

Electroencephalogram Interpretable Signal Processing for Anesthesia Predictive Monitoring with Knowledge Distillation

Abstract:

Anesthesia monitoring is a critical aspect of clinical care and psychiatry. Accurate and consistent monitoring of the depth of the anesthetic is essential, not only to ensure patient safety during sedation but also to avoid under or overdosing, which can lead to intraoperative awareness, prolonged recovery times, or adverse psychological outcomes. Despite advances in clinical monitoring tools, the identification of transitions in brain states under anesthesia remains complex. With the increase in neurological disorders globally[1], this highlights the need for reliable methods to monitor brain states during anesthesia. EEG signals offer a promising window into brain function, but this requires trained experts to be able to read and interpret these signals. Currently, there is a shortage of experts or neurologists [2] who can read and interpret EEG due to the complex nature of the EEG signals, which may include noise from muscle activity, clinical environment or even equipment used to collect the signals, making it hard to find meaningful brain activity and the high dimensionality of the data.

Machine learning and deep learning models have shown assured benefits by identifying complex, nonlinear patterns in even huge datasets and recognizing brain states. However, these models are black box in nature, and their lack of interpretability is still a limitation to their usage in healthcare.

Furthermore, the absence of important domain-specific standardized features further limits the clinical adoption of AI-based tools.

Machine learning and deep learning approaches are explored to overcome computational challenges in analyzing brain oscillations using electroencephalography (EEG) data taken from patients administered ketamine. The main goals are to categorize various neural oscillation patterns (delta, theta, alpha, beta, and gamma), monitor anesthesia stages during ketamine administration, and forecast therapy responses for depression by implementing machine and deep learning models to achieve reliable classification of brain states, allowing personalized predictions of treatment efficacy. The performance of the models is assessed based on the output of the different evaluation metrics (accuracy, precision, F1 score, MAE, MSE). KNN followed by the Random Forest model slightly outperformed other models (SVM, Naive Bayes, GBM, XGBoost, LightGBM, CatBoost, Graph Attention Networks(GATs) and graph convolutional networks (GCNs)), showing a 99.78% average accuracy for various data split options. GATs showed consistent results both before and after hyperparameter tuning on both datasets. Knowledge distillation was applied using the GATs model as a teacher model to train a random forest model (student model). The distilled random forest model showed an accuracy of 92.06%.

Furthermore, XAI techniques were used, specifically SHAP, GNNExplainer, and LIME, to evaluate models for feature representation learning and to assess the interpretability of implemented AI learning models. This study has the potential to improve our knowledge of brain oscillations as biomarkers to monitor anesthesia and predict therapy responses, which could ultimately result in safer and more efficient personalized treatment.

[1]World Health Organization (2024) Over 1 in 3 people affected by neurological conditions, the leading cause of illness and disability worldwide. 14 March. Available at:

<https://www.who.int/news-room/detail/14-03-2024-over-1-in-3-people-affected-by-neurological-conditions> (Accessed: 14 September 2024).

[2]N. Aderinto, "Unlocking the hidden burden of epilepsy in Africa: Understanding the challenges and harnessing opportunities for improved care," Health Science Reports, vol. 6, no. 4, 2023.