Predicting Cortical Signatures of Consciousness using Dynamic Functional Connectivity Graph-Convolutional Neural Networks

A. Grigis[1,a], J. Tasserie[1,a,b,c], V. Frouin[a], B. Jarraya[a,b,d] and L. Uhrig[a,b]

[a] Université Paris-Saclay, CEA, NeuroSpin, 91191, Gif-sur-Yvette, France.
[b] Cognitive Neuroimaging Unit, Institut National de la Santé et de la Recherche Médicale U992, Gif-sur-Yvette, France.
[c] Université Paris Sud, Université Paris-Saclay, Orsay, France.
[d] University of Versailles, Université Paris-Saclay, Neurosurgery Department, Foch Hospital, Suresnes, France.
[1] These authors equally contributed to the study.

Abstract

Decoding the levels of consciousness from cortical activity recording is a major challenge in neuroscience. Using clustering algorithms, we previously demonstrated that resting-state functional MRI (rsfMRI) data can be split into several clusters also called "brain states" corresponding to "functional configurations" of the brain. Here, we propose to use a supervised machine learning method based on artificial neural networks to predict functional brain states across levels of consciousness from rsfMRI. Because it is key to consider the topology of brain regions used to build the dynamical functional connectivity matrices describing the brain state at a given time, we applied BrainNetCNN, a graph-convolutional neural network (CNN), to predict the brain states in awake and anesthetized non-human primate rsfMRI data. BrainNetCNN achieved a high prediction accuracy that lies in *[0.674, 0.765]* depending on the experimental settings. We propose to derive the set of connections found to be important for predicting a brain state, reflecting the level of consciousness. The results demonstrate that deep learning methods can be used not only to predict brain states but also to provide additional insight on cortical signatures of consciousness with potential clinical consequences for the monitoring of anesthesia and the diagnosis of disorders of consciousness.

Keywords: deep learning, graph-convolutional network, cortical signatures of consciousness, fMRI, non-human primate.

1. Introduction

A major challenge in fundamental and clinical neuroscience is to predict the level of consciousness from brain activity recording such as EEG or functional MRI (fMRI). Decoding levels of consciousness from neural activity is key to better understand the mechanism by which anesthetics induce consciousness loss, to monitor the depth of anesthesia in an objective manner, and help diagnosing disorders of consciousness with higher accuracy [11]. Instead of a bottom up investigation starting from molecules and cells, an information-processing analysis of brain functional imaging studies during the awake state or loss of consciousness induced by various anesthetics was proposed in non-human primates using dynamical resting state [1, 12]. The resting-state fMRI (rsfMRI) data were split into several clusters also called "brain states" corresponding to "functional configurations" of the brain, and allows a decomposition of the consciousness conditions of a subject in terms of functional patterns. Briefly, a set of k Brain States $\{BS\}^{k,}$ corresponding to functional configurations, is obtained from rsfMRI in non-human primates in the awake state and under different anesthetics. Then each $\{BS\}^{k}$ is ranked according to its similarity with an anatomical connectivity matrix defined in [5]. In [1, 12] the authors demonstrated that dynamic long-range functional correlation matrices reduce to the underlying structural anatomical connectivity matrix during anesthesia (Fig. 1). In other words, the brain states $\{BS\}^{k}$ are directly related to the level of

consciousness.

The methodology described above includes the entire dataset for the extraction and annotation of $\{BS\}^k$. Although the set of $\{BS\}^k$ is robust and stays stable when a new rsfMRI sequence is added to the database, the analysis process resumes the whole clustering, with potentially an appropriate initialization. Leveraging the increasing number of rsfMRI runs available, the objectives of this work is three-folds and will not require to recompute the $\{BS\}^k$ set. First we aim at confirming the interest of $\{BS\}^k$ as labels to characterize the succession of states occurring under a given arousal condition. Using supervised machine learning methods we will train a classifier that can be used to predict one brain state from small temporal windows of rsfMRI runs in completely new samples [6]. In this work, we will use neural networks, and more specifically the BrainNetCNN graph-convolutional network classifier. Convolutional neural networks are not able to handle the spatial neighborhood of the ROIs used to build the connectivity matrices. BrainNetCNN is composed of specific edge-to-edge, edge-to-node and node-to-graph convolutional filters that will better reflect the topological locality of the regions used to generate the connectivity matrices. Second, we will challenge the data and learning process to model the brain dynamic oscillating from one state to another. Finally, we will leverage the embedded automatic differentiation of the deep learning framework, to uncover which connections were used by the classifying network to be predictive of brain states. These maps of predictive connections, which are duals of the $\{BS\}^k$, yield some insights on both the brain states and the cortical signature of consciousness, and can be considered (at some extent) as proxies of cortical signatures of consciousness.

1. Material and Method

Material: This study uses the dataset described in [1]. Data were acquired either in the awake state or under different anesthetic conditions (ketamine, propofol, or sevoflurane anesthesia) in five macaque monkeys. Levels of anesthesia were defined by a clinical score (the monkey sedation scale) and continuous electroencephalography monitoring. 156 resting state functional MRI runs with corresponding structural MRI were acquired on a Siemens 3-Tesla with a customized single transmit-receiver surface coil. The spatial prepossessing was performed with the NSM pipeline described in [3, 13], with specific time-series denoising operations [1, 12]. Specifically, voxel

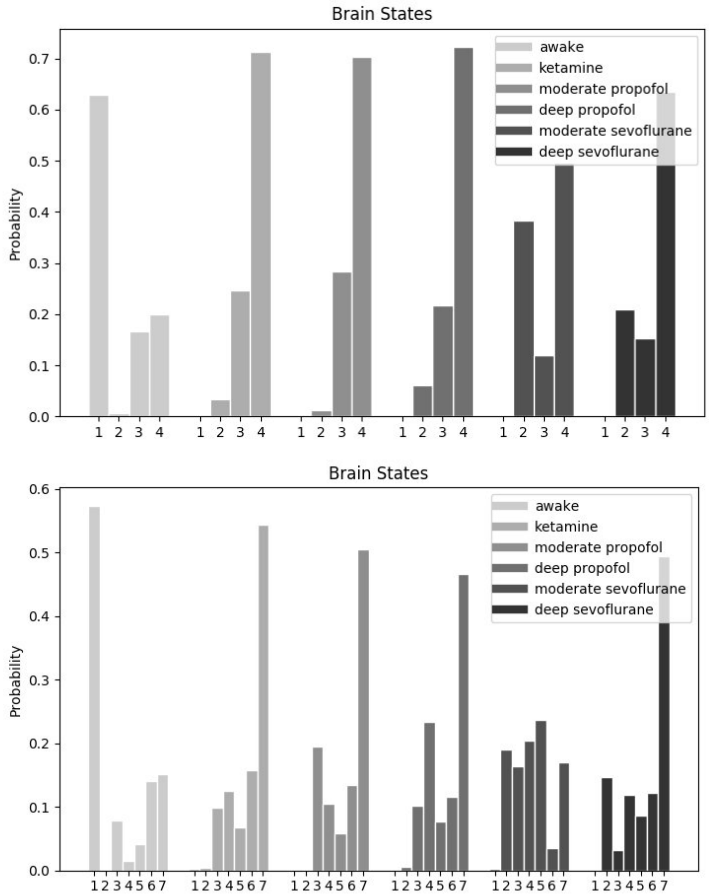

Fig. 1: Probability distributions to observe in a rsfMRI time window, one of ranked brain states $\{BS\}^k$ for the awake and anesthesia conditions for k=4 and k=7. Each bar represents the within-condition probability of occurrence of a state.

time-series were filtered with low-pass (0.05-Hz cutoff) and high-pass (0.0025-Hz cutoff) filters and a zero-phase fast-Fourier notch filter (0.03 Hz) to remove an artefactual pure frequency present in all the data. The variance of each time-serie was normalized, thus, covariance matrices correspond to correlation matrices. Sliding windows are used to compute dynamic connectivity matrices reflecting the brain state at a given time [1]. Windowed segments of each time-serie were computed with a Hamming window (84 s width, which corresponds to 35 scans of 2.4 ms each), and a sliding step of one scan, resulting in 464 windows per run. From the windowed rsfMRI, dynamic functional connectivity matrices were derived using the CoCoMac template with 82 cortical regions (41 cortical regions within each hemisphere) [1, 5]. The windowed rsfMRI time-series from each region were converted into a connectivity matrix by projection into tangent space, capturing connectivity aspects from both the correlation and partial correlation. The corresponding dynamic

connectivity matrices led to a data array of shape 156x464x82×82 which was Fisher transformed to form a set of 156x464 samples. Each sample is more than a list of independent numbers and encodes the functional connectivity patterns. Dominant recurrent patterns of brain correlations (the $\{BS\}^k$ set) were determined using an unsupervised k-means clustering with k states (or classes), k in *[3, 9]*. To avoid double dipping, ie. the use of the same dataset for the training and the testing , the k-means was performed on a subset of data, referred in the following as the train set. The labels associated to the remaining data, referred in the following as the test set, were recovered by selecting the closest brain state $\{BS\}^k$. To this end, the correlation coefficients between the test set dynamic functional connectivity matrices and each $\{BS\}^k$ were computed. This led to a label array of shape 156x464 that will form the empirical ground truth dataset.

Data partitioning: As the dataset is composed of 5 monkeys, with 156 rsfMRI runs in different arousal conditions (31 runs for the awake state, 25 runs for ketamine anesthesia, 25 runs for moderate propofol sedation, 30 runs for deep propofol anesthesia, 25 runs for moderate sevoflurane sedation, and 20 runs for deep sevoflurane anesthesia), we propose to leave one monkey out as the test set (35 runs). The four other monkeys form the train set (121 runs). We further checked that both train and test sets have matching proportions of arousal conditions and brain states labels. The test set was set aside and not used during training. The train set was further split into validation and training folds using a 3-fold stratified cross validation approach.

Model architecture: The BrainNetCNN [2] is designed specifically for network data input, and in our application will enforce the functional connectivity patterns. The network architecture is composed of convolutional layers followed by fully connected layers. Among the proposed configurations, we selected the E2Enet-sml which consists in removing one edge-to-edge layer, and two of the fully connected layers (Fig. 2). This configuration has shown good performance with a restricted number of parameters to be estimated. Precisely, E2Enet-sml has an edge-to-edge layer composed of 32 1×82 and 32 82×1 filters producing feature maps of size 32x82x82. This is followed by an edge-to-node layer with 64 1×82×32 filters producing feature maps of size 64×82x1, an node-to-graph layer with feature maps of size 1×1×30, and a fully connected layer with an output of size k. Increasing the number of feature maps with each layer is a common strategy for CNNs to compensate for the reductions along the other dimensions. Every layer uses very leaky rectified linear units as an activation function, where a negative slope of 1/3 is assigned. For training, we employed a dropout of 0.5 before the node-to-graph layer as shown in Fig. 2. We used an Adam optimizer, a mini-batch of size 100, a weight decay of 5e-3, a learning rate of 1e-5, and a negative log likelihood training loss.

Maps of Predictive Connections: In order to uncover which connections were learned by the BrainNetCNN to be predictive of consciousness, we used the method of Simonyan et al. [4], which computes saliency maps. Saliency maps are a popular visualization tool for gaining insight into why a deep learning model made an individual decision, such as classifying an image or in our case a connectivity matrix. Saliency maps are usually rendered as a heatmap, where hotness corresponds to regions that have a big impact on the model's final decision. In our application they are helpful to extract proxies of cortical signature of consciousness. These consciousness markers are rendered using a circular graph layout where the 41 cortical regions within each hemisphere are displayed around a circle, and the strongest connections linked (>95th percentile).

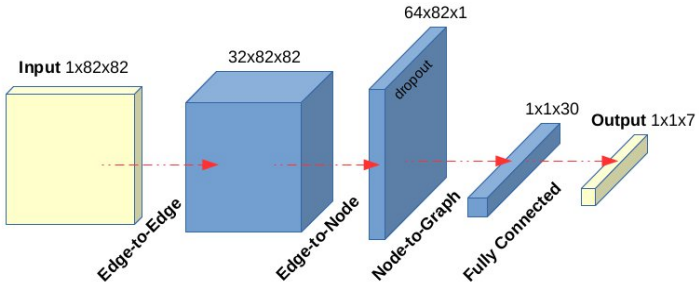

Fig. 2: Schematic representation of the E2Enet-sml architecture.

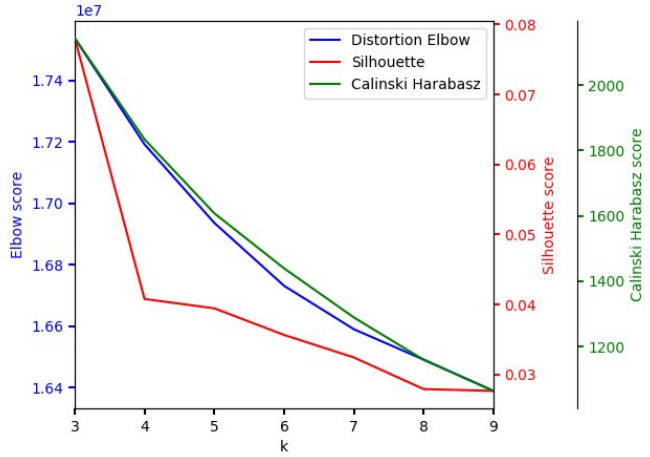

Fig. 3: Determining the optimal number of brain states for the k-means clustering using the Elbow, the Silhouette, and the Calinski & Harabasz scores.

2. Results

Optimal number of the brain states: Prior to the presentation of the results on the three objectives of our work, we detail our sound estimation of the optimal number of brain states in the k-means. The basic idea behind the k-means consists in defining k clusters such that the within-cluster variations are minimum. Different metrics can be employed to select the optimal number of brain states among which we have retained the Elbow [7], the Silhouette [7, 8], and the Calinski & Harabasz [9] scores. The Elbow value is defined as the Within-Cluster Sum of Squared (WCSS) error for different values of k. The optimal k corresponds to the point of inflection on the WCSS versus k curve. Note that the Elbow score is more a decision rule than a metric. The Silhouette value measures how similar a sample is to its own cluster compared to others and lies in *[-1, 1]*. The Calinski & Harabasz value is defined as the ratio between the within-cluster and the between-cluster dispersions. The last two scores have to be maximized with respect to k. From the proposed experiment, we found no clear rule to select the optimal number of brain states for the k-means clustering (Fig. 3). The Silhouette and the Calinski & Harabasz indicators would select the smallest k value. However the choice of the k value is actually the choice of the number of brain states that needs to be driven by biomarkers.

Performance of the classifier: Thus, we decided to monitor the prediction accuracy through different values of k (or targeted brain states $\{BS\}^k$) as shown in Fig. 4. In a k-classes classification problem, the theoretical chance level is 1/k. As described in [10], this threshold holds for an infinite number of samples, and the smaller the sample size, the more likely chance performance differs from this theoretical chance level. In our application the sample size is large (56,144 samples), and we assume that the chance level is close to its theoretical value. For all k values, we noticed that the accuracy is much higher than the theoretical chance level, demonstrating the interest of $\{BS\}^k$ as labels to characterize the succession of states occurring under a given arousal condition. It can be also highlighted that smaller k values give better prediction results regardless of the chance level. The overall predictions are satisfactory and lie in *[0.674, 0.765]*.

For the remaining part of the paper we retained a number of brain states that are biologically credible, ie. k=4 or k=7 [1]. We chose the accuracy as metric to select the best trained model. Then, the Receiver Operating Characteristic (ROC) metric is used to evaluate the BrainNetCNN classifier output quality. ROC curves show the true positive rate on the y-axis, and the false positive rate on the x-axis. An area under the curve (AUC) of 1 corresponds to the ideal case where the false positive rate is zero, and the true positive rate of is one. In the context of multi-label classification, it is necessary to binarize the predicted brain states outputs. First, one ROC curve is drawn per label. We also performed micro-averaging by considering each element of the label indicator matrix as a binary prediction, and macro-averaging by giving equal weight to the classification of each label. Using this validation strategy, we demonstrate that the BrainNetCNN predicts the brain states with a high reproducibility (>0.92 AUC) as shown in Fig. 5.

Towards modeling the brain states dynamic: The difference between the AUC and the accuracy is illustrated in Fig. 6 where we displayed the predicted and the true labels as well as the network estimated probabilities for a test set run acquired in the awake condition. According to Fig. 1, in the awake condition we expect a prevalence of state 1, with residual states 3 and 4 for k=4, and 3, 5, 6 and 7 for k=7. It is clear that the network decisions differ at state transitions modeled by the network estimated probabilities. By challenging the data and the learning process we are able to model the brain dynamic oscillating from one state to another.

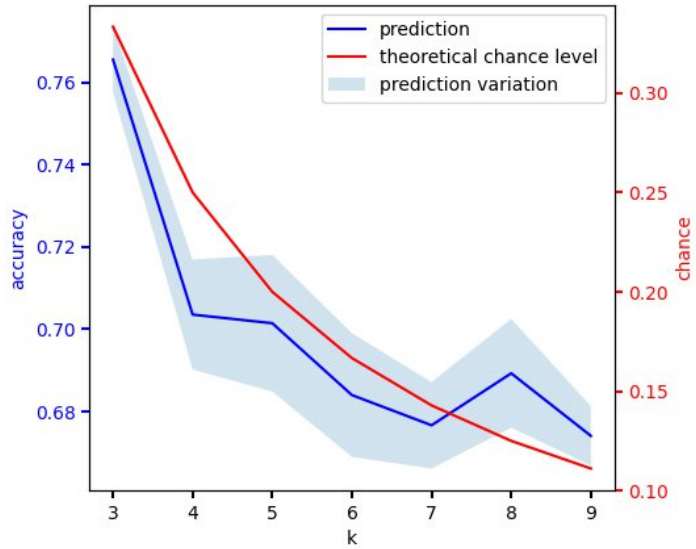

Fig. 4: Determining the optimal number of brain states for the prediction using the accuracy as a reference metric: the mean and standard deviation, computed across the training folds, are displayed. The theoretical chance levels are also displayed, and are always much lower than the accuracy values.

Maps of predictive connections: The proxies of the cortical signatures of consciousness were also computed for each brain state by averaging the within-brain state maps of predictive connections generated for each dynamic connectivity matrix (Fig. 7). The brain states are characterized by the dynamical exploration of a rich, fluctuating set of functional configurations. These dynamical properties constitute a marker of consciousness.

3. Discussion and Conclusion

The BrainNetCNN graph-convolutional model is able to predict the brain states with a good reproducibility and accuracy. We noticed that the proposed network tends to shift the brain state transitions, explaining the observed differences between the accuracy metric and the AUCs. The gold standard is an empirical ground truth generated with the result of a k-means clustering. K-means clustering consists of linear decision boundaries, while the proposed network uses a cascade of many layers of nonlinear processing units for features extraction. This new latent space contains features that are likely to be linearly separable. The proposed network is able to learn a complex representation of the data that we think enables more accurate brain states prediction especially during transitions. More than a simple prediction tool, the proposed network is able to model the brain dynamic oscillating from one state to another.

As presented in [1], the entropy of the state transition matrix, the distance of each state to the underlying anatomical connectivity matrix, and presence of negative correlations are useful candidates when trying to find signatures of consciousness. We illustrate how the maps of predictive connections can disentangle which connections are useful to discriminate between different levels of wakefulness, and bring a significant insight about the brains states. These maps should help in understanding the brain states signature of consciousness. In a future work we want to apply this technique to clinical data of anesthesia and disorders of consciousness.

In a future work we also want to systematically compare the performance of different classifiers based on Machine Learning (ML) or deep learning models to predict brain states. We believe that graph-convolutional networks as the BrainNetCNN will outperform models that do not take into account the topological locality of the regions used to generate the connectivity matrices.

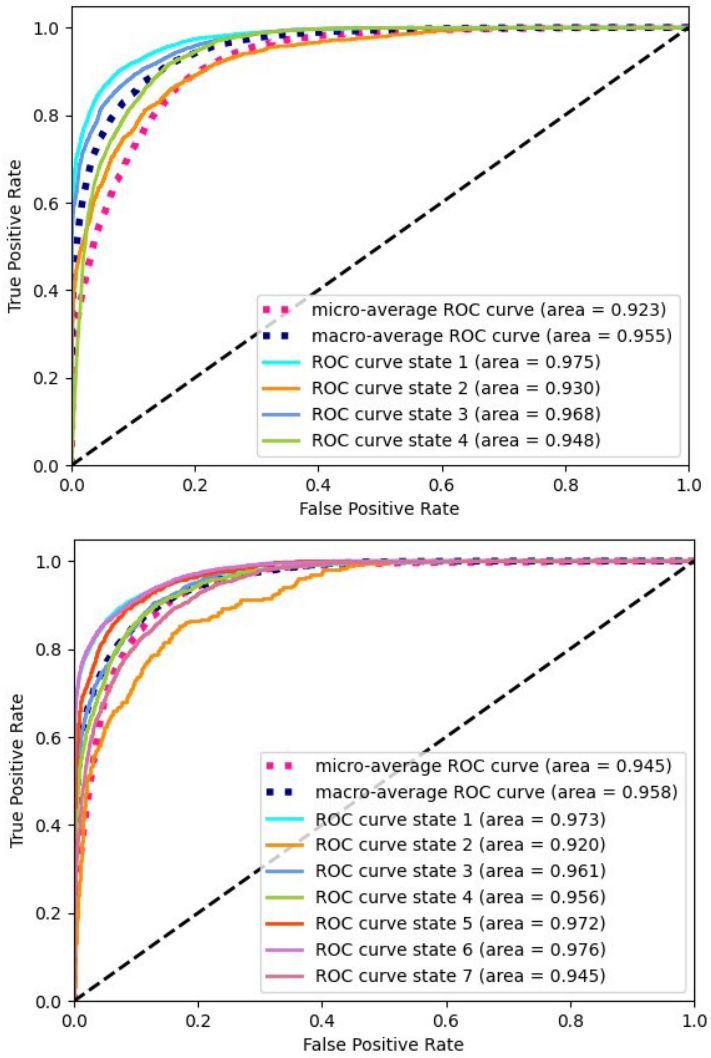

Fig. 5: Evaluation of the BrainNetCNN classifier outputs quality using Receiver Operating Characteristic (ROC) curves for k=4 and k=7.

References

1. Resting-state Dynamics as a Cortical Signature of Anesthesia in Monkey, L. Uhrig et al., Anesthesiology, July 2018.
2. BrainNetCNN: Convolutional Neural Networks for Brain Networks; Towards Predicting Neuro-development, J. Kawahara, NeuroImage, September 2016.
3. Pypreclin: An automatic pipeline for macaque functional MRI preprocessing, J. Tasserie, NeuroImage, Ferbruary 2020.

4. Deep inside convolutional networks: Visualising image classification models and saliency maps, Karen A. Simonyan et al., arXiv preprint arXiv 2013.
5. CoCoMac 2.0 and the future of tract-tracing databases, R. Bakker et al., Front Neuroinform 2012.
6. Supervised Machine Learning: A Review of Classification Techniques, S. Kotsiantis, Informatica (Ljubljana). 2007.
7. Review on Determining of Cluster in K-means Clustering, T. Kodinariya, International Journal of Advance Research in Computer Science and Management 2013.
8. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis, P.J. Rousseeuw, Computational and Applied Mathematics, 1987.
9. A dendrite method for cluster analysis, T. Calinski and J. Harabasz, Communications in Statistics, 1974.
10. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy, E. Combrisson E, and K. Jerbi, Journal of Neuroscience Methods, 2015.
11. Conscious Processing and the Global Neuronal Workspace Hypothesis, G.A .Mashour et al., Neuron, 2020.
12. Signature of consciousness in the dynamics of resting-state brain activity, P. Barttfeld, et al., PNAS, 2014.
13. A Hierarchy of Responses to Auditory Regularities in the Macaque Brain, L. Uhrig, et al., The Journal of Neuroscience, 2014.

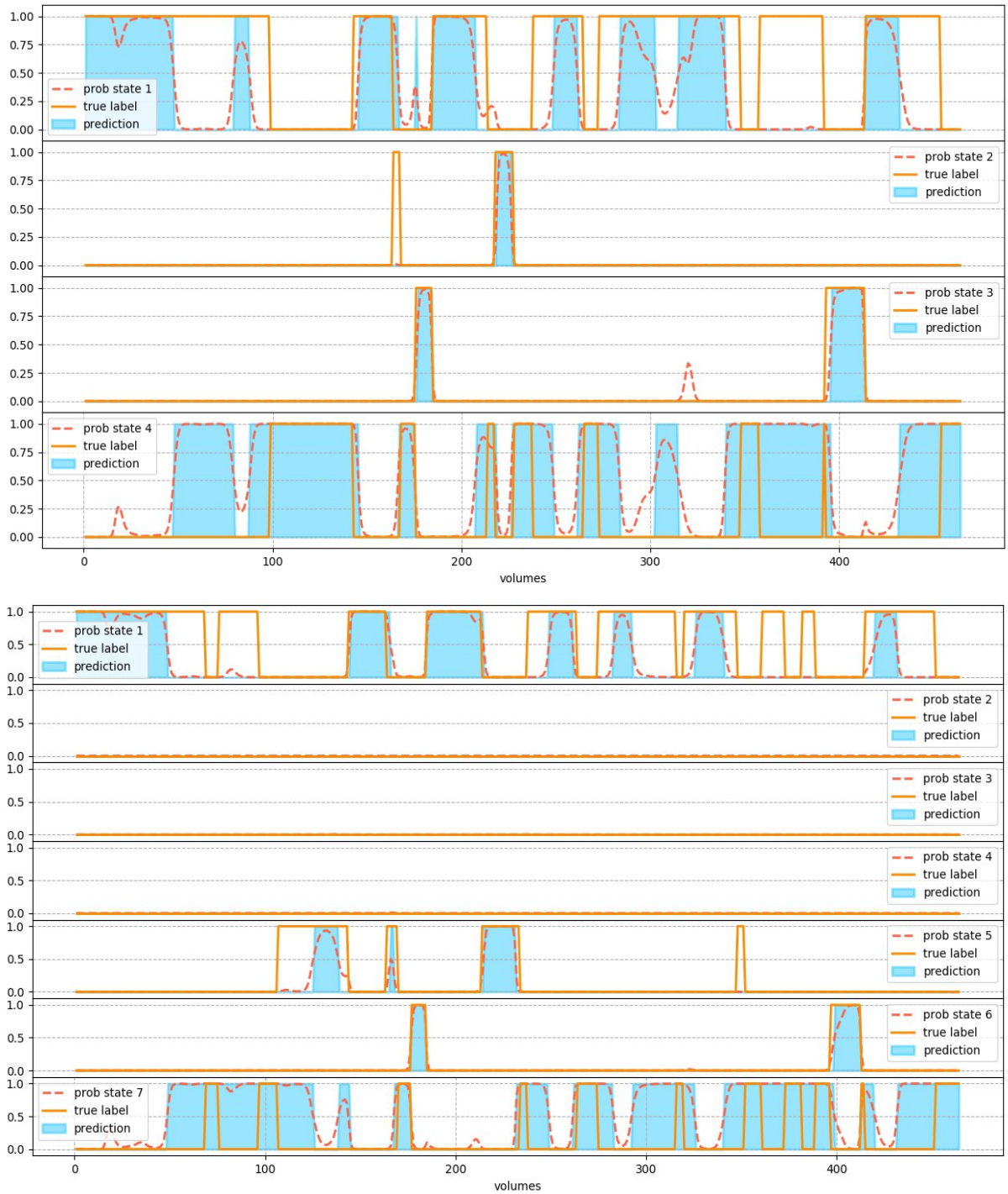

Fig 6: The prediction transitions with the associated probabilities for a test set run acquired in the awake condition for k=4 and k=7.

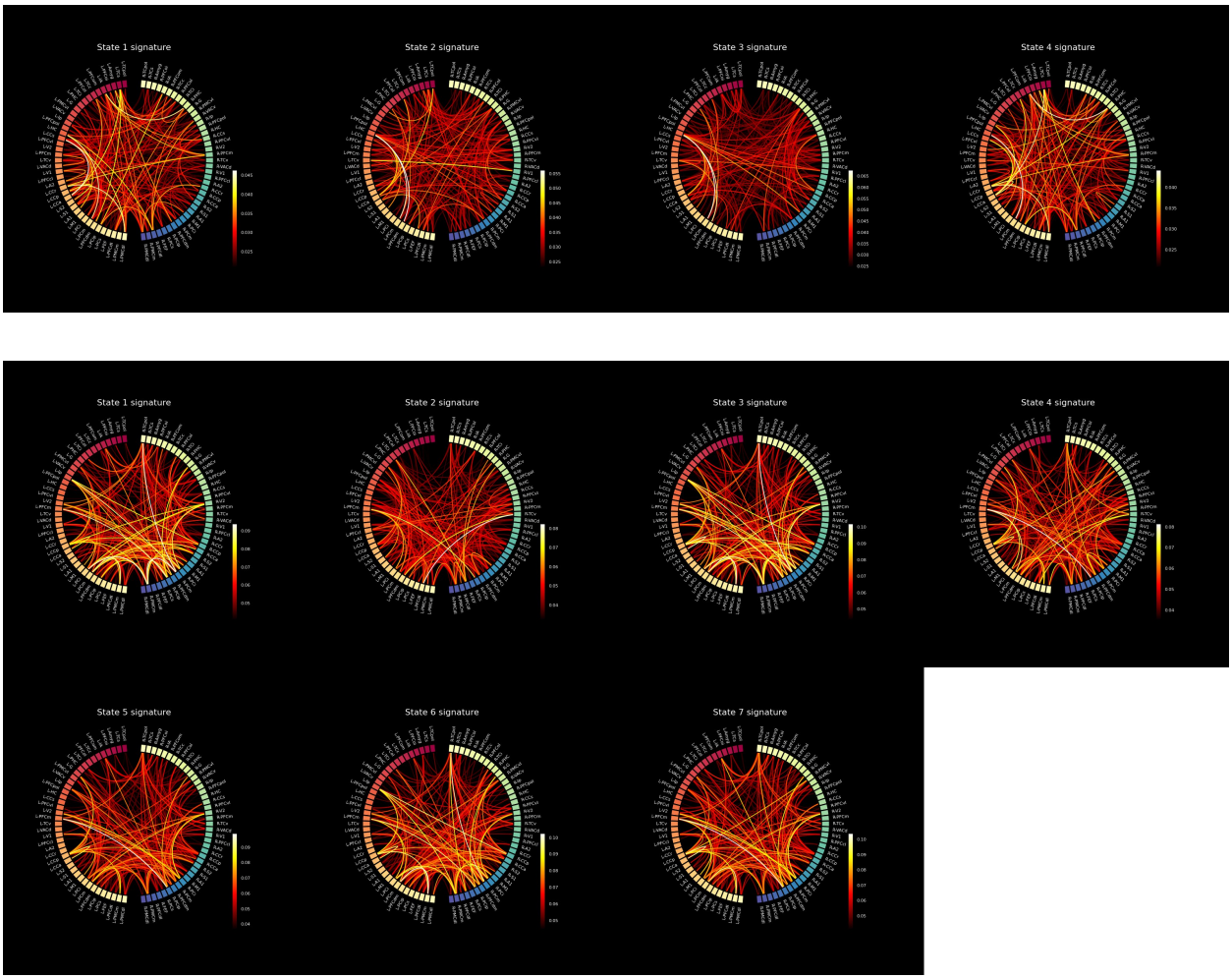

Fig. 7: Within-brain states marker of consciousness, for k=4 and k=7, rendered using a circular graph where the 41 cortical regions within each hemisphere are displayed, and the strongest connections linked (>95th percentile).