakefulness). In contrast, awareness refers to the subjective experience, such as perceiving a blue triangle versus a red circle<sup>5</sup>. Typically, at the clinical level, arousal is indicated by the opening of the eyes, and awareness is inferred by the ability to follow commands.

Various levels of consciousness exist in physiological, pharmacological, and pathological modifications of consciousness (Table 1). In non-rapid eye movement (NREM) sleep with no subsequent reports of subjective experiences, both arousal and awareness are low. However, in rapid eye movement (REM) sleep with dreams, arousal is low but awareness can reach high levels<sup>5,6</sup>. Under general anesthesia, propofol and xenon predominantly induce states similar to NREM sleep without subjective experiences, whereas ketamine induces dream-like experiences similar to REM sleep with subjective reports upon awakening<sup>3,7</sup>. Additionally, recent findings suggest that a minority of patients under anesthesia may also be conscious of their environment during a surgical procedure, which is referred to as connected consciousness. These patients exhibit the ability to follow commands using the isolated forearm technique<sup>8</sup>, without recollection upon awakening. Unresponsive wakefulness syndrome (UWS) describes patients who recover with their eyes open, but only demonstrate reflex behaviors, while patients in a minimally conscious state (MCS) present reproducible non-reflex movements<sup>9</sup>. In both UWS and MCS patients, arousal is high; however, unlike UWS, MCS patients also show signs of awareness that can be considered as a sign of volitional behavior<sup>10,11</sup>. However, certain UWS patients (assessed several times with the Coma Recovery Scale-Revised (CRS-R)<sup>12</sup> may present brain activity similar to MCS patients (e.g., high brain metabolism measured by positron emission tomography). This peculiar state has been termed non-behavioral MCS or MCS\*<sup>13,14,15</sup>.

An effective measure of consciousness, labeled perturbational complexity index (PCI), was developed from electroencephalographic (EEG) responses to direct and noninvasive cortical perturbation with transcranial magnetic stimulation (TMS)<sup>16</sup>. PCI quantifies the complexity of deterministic patterns of significant cortical activation evoked by TMS. This index was validated in a large benchmark population to derive an empirical cutoff ( $PCI^* = 0.31$ ) that reliably discriminates between unconsciousness ( $PCI_{max} \leq PCI^*$ : NREM sleep; midazolam-, propofol- and xenon-induced anesthesia) and consciousness ( $PCI_{max} > PCI^*$ :

REM sleep; wakefulness; ketamine-induced anesthesia; and conscious brain-injured patients)<sup>17</sup>. However, PCI cannot discriminate REM sleep or ketamine-induced anesthesia from healthy wakefulness. In addition, multiple trials are required to compute PCI<sup>16,17</sup>. A few studies have attempted to develop an objective measure of consciousness from resting-state EEG brain activity<sup>3,4,18,19</sup>. Interestingly, the spectral exponent, which quantifies the slope of power spectral density of resting-state EEG activity, is another measure of consciousness that is highly correlated with PCI and allows distinguishing between ketamine and propofol or xenon-induced anesthesia<sup>3</sup>. In addition, when low- (1–20 Hz) and high-band (20–40 Hz) spectral exponents are jointly considered, ketamine-induced anesthesia can be distinguished from wakefulness, and the xenon and propofol-induced anesthesia conditions are partially superimposed in the spectral exponent; consequently, these are difficult to distinguish<sup>3</sup>. Recently, a decrease in high-frequency oscillations and an increase in low-frequency power in the primary sensory, motor, and visual cortices were observed during REM sleep when compared to healthy wakefulness<sup>20</sup>. The quantification of the spectral slope between 30 and 50 Hz was also proposed for discriminating REM sleep from healthy wakefulness<sup>4</sup>. However, this measure did not differentiate REM from NREM sleep; thus, it distinguishes between different arousal levels but not awareness levels. Therefore, an alternative measure to simultaneously disentangle the two components of consciousness, requiring fewer trials, would be a valuable and necessary tool.

The classical neurophysiological approach for calculating PCI, power spectral density, and spectral exponent relies on many epochs to improve the reliability of statistical estimates of these indices<sup>21</sup>. However, these methods are only suitable for investigating the averaged brain states and they can only clarify general neurophysiological aspects. Machine learning (ML) allows decoding and identifying specific brain states and discriminating them from unrelated brain signals, even in a single trial in real-time<sup>22</sup>. This can potentially transform statistical results at the group level into individual predictions<sup>9</sup>. A deep neural network, which is a popular approach in ML, has been employed to classify or predict brain states using EEG data<sup>23</sup>. Particularly, a convolutional neural network (CNN) is the most extensively used technique in deep learning and has proven to be effective in the classification of EEG data<sup>24</sup>. However, a CNN has the drawback that it cannot provide information on why it made a particular prediction<sup>25</sup>. Recently, layer-wise relevance propagation (LRP) has successfully demonstrated why classifiers such as CNNs have made a specific decision<sup>26</sup>. Specifically, the relevance score resulting from the LRP indicates the contribution of each input variable to the classification or prediction decision. Thus, a high score in a particular area of an input variable implies that the classifier has made the classification or prediction using this

**Table 1 Schematic representation of different states of consciousness according to low or high arousal and awareness: the plus sign indicates high arousal or awareness, whereas the minus sign indicates low arousal or awareness.**

Condition	State	Arousal	Awareness
Physiology	Healthy wakefulness	+	+
	REM sleep with dreams	–	+
	NREM sleep without dreams	–	–
Pharmacology	Anesthesia induced with ketamine	–	+
	Anesthesia induced with propofol or xenon	–	–
Pathology	MCS	+	+
	MCS*	+	+
	UWS	+	–

REM rapid eye movement, NREM non-rapid eye movement, MCS minimally conscious state, MCS\* non-behavioral MCS, UWS unresponsive wakefulness syndrome. Note that the anesthesia-induced with propofol and xenon mentioned here does not include the use of the isolated forearm technique.

feature. For example, neurophysiological data suggest that the left motor region is activated during right-hand motor imagery<sup>27</sup>. The LRP indicates that the neural network classifies EEG data as right-hand motor imagery because of the activity of the left motor region<sup>28</sup>. Therefore, the relevance score was higher in the left motor region than in other regions. Thus, it is possible to interpret the neurophysiological phenomena underlying the decisions of CNNs using LRP.

In this work, we develop a metric, called the explainable consciousness indicator (ECI), to simultaneously quantify the two components of consciousness—arousal and awareness—using CNN. The processed time-series EEG data were used as an input of the CNN. Unlike PCI, which relies on source modeling and permutation-based statistical analysis, ECI used event-related potentials at the sensor level for spatiotemporal dynamics and ML approaches. For a generalized model, we used the leave-one-participant-out (LOPO) approach for transfer learning, which is a type of ML that transfers information to a new participant not included in the training phase<sup>24,27</sup>. The proposed indicator is a 2D value consisting of indicators of arousal (ECI<sup>aro</sup>) and awareness (ECI<sup>awa</sup>). First, we used TMS-EEG data collected from healthy participants during NREM sleep with no subjective experience, REM sleep with subjective experience, and healthy wakefulness to consider each component of consciousness (i.e., low/high arousal and low/high awareness) with the aim to analyze correlations between the proposed ECI and the three states, namely NREM, REM, and wakefulness. Next, we measured ECI using TMS-EEG data collected under general anesthesia with ketamine, propofol, and xenon, again with the aim to measure correlation with these three anesthetics. Before anesthesia, TMS-EEG data were also recorded during healthy wakefulness. Upon awakening, healthy participants reported conscious experience during ketamine-induced anesthesia and no conscious experience during propofol- and xenon-induced anesthesia. Finally, TMS-EEG data were collected from patients with disorders of consciousness (DoC), which includes patients diagnosed as UWS and MCS patients. We hypothesized that our proposed ECI can clearly distinguish between the two components of consciousness under physiological, pharmacological, and pathological conditions.

To verify the proposed indicator, we next compared ECI<sup>awa</sup> with PCI, which is a reliable index for consciousness. Then, we applied ECI to additional resting-state EEG data acquired in the anesthetized participants and patients with DoC. We hypothesize that if CNN can learn characteristics related to consciousness, it could calculate ECI accurately even without TMS in the proposed framework. In terms of clinical applicability, it is important to use the classifier from the previous LOPO training of the old data to classify the new data (without additional training). Therefore, we computed ECI in patients with DoC using a hold-out approach<sup>29</sup>, where training data and evaluation data are arbitrarily divided, instead of cross-validation. Finally, we investigated why the classifier generated these decisions using LRP to interpret ECI<sup>30</sup>. We show that proposed ECI using interpretable deep learning distinguishes arousal and awareness between normal consciousness, sleep, anesthesia, and patients with DoC. Furthermore, we show that the parietal region is most closely related to quantifying arousal and awareness in altered states of consciousness.

## Results

**Overview of the calculation of ECI.** We used TMS-EEG data in three conditions: (i) sleep, (ii) general anesthesia, and (iii) severely brain-injured patients (Fig. 1a). Figure 1b shows the framework for calculating ECI to distinguish between low and high states in each arousal and awareness. To explore the optimal input and

**Table 2 Averaged single-trial classification accuracy (%) in physiological, pharmacological, and pathological conditions for TMS-EEG: this represents the accuracy  $\pm$  standard deviation.**

Target domain	Source domain	Arousal	Awareness
Sleep	Sleep	87.79 $\pm$ 2.50	91.95 $\pm$ 4.74
	Sleep + Ane	87.23 $\pm$ 2.99	89.96 $\pm$ 5.48
	Sleep + DoC	80.73 $\pm$ 5.05	89.60 $\pm$ 4.26
	Sleep + Ane + DoC	84.01 $\pm$ 3.46	91.14 $\pm$ 4.29
Ane	Ane	79.01 $\pm$ 10.61	80.20 $\pm$ 10.06
	Ane + Sleep	82.58 $\pm$ 6.92	87.78 $\pm$ 6.46
	Ane + DoC	69.99 $\pm$ 11.89	82.22 $\pm$ 10.66
	Ane + Sleep + DoC	72.68 $\pm$ 17.22	85.61 $\pm$ 9.09
DoC	DoC	-	75.84 $\pm$ 14.71
	DoC + Sleep	75.94 $\pm$ 18.14	79.44 $\pm$ 15.51
	DoC + Ane	83.12 $\pm$ 12.79	75.30 $\pm$ 11.99
	DoC + Sleep + Ane	66.29 $\pm$ 19.02	78.78 $\pm$ 12.98

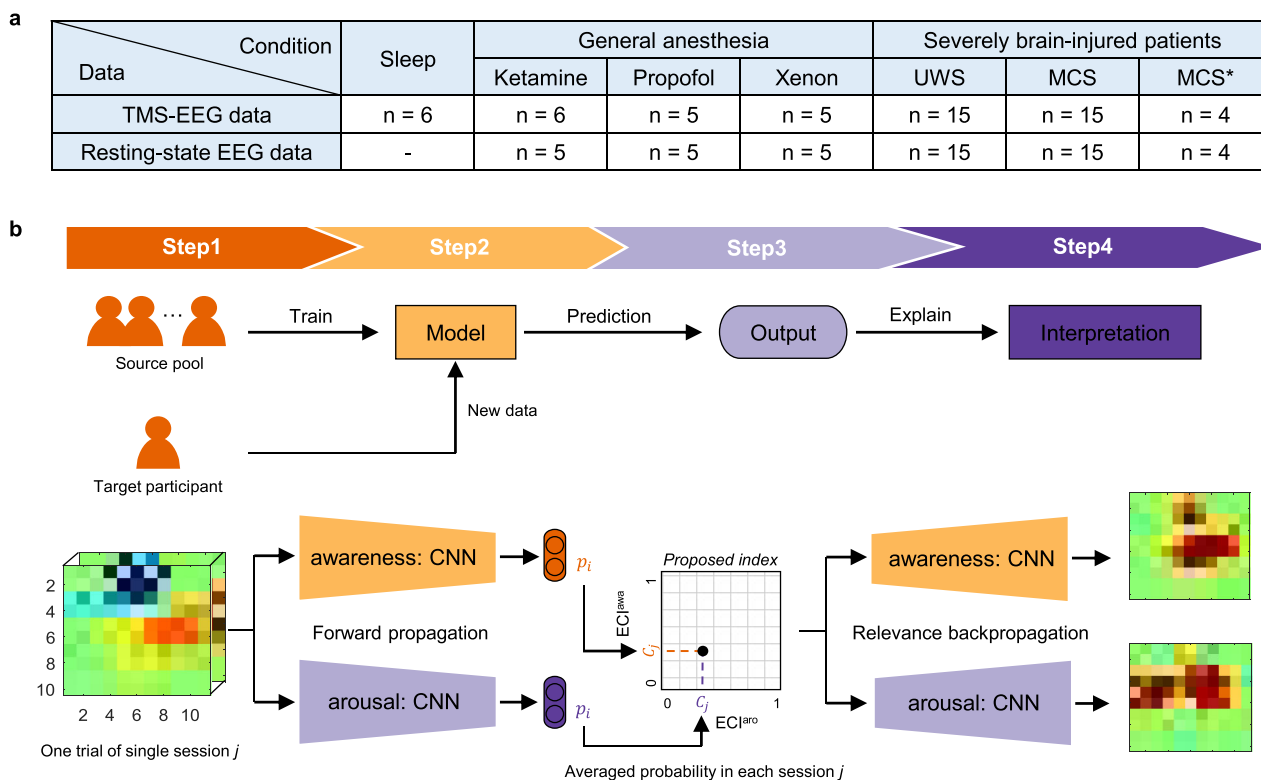
Ane anesthesia domain, DoC patients with disorders of consciousness domain.

The target domain implies the condition with the target participant to be tested for calculating explainable consciousness indicator (ECI) using convolutional neural network (CNN) with spatiotemporal information, and the source domain implies the conditions included in training for learning classifiers.

classifier, we compared the single-trial classification performance in each component (see Supplementary Notes 1–2, Supplementary Figs. 1–3, and Supplementary Tables 1–3). For an input in the classifier, we converted time-series EEG data to 3D data (2D meshes according to the spatial information + 1D vector according to temporal information). The CNN model was used to distinguish low from high state in each arousal and awareness. Next, the interclass probability in the new target participant was calculated using the trained model. Finally, ECI was measured by averaging the probability of a high state in each arousal and awareness over a single session.

To improve the classification performance, we trained the models with domain transfer learning, which uses the knowledge learned in one domain to improve generalization, based on similarity in the sense of pooling the training data across domains (Supplementary Fig. 4). When referring to domain transfer learning using the EEG signals in this study, the domain refers to the clinical condition in which the EEG signals were acquired. Precisely, three domains were considered here: sleep, anesthesia, and patients with DoC. In domain transfer learning, the target domain indicates a state that contains a single session for calculating ECI, whereas the source domain indicates those sessions that are included in the training phase. Therefore, the information (knowledge) trained in the source domain is applied when testing the target domain, and this is described as the transfer of learned information for domain transfer learning. In the LOPO approach, the data from all the participants in the source domain, except for the target participant, were used for training (Supplementary Fig. 3b). Note that the data for training and testing did not overlap. We used all three domains to classify high and low states in both arousal and awareness (Table 2). Consequently, the classification performance was higher when training with a closer domain (see Supplementary Notes 3–4 and Supplementary Table 4). Therefore, we trained these domains together when calculating ECI in sleep and anesthesia domains, but trained the DoC domain along with the anesthesia domain when calculating ECI in patients with DoC.

Then, we applied ECI to resting-state EEG data in two conditions: (i) general anesthesia and (ii) severely brain-injured patients (Fig. 1a). Similarly, we classified low and high states in resting-state EEG data for arousal and awareness using domain



**Fig. 1 Overview of the study.** **a** Data description. The same participants participated in transcranial magnetic stimulation-induced electroencephalography (TMS-EEG) and resting-state electroencephalography (EEG) measurements. The sleep condition did not include resting-state EEG, and one participant under ketamine-mediated anesthesia was missing in resting-state EEG. **b** Schematic framework for determining the explainable consciousness indicator (ECI). In step 1, raw EEG signals were converted into a spatio-spectral or spatiotemporal 3D matrix. In step 2, the converted 3D feature was used on a convolutional neural network in the two components of consciousness: arousal and awareness. In each arousal and awareness state, the EEG data were trained as two classes (low versus high). For example, for awareness, rapid eye movement (REM) sleep with subjective experience (i.e., dreaming) and healthy wakefulness belong to the same class in terms of high awareness; however, for arousal, non-rapid eye movement (NREM) with no subjective experience and REM sleep with subjective experience belong to the same class in terms of low arousal. The output  $p_i$  indicates the probability in the trial  $i$  of arousal and awareness. For the training and test phase, we used the leave-one participant-out approach as transfer learning. Therefore, the EEG data in the source pool were used for training and the data of target participants was predicted for arousal or awareness. The source pool contains data corresponding to the source domain except for the target participant. In step 3, the interclass probability for each arousal and awareness was averaged for calculating ECI in each session  $j$ . The averaged probability  $C_j$  is  $ECI^{arousal}$  and  $ECI^{awareness}$  on the x- and y-axes, respectively. Therefore, we represented the 2D consciousness indicator for the two components of consciousness. In the final step, we checked which brain signals the model has learned and why it made such a decision using layer-wise relevance propagation (LRP). Through this step, we could interpret the proposed indicator.  $ECI^{awareness} = ECI$  in awareness component;  $ECI^{arousal} = ECI$  in arousal component.

**Table 3 Averaged single-trial classification accuracy (%) in pharmacological and pathological conditions for resting-state EEG: this represents the accuracy ± standard deviation.**

Target domain	Source domain	Arousal	Awareness
Ane	Ane	89.91 ± 8.01	90.14 ± 6.96
	Ane + DoC	73.33 ± 14.25	88.62 ± 8.83
DoC	DoC	-	73.03 ± 17.26
	DoC + Ane	86.04 ± 12.35	73.68 ± 15.42

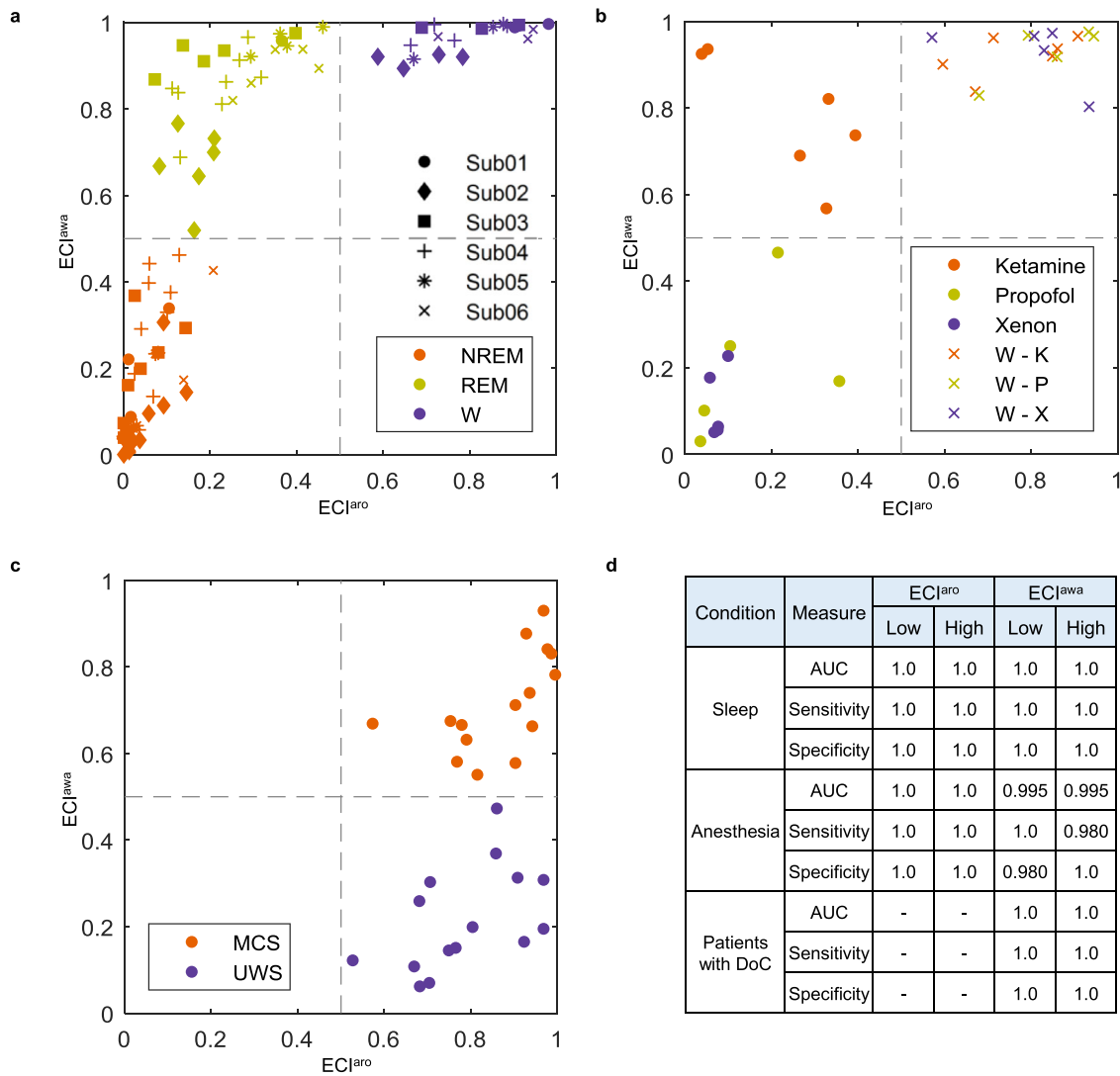
The target domain implies the condition with the target participant to be tested for calculating ECI using CNN with spatiotemporal information, and the source domain implies the conditions to be included in training for learning classifiers.

transfer learning (Table 3). This result was similar to the classification performance of TMS-EEG data, considering that there was no resting-state EEG during sleep in the same participants (see Supplementary Note 4). Based on the results of domain transfer learning, we used only the anesthesia domain for calculating ECI under anesthesia; however, the DoC and

anesthesia domains were used for calculating ECI in patients with DoC.

**ECI in TMS combined with electroencephalography.** Figure 2a shows ECI for each TMS session during sleep and wakefulness. This is a 2D indicator, ranging from 0 (low) to 1 (high) for both arousal and awareness. The cutoff was set to 0.5 for both arousal and awareness, as it is the mean probability for the two-class classification (low versus high). ECI in NREM sleep showed low arousal and awareness, whereas REM sleep had low arousal with high awareness. ECI in healthy wakefulness had both high arousal and awareness. We performed a receiver operating characteristic (ROC) curve analysis and determined that the area under the curve (AUC), sensitivity, and specificity for low and high arousal when using ECI were all equal to 1.0. The AUC, sensitivity, and specificity for low awareness were 0.995, 1.0, and 0.980, respectively, whereas, for high awareness, values of 0.995 for AUC, 0.980 for sensitivity, and 1.0 for specificity were obtained with ECI (Fig. 2d).

We measured ECI using three anesthetic drugs (ketamine, propofol, and xenon) and wakefulness before anesthesia (Fig. 2b).



**Fig. 2 Characteristics of ECI in TMS-EEG.** In ECI, the symbols show the average ECI value of each transcranial magnetic stimulation (TMS) session, and the gray dashed lines indicate the optimal cutoff (0.5) dividing the space into high and low states of ECI. **a** In sleep and normal wakefulness, we depicted P01-P06 by circles, diamonds, squares, plus signs, asterisks, and cross signs, respectively. Orange indicates NREM sleep with no subjective experience; copper indicates REM sleep with subjective experience (i.e., dreaming); moreover, purple indicates normal wakefulness. **b** In anesthesia and wakefulness before anesthesia, the orange, copper, and purple dots indicate the use of ketamine, propofol, and xenon, respectively. In addition, cross markers indicate normal wakefulness before each anesthetic. **c** In patients with disorders of consciousness, the orange and purple dots indicate patients in the minimally conscious state (MCS) and with unresponsive wakefulness syndrome (UWS), respectively. **d** In each condition, the classification performance for ECI<sup>aro</sup> and ECI<sup>awa</sup> was measured. W normal wakefulness; W - K = healthy wakefulness before ketamine; W - P = healthy wakefulness before propofol; W - X = healthy wakefulness before xenon.

ECI in ketamine-induced anesthesia demonstrated low arousal and high awareness, whereas propofol- and xenon-induced anesthesia showed low arousal and awareness. Also, all periods of wakefulness showed high arousal and awareness. As a result of the ROC analysis, classification using ECI achieved an AUC, sensitivity, and specificity of 1.0 for all parameters for both arousal and awareness (Fig. 2d).

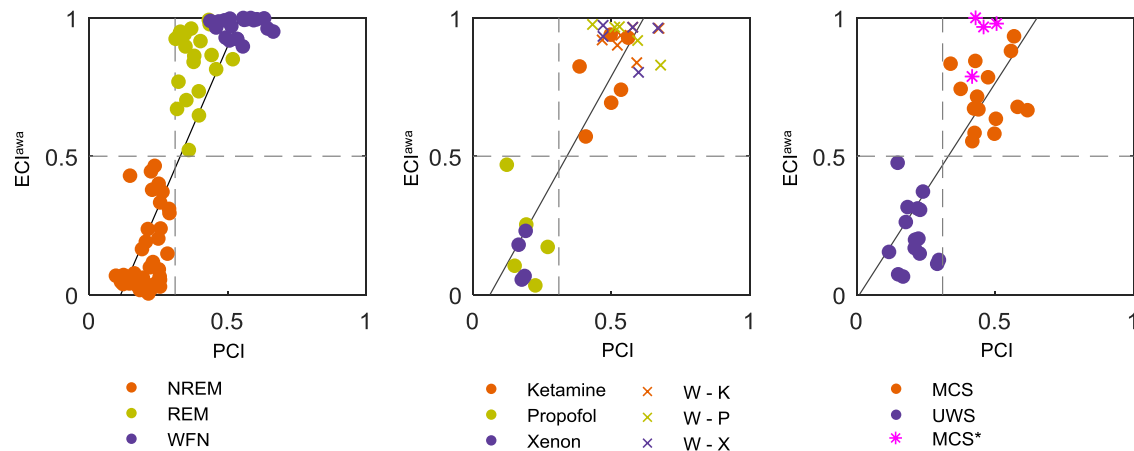
In patients with DoC, ECI indicated high arousal and awareness in MCS patients and high arousal and low awareness in UWS patients (Fig. 2c). AUC, sensitivity, and specificity of ECI were 1.0 for both high and low states of awareness (Fig. 2d). ROC analysis for arousal was not conducted because both UWS and MCS patients were considered to have high arousal. We additionally applied ECI to four MCS\* patients; these cases were not included in the training phase. Similar to patients with DoC, anesthesia and DoC domains were selected as source domains in

the training phase. For the MCS\* patients, ECI successfully predicted high arousal and awareness, as expected (Supplementary Fig. 5).

**Relationship with PCI.** We calculated PCI in TMS-EEG sessions that had at least 80 trials with healthy participants under sleep, anesthesia, and brain-injury conditions. PCI in all three conditions was consistent with the optimal cutoff (0.31<sup>16,17</sup>) that maximizes the accuracy of the distinction between consciousness and unconsciousness in a benchmark population. For ECI, the optimal cutoff of 0.5 perfectly distinguished low or high states of arousal and awareness in the physiological, pharmacological, and pathological conditions (see Supplementary Note 5).

We investigated the relationship between ECI<sup>awa</sup> and PCI (Fig. 3). During sleep, a positive correlation between ECI<sup>awa</sup> and





**Fig. 3 Correlation between  $ECI^{awa}$  and PCI from TMS-EEG.** During sleep and healthy wakefulness (left), under anesthesia, and wakefulness before anesthesia (middle), and in severely brain-injured patients (right),  $ECI^{awa}$  was compared to PCI. The gray horizontal and vertical dashed lines represent the optimal cutoff of  $ECI^{awa}$  and PCI to discriminate between low and high awareness, respectively. The solid lines represent linear fits to the data. W healthy wakefulness, W - K healthy wakefulness before ketamine, W - P healthy wakefulness before propofol, W - X healthy wakefulness before xenon, MCS patients in a minimally conscious state, UWS patients with unresponsive wakefulness syndrome, MCS\* non-behavioral MCS.

PCI was observed ( $r = 0.872$ ,  $p < 0.001$ ). Similarly,  $ECI^{awa}$  during anesthesia and in brain-injured patients showed a strong correlation with PCI (anesthesia:  $r = 0.885$ ,  $p < 0.001$ ; brain injury:  $r = 0.770$ ,  $p < 0.001$ ).  $ECI^{awa}$  and PCI, therefore, matched for all states.

**ECI in resting-state electroencephalography.** ECI results using resting-state EEG data were similar to those when using TMS-EEG results (Fig. 4). In anesthesia, ECI using resting-state EEG statistically correlated with ECI using TMS-EEG ( $ECI^{aro}$ :  $r = 0.848$ ,  $p < 0.001$ ;  $ECI^{awa}$ :  $r = 0.938$ ,  $p < 0.001$ ). Similarly, in patients with DoC, there was a positive correlation ( $ECI^{aro}$ :  $r = 0.534$ ,  $p = 0.02$ ;  $ECI^{awa}$ :  $r = 0.832$ ,  $p < 0.001$ ). Similarly, we applied this method to four MCS\* patients to verify ECI. As expected, similar to TMS-EEG results, accurate predictions were obtained in four MCS\* patients (Supplementary Fig. 5).

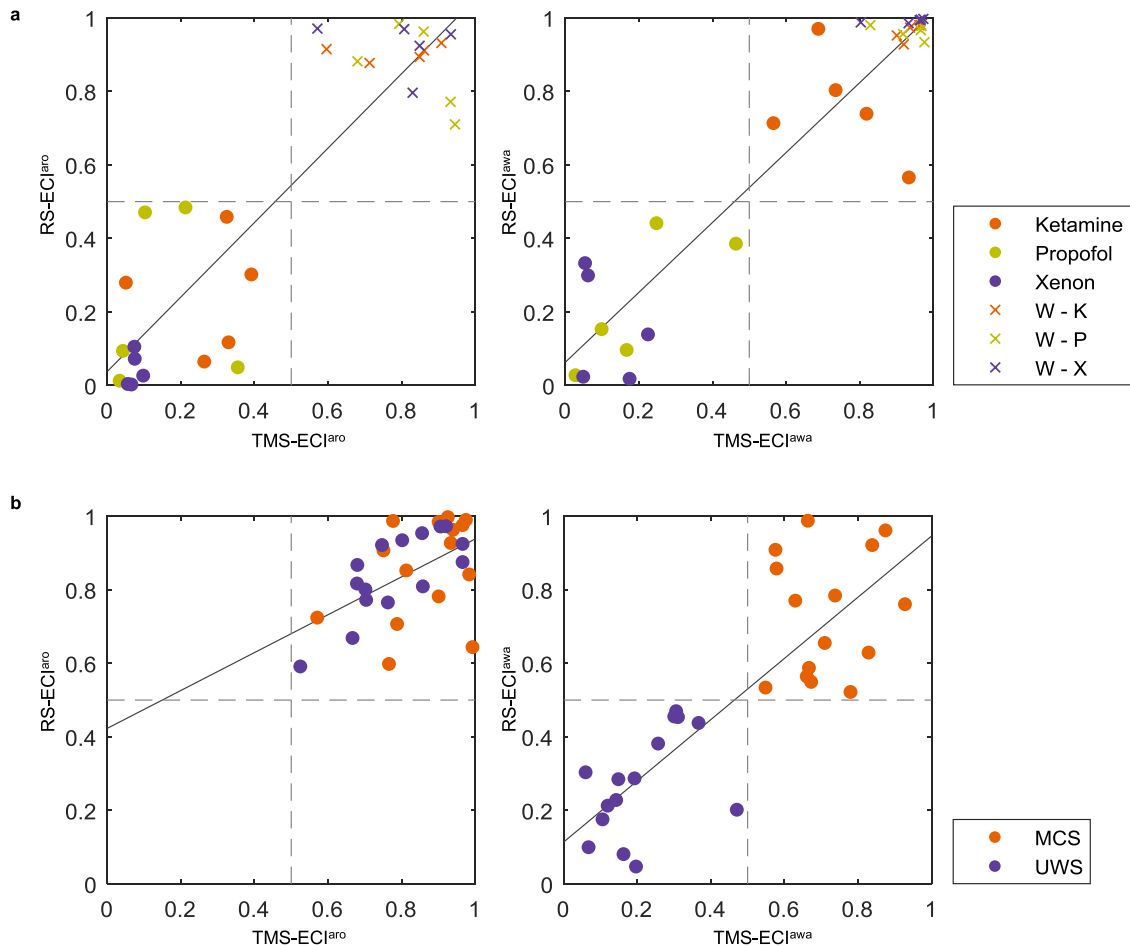
**The practicality of calculating ECI.** We explored the possibility of calculating ECI with a limited number of trials during sleep and wakefulness. The classification performance was compared through the calculation of ECI from a single trial up to the standard number of trials (i.e., 80 trials similar to  $PCI^{16}$ ) (Fig. 5). In the sleep and healthy wakefulness conditions, the performance reached a specificity of 0.853, a sensitivity of 0.884, and an AUC of 0.931 when using single trials (based on one TMS pulse), but from 2 trials, it was above 0.9 for specificity, sensitivity, and AUC. In the anesthesia and patients with DoC conditions, the detailed performance from 1 to 80 trials is shown in Supplementary Note 6.

Figure 6 shows the possibility of awareness being high for the first participant in each of these conditions: sleep, anesthesia (ketamine, propofol, and xenon), and patients with DoC (UWS and MCS). For instance, as NREM sleep is considered to have low awareness, in P01, it was correctly predicted when the probability in a single trial was less than 0.5. Notably, NREM sleep showed that 17 trials were incorrectly predicted with a probability higher than 0.5. In addition, 1 out of 80 trials in both REM sleep and healthy wakefulness showed a value less than 0.5 and were incorrectly predicted. This indicates that ECI can be predicted as somewhat low or high, even in a single trial. There was a clear spatiotemporal difference between correct and incorrect trials only in parietal regions (Supplementary Fig. 6). No significant differences between both types of trials were observed over frontal

and temporal regions. However, in parietal regions, TMS-evoked potentials at 350–400 ms were significantly higher in incorrect trials than in correct trials. In a single trial, these different patterns resulted in misprediction. Nevertheless, the effect of the incorrect trials (i.e., a failure to predict) was eliminated because ECI was calculated by averaging the interclass probability in a single trial. The probability in a single trial is shown in Supplementary Figs. 7–9 for other participants in all conditions.

To demonstrate that the method can be easily applied to a new set of patients (without additional training) to identify their state of consciousness in a clinical setting, we computed ECI using the hold-out approach. The dataset in the patients with DoC was split between the training and evaluation sets with respective ratios of 0.75 and 0.25. In other words, two MCS patients and five UWS patients were completely excluded from training, and their ECI was calculated. Consequently, ECIs using conventional LOPO and the hold-out approaches showed high positive correlation ( $ECI^{aro}$ :  $r = 0.702$ ,  $p = 0.005$ ;  $ECI^{awa}$ :  $r = 0.886$ ,  $p < 0.001$ ) (Supplementary Figure 10). In both TMS-EEG and resting-state EEG, the ECI using the hold-out method indicated high arousal and awareness in MCS patients, whereas high arousal and low awareness in UWS patients. This was consistent with 0.5 typical cutoffs. These results show that the proposed method generalizes to new data without retraining the classifier.

**Interpretation for calculating ECI.** We further checked what the classifier learned through CNN and how it was able to derive those results using LRP. This algorithm describes the predictions of CNN in a given dataset using relevance scores<sup>30</sup>. Figure 7 shows the relevance scores for arousal and awareness among the frontal, temporal, and parietal regions at the scalp level for calculating ECI using TMS-EEG. A high relevance score implies that the trained model recognized the brain region that determines whether it is low or high arousal and awareness. Specifically, brain regions with higher relevance scores indicate that brain signals over that region contributed more to the decision of the classifier on whether arousal and awareness were low or high. In the three conditions (sleep, anesthesia, and patients with DoC), the relevance score over the parietal region was higher than those in the frontal and temporal regions at the group level for both arousal and awareness. The detailed statistical results are reported in Supplementary Table 5. However, since most TMS sites targeted the parietal cortex, it can be argued that the relevance of



**Fig. 4 Relationship between ECI using TMS-EEG and resting-state EEG.** The x-axis represents ECI using TMS-EEG, and the y-axis represents ECI using resting-state EEG for arousal (left) and awareness (right). The gray dashed lines indicate the optimal cutoff (0.5) dividing the space into high and low states of ECI. The solid lines represent linear fits to the data. **a** In anesthesia and wakefulness before anesthetics, the orange, copper, and purple dots indicate the use of ketamine, propofol, and xenon, respectively. In addition, cross markers indicate normal wakefulness before each anesthetic. **b** In patients with disorders of consciousness, the orange and purple dots indicate MCS and UWS patients, respectively.

parietal regions to correctly classify datasets may be biased. Therefore, we investigated the relevance scores with only non-parietal stimulations in patients with DoC. As a result, the parietal region had statistically higher relevance scores than the frontal and temporal regions in both arousal and awareness (Supplementary Fig. 11 and Table 6). Similarly, there was a higher relevance score over the parietal region in both arousal and awareness when calculating ECI from resting-state EEG (Supplementary Fig. 12 and Table 7).

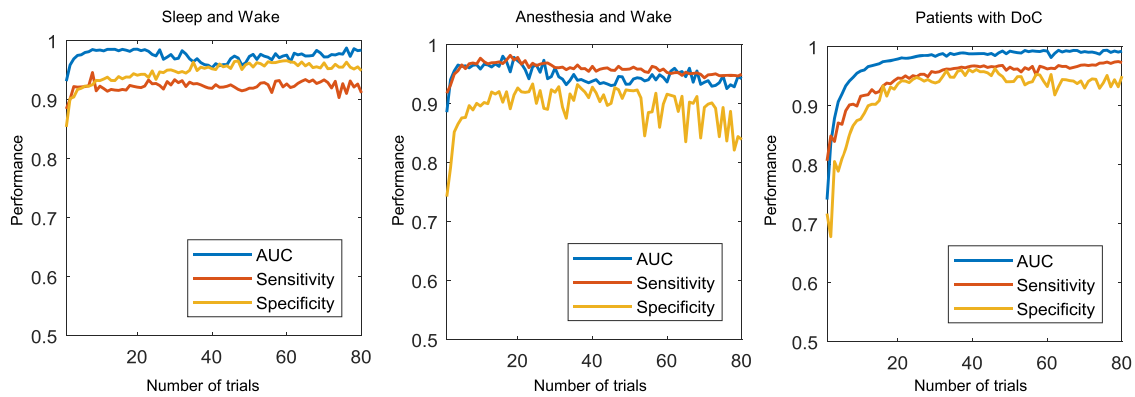
We additionally compared the classification performance of  $ECI^{awa}$  among patients with DoC using TMS-EEG data when we excluded electrodes in different brain regions from the input during classification. Consequently, AUC was 1.0 when using all electrodes; however, AUC values were 0.867 and 0.680 when removing frontal and parietal electrodes, respectively (Supplementary Fig. 13).

## Discussion

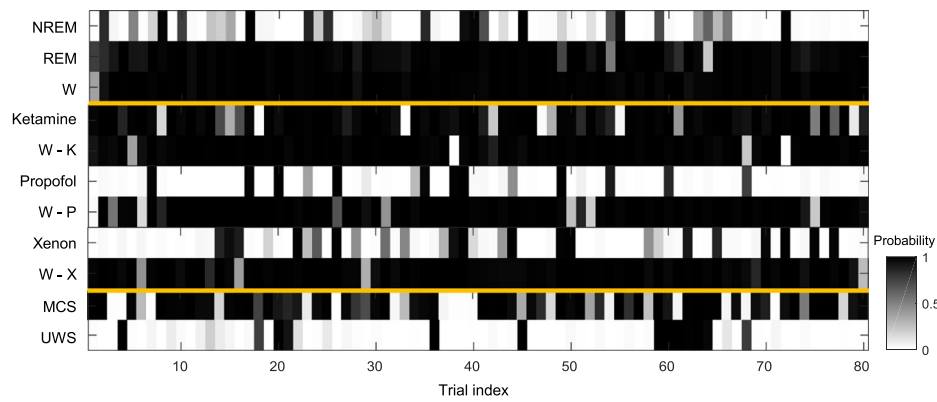
We show that ECI clearly distinguishes between low and high states of arousal and awareness in TMS-EEG results under sleep, anesthesia, and patients with DoC. Our results suggest that this proposed indicator could similarly be used for resting-state EEG data without TMS under anesthesia and patients with DoC, yielding the same degree of accuracy. In addition, a high correlation with PCI, which measures the integrated EEG response of

the thalamocortical system to a direct perturbation induced by TMS<sup>16</sup>, proves that  $ECI^{awa}$  is reliable using TMS-EEG data. It also shows that the two measures calculated entirely independently using different methods resulted in the same conclusion, which is a sign that deep learning is indeed a valid and reliable approach. For an ML-based indicator, ECI can be calculated using very few trials. Furthermore, because the classifier learned specific features of the data on its own, our indicator can be computed regardless of whether TMS is applied or its location. Therefore, ECI is a significantly practical and reliable indicator to evaluate levels of consciousness under various conditions. Our analyses using LRP highlighted the major role of the parietal region in determining consciousness, as the classifier primarily uses brain activity in this lobe for predicting low and high states of arousal and awareness.

TMS-EEG responses under sleep exhibited well-known phenomena<sup>31</sup>. In wakefulness, TMS generates a series of low-amplitude high-frequency activities related to cortical flow in long-range connections<sup>32</sup>. A similar long-lasting response is evoked during REM sleep with subjective experience<sup>33</sup>. During NREM sleep with no subjective experience, TMS triggers larger, low-frequency activity that quickly dissipates<sup>32</sup>, which is the hallmark of bistability in the thalamocortical system. Cortical effective connectivity is also broken down during NREM sleep<sup>33,34</sup>. The brain response to TMS perturbation was already



**Fig. 5 Performance of  $ECI^{awa}$  according to the number of trials in the ECI calculation.** During sleep and healthy wakefulness (left), under anesthesia and wakefulness before anesthesia (middle), and in severely brain-injured patients (right). The area under the curve, sensitivity, and specificity were measured for calculating  $ECI^{awa}$  when going from single trials to 80 trials.



**Fig. 6 Interclass probability in the representative participant for  $ECI^{awa}$ .** We depicted the probability of a representative participant (P01) in all conditions (sleep, anesthesia, DoC). P01 was the first participant of each list, randomly chosen. Each colored box indicates the probability that the corresponding trial is considered as high awareness in each participant. If it was a perfect prediction in one trial, during sleep and healthy wakefulness, NREM sleep with no subjective experience (low awareness) has a probability of less than 0.5, whereas REM sleep and normal wakefulness (high awareness) have probabilities of more than 0.5. Under anesthesia and wakefulness, ketamine and wakefulness before anesthesia are high awareness, whereas propofol and xenon have low awareness. MCS and UWS patients have high and low awareness, respectively.

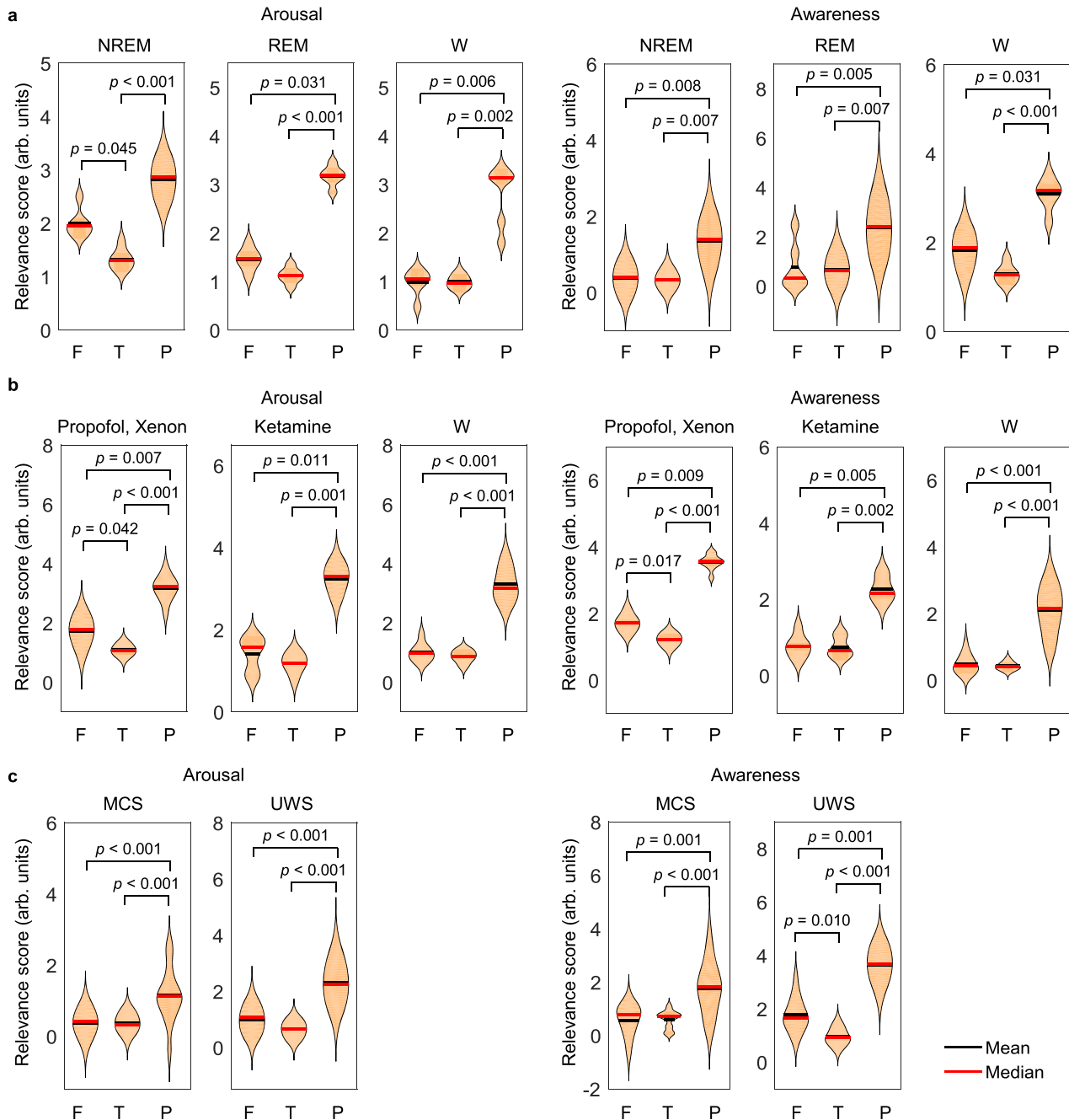
used to distinguish the levels of awareness, irrespective of sensory processing and motor responses under physiological, pharmacological, and pathological conditions<sup>34–36</sup>. Therefore, this TMS-evoked response was applied to our end-to-end CNN framework. Similar to several studies using ML, we observed higher two-class (low or high) classification accuracy using CNN when compared to linear discriminant analysis (LDA) and support vector machine (SVM) for both arousal and awareness. This suggests that our framework is especially relevant for EEG results, which possess several nonlinear features. The classification performance of spatiotemporal information was higher than that of spatio-spectral information. Using our framework, it was shown that temporal information discriminates different levels of consciousness more clearly than spectral information, as the functional connectivity associated with consciousness changes in both space and time<sup>37</sup>. However, this does not imply that temporal information is more important than spectral information for distinguishing consciousness. Temporal information has more distinct characteristics than spectral information for predicting the state of consciousness in the proposed framework. Nevertheless, PCI also used spatiotemporal dynamics in TMS-evoked responses<sup>16</sup>, which is significantly important for distinguishing consciousness.

We applied transfer learning to a single domain as well as multiple domains. In the sleep domain, classification performance

was high when trained on only the sleep domain or together with the anesthesia domain. Due to a large number of trials in the sleep domain, training was performed satisfactorily using only sleep data. The distance of the averaged TMS-evoked potentials under the domains of anesthesia and sleep was significantly close, which indicated that the two domains were highly similar<sup>38</sup>. Thus, a close distance implies that the two domains have similar patterns, and the classification performance indeed increased when these domains were trained together. In the anesthesia and DoC domains, the trials of a single domain were not sufficient; thus a higher classification performance was achieved when trained with similar domains. In addition, brain signals change over time, even when recorded with the same participants because of physiological and psychological differences over time<sup>27</sup>. Therefore, participant-independent learning such as the LOPO cross-validation is considerably difficult, as opposed to participant-dependent learning<sup>24</sup>. We solved these problems via transfer learning utilizing multiple domains and participant-independent ECI.

ECI, by averaging interclass probability, distinguished whether each state was low or high in both arousal and awareness. Because of its inherently high two-class performance, it was possible to calculate the discriminable ECI using a few trials in a single session. The criteria used for determining  $ECI^{aro}$  and  $ECI^{awa}$  were different, especially for REM sleep, ketamine-induced anesthesia,





**Fig. 7** Relevance scores from LRP in TMS-EEG. **a** Sleep and normal wakefulness, **b** anesthesia and wake, and **c** patients with DoC. The violin plots depict the average relevance scores over the frontal, temporal, and parietal regions in all participants. The exact  $p$ -value corresponding to the significance level was shown using two-sided multiple  $t$ -tests with Fisher’s least significant differences method for multiple comparisons. [arb. units] denotes an arbitrary unit. F frontal region, T temporal region, P parietal region.

and UWS patients. Although the dataset was the same, these states were trained by different labels depending on arousal and awareness. For instance, REM sleep is high in awareness but low in arousal. The proposed classifier learns models for these different criteria by training itself based on the criteria of arousal and awareness. Specifically, because the learning ability of the CNN is derived from the automatic extraction of complicated representations from EEG signals<sup>39</sup>, it can properly distinguish between both states, even if the same state has different labels depending on the criteria. We further used data from MCS\* patients as verification samples. MCS\* patients were correctly

predicted, since data from UWS and MCS patients were included during training, and the classifier is independent of behavior.

According to the LRP, we observed higher relevance scores in the parietal regions, compared to other regions at the scalp level. This was observed for all data, including TMS-EEG and resting-state EEG results. The brain regions that led to decision-making were similar in arousal and awareness. However, this does not imply that arousal and awareness are supported by the same underlying neurophysiological mechanisms. The relevance scores simply explain the patterns of cortical EEG activity resulting in this classification.

In sleep and healthy wakefulness, when arousal and awareness were high, they were highly relevant in the parietal region. This EEG feature, which distinguishes high and low states in arousal and awareness, can be interpreted in line with the posterior hot zone of consciousness<sup>40</sup>. Local changes in this parietal region are associated with the occurrence of dreaming and unconscious sleep<sup>41,42</sup>, and our framework may learn from EEG pivotal features recorded in this area. The importance of the posterior hot zone has already been emphasized using the within-state sleep paradigm<sup>43,44</sup>. Similar to sleep, we observed high relevance scores for the parietal region in the domains of both anesthesia and patients with DoC. This implies that EEG activity in this region had a decisive effect in determining the high and low states of arousal and awareness. The increased slow-wave activity was observed under propofol- and xenon-induced anesthesia when compared to healthy wakefulness before drug administration<sup>36</sup>. In addition, just as cortical neurons induced bistable changes during NREM sleep, TMS in propofol-mediated anesthesia-induced low-amplitude, low-frequency, positive-negative potentials, and TMS in xenon-mediated anesthesia caused a significantly large amplitude but stereotyped positive-negative deflection<sup>36</sup>. Moreover, under ketamine-induced anesthesia, TMS-evoked response was determined to be similar to REM sleep, which features dreaming during a low state of arousal<sup>36</sup>. Previous studies have shown that the change in slow waves induced by propofol is primarily observed in the posterior hot zone<sup>45</sup> and the posterior main hub is disrupted during anesthesia-induced alteration of consciousness<sup>7</sup>. In UWS patients, TMS triggered a local and slow response similar to NREM sleep and general anesthesia, whereas MCS patients showed complex TMS-evoked responses<sup>33</sup>. Similarly, differences in alpha connectivity between the UWS and MCS patients are apparent within the posterior hot zone<sup>46</sup>. That is, when determining whether arousal and awareness are low or high, our classifiers used differences and changes in EEG activity over the parietal region to make decisions. Particularly, similar results were observed in patients with DoC using several TMS target sites, and in sleep and anesthesia domains, where the parietal region was primarily stimulated. This suggests that our findings regarding the parietal region are unrelated to the TMS target site. Thus, our trained model used neurophysiological features to classify whether arousal and awareness are low or high. This indicates an appropriate design of our model as the classification decision was primarily based on EEG signals over the parietal region, which is suggested to be a hot spot of consciousness, compared to the frontal region<sup>43</sup>. The difference in the parietal region could be clearly identified through correct trials during sleep and healthy wakefulness. It is meaningful that the frontal region contributed less than parietal regions in the context of the controversy regarding the spatial localization of the neural correlates of consciousness<sup>40,45</sup>. Considering the subcortical influences related to striatal-thalamic circuits, it has been recently observed that the parietal region contributes more to the levels of consciousness than the frontal region<sup>47</sup>. The implication of the parietal cortex in consciousness has also been demonstrated in other neuroimaging modalities, such as functional magnetic resonance imaging<sup>48</sup> and magnetoencephalography<sup>49</sup>.

This study does have certain limitations. First, our sample size was relatively small. In the future, it will be necessary to further test the reliability of the proposed indicator with larger cohorts and validate it at the clinical level before implementing it in a clinical setting. Sleep experiments would also have to be applied to more participants in the future. Second, we explored the possibility of calculating ECI with a minimal amount of data, up to a single trial. However, we did not attempt to measure ECI in real time. Thus, in the future, ECI could be calculated in real-time

for practical application. Third, ECI does not differentiate between physiological, pharmacological, and pathological conditions, but distinguishes between high and low states of arousal and awareness. ECI can thus distinguish between REM sleep (or ketamine) and wakefulness. It may also be difficult to select the model to use when calculating ECI since the domain has to be known beforehand. Nevertheless, if a single domain has sufficient trials, the LOPO approach would be the most accurate. Another limitation might be related to the possible contamination of TMS-EEG data by auditory and somatosensory components. As in previous studies<sup>17,35,36</sup>, to avoid auditory and somatosensory co-stimulation, participants wore earphones with noise masking and a thin foam between the scalp and the TMS coil that was used. Although it is difficult to systematically rule out the contribution of sensory co-stimulation in every measurement, the application of effective noise-masking procedures<sup>50</sup> and the real-time monitoring of data quality during the acquisition<sup>51</sup> may significantly mitigate this issue<sup>52</sup>. Finally, ECI indicates whether arousal and awareness are low or high and cannot be considered functional. It, therefore, should be developed into a functional index.

In conclusion, we proposed ECI as a neurophysiological indicator to simultaneously discriminate the levels of arousal and awareness in modified states of consciousness. This tool allows disentangling the levels of consciousness, with a single measure, in different clinical settings such as monitoring surgical interventions (i.e., anesthesia-induced states) and diagnosing patients with DoC. This indicator was validated under different physiological, pharmacological, and pathological conditions, and it reliably disentangled low levels from high levels of both arousal and awareness. Besides, the proposed ECI is considerably accessible and practical, as it can be applied to resting-state EEG without TMS, and requires fewer trials. Therefore, the proposed indicator can be a reliable discriminator and valuable tool as an objective measure of consciousness. As parietal regions appear to be the most relevant for classification, an EEG configuration around that area could be sufficient if ECI is used in clinical practice. These findings could be useful in diagnosing severely brain-injured patients and monitoring their levels of consciousness in real-time, especially in clinical settings where time constraints preclude long-duration assessment. The proposed reliable ECI can provide insights into the classification of conscious levels using deep learning and neural correlates of consciousness.

## Methods

**Datasets.** The sleep dataset included six healthy participants (five males, aged  $23.7 \pm 3.2$  years), as previously reported by Nieminen et al.<sup>43</sup> for NREM data and Lee et al.<sup>53</sup> for REM data. The inclusion criteria included (i) between 18 and 75 years of age and (ii) in good general health. The exclusion criteria were as follows: (i) neurological, psychiatric, mood, and sleep disorders, (ii) contraindications for TMS (e.g., history of seizures), and (iii) psychotropic medication. All participants provided written consent, and the experimental paradigm was approved by the Institutional Review Board (IRB) at the University of Wisconsin–Madison (HSC-2013-0019). Sleep stages were manually scored every 30 s following the American Academy of Sleep Medicine Scoring Manual. After 3 min or more, when the participant entered a specific sleep stage, TMS was applied over the parietal cortex using a navigated brain stimulation system (eXimia Navigated Brain Stimulation, Nexstim Plc, Finland). Supplementary Table 8 lists the TMS target site and the number of sessions and trials. The participants were awoken by an alarm sound that lasted for 1.5 s after each session. They were then asked if they had had conscious experience. The TMS-EEG experiments were performed over a period of four or five nights per participant.

The anesthesia data were previously published by Sarasso et al.<sup>36</sup>. Sixteen healthy participants (eight males, aged 18–28 years) were included under ketamine ( $n = 6$ ), propofol- ( $n = 5$ ), and xenon-induced ( $n = 5$ ) anesthesia. The inclusion criteria included (i) older than 18 years and (ii) stability of vital parameters. The exclusion criteria were as follows: (i) neurological, cardiovascular, psychiatric, and mood disorders, (ii) contraindications for TMS (e.g., history of seizures, metal implants such as a pacemaker), and (iii) medical conditions that were incompatible with the anesthesia and/or the TMS procedure. This experimental protocol was approved by IRB at the University of Liège (2009/153 (ketamine), 2007/191

(propofol), and 2009/242 (xenon)); moreover, all participants provided written informed consent. TMS was applied over the left parietal or motor regions after participants reached deep unresponsiveness (a score equal to 5 in the Ramsay scale, which corresponds to no response to external stimuli) following standard anesthetic procedures<sup>36</sup>. The stimuli target site and the number of trials are listed in Supplementary Table 9. In addition, upon waking up from anesthesia, reports about the conscious experience during anesthesia were collected. Conclusively, the participants reported little conscious experience during propofol- and xenon-induced anesthesia, but vivid dreams were experienced during ketamine-induced anesthesia.

For patients with severe brain injury, the data of six UWS patients (2 males, 4 traumatic brain injuries, time since the injury of 10.6 months (1–47), age:  $36.2 \pm 28.6$  years) and ten MCS patients (7 males, 5 traumatic brain injuries, time since the injury of 65.7 months (1–343), age:  $44.6 \pm 20.5$  years) were previously reported by Bodart et al.<sup>54,55</sup> and Rosanova et al.<sup>56</sup>. This study was approved by the Ethics Committee of the Medicine Faculty of the University of Liège (ref 2009/52) and written informed consent was obtained from legal representatives of all patients. All of them fell into a coma due to brain injury and presented a prolonged state of impaired consciousness. The inclusion criteria included (i) older than 18 years and (ii) diagnosis of DoC following a severe acquired brain injury. The exclusion criteria for patients were as follows: (i) patients having significant neurological, neurosurgical, or psychiatric disorders prior to the brain injury that leads to DoC, (ii) patients having any contraindication to TMS–EEG or magnetic resonance imaging (electronic implanted devices, active epilepsy, external ventricular drain), and (iii) patients who were not medically stable. Accredited experts performed repeated CRS–R for each patient, including on the day of the TMS–EEG examination and before the fluorodeoxyglucose-positron emission tomography (FDG–PET) scan. The FDG–PET is a reliable and sensitive tool to detect MCS\* patients based on the previous literature<sup>15</sup>. MCS\* patients were the patients who were diagnosed with a UWS with the CRS–R at the bedside but diagnosed as an MCS based on the FDG–PET data (that is, patients showing relative metabolic preservation of the frontoparietal network based on a subjective visual assessment of the Statistical Parametric Mapping analysis<sup>15</sup>). TMS–EEG data were acquired similar to previous studies<sup>16,35</sup>. Moreover, added data were newly included as follows: 9 UWS patients (6 males, 5 traumatic brain injuries, time since the injury of 6.2 months (1–13), age:  $41.4 \pm 21.1$  years), 5 MCS patients (4 males, 2 traumatic brain injuries, time since the injury of 63.0 months (2–169), age:  $32.4 \pm 14.0$  years), and 4 MCS\* patients (2 males, 3 traumatic brain injuries, time since injury 18.5 months (3–52), age:  $36.3 \pm 10.5$  years). This study was approved by the Ethics Committee of the Medicine Faculty of the University of Liège (ref 2012/55) and all legal representatives of patients provided written informed consent before the experiments; moreover, newly recorded data used exactly the same procedure as before. These added data were recently acquired by Dr. Olivia Gosseries & Pr. Steven Laureys team at the University of Liège. For the added data, the inclusion and exclusion criteria for patients were exactly the same as the previously used data. These data are part of a bigger study conducted in the frame of the Human Brain Project. The final dataset consisted of 15 UWS patients, 15 MCS patients, and 4 MCS\* patients. The detailed demographic and clinical information of severely brain-injured patients is listed in Supplementary Table 10. The TMS target site was selected using a neuronavigation system over the parietal, motor, or premotor regions, avoiding structural lesions using the magnetic resonance imaging data of the patient. The stimuli target sites for all participants are listed in Supplementary Table 11. The participants remained awake or were kept awake using the CRS–R arousal protocol in between TMS stimulation<sup>12</sup>.

For the four datasets, EEG data were recorded using a 60-channel TMS-compatible amplifier and a two-channel electrooculogram (Nexstim eXimia, Nexstim Plc, Finland) with a 1450 Hz sampling rate. During all the sessions, earphones presenting white noise were used to reduce the noise of the TMS pulses and we used a thin foam between the scalp and the TMS coil to avoid somatosensory evoked potentials. These pulses were presented at random intervals of 2–2.3 s using a figure-of-eight coil. The maximum electric field was between 100 and 130 V/m at the TMS target site. In particular, the TMS stimulation was always performed in the medial half of one hemisphere, to avoid any muscle artifacts.

Finally, we used resting-state EEG without TMS for the domains of anesthesia and severely brain-injured patients. These data were acquired using the same participants as in the TMS–EEG experiments. We used the ketamine- ( $n = 5$ ), propofol- ( $n = 5$ ), and xenon-induced anesthesia data ( $n = 5$ ) as previously reported by Sarasso et al.<sup>36</sup>. For at least 3 min before the TMS–EEG experiments, resting-state EEG data were recorded during anesthesia and each state of wakefulness before the anesthesia. With regard to the severely brain-injured patients, 5-min resting-state EEG for MCS patients ( $n = 15$ ) and UWS patients ( $n = 15$ ) was included. In addition, four MCS\* patients were added along with the same participants for whom TMS–EEG was recorded. The sampling rate was 1450 Hz. Finally, among previously published data, the data with a signal-to-noise of 1.4 or less were excluded from the analysis<sup>16</sup>.

**Data preprocessing.** TMS–EEG data were preprocessed using the SiSyPhus Project MATLAB program (University of Milan, Italy) and the EEGLAB toolbox<sup>57</sup>. The signals were down-sampled to 362.5 Hz and band-pass filtered between 0.5 and 45 Hz using a second-order Butterworth filter. The signals of –400 to 1000 ms

were segmented and baseline-corrected using the 400 ms baseline before the TMS pulses. Bad channels were manually detected and interpolated using superfast spherical interpolation for artifact removal. We also discarded the components related to eye movements using independent component analysis and removed trials setting a threshold of  $\pm 100 \mu\text{V}$  affected by ocular artifacts, other artifacts, or noise. The data were re-referenced to an average reference<sup>58,59</sup>.

Resting-state EEG data were processed using the EEGLAB toolbox<sup>57</sup>. Preprocessing was performed using a process similar to that used with TMS–EEG data, with segmentation being performed every 1 s. The number of trials for each session we used is listed for patients under anesthesia (Supplementary Table 12) and severely brain-injured patients (Supplementary Table 13).

### Proposed framework for calculating an ECI. Step 1—Extraction of EEG features:

In all trials of all TMS–EEG data, we used the 200–400 ms time window of data after the TMS regardless of the lateralization or target site of the TMS in sleep, anesthesia, and for patients with severe brain injury. More details related to this deliberate choice are reported in Supplementary Note 2. In resting-state EEG, only the first 200 ms of data were used from the segmented 1 s of data. In the first step, EEG data were converted from 2D raw signals to 3D input. To preserve the spatial information and characteristics of EEG, we used spatio-spectral and spatiotemporal 3D features. The raw EEG signals at time index  $t$  are measured in a 1D data vector  $r_t = [s_t^1, s_t^2, s_t^3, \dots, s_t^n]^T$ , where  $s_t^i$  is the acquisition data by the  $i$ th electrode channel at timestamp  $t$ .  $n$  indicates the number of electrode channels. However, these simple signals do not capture all the spatial information characteristics in the brain. Therefore, we converted 1D data vectors to 2D EEG data meshes using the spatial information of the electrode location. Zero was inserted in the place of a null electrode in 2D matrices at time index  $t$ <sup>60</sup>. Finally, we calculated 3D data by adding spectral or temporal information (Supplementary Fig. 14). Specifically, spectral information was divided into 5 frequency bands: delta (1.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma bands (30–40 Hz)<sup>44</sup>. Finally, a  $10 \times 11 \times 5$  matrix using spectral information and a  $10 \times 11 \times 72$  matrix using temporal information were used as the final CNN inputs (here,  $10 \times 11$ : the converted matrix of spatial information; 5: delta, theta, alpha, beta, and gamma bands of spectral information; 72: 200–400 ms of temporal information).

Step 2—Calculation of interclass probability using CNN: The model was trained to distinguish both arousal and awareness in terms of whether they were low or high. We first calculated the domain similarity based on the cosine distance for domain transfer learning. The cosine distance  $D$  between domains  $A$  and  $B$  is defined as follows:

$$D(A, B) = 1 - \frac{f_A^T f_B}{\|f_A\| \|f_B\|} \quad (1)$$

where  $\|\bullet\|$  indicates the norm of a vector. This similarity was used when selecting the source domain for domain transfer learning<sup>38</sup>. In calculating the similarity according to the states of arousal and awareness, the labels in each state are different. During sleep, NREM sleep with no subjective experience and REM sleep with subjective experience were learned as low arousal, whereas healthy wakefulness was learned as high arousal. Conversely, in awareness, REM sleep with subjective experience and healthy wakefulness indicated by open eyes were learned as high awareness, while NREM sleep with no subjective experience was learned as low awareness. Specifically, we used stage 3 of NREM sleep for clear feature extraction of deep sleep. Under general anesthesia and in healthy wakefulness before anesthesia, the arousal state was divided as follows: (i) low, when under ketamine, propofol, and xenon-induced anesthesia, and (ii) high, when in healthy wakefulness before anesthesia. Conversely, the awareness state was divided into low (under propofol and xenon-induced anesthesia) and high (under ketamine-induced anesthesia and healthy wakefulness before anesthesia). Finally, in patients with DoC, UWS and MCS patients were distinguished in terms of awareness: low (UWS patients) and high (MCS patients). The UWS and MCS patients corresponded to high arousal.

Deep learning was conducted in a MATLAB environment powered by a TITAN V GPU. We used the LRP toolbox<sup>30</sup> for CNN classification and interpretation. The CNN was applied to the two components of consciousness (arousal and awareness). In each architecture, we inserted five convolutional layers with 2D filters for the deep neural network. The first layer with 100 filters and the second layer with 80 filters featured kernel sizes of  $3 \times 3$  and  $2 \times 2$  with stride  $1 \times 1$ , respectively. Then, a max-pooling layer with a pool size of  $2 \times 2$  and stride  $1 \times 2$  was added. Similarly, two convolutional layers with kernel sizes of  $3 \times 3$  (with stride  $1 \times 1$ ) and  $2 \times 2$  (with stride  $2 \times 1$ ) were subsequently used. After max-pooling with a pool size of  $2 \times 1$  and stride  $2 \times 1$ , a final convolutional layer comprising two filters with a kernel size of  $1 \times 1$  and stride  $1 \times 1$  was incorporated. Finally, the generated feature maps were flattened into a 1D vector. A softmax layer was used for classification. In the softmax layer, each element indicates the probability that the original input belongs to the corresponding class. In this training procedure, the parameters in the deep neural network were learned through back-propagation. The activation function in each convolutional layer was a rectified linear unit. The detailed CNN architecture is presented in Supplementary Table 14. The Adam optimizer was used with an initial learning rate of 0.005, and the learning rate was updated to sublinear for learning rate decay during an evaluation step of training. Specifically, we used hyperparameters as values of 0.9 for  $\beta_1$ , 0.999 for  $\beta_2$  for Adam



optimizer<sup>61</sup>. The batch size was 25 for training, and the maximum number of training iterations was five times the number of training data. Consequently, the output of this architecture was the probability of each class.

For a comparison with other classifiers, we also considered an LDA classifier<sup>62</sup> and SVM with polynomial kernel function<sup>63</sup> using the same input data as fair baseline methods. The classification performance was measured in LOPO-nested cross-validation for the generalized neural network. This method is a special case of  $k$ -fold cross-validation. Specifically, all participants except one (the target participant) were used for training, and the target participant was then tested using the classifier. In the training phase, 80% of the datasets were used to learn the classifier, and the remaining 20% were reserved for validation. This process was repeated for each participant. Further, internal validation sets (inner cross-validation) were performed to choose the hyperparameters of the model<sup>64</sup>. Thus, the same hyperparameters were selected in the external validation sets<sup>65</sup>. The LOPO cross-validation procedure uses data efficiently and can reduce overfitting. It also provides unbiased estimates of the averaged classification error for all possible training sets<sup>66</sup>. The same LOPO-nested cross-validation is applied in LDA and SVM using the Berlin brain-computer interface toolbox<sup>67</sup>. Three possible values (0.0001, 0.01, and 1) were chosen as penalties for misclassification in the SVM model<sup>65</sup>. For spatio-spectral input and spatiotemporal input 1 and 0.0001 were used respectively.

**Step 3—Calculation of ECI in a single session:** We obtained each interclass probability according to two components (arousal and awareness) at the previous step. In each TMS session, the interclass probability was averaged to calculate an ECI. This approach has the advantage of being able to offset outliers.

$$C_j = \frac{1}{N} \sum_{i=1}^N p_i \quad (2)$$

Here,  $p_i$  is the probability of high arousal or high awareness from each trial  $i$  among the probability values (high versus low) for the two classes from the softmax function in the CNN.  $N$  is the number of trials in a single session  $j$ . The averaged interclass probability  $C_j$  from arousal is the value of the  $x$ -axis as ECI<sup>aro</sup>, and the averaged interclass probability  $C_j$  from awareness becomes the value of the  $y$ -axis as ECI<sup>awa</sup>. Consequently, ECI is expressed as a 2D indicator representing both arousal and awareness simultaneously. As mentioned earlier, we used the LOPO cross-validation. It is to be noted that only data from one participant were used as the test, and the data from the remaining participants were used for training. Finally, the test and training data did not demonstrate an overlap at all.

**Step 4—Interpretation using LRP:** We used LRP based on a backward-propagation mechanism for interpretability of the deep neural networks. This calculated the pixel-based decomposition process<sup>28</sup>.

$$\sum_{p=1}^d R_p = f(x) \quad (3)$$

Here,  $x = (x_1, \dots, x_d)$  indicates an input vector and  $f(x)$  is the model output. The relevance score  $R_p$  is the decomposition of the prediction for the input  $x_p$ . This score is calculated through the backward propagation of the model input. Therefore, relevance scores describe a single nonlinear decision for the output corresponding to each input<sup>68</sup>. Through this method, we can observe not only the interpretation of classification decisions but also what features the model has learned<sup>28</sup>. To investigate which brain regions and signals caused these classification results, we compared the relevance scores from the LRP by dividing them into the following three regions<sup>69</sup>: the frontal (Fp1–2, Fpz, AF1–2, AFz, F1–8, and Fz), temporal (FT9–10, T7–8, TP9–10), and parietal (CP1–6, CPz, P1–4, P7–8, and Pz) regions. We focused on the frontal, temporal, and parietal regions when comparing the relevance score resulting from the LRP as there is an ongoing debate regarding which brain area, i.e., the front versus the back, is related to consciousness<sup>40</sup>. We also included the temporal region as activation in the NREM sleep increases in this region<sup>70</sup>.

Additional EEG analyses are shown in Supplementary Methods as follows: (i) performance according to the number of trials in the ECI calculation, (ii) comparison of the difference between correct and incorrect trials, and (iii) classification performance using EEG signals excluding frontal or parietal electrodes.

**Comparison with PCI.** We compared ECI<sup>awa</sup> with the PCI values computed following the same procedure described in<sup>16</sup>. PCI measures the complexity of the spatiotemporal patterns of cortical activity significantly evoked by TMS<sup>16</sup>. PCI ranges between 0 (minimum complexity) and 1 (maximum complexity). Previous extensive validation of PCI provided an empirical cutoff (PCI\* = 0.31) to discriminate between consciousness and unconsciousness<sup>17</sup>.

**Statistical analysis.** We used the Kruskal–Wallis test (nonparametric one-way analysis of variance) to analyze the differences in the classification accuracy; moreover, two-sided multiple  $t$ -tests were used for post-hoc analysis using Fisher's least significant differences method for multiple comparisons to compare the classification performance of the three classifiers (LDA, SVM, and CNN) of sleep data for each component (arousal and awareness) and at three-time ranges (0–200, 200–400, 400–600 ms) of spatiotemporal information. The Kruskal–Wallis test was also performed to compare

the classification performance using transfer learning. Similarly, Fisher's least significant differences method was applied after the two-sided multiple  $t$ -tests.

To investigate the discrimination of ECI in each state of consciousness, the feedforward network was trained with 20 hidden layers using the LOPO approach. For each output class, the AUC, sensitivity, and specificity were calculated using ROC analysis.

The Kruskal–Wallis test was employed to investigate if there were any differences in relevance scores among brain regions from LRP under sleep. We performed two-sided multiple  $t$ -tests using Fisher's least significant differences method for multiple comparisons. In addition, the Kruskal–Wallis test was performed to explore the differences in relevance scores from LRP under the condition of anesthesia and for severely brain-injured patients. For post-hoc analysis, two-sided multiple  $t$ -tests were performed using Fisher's least significant differences method for multiple comparisons.

Finally, we used Pearson's correlation to investigate the relationship between ECI<sup>awa</sup> and PCI. In this study, all significances were indicated by  $\alpha = 0.05$ .

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All data (sleep dataset, anesthesia dataset, already published brain injury dataset, and new added brain injury dataset) generated and used in this study have been deposited in a local database and are available upon reasonable request to Olivia Gossesries. In addition, resting-state EEG signals during anesthesia and wake are available online upon request at the repository Zenodo (<https://doi.org/10.5281/zenodo.806176>). The TMS–EEG data of some brain injury patients (published and new datasets) will also be freely available through EBRAINS within 2022 with no restriction to access (<https://doi.org/10.25493/G8E3-DQE>). The sleep dataset was previously reported by Nieminen et al.<sup>43</sup> for NREM data and Lee et al.<sup>53</sup> for REM data. The anesthesia data were also previously published by Sarasso et al.<sup>36</sup>. The published brain injury dataset was previously reported by Bodart et al.<sup>54,55</sup> and Rosanova et al.<sup>56</sup>. However, new added brain injury dataset has not yet been published. The raw EEG data are protected and are not made publicly available owing to data privacy laws, but are available from the corresponding author upon reasonable request. Source data are provided with this paper.

## Code availability

Source code generated and used for this study is publicly available for download at <https://github.com/MinjiLee-ku/ECI> and <https://doi.org/10.5281/zenodo.5760787> (ref.<sup>71</sup>). Source code for CNN and LRP is freely available at [https://github.com/sebastian-lapuschkin/lrp\\_toolbox](https://github.com/sebastian-lapuschkin/lrp_toolbox). Source code for violin plot is available from <https://www.mathworks.com/matlabcentral/fileexchange/45134-violin-plot>. Source code for shaded error bar is available from <https://github.com/raacampbell/shadedErrorBar>.

Received: 8 August 2020; Accepted: 25 January 2022;

Published online: 25 February 2022

## References

- Sanders, R. D., Tononi, G., Laureys, S. & Sleigh, J. W. Unresponsiveness unconsciousness. *Anesthesiology* **116**, 946–959 (2012).
- Darracq, M. et al. Evoked alpha power is reduced in disconnected consciousness during sleep and anesthesia. *Sci. Rep.* **8**, 16664 (2018).
- Colombo, M. A. et al. The spectral exponent of the resting EEG indexes the presence of consciousness during unresponsiveness induced by propofol, xenon, and ketamine. *Neuroimage* **189**, 631–644 (2019).
- Lendner, J. D. et al. An electrophysiological marker of arousal level in humans. *eLife* **9**, e55092 (2020).
- Mashour, G. A. & Hudetz, A. G. Neural correlates of unconsciousness in large-scale brain networks. *Trends Neurosci.* **41**, 150–160 (2018).
- Casarotto, S. et al. Exploring the neurophysiological correlates of loss and recovery of consciousness: perturbational complexity in *Brain Function and Responsiveness in Disorders of Consciousness* (ed Monti, M. M.) 93–104 (Springer, 2016).
- Bonhomme, V. et al. General anesthesia: a probe to explore consciousness. *Front. Syst. Neurosci.* **13**, 36 (2019).
- Sanders, R. D. et al. Incidence of connected consciousness after tracheal intubation: a prospective, international, multicenter cohort study of the isolated forearm technique. *Anesthesiology* **126**, 214–222 (2017).
- Noirhomme, Q., Brecheisen, R., Lesenfants, D., Antonopoulos, G. & Laureys, S. “Look at my classifier's result”: disentangling unresponsive from (minimally) conscious patients. *Neuroimage* **145**, 288–303 (2017).
- Giacino, J. T. et al. The minimally conscious state: definition and diagnostic criteria. *J. Neurol.* **58**, 349–353 (2002).

11. Gosseries, O., Di, H., Laureys, S. & Boly, M. Measuring consciousness in severely damaged brains. *Annu. Rev. Neurosci.* **37**, 457–478 (2014).
12. Giacino, J. T., Kalmar, K. & Whyte, J. The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. *Arch. Phys. Med. Rehabil.* **85**, 2020–2029 (2004).
13. Thibaut, A. et al. Preservation of Brain Activity in Unresponsive Patients Identifies MCS Star. *Ann Neurol* **90**, 89–100 <https://doi.org/10.1002/ana.26095> (2021).
14. Gosseries, O., Zasler, N. D. & Laureys, S. Recent advances in disorders of consciousness: focus on the diagnosis. *Brain Inj.* **28**, 1141–1150 (2014).
15. Stender, J. et al. Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. *Lancet* **384**, 514–522 (2014).
16. Casali, A. G. et al. A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* **5**, 198ra105–198ra105 (2013).
17. Casarotto, S. et al. Stratification of unresponsive patients by an independently validated index of brain complexity. *Ann. Neurol.* **80**, 718–729 (2016).
18. Gosseries, O. et al. Automated EEG entropy measurements in coma, vegetative state/unresponsive wakefulness syndrome and minimally conscious state. *Funct. Neurol.* **26**, 25 (2011).
19. Engemann, D. A. et al. Robust EEG-based cross-site and cross-protocol classification of states of consciousness. *Brain* **141**, 3179–3192 (2018).
20. Baird, B. et al. Human rapid eye movement sleep shows local increases in low-frequency oscillations and global decreases in high-frequency oscillations compared to resting wakefulness. *eNeuro* **5**, 4 (2018).
21. Müller, K.-R. et al. Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. *J. Neurosci. Methods* **167**, 82–90 (2008).
22. Lemm, S., Blankertz, B., Dickhaus, T. & Müller, K.-R. Introduction to machine learning for brain imaging. *Neuroimage* **56**, 387–399 (2011).
23. Liu, Q. et al. Spectrum analysis of EEG signals using CNN to model patient's consciousness level based on anesthesiologists' experience. *IEEE Access* **7**, 53731–53742 (2019).
24. Fahimi, F. et al. Inter-subject transfer learning with end-to-end deep convolutional neural network for EEG-based BCI. *J. Neural Eng.* **16**, 026007 (2019).
25. Webb, S. Deep learning for biology. *Nature* **554**, 555–557 (2018).
26. Montavon, G., Binder, A., Lapuschkin, S., Samek, W. & Müller, K.-R. Layer-wise relevance propagation: an overview in explainable AI: interpreting, explaining and visualizing deep learning (eds Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., Müller, K.-R.) 193–209 (Springer, 2019).
27. Kwon, O.-Y., Lee, M.-H., Guan, C. & Lee, S.-W. Subject-independent brain-computer interfaces based on deep convolutional neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **31**, 3839–3852 (2020).
28. Sturm, I., Lapuschkin, S., Samek, W. & Müller, K.-R. Interpretable deep neural networks for single-trial EEG classification. *J. Neurosci. Methods* **274**, 141–145 (2016).
29. Lotte, F. et al. A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update. *J. Neural Eng.* **15**, 031005 (2018).
30. Lapuschkin, S., Binder, A., Montavon, G., Müller, K.-R. & Samek, W. The LRP toolbox for artificial neural networks. *J. Mach. Learn. Res.* **17**, 3938–3942 (2016).
31. Massimini, M. et al. Triggering sleep slow waves by transcranial magnetic stimulation. *Proc. Natl. Acad. Sci. USA* **104**, 8496–8501 (2007).
32. Massimini, M., Tononi, G. & Huber, R. Slow waves, synaptic plasticity and information processing: insights from transcranial magnetic stimulation and high-density EEG experiments. *Eur. J. Neurosci.* **29**, 1761–1770 (2009).
33. Napolitani, M. et al. Transcranial magnetic stimulation combined with high-density EEG in altered states of consciousness. *Brain Inj.* **28**, 1180–1189 (2014).
34. Massimini, M. et al. Breakdown of cortical effective connectivity during sleep. *Science* **309**, 2228–2232 (2005).
35. Rosanova, M. et al. Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. *Brain* **135**, 1308–1320 (2012).
36. Sarasso, S. et al. Consciousness and complexity during unresponsiveness induced by propofol, xenon, and ketamine. *Curr. Biol.* **25**, 3099–3105 (2015).
37. Luppi, A. I. et al. Consciousness-specific dynamic interactions of brain integration and functional diversity. *Nat. Commun.* **10**, 4616 (2019).
38. Jeon, E., Ko, W. & Suk, H.-I. Domain adaptation with source selection for motor-imagery based BCI. in *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*. 1–4 (IEEE).
39. Lawhern, V. J. et al. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng.* **15**, 056013 (2018).
40. Koch, C., Massimini, M., Boly, M. & Tononi, G. Posterior and anterior cortex—where is the difference that makes the difference? *Nat. Rev. Neurosci.* **17**, 666 (2016).
41. Siclari, F. & Tononi, G. Local aspects of sleep and wakefulness. *Curr. Opin. Neurobiol.* **44**, 222–227 (2017).
42. Siclari, F., Bernardi, G., Cataldi, J. & Tononi, G. Dreaming in NREM sleep: a high-density EEG study of slow waves and spindles. *J. Neurosci.* **38**, 9175–9185 (2018).
43. Nieminen, J. O. et al. Consciousness and cortical responsiveness: a within-state study during non-rapid eye movement sleep. *Sci. Rep.* **6**, 30932 (2016).
44. Lee, M. et al. Connectivity differences between consciousness and unconsciousness in non-rapid eye movement sleep: a TMS-EEG study. *Sci. Rep.* **9**, 5175 (2019).
45. Lee, M. et al. Network properties in transitions of consciousness during propofol-induced sedation. *Sci. Rep.* **7**, 16791 (2017).
46. Chennu, S. et al. Brain networks predict metabolism, diagnosis and prognosis at the bedside in disorders of consciousness. *Brain* **140**, 2120–2132 (2017).
47. Afrasiabi, M. et al. Consciousness depends on integration between parietal cortex, striatum, and thalamus. *Cell Syst.* **12**, 363–373 (2021).
48. Vanhaudenhuyse, A. et al. Default network connectivity reflects the level of consciousness in non-communicative brain-damaged patients. *Brain* **133**, 161–171 (2010).
49. Andersen, L. M., Pedersen, M. N., Sandberg, K. & Overgaard, M. Occipital MEG activity in the early time range (<300 ms) predicts graded changes in perceptual consciousness. *Cereb. Cortex* **26**, 2677–2688 (2016).
50. Russo, S. et al. TAAC-TMS Adaptable Auditory Control: a universal tool to mask TMS clicks. *J. Neurosci. Meth.*, <https://doi.org/10.1016/j.jneumeth.2022.109491> (2022).
51. Casarotto, S. et al. The rt-TEP tool: real-time visualization of TMS-evoked potentials to maximize cortical activation and minimize artifacts. *J. Neurosci. Meth.* **370**, 109486 (2022).
52. Belardinelli, P. et al. Reproducibility in TMS-EEG studies: a call for data sharing, standard procedures and effective experimental control. *Brain Stimul.* **12**, 787–790 (2019).
53. Lee, M. et al. Graph theoretical analysis of cortical networks based on conscious experience. in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 3373–3376 (IEEE).
54. Bodart, O. et al. Measures of metabolism and complexity in the brain of patients with disorders of consciousness. *Neuroimage Clin.* **14**, 354–362 (2017).
55. Bodart, O. et al. Global structural integrity and effective connectivity in patients with disorders of consciousness. *Brain Stimul.* **11**, 358–365 (2018).
56. Rosanova, M. et al. Sleep-like cortical OFF-periods disrupt causality and complexity in the brain of unresponsive wakefulness syndrome patients. *Nat. Commun.* **9**, 4427 (2018).
57. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).
58. Bertrand, O., Perrin, F. & Pernier, J. A theoretical justification of the average reference in topographic evoked potential studies. *Electroencephalogr. Clin. Neurophysiol.* **62**, 462–464 (1985).
59. Ludwig, K. A. et al. Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *J. Neurophysiol.* **101**, 1679–1689 (2009).
60. Zhang, D. et al. Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface. in *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*. 1703–1710.
61. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. Preprint at [arXiv https://arxiv.org/abs/1412.6980](https://arxiv.org/abs/1412.6980) (2014).
62. Blankertz, B., Lemm, S., Treder, M., Haufe, S. & Müller, K.-R. Single-trial analysis and classification of ERP components—a tutorial. *Neuroimage* **56**, 814–825 (2011).
63. Smits, G. F. & Jordaan, E. M. Improved SVM regression using mixtures of kernels. in *2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*. 2785–2790 (IEEE).
64. Rueda-Delgado, L. et al. Brain event-related potentials predict individual differences in inhibitory control. *Int. J. Psychophysiol.* **18**, 30870–30875 (2019).
65. Korjus, K., Hebart, M. N. & Vicente, R. An efficient data partitioning to improve classification performance while keeping parameters interpretable. *PLoS ONE* **11**, e0161788 (2016).
66. Thiery, T. et al. Long-range temporal correlations in the brain distinguish conscious wakefulness from induced unconsciousness. *Neuroimage* **179**, 30–39 (2018).
67. Krepki, R., Blankertz, B., Curio, G. & Müller, K.-R. The Berlin Brain-Computer Interface (BBCI)—towards a new communication channel for online control in gaming applications. *Multimed. Tools Appl.* **33**, 73–90 (2007).
68. Lapuschkin, S. et al. Unmasking Clever Hans predictors and assessing what machines really learn. *Nat. Commun.* **10**, 1096 (2019).
69. Tóth, B. et al. EEG network connectivity changes in mild cognitive impairment—preliminary results. *Int. J. Psychophysiol.* **92**, 1–7 (2014).



70. Nir, Y., Massimini, M., Boly, M. & Tononi, G. Sleep and consciousness in *Sleep and Consciousness* (ed Cavanna, A. E., Nani, A., Blumenfeld, H. & Laureys, S.) Chapter 9, 133–182 (Springer Berlin Heidelberg, 2013).
71. Lee, M. et al. Quantifying arousal and awareness in altered states of consciousness using interpretable deep learning. MinjiLee-ku/ECI: First release of ECI\_update. <https://doi.org/10.5281/zenodo.5760787> (2021).

### Acknowledgements

This work was supported by the Institute for Information and Communications Technology Planning and Evaluation (IITP) funded by the Korean government (Nos. 2017-0-00451; 2017-0-01779; 2019-0-00079; 2019-0-01371; and 2021-0-02068), the University and University Hospital of Liège, Belgian National Fund for Scientific Research (F.R.S.-FNRS), the Italian Ministry of Health, GR-2016-02361494 (to S.C.), the Canadian Institute for Advanced Research (CIFAR) (to M.M.), European Union's Horizon 2020 Framework Program for Research and Innovation under the Specific Grant Agreement (No. 945539, Human Brain Project SGA3) (to M.M. and S.L.), BIAL Foundation, AstraZeneca Foundation, Fund Generate, King Baudouin Foundation, DOCMA project [EU-H2020-MSCA-RISE--778234], James McDonnell Foundation, Mind Science Foundation, Fondazione Europea di Ricerca Biomedica, National Institutes of Health (No. R01MH064498), Academy of Finland (Nos. 265680 and 294625), Tiny Blue Dot Foundation (to M.M.), and grant EraPerMed JTC 2019 "PerBrain" (to M.R.). L.R.D.S. and R.P. are PhD fellows, O.G. and A. T. are research associates, and S.L. is research director at the F.R.S.-FNRS. We thank S. Lapuschkin for sharing the code; further, we thank all the healthy participants, patients, and their families who participated in this study.

### Author contributions

O.G., J.O.N., M.B., S.L., M.M., and G.T. designed the experiments. O.G., A.W., A.B., L.R.D.S., J.O.N., R.P., V.B., M.B., O.B., J.A., A.T., M.R., S.C., and M.M. performed the experiments. M.L. and S.-W.L. designed the methodology and analyzed the data. M.L. drafted the manuscript with the help of O.G. All authors revised the manuscript critically and contributed to the important intellectual content.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-28451-0>.

**Correspondence** and requests for materials should be addressed to Olivia Gosseries or Seong-Wan Lee.

**Peer review information** *Nature Communications* thanks Christof Koch and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022