Computer Science and Engineering

DSCI341501 - Fund. of Machine Learning - Spring 2025

Project: Sleep Stages Prediction

Michael Reda - 900203291
Freddy Amgad - 900203088

Table of Contents

Introduction

Disorders can severely impact sleep as a whole and, in return, disrupt health, daily functioning, and even professional life, insomnia, for instance. Multiple approaches can be utilized to address sleep-related issues, yet accurate classification of sleep stages is one of the most fundamental keys, which include light and deep sleep along with REM (rapid eye movement) sleep. Conventionally, PSG, or polysomnography, is used to conduct sleep stage monitoring, which is intricate, invasive, and usually difficult. PSG requires a clinical setting, making it all the more multifaceted and complex.

The development of new technologies for wearable sensors makes it possible to obtain physiological and movement-based metrics like heart rate and acceleration, which can be used as approximate indicators for determining sleep stages. This project, Sleep Stage Prediction from Heart Rate and Motion Data, seeks to use these metrics obtained from wearables to create a model that predicts the stages of sleep using machine learning. The main goal of the project is to create an easy and non-invasive method for determining sleep stages compared to PSG by using data from the Sleep-Accelerometry Dataset.

The issue we intend to discuss here has two sides. Firstly, there is a demand for new simpler methods to track sleep stages outside clinical settings. Secondly, the process of pulling insights from chaotic multi-sensor data presents unique problems. This project applies machine learning techniques to achieve high precision in sleep stage classification, which could enhance personalized sleep health tracking and facilitate the early detection of sleep disorders.

This document describes the techniques applied in data preparation and analysis, outlines the results for different implemented models, and discusses the findings.

Methods

In order to achieve **high** accuracy, we **analyze** and investigate the data we have. Then, we decide to perform several **manipulations** to achieve higher **accuracy:**

1. **Data Synchronization:**
   Raw data from heart rate, acceleration, steps, and sleep stage labels were collected asynchronously. To align these heterogeneous data sources temporally, a nearest-neighbor time merging strategy was applied. This ensured that measurements from different sensors corresponded to the same time intervals, enabling reliable multi-modal feature extraction.

2. **Missing Data Treatment:**
   After merging, any samples containing missing values were removed to maintain data integrity and avoid introducing bias or noise into the training process.

3. **Signal Smoothing:**
   The heart rate measurements were smoothed using a rolling average filter with a window size of five samples. This reduced transient noise and improved the stability of the heart rate signal.

4. **Feature Engineering:**
   - The magnitude of the acceleration vector was calculated from the three-axis acceleration components to represent overall motion intensity.

- Additional temporal features, including rolling mean of heart rate and rolling standard deviation of acceleration magnitude over fixed time windows, were computed to capture short-term trends and variability.

5. **Outlier Removal:**

   To mitigate the influence of anomalous data points, physiological and sensor thresholds were applied:

   - Heart rate values outside the physiologically plausible range (30 to 220 beats per minute) were discarded.
   - Samples with unrealistically high step counts (≥ 1000) or acceleration magnitudes (≥ 50) were also excluded.

6. **Label Simplification:**

   The original sleep stage labels included six categories, with stages 3 and 4 often exhibiting similar characteristics. For modeling purposes and to reduce class imbalance, stage 4 was merged into stage 3, resulting in five distinct sleep stages.

7. **Feature Normalization:**

   All features were standardized to zero mean and unit variance using z-score normalization. This step ensured that all inputs contributed proportionally during model training, especially benefiting algorithms sensitive to feature scales.

8. **Imbalance Correction:**

   Sleep stage data is inherently imbalanced due to varying durations of each stage. To address this, Synthetic Minority Oversampling Technique (SMOTE) was applied on the training set to generate synthetic examples for underrepresented classes, enhancing model generalization and accuracy.

Experimental Design

To evaluate the performance of our sleep stage prediction system and understand the impact of various factors on model accuracy, we conducted a series of experiments structured as follows:

1. **Feature Set Evaluation:**
   The first experiment aimed to assess how the number of features used influences model accuracy. By incrementally increasing the number of features and observing the resulting performance, we were able to determine whether additional features contribute positively to the prediction accuracy or introduce noise that may degrade performance.

2. **Model Comparison:**
   In the second experiment, we evaluated multiple machine learning models to identify which architecture delivers the highest accuracy in sleep stage classification. By comparing performance metrics such as accuracy. We gained insights into the suitability of different models for our dataset and investigated the reasons behind the varying performance across algorithms.

Results

To assess the effectiveness of different physiological signals and model architectures in predicting sleep stages, we conducted multiple experiments using individual features, combinations of features, and all features combined. The performance of each configuration was measured primarily through accuracy, with results summarized below.

1. Single Feature Analysis

- **Heart Rate Only**
  When using heart rate as the sole input feature, the best performance was achieved by the K-Nearest Neighbors (KNN) model with an accuracy of **42.76%**, followed by Linear SVM (**34.01%**) and Logistic Regression (**33.95%**). Most models performed moderately, with Neural Network (MLP) at **25.86%** and CatBoost at **24.63%**.

- **Steps Rate Only**
  The models showed relatively poor performance when using only step rate, with KNN again leading at **21.53%**, while most other models yielded accuracy below 18%. This suggests that step rate alone may not be a strong predictor of sleep stages.

- **Acceleration Only**
  Using acceleration data exclusively, Linear SVM achieved the highest accuracy at **32.47%**, followed by Kernel SVM (**29.62%**) and KNN (**22.97%**). Other models demonstrated limited predictive capability using

this feature alone.

2. Two-Feature Combinations

- **Acceleration + Steps Rate**
  A marginal improvement was observed using this combination. CatBoost achieved **27.62%** accuracy, while Random Forest and KNN followed closely with **26.87%** and **26.31%**, respectively. This suggests a slight benefit in combining features, though still limited.

- **Acceleration + Heart Rate**
  A notable improvement in performance was seen in this combination. Linear SVM and Random Forest both achieved around **34%**, while CatBoost and Logistic Regression followed closely. This indicates that heart rate adds valuable information when paired with acceleration data.

- **Steps + Heart Rate**
  The KNN model yielded the highest accuracy among all two-feature combinations with **45.10%**, followed by Linear SVM (**35.78%**) and Random Forest (**33.28%**), indicating a strong correlation between these features and sleep stage patterns.

All Features Combined (Heart Rate, Step Rate, Acceleration)

When all three physiological signals were used together for training and prediction, the models demonstrated a substantial improvement in accuracy, highlighting the importance of multi-feature integration:

- **Random Forest**: **85.95%**
- **K-Nearest Neighbors (KNN)**: **84.05%**
- **CatBoost**: **77.04%**
- **Neural Network (MLP)**: **70.24%**
- **Kernel SVM (RBF)**: **45.64%**
- **Linear SVM**: **36.60%**
- **Logistic Regression**: **34.02%**
- **Naive Bayes**: **16.79%**

These results clearly demonstrate that combining all three physiological features significantly enhances prediction accuracy. In particular, ensemble models such as Random Forest and CatBoost, as well as KNN, performed best in the multi-feature scenario, validating the importance of feature richness and model choice in sleep stage prediction tasks.

Conclusion

This project set out to develop a machine learning-based system capable of predicting sleep stages using non-invasive data collected from wearable sensors. By leveraging physiological signals such as heart rate, step rate, and acceleration, we explored various feature combinations and machine learning models to determine the most effective approach.

Our experiments revealed that while single features offer limited predictive power, combining multiple modalities significantly improves classification performance. The best results were obtained using all three features together, with the Random Forest and K-Nearest Neighbors models achieving accuracies of 85.95% and 84.05%, respectively. These findings underscore the importance of multi-sensor data fusion and the selection of appropriate model architectures for complex biomedical classification tasks.