Executive Summary

Sleep Stage Classification Using EEG signals

Hungry Stayers

Objective

The **goal** of this project is to develop a machine learning model to classify segments of EEG signals collected from animals into one of three sleep stages (classes). The dataset, combined altogether, contains 138240 rows and 5001 columns, which poses a challenge in efficient data processing and model training.

Pre-processing

We reduced the number of columns from 5000 to 500 by uniformly sampling one out of every 10 data points. General patterns of the data are reserved after this reduction.

We split the data into training and test sets. Assuming the future data to be similar recordings of new animals, we treated 2 animals as if they were new (unseen to the model) and trained our model using data for the remaining 6 animals.

Methods

We implemented several different methods including Logistic Regression (LR), K-Nearest Neighbors (KNN), Neural Networks (NN), Random Forest (RF), Recurrent Neural Networks (RNN), and Long Short-Term Memory(LSTM RNN). Each method is evaluated on the test set and compared based on accuracy, precision, and recall.

Results

The best accuracy of 79.73 % is achieved by Random Forest. The 2nd best accuracy was 58.87 % achieved by KNN. Most models perform poorly on Class 0 due to class imbalance in the dataset. RF performed best on Class 1 and Class 2 with high precision and recall.

We attempted to mitigate the class imbalance problem by oversampling but it did not work well. We also tried to improve model performance by considering different experimental conditions which turned out to be not very helpful.

Conclusion

We built multiple machine learning models to get 3 sleep stages. After reducing the dataset uniformly, we used 6 animal data to train and 2 to test. The results of which were compared amongst different models. Multiple ways of improvement were tested including class imbalance and running different models for different experimental conditions.