Predicting Sleep Stage From Human fMRI Data

Hoang Son Do [1] **Hong Niu** [2] **Weixiong Jiang** [3]

Abstract

Sleep has long been believed to play crucial roles in human physiology, modulating responses ranging from supporting brain function to maintaining physical health. Yet despite wide recognition of the importance of sleep and significant progress in understanding the benefits of sleep for memory and cognition, the underlying neurocognitive mechanisms of sleep remain unclear. In this study, we apply machine learning models ranging from support vector machines, gradient boosting, and random forests, to study the relationship between brain regions and patient sleep stages by developing several classifiers using fMRI data.

1. Introduction

Human brains experience varying states of consciousness during different stages of sleep each night (1), which have been classified into rapid eye movement (REM) and non-rapid eye movement (NREM) sleep. NREM sleep is further divided into three stages, ranging from light sleep N1 to deep sleep N3 (2). Different sleep stages are characterized by unique temporal attributes, such as spindles, slow-wave activities, or REM activity, which can be monitored through electroencephalography (EEG) and polysomnography (1). Sleep stages have long been believed to play a crucial role in the reorganization and restoration of cerebral physiology (3), and the reorganization of attention and memory (4). Specifically, memory consolidation during sleep has been well-documented in research, with people recalling more information after sleep than after an equivalent period of wakefulness (5). Despite significant progress in understanding the benefits of sleep for memory and cognitive functioning, the underlying neurocognitive mechanisms remain unclear, and measures of quantifying the sleep stages are still a challenge. Understanding these processes provides valuable insights into brain reorganization and memory formation and has far-reaching implications for aging and mental health (6) (7). In this study, we developed different machine learning methods to classify sleep stages and extract biomarkers to characterize each sleep stage, which is very meaningful for understanding human sleep mechanics.

[1]Department of Computer Science, University of North Carolina, Chapel Hill, USA [2]Department of Biostatistics, University of North Carolina, Chapel Hill, USA [3]Department of Radiology, University of North Carolina, Chapel Hill, USA. Correspondence to: Hoang Son Do <mattdo@email.unc.edu>, Hong Niu <hong.niu@unc.edu>, Weixiong Jiang <wxjiang@med.unc.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

2. Data

Our data consists of fMRI imaging studies over 40 young adult males (24.46 ± 3.56 years), with each sleep session recorded for for 1 - 2 hours at a time. This longitudinal data was then divided up into 2 minute chunks, resulting in 1154 total samples. For each sample, we have the true sleep stage (1-4) as labeled by clinicians, with 16382 total pairs of fMRI interaction signals between distinct brain regions.

2.1. Preprocessing

To ensure that all features contribute equally to the analysis, we applied normalization using the StandardScaler from the scikit-learn library. This step transforms the features by scaling them to have zero mean and unit variance. Normalization is essential, especially when the dataset contains features with different scales and units, as it prevents the model from being biased towards features with larger magnitudes.

2.2. Dimensionality reduction

High-dimensional datasets can pose challenges for machine learning models, such as increased computational complexity, increased noise in the datasets, as well as suffering from the curse of dimensionality. In our dataset in particular, the number of features is significantly greater even than the number of samples. To address this issue, we applied Principal Component Analysis (PCA) for dimensionality reduction. PCA is a linear transformation technique that projects the original features onto a lower-dimensional space while preserving a specified amount of variance in the data. By reducing the dimensionality of the data, PCA can help improve the computational efficiency of our models and

mitigate the risk of overfitting. Additionally, PCA can be particularly useful when the dataset contains redundant or correlated features.

2.3. Data Imbalance

Our sleep stage dataset exhibits an imbalanced distribution of samples across the different stages. The number of samples for each sleep stage is as follows:

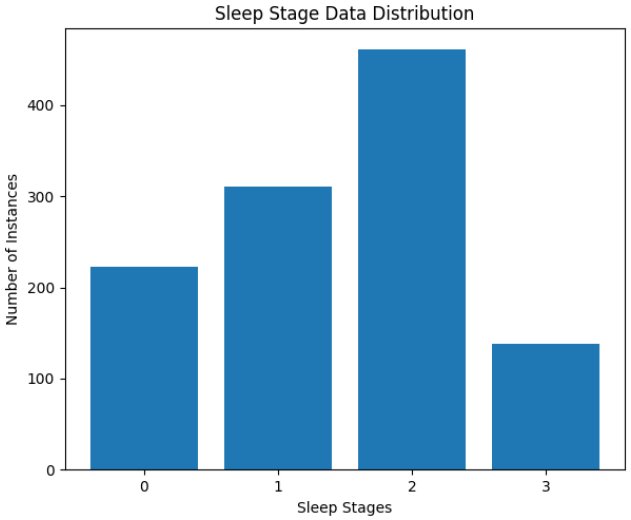

Figure 1. Distribution of sample sizes by sleep stage

We observe that Stage 2 has the highest number of samples, while Stage 3 has the lowest; nothing the imbalance in the distribution of sleep stages, this fact can introduce biases in our models during training and impact their performance.

3. Models

In this project, we aimed to analyze sleep stage data using various machine learning models. The models we employed for our analysis include Gradient Boosting, Support Vector Machines (SVM), and Random Forests. Each of these models offers a unique set of characteristics that can contribute to the accurate classification of sleep stages.The Python scikit-learn package was used in the implementation of each of these models.

3.1. Training and Testing

We used a dataset containing sleep stage information which was split into features and labels. The features represent the measured variables, and the labels correspond to the sleep stages (1-4). We applied dimensionality reduction using PCA to optimize the performance of our models, given the large size of the dataset. To assess the generalization capability of our models, we performed Leave-One-Subject-Out (LOSO) cross-validation. This method is particularly useful when the data is structured with subject-based data, ensuring that each subject appears in the test set exactly once. For our case, these subjects correspond to the subject IDs. This approach simulates a realistic scenario where the model is tested on unseen subjects, ensuring an unbiased evaluation of its performance.

3.2. Support Vector Machines

Support Vector Machines (SVM) is a powerful algorithm used for classification and regression tasks. In the context of sleep stage classification, SVM aims to find the optimal hyperplane that maximizes the margin between different sleep stages in the feature space. This is achieved by transforming the data into a higher-dimensional space, where the classes are more easily separable. The SVM model was trained with a radial basis function (RBF) kernel, which allows the SVM to handle non-linear relationships between the features and the sleep stages, which makes it a powerful algorithm for sleep stage classification. The RBF kernel is defined as:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{2\sigma^2}\right)$$

here, the RBF function calculates the similarity between two data points $\mathbf{x}$ and $\mathbf{x}'$ via the squared Euclidean distance. The σ parameter is the variance and our hyperparameter (8).

3.3. Gradient Boosting

Gradient Boosting is an ensemble method that combines multiple weak learners (typically decision trees) to create a more powerful model. The weak learners are trained sequentially, with each subsequent learner focusing on correcting the errors of the previous one. This can be expressed as $F_m(x) = F_{m-1}(x) + \beta_m h_m(x)$, where $F_m(x)$ is the boosted model at iteration m, $h_m(x)$ is the weak learner, and β_m is the learning rate (9). This allows the model to adapt to complex patterns in the data iteratively. We used the XGBoost implementation (10), which provides several optimizations for improved speed and performance.

3.4. Random Forests

Random Forests is an ensemble learning method that constructs multiple decision trees during the training phase. The final prediction is obtained by aggregating the predictions of individual trees, which can be expressed as $y(x) = \frac{1}{N}\sum_{n=1}^{N} T_b(x)$, where $T_n(x)$ is the output of the n-th tree and the total number of trees is denoted by N (11). This method performs majority vote for classification tasks. Random Forests can efficiently handle large datasets with high-dimensional feature spaces and provide built-in

feature importance measures. This model is known for its robustness against overfitting and its ability to handle both linear and non-linear relationships between variables.

4. Results

Due to the heavy imbalance present in our original dataset, we first sought to classify stage 1 sleep against stage 2 sleep, since these two categories had the most balanced number of samples. In all cases, we first applied PCA retaining 50%, 70%, and 95% of the variance, followed by three different classification methods. This corresponded to retaining the first 27, 300, and 389 principle components. Below we report validation metrics for each pair of dimension reduction and classification method.

4.1. SVM

Here we used PCA followed by SVM classification using the RBF kernel within scikit-learn. It can be seen that the highest cross-validated accuracy was when PCA retaining 95% of the variance was used prior to classification. Furthermore, this method resulted in the highest scores across the board in precision, recall and F1 score as well, suggesting SVM performed best when most of the variance was retained.

Table 1. SVM-RBF

Metric	PCA=0.50	PCA=0.70	**PCA=0.95**
Accuracy	0.62	0.65	0.65
Precision	0.74	0.72	0.76
Recall	0.62	0.65	0.65
F1-score	0.61	0.64	0.63

4.2. Random Forest

Here we used PCA followed by Random Forest classification within the scikit-learn package, using 256 estimators with a max depth of 3. The results are presented below. It can be seen that although accuracy is relatively constant across the PCA dimensions, we see improved precision and comprable recall when the fewest number of PCA components are retained at 50% of the total variance. Again, all scores presented are cross validated using leave one group out CV.

4.3. XGBoost

Here we applied PCA followed by the XGboost classifier 256 estimators and a max depth of 3. Again, in each test, we used leave one group out cross validation, completely removing one subject each time since samples derived from the same subject are likely to be correlated. Similar to the results of the random forest classifier, we find that the best overall results occur when the fewest number of dimensions are retained, however, in this case, we do find improved accuracy scores across the board.

Table 2. Random Forest

Metric	**PCA=0.50**	PCA=0.70	PCA=0.95
Accuracy	0.65	0.65	0.65
Precision	0.68	0.57	0.65
Recall	0.65	0.65	0.65
F1-score	0.62	0.59	0.63

Table 3. XGBoost

Metric	**PCA=0.50**	PCA=0.70	PCA=0.95
Accuracy	0.67	0.61	0.62
Precision	0.78	0.80	0.73
Recall	0.67	0.61	0.62
F1-score	0.67	0.62	0.61

5. Discussion

In this study, we applied common machine learning classification models, Support Vector Machines, Random Forests, and Gradient Boosting, paired with PCA dimensionality reduction to classify human sleep stages from fMRI functional brain data. In our final models, we focused on classifying stage 1 sleep against stage 2 sleep due to the inherent imbalances present in the data. The final sample sizes were 224 samples for stage 1 and 331 for stage 2, resulting in a roughly 40/60% split. We find that our best model, gradient boosting with 50% of the variance retained after applying PCA offered modest improvements over a naive classifier, with a cross validated accuracy of 67%. This is not a completely surprising result, as retaining the fewest number of dimensions resulted in a dimensionality reduction from 16382 features to 27. This is particularly important in our extremely high-dimensional dataset, where the number of features is significantly greater than the number of samples. Despite this, we find that we are able to obtain moderate results in predicting sleep stage, suggesting activation of brain regions is potentially significant in understanding human sleep patterns.

In the future it will be important to both try to improve our results, as well as better understand the relationships explaining our results. We could try more advanced dimensionality reduction methods since this is a critical challenge in our dataset. We may also seek to apply more interpretable methods to get a better understanding of the significance of each feature. Finally, we seek to extend our classification models to predicting all 4 sleep stages from each other in future

work, which will likely require either feature engineering or the collection of new data to balance out the dataset.

References

[1] Yi-Chia Kung, Chia-Wei Li, Shuo Chen, Sharon Chia-Ju Chen, Chun-Yi Z Lo, Timothy J Lane, Bharat Biswal, Changwei W Wu, and Ching-Po Lin. Instability of brain connectivity during nonrapid eye movement sleep reflects altered properties of information integration. *Human Brain Mapping*, 40(11):3192–3202, 2019.

[2] Richard B Berry, Rita Brooks, Charlene Gamaldo, Susan M Harding, Robin M Lloyd, Stuart F Quan, Matthew T Troester, and Bradley V Vaughn. Aasm scoring manual updates for 2017 (version 2.4), 2017.

[3] Eugene Aserinsky and Nathaniel Kleitman. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, 118(3062):273–274, 1953.

[4] Peter Hu, Melinda Stylos-Allan, and Matthew P Walker. Sleep facilitates consolidation of emotional declarative memory. *Psychological science*, 17(10):891–898, 2006.

[5] Sabrina Berres and Edgar Erdfelder. The sleep benefit in episodic memory: An integrative review and a meta-analysis. *Psychological Bulletin*, 147(12):1309, 2021.

[6] Gordon B Feld and Susanne Diekelmann. Building the bridge: Outlining steps toward an applied sleep-and-memory research program. *Current Directions in Psychological Science*, 29(6):554–562, 2020.

[7] Bryce A Mander, Joseph R Winer, and Matthew P Walker. Sleep and human aging. *Neuron*, 94(1):19–36, 2017.

[8] Jean-Philippe Vert, Koji Tsuda, and Bernhard Schölkopf. A primer on kernel methods. *Kernel methods in computational biology*, 47:35–70, 2004.

[9] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[10] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[11] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.