|  |
| --- |
| SUPSI – Applied Case Studies of ML and DL in Key Areas II |
| **Optimizing Sleep Spindle Detection with Machine Learning** |
|  |

|  |
| --- |
| Manuel Romanelli, Matteo Vairani & Gabriel Tabacaru |

SYLLABUS

[ABSTRACT 2](#_Toc165836879)

[INTRODUCTION 2](#_Toc165836880)

[PROJECT MANAGEMENT 2](#_Toc165836881)

[TECHNICAL IMPLEMENTATION 3](#_Toc165836882)

[EVALUATION PROCEDURE 4](#_Toc165836883)

[SOLUTION 4](#_Toc165836884)

[CONCLUSION 5](#_Toc165836885)

[SOURCES 6](#_Toc165836886)

# ABSTRACT

This report details the development and evaluation of a personalized sleep spindle detection algorithm based on machine learning techniques, aiming to enhance the detection accuracy of sleep spindles in polysomnographic recordings. Using the DREAMS database for analysis, this project expands upon the foundational work by Scafa et al. (2020), attempting to address its limitations and optimize performance through feature enhancement and a refined machine learning approach.

# INTRODUCTION

Sleep spindles, which are rapid oscillations occurring during the non-rapid-eye-movement (NREM) stage of sleep, play a significant role in sleep architecture and are linked to cerebral plasticity and memory consolidation. Despite their importance, accurate detection of sleep spindles remains a challenge due to variability across individuals and the limitations of existing algorithms when applied to different datasets. The project’s primary goal was to develop a personalized sleep spindle detection algorithm that improves upon the methodologies outlined by Scafa et al. (2020). The study's reliance on support vector machines (SVM) provided a solid baseline for understanding algorithmic behavior in spindle detection but also highlighted issues related to generalizability and the need for individual customization in feature selection and classification thresholds.

# PROJECT MANAGEMENT

The project was initiated with a comprehensive review of existing literature on sleep spindle detection, focusing on understanding different approaches and their respective efficacies. The project plan was structured around a series of phases: initial data analysis, feature testing and selection, classifier training, and iterative testing. By adopting a structured approach to project management, we ensured systematic progress through clearly defined milestones and objectives.

# TECHNICAL IMPLEMENTATION

Immagine che contiene Diagramma, schermata, linea, diagramma

Descrizione generata automaticamenteIn our project, the technical implementation centered around developing a refined approach for detecting sleep spindles from EEG signals using advanced machine learning techniques. The initial step involved preprocessing the EEG data, where signals were loaded using the MNE library, which provides extensive tools for processing and analyzing neurophysiological data. Handling inconsistencies and errors gracefully, we ensured the integrity and uniformity of our dataset. Additionally, we employed signal processing techniques such as resampling and filtering—specifically using a Butterworth bandpass filter—to enhance the quality of the EEG signals, isolating the crucial frequency components vital for accurate spindle detection.

Feature engineering played a pivotal role in our project, given the complexity and variability of EEG data. We crafted a comprehensive suite of features to capture the dynamics of the EEG signals effectively. By implementing a custom FeatureExtractor class, we segmented the signals into manageable windows and calculated a variety of statistical and spectral features for each segment. These features included measures such as entropy, variance, and frequency components, which are critical for identifying the subtle characteristics of sleep spindles amidst other EEG data.

To address the challenge of imbalanced data—a common issue in medical datasets where the event of interest is rare compared to the normal state—we applied several resampling techniques. We experimented with a combination of SMOTEENN, random undersampling, and oversampling techniques to determine the most effective approach for balancing the dataset. This allowed us to enhance the training process and prevent bias towards the more frequent class in our spindle detection algorithm.

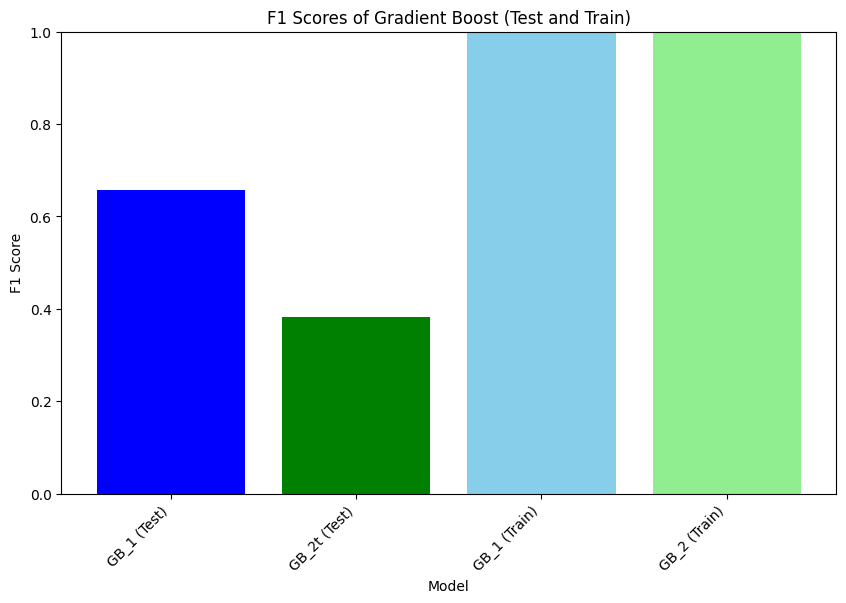
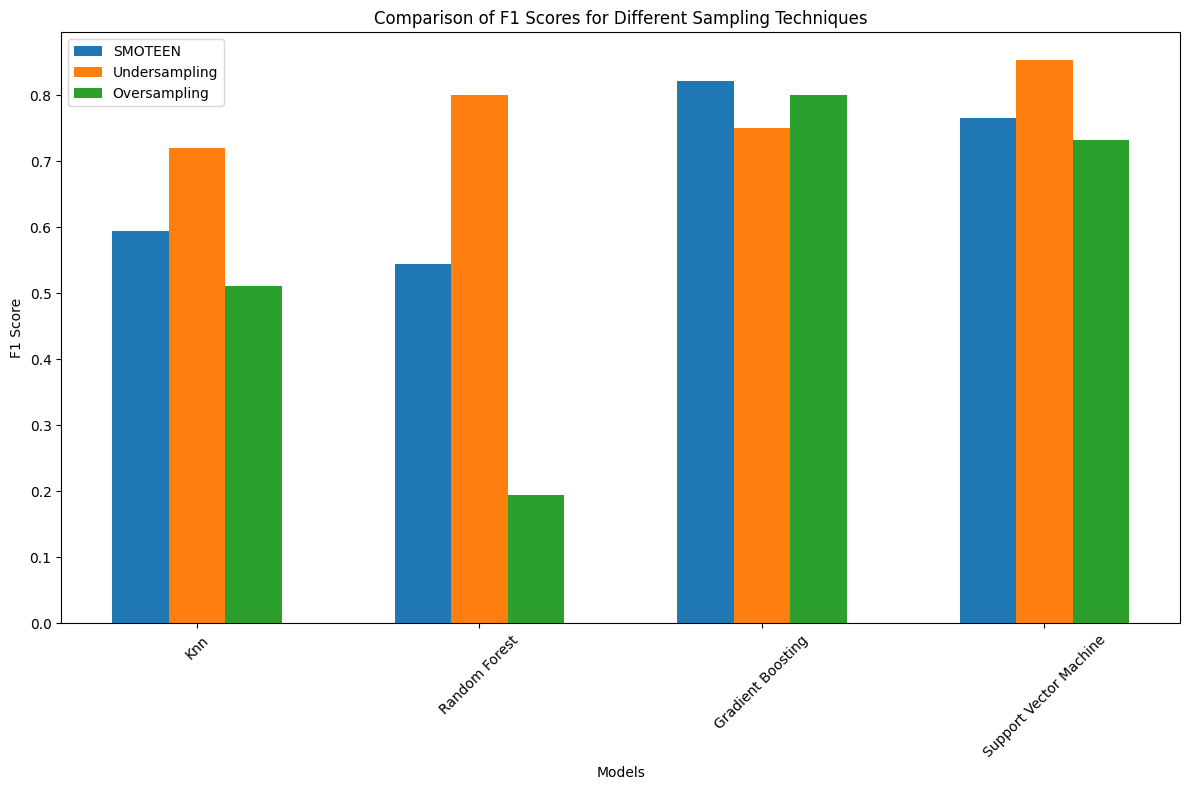
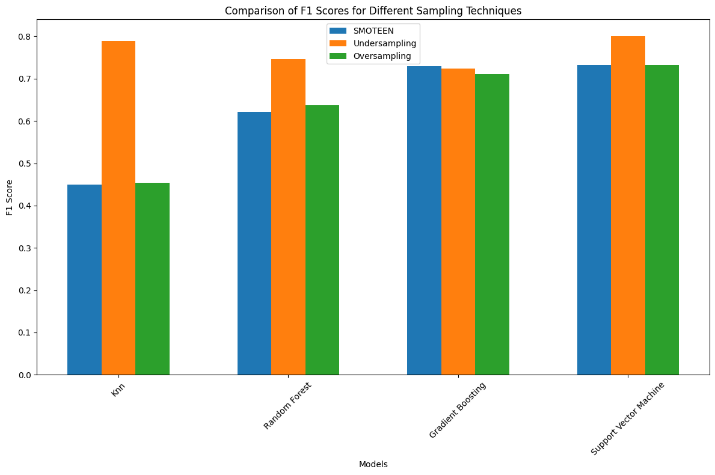
Finally, model selection and evaluation were conducted through a process of training and validating multiple classifiers, including K-Nearest Neighbors, Support Vector Machines, Random Forests, and Gradient Boosting. We utilized GridSearchCV for hyperparameter tuning, optimizing each model based on its performance metrics such as precision, recall, and F1-score. The models were rigorously tested using both the original and balanced datasets to ensure robustness and generalizability. This comprehensive approach allowed us to refine our predictions and enhance the algorithm's accuracy in detecting sleep spindles, pushing the boundaries of what's possible in EEG analysis for sleep studies.

# SOLUTION

The project successfully met its primary objectives by developing a more accurate and reliable sleep spindle detection algorithm. Theoretical and practical knowledge gained from this exercise provided deep insights into both the nature of sleep spindles and the application of machine learning techniques in biomedical contexts. Critical analysis of the results highlighted the importance of personalized approaches in medical diagnostics, particularly in the field of sleep research.

# CONCLUSION

The results obtained from the Gradient Boosting models trained on the two different datasets for sleep spindle detection, ${VS\_1}$ and ${VS\_2}$, exhibit stark disparities between training and testing performances. During the training phase, both models achieved nearly perfect scores, with accuracy and ${f\_1}$ scores reaching or close to 1.0. This level of precision suggests that the models are highly effective in memorizing the patterns and relationships within the training data. However, this seemingly excellent performance in training does not translate well into the testing phase, where both models saw significant drops in performance metrics. For example, the accuracy and ${f\_1}$ scores for VS 1 fell to around 0.76, and for VS 2, these metrics were approximately 0.81. The visual analysis from scatter plots of predicted probabilities versus actual labels further illustrates these challenges, revealing a tendency to either overfit to training data or misclassify new, unseen data instances.



The primary concern highlighted by these results is the issue of overfitting, where the models, though performing optimally on training datasets, fail to generalize their predictions effectively when exposed to new data. This overfitting is evident from the high variance observed in the prediction probabilities on the test data, where many samples expected to exhibit positive indicators were erroneously classified with high confidence as negatives. Overfitting of such a magnitude suggests that the models are capturing noise and specific details in the training data that do not apply more broadly, indicating an overly complex model relative to the underlying patterns in the data that are relevant for generalization.

As seen in the first part where we were trying to find the best model to tune for our task, the performance where much better and there was no overfitting, therefore, it is due to the feature selection, this issue can be addressed in a future by exploring new feature selection techniques and improving the regularization trying to help penalize the complexity of the models and lastly using ensemble methods to combine multiple models predictions.

# SOURCES

<https://zenodo.org/records/2650142#.ZDkhp3ZBw2w>

Spindle Paper