GENERALIZED PREDICTION OF UNCONSCIOUSNESS DURING PROPOFOL ANESTHESIA USING 3D CONVOLUTIONAL NEURAL NETWORKS

Konstantinos Patlatzoglou[1], Srivas Chennu[1], Olivia Gosseries[2], Vincent Bonhomme[3], Audrey Wolff[2] and Steven Laureys[2]

[1] *School of Computing, University of Kent, Chatham Maritime, Kent, UK (e-mail: kp356@kent.ac.uk).*
[2] *Coma Science Group, GIGA Consciousness, University and University Hospital of Liege, Liege, Belgium*
[3] *Department of Anesthesia and Intensive Care Medicine, CHU University Hospital of Liege, Liege, Belgium*

KEYWORDS: Anesthetics, Consciousness, Humans, Neural Networks, Computer, Propofol, Unconsciousness

ABSTRACT

Neuroscience has generated a number of recent advances in the search for the neural correlates of consciousness, but these have yet to find valuable real-world applications. Electroencephalography under anesthesia provides a powerful experimental setup to identify electrophysiological signatures of altered states of consciousness, as well as a testbed for developing systems for automatic diagnosis and prognosis of awareness in clinical settings. In this work, we use deep convolutional neural networks to automatically differentiate sub-anesthetic states and depths of anesthesia, solely from one second of raw EEG signal. Our results with leave-one-participant-out-cross-validation show that behavioral measures, such as the Ramsay score, can be used to learn generalizable neural networks that reliably predict levels of unconsciousness in unseen transitional anesthetic states, as well as in unseen experimental setups and behaviors. Our findings highlight the potential of deep learning to detect progressive changes in anesthetic-induced unconsciousness with higher granularity than behavioral or pharmacological markers. This work has broader significance for identifying generalized patterns of brain activity that index states of consciousness.

CLINICAL RELEVANCE

In the United States alone, over 100,000 people receive general anesthesia every day, from which up to 1% is affected by unintended intraoperative awareness [1]. Despite this, brain-based monitoring of consciousness is not common in the clinic, and has had mixed success [2]. Given this context, our aim is to develop and explore an automated deep learning model that accurately predicts and interprets the depth and quality of anesthesia from the raw EEG signal.

I. Introduction

The science of consciousness has received growing interest over the past decades, as brain imaging, neuroscience and computer science have generated new empirical tools and computational methods. One of the research questions arising from this problem, pertains to the electrophysiological signatures of the full neural correlates of consciousness. The investigation of these signatures can contribute not only to our theoretical understanding, but also to biomedicine, where automated detection of these signatures can improve monitoring of depth of anesthesia (DOA), and inform better diagnosis and prognosis for patients with disorders of consciousness. To this end, electroencephalography (EEG) during anesthesia provides a valuable experimental paradigm to tackle this research question, while also being a viable clinical tool, which is simple, convenient and widely accessible in hospitals.

One of the challenges encountered in applying EEG metrics of DOA, is that they require expert analysis and interpretation, while the currently available partial or fully automated DOA systems have not produced reliable results [2]. Here, deep learning with artificial neural networks offers a promising methodology in terms of its predictive accuracy, computational efficiency, automation and generalizability. Recent advances in EEG classification using deep learning, have demonstrated its ability to produce state-of-the-art performance [3, 4]. In our own work, we have highlighted the potential of deep convolutional neural networks to discover and utilize temporal features in raw EEG, for classifying anesthetic states [5].

In this paper, we demonstrate the power of deep learning to generalize and estimate levels of unconsciousness in unseen sub-anesthetic states and experimental contexts. In these contexts, unconsciousness due to propofol anesthesia is determined by either behavioral or pharmacological measures, the former of which is primarily used in clinical settings. However, as neither behavioral nor pharmacological measures can yet provide inconvertible evidence for consciousness in the brain [6, 7], we test the reliability of our neural network model in both experimental contexts. Our results show that reliable models can be learnt to predict behavioral states of anesthesia, and to generate robust predictions in unseen data.

II. Methods

Our methodology focused on the investigation of two datasets acquired and reported in previous studies [7, 8]. The studies were approved by the Ethics Committee of the Faculty of Medicine of the University of Liege [8] and the Cambridgeshire 2 Regional Ethics Committee [7] respectively, with written informed consent from all participants. In both studies, propofol was administered to healthy participants in order to induce two progressively deeper levels of sedation. In the first study, these levels were defined by a behavioral response of each participant according to a standardized clinical scoring scale. In the second study, specific propofol concentrations were targeted in each state, irrespective of the participant's behavioral response. Our aim here was

not to find the optimal network parameters for each given task, but rather to have an exploration and evaluation framework for the model's generalization capabilities.

A. EEG DATASETS

1) DATASET 1

Dataset 1 acquisition is described in detail in [8]. Briefly, fifteen minutes of spontaneous high-density EEG (256- channel EGI Hydrocel GSN) was recorded from 9 participants (mean age 22 $\pm$ 2 y, 4 males) during anesthesia, at 4 states: Wakefulness, Sedation, Loss of Consciousness (LOC) and Recovery. Each state was determined based upon reaching and sustaining the desired Ramsay score, assessed twice by a verbal command. The desired Ramsay score for Wakefulness and Recovery was 2 (clear response to command), Ramsay 3 for Sedation (clear but slow response) and Ramsay 6 for LOC (no response). Throughout the experiment, computer-controlled intravenous infusion was used to obtain constant effect-site concentrations of propofol (alongside the pharmacokinetic Marsh model). Arterial blood samples were also taken immediately before and after each anesthetic state, for subsequent determination of propofol levels (average concentrations were 1.91 $\pm$ 0.52 µg/ml for *Sedation,* and 3.87 $\pm$ 1.39 µg/ml for LOC). The experimental design is depicted in Fig. 1.

2) DATASET 2

Dataset 2 acquisition is described in detail in [7]. Briefly, seven minutes of spontaneous high-density EEG (128- channel EGI Hydrocel GSN) was recorded from 20 participants (mean age = 30.85, SD = 10.98, 9 males) during anesthesia, at 4 different states: *Wakefulness*, *Mild Sedation*, *Moderate Sedation* and *Recovery*. For each state, a desired plasma concentration was targeted and controlled by a computerized syringe driver that determined and maintained the required infusion rate of propofol (Marsh model). The targeted plasma level for *Mild Sedation* was 0.6 µg/ml (relaxed but still responsive) and 1.2 µg/ml for *Moderate Sedation*. For *Recovery*, EEG measurements were taken 20 minutes after cessation of infusion, to ensure that propofol concentrations would approach zero (based on pharmacokinetic simulation). Blood samples were taken at the beginning and end of each state, in order to measure propofol levels and characterize inter-individual variability, as well as to confirm similarity to target concentrations. The experimental design is depicted in Fig. 2.

After each resting state period, a simple behavioral task was performed that involved a fast discrimination (button press) between two possible auditory stimuli. Each participant's hit rates (percentage of correct responses) and reaction times were recorded for further analysis. Based on binomial modelling of their hit rates, two subgroups of 13 and 7 participants were identified and characterized as *Responsive* and *Drowsy*, which reflected the behavioral impairment of the participant-specific effect of the drug.

B. EEG PRE-PROCESSING

Minimal pre-processing steps were applied to the original EEG recordings, in order to preserve the information in the signal and enable the neural network to extract relevant features. Twenty channels were selected from both datasets, based on the 10-20 system, namely: Fp1, Fp2, F7, F3,

Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, Oz, and O2. The data were filtered from 0.5 to 40 Hz (windowed FIR design), down-sampled from 1000 Hz to 100 Hz, and rereferenced to the average. No artifacts were rejected, ensuring that the pre-processing required no manual intervention and could potentially be deployed in real-time in a clinical setting. The above pre-processing steps were implemented using the MNE-python library.

The input to the deep learning model described below were 1-second non-overlapping epochs of the raw EEG signal. Within each epoch, data were normalized using the robust standardization implemented in Scikit-Learn.

C. CONVOLUTIONAL NEURAL NETWORK

The model used in this work employed a novel 3D convolutional neural network design, which innovated upon our previous model in [5]. In this design, the spatial configuration of the EEG channels is transformed into a 2D grid that preserves the spatial continuity and structure of the channels, resulting in a 3D representation per sample (5 x 5 x 100) (similar approaches found in [9, 10]). This approach enables us to exploit the potential of the convolution operation simultaneous in time and space. It also has the prospective of better model interpretability, by preserving the spatio-temporal relations understood in clinical EEG. Finally, and equally importantly, converting the data from channel space to a 2D grid enables a model design that is independent of the specific EEG montage.

The precise architecture of the convolutional neural network can be seen in Fig. 3. All activation functions are ReLU units, except for the output neuron which is linear.

Figure 1. *The experimental design of Dataset 1. Anesthetic induction was guided by the desired Ramsay score on each of the 4 states.*
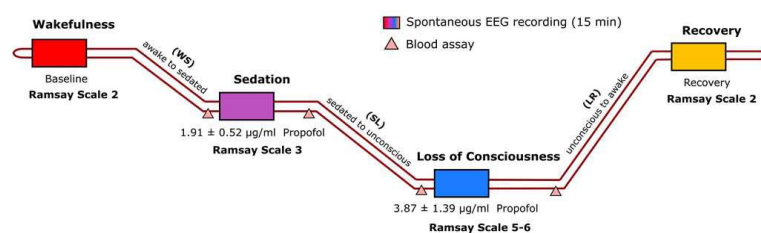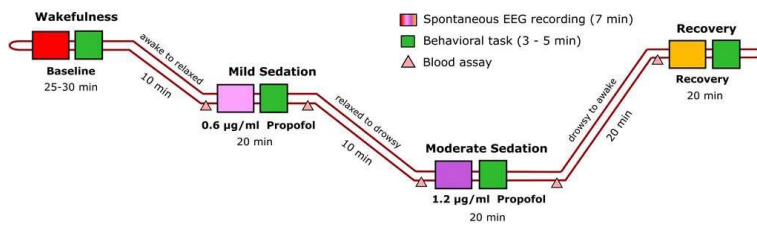

D. MODEL TRAINING

The model was trained using the data from the 1st experimental design, in a regression-on-Ramsay-score optimization problem. Behavioral measures are better suited for the given ground-truth, in contrast to pharmacokinetic models, which are unable to associate targeted propofol infusion rates with specific states and levels of unconsciousness (already known from clinical practice). We experimented with the inclusion of an increasing number of states during training, from the two prominent states of *Wakefulness* and *LOC*, to the sub-anesthetic *Sedation*, and finally *Recovery*, which introduced more and more subtle changes.

The evaluation of the model was based on a *Leave-One- Participant-Out* cross validation, as a way to test the generalizability to unseen participants and to avoid overfitting (often missed in EEG studies). Each input sample was labeled with the Ramsay score of its corresponding state.

The model was trained using the mean-squared-error (MSE) loss and evaluated with the mean-absolute-error (MAE) for its performance. Weight initialization was done with the Xavier normal initializer implemented in Keras. A batch size of 100 was used, and 10 runs of the whole dataset, which ensured performance convergence. All experiments were implemented in Python 3 using Keras/Tensorflow and run on a CUDA NVIDIA GPU (Tesla P100).

Figure 2. *The experimental design of Dataset 2. Anesthetic induction was guided by the drug concentrations targeted on each of the 4 states.*

Figure 3. *The 3D convolutional Neural Network Architecture.*
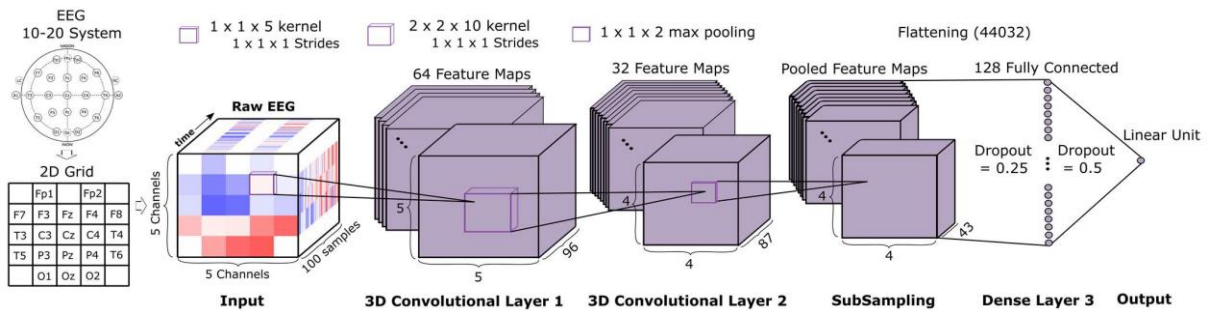
III. Results

A. MODEL EVALUATION

The results for Dataset 1 are summarized in Table I, combining three experiments with increasing number of states included in model training. Training with regression on Ramsay scores revealed an estimation bias, which was affected by a number of factors, such as the number of samples in each state, the given states (ground-truth), and the assumption of the Ramsay scale to represent changes in levels of consciousness linearly. Notably, Sedation was significantly overestimated (MAE: 1.12) in the 2-state case (as test state), while LOC was significantly underestimated (MAE: 1.07) in the 4-state case, when looking at the average Ramsay scores (3.45 for Sedation and 5.08 for LOC, respectively). Importantly, when trained only on 3 states (Wakefulness-Sedation-LOC), the model reached an overall MAE of 0.56, with the most balanced per-state error (Fig. 4). Hence, this optimal model was used for all subsequent testing and evaluation.

Table I. *Exp 1. Regression On Ramsay Score*

States in Training	Mean Absolute Error

	Per State	Total
Wakefulness,LOC	(0.36>*1.12*>0.67>*0.38*)	0.62
Wakefulness ,Sedation,LOC	**(0.31>0.58>0.87>*0.47*)**	**0.56**
Wakefulnes,Sedation,LOCRecovery	(0.23>0.59>1.07>0.25)	0.54

Unseen test states are underlined. Optimal model is highlighted in bold.

Figure 4. *Ramsay score predictions for the unseen test subjects of Exp. Design 1 (average of 9 subjects). Within-subject predictions are robust and consistent in time, when trained in the 3 states highlighted by black boxes.*
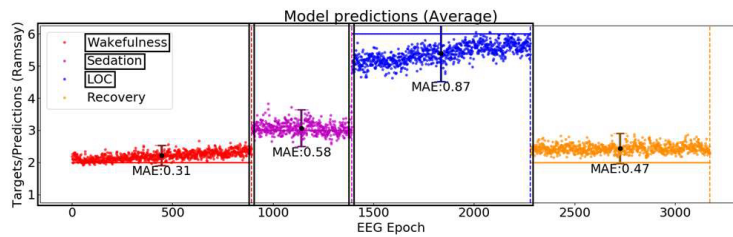

B. MODEL PREDICTIONS ON TRANSITIONAL STATES

We acquired EEG recordings during inter-state transitional periods from 3 participants in the 1st experimental design - namely during WS (Wakefulness to Sedation), SL (Sedation to LOC) and LR (LOC to Recovery). With these data, we were able to test our model's generalizability to both stable and unstable sub-anesthetic states of unconsciousness. Ramsay score estimations (Fig. 5) showed a clear increasing trend within WS and SL (and up to the end phase of LOC), followed by a rapid decrease during LR, accurately tracking the anesthetic induction protocol during the experiment.

## C. MODEL PREDICTIONS ON EXPERIMENTAL DESIGN 2

In order to further show the generalizability of our model trained with the 3 states from Dataset 1, we used Dataset 2 as a test set. Again, predicted Ramsay scores were consistent in levels and increased over time, showing an average range of 2 to 4, in agreement with behavioral reports [7]. As per the original study, we divided participants in Dataset 2 based on the two subgroups - *Responsive* and *Drowsy* - defined solely by the anesthetic-induced change in their individual behavioral states. The model was able to predict the Ramsay scores corresponding to these behavioral states (Fig. 6). This was confirmed statistically by entering the average Ramsay score for each participant and each state, into a mixed ANOVA model (fitrm/ranova, MATLAB) with one non-repeated measure (Participant Group - *Responsive* or *Drowsy*) and one repeated measure (Level of Sedation - *Wakefulness*, *Mild Sedation*, *Moderate Sedation, Recovery*). We found a significant interaction between participant group and sedation level ($F(3) = 8.8$, $p = 0.0007$) in the predicted Ramsay scores. This interaction effect was driven by a significant effect of group during the *Moderate Sedation* level ($p = 0.00006$). Taken together, these effects highlight that the model generalized to correctly predict Ramsay scores in experimental design 2.

Figure 5. *Ramsay score predictions of the 3-state model for all transitional states of Exp. Design 1 (median subject and average of 3 subjects). Trained states are highlighted by black boxes. The changes in levels of consciousness are consistent over time, in agreement with the anesthetic paradigm.*
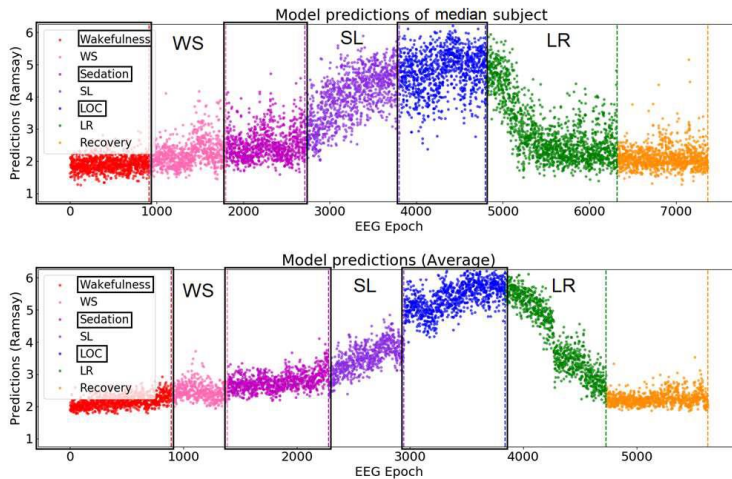
Figure 6. *Ramsay score predictions (average) of the two subgroups from Exp. Design 2. Group 1 (Responsive) and Group 2 (Drowsy) predictions reveal the behavioral traits found in the original study.*
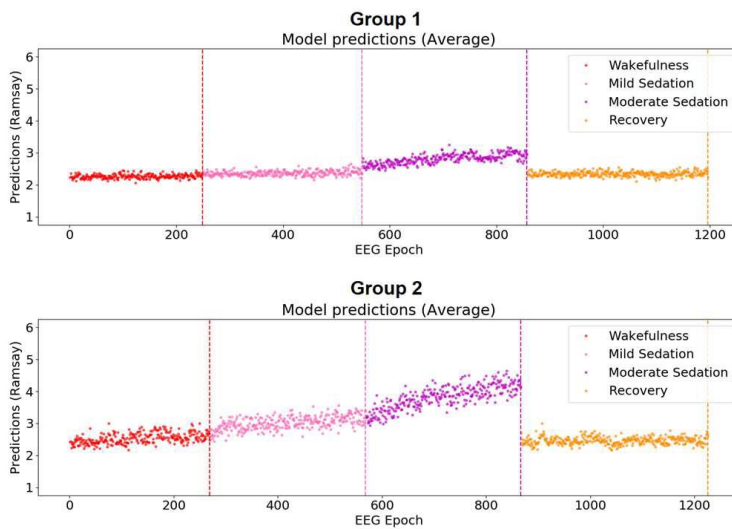
IV. Discussion

The challenge we have addressed in this work is relevant to both research into consciousness and clinical practice of depth of anesthesia monitoring, with various techniques developed in order to tackle each (e.g. PCI, Algorithmic Complexity, SWAS). Here, we have shown the ability of deep learning to accurately detect progressive, fine-grained changes in levels of unconsciousness, straight from one second of raw EEG.

The inter-individual variability of the pharmacodynamics impact of propofol, which has been previously reported [7], provides further motivation for the development of more accurate metrics that can be passively and non-invasively obtained during anesthesia. Here, we have

shown the ability of deep learning to learn EEG signatures that track levels of consciousness (almost) linearly, corresponding to the Ramsay scale.

Additionally, we have shown that our model generalized accurately to novel, unseen states. This is highlighted by the model predictions in Dataset 2, where behavioral unconsciousness was determined by auditory discrimination rather than the Ramsay scale. It is also highlighted in the 3- state model trained with Dataset 1, when tested on the unseen Recovery state. Though behaviorally equivalent to Wakefulness (Ramsay score 2), our model slightly overestimated the level of unconsciousness (see Fig. 4), which in turn revealed the EEG signature of residual propofol during Recovery. This was also evidenced by the blood samples taken at the beginning of Recovery in Dataset 2 (average concentration of 0.29 µg/ml, for 20 participants).

Hence in summary, our model was able to generalize to unseen participants (with *leave-one-participant-out* cross-validation), unseen transitional states, and even to unseen experimental setups (by generalization from Dataset 1 to Dataset 2). Future work should focus on extending generalization to novel clinical anesthetics. This would serve as an important demonstration of the ability of EEG to track behavioral consciousness (connected consciousness), and maybe even intraoperative awareness (disconnected consciousness, as revealed from retrospective reports after certain anesthetics, such as ketamine [11]), in a broad set of clinically relevant use cases.

Acknowledgment

References

[1]   Avidan, M.S., Jacobsohn, E., Glick, D., Burnside, B.A., Zhang, L., Villafranca, A., Karl, L., Kamal, S., Torres, B., O'Connor, M., Evers, A.S., Gradwohl, S., Lin, N., Palanca, B.J., Mashour, G.A.: Prevention of Intraoperative Awareness in a High-Risk Surgical Population. N. Engl. J. Med. (2011). doi:10.1056/NEJMoa1100403

[2]   Kreuzer, M.: EEG Based Monitoring of General Anesthesia: Taking the Next Steps. Front. Comput. Neurosci. 11, 1-7 (2017). doi:10.3389/fncom.2017.00056

[3]   Heilmeyer, F.A., Schirrmeister, R.T., Martin, V.: A large-scale evaluation framework for EEG deep learning architectures.

[4]   Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T.H., Faubert, J.: Deep learning-based electroencephalography analysis: A systematic review, (2019)

[5]   Patlatzoglou, K., Chennu, S., Boly, M., Noirhomme, Q., Bonhomme, V., Brichant, J.F., Gosseries, O., Laureys, S.: Deep neural networks for automatic classification of anesthetic-induced unconsciousness. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 216-225 (2018)

[6]   Sanders, R.D., Tononi, G., Laureys, S., Sleigh, J.W.: Unresponsiveness ^ unconsciousness. Anesthesiology. 116, 946-959 (2012). doi:10.1097/ALN.0b013e318249d0a7

[7]   Chennu, S., O'Connor, S., Adapa, R., Menon, D.K., Bekinschtein, T.A.: Brain Connectivity Dissociates Responsiveness from Drug Exposure during Propofol-Induced Transitions of Consciousness. PLoS Comput. Biol. 12, 1-17 (2016). doi:10.1371/journal.pcbi.1004669

[8]   Murphy, M., Bruno, M.-A., Riedner, B.A., Boveroux, P., Noirhomme, Q., Landsness, E.C., Brichant, J.-F., Phillips, C., Massimini, M., Laureys, S., Tononi, G., Boly, M.: Propofol Anesthesia and Sleep: A High-Density EEG Study. Sleep. 34, 283-291 (2011). doi:10.1093/sleep/34.3.283

[9]   Cecotti, H., Jha, G.: 3D Convolutional Neural Networks for Event- Related Potential detection. 2019 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. 4160-4163 (2019). doi:10.1109/embc.2019.8856485

[10]   Wilaiprasitporn, T., Ditthapron, A., Matchaparn, K., Tongbuasirilai, T., Banluesombatkul, N., Chuangsuwanich, E.: Affective EEG-Based Person Identification Using the Deep Learning Approach. IEEE Trans. Cogn. Dev. Syst. (2019). doi:10.1109/TCDS.2019.2924648

[11]   Bonhomme, V., Staquet, C., Montupil, J., Defresne, A., Kirsch, M., Martial, C., Vanhaudenhuyse, A., Chatelle, C., Larroque, S.K., Raimondo, F., Demertzi, A., Bodart, O., Laureys, S., Gosseries, O.: General Anesthesia: A Probe to Explore Consciousness. Front. Syst. Neurosci. 13, (2019). doi:10.3389/fnsys.2019.00036