Automated Sleep Staging System with EEG Signal using Machine Learning Techniques

Dr SANTOSH SATAPATHY

Information and Communication Technology
School of Technology
Pandit Deendayal Petroleum University (PDPU)

April 7, 2025

Table Content

Abstract

- Sleep stage classification is crucial for diagnosing sleep-related disorders.
- Processing raw EEG signals is computationally expensive and resource-intensive, especially with large datasets.

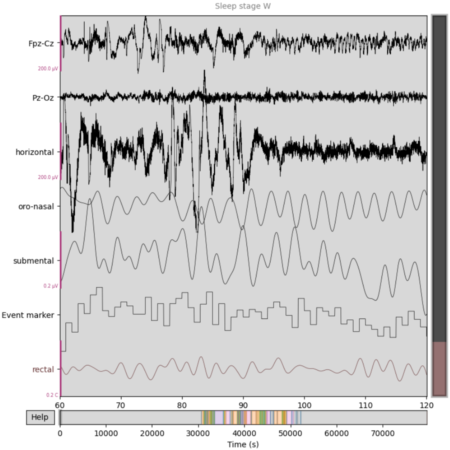

Figure: Visualization of EEG signal.

Abstract

- Extracting, segmenting, and labeling PSG data.
- Identifying and analyzing transitions between sleep stages.
- Using annotated descriptions for validation.

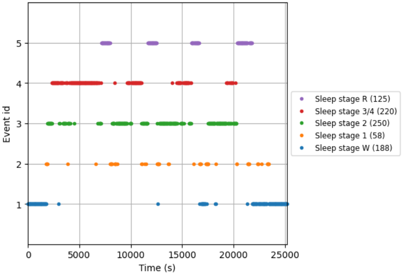

Figure: Sleep stage transition plot.

Sleep Stages and Characteristics

Sleep Stage	Frequency Range (Hz)	Description
Wake (Beta)	12-30	Active, alert state; engaged in cognitive activities.
N1 (Light Sleep)	4-8	Transition stage between wakefulness and sleep; easy to wake up.
N2 (Moderate Sleep)	4-6	Sleep spindles and K-complexes
N3 (Deep Sleep)	0.5-4	Slow-wave sleep; restorative processes occur.
REM (Theta)	4-6	Associated with dreaming; brain activity resembles wakefulness.

Problem Statement

Multi-Channel EEG Analysis for Advanced Sleep Disorder Characterization Using Machine Learning

Motivation

- Enhanced efficiency and scalability for sleep disorder diagnosis.
- Integration potential with consumer-grade devices like wearables for sleep monitoring.
- Opportunities to improve sleep quality analysis using advanced deep learning techniques and personalized health recommendations & disease classification.

Methodology

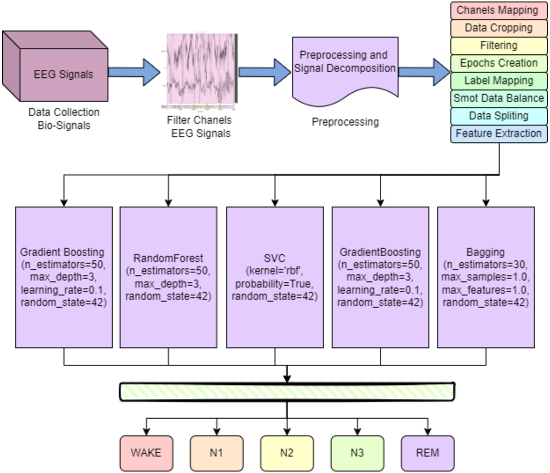

Figure: Proposed Architecture Of Machine Learning

Methodology

Preprocessing Workflow

- **Loading Data**: EEG signals and sleep stage annotations are loaded from the Sleep-EDF dataset using MNE.
- **Filtering**: Applied band-pass filtering (0.3–30 Hz) to remove noise and artifacts.
- **Sleep Stage Event Extraction**: Extracted sleep stage events, filtered out irrelevant annotations, and mapped data into corresponding sleep stages.
- **Feature Extraction**: Converted EEG epochs into numerical feature vectors.
- **Data Balancing**: Applied Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance.

Methodology

Sleep Stages Data Visualization

Table: Summary of Sleep Stage Data

Feature (x)	**Label (y)**
[0.059, 0.596, -0.193, ..., -0.601, 0.201]	W
[-0.022, -0.107, -0.135, ..., 0.038, 0.103]	W
[... (Additional data)]	N1

Note: Data truncated for brevity, showing representative examples of preprocessed input and labels.

Methodology: Class Mapping

Class Mapping Process

- Original sleep stage labels were transformed using a predefined mapping:
  - `'1'` → `'N1'` (Light Sleep)
  - `'2'` → `'N2'` (Intermediate Sleep)
  - `'3'`, `'4'` → `'N3'` (Deep Sleep)
  - `'W'` → `'Wake'`
  - `'R'` → `'REM'` (Rapid Eye Movement)
  - `'e'` → Removed (unnecessary class)

Mapping ensures consistency in class representation across the dataset.

Methodology: Cleaning and Encoding

Cleaning and Label Encoding

- **Mask Application:**
  - Instances of unnecessary classes (`'e'`) were removed using masks.
- **Label Encoding:**
  - Transformed remaining classes into numerical labels using `LabelEncoder`.

Cleaning and encoding prepared the data for machine learning models.

Methodology 1 :Random Forest Classification

Key Steps:

- **Data Splitting:** 80% training, 20% testing.
- **Random Forest Training:** Model trained on extracted features.
- **Prediction:** Classified sleep stages on test data.

Important Parameters:

- **Number of Trees:** 100 estimators
- **Train-Test Split:** 80% train, 20% test
- **Random State:** 42 (for reproducibility)

Methodology 2: Gradient Boosting Classification

Key Steps:

- **Dimensionality Reduction:** Applied PCA (95% variance retained).
- **Gradient Boosting Training:** Model trained on reduced feature set.
- **Prediction:** Classified sleep stages on test data.

Important Parameters:

- **PCA Components:** Retained 95% variance
- **Gradient Boosting:** 30 estimators, max depth = 3
- **Learning Rate:** 0.1
- **Random State:** 42 (for reproducibility)

Methodology 3: Ensemble Learning Classification

Key Steps:

- **Dimensionality Reduction:** PCA (95% variance retained).
- **Ensemble Model Training:** Gradient Boosting, RF, and SVM.
- **Voting Strategy:** Used soft voting for final classification.

Important Parameters:

- **PCA Components:** Retained 95% variance
- **Voting Type:** Soft Voting
- **Gradient Boosting:** 50 estimators, max depth = 3, learning rate = 0.1
- **Random Forest:** 50 estimators, max depth = 3
- **SVM:** RBF kernel, probability enabled
- **Bagging Classifier:** 30 estimators
- **Random State:** 42 (for reproducibility)

Methodology 4: Bagging Classifier

Key Steps:

- **Dimensionality Reduction:** PCA (95% variance retained).
- **Bagging Model Training:** Trained multiple base models on resampled data.
- **Bootstrap Aggregation:** Combined multiple predictions to improve stability.

Important Parameters:

- **PCA Components:** Retained 95% variance
- **Number of Base Models:** 30 estimators
- **Max Samples:** 100% of the dataset
- **Max Features:** 100% of features per base model
- **Random State:** 42

Conclusion

This study demonstrates the potential of machine learning in developing efficient and lightweight models while maintaining strong classification performance. The results highlight that machine learning models can achieve high accuracy and reliability, making them well-suited for deployment on lightweight and edge devices. While deep learning models tend to offer superior accuracy, traditional machine learning approaches prove to be a highly efficient alternative, ensuring optimal performance without the computational overhead. These findings suggest that machine learning can be effectively utilized in real-world applications where resource constraints are a critical factor.

References

- Yan, R., Zhang, C., Spruyt, K., Wei, L., Wang, Z., Tian, L., . . . Cong, F. (2019). Multi-modality of polysomnography signals' fusion for automatic sleep scoring. *Biomedical Signal Processing and Control, 49*, 14–23. doi:10.1016/j.bspc.2018.10.001
- Zhou, J., Tian, Y., Wang, G., Liu, J., Wu, D., Xu, W., . . . Hu, Y. (2020). Automatic Sleep Stage Classification with Single Channel EEG Signal Based on Two-layer Stacked Ensemble Model. *IEEE Access, 1–1*. doi:10.1109/access.2020.2982434
- Shen, H., Ran, F., Xu, M., Guez, A., Li, A., & Guo, A. (2020). An Automatic Sleep Stage Classification Algorithm Using Improved Model Based Essence Features. *Sensors, 20*(17), 4677. doi:10.3390/s20174677
- Huang, W., Guo, B., Shen, Y., Tang, X., Zhang, T., Li, D., & Jiang, Z. (2019). Sleep staging algorithm based on multichannel data adding and multi feature screening. *Computer Methods and Programs in Biomedicine, 105253*. doi:10.1016/j.cmpb.2019.105253
- Ghimatgar, H., Kazemi, K., Helfroush, M. S., & Aarabi, A. (2019). An automatic single-channel EEG-based sleep stage scoring method based on hidden Markov model. *Journal of Neuroscience Methods, 108320*. doi:10.1016/j.jneumeth.2019.108320
- Ghimatgar, H., Kazemi, K., Helfroush, M. S., Pillay, K., Dereymaeker, A., Jansen, K., . . . Aarabi, A. (2020). Neonatal EEG sleep stage classification based on deep learning and HMM. *Journal of Neural Engineering*. doi:10.1088/1741-2552/ab965a

Questions and Answers!

Thank You!