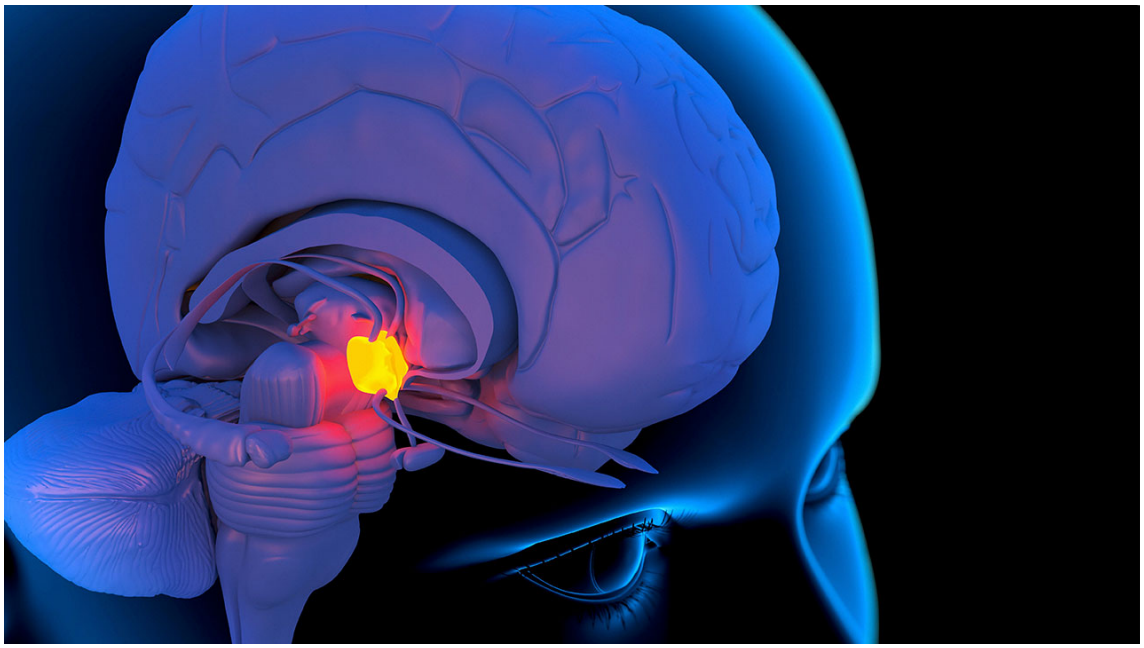
Automatic dream emotion recognition during REM sleep based on minimally invasive EEG

Assistant Engineer Internship Report



Anatole LE BOUDEC–MOLENAAR
Robotic Engineering
FISE 2027
anatole.le_boudec-molenaar@ensta.fr

Septembre 2026

Abstract

Dream emotions remain poorly understood, yet they provide a unique opportunity to study emotional processing during sleep, when the brain is largely disconnected from external stimuli. This report presents two complementary contributions toward the automatic recognition of dream emotions from minimally invasive EEG recordings. First, a sleep-stage classification system was developed and evaluated on the Sleep-EDF Expanded database, containing recordings from 153 subjects. The proposed approach achieved performance comparable to recent single-channel EEG methods. Second, dream emotion recognition was investigated using the REM_Turku dataset, which includes REM sleep recordings from 18 subjects. A rigorous subject-independent validation strategy was used to assess whether the identified EEG patterns could generalize to new individuals. The results show that a reliable EEG signature associated with anger can be identified, whereas the relationships observed for other emotions do not generalize consistently across subjects. Overall, these findings highlight both the potential of EEG for studying dream emotions and the methodological challenges involved in developing reliable subject-independent emotion recognition systems.

Resume

Les émotions ressenties dans les rêves restent encore peu comprises, alors qu'elles offrent une occasion unique d'étudier le traitement des émotions pendant le sommeil, lorsque le cerveau est largement déconnecté des stimuli extérieurs. Ce rapport présente deux contributions complémentaires à la reconnaissance automatique des émotions oniriques à partir d'un EEG peu invasif. La première consiste en le développement d'un système de classification des stades du sommeil, évalué sur la base de données Sleep-EDF Expanded comprenant 153 sujets. Les performances obtenues sont comparables à celles des approches récentes utilisant un seul canal EEG. La seconde contribution porte sur l'étude de la reconnaissance des émotions des rêves à partir de la base de données REM_Turku, qui regroupe des enregistrements de 18 sujets en sommeil paradoxal. Une méthodologie de validation rigoureuse, indépendante des sujets, a été mise en place afin d'évaluer la capacité des résultats à se généraliser à de nouveaux individus. Les analyses montrent qu'une signature EEG associée à la colère peut être identifiée de manière fiable, tandis que les relations observées pour les autres émotions ne se généralisent pas de façon suffisamment robuste. Ces résultats mettent en évidence à la fois le potentiel de l'EEG pour l'étude des émotions oniriques et les défis méthodologiques liés au développement de systèmes de reconnaissance des émotions généralisables.

Abbreviations

List of all abbreviations in alphabetic order:

- **BiLSTM** Bidirectional Long Short-Term Memory
- **CNN** Convolutional Neural Network
- **DE** Dream Experience
- **DEWR** Dream Experience Without Recall
- **EEG** Electroencephalogram / Electroencephalography
- **EMG / EOG** Electromyogram / Electrooculogram
- **FAA** Frontal Alpha Asymmetry
- **FDR** False Discovery Rate
- **ICA** Independent Component Analysis
- **LOSO** Leave-One-Subject-Out
- **LSTM** Long Short-Term Memory
- **mDES** Modified Differential Emotions Scale
- **NE** No Experience
- **NREM** Non-Rapid Eye Movement (sleep)
- **PCA** Principal Component Analysis
- **PSD** Power Spectral Density
- **PSG** Polysomnography / Polysomnographic
- **RF** Random Forest
- **REM** Rapid Eye Movement (sleep)
- **SVM** Support Vector Machine
- **XGBoost** Extreme Gradient Boosting

Contents

Abstract **1**

Abreviations **3**

1 Introduction **6**
1.1 Motivation and Goal . . . 6
1.2 Objectives . . . 6
1.3 Report Outline . . . 6

2 Theory **8**
2.1 Brain anatomy . . . 8
2.2 EEG . . . 9
2.3 Sleep stages . . . 9
2.4 Dreams . . . 10
2.5 Machine Learning and Deep Learning . . . 10
2.5.1 Machine Learning & Deep Learning Approaches . . . 10
2.5.2 LOSO . . . 11
2.5.3 Evaluation Metrics . . . 11

3 State of the Art **13**
3.1 Dreams . . . 13
3.1.1 Decoding Experience in NREM Sleep . . . 13
3.1.2 EEG Correlates of Dream Recall . . . 13
3.2 Dream Emotion Recognition . . . 14
3.2.1 DEED . . . 14
3.2.2 Sikka . . . 14

4 Methods **15**
4.1 Datasets . . . 15
4.1.1 Sleep-EDF expanded . . . 15
4.1.2 REM_Turku . . . 15
4.2 Feature extraction and validation strategy . . . 17
4.3 Code Framework . . . 18
4.4 Common EEG preprocessing . . . 19

5 Results and Discussion **21**
5.1 Sleep Staging . . . 21

5.1.1 Progression of models 21
5.1.2 Classical machine learning 22
5.1.3 Deep learning 22
5.1.4 Per-stage results 22
5.1.5 Comparison with the literature 24
5.2 Emotion Recognition 25
5.2.1 Replication attempt: frontal alpha asymmetry and anger 25
5.2.2 Broader feature search 26
5.2.3 Mapping across the 20 mDES emotion categories 29
5.2.4 Multivariate exploration 31
5.2.5 Comparison with the original study (Sikka et al., 2019) 31
5.2.6 Limitations and Perspectives 32

6 Conclusion **33**

CHAPTER 1

Introduction

1.1 Motivation and Goal

Emotions play a central role in human cognition, decision-making, and mental health. While emotion recognition from EEG has been extensively studied during wakefulness, sleep — and REM sleep in particular — remains comparatively unexplored, despite being the physiological stage most closely associated with dreaming and emotional memory consolidation [39, 31]. Automatic recognition of dream emotions from EEG could open new avenues for sleep and mental health monitoring, using lightweight, minimally invasive hardware rather than full polysomnographic montages. This project addresses this goal in two stages: first, by building a robust sleep-staging pipeline capable of identifying REM sleep from EEG alone, and second, by investigating whether the emotional content of dreams occurring during REM sleep can be decoded from EEG signals recorded at the moment of awakening.

1.2 Objectives

The objectives of this internship were threefold:

- Develop and evaluate a sleep-staging pipeline on a standard EEG dataset, progressing from classical machine learning to deep learning architectures.
- Identify a suitable public dataset combining REM sleep EEG with emotion-annotated dream reports, and establish a rigorous, subject-independent methodology for testing EEG-emotion relationships given the typically small sample sizes available in this research area.
- Investigate whether specific discrete emotions (as opposed to general valence or arousal) are associated with distinguishable EEG correlates during REM sleep, and assess how these findings relate to prior published results on the same dataset.

1.3 Report Outline

Chapter 2 – Theory: Introduces the theoretical background required for this work, including brain anatomy, EEG fundamentals, sleep stages, dreaming, and the machine learning concepts

used throughout the project.

Chapter 3 – State of the Art: Reviews previous research on dream decoding and emotion recognition from EEG, with a particular focus on dream detection, public dream datasets, and existing machine learning approaches.

Chapter 4 – Methods: Describes the datasets used, the EEG preprocessing pipeline, and the software framework developed for both sleep staging and dream emotion recognition.

Chapter 5 – Results and Discussion: Presents the performance of the proposed methods for sleep staging and dream emotion recognition, followed by an analysis of the results, comparisons with the literature, and a discussion of their limitations.

Chapter 6 – Conclusion: Summarizes the main contributions of this internship, discusses the limitations of the proposed approaches, and outlines possible directions for future work.

CHAPTER 2

Theory

2.1 Brain anatomy

The human brain is the control center of the nervous system. It processes sensory information, controls movement, generates thoughts, stores memories, and influences emotions and behavior. The brain consists of three main structures: the cerebrum, the cerebellum, and the brain stem.

The cerebrum constitutes the largest part of the brain. It is divided into four lobes, each serving specific functions.

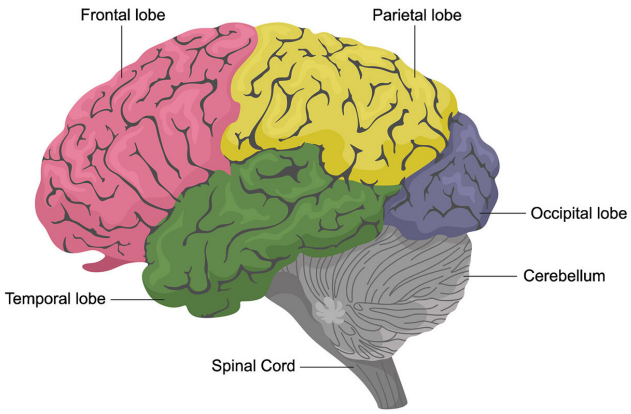


Figure 2.1: Brain anatomy

The frontal lobe is involved in speech production, voluntary movement control, understanding and responding to others' emotions, personality formation, motivation, and attention management.

The parietal lobe processes sensory information such as touch, pressure, pain, temperature, and tension. It also contributes to spatial perception, body orientation, written language comprehension, and mathematical reasoning.

The occipital lobe is primarily responsible for processing visual information. It enables depth perception, color recognition, distance estimation, and the recognition of faces and objects.

The temporal lobe plays an essential role in long-term memory, emotions, sound processing, and speech interpretation. It also forms an important part of the limbic system [18].

2.2 EEG

An electroencephalogram (EEG) is a technique used to record the electrical activity of the brain via electrodes placed on the scalp. The analysis of brain signals enables the identification of various neurological conditions, such as epilepsy, certain brain tumors, or sleep disorders.

In 1875, Richard Caton became the first to record electrical activity in the cortex of animals [6]. In the 1920s, Hans Berger achieved the first non-invasive recording of electrical activity in the human brain, marking the beginning of modern electroencephalography [3].

Today, five main categories of brain waves are recognized: delta, theta, alpha, beta, and gamma waves [36]. Each is associated with a specific state of brain activity.

In the context of this project, EEG is used to record brain activity during sleep to identify different sleep stages based on variations in brain wave patterns.

2.3 Sleep stages

A night of sleep consists of multiple cycles averaging 90 to 110 minutes in duration. An adult typically goes through four to six cycles per night. Each cycle alternates between non-rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep. Each stage possesses distinct characteristics regarding both brain activity and the body's physiological state [4].

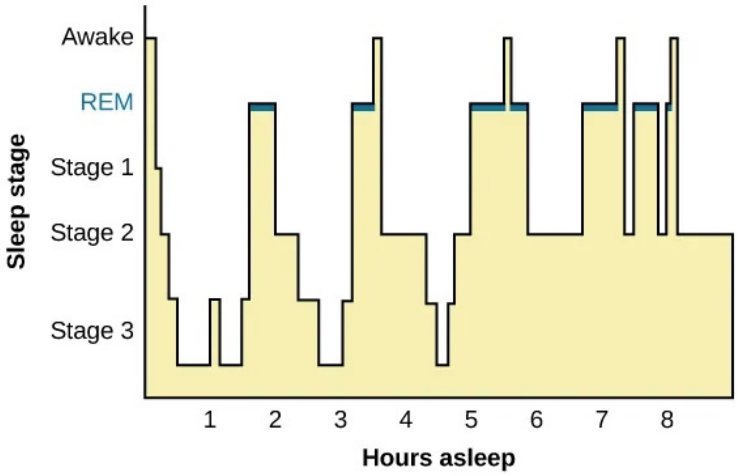


Figure 2.2: Multiple cycles during sleep

Wakefulness (W) : The brain mainly exhibits alpha and beta waves. Muscle tone is normal, and the individual is conscious of their surroundings.

Stage N1 (∼5%) : This is the lightest stage of sleep. Alpha waves progressively disappear, replaced by low-amplitude theta waves. This stage represents the transition between wakefulness and sleep. (EEG: low-amplitude theta waves.)

Stage N2 (∼45%) : Heart rate slows, body temperature decreases, and sleep becomes deeper. This stage is characterized by the appearance of sleep spindles and K-complexes. (EEG: sleep spindles and K-complexes.)

Stage N3 (∼25%) : Stage N3 corresponds to deep sleep. Brain activity is dominated by delta waves, which are low-frequency and high-amplitude. This phase is particularly critical for physical recovery. (EEG: delta waves.)

REM sleep (∼25%) : REM sleep is primarily associated with dreaming [1, 11]. Although brain activity closely resembles that observed during wakefulness, skeletal muscles are almost completely paralyzed, except for the eye muscles and the diaphragm. (EEG: beta waves, similar to those observed during wakefulness.)

2.4 Dreams

Dreams occur mainly during REM sleep, although they can also occur during other sleep phases [11, 23, 33]. They correspond to complex mental activity during which the brain integrates memories, emotions, and sensory inputs [31]. The precise mechanisms governing their formation remain only partially understood today.

2.5 Machine Learning and Deep Learning

2.5.1 Machine Learning & Deep Learning Approaches

To perform sleep staging and emotion recognition from EEG signals, two distinct paradigms are evaluated: classical feature-based machine learning classifiers and end-to-end deep learning architectures.

Feature-Based Machine Learning

Classical pipelines rely on manually engineered features extracted in the time, frequency, and time-frequency domains from multi-channel EEG epochs.

- **Random Forest (RF):** An ensemble of decision trees trained via bootstrap aggregation (bagging). RF [5] mitigates overfitting on high-dimensional feature spaces and serves as a robust baseline for both sleep staging and dream emotion classification.
- **Support Vector Machine (SVM):** A kernel-based classifier [10] optimizing hyperplane separation with maximum margin. Using Radial Basis Function (RBF) kernels, SVM projects non-linearly separable EEG power spectral density features into higher dimensions.
- **Extreme Gradient Boosting (XGBoost):** A gradient-boosted decision tree framework [8] that builds weak learners sequentially to minimize residual errors. It handles feature dependencies and missing values effectively in multivariate EEG feature matrices.

End-to-End Deep Learning

Unlike conventional methods, deep architectures directly process raw single-channel or multi-channel EEG waveforms ($x \in R^{C \times T}$), bypassing manual feature extraction.

- **1D Convolutional Neural Networks (1D CNN):** Utilizes temporal 1D kernels [21] to automatically filter raw 30 s EEG epochs, capturing local morphological features (e.g., sleep spindles, K-complexes, delta waves) across varying frequency bands.
- **Hybrid CNN-BiLSTM:** Combines spatial/local temporal feature learning via a 1D CNN with sequence-level context modelling via Bidirectional Long Short-Term Memory (BiLSTM) units [16, 32, 35]. While the CNN extracts epoch-level feature representations, the BiLSTM models inter-epoch transition dynamics (e.g., progression across sleep stages) over consecutive time windows.

2.5.2 LOSO

Leave-One-Subject-Out (LOSO) cross-validation is a evaluation strategy where all data from an entire subject (a participant, patient, or user) is held out as the test set, while the model is trained on data from all remaining subjects. This procedure is repeated iteratively so that every subject serves as the test set exactly once. This technique is crucial for assessing a model's ability to generalize to unseen individuals, eliminating the risk of data leakage inherent to pooled or repeated measurements [38].

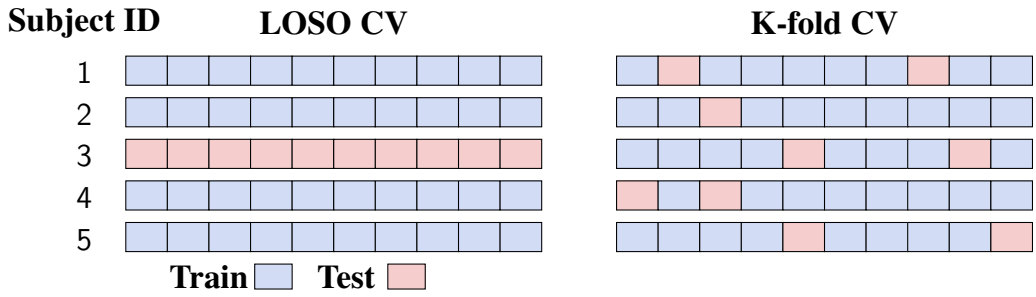


Figure 2.3: Comparison between leave-one-subject-out and k-fold cross-validation.

2.5.3 Evaluation Metrics

Due to the inherent class imbalance in EEG datasets—for instance, REM sleep accounts for only 7% of epochs while stage N2 represents 29%—standard Accuracy alone is insufficient as it over-penalizes minority classes. Classification performance is thus evaluated using a combination of confusion-matrix-derived metrics and chance-corrected agreement indices.

Table 2.1 summarizes the primary evaluation metrics computed per class. For multi-class settings (such as 5-stage sleep scoring), macro-averaged Precision, Recall, and F_1-score are reported to weight each stage equally regardless of its frequency.

Table 2.1: Summary of standard evaluation metrics derived from confusion matrix counts.

Metric	Description & Focus	Formula
Accuracy	Overall correct predictions ratio. Sensitive to class imbalance.	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	Reliability of positive predictions (minimizes false alarms).	$\frac{TP}{TP+FP}$
Recall	Sensitivity; proportion of actual positive samples identified.	$\frac{TP}{TP+FN}$
Macro F_1-score	Harmonic mean of Precision and Recall; robust baseline for imbalanced data.	$2 \cdot \frac{Precision \cdot Recall}{Precision+Recall}$

Cohen's Kappa (κ)

To explicitly account for chance agreement—a critical factor in heavily imbalanced multi-class sleep scoring—Cohen's Kappa (κ) [9] is selected as the primary benchmark metric:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where p_o represents the observed agreement (overall Accuracy) and $p_e = \frac{1}{N^2} \sum_k n_{k,\text{row}} n_{k,\text{col}}$ denotes the expected chance agreement based on marginal distributions.

Unlike standard F_1-scores that require post-hoc averaging schemes, Cohen's Kappa naturally extends to multi-class classification and penalizes random guessing models. Following the standard interpretation by Landis and Koch [20], $\kappa > 0.60$ represents substantial agreement, whereas $\kappa > 0.80$ indicates almost perfect agreement. It serves as the main comparative metric throughout Chapter 5.

CHAPTER 3

State of the Art

3.1 Dreams

3.1.1 Decoding Experience in NREM Sleep

This thesis tackles the binary classification of dream experience (DE) versus no experience (NE) during NREM stage-2 sleep, using the high-density Tononi dataset [33] (36 subjects, 256 channels). Combining multivariate empirical mode decomposition (MEMD) [29] with power spectral density (PSD) features, the author reaches 97.9% classification accuracy (Kappa = 0.96), with beta-band power contributing most to the classification, ahead of delta, theta and alpha. A greedy, PSD-guided channel-selection algorithm shows that a 64-channel montage preserves over 95% of full-montage performance, and that even 20 channels exceed 90% accuracy — roughly a 92% reduction in electrode count, a result directly relevant to portable Brain-Computer Interface hardware. However, LOSO validation reveals a substantial generalization gap (mean accuracy around 59%, Kappa $\approx$ 0.21), only partly offset by adding gamma-band power and metadata (age, sex). A complementary sex-classification experiment using only dream epochs reaches 77% accuracy (Kappa = 0.49) and reveals distinct spectral topographies between sexes: heightened cerebellar delta/theta activity in women versus more symmetrical frontotemporal delta hotspots in men. The thesis stresses that leakage-aware, subject-independent validation is essential, since models performing well under random cross-validation can degrade sharply on unseen subjects.

3.1.2 EEG Correlates of Dream Recall

This thesis represents an initial step toward an automated "dream decoder" that identifies, from EEG alone, whether a sleeping subject is having a dream experience. Using two public datasets from the DREAM database [41] — the Zhang & Wamsley (ZW) dataset [42] (58 channels, NREM and REM sleep) and the Tononi dataset [33] (256 channels, NREM sleep) — the authors combine PCA and common spatial patterns (CSP), as well as PSD, as feature-extraction methods, with MEMD filtering applied to improve classification. Results are strong across both datasets: up to 99.3% accuracy (Kappa = 0.987) on the Tononi dataset using MEMD, PSD and XGBoost, and up to 99.8% accuracy on the ZW dataset during REM sleep using MEMD, PCA, CSP and k-nearest neighbors; NREM classification on ZW using the same pipeline reaches 94% accuracy (Kappa = 0.84). Permutation-based channel-importance analysis points toward parietal and frontal regions as relevant for distinguishing

dreaming from non-dreaming states, though the authors are careful to note that firm anatomical conclusions are not yet possible, and they flag several methodological caveats around data leakage (segment-based train/test splits, MEMD applied before channel selection) as priorities for future work. Two papers based on this thesis were submitted to the ESANN 2024 conference [25, 37].

3.2 Dream Emotion Recognition

3.2.1 DEED

This paper [22] introduces DEED, the first publicly available EEG dataset built specifically for dream emotion research, collected from 38 participants across 89 full nights of sleep (yielding 533 usable dream segments). The protocol combines emotional induction through film viewing while awake with repeated REM awakenings, dream reports, and subjective mood ratings. The authors report a significant correlation between alpha-band power and subjective dream-emotion ratings, along with hemispheric asymmetry differences (measured at F3–F4, FT7–FT8, and T7–T8) between positive and negative dream emotions. They further validate the dataset by testing several machine-learning classifiers (kNN, SVM, LightGBM, GBDT, RF) on PSD, differential entropy, and CSP features, reaching up to about 88% accuracy for binary (positive/negative) emotion classification on 20-second segments.

3.2.2 Sikka

This study [34] examines whether EEG frontal alpha asymmetry (FAA, defined here as the F4–F3 power difference) is linked to affective experience in REM sleep dreams. Seventeen participants slept two nights in the laboratory and were woken five minutes after each REM onset to report their dream and rate its emotional content. FAA recorded both during REM sleep and during evening pre-sleep wakefulness significantly predicted the intensity of anger reported in the following dream (but not interest), an effect driven specifically by greater alpha power over the right frontal electrode (F4). The authors interpret this through the "supervisory control" account of FAA [30]: reduced right frontal activity would reflect a weaker capacity to inhibit strong affective states, not only while awake but also while dreaming, suggesting continuity of the neural mechanisms underlying emotion regulation across wakefulness and REM sleep.

CHAPTER 4

Methods

4.1 Datasets

4.1.1 Sleep-EDF expanded

The Sleep-EDF Database Expanded (sleep-edfx), hosted on PhysioNet, is an expanded version of the original 2002 Sleep-EDF dataset [19, 13], including the Sleep-Cassette (SC) subcategory with 153 overnight polysomnographic (PSG) recordings from healthy subjects aged 21 to 35 and patients with mild difficulties falling asleep, monitored over two consecutive nights. The EDF files contain EEG signals (two channels, 100 Hz), EOG (one channel, 100 Hz), EMG, and sometimes additional signals such as temperature or respiration. Each recording is accompanied by a manually annotated hypnogram according to the Rechtschaffen and Kales (R&K) [28] or American Academy of Sleep Medicine (AASM) [4] rules, with 30-second epochs classified as Wake (W), REM, light and deep sleep stages (N1, N2, N3) of NREM, and unscored segments.

Label	Sleep stage	Count	%
0	Awake	3698	16.5
1	N1	5944	26.5
2	N2	6378	28.4
3	N3	4553	20.3
4	REM	1560	7
	Unknown	174	0.8
	Movement	123	0.5

Table 4.1: Distribution of sleep stages in the dataset

4.1.2 REM_Turku

The REM_Turku dataset, originating from a study published in the Journal of Neuroscience [34], comes from the DREAM database [41] and was collected in a sleep laboratory. Participants were systematically awakened after 5 minutes of rapid eye movement (REM) sleep during the phasic phase, and were then asked to describe their last mental image and provide a detailed dream report. They subsequently rated their emotions using the mDES

questionnaire [12]. If no dream was recalled (complete absence or "white dream"), the mDES was not completed. This protocol was repeated throughout the night, until the final awakening between 5:30 a.m. and 8:30 a.m. The dataset therefore contains dream reports, affective ratings, and associated physiological data. Dream reports were categorized into three types: Dream Experience (DE), No Experience (NE), and Dream Experience Without Recall (DEWR).

Type	Number of reports	%
DE	125	62.5
NE	45	22.5
DEWR	30	15.0

Table 4.2: Distribution of dream types in the dataset

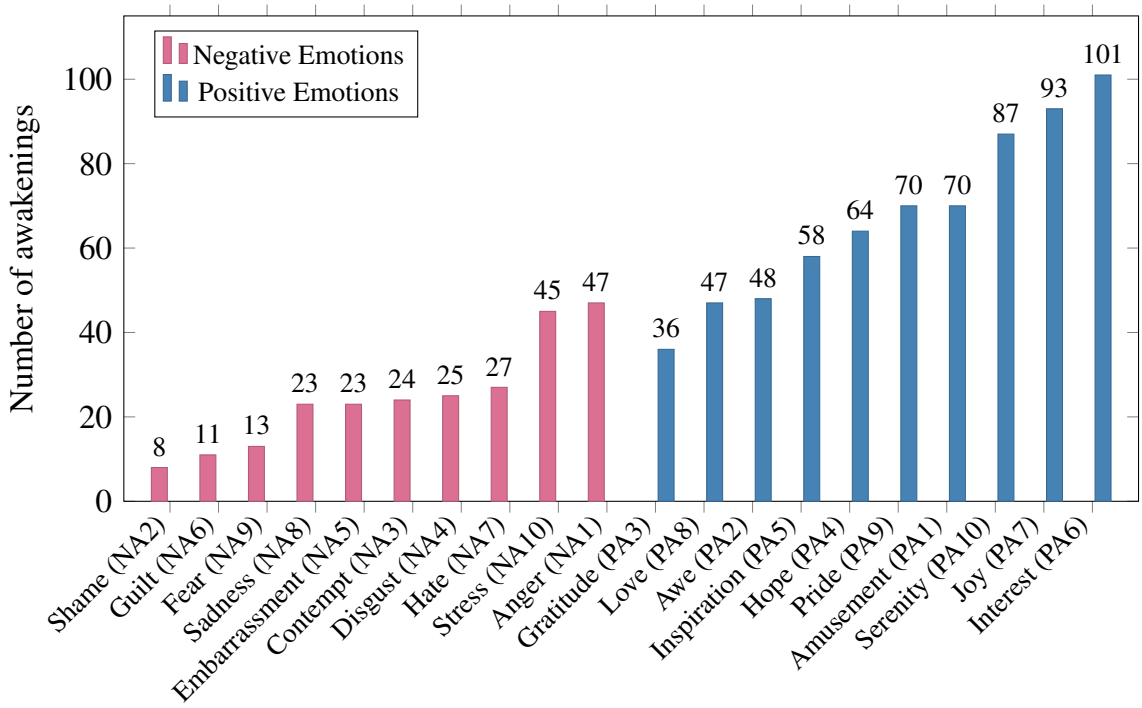


Figure 4.1: Distribution of positive (PA) and negative (NA) emotions upon awakening

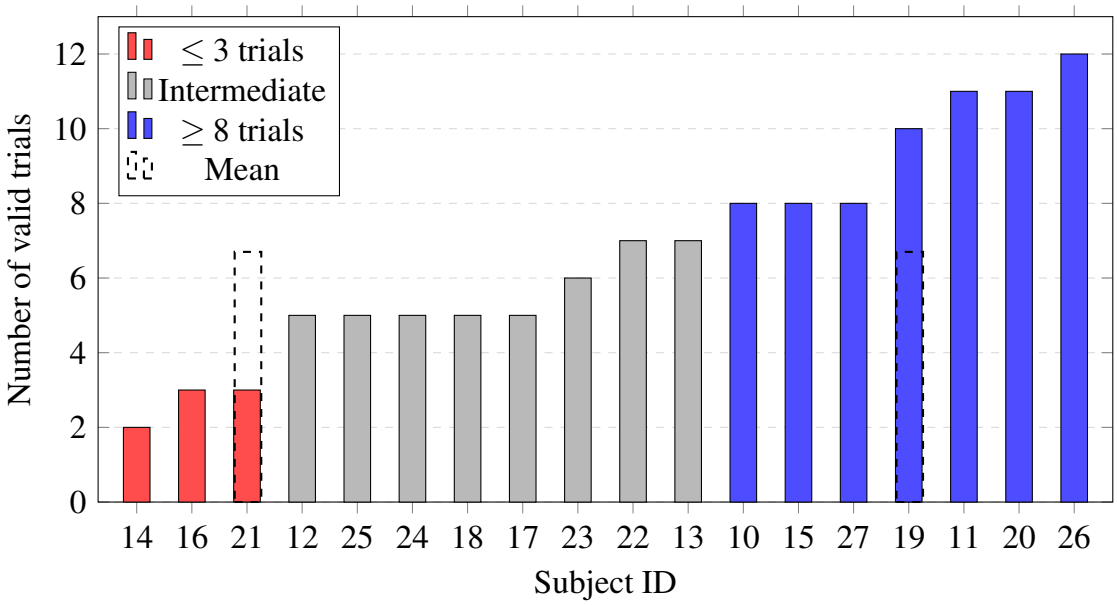


Figure 4.2: Distribution of valid REM trials per participant used in the Leave-One-Subject-Out (LOSO) evaluation

4.2 Feature extraction and validation strategy

Among the 125 Dream Experience (DE) reports available in the REM_Turku dataset (Table 4.2), 121 REM awakenings had usable, artifact-free EEG recordings and were retained for feature extraction; the remaining 4 were excluded due to corrupted or incomplete EDF segments. For each of these 121 REM awakenings, EEG signals were preprocessed using a standardized pipeline consisting of a 1–45 Hz band-pass filter, a 50 Hz notch filter, and conservative ICA-based ocular artifact removal [17]. Note that the binary classification experiments reported in Section 5.2 (e.g., Anger, Pride) further restrict this set to 113 awakenings, after excluding trials with missing mDES ratings for the emotion under study; this explains the slightly different sample sizes (n = 113–121) reported across analyses in this chapter. Power spectral density (PSD) was estimated using Welch's method [40] across five canonical frequency bands (delta, theta, alpha, beta, gamma) for each of the 24 EEG channels, yielding 120 spectral features. Five frontal alpha asymmetry (FAA) indices [34, 30] were additionally computed as the log-power difference between electrodes F4 and F3 for each frequency band, resulting in a total of 125 EEG features per REM awakening.

Given the relatively small number of participants (n = 18) compared with the number of extracted features, two methodological safeguards were applied throughout the study. First,LOSO cross-validation [38] was used to evaluate model generalization on entirely unseen participants rather than unseen trials from previously observed individuals. Before classification, each EEG feature was independently standardized within each participant using z-score normalization to reduce stable inter-individual differences in EEG power while preserving within-subject emotional variability. Second, statistical significance was assessed using permutation testing (500 iterations). At each permutation, emotion labels were randomly shuffled and the complete LOSO classification pipeline was repeated from

model training to evaluation, producing an empirical null distribution from which p-values were estimated [24].

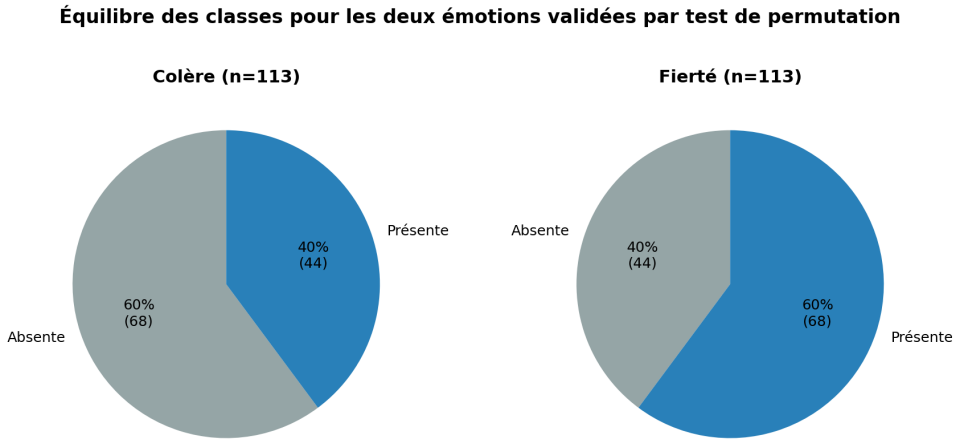


Figure 4.3: Class balance for the two emotions subjected to full validation (Anger, Pride).

4.3 Code Framework

All signal processing and machine learning components developed during this internship were implemented in Python. Although the final objectives differed between sleep staging and emotion recognition, both studies relied on a common software architecture built around a shared EEG preprocessing pipeline. This design ensured methodological consistency while allowing task-specific feature extraction and machine learning approaches.

The implementation mainly relied on the following scientific libraries:

- **MNE-Python**: EEG preprocessing (channel typing, filtering, ICA) and signal manipulation [14];
- **scikit-learn**: feature selection, classical machine learning algorithms (Random Forest, Support Vector Machine and Logistic Regression), cross-validation and statistical evaluation [27];
- **PyTorch**: implementation and training of the CNN–Bidirectional LSTM architecture [26] presented in Section 5.1.3.

Figure 4.4 summarizes the overall software architecture developed during this internship. Both projects share an identical preprocessing stage before diverging toward dedicated pipelines for sleep staging and emotion recognition.

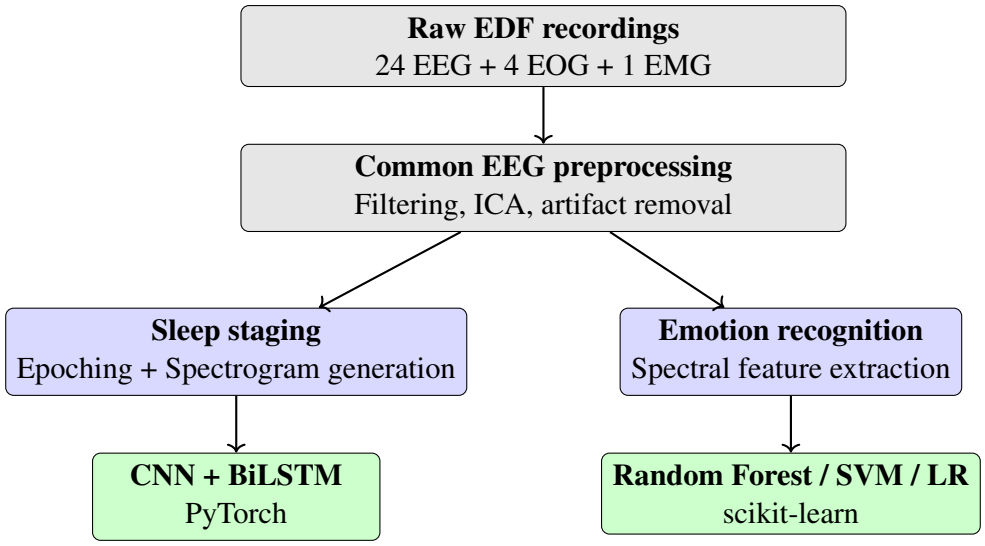


Figure 4.4: Overall software framework developed during the internship. Both studies share a common preprocessing pipeline before following task-specific analysis workflows.

4.4 Common EEG preprocessing

Although the downstream analyses differed between sleep staging and emotion recognition, both relied on the same preprocessing workflow. A common pipeline was therefore implemented to ensure methodological consistency throughout the internship.

For each REM awakening, the raw EDF recording was imported into MNE-Python and channel types were first assigned to distinguish EEG, EOG and EMG signals. The EEG data were then band-pass filtered between 1 and 45 Hz using a zero-phase FIR filter, followed by a 50 Hz notch filter to suppress power-line interference.

Ocular artifacts were removed using Independent Component Analysis (ICA). Twenty independent components were estimated, after which components correlated with the EOG channels were automatically detected using `ica.find_bads_eog()`. Only components clearly corresponding to eye movements were rejected, providing a conservative artifact correction while preserving physiological brain activity.

The resulting cleaned EEG recordings constituted the common input for both studies. In the sleep staging task, these signals were transformed into time-frequency representations before being processed by the deep learning model. For emotion recognition, spectral EEG features were extracted and used for statistical analyses and classical machine learning models.

Figure 4.5 summarizes the complete preprocessing workflow.

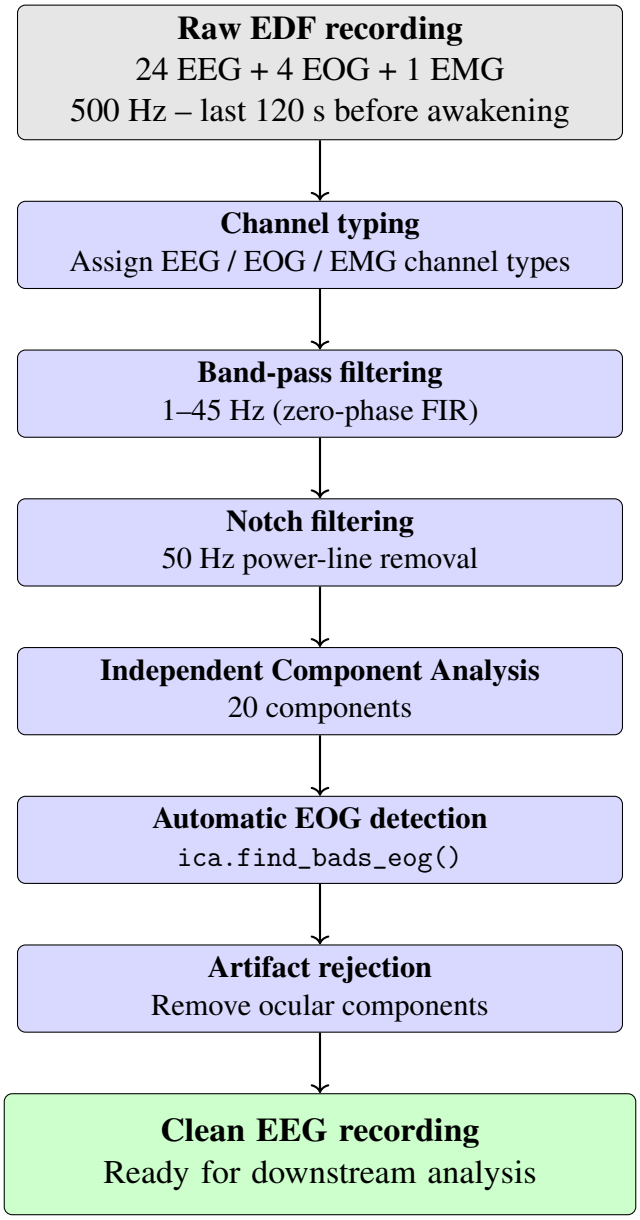


Figure 4.5: Common EEG preprocessing pipeline applied before both sleep staging and emotion recognition analyses.

CHAPTER 5

Results and Discussion

5.1 Sleep Staging

The sleep staging pipeline was developed and evaluated on the Sleep Cassette subset of the Sleep-EDF Expanded database, comprising 153 healthy subjects recorded at home. Table 4.1 reports the raw label distribution across the full database (22,430 epochs). After excluding the *Unknown* and *Movement* labels and, following standard practice for the Sleep-EDF Expanded database, cropping the extended wake periods preceding and following each recorded sleep period (during which the subject was not wearing the device or was awake for an extended time), the working dataset was reduced to 20,084 epochs of 30 seconds each, distributed across five stages: Wake (16%), N1 (25%), N2 (29%), N3 (22%), and REM (7%). The strong class imbalance, particularly for REM, represented the main methodological challenge throughout this work.

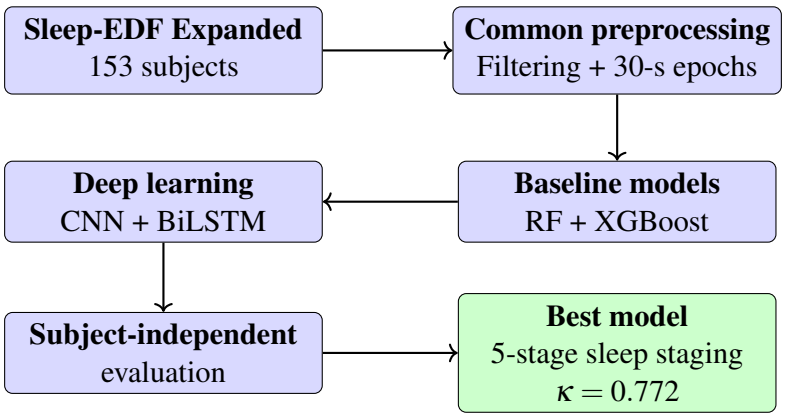


Figure 5.1: Overall workflow of the sleep staging study.

5.1.1 Progression of models

Several progressively more sophisticated models were developed during the internship. Rather than directly adopting a deep learning architecture, classical machine learning approaches were first investigated in order to establish robust baselines and identify the main methodological limitations. The successive improvements are summarized in Table 5.1.1.

Model	Kappa	Notes
Random Forest (2 classes)	0.398	Baseline, binary REM/NREM
Random Forest (5 classes + context)	0.628	Temporal context window of 3 epochs
XGBoost (5 classes)	0.639	Marginal improvement over RF
CNN 1D alone	0.533	Raw signal, no manual features
CNN + Bidirectional LSTM	0.772	Best result obtained

Table 5.1: Models performance

5.1.2 Classical machine learning

A Random Forest classifier was first trained on 25 hand-crafted spectral and temporal features extracted from each epoch, including band power in the delta, theta, alpha, sigma and beta ranges, as well as Hjorth parameters [15] and spectral entropy. To capture the sequential nature of sleep, a temporal context mechanism was added, concatenating the features of the three preceding epochs to each sample, expanding the feature vector from 25 to 100 dimensions. This alone pushed the Kappa from 0.421 to 0.628, confirming that sleep stage transitions carry strong discriminative information. SMOTE oversampling [7] was applied exclusively on the training set to mitigate the REM class imbalance.

5.1.3 Deep learning

The final architecture combines a 1D CNN encoder with a bidirectional LSTM, inspired by the DeepSleepNet framework [35]. The CNN processes each 30-second epoch (3,000 time points at 100 Hz) and produces a 256-dimensional embedding, capturing local EEG patterns such as sleep spindles and delta waves. The LSTM then models the temporal sequence of embeddings across an entire night, allowing the model to exploit inter-epoch dependencies that the classical approach only partially captured through the fixed context window. The model was trained using a subject-wise train/test split (108 subjects for training, 27 for testing) to prevent data leakage, and an early stopping strategy based on validation Kappa was applied.

5.1.4 Per-stage results

The CNN+LSTM model achieves its best performance on N3 (F1 = 0.90) and N2 (F1 = 0.84), where the EEG signature is most distinctive. Wake (F1 = 0.82) and N1 (F1 = 0.79) are also well classified. REM remains the most challenging stage with an F1 of 0.59, primarily due to its visual similarity to Wake on a single EEG channel and its low prevalence in the dataset. Notably, the recall for REM is only 0.46, meaning the model misses more than half of the actual REM episodes. This limitation is inherent to single-channel EEG systems: clinical scoring relies additionally on EOG for eye movements and EMG for muscle atonia, both of which are strong discriminators between REM and Wake.

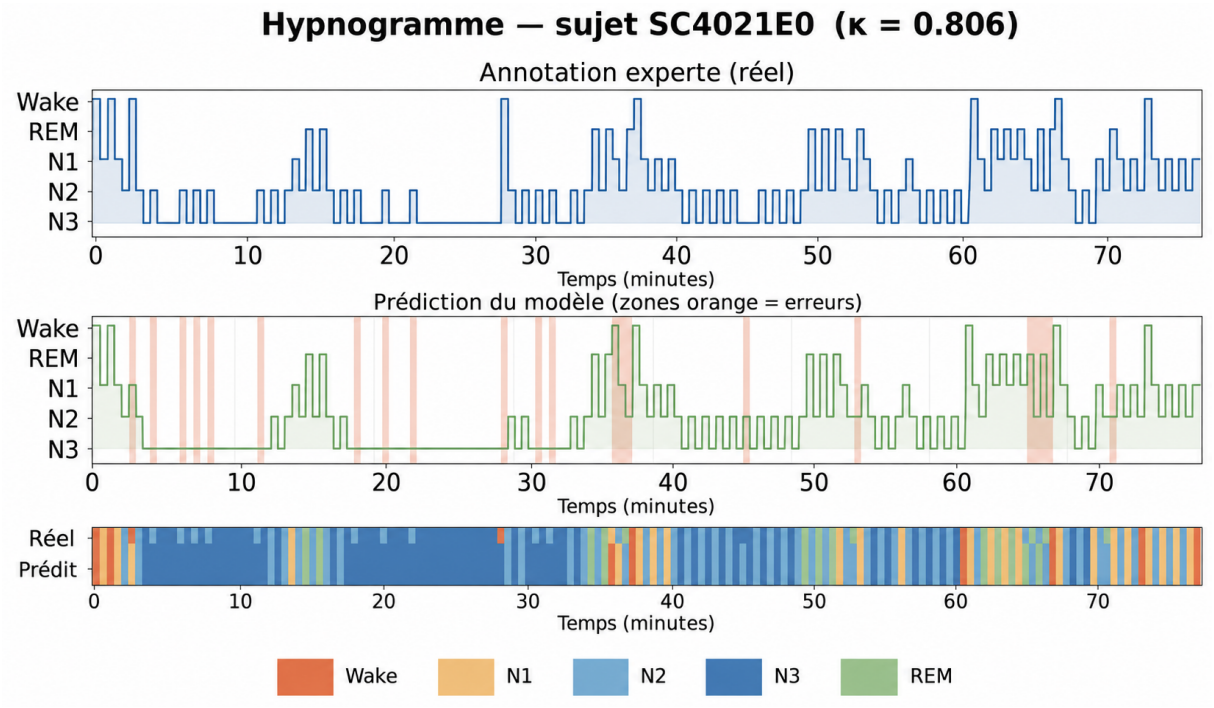


Figure 5.2: Ground-truth hypnogram and model prediction for subject SC4021E0. Orange bands in the predicted hypnogram indicate misclassified epochs.

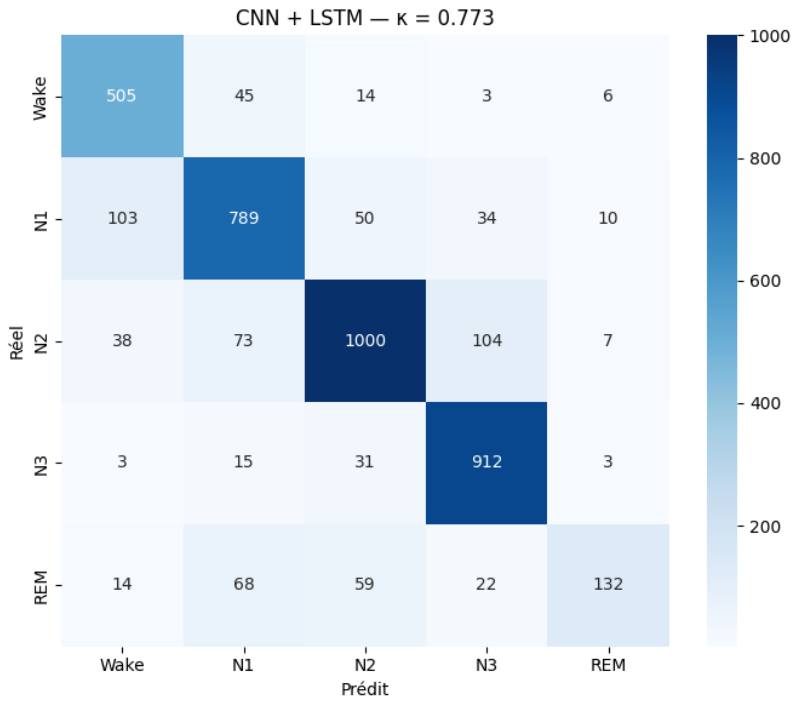


Figure 5.3: Confusion matrix of the CNN + BiLSTM model on the held-out test subjects ($\kappa = 0.772$). REM is the stage most frequently confused with N1 and N2, consistent with the lower recall reported above.

5.1.5 Comparison with the literature

The obtained Kappa of 0.772 is competitive with published Random Forest baselines (typically 0.65–0.75) and approaches state-of-the-art deep learning systems trained on multi-channel PSG recordings (0.80+). The gap with the best published results is largely attributable to the use of a single EEG channel, as opposed to the full polysomnographic montage used in most reference studies.

5.2 Emotion Recognition

Having established a robust sleep staging pipeline, the second objective of this internship was to investigate whether EEG activity recorded immediately before REM awakenings also contains information about dream emotions. Unlike sleep staging, which relies on temporal sequence modelling, emotion recognition was formulated as a feature-based classification problem. The analysis therefore followed a different methodological strategy summarized in Figure 5.4.

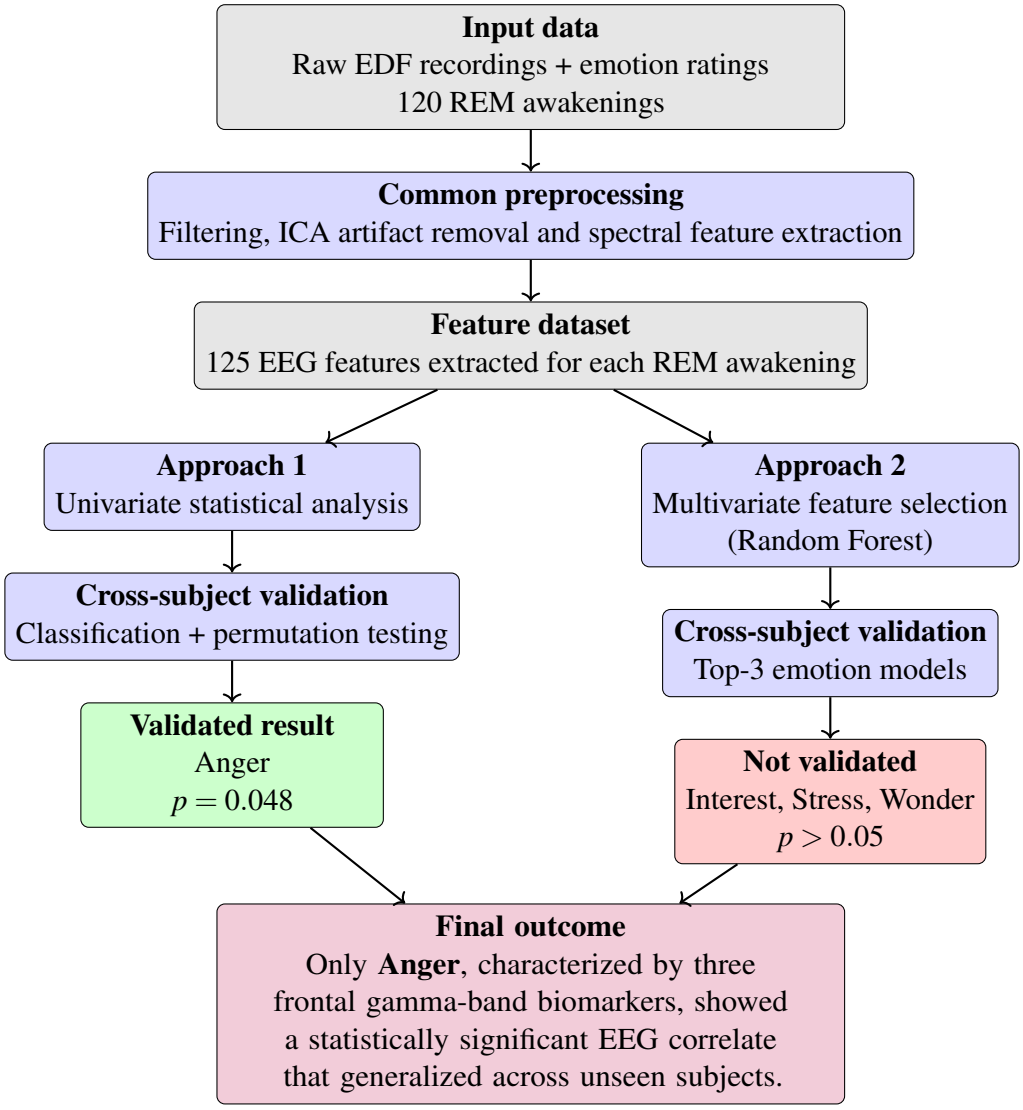


Figure 5.4: Overall workflow of the emotion recognition study. After a common preprocessing and feature extraction stage, the extracted EEG features were analyzed using two complementary strategies: a univariate statistical approach and a multivariate machine learning approach. Both strategies were independently validated on unseen subjects.

5.2.1 Replication attempt: frontal alpha asymmetry and anger

As a first step, the relationship between frontal alpha asymmetry (FAA) and anger reported by Sikka et al. [34] was directly evaluated using the five FAA features as input to both

Random Forest and Support Vector Machine classifiers. Neither model outperformed a chance-level baseline under LOSO validation (accuracy 49.2%–51.7% versus 59.2% for a stratified Dummy classifier). Although this represents a negative replication, it does not necessarily contradict the original findings, which were obtained using pooled correlation analyses rather than subject-independent prediction. Unlike the sleep-staging task (Section 5.1), which relies on Cohen's Kappa to account for class imbalance across five stages, emotion recognition results are reported using accuracy, F1-macro, and empirical p-values obtained via permutation testing. This choice reflects the binary nature of each emotion classification task and provides a data-driven, rather than theoretical, correction for chance level — directly accounting for the specific class distribution and subject structure of each LOSO fold.

5.2.2 Broader feature search

Following the unsuccessful FAA replication, a data-driven search was conducted across all 125 extracted EEG features. Spearman correlations were computed independently between each feature and the anger score, followed by Benjamini-Hochberg false discovery rate (FDR) correction [2] for multiple comparisons.

Only three EEG features survived FDR correction, all corresponding to gamma-band activity over highly frontal electrodes (Fp1, Fp2 and AF4; $r = 0.30$–0.32, $p_{FDR} \approx 0.048$). Notably, these were the only three significant features found when testing anger specifically against the 125 EEG measures, suggesting a localized association between frontal gamma activity and dream-related anger. As shown in Section 5.2.3, other emotions (Pride, Stress) exhibit a considerably larger number of univariately significant features; however, only anger's association survives the subject-independent LOSO validation described below, illustrating that univariate significance alone is not sufficient evidence of a generalizable EEG correlate.

A Random Forest classifier trained exclusively on these three features achieved an accuracy of 61.1% and an F1-macro score of 0.597 under LOSO validation, compared with 55.8% accuracy and an F1-macro of 0.550 for the Dummy baseline. Permutation testing [24] confirmed that this improvement exceeded empirical chance level (p = 0.048). This p-value should nonetheless be interpreted with caution: it lies close to the conventional 0.05 threshold, was estimated from only 500 permutations, and anger was tested for generalization through two related pipelines (univariate feature selection and, in Section 5.2.4, multivariate feature selection). No formal correction for this combined testing was applied, so the result is best regarded as a promising but not definitive finding that would benefit from replication on an independent, larger sample.

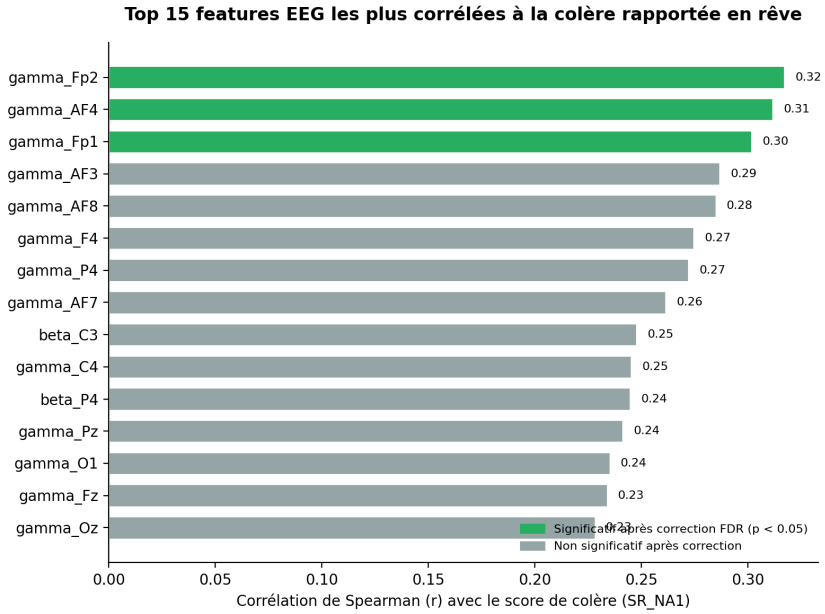


Figure 5.5: Top 15 EEG features most correlated with anger. Features highlighted in green survived FDR correction across 125 statistical tests.

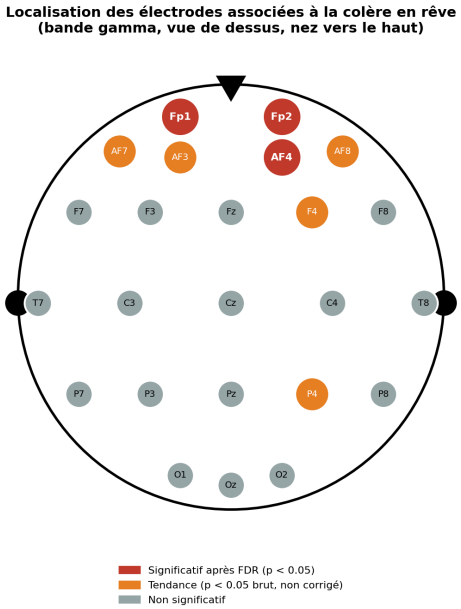


Figure 5.6: Scalp localization of the EEG features significantly associated with anger.

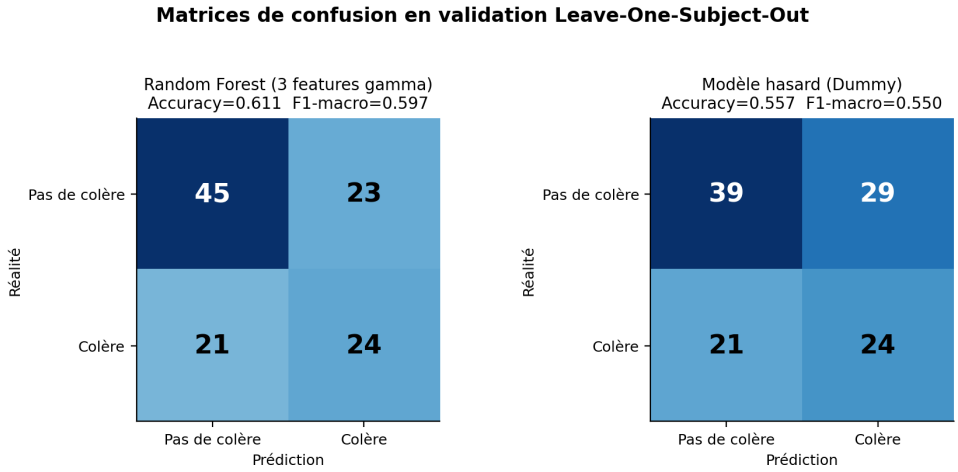


Figure 5.7: Confusion matrices comparing the Random Forest classifier (three gamma features) with the Dummy baseline under LOSO validation.

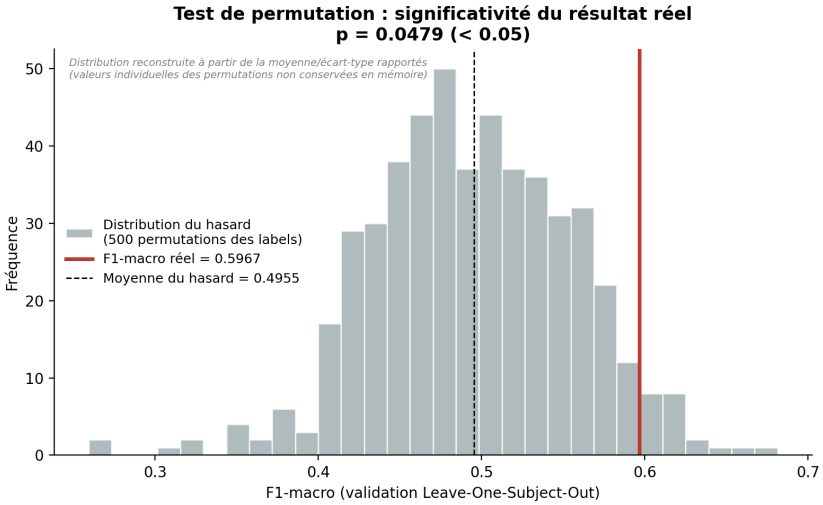


Figure 5.8: Permutation test (500 iterations) for anger classification. The observed F1-macro significantly exceeded the empirical chance distribution (p = 0.048).

5.2.3 Mapping across the 20 mDES emotion categories

The same correlation and FDR-correction procedure was repeated independently for each of the 20 discrete emotions assessed by the mDES questionnaire. Four emotions showed at least one significant EEG feature after correction: Anger (3 features), Pride (25), Stress (22) and Love (6). The remaining sixteen emotions showed no statistically significant EEG correlates at this exploratory stage.

Pride, Stress and Love were subsequently evaluated using the same validation framework applied to anger, namely LOSO classification followed by permutation testing. Despite Pride exhibiting the strongest pooled correlation observed across all emotions ($r = -0.42$), none of the three candidates achieved statistical significance (Pride: p = 0.291; Stress: p = 0.156; Love: p = 0.407). Likewise, an additional analysis based on a global arousal score, computed as the sum of all twenty emotion ratings, also failed to generalize (p = 0.681).

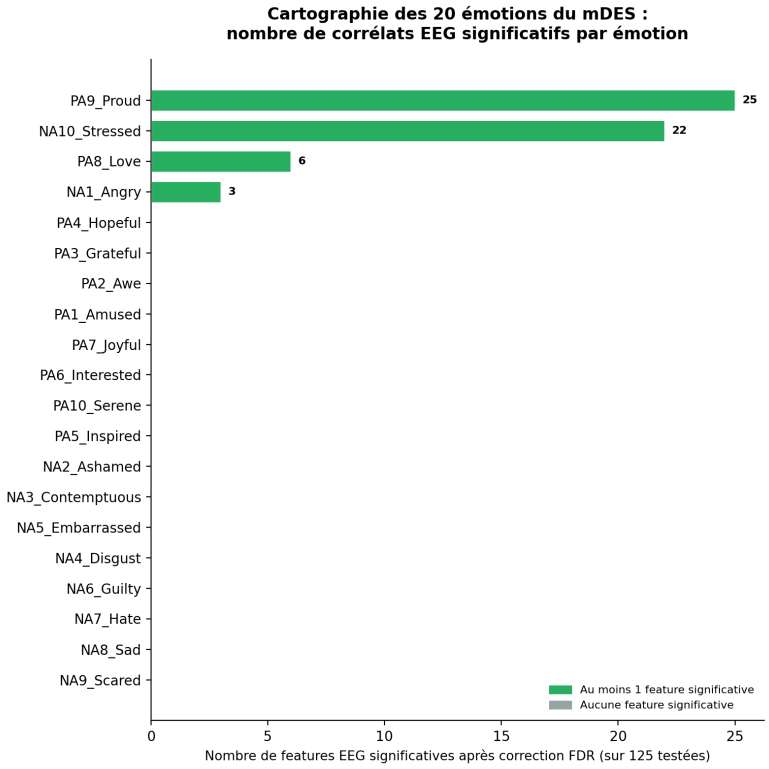


Figure 5.9: Number of significant EEG features (after FDR correction) identified for each of the 20 mDES emotion categories.

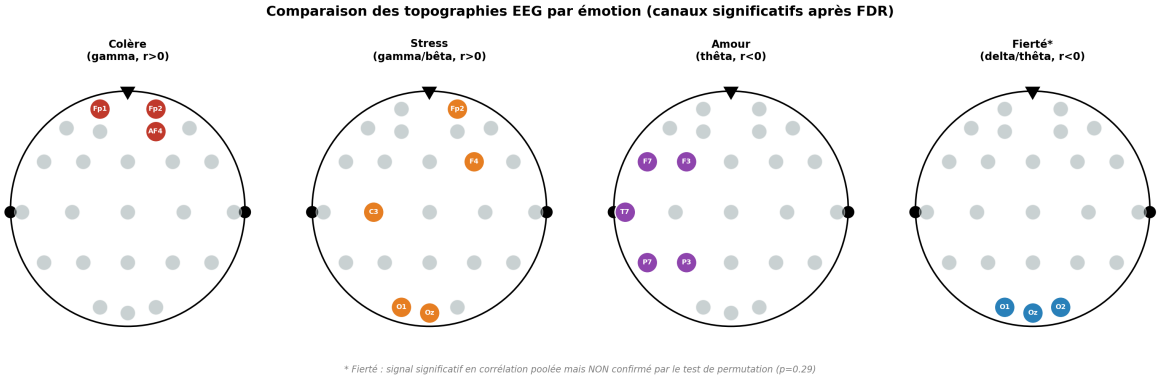


Figure 5.10: Comparison of EEG topographies across the four candidate emotions. Significant electrodes after FDR correction are highlighted.

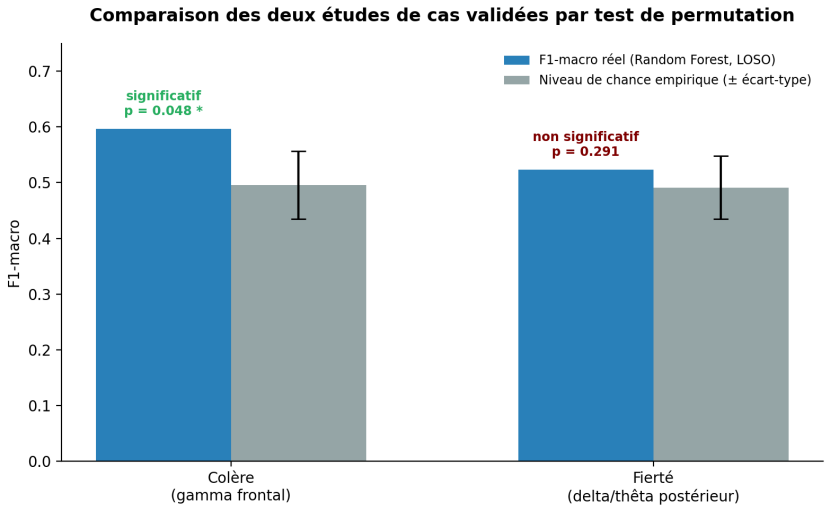


Figure 5.11: Comparison between observed F1-macro and empirical chance level for the fully validated emotion classification experiments.

These findings indicate that pooled correlation strength alone is not a reliable indicator of subject-independent prediction performance. Strong correlations may instead reflect stable between-subject differences rather than neural signatures directly associated with emotional experience during individual dreams.

5.2.4 Multivariate exploration

Because univariate correlation screening may overlook distributed or non-linear EEG patterns, a multivariate feature-selection strategy was also investigated. For each emotion, the ten most informative EEG features were selected using mutual information. Importantly, feature selection was recomputed independently within every LOSO training fold to avoid information leakage. Random Forest classifiers were then trained using the selected features.

The same nested strategy was maintained during permutation testing, with feature selection repeated independently within every permutation and every LOSO fold to ensure an unbiased estimate of statistical significance.

This multivariate screening identified several additional candidate emotions showing performance above empirical chance before statistical validation, including Interest, Stress, Awe and Pride. However, subsequent permutation testing failed to confirm these effects (Interest: $p = 0.264$; Stress: $p = 0.697$; Awe: $p = 0.498$). Consequently, anger remained the only emotion showing a statistically significant and generalizable EEG signature across both the univariate and multivariate analytical frameworks.

5.2.5 Comparison with the original study (Sikka et al., 2019)

The original publication [34] reported substantially stronger pooled correlations between frontal alpha asymmetry and dream emotions ($r = 0.638$ for anger and $r = 0.572$ for hate) than those observed in the present analyses (approximately $r = 0.02$–0.05). A likely explanation lies in the unit of analysis. Whereas the original study appears to have computed correla-

tions after averaging observations at the participant level (n = 18), the present work relied on individual REM awakenings (n = 113–120), thereby introducing substantially greater trial-to-trial variability. Consequently, direct comparison between the reported correlation coefficients should be interpreted cautiously.

Overall, the analyses demonstrate the importance of combining exploratory feature screening with rigorous subject-independent validation. Across hypothesis-driven replication, univariate screening, multivariate feature selection, LOSO classification and permutation testing, only frontal gamma activity associated with anger consistently generalized to unseen participants. These findings suggest that reliable EEG correlates of dream emotions are likely to be more specific and less widespread than suggested by pooled correlation analyses alone.

5.2.6 Limitations and Perspectives

Several limitations should be kept in mind when interpreting these results. First, the REM_Turku dataset includes only 18 participants and 113–121 usable REM awakenings, which limits statistical power and makes LOSO estimates sensitive to individual subjects; the anger finding ($p = 0.048$) should therefore be treated as a preliminary result requiring replication on a larger, independent cohort before any clinical or applied interpretation. Second, the sleep-staging pipeline relies on a single EEG channel and consequently underperforms on REM detection (recall = 0.46) compared with full polysomnographic montages that include EOG and EMG; incorporating even a single auxiliary EOG channel could plausibly close much of this gap. Third, the emotion-recognition analysis was restricted to spectral power and frontal asymmetry features; it did not explore connectivity measures, non-linear dynamics, or deep-learning representations directly learned from raw EEG, which future work could investigate now that a validated sleep-staging pipeline is available to automatically select REM segments. Finally, combining the sleep-staging and emotion-recognition pipelines into a single end-to-end system – detecting REM sleep and decoding dream emotions from EEG alone, without relying on scheduled awakenings – remains an open direction toward the lightweight, minimally invasive monitoring envisioned in the introduction.

CHAPTER 6

Conclusion

Overall, the analyses demonstrate the importance of combining exploratory feature screening with rigorous validation procedures. Across multiple analytical strategies—including hypothesis-driven replication, univariate feature screening, multivariate feature selection, subject-independent classification, and permutation testing—only frontal gamma activity associated with anger consistently generalized to unseen participants. These findings suggest that robust EEG correlates of dream emotions are likely to be more specific and less widespread than suggested by pooled correlation analyses alone.

Bibliography

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

[2] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995.

[3] Hans Berger. Über das elektrenkephalogramm des menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87(1):527–570, 1929.

[4] Richard B. Berry, Rita Brooks, Charlene E. Gamaldo, Susan M. Harding, Carole L. Marcus, and Bradley V. Vaughn. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.0*. American Academy of Sleep Medicine, Darien, IL, 2012.

[5] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[6] Richard Caton. The electric currents of the brain. *British Medical Journal*, 2(765):278, 1875.

[7] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

[8] 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.

[9] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

[10] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[11] William Dement and Nathaniel Kleitman. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology*, 53(5):339–346, 1957.

[12] Barbara L. Fredrickson. Positive emotions broaden and build. In Patricia Devine and Ashby Plant, editors, *Advances in Experimental Social Psychology*, volume 47, pages 1–53. Academic Press, 2013.

[13] Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[14] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7:267, 2013.

[15] Bo Hjorth. EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3):306–310, 1970.

[16] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[17] Aapo Hyvärinen and Erkki Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4–5):411–430, 2000.

[18] Khalid H. Jawabri and Sandeep Sharma. Physiology, cerebral cortex functions. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), 2024.

[19] Bob Kemp, Aeilko H. Zwinderman, Bert Tuk, Hilbert A. C. Kamphuisen, and Josefien J. L. Oberýe. Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG. *IEEE Transactions on Biomedical Engineering*, 47(9):1185–1194, 2000.

[20] J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.

[21] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[22] Wanqiu Liu, Yupeng Zhang, Pengcheng Ma, Lei Zheng, Di Zhou, Zhengbo Chen, Mingming Shen, Yongchun Cai, Zhengyi Lu, Qiao Liu, Junwen Luo, and Xiaoan Wang. DEED: A dataset for dream-related emotion research. *bioRxiv*, 2022. Preprint.

[23] Yuval Nir and Giulio Tononi. Dreaming and the brain: From phenomenology to neurophysiology. *Trends in Cognitive Sciences*, 14(2):88–100, 2010.

[24] Markus Ojala and Gemma C. Garriga. Permutation tests for studying classifier performance. *Journal of Machine Learning Research*, 11:1833–1863, 2010.

[25] Mithila Packiyanathan, André Torvestad, Luis Alfredo Moctezuma, and Marta Molinas. Exploring high- and low-density electroencephalography for a dream decoding brain-computer interface. In *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2024)*, pages 565–570, Bruges, Belgium, 2024. i6doc.com.

[26] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, pages 8024–8035, 2019.

[27] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[28] Allan Rechtschaffen and Anthony Kales. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. U.S. Government Printing Office, Washington, DC, 1968.

[29] Naveed Rehman and Danilo P. Mandic. Multivariate empirical mode decomposition. *Proceedings of the Royal Society A*, 466(2117):1291–1302, 2010.

[30] Samantha J. Reznik and John J. B. Allen. Frontal asymmetry as a mediator and moderator of emotion: An updated review. *Psychophysiology*, 55(1):e12965, 2018.

[31] Serena Scarpelli, Chiara Bartolacci, Aurora D'Atri, Maurizio Gorgoni, and Luigi De Gennaro. The functional role of dreaming in emotional processes. *Frontiers in Psychology*, 10:459, 2019.

[32] Mike Schuster and Kuldip K. Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997.

[33] Francesca Siclari, Benjamin Baird, Lampros Perogamvros, Giulio Bernardi, Joshua J. LaRocque, Brady Riedner, Melanie Boly, Bradley R. Postle, and Giulio Tononi. The neural correlates of dreaming. *Nature Neuroscience*, 20(6):872–878, 2017.

[34] Pilleriin Sikka, Antti Revonsuo, Valdas Noreika, and Katja Valli. EEG frontal alpha asymmetry and dream affect: Alpha oscillations over the right frontal cortex during REM sleep and presleep wakefulness predict anger in REM sleep dreams. *Journal of Neuroscience*, 39(24):4775–4784, 2019.

[35] Akara Supratak, Hao Dong, Chao Wu, and Yike Guo. DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11):1998–2008, 2017.

[36] Michal Teplan. Fundamentals of EEG measurement. *Measurement Science Review*, 2(2):1–11, 2002.

[37] André Torvestad, Mithila Packiyanathan, Luis Alfredo Moctezuma, and Marta Molinas. Unveiling dreams: Moving towards automatic dream decoding via PSD-based EEG analysis and machine learning. In *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2024)*, pages 583–588, Bruges, Belgium, 2024. i6doc.com.

[38] Gaël Varoquaux, Pradeep Reddy Raamana, Denis A. Engemann, Andrés Hoyos-Idrobo, Yannick Schwartz, and Bertrand Thirion. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, 145:166–179, 2017.

[39] Matthew P. Walker and Els van der Helm. Overnight therapy? the role of sleep in emotional brain processing. *Psychological Bulletin*, 135(5):731–748, 2009.

[40] Peter D. Welch. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.

[41] William Wong, Rubén Herzog, Kátia Cristine Andrade, Thomas Andrillon, Draulio Barros de Araujo, Isabelle Arnulf, Somayeh Ataei, Giulia Avvenuti, Benjamin Baird, Michele Bellesi, Giulio Bernardi, Karen R. Konkoly, Valdas Noreika, Antti Revonsuo, Francesca Siclari, Pilleriin Sikka, Katja Valli, Erin J. Wamsley, Jennifer M. Windt, Jing Zhang, Naotsugu Tsuchiya, et al. A dream EEG and mentation database. *Nature Communications*, 16(1):7495, 2025.

[42] Jing Zhang and Erin J. Wamsley. EEG predictors of dreaming outside of REM sleep. *Psychophysiology*, 56(7):e13368, 2019.