



NTNU – Trondheim
Norwegian University of
Science and Technology

**Towards a Dream Decoder:
Classification and Analysis of Dreamful Experiences
During Sleep Using Electroencephalography Signals**

**Mithila Packiyanathan &
André Torvestad**

Department of Engineering Cybernetics
Norwegian University of Science and Technology

Submission date: June 4, 2024

Supervisor: Marta Molinas

Co-Supervisor: Luis Alfredo Moctezuma

Preface

This master's thesis is our final work for the Master of Technology at the Department of Engineering Cybernetic of the Norwegian University of Science and Technology (NTNU).

The project builds on the idea of Professor Marta Molinas, our supervisor throughout this work. Our part in the dream decoder project started with two individual specialization projects [1, 2], conducted during the fall of 2023. Here we initiated our work on dream and sleep research, providing us with an initial understanding of the problem at hand.

Based on findings from this thesis, two papers have been submitted to the 32th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2024). They are available in appendix B and appendix C.

Trondheim, June 4, 2024

Mithila Packiyathan & André Torvestad

Acknowledgement

Through this project we have delved into the field of biomedical cybernetics, where engineering and biology combine. We are grateful for the opportunity to be a part of the dream decoder project, and proud of having made a contribution to such a fascinating field.

We want to express our gratitude to our supervisor, Professor Marta Molinas. She has been a great resource throughout the project, providing guidance and suggestions of the highest quality. Her enthusiasm and excitement for this project has been highly contagious, and we are thankful to her for including us in the dream decoder project.

We also wish to thank our co-supervisor, Dr. Luis Alfredo Moctezuma, for his guidance during the last ten months. His technical insight and readiness to assist has been invaluable throughout the project, and we are deeply grateful for the opportunity to learn from such a skilled researcher.

Abstract

Dreams reflect our general well-being, with symptoms of mental health disorders often manifesting in our dream content. Dream analysis can be a powerful tool for identifying early symptoms and monitoring mental health disorders. Developing a tool capable of automatic retrieval of dream content from sleeping subjects - a dream decoder - would allow for the retrieval of many more dream experiences, allowing better understanding of a subject's mental well-being.

Brain activity can be measured using electroencephalography (EEG), a relatively inexpensive, non-invasive technique with high temporal resolution. With this technique, we can investigate recordings of brain activity during dream experiences and compare them to recordings of no experience, investigating what dreams are.

This thesis is an initial contribution towards the dream decoder. The first step is to identify when a subject is experiencing a dream. In this work, we experiment with deep learning (DL) and machine learning (ML) to investigate to what extent it is possible to identify dream experiences from EEG signals. Principal component analysis (PCA) and common spatial pattern (CSP) in combination, and power spectral density (PSD), are used for feature extraction. Multivariate empirical mode decomposition (MEMD) is applied to improve the classification. Furthermore, we attempt to define the most important regions for identifying dream experiences using permutation-based channel selection and PSD representation, from which channel subsets are identified for further classification.

Two datasets are used: The Zhang & Wamsley (ZW) and the Tononi dataset, both publicly available through the Dream EEG and Mentation (DREAM) database. They contain EEG data collected before awakening, paired with a verbal report of any dream experiences from which the recordings are labeled. Recordings are made during non-rapid eye movement (NREM) and rapid eye movement (REM) sleep.

The results indicate that dream experiences can be identified from EEG signals using ML. For the Tononi dataset using 256 EEG channels, we obtain up to 0.993 accuracy, 0.994 F-score, and 0.987 kappa score using MEMD, PSD, and the extreme gradient boosting (XGB) classifier. For the ZW data using 58 EEG channels, during NREM and REM sleep, we obtain 0.94 accuracy, 0.96 F-score and 0.84 kappa, and accuracy 0.998, F-score 0.998 and kappa 0.995, using MEMD, PCA, CSP, and the k-nearest neighbor (KNN) classifier. Firm conclusions on important brain regions are not readily available, but the results indicate areas in the parietal and frontal lobe.

Sammendrag

Vår mentale helse og symptomer på psykiske lidelser reflekteres ofte i drømmer. Drømmeanalyse kan være et kraftig verktøy for å identifisere tidlige symptomer og overvåke psykiske lidelser. Å utvikle et verktøy som automatisk kan hente drømmeinnhold fra personer som sover – en drømmedekoder – ville gjort det mulig å samle inn data fra drømmer, og derav gi en bedre forståelse for en persons mentale helse.

Hjerneaktivitet kan måles ved hjelp av elektroencefalografi (EEG). EEG måles ved å plassere metal-elektroder, også kalt kanaler, på hodebunnen for å måle de elektriske strømmene som forplanter seg gjennom hodeskallen fra hjernen. Det er en ikke-invasiv, relativt billig teknikk med høy temporal oppløsning som ofte brukes i nevrovitenskapelig forskning og for utvikling av hjerne/maskin-grensesnitt. Med denne teknikken kan vi undersøke hjerneaktivitet under drømmeopplevelser og sammenligne dem med hjerneaktivitet uten opplevelser, for å undersøke hva drømmer er.

Denne oppgaven er et første bidrag til utvikling av drømmedekoderen. For å få til dette er det første skrittet å identifisere når en person drømmer. I denne oppgaven utforsker vi med maskinlæring og metoder for utregning av karakteristikk, for å undersøke i hvilken grad det er mulig å identifisere drømmeopplevelser fra EEG-signaler. Prinsipal komponentanalyse (PCA) i kombinasjon med "common spatial pattern" (CSP), og spektraltetthet (PSD) er anvendt. "Multivariate empirical mode decomposition" (MEMD) er anvendt for å forbedre klassifiseringen. Videre prøver vi å kartlegge de viktigste hjernedelene for å identifisere drømmeopplevelser ved å bruke permutasjonsbasert kanalvalg og spektraltetthetsrepresentasjon, hvorfra vi velger ut et begrenset antall kanaler for videre klassifisering.

To datasett brukes: Zhang & Wamsley- og Tononi-datasettet, begge gjort offentlig tilgjengelige gjennom "Dream EEG and Mentation" (DREAM)-databasen. De inneholder EEG-data samlet inn før oppvåkning, kombinert med en muntlig rapport om eventuelle drømmer som hver måling er merket med. Målinger gjøres under både non-rapid eye movement (NREM) og rapid eye movement (REM) søvn.

Resultatene indikerer av drømmeopplevelser kan identifiseres fra EEG-signaler ved å bruke maskinlæring. For Tononi-datasettet får vi 0.993 nøyaktighet, 0.994 F-tall, og 0.987 kappa ved å bruke MEMD, PSD, og "extreme gradient boosting". For Zhang & Wamsley-datasettet, for henholdsvis NREM og REM søvn, får vi 0.94 nøyaktighet, 0.96 F-tall, og 0.84 kappa, og 0.998 nøyaktighet, 0.998 F-tall, og 0.995 kappa ved å bruke MEMD, PCA, CSP, og k -nærmeste naboer. Hvilke hjerneområder som er viktigst kan ikke fastslås helt, men resultatene indikerer områder i parietal- og frontallappen som betydningsfulle.

Contents

Preface	i
Acknowledgement	ii
Abstract	iii
Sammendrag	iv
List of Figures	viii
List of Tables	x
List of Abbreviations	xiii
1 Introduction	1
1.1 Background & Motivation	1
1.2 Problem Description	2
1.3 Objectives	2
1.4 Approach	2
1.5 Outline	3
2 Theoretical Background	4
2.1 Brain Anatomy	4
2.2 Human Sleep	5
2.3 Dreams	7
2.3.1 The Nature of Dreams	7
2.3.2 Models of the Dreaming Brain	7
2.4 EEG Signals	9
2.4.1 Artifacts	9
2.5 Multivariate Empirical Mode Decomposition	10
2.5.1 Noise-assisted Multivariate Empirical Mode Decomposition	11
2.6 Feature Extraction	12
2.6.1 Power Spectral Density and the Fast Fourier Transform	12
2.6.2 Principal Component Analysis	13
2.6.3 Common Spatial Pattern	13
2.7 Machine Learning	14
2.7.1 Machine Learning Algorithms	14
2.7.2 Deep Learning	16
2.8 Performance Evaluation of Classifications	17
2.9 Data Augmentation	20

2.10 Channel Selection	20
3 Literature Review	22
3.1 Prediction of Dream Experience	22
3.1.1 NREM and REM Sleep	22
3.1.2 NREM Sleep	23
3.1.3 REM Sleep	24
3.2 Brain Activity for Specific Dream Content	24
3.3 Frequency Activity in Sleep	25
4 Methodology	27
4.1 Materials	27
4.1.1 The Zhang & Wamsley (ZW) Dataset	27
4.1.2 The Tononi Dataset	28
4.1.3 Deep Learning Model	29
4.1.4 Development Tools	29
4.2 Data Filtering Using MEMD	30
4.2.1 The ZW Dataset	31
4.2.2 The Tononi Dataset	31
4.3 Feature Extraction and ML Algorithms	32
4.3.1 PCA, CSP, and ML	32
4.3.2 PSD-Based Classification	33
4.4 Deep Learning Classification	33
4.4.1 The ZW Dataset	34
4.4.2 The Tononi Dataset	34
4.5 Data Augmentation Using MEMD	34
4.5.1 ZW with Data Augmentation	35
4.6 PSD Analysis on MEMD-Filtered Data	36
4.6.1 The ZW Dataset	36
4.6.2 The Tononi Dataset	36
4.7 Channel Selection	36
4.7.1 The ZW Dataset	37
4.7.2 The Tononi Dataset	38
5 Results	39
5.1 Classification Using the ZW Dataset During NREM Sleep	39
5.1.1 Classification Using 58 Channels	39
5.1.2 Channel Importances and PSD Analysis	42
5.1.3 Classification Using Channel Subsets	43
5.2 Classification Using the ZW Dataset During REM Sleep	45
5.2.1 Classification Using 58 Channels	45
5.2.2 Channel Importances and PSD Analysis	47
5.2.3 Classification Using Channel Subsets	50
5.3 Classification Using the Tononi Dataset	51

5.3.1	Classification Using 256 Channels	51
5.3.2	Channel Importances and PSD Analysis	54
5.3.3	Classification Using Channel Subsets	56
6	Discussion	61
6.1	Classification Methods	61
6.2	The Effect of MEMD Filtering	61
6.2.1	Data Augmentation With MEMD	62
6.3	Classification, Channel Importances, and PSD Analysis with NREM Data	63
6.3.1	Classification Results	63
6.3.2	Permutation-Based Channel Selection	64
6.3.3	PSD Analysis	65
6.3.4	Combined Analysis	66
6.4	Classification, Channel Importances, and PSD Analysis with REM Data	67
6.4.1	Classification Performances	67
6.4.2	Permutation-Based Channel Selection	67
6.4.3	PSD-Based Analysis	67
6.4.4	Combined Analysis	69
6.5	NREM versus REM	69
6.5.1	Classification Performances	69
6.5.2	Permutation-Based Channel Selection	70
6.5.3	PSD-Based Analysis	70
6.6	Limitations	71
6.6.1	Low-Frequency Components in the ZW Dataset	71
6.6.2	Segmenting Before Train-Test Data Split	71
6.6.3	Segmenting and Channel Selection Before MEMD Filtering	72
6.6.4	Recall of DE, Not Necessarily DE	72
6.7	Future Work	73
6.7.1	Single-Subject Classification	73
6.7.2	Optimal Channel Selection Algorithm	73
6.7.3	Working Towards a Brain Computer Interface System	73
7	Conclusion	74
	Bibliography	75
A	Additional Classification Results	85
A.1	PSD and ML Classification: Including DEWR in Tononi Data	85
A.2	Channel Subset Classification with Tononi Data	86
B	ESANN2024 Conference Paper: <i>Unveiling Dreams: Moving Towards Automatic Dream Decoding via qualitative EEG Analysis and Machine Learning</i>	88
C	ESANN2024 Conference Paper: <i>Exploring High- and Low-Density Electroencephalography for a Dream Decoding Brain-Computer Interface</i>	95

List of Figures

2.1	Anatomy of the human brain: The four brain lobes.	5
2.2	The empirical mode decomposition algorithm explained step by step.	11
2.3	The architecture of an artificial neural network.	16
2.4	Common layers of a convolutional neural network.	17
2.5	The structure of a confusion matrix.	19
4.1	Distribution of the classes <i>No experience</i> and <i>Dream experience</i> , in the NREM and REM sub-datasets of Zhang and Wamsley.	28
4.2	Architecture of the EEGNeX Model.	30
4.3	Process of obtaining MEMD-filtered signal from raw signal.	31
5.1	Classification performance using Zhang and Wamsley NREM data for <i>Dream experience</i> vs <i>No experience</i> with and without data augmentation, using different machine learning algorithms and the EEGNeX model.	40
5.2	Classification performance using Zhang and Wamsley NREM data for <i>Dream experience</i> vs <i>No experience</i> with raw or MEMD-filtered data, using different feature extraction techniques in combination with machine learning algorithms.	41
5.3	Classification performance using Zhang and Wamsley NREM data for <i>Dream experience</i> vs <i>No experience</i> with raw or MEMD-filtered data, using EEGNeX.	41
5.4	Topographic map showing the channel importances for Zhang and Wamsley NREM data, using raw or MEMD-filtered data, extracted using permutation-based channel selection.	42
5.5	Topographic map showing average PSD ratios (<i>Dream experience</i> / <i>No experience</i>) for MEMD-filtered Zhang and Wamsley NREM data.	43
5.6	Classification performances using Zhang and Wamsley NREM data, for <i>Dream experience</i> vs <i>No experience</i> , with raw or MEMD-filtered data, using machine learning algorithms and the EEGNeX model. Permutation-based channel selection is used for choosing the channel subsets.	44
5.7	Classification performances using Zhang and Wamsley NREM data, for <i>Dream experience</i> vs <i>No experience</i> , with MEMD-filtered data, using PSD and XGB. The channels are chosen by the highest average PSD ratio.	45
5.8	Classification performance using Zhang and Wamsley REM data for <i>Dream experience</i> vs <i>No experience</i> with raw or MEMD-filtered data, using different feature extraction techniques in combination with machine learning algorithms.	46

5.9	Classification performance using Zhang and Wamsley NREM data for <i>Dream experience</i> vs <i>No experience</i> with raw or MEMD-filtered data, using the EEGNeX model.	46
5.10	Topographic map showing the channel importances using raw or MEMD-filtered data of Zhang and Wamsley REM data, extracted using permutation-based channel selection.	47
5.11	Topographic map showing average PSD ratios using MEMD-filtered data of Zhang and Wamsley REM data.	47
5.12	Topographic map showing average PSD ratios using MEMD-filtered data of Zhang and Wamsley REM data, after removing the first IMF before constructing the MEMD-filtered data.	49
5.13	Classification performances using Zhang and Wamsley REM data, for <i>Dream experience</i> vs <i>No experience</i> , with raw or MEMD-filtered data, using machine learning algorithms and the EEGNeX model. Permutation-based channel selection is used for choosing the channel subsets.	50
5.14	Classification performances using Zhang and Wamsley REM data, for <i>Dream experience</i> vs <i>No experience</i> , with MEMD-filtered data, using PSD and KNN. The channels are chosen by the highest average PSD ratio.	51
5.15	Classification performance using Tononi NREM data for <i>Dream experience</i> vs <i>No experience</i> with raw or MEMD-filtered data, using different feature extraction techniques in combination with machine learning algorithms.	52
5.16	Classification performance using Tononi NREM data for <i>Dream experience</i> vs <i>No experience</i> with MEMD-filtered data, using CSP in combination with machine learning algorithms.	53
5.17	Classification performance using Tononi NREM data for <i>Dream experience</i> vs <i>No experience</i> with raw data, using the EEGNeX model.	53
5.18	Classification performance using Tononi NREM data for <i>Dream experience</i> vs <i>No experience</i> with raw or MEMD-filtered data, using PSD in combination with machine learning algorithms.	54
5.19	Classification performance using Tononi NREM data when including <i>Dream experience without recall</i> , with raw data, using different feature extraction techniques in combination with machine learning algorithms and the EEGNeX model.	55
5.20	Topographic map showing the channel importances using raw or MEMD-filtered data of Tononi NREM data, extracted using permutation-based channel selection.	55
5.21	Topographic map showing average PSD ratios using MEMD-filtered 2-second-long segments of Tononi NREM data.	56
5.22	Topographic map showing average PSD ratios using MEMD-filtered 73-second-long segments of Tononi NREM data.	56
5.23	Classification performance of channel subsets of Tononi NREM data, for <i>Dream experience</i> vs <i>No experience</i> , on raw data. Channels selected by permutation-based channel selection	57
5.24	Classification performances using Tononi N2 data, for <i>Dream experience</i> vs <i>No experience</i> , with MEMD-filtered data, using PSD and XGB. The channels are chosen by the highest average PSD ratio.	58

5.25 Classification performance using MEMD-variations on 10 channels with NREM data from Tononi, using 2-second-long input segments.	59
5.26 Topographic maps showing average PSD of <i>Dream experience</i> of Zhang and Wamsley REM, using MEMD-filtered data.	60
5.27 Topographic maps showing average PSD of <i>No experience</i> of Zhang and Wamsley REM, using MEMD-filtered data.	60

List of Tables

2.1	Associated frequency bands of some brain waves.	10
2.2	Interpretation of kappa score ranges.	19
4.1	Sleep stage and class distribution for the Zhang and Wamsley dataset.	28
4.2	Epoch and batch sizes for the classification tasks within the Zhang and Wamsley dataset	34
5.1	Summary of IMFs in the <i>Dream experience</i> class for different frequency bands in Zhang and Wamsley NREM data.	43
5.2	Summary of IMFs in the <i>No experience</i> class for different frequency bands in Zhnag and Wamsley NREM data.	43
5.3	Summary of IMFs in the <i>Dream experience</i> class for different frequency bands in Zhang and Wamsley REM data.	49
5.4	Summary of IMFs in the <i>No experience</i> class for different frequency bands in Zhang and Wamsley REM data.	49
A.1	Performance metrics for PCA, CSP and classifiers KNN, XGB, RBF SVM, and RF on low-pass filtered data from Tononi Serial Awakenings. Class combination: DE vs NE and DEWR.	85
A.2	Performance metrics for PCA, CSP and classifiers KNN, XGB, RBF SVM, and RF on low-pass filtered data from Tononi Serial Awakenings. Class combination: DE and DEWR vs NE.	85
A.3	Performance metrics for PSD and classifiers KNN, XGB, SVM, and RF on low-pass filtered data from Tononi Serial Awakenings. Class combination: DE vs NE and DEWR. . .	86
A.4	Performance metrics for PSD and classifiers KNN, XGB, and RBF SVM on low-pass filtered data from Tononi Serial Awakenings. Class combination: DE and DEWR vs NE. . .	86
A.5	Performance metrics for PCA, CSP and classifiers KNN, XGB, RBF SVM, and RF, and EEGNeX on subset of low-pass filtered data from Tononi Serial Awakenings. Channel subset: 50 channels chosen by permutation-based channel selection.	86
A.6	Performance metrics for PCA, CSP and classifiers KNN, XGB, RBF SVM, and RF, and EEGNeX on subset of low-pass filtered data from Tononi Serial Awakenings. Channel subset: 30 channels chosen by permutation-based channel selection.	87
A.7	Performance metrics for PCA, CSP and classifiers KNN, XGB, RBF SVM, and RF, and EEGNeX on subset of low-pass filtered data from Tononi Serial Awakenings. Channel subset: 20 channels chosen by permutation-based channel selection.	87

A.8 Performance metrics for PCA, CSP and classifiers KNN, XGB, RBF SVM, and RF, and EEGNeX on subset of low-pass filtered data from Tononi Serial Awakenings. Channel subset: 10 channels chosen by permutation-based channel selection.	87
--	----

List of Abbreviations

EEG	electroencephalography
PGO	ponto-geniculo-occipital
PSG	polysomnography
BCI	brain computer interface
EOG	electrooculogram
EMG	electromyogram
fMRI	functional magnetic resonance imaging
REM	rapid eye movement
NREM	non-rapid eye movement
DE	dream experience
NE	no experience
DEWR	dream experience without recall
PTSD	post-traumatic stress disorder
ML	machine learning
DL	deep learning
CNN	convolutional neural network
DNN	deep neural network
CSP	common spatial patterns
PCA	principal components analysis
ADAM	adaptive moment estimation
XGB	extreme gradient boosting
GB	gradient boosting
KNN	k-nearest neighbors
RF	random forest
SVM	support vector machine
ANN	artificial neural network
AI	artificial intelligence
DWT	discrete wavelet transform
NSGA	non-dominated sorting genetic algorithm
MEMD	multivariate empirical mode decomposition
IMF	intrinsic mode function
FFT	fast Fourier transform
DFT	discrete Fourier transform

EMD empirical mode decomposition

PSD power spectral density

NA-MEMD noise-assisted multivariate empirical mode decomposition

ICA independent component analysis

PFI permutation feature importance

DEED dream emotion evaluation dataset

RBF radial basis function

ZW Zhang & Wamsley

Chapter 1

Introduction

1.1 Background & Motivation

The United Nations Sustainable Development Goal 3 states "Ensure healthy lives and promote well-being for all at all ages"¹. The right to mental health is a basic human right, but mental health disorders are still a worldwide problem. In 2019, 1 in 8 people worldwide lived with a mental disorder. Anxiety and depression, the two most common disorders, affected 301 million and 280 million lives, respectively. Other major disorders are bipolar disorder, post-traumatic stress disorder (PTSD), and schizophrenia [3].

Symptoms of mental health disorders have been shown to correlate with dream content [4]. Frequently experiencing nightmares is correlated with psychotic or mood disorders, higher suicidality scores and more frequent suicide attempts, and is a symptom of PTSD [5]. Anxiety, depression, and stress are also correlated with specific dream content, as shown in multiple studies [6–10]. Analysis of dream content could provide valuable insight into a subjects mental health and well-being, allowing for better treatment from healthcare professionals.

Dreaming is a state of consciousness where the brain constructs an imagined world in which we have subjective, conscious experiences similar to wakefulness. Dreams occur during sleep, divided into non-rapid eye movement (NREM) and rapid eye movement (REM) sleep. Dreams differ between these two states: NREM dreams are reported as conceptual and thought-like, whereas REM dreams are vivid, emotionally charged experiences [11–13]. Capturing dreams is a difficult task: Although we experience multiple throughout the night, come morning most are forgotten. At most, one dream is captured at awakening. To retrieve multiple dreams in one night the subject must be awakened throughout the night to provide verbal reports of their experiences, a procedure which has negative impact on sleep quality.

Developing a technological framework for decoding dream content on a sleeping subject - a dream decoder - would therefore be a valuable tool for capturing early symptoms of mental health conditions and to better understand the mental well-being of a subject. Advances within neuroscience have shown that recordings of brain activity, such as electroencephalography (EEG) or functional magnetic

¹United Nations. The 17 goals. Available at: <https://sdgs.un.org/goals> (Accessed: 27.05.2024)

resonance imaging (fMRI), reflect what is happening within our minds; Sleep state, emotions, and even the content of our minds can be identified using machine learning (ML) [14–16]. The development of a dream decoder could thus be within reach. Such a device, capable of retrieving dream content during sleep, would help psychologists and healthcare professionals diagnose and monitor subjects with mental health disorders such as depression, anxiety, and PTSD. It would contribute to the United Nations Sustainable Development Goal 3, by bettering the mental health of people.

1.2 Problem Description

Mental health conditions are a serious concern in modern society and tools for combating them are crucial. One such tool is dream analysis, as symptoms of mental disorders such as anxiety, depression, and PTSD have been shown to manifest in dream content. This requires retrieval of dream content, which today is only possible through awakening the subject throughout the night.

A dream decoder, a device capable of retrieving dream content through interpretation of brain activity, would allow for automatic collection of dream content. The process of dream analysis would speed up drastically, allowing for earlier treatment and better understanding of mental health conditions.

In this thesis we focus on the first steps for the conception of a dream decoder. First, we investigate whether dreams can be identified through EEG data by experimenting with ML techniques. Second, we make an effort into identifying which brain regions are most important for identifying dreams.

1.3 Objectives

The objective of the thesis is summarized in the following items:

1. Explore the possibility of identifying dreams based on sleep EEG data.
2. Identify important brain regions for dream experiences.

1.4 Approach

Two datasets have been used for this work: The Tononi dataset and the Zhang & Wamsley (ZW) dataset. They are made available through the Dream EEG and Mentation (DREAM) database, which contains datasets related to dream research [17]. Both contain EEG data of brain activity collected during sleep preceding awakening, paired with a verbal dream report. For the ZW dataset, the recordings are labelled as dream experience (DE) or no experience (NE). They are made using 58 EEG channels during both NREM and REM sleep. For the Tononi dataset, the recordings are labelled DE, NE, or dream experience without recall (DEWR). They are made using 256 EEG channels during NREM sleep.

We have used ML and deep learning (DL) to investigate the possibility of classifying between primarily DE- and NE-labelled recordings. Two feature extraction methods are tested: 1) principal components analysis (PCA) and common spatial patterns (CSP) in a single pipeline, and 2) power spectral density (PSD) calculation. Multivariate empirical mode decomposition (MEMD) filtering has been

applied in an effort to improve the classification performances. Further analysis of the differences between DE- and NE-labelled recordings, both spectral and spatial, is made with PSD analysis and permutation-based channel selection. Classification is performed again on channel subsets selected from the analysis.

1.5 Outline

Further on, the thesis is organized as follows:

Chapter 2 covers the theoretical background necessary for comprehending the work done in this thesis. It starts with an overview of brain anatomy, sleep, dreams, and EEG data, to provide the reader with an understanding of what dreams are and how brain activity can be understood as signals. Thereafter, techniques for signal processing and feature extraction is provided. Following this, the ML algorithms used in this thesis, DL, and performance evaluation is explained. At last, theory on data augmentation and channel selection is provided.

Chapter 3 is an overview of state-of-the-art studies regarding prediction of dream experiences, brain activity for specific dream content, and frequency activity in sleep.

Chapter 4 describes the datasets, the DL model, and the methods used in this work. Each method is described first in general terms, before specifying how it is applied for each dataset.

Chapter 5 contains the results obtained after applying each method. It is divided into three main categories: The ZW dataset during NREM sleep, the ZW dataset during REM sleep, and the Tononi dataset. For each category we present classification results using all available channels, frequency analysis, and classification results from channel subsets.

Chapter 6 provides a discussion of the results by themselves and related to theory and literature review, as well as possible limitations and future work.

Chapter 7 concludes the thesis.

Appendix A contains additional classification results, when only the top performing classifier for a method has been shown in the results section for comparison between methods.

Appendix B and appendix C contain papers submitted for the 32th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Both are based on findings from this thesis.

Chapter 2

Theoretical Background

2.1 Brain Anatomy

The human brain is a remarkable and complex organ, serving as the epicenter of the nervous system and the cornerstone of our thoughts, emotions, and behaviors. The brain consists of the cerebrum, the cerebellum, and the brain stem. Together with the spinal cord, it comprises the central nervous system, the system responsible for controlling our bodily functions and processing sensory information. The brain is also responsible for our cognitive abilities, such as thought and emotional experiences. These abilities occur in the cerebral cortex, the grey matter covering the cerebrum. The cerebral cortex is divided into the frontal, parietal, occipital, and temporal lobe, as shown in fig. 2.1, each of which involved in specific tasks [13, 18].

The frontal lobe is located in the front of the hemispheres and is involved in important functions. Among them are aspects related to personality, such as prospective memory, future planning, and decision making. Speech and language are also engaged in the frontal lobe. The motor cortex is located in the posterior of the frontal lobe and is tasked with movement control and movement planning [13, 18].

The parietal lobe, located posterior to the frontal lobe and superior to the temporal lobe, is tasked with sensory information processing and somatosensory association. Most sensory information is interpreted in this region. It creates the spatial understanding of the surrounding environment and how we are located to it [13, 18].

The temporal lobes, located in the temporal regions of the cortex, are engaged in sound processing and declarative memory. Sounds are linked to previously known meanings based on tones, in a process called phonological representation. Declarative memory is the conscious recollection of concepts and events, as well as objects. The temporal lobe is also engaged in determining the content of vision, based on signals from the visual cortex [13, 18].

The occipital lobe contains the visual cortex and is tasked with processing and interpreting visual stimuli. After processing, the information is sent to other regions of the brain for analysis and further interpretation. The occipital lobe is located in the back of the cortex [13, 18].

The left and right hemisphere have different functionalities. The left hemisphere controls the right

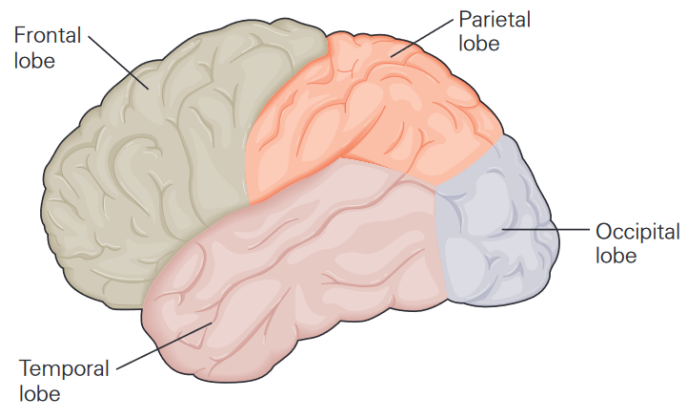


Figure 2.1: Anatomy of the human brain: The four brain lobes. Adapted from [13].

side of the body, speech, and research suggests that the left brain is associated with logical thinking [19, 20]. It is connected with objectivity, reasoning and for being analytical. The right hemisphere controls the left side of the body, does also contain some neurons needed for language, however, this part of the brain is necessary to appreciate semantic subtleties of language, as well as emotional content of speech. The right side is associated with imagination, intuition, creativity and emotions [13, 21].

2.2 Human Sleep

A person spends on average around 25 years of life asleep. Why we sleep and its functions have been debated for ages, but many questions are still open for debate. Sleep, which was previously considered a state of rest and inactivity, is now recognized as a highly organized and important state [13].

According to the hypothesis of synaptic homeostasis, sleep is necessary due to synaptic firings in the brain during wakefulness [22, 23]. These firings cause synaptic potentiation in several cortical circuits, as signals are transmitted between neurons. This causes saturation in the brain's ability to function. The role of sleep is to downscale synaptic strength so that homeostasis is restored. This is a crucial process, as increased synaptic strength has various costs, such as higher energy consumption, greater cellular stress, and changes in support cells. Downscaling, or renormalization of synaptic strength, reduces the burden of wakefulness.

Human sleep is divided into two main states: NREM and REM, recognized by their different characteristics. NREM sleep ranges from light to deep sleep, with gradually more synchronous brain activity and an increased threshold for arousal. REM sleep is recognized by the appearance of rapid eye movements, from which the name is derived. A sleeping subject alternates between these stages throughout the duration of sleep. Throughout a night, lasting approximately 7 hours, the cycle occurs 4-6 times, each cycle lasting approximately 90-110 minutes [13, 24, 25].

In the case of sleep scoring by a human expert, sleep stages are scored using polysomnography (PSG), where EEG, electrooculogram (EOG), and electromyogram (EMG) electrodes are used to monitor

physiological parameters of a subject. The scoring is made on 30-second long epochs, sequentially divided from the start of the trial. The quality of the collected data is dependent on the application of electrodes at the beginning of the trial, making it an important part of sleep studies. Sleep is scored in five categories, presented as follows [25, 26].

- Stage W (wakefulness): The subject is awake or in the early stages of drowsiness. An epoch is scored as W when more than 50% contains either alpha rhythms over the occipital, eye blinks, rapid eye movements associated with normal or high chin muscle tone, or reading eye movements. [26].
- Stage N1 (NREM 1): This is the lightest sleep stage, indicating the transition from wakefulness to sleep. There is a decrease in the high-frequency activity related to wakefulness, with the EEG showing low amplitude, mixed frequency activity. An epoch is scored as N1 when this activity replaces the alpha rhythm for more than 50% of the epoch [13, 25, 26].
- Stage N2 (NREM 2): The subject is fully asleep, muscle tone decreases, respiration becomes slower and more regular, and body temperature falls. It is characterized by the appearance of K-complexes and sleep spindles in the EEG. K-complexes are recognized as a sharp, negative wave followed by a slower, positive component. Sleep spindles are trains of sinusoidal waves with frequency 11-16 Hz and a duration above 0.5 seconds. An epoch is scored as N2 if one or more K-complexes or sleep spindles are present [13, 25, 26].
- Stage N3 (NREM 3): This state is considered restorative, marked by decreased heart rate and blood pressure, and stable breathing. An epoch is scored as N3 when more than 20% of the epoch consists of slow wave EEG activity, which are slow waves of frequency 0.5-2 Hz and peak-to-peak amplitude above 75 μ V. The slow waves occurring in NREM sleep are not uniformly distributed across the cortical surface, but rather asynchronously and locally distributed across brain regions [24–27].
- Stage R (REM): An epoch is defined as stage R when it contains the following phenomena: low-amplitude, mixed frequency EEG activity without K-complexes or sleep spindles, low chin EMG tone for the majority of the epoch and concurrent with REMs, and REMs at any position within the epoch. Sawtooth waves, serrated waves of 2-6 Hz over the central region, often precede burst of REM. Recent studies have also shown local low-frequency oscillations below 10 Hz appearing during REM sleep, and dreams in REM being associated with the local decreases of slow wave activity in posterior brain regions [25–27].

During REM sleep the brain activity of the subject increases. The activation is especially prominent in the motor cortex, although the outputs of most signals are blocked. The output to the eyes is not blocked, however, resulting in the characteristic REMs. The proportion of REM sleep decreases with age; a newborn spends more than 50% of sleep in REM sleep, whereas adults spend around 20% [13, 28].

2.3 Dreams

2.3.1 The Nature of Dreams

Dreams are a state of consciousness and refer to the presence of subjective experiences during sleep. They appear in all sleep stages, although with different characteristics, and often reflect everyday events. While dreaming the subject creates an imagined world similar to, and experienced like, the real one. Both contain similar perceptual modalities, especially visual and auditory, and have the same content. Dream experiences differ, from ideas and kinds of thought to highly emotional and perceptual experiences. There is a correlation between the waking and dreaming self with respect to mood, imaginativeness, and concerns, and dreams mostly reflect the events of everyday life [12, 13, 29].

Dream reports collected after NREM and REM sleep differ in their content and characteristics. Dreams during REM sleep are reported as complex, hallucinatory experiences as vivid as during wakefulness. These experiences are often motoric, associated with both emotional content and sensory modalities, and have a high level of bizarreness. They are also less connected to everyday life and often story-like. The frequency of dream recollection when awakened in REM sleep is reported to be approximately 80-90% [11–13, 29].

NREM sleep dreams are often characterized by thought-like experiences, especially early in the night. They contain less of the characteristics associated with REM sleep dreams and are instead reported as more realistic, often containing material from recent real-world events. REM sleep dream characteristics during NREM sleep dreams do however increase throughout the night, being more present in reports collected after late night NREM sleep. Around 5-10% of NREM dream reports are indistinguishable from REM dream reports in perceptual and emotional content. Dream recall is less frequent from NREM sleep, reported at approximately 50% [11–13, 29].

2.3.2 Models of the Dreaming Brain

The activation-synthesis hypothesis proposed by Hobson and McCarley suggests that random sensory signals originating from the brainstem, when transmitted to the forebrain, give rise to dreams [30]. According to this hypothesis, the forebrain automatically generates dreams by processing information from brainstem circuits and comparing it with stored memories. Furthermore, during the generation of dreams, the brainstem exhibits increased activity, particularly in the sensorimotor circuits associated with sensations and movement.

Okuma proposed the hypothesis of sensory image-free association stating that during the first REM burst, impulses originating in pontine reticular formation stimulate the entire brain, including the sensory and visual systems, as well as the hippocampus [31]. As a result, the subject accesses and consciously observes the sensory representations stored in memory. Unintentional or surprising associations are then formed with the initial sensory image. As multiple REM bursts occur, more uncorrelated sensory images are provoked, and the previous images are combined into a story: a dream. The dreamer will incorporate sensory images into the dream, regardless of how it fits, explaining bizarre dreams. Furthermore, both external and internal sensory stimuli may be integrated into the content of the dream.

To explain the connection between dreams and what happens in our brain activity, Hobson updated the activation-synthesis model to a three-dimensional state space model [32]. The model is referred to as AIM, and stands for activation, information flow, and mode of information processing. The dimensions are meant to capture the information processing capacity of the system, how much of the information which is processed that comes from the outside world, whether it is reflected in behavior, and lastly, in which way the information in the system is processed. The model helps to understand how our consciousness changes when dreaming in different stages of sleep and wake. Using the model, they show that the conscious states of waking, NREM and REM sleep differ in three clear and measurable ways. The three measurable ways are the three parameters forming the axes of the state space model, AIM; activation, information flow and mode of information processing. Using these three parameters, it is possible to distinguish the conscious states of wake, NREM and REM.

Nielsen presented a theory explaining why dreams can be recalled from NREM sleep, and not only from REM sleep [33]. Nielsen argues that dream recall from NREM sleep can be explained by "covert" REM sleep processes, i.e. processes known from REM sleep appear during NREM sleep. Although evidence has pointed out that cognitive activity, or mentation, appears in all sleep stages, Nielsen distinguishes dreams from mentation. He defines dreams to contain sensory stimulation, whereas mentation is thought-like. Mentation is the most common form of dream recall from NREM sleep, whereas dreams with sensory stimulation are more common from REM sleep. Nielsen further hypothesizes that the more apparent the dream recall from NREM sleep is, the more covert REM sleep processes are active. This suggests that the more vivid the dream recall from NREM sleep is, the more the observed cognitive activity will resemble that of REM sleep, and less like NREM sleep without dream recall. Nielsen points to several findings that could support his hypothesis, including the state of intermediate sleep (sleep transitioning from NREM to REM sleep), processes during sleep onset that resemble REM sleep, and that covert REM sleep occurrences are more likely during REM sleep deprivation.

Another theory states that dreaming and REM sleep are controlled by different brain mechanisms, presented by Solms [34]. Solms states that, while REM sleep is controlled by the cholinergic brain stem mechanisms, dreaming is controlled through a forebrain mechanism, presumably dopaminergic. The argument is that dreaming can be altered without affecting REM sleep and vice versa. Dreaming can be preserved with brain stem lesions that affect REM sleep and can be eliminated by forebrain lesions that do not affect REM sleep. The theory states that dreaming is a two-stage process; a cerebral activation, and if activation engages the dopaminergic circuits of the ventromesial forebrain, dreaming occurs. The relation between high occurrence of dream recall during REM sleep is explained with REM sleep being a state of enhanced cerebral activity.

A recent suggestion is to consider dreams in the context of off-line memory consolidation that occurs during sleep. This point of view was proposed by Stickgold and Wamsley in 2017 [11]. Dreaming is believed to involve the reactivation of memories and emotions from previous experiences. However, the theory emphasizes that dreams are not the same as replays of specific events from the day; rather, they consist of fragments of experiences combined with facts, memories, and images stored in the brain. Dreams are believed to be the result of developed mechanisms that facilitate sleep-dependent memory consolidation and integration.

2.4 EEG Signals

The brain consists of millions of neurons that play a crucial role in transmitting information between the human body and the brain itself. It also regulates both internal and external sensory responses, as well as motor functions [35].

Caton discovered the existence of electrical brain activity, known as EEG, in 1875 [36]. Later, in the 1920s, Hans Berger demonstrated that EEG could be recorded from the human scalp [37]. EEG is the noninvasive measurement of the brain's electric fields, achieved by placing electrodes on the scalp to record voltage potentials generated by current flow into and around neurons [38].

The main objective of EEG is the evaluation of dynamic cerebral functioning, which means real-time assessment and monitoring of brain activity and functioning, and can be used to analyze brain functions in health and disease [36]. The primary source of these signals is believed to be generated by cortical pyramidal neurons located in the cerebral cortex, which are oriented perpendicularly to the brain's surface. The detectable neural activity that gives rise to EEG is the collective sum of inhibitory or excitatory postsynaptic potentials originating from thousands of pyramidal cells near each electrode [39].

Therefore, EEG signals can be understood as the collective electrical activities of groups of neurons. Since neurons are excitable cells with intrinsic electrical properties, their activity generates electrical and magnetic fields. These fields can be detected from various distances from the sources; with electrodes at a short distance, from the cortical surface, or at a longer stance (for example, from the scalp) [40].

The sinusoidal rhythmic activity observed on the EEG is believed to represent oscillatory communication between the cortex and deeper subcortical structures. These communication loops manifest when the cortex is at rest or not engaged in specific tasks. However, when the cortex is tasked with performing a function, its electrical activity becomes desynchronized and faster electrical rhythms with lower amplitude become dominant until the cortex completes the task and returns to resting state [39].

Observing EEG recordings reveals that brain activity manifests as brain waves with sinusoidal patterns. These waveforms can be categorized on the basis of various properties, including frequency, amplitude, shape, and the placement of electrodes on the scalp. Generally, the waves are measured from peak to peak, with amplitudes ranging between 0.5 to 100 μV . Among these waves, certain frequency bands have emerged as fundamental components of brain activity. These bands are shown in table 2.1 [35, 41].

2.4.1 Artifacts

Artifacts are unwanted components that appear from other sources. They incorporate into the recorded EEG data, misleading the actual cerebral activity present. There are two main divisions of artifacts; physiological artifacts emerging from internal sources and non-physiological artifacts emerging from external sources. Eye movements, muscle activity and cardiac activity are often causes of physiological artifacts that could be present in EEG data. While, non-physiological artifacts are often caused by environmental interference and electrode issues, such as electrode movement and poor electrode

Brain wave	Frequency band
Slow wave	0.5-2 Hz
Delta	0-4 Hz
Theta	4-8 Hz
Alpha	8-13 Hz
Beta	13-25 Hz
Gamma	> 25 Hz

Table 2.1: Associated frequency bands of some brain waves [26, 42].

contact [43, 44].

In practice, biological and environmental electrical artifacts often disrupt EEG signals, making it challenging to accurately identify normal rhythms and pathological patterns. Fortunately, skilled and attentive observers can recognize distinguishing characteristics in these artifacts. Moreover, the approaches for using EEG signals involves directly using the raw signals from the scalp, or preprocess the collected data. By performing preprocessing, it becomes possible to eliminate artifacts and discard unreliable EEG channels. This approach also enables the extraction of brain activity from the noisy EEG signals [39, 45].

2.5 Multivariate Empirical Mode Decomposition

The MEMD algorithm is an extension of the empirical mode decomposition (EMD) algorithm. Due to the non-linear and non-stationary nature of EEG signals, conventional analytical tools like fast Fourier transform (FFT) are not typically used with EEG signals. However, EMD, developed by Huang et al. [46], being a data-driven algorithm, proves to be more suitable for handling nonstationary signals that undergo changes in frequency structure over short time periods. The application of EMD offers the advantage of capturing similar underlying features as the real EEG signals [47].

The EMD algorithm decomposes the original signal into a finite set of amplitude or frequency modulated components, named intrinsic mode function (IMF)s. Each IMF represents a nonlinear oscillation of the real signal in the temporal domain, thereby preserving features across both temporal and frequency domains. IMFs are defined according to two conditions; 1) the number of extrema and zero crossings must be equal to each other or differ at most by one, 2) the mean value of the envelopes defined by the local maxima and the local minima is zero [47, 48].

The IMFs are computed through an iterative process called the sifting algorithm. After extracting the first IMF, the residual $r(k) = x(k) - d(k)$ is used for extracting the remaining IMFs until the remaining residual is a monotonic function or a function only containing one extremum. The algorithm is depicted in fig. 2.2. Using the EMD, the original signal can be represented as a set of N IMFs, $\{c_i(k)\}_{i=1}^N$ and a residue signal, $r(k)$, as shown in eq. (2.1) [48, 49].

$$x(k) = \sum_{i=1}^N c_i(k) + r(k) \quad (2.1)$$

As the EMD algorithm sequentially removes the most oscillatory components of the signal, until a

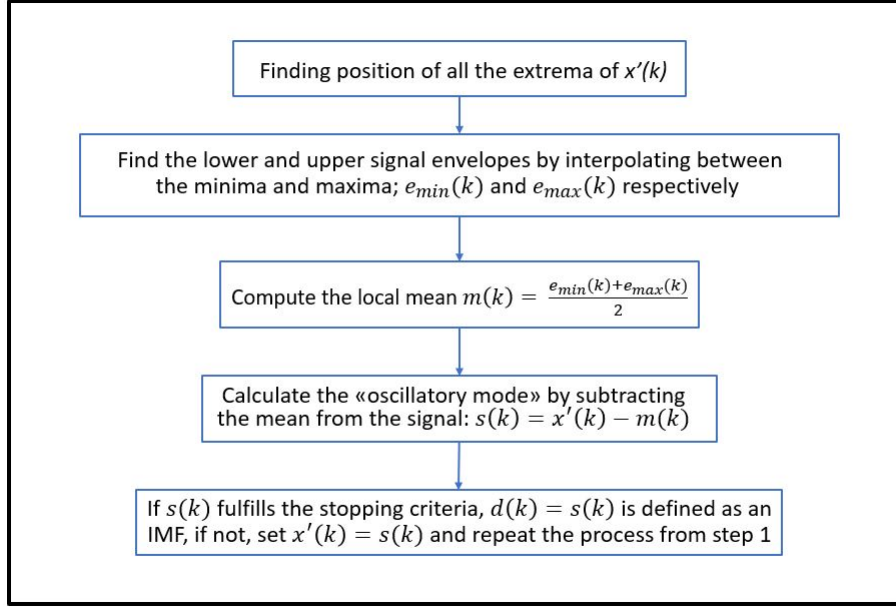


Figure 2.2: The EMD algorithm explained step by step [48].

residual signal is left, the algorithm often extracts IMFs going from high to low frequencies. That is, the first IMF extracted usually contains the highest frequency components. The EMD algorithm is processed channel-wise, using only one channel at a time. The properties of EMD has made it appropriate in the analysis of non-linear and non-stationary signals [48].

In contrast to EMD, the MEMD ensures a consistent number of IMFs across the channels in the signals. Since EMD is a data-driven algorithm, processing signals channel-wise, the number of generated IMFs may vary depending on the characteristics of the signals. A varying number of IMFs leads to desynchronization of the weights associated with each IMF. The desynchronization makes it challenging for the system to properly learn and distinguish between different classes or categories, and will therefore affect the performance of the classification system negatively. The consistent number of IMFs generated by the MEMD are therefore contributing to improving the classification performance and preventing weight desynchronization [50].

The MEMD algorithm operates by identifying common scales in all EEG channels and extracts the most relevant signals while also adaptively eliminating noise. Unlike the EMD algorithm, which computes the local mean of the original signal using the average of the upper and lower envelopes, MEMD uses multiple n-dimensional generated envelopes. The envelopes are generated by first taking signal projections along different directions in n-dimensional spaces. Extremes are then computed from these projections and interpolated component-wise to obtain the desired multidimensional envelopes of the signal. In addition, the local mean is computed using the average of these generated envelopes. This approach allows for a precise estimation of local means, facilitating the extraction of IMFs from multivariate signals [48].

2.5.1 Noise-assisted Multivariate Empirical Mode Decomposition

The noise-assisted multivariate empirical mode decomposition (NA-MEMD) algorithm is an extension of the MEMD algorithm, forming a multivariate signal, which consists of input data and noise in

separate channels. The MEMD algorithm is then applied to the data containing the noise channels. The resulting IMFs corresponding to the noise are subsequently discarded, leaving only the IMFs that represent the components of the original signal [51].

The NA-MEMD is mainly introduced due to dealing with some of the limitations of EMD. The EMD needs a sufficient number of local maxima and minima to work effectively. Moreover, the mode-mixing problem of EMD where the algorithm mix different signal components into a single IMF, and the fact that it function for univariate data, meaning that it only works for one channel at a time, are all challenges of the EMD algorithm. Adding noise in a controlled manner, instead of adding it directly to the data has given the advantage of being able to separate components of the signal more effectively, with more accurate frequency estimates, while reducing the mode-mixing problem. The application of NA-MEMD is therefore beneficial if prone to the mode-mixing problem, and it has also proven to be working well in brain computer interface (BCI) systems [51].

2.6 Feature Extraction

Feature extraction is used as a technique to reduce dimensionality and capture the most relevant information from the data [52]. The extracted features give varying performances when input to a classifier. Ineffective features selected for model training will result in lower model performance, while selecting the right features that accurately represent the underlying EEG signal will improve model performance [53].

Feature extraction can be performed manually or automatically. Manual feature extraction techniques are often time-consuming and are generally used in combination with ML classifiers such as k-nearest neighbors (KNN) and support vector machine (SVM). Automatic feature extraction is often present in DL methods [54].

Some hand-crafted feature extraction techniques are PSD, PCA and CSP.

2.6.1 Power Spectral Density and the Fast Fourier Transform

PSD, or power spectrum, is a common method for EEG analysis and feature extraction [55–57]. PSD reflects the distribution of the signal power over frequency. It is calculated from the discrete Fourier transform (DFT) as

$$PSD(k) = \frac{1}{N} |\hat{X}(k)|^2, \quad (2.2)$$

where N is the number of samples in the signal and $\hat{X}$ the DFT of the signal.

The DFT is a transformation of the signal from the time domain to the frequency domain, used for power spectrum analysis [58]. The DFT of a signal is widely computed using the FFT, an algorithm that reduces the computational power. The DFT of a signal x is defined as

$$\hat{X}(k) = \sum_{n=0}^{N-1} x(n) e^{-2\pi j n / N}. \quad (2.3)$$

2.6.2 Principal Component Analysis

PCA is a technique used to transform data into principal components. These components are a linear combination of the initial data, and reducing the number of them is often the main goal of PCA. Principal components are calculated from the initial data $\mathbf{x}$ by

$$\mathbf{x}_{pca} = A^T \mathbf{x}, \quad (2.4)$$

where A is a matrix of orthonormal eigenvectors of S , calculated by

$$|S - \lambda I| = 0, \quad (2.5)$$

where I is the identity matrix and λ are eigenvalues.

To reduce the dimensionality of the data, the number of principal components is chosen based on the amount of information loss accepted for a given problem. Each principal component carries a certain variance. Dimension reduction is done by selecting a number of principal components that accumulate to a certain percentage of the total variance [59, 60].

2.6.3 Common Spatial Pattern

CSP is a technique for transforming multichannel data into a low-dimensional spatial subspace [61, 62]. The technique is data-driven and supervised. Spatial filters which maximize the variance for one class and minimizes it for the other are calculated, making it easier to classify between them. The initial multichannel data is transformed into the spatial subspace by:

$$\mathbf{X}_{CSP} = W^T \mathbf{X}, \quad (2.6)$$

where W is calculated by

$$\begin{aligned} W \Sigma_1 W^T &= D, \\ W \Sigma_2 W^T &= I - D \end{aligned} \quad (2.7)$$

where Σ_i is the covariance matrix for class i , D is a diagonal matrix, and I is the identity matrix. This problem is solved by finding a matrix P such that

$$P(\Sigma_1 + \Sigma_2)P^T = I. \quad (2.8)$$

This process is called *whitening* the matrix $\Sigma_1 + \Sigma_2$. This is used to define $S_l = P \Sigma_l P^T, l \in \{1, 2\}$, whereby an orthogonal matrix R and D is calculated such that

$$\begin{aligned} S_1^T &= RDR^T, \\ S_2^T &= R(I - D)R^T, \end{aligned} \tag{2.9}$$

where S_2 follows from the definition $S_1 + S_2 = I$. The spatial filter matrix is then calculated as

$$W = R^T P. \tag{2.10}$$

2.7 Machine Learning

ML emerged as a field of study from the traditional statistics and the artificial intelligence (AI) communities. ML is a form of applied statistics, where the parameters of an algorithm are optimized for a given task on a given set of data. Unlike AI tasks that aims at replicating human thought processes for solving complex problems, the ML tasks learn from data to perform these complex tasks. The algorithms convert empirical data into practical models, and these models are not explicitly programmed, they rather undergo training to learn [63–66].

Multiple tasks can be performed using ML, with a common one being classification. In a classification task, the algorithm is tasked with deciding which of k categories a given data sample belongs to. The algorithm produces a function $f : \mathbb{R}^n \rightarrow \{1, \dots, k\}$ for k categories. The function is found by training the algorithm, where it iterates over a portion of the dataset over multiple epochs. Classification is a form of supervised learning, which is when the algorithm learns to associate a given output with a given input. Classification between two categories is referred to as binary classification, while classification of multiple categories is referred to as multi-class classification [66].

ML algorithms have parameters which affect the performance of the algorithm. The parameters available depend on the model architecture. They can be understood as a set of weights updated when training an algorithm. Hyperparameters are parameters not adapted by the algorithm itself, but must instead be defined by the developer. There are many techniques for choosing the parameters, such as using previous experience or values used in similar experiments [66].

2.7.1 Machine Learning Algorithms

There are a multitude of various ML algorithms, each with its own architecture for learning. Different algorithms have different strengths and weaknesses depending on the task at hand. This section provides an introduction to the algorithms used in this thesis.

K-Nearest Neighbors

KNN is an ML technique used for classification and regression tasks. It operates as a nonparametric learning algorithm, thus avoiding constraints imposed by a fixed number of parameters, often considered to have none. Instead, the algorithm uses a straightforward function of the training data.

Unlike some other algorithms, it does not involve a defined training stage. Rather, during the test, it identifies the KNN of the test sample x within the training data X . Using this information, it generates

the output y corresponding to the test input x , achieved by averaging the corresponding y values in the training set [66].

Extreme Gradient Boosting

Extreme gradient boosting (XGB) is an ML algorithm that is based on the foundations of decision trees. The algorithm constructs short and simple decision trees iteratively, and because of its high bias, each tree is often referred to as a "weak learner". When constructing the subsequent tree, the objective is to train it to predict what the previous weak learner couldn't achieve. In this manner, each weak learner aims to correct the errors of its predecessor. This iterative process continues until the predefined stopping criteria are met. XGB therefore provides fast training and the ability to be parallelized [67].

Support Vector Machine

SVM is an ML algorithm similar to logistic regression, however, this technique does not provide probabilities. It is driven by the linear function

$$w^T x + b \quad (2.11)$$

and yields a class identity as its output. The output is predicted as positive if the linear function 2.11 produces a positive value and negative if it gives a negative value.

SVM can be represented in terms of dot products between examples. For instance, using SVM, the expression 2.11 can be reformulated as depicted in eq. (2.12). Here, $x^{(i)}$ denotes a training sample, and α represents a vector of coefficients. This formulation allows the substitution of x with the output of a specified feature function $\phi(x)$. Subsequently, this function can be employed to give the dot product in eq. (2.13).

$$w^T x + b = b + \sum_{i=1}^m \alpha_i x^T x^{(i)} \quad (2.12)$$

$$k(x, x^{(i)}) = \phi(x) \cdot \phi(x^{(i)}) \quad (2.13)$$

The outcome of eq. (2.13) is referred to as a kernel. The Gaussian kernel, expressed in eq. (2.14)

$$k(u, v) = \mathcal{N}(u - v; 0, \sigma^2 I) \quad (2.14)$$

is one of the most commonly used kernel, and is also known as the radial basis function (RBF) kernel. Conceptually, the RBF kernel can be compared to a template matching process. If a training example x is linked with a training label y , it effectively serves as a template for class y . The RBF kernel has a large response when a test point x' is close to x , in terms of Euclidean distance. This indicates that x' is very similar to the x template. Following this, the model places a large weight on the associated training label y . In this way, the prediction will combine these training labels, weighed by the similarity of the associated training examples [66].

Random Forest

Random forest (RF) is an ensemble method that operates by training numerous decision trees simultaneously through bootstrapping, followed by aggregation. The decision trees are trained independently in parallel, each on different subsets of the training data, utilizing various subsets of available features. The uniqueness of each decision tree is ensured by bootstrapping, thereby reducing the overall variance of the classifier. Moreover, when combining the decisions of the trees to make the final decision, the classifier consequently demonstrates good generalization capabilities [68].

2.7.2 Deep Learning

DL represents learning methods from data and is a subset of ML, where the algorithm is a deep neural network (DNN). A DNN is inspired of biological computational networks, called artificial neural network (ANN)s. ANNs consist of an input layer, one or several hidden layers and an output layer, as shown in fig. 2.3 [66, 69].

The term "deep", in DL originates from the learning process being done through multi-layer neural networks [70]. A DNN is composed of two or more hidden layers and multiple mathematical functions, connected in a chain. The function f is approximated by

$$f(x) = f^n(f^{(n-1)}(\dots(f^{(1)}(x; \theta))))), \quad (2.15)$$

for an n-layered network. Each function layer is composed of unit functions, called neurons, working in parallel. The algorithm learns the value of the parameters θ which result in the best function approximation. The learning is performed, updating the mathematical relationships, using an algorithm called backpropagation. The backpropagation algorithm calculates the derivatives inside the DNN. The learned relationships during training can further on be used to make predictions of the data input to the DNN [66].

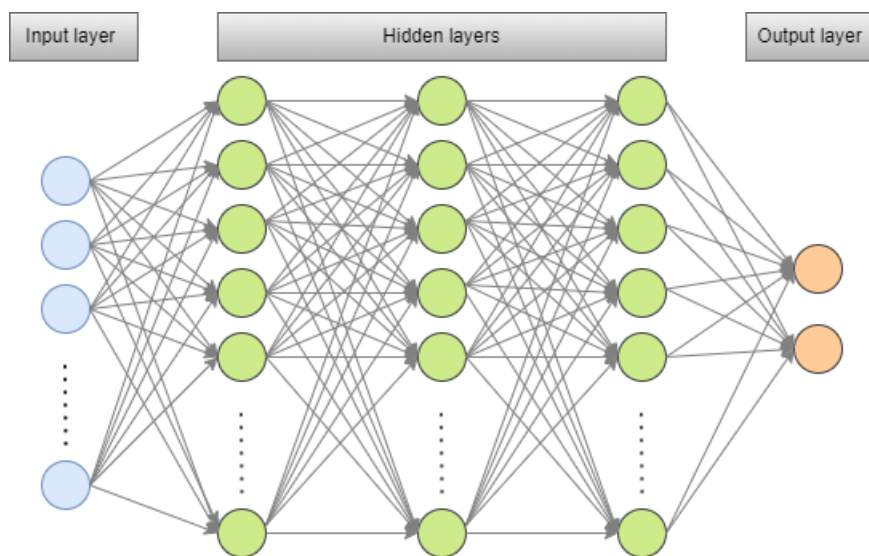


Figure 2.3: The architecture of an ANN.

Convolutional Neural Network

A convolutional neural network (CNN) is a specialized neural network used for processing data with a grid-like topology. CNNs have been successful in practical applications such as computer vision or time-series analysis. The common layers of a CNN is depicted in fig. 2.4, with a $N \times M$ matrix as input and 3 classes as output.

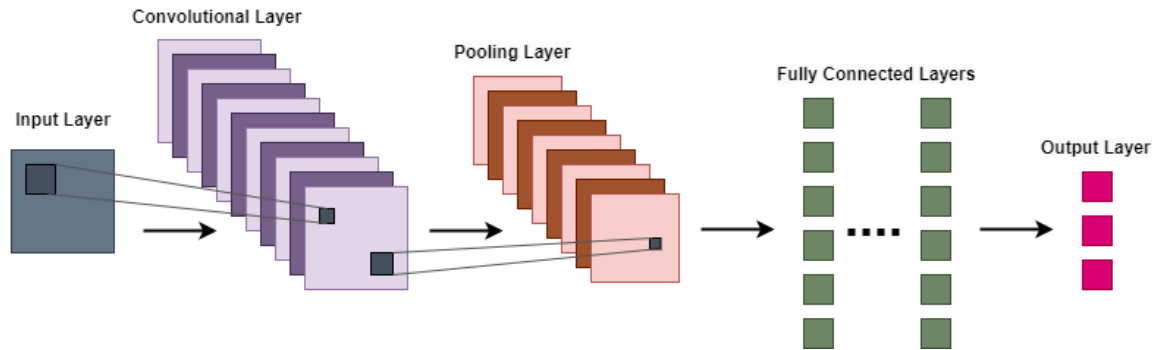


Figure 2.4: The architecture of a CNN.

These networks uses convolution, a type of linear mathematical operation. Data is given as an input to the convolutional layers, which return a feature map as an output. The feature map is obtained doing a convolution product between a matrix representation of a portion of the data and another matrix with a set of learnable parameters, referred to as a kernel. The main motivation of using convolution is sparse interactions, parameter sharing and equivariant representation. Sparse interactions typically refers to the process of making the kernel smaller than the input, so that fewer parameters are stored, reducing the memory requirements. Parameter sharing is the use of the same parameter for more than one function in the model. Equivariant representation refers to the property of the output changing in the same way as the input changes [66].

Another operation frequently used in convolutional networks is pooling. The main function of pooling is to provide translation invariance by reducing the original size of the data and the feature maps. The pooling function reduces the size by replacing the output of the net at a certain location with summary statistics of the nearby outputs [66, 71].

In the final layer of the CNN, a classification layer takes the feature map as input and returns the predicted class. The classification layer is referred to as fully connected, since all inputs of the layer influence all the outputs of the layer [66].

2.8 Performance Evaluation of Classifications

The goal is to create the best possible real-world applicable algorithm, but for most applications, it is impossible to achieve zero error. Algorithms are often evaluated in benchmarks, where different algorithms are measured against each other on the same problem. They can also be evaluated on the basis of a required threshold based on safety, cost-efficiency, or consumer-appeal. Evaluation is made with performance metrics. Choosing the right metric for a specific problem is vital for understanding the performance of an algorithm [66].

To ensure that an algorithm works in the real world, a portion of the available data is separated for test purposes. This data is called the test set, and the other portion is the training set. The separation of the training and test set ensures generalization and the ability to perform well on new data. During training, the algorithm attempts to reduce the training error by updating its parameters to fit the labels of the training set. The algorithm is evaluated by predicting the labels of the data samples in the test set and calculating the test error [66].

Two challenges arise during training algorithms: underfitting and overfitting. Underfitting occurs when the algorithm is unable to sufficiently reduce the training error. Overfitting occurs when the model performance is much higher for training data than for test data. These issues are combated by altering the capacity of the algorithm. When underfitting occurs, the capacity of the algorithm is not high enough - the model is too simple. Overfitting occurs when the capacity is too high. The "no free lunch" theorem states that no ML algorithm is universally better than any other. An algorithm performs best when the capacity is appropriate for the complexity of the task. Finding the appropriate capacity is an important task when developing ML algorithms [66].

For DL algorithms, this can be monitored during training by separating a part of the training set in a validation set. After each training epoch, the performance is evaluated using the validation set. This is used to guide the selection of hyperparameters and for introducing regularization techniques, such as early stopping criteria and the reduce learning rate algorithm. Early stopping monitors the performance of the validation data and stops training when no further performance increase is made after a predefined number of epochs. Reduce learning rate monitors in the same way, but reduces the learning rate when the performance increase is insufficient [66].

A possible problem when dividing the available data into train and test sets is that the test set is too small, resulting in a greater statistical uncertainty around the estimated performance. A lucky division could result in an unrealistically high performance estimation, and vice versa. Cross-validation is introduced to combat this; With k -fold cross-validation training and testing is performed over k divisions. The test error is then estimated as the average test error over k folds, resulting in a lower statistical uncertainty. All available data is divided into k random partitions; during trial i , partition i is used as a test set and the rest as a train set [66].

The most common performance metrics are calculated using a confusion matrix, which contains four counts: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), as shown in fig. 2.5 [72].

Accuracy, precision, and recall are calculated with these counts. Accuracy is the proportion of correctly predicted samples over total samples (see (2.16)). Precision is the proportion of correctly predicted positive samples over all predicted positive samples (see (2.17)). Recall is the proportion of correctly predicted positive samples over all true positive samples (see (2.18)) [66, 72].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.16)$$

$$Precision = \frac{TP}{TP + FP} \quad (2.17)$$

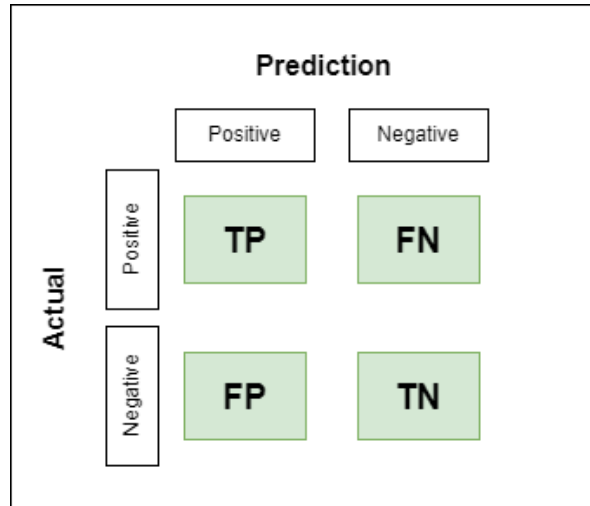


Figure 2.5: The structure of a confusion matrix.

$$Recall = \frac{TP}{TP + FN} \quad (2.18)$$

The F-score evaluates the relationship between precision and recall. These two metrics are often sacrificed for each other, depending on whether the importance is to be precise or to identify all positive samples. The F-score summarizes the precision and recall performance in a single metric, as given in (2.19).

$$F - score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (2.19)$$

The area under receiving operating characteristic curve (AUROC) is a metric typically used for evaluating binary classification, but can also be used to evaluate multiclass classification. The curve describes the relationship between false positive rate (FPR) and true positive rate (TPR). The metric ranges between 0 and 1, where 1 indicates a perfect performance and 0.5 a random performance. AUROC can be misleading for highly imbalanced datasets and should then be evaluated together with other metrics [72].

Cohen's kappa, known as the kappa score, measures interrater reliability. The metric measures the agreement between the true and the predicted labels, while taking into account the expected performance of the algorithm. The kappa score ranges from -1 to 1, where 1 indicates a perfect performance and 0 a random performance. The interpretation of the range of kappa score is depicted in table 2.2 [73].

Kappa Score Range	Interpretation
0.00 - 0.20	Poor Agreement
0.21 - 0.40	Slight Agreement
0.41 - 0.60	Fair Agreement
0.61 - 0.80	Substantial Agreement
0.81 - 1.00	Almost Perfect Agreement

Table 2.2: Interpretation of kappa score ranges [73].

2.9 Data Augmentation

Although DNNs have become a popular choice for signal classification, it comes with the challenge of collecting large datasets. Annotation processes, such as when collecting dream reports, are time-consuming and require human experts for sleep staging in real time.

Performing classification without enough representative data, meaning for instance an unbalance in number of samples for each class, may lead to overfitting or weak generalizability of the classifier. Thus, data augmentation emerges as a promising technique to enhance the performance of decoding models and the most effective way of solving the limitations of obtaining sufficient training data [74, 75].

Data augmentation is a type of regularization technique that aims to improve generalization. The primary objective of regularization methods is to incorporate information into the underlying ML model, enabling it to better encapsulate the general characteristics of the data. In line with this, data augmentation aims at increasing the robustness and accuracy of ML models, thereby giving models with well performance on small and poorly representative data. Data augmentation mitigates the limitations of unbalanced datasets by using appropriate data augmentation techniques, where existing training data are used to generate more training samples [75].

2.10 Channel Selection

EEG data is collected from electrodes, called channels, positioned across the scalp to capture electrical activity. Having many channels may effect the classification performance by decreasing the performance due to irrelevant or redundant information. To address this issue, channel selection algorithms play a crucial role in identifying optimal EEG channels for a specific application [76, 77].

The application of automatic channel selection has three main purposes. First, it reduces computational complexity by choosing the relevant channels and, thereby, extracting the most important features. Second, it reduces overfitting by decreasing the amount of memorization of noisy and unnecessary data from the channels, thereby improving performance. Third, it is applied to reduce the setup time for some applications, such as BCIs [76, 77].

Permutation-based channel selection is a technique where channels are selected for subsets by their individual importance. It is a computationally cheap method which identifies the most informative channels. The algorithm for calculating each channel importance is called permutation feature importance (PFI). The steps of the algorithm are as follows [78, 79]:

1. Create a baseline model using training data.
2. Calculate the baseline performance by evaluating the model on the test data.
3. For each feature i , randomly shuffle the data within (permute), for all samples, to corrupt the feature.
4. Use the baseline model to evaluate the performance on the test data with the corrupted feature. The importance of feature i is the difference between baseline performance and performance on the corrupted feature.

Channel selection can also be done using a multi-objective evolutionary algorithm. Evolutionary algorithms are population-based stochastic search algorithms which mimic natural evolution. A set of possible solutions are formed into a population, where they are evaluated and ranked depending on algorithm. The best solutions go through crossovers and mutates to create a new generation of populations, which are again evaluated and ranked. This process iterates until a termination criteria is fulfilled [80].

Non-dominated sorting genetic algorithm (NSGA)-II is an optimization evaluation algorithm which determines a set of optimal solutions for a multi-objective problem on a single simulation run. It uses non-dominated ranking to evaluate and rank the solutions. A solution is considered non-dominated if there is no other solution with a better evaluation value [81].

Chapter 3

Literature Review

3.1 Prediction of Dream Experience

3.1.1 NREM and REM Sleep

Siclari et al. investigated the neural correlates of dreaming during both NREM and REM sleep [82]. Using the serial awakening paradigm [83], participants were awakened at 15-30 minute intervals and asked if they, right before awakening, had been experiencing anything, where experience was defined as "any kind of mental activity". Three categories emerged from the reported cases: dream experience (DE), dream experience without recall (DEWR), and no dream experience (NE). Data were captured using 256 high-density EEG channels. The researchers calculated the power spectral density at source level, in both a low frequency (1 - 4Hz) and a high frequency (20 - 50Hz) band to investigate differences between the reports. For NREM sleep, differences were found in both frequency bands. Both DE and DEWR exhibited a decrease in low-frequency power in the parietal-occipital region, compared to NE. The same parietal-occipital region also showed an increase in high-frequency power, when comparing DE to NE. No significant difference was found between DE and DEWR. For DE versus DEWR, an increase in high-frequency power was found in the medial and lateral frontoparietal regions. No differences were found between DEWR and NE. For REM sleep, differences were also found in both frequency bands. DE showed both a decrease in low-frequency power in the parieto-occipital region and an increase in high-frequency power in frontal and lateral regions, compared to NE. The findings thus highlight a parieto-occipital region as a potential neural correlate for dreaming in both NREM and REM sleep. For REM sleep, dreaming could also be correlated to the frontal and lateral regions.

The study conducted by Zhang and Wamsley investigated the EEG correlates of dreaming [84]. They used high-density EEG recordings, with 58 electrodes on each of the 40 participants. The subjects were awakened at different stages of sleep and asked to provide a dream report. The awakenings were done during sleep onset, REM sleep, and N2 sleep. The correlates of dreaming found calculating the mean power spectral density in five a priori frequency bands; slow oscillations (frequencies below 1 Hz that remained after 0.1 Hz high pass filtering), delta (1-4Hz), theta (4-7Hz), alpha (8-12Hz) and beta (13-35Hz). The results unveiled distinct patterns in the EEG activity. Dream imagery during the transition to sleep was associated with increased absolute power in frequencies below 7 Hz. During periods of decreased relative power below 1 Hz and increased relative power above 4 Hz, dreaming

occurred in N2. However, for REM sleep, no EEG predictors of dreaming were observed.

3.1.2 NREM Sleep

Another study performed by Siclari et al. investigates dreaming during NREM sleep [85], as an extension to the previous research investigating the neural correlates of dreaming [82]. They wanted to identify how the hallmarks of NREM sleep, sleep spindles and slow waves relate to dream experiences. The study used the serial awakening paradigm on 14 subjects using high-density EEG of 256 electrodes [83]. They used PSD for calculating spectral power changes associated with different types of slow waves. Source modeling was used for evaluating the cortical localization of high-frequency power increases after slow waves. They showed that reports of dreaming had fewer, smaller and shallower slow waves, and faster spindles, compared to reports of no experience. This was especially prominent in the central and posterior cortical regions. They also found an indicator of successful recall of dream content; a minority of very steep and large slow waves in the frontal areas, occurring during decrease of slow wave activity and linked with increase of high-frequency power.

Scarpelli et al. investigated the EEG correlates of dream recall from NREM sleep awakenings [86]. The researchers wanted to test two hypotheses: Whether dream recall share similar brain mechanisms to those involved in encoding of episodic memory during wakefulness, or if dream recall depend on the amount of synchronized EEG activity during sleep preceding the awakening. Data was collected from 14 healthy subjects, who were asked to maintain a regular sleep-wake habit and to keep a sleep and dream diary for the week preceding the experiments. The subjects were awakened in N2 sleep and filled out a sleep and dream report. The experiment did not consider when the subject could remember to have dreamt, but not the content - defining this as no recall. The data was collected by PSG recordings using an EEG cap with 28 EEG channels and a sampling rate of 250 Hz. The recordings were high-pass filtered with a 3-second time constant and low-pass filtered at 30 Hz. For the analysis the power spectra for each channel were computed using FFT, log-transformed, and divided into delta, theta, alpha, sigma, and beta bands for statistical analysis. The researchers performed paired *t*-tests to see whether dream recall and no recall differed. The main finding from the study is that dream recall is predicted by a lower delta wave activity over the left hemisphere compared to no recall. The work points to that dream recall is dependent on the amount of synchronized EEG activity and that dream recall is more likely when subjects exhibited greater brain activation.

A study aiming to identify dreamful or dreamless sleep using unsupervised ML, based on feature extraction of EEG spectral power, was conducted by Wong et al. [87]. They employed the same serial-awakening paradigm as Siclari on nine participants [82, 83]. Two teams were involved in the experiments: the data team, responsible for measuring the participants' EEG activity during sleep and collecting dream reports, and the analysis team, tasked with predicting the presence of a dream experience during NREM (N2 and N3) sleep. The data team provided 54 one-minute PSG recordings of NREM sleep stages 2-3 in balanced proportions, with half containing dreamful sleep and the other half containing dreamless sleep. The analysis team attempted to separate dreamful and dreamless sleep using an evidence accumulation clustering approach with modifications. The results showed that EEG spectral power could not be effectively used to detect specific patterns distinguishing the classes.

3.1.3 REM Sleep

A study performed on the dream emotion evaluation dataset (DEED) dataset has been conducted by Moctezuma [88]. The experiments were done using 6 EEG channels on 38 subjects who were awakened after REM sleep, followed by providing a dream report. For the classification, feature extraction using discrete wavelet transform (DWT) on the data was fed into multiple classical ML algorithms, and raw data was fed into the CNN model, EEGNet. The study considered three separate classification tasks: Neutral emotion versus positive/negative emotion, positive versus negative emotion, and dreamless versus dreamful sleep. The resulting accuracies were 0.85 ± 0.02 for dream versus dreamless sleep, 0.64 ± 0.03 for positive versus negative emotional dreams, and 0.66 ± 0.02 for neutral versus negative/positive emotions in dreams. The results show that EEGNet gave the best metric scores and the lowest standard deviation. Furthermore, the NSGA-II algorithm was used for optimizing the selection of EEG channels. The results were almost identical to the accuracies obtained in the classification of dreamless and dreamful sleep [88].

The findings of Siclari received criticism from Ruby, who stated that neural correlates of dreaming have not yet been identified [89]. One of the points highlighted was that the serial awakening paradigm did not consider the possibility of a subject experiencing a dream without any recall of the experience. This is different to Siclari's definition of DEWR, which are dream recalls without recall of specific content. Such instances are merged with NE. Consequently, dreamful and dreamless sleep might have gotten mixed in the no-dream collection. Furthermore, Ruby points to that previous and later EEG studies using the serial-awakening paradigm have resulted in different and inconsistent results.

3.2 Brain Activity for Specific Dream Content

Horikawa et al. conducted a study to investigate whether visual experience during sleep onset (sleep stage N1 and N2) is represented by brain activity patterns shared by stimulus perception during wakefulness [90]. The study involved three subjects, using fMRI and PSG for data collection and sleep scoring. The subjects were awakened during sleep onset and provided a verbal report describing their visual experience before awakening. The procedure was repeated until minimum 200 awakenings for each subject was attained. The subjects were awakened on average every 342 seconds. Words describing visual objects or scenes were extracted from the verbal reports and mapped to a lexical database where semantically similar words are grouped together as synsets in a hierarchical structure. The words were grouped into base synsets that appeared in at least 10 reports from each participant. The data collected from the fMRI was labelled with a visual content vector where each element indicated the presence or absence of a base synset in the verbal report. A ML decoder was constructed by training SVMs on fMRI data measured while the subject viewed images related to the synsets and subsequently used for prediction. The results indicate that content-specific patterns in brain activity is highly consistent between perception and sleep-onset imagery. Neural measurements could thus be used to uncover the subjective visual contents of dreaming.

A study conducted by Dresler et al. combined brain imaging and lucid dreaming to show that a pre-defined motor task performed in a dream elicits similar neuronal activation as when performing the task during wakefulness [91]. Six subjects, who had trained and practiced lucid dreaming for several

years, participated in the study. Data was collected using fMRI and near-infrared spectroscopy during early morning REM sleep. The subjects were tasked with indicating a lucid dream by performing a left-right-left-right (L-R-L-R) eye movement. They should then clench the left hand 10 times, perform a new L-R-L-R movement, and clench the right hand. The sequence was performed for as long as possible, with the L-R-L-R indicating a hand switch. Two subjects succeeded in performing the task during lucid REM sleep. The subjects performed both actually executed and imagined hand clenchings during wakefulness for comparison. The results show that hand clenching during lucid dreaming elicits neuronal activity in the sensorimotor cortex. This activity is similar to that of an imagined hand clenchings, which overlaps with the elicited activity during executed hand clenchings. The study indicates that specific motoric contents of REM dreaming can be visualized by neuroimaging.

Siclari et al. investigated whether differences in high-frequency power could identify specific contents of DEs during REM sleep [82]. They found a correlation between high frequency and the dimension of thinking in the frontal regions of the brain and the dimension of perception in the parietal, occipital and temporal regions. Specifically, they found that DEs containing faces were associated with increased high-frequency activity in a temporo-occipital region, which during wakefulness is associated with face perception. Increased high-frequency activity in the right posterior parietal cortex, an area involved in spatial perception, was found for DEs with a definite spatial setting. Moreover, DEs with reports of the sense of movement were associated with high-frequency activity in an area involved in the perception of biological motion and in viewing body movements. Additionally, DEs containing speech were found to have increased high-frequency activity over a left posterior region corresponding to an area associated with speech and language called Wernicke's area.

3.3 Frequency Activity in Sleep

A study investigating the presence and cortical distribution of negative delta waves in REM sleep was conducted by Bernardi et al. [92]. 16 subjects participated using high-density EEG, 256 channels. They used independent component analysis (ICA) to remove potential ocular, muscular and electrocardiograph artifacts. For the detection of delta waves, they used an algorithm based on the analysis of consecutive signal zero-crossings. Following this analysis, they found two clusters of delta waves with distinct characteristics; 1) a frontal-central cluster with 2.5-3.0 Hz called sawtooth waves, 2) a medial-occipital cluster with frequencies below 2 Hz. The sawtooth waves were found to occur in bursts and were associated with increased gamma activity and rapid eye movements. They share many characteristics with ponto-geniculo-occipital (PGO) waves. These waves were found to have an occipital-temporal and a frontal-central component, using source modeling. The slower waves, however, displayed a negative correlation with gamma activity and were also found in NREM sleep. The findings indicate a spatially and temporally heterogeneous state of REM sleep.

Vijayan et al. performed a study exploring activity in brain regions during REM sleep [93]. They used EEG, EMG and EOG for sleep scoring and recordings characterised as REM sleep were used for further analysis. The recordings were taken on total five subjects diagnosed with epilepsy. The subjects were monitored using intracranial depth electrodes, where the distal-most contacts focused on the medial temporal and frontal regions. The results showed rhythmic waves of neuronal activity in the frontal area, where theta and beta waves dominated. The waves were especially prominent in the frontal

lobe called the dorsolateral prefrontal cortex, and the anterior cingulate cortex.

A review summarizing the role of PGO waves in dream formation is presented by Gao et al. [94]. PGO waves originate in the pons and propagate to the lateral geniculate nucleus of the visual thalamus and the occipital cortex. Their pathway could be interpreted by the brain as visual information, leading to the visual hallucinations experienced as dreams. PGO waves start to appear during the transition from NREM to REM, and are then persistent throughout REM. The authors suggest that the PGO waves may have a role in physiological functions during REM sleep such as memory consolidation, unlearning, brain development and plasticity, and mood regulation.

Chapter 4

Methodology

This chapter documents the materials used and the methods applied for the project. Each method is first presented in general terms, then detailed for each dataset.

4.1 Materials

4.1.1 The Zhang & Wamsley (ZW) Dataset

The Zhang & Wamsley (ZW) dataset is presented by Zhang and Wamsley [84, 95] and made available through the DREAM database [17]. The data was collected from 28 subjects sleeping in the laboratory, where they were awakened throughout the night to provide a dream report. The subjects were instructed to verbally describe "everything that was going through your mind just before I called" whenever awakened. They were encouraged to give all details on their experience, regardless of considering whether it was a dream or not. If they could not remember an experience, they reported this.

The participants' sleep onset was determined through online PSG monitoring, and they were subsequently awakened up to 13 times to report their experience. Of these awakenings, up to 10 occurred within the initial hour of the night, shortly after the onset of sleep. As a consequent, the reports from the sleep onset phase encompassed different NREM sleep stages. Later on, the subjects were awakened after at least 10 minutes of PSG-defined N2 and 5 minutes of PSG-defined REM. In this way, one N2 sleep and one REM sleep report were collected. Upon morning awakening, a final subjective report was reported, irrespective of the sleep stage.

The EEG recordings were conducted using a high-density electrode cap containing 58 EEG electrodes positioned according to the international 10-10 system. Data were captured at a sampling frequency of 400 Hz using a Grass-Telefactor AURA amplifier equipped with a high-pass filter set at 0.1 Hz. The average between the left and right mastoids, channel A1 and A2, are used to reference the EEG channels.

They collected 308 samples in total, each minimum 73 seconds long, with a label assigned as either NE or DE. The label was decided by two independent judges, who did not know the sleep stage, and followed the definition of dreams used in the study; a mental experience recalled from sleep. Of these, 181 were in during N1 and N2, referred to as NREM, 26 in REM and the remaining were collected at

the last epoch before the report was "wake" and in N3. Overall proportions of the sleep stages and the class distribution of ZW dataset are illustrated in table 4.1.

ZW Dataset			
Sleep Stage	Percentage	No Experience	Dream Experience
N1	25%	23%	77%
N2	34%		
N3	2%		
REM	8%		

Table 4.1: Sleep stage and class distribution for the ZW dataset [95].

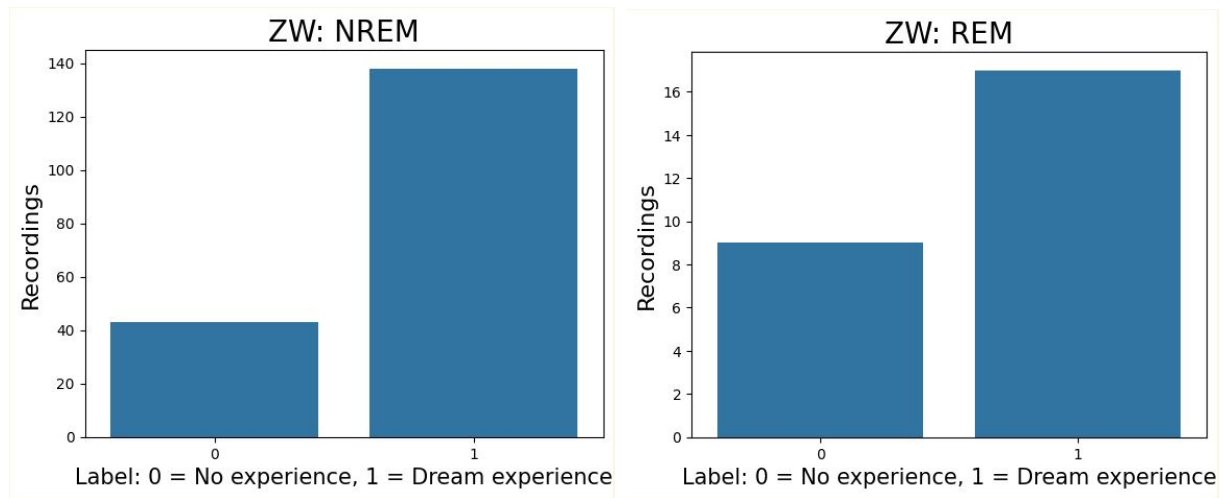


Figure 4.1: Distribution of the classes *No experience* and *Dream experience*, in the NREM and REM sub-datasets of ZW.

Furthermore, in all experiments, electrodes A1 and A2 were selected as reference nodes. The experiments were divided into two parts: one focuses on classifying DE and NE during NREM sleep, while the other focuses on REM sleep. To accomplish this, the entire dataset was partitioned into the sleep stages N1, N2, and REM, by extracting the labels corresponding to the sleep stages. Subsequently, N1 and N2 were merged and collectively referred to as NREM. The unbalanced class distributions for each of the sub-datasets are depicted in fig. 4.1.

Each recording in the sub-datasets was segmented to obtain more data. The length of each segment was determined through analysis testing with various segment lengths: 0.25, 0.5, 1, 2, 3, 4, 5, and 10 seconds. Among these, 0.5-second segments were chosen for the NREM dataset, while 2-second segments were chosen for the REM dataset, as they gave the highest performances.

After the segmenting, the NREM sub-dataset contained 26458 segments in total, while the REM sub-dataset contained 936 in total.

4.1.2 The Tononi Dataset

The dataset Tononi Serial Awakenings, referred to as the Tononi dataset, is presented by Siclari et al. [82, 96]. It is made available through the DREAM database [17]. The experiment was conducted with

the serial awakening paradigm, where the subject sleeping in the laboratory were awakened during N2 sleep throughout the night. Subjects were awakened by a computerized sound that lasted 1.5 seconds. The subject then reported whether they had a dream experience or not through an intercom, which was noted by the researchers. The reports were divided into three classes: Dream experience (DE), dream experience without recall (DEWR), and no experience (NE). The class distribution is 42% DE, 19% DEWR, and 39% NE.

The dataset contains a total of 261 recordings from 36 subjects. The number of recordings and subjects available in the dataset differs from the numbers presented in the dataset description [96], which report 240 recordings from 32 subjects. Recordings were performed using a 256-channel high-density EEG system, with a sampling rate of 500 Hz. Sleep stage scoring was performed over 30-second epochs. Each recording consists of the last 120 seconds before awakening.

For further work, the recordings were divided into 2-second long segments, low-pass filtered with a cutoff frequency of 45 Hz, and resampled at 150 Hz. These values were shown through experimenting to give the best classification performance.

4.1.3 Deep Learning Model

EEGNeX is a CNN-based DL model developed for EEG-based classification tasks [97]. During development the researchers identified four key pillars for their design: Spatial filtering, spatial feature extraction, temporal feature extraction, and model parameter space and performance stability. The design is a continuation of the EEGNet model [98]. EEGNeX has a three-part architecture excluding input and classification parts, as seen in fig. 4.2. The first part, responsible for spectral feature extraction, is composed of two two-dimensional convolutional layers. The second part is a depth-wise two-dimensional convolutional layer, acting as a spectral filter. The third part performs spatial feature learning with two two-dimensional convolutional layers. The classification part consists of a single dense layer and an output.

4.1.4 Development Tools

Software

The code for this project is developed using the Python programming language, which allows the use of multiple toolboxes related to data engineering and ML. The work requires handling multidimensional time series in the form of EEG signals, making Python a suitable choice. For data processing, the most used toolboxes are pandas, numpy, scipy, and MNE-python [99–102]. ML and DL algorithms have been created mainly using scikit-learn, tensorflow, and keras [103–105]. Github is used to facilitate collaboration throughout development and to make the project available after finishing.

Computational Resources

Experiment configurations with high computational complexity were performed using Idun, a computing cluster at NTNU for rapid testing of high-performance computing software [106]. The cluster is equipped with 185 compute nodes, 8264 CPU cores, and 210 GPUs¹. The combination of ML tasks

¹As of 02.06.2024: NTNU (2024) Idun - High Performance Computing. Available at: <https://www.hpc.ntnu.no/idun> (Accessed: 02.06.2024)

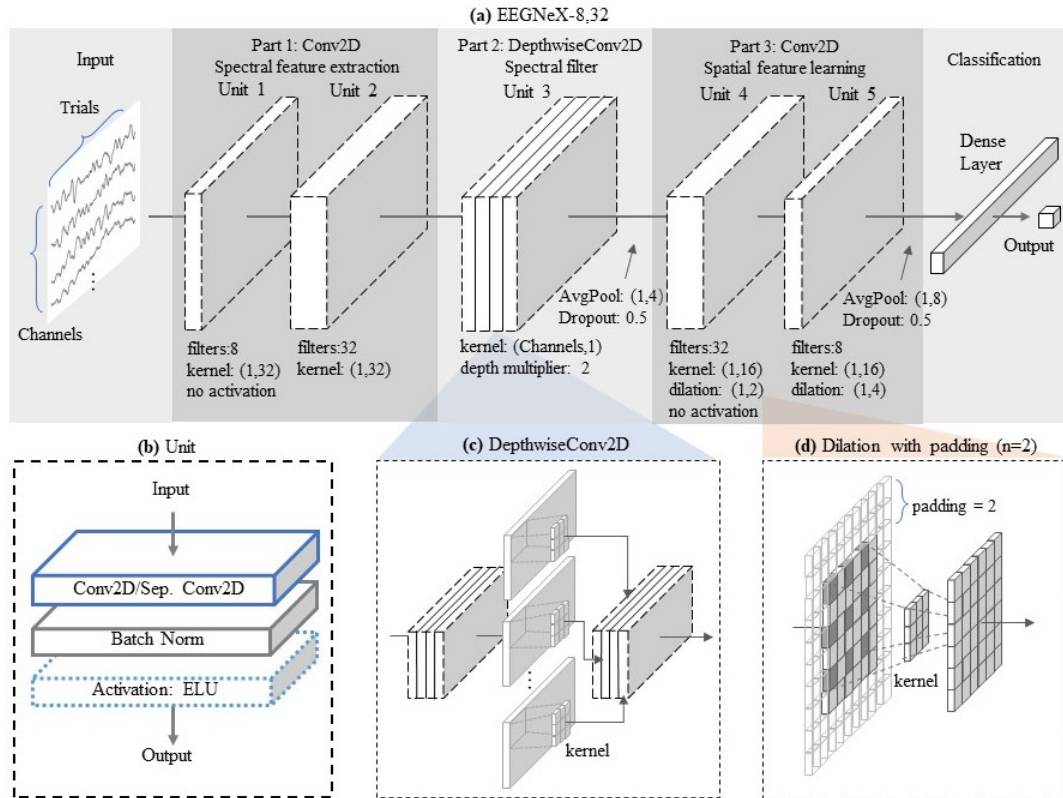


Figure 4.2: Architecture of the EEGNeX Model [97].

and EEG data with a high number of channels is computationally heavy, making Idun a valuable resource.

Artificial Intelligence Tools

The AI tools Copilot by Microsoft², ChatGPT by OpenAI³, and Github Copilot have been used during the development of this project. Copilot by Microsoft and ChatGPT have been accessed through the web application, where they have been used primarily for code-assisting in the form of generating examples on how to use specific toolboxes and for debugging, but also for explaining technical concepts. Github Copilot is used as a plug-in feature in Visual Studio Code, where it suggests code completions and can develop simple code based on comments. These tools have proven valuable in the project, as they allow asking direct questions about specific topics using natural language. All information obtained is cross-checked with other sources, as the tools can provide false information.

4.2 Data Filtering Using MEMD

The MEMD was employed on the data as a filter with the aim of enhancing classification performance. This filter eliminates unnecessary frequency components by constructing a new signal based on a subset of IMFs. Three IMFs closest to the original signal, determined by Euclidean distance, were selected. The number of IMFs was determined experimentally to maintain consistency. These three

²Microsoft (2024) Copilot. Available at: <https://copilot.microsoft.com>

³OpenAI (2024) ChatGPT. Available at: <https://chatgpt.com>

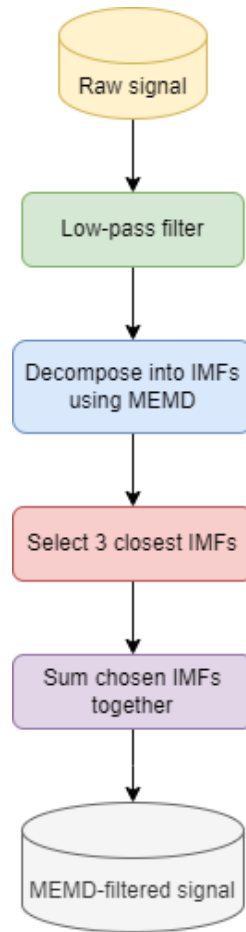


Figure 4.3: Process of obtaining MEMD-filtered signal from raw signal.

IMFs were then combined to reconstruct the MEMD-filtered data. In the case of the signal containing only two IMFs, the third IMF contains only zeroes. The process of obtaining the MEMD-filtered signal for one signal is shown in 4.3.

All experiments using MEMD-filtered data are based on applying MEMD on all channels, except for one experiment in Tononi, exploring the use of MEMD on only a subset of 10 channels. This was done to explore the differences of using MEMD before and after channel subsets were chosen.

4.2.1 The ZW Dataset

The process of using MEMD as a filter was applied to both the NREM and REM sub-datasets of ZW. The MEMD was applied on the total recording of 73 seconds, obtaining the MEMD-filtered data. The MEMD-filtered data was subsequently segmented into 0.5 seconds and 2 seconds for NREM and REM data respectively before being used as input for the classification tasks.

4.2.2 The Tononi Dataset

For the Tononi dataset MEMD was applied on the 2 second long segments before being used as input to the classification tasks. Because of the computational load, only 2000 2-second segments were used.

An additional effort, for comparison to the ZW method, was made by applying MEMD to the last 73 seconds of 36 recordings with all 256 channels, before segmenting the MEMD data into 2 seconds. Channel subsets of 29 and 10 channels were used as input to the classifiers.

It was in interest to apply the same methods across two different datasets, to see the effect of various methods, such as applying MEMD-filtered data. In order to compare the experiments performed on Tononi with the experiments on ZW, it was necessary to have as many similar configurations as possible. Meaning that, since the experiments of ZW were performed using 73 seconds recordings, it was necessary to use the same for Tononi.

The NA-MEMD algorithm was applied on the Tononi data using 2-second segments on 10 channels. This was done with the intention of increasing the performance, compared to only using the MEMD algorithm as a filter. For this, 4 channels of noise were generated and concatenated with the other 10 channels. This resulted in 14 channels of data, that were used as input to the MEMD algorithm, extracting the IMFs of the data. Further on, the three closest IMFs in terms of Euclidean distance were extracted and used for classification, just as when obtaining the MEMD-filtered signal.

4.3 Feature Extraction and ML Algorithms

Two main feature extraction methods are used: 1) Data transformation with PCA and CSP, 2) PSD calculation, both on raw and MEMD-filtered data. Classifications are performed with classical ML algorithms. For all methods, the metrics accuracy, F-score, precision, recall, AUROC, and kappa score are calculated. The classifications are performed using five-fold cross-validation with a train-test split ratio of 80-20.

4.3.1 PCA, CSP, and ML

The first method used for classification is with PCA and CSP for feature extraction, before input into ML algorithms. CSP is selected as it has shown to be useful in previous research on EEG classification tasks [107–109]. PCA and CSP are defined as two models, which are trained to fit the data at hand. The data is separated into train and test sets, where the train set is first used as input to a PCA model. The PCA learns from the features and returns PCA-filtered data. This data is used as input to a CSP model, which learns from the PCA-filtered data and returns CSP-filtered data. The data is then standardized by subtracting the mean and scaling to unit variance. The models are now fitted to the data and is used to transform the test set.

For both the PCA and the CSP the number of components returned are equal to the number of channels. Although PCA is commonly used to reduce the dimensionality of the data, this proved, during experimenting, to reduce the classification performance. This results in a higher computational load, but without a performance trade-off.

Classification is performed using KNN, XGB, and SVM with RBF kernel. For KNN, classification is performed for $k \in (1, 10)$ and the k with the highest performance is selected.

The ZW Dataset

PCA, CSP and ML was used on ZW, both for NREM and REM data in order to classify DE versus NE. For both sub-datasets, the method was applied using raw or MEMD-filtered data.

The Tononi Dataset

The PCA, CSP, and ML pipeline was applied for both raw and MEMD-filtered data. For the raw data, three class configurations are used: 1) DE versus NE, 2) DE versus NE and DEWR, and 3) DE and DEWR versus NE. For the MEMD-filtered data, only DE versus NE is investigated. To investigate the role of PCA-filtering in the PCA, CSP, and ML pipeline on the MEMD filtered data, the MEMD-filtered data is also used directly as input to CSP, without using PCA.

4.3.2 PSD-Based Classification

The other approach is to use PSD as a feature for classification, as previous studies have shown differences between DE and NE based on this feature [27, 82, 84, 86]. PSD is calculated for each time-series in the data, and the average value across all frequencies is used as the feature. For the raw data, the average PSD value is calculated over each channel. For the MEMD-filtered data, the average PSD is calculated over each IMF in each channel. The average PSD values are then used as input to the classifiers.

The classifiers are KNN, XGB, and SVM with RBF kernel. For KNN, classification is performed for $k \in (1, 10)$ and the k with the highest performance is selected.

RF classifier has not been used for this approach due to observations of performances in previous experiments.

The ZW Dataset

In order to classify DE versus NE using NREM or REM sub-datasets of ZW, only the MEMD-filtered data was applied when using PSD and ML. This is due to observations from previous experiments.

The Tononi Dataset

The PSD and ML pipeline was applied for both raw and MEMD-filtered data. For the raw data, three class configurations are used: 1) DE versus NE, 2) DE versus NE and DEWR, and 3) DE and DEWR versus NE. For the MEMD-filtered data, only DE versus NE is investigated.

4.4 Deep Learning Classification

The chosen DL algorithm for this project is EEGNeX. This model was shown in a benchmark study to outperform other state-of-the-art models in EEG-based classification tasks, demonstrating an increase in accuracy from 2.1% to 8.5% [97].

All tasks were performed using the default parameters of the algorithm, with the adaptive moment estimation (ADAM) optimizer and categorical cross-entropy as the loss function. The hyperparameters were tested through experimenting for each task. The metrics accuracy, F-score, precision, recall,

ZW dataset		
Sub-datasets	Epoch size	Batch size
NREM	80	16
REM	83	16

Table 4.2: Epoch and batch sizes for the classification tasks within the ZW dataset

AUROC, and kappa score are calculated for performance evaluation. Each task was performed using five-fold cross-validation with a train-test split ratio of 80-20 to reduce the statistical uncertainty of the metrics.

4.4.1 The ZW Dataset

EEGNeX was applied to both NREM and REM sub-sets of ZW, using raw or MEMD-filtered data as input, to classify DE versus NE.

The epoch and batch sizes chosen for the sub-datasets within the ZW dataset are presented in table 4.2. The epoch size is determined based on an analysis of several runs of the training.

The batch size was determined through hyperparameter optimization, leveraging the Python library Hyperopt. Hyperopt was applied for this, as it serves as a valuable tool for doing distributed asynchronous hyperparameter optimization [110].

Early stopping and reduce learning on plateau were also tested, but discarded due to no increase in performance.

4.4.2 The Tononi Dataset

The hyperparameters to choose were the number of epochs and the batch size. Each classification was run over 120 epochs and a batch size of 64. The number of epochs was selected through experimenting, with the objective of finding an appropriate value which resulted in neither over- nor underfitting. The batch size was selected as the highest possible power of two to increase computational time, while respecting the program's memory capacity. Two regularization algorithms, early stopping and reduce learning on plateau, were tested and discarded as they failed to improve performance.

Based on the three classes in the dataset, three configurations are chosen for classification: 1) DE versus NE, 2) DE versus NE and DEWR, and 3) DE and DEWR versus NE.

4.5 Data Augmentation Using MEMD

Maintaining dataset balance is important to prevent biases toward classes with a higher number of samples. To achieve this, data augmentation techniques are applied to rebalance the imbalanced datasets. The objective is to generate additional artificial samples for the underrepresented class, thus ensuring a more equal distribution across classes.

In this project, MEMD is used for data augmentation. The MEMD technique is employed to decompose real EEG signals into their respective IMFs, combining one IMF from each segment to create a

new artificial augmented data sample. In this way, MEMD serves as a tool for data augmentation.

A set of random segments, equal to the number of IMFs in the artificial segment, from the NE class within the training dataset was chosen for further processing. Although the MEMD algorithm maintains a consistent number of IMFs across channels, this consistency is not ensured across segments. Each segment could vary in the number of extracted IMFs, and it were therefore necessary to pre-define a number of IMFs for the new artificial segments. This quantity was determined through an observation of IMF extraction from multiple segments and choosing the number corresponding to the majority of the segments. Based on this, 6 IMFs were designated as the preferred count for each segment. Meaning, during the selection of segments for further processing, they were eliminated if containing less than 4 or above 6 IMFs. This aimed to ensure that the resulting augmented segments would come from similar segments, since they had almost the same number of IMFs across segments. In instances where segments contained only 4 or 5 IMFs, IMFs of zeros were added.

Following the extraction of IMFs from segments, they were sorted into a list encompassing all IMFs across channels and segments. Subsequently, one random IMF from each segment for each channel was selected, ensuring that the 6 IMFs for each channel did not originate from the same segment, nor the same IMF number. Meaning that, for instance in channel 1, IMF 1 could be picked from segment 2, IMF 2 from segment 5, IMF 3 from segment 1, IMF 4 from segment 3 etc. Finally, the selected IMFs for each channel were summed together, resulting in a new complete artificial segment.

This process was iterated multiple times, with each iteration involving the selection of train and test splits through random partitioning of the original dataset. Consequently, the augmented segments were effectively constructed from the entirety of the dataset. Thereafter, for classification purposes, the original data was divided into train and test sets, with the augmented data incorporated into the train set to ensure balance. Following this, classification was executed on the balanced training set and evaluated on a subset of the original dataset, the test set. This means that the augmented data samples used in training could be constructed of IMFs from the test set.

4.5.1 ZW with Data Augmentation

The technique of MEMD for data augmentation has been employed on the NREM sub-dataset from ZW. As illustrated in fig. 4.1, it exists an imbalance in the distribution of NE and DE. Consequently, data augmentation has been used to generate artificial NE samples, thereby balancing the sub-dataset.

The number of augmented segments were chosen based on the difference between the number of DE samples and NE samples in the train data:

$$\text{needed_segments} = \text{number_of_DE} - \text{number_of_NE} \quad (4.1)$$

If the amount of needed segments was bigger than the total number of augmented segments, we chose the total number of augmented samples instead. The total number of augmented samples were 12171 samples. This was repeated 5 times, for each fold in the 5-fold cross validation.

4.6 PSD Analysis on MEMD-Filtered Data

An effort to understand the difference between DE and NE is made using MEMD-filtering and PSD calculation. By evaluating the power produced in the channels we identify where brain activity is occurring for both DE and NE. Each IMF in the MEMD-filtered data is sorted into one of four frequency bands: delta, theta, alpha, and beta, defined in table 2.1. The gamma band is not included in the analysis, since previous studies have not included gamma, and the sleep scoring manual provided by AASM do not mention it [26, 82]. The sorting is made by finding the frequency corresponding to the maximum value of the PSD of the IMF. The grand average PSD for all IMFs is calculated for each band, for each channel, is made for both DE and NE. The ratio between the PSD of each class, calculated as

$$Ratio = \frac{PSD(X_{DE})}{PSD(X_{NE})} \quad (4.2)$$

is presented in a topographic map. They display the PSD ratios across all channels, for each of the four frequency bands.

4.6.1 The ZW Dataset

The topographic map for the PSD ratio for NREM data from ZW is created using 73 second long segments into the MEMD algorithm. For this data we have investigated the recordings to identify potential outliers, and removed them before creating the topographic map. This is done by plotting and visually inspecting the power spectra of all channels for each recording. The remaining recordings are used to calculate the PSD ratio for the topographic map.

For the REM data from ZW, we also used 73 second long segments to create the MEMD-filtered signals, however, the analysis of potential outliers in the signals was not performed.

4.6.2 The Tononi Dataset

For the Tononi dataset two topographic maps are presented: The first where the MEMD-filtering is applied on 2 second segments, and the second where MEMD-filtering is applied on 73 second segments. This is done to investigate how the length of the data before applying MEMD affects the topographic map.

The topographic maps are plotted for PSD ratios between 0 and 2, where 1 equals equal ratio. This is to clarify when the PSD of a DE sample is stronger than of NE, and vice versa.

4.7 Channel Selection

Using a subset of EEG channels instead of high-density EEG proves beneficial in reducing the dimensionality of the data and the computational cost for training the ML models. Employing all channels can be overly time-consuming, making computations unfeasible, especially for BCIs or real-time classification using EEG signals. In addition, the importance of the channels could be reflected in important brain regions to classify dream experiences. Hence, channel selection has been implemented

[111]. In this project the main channel selection algorithm is permutation-based channel selection, but an alternative channel selection using PSD ratio has also been made.

Permutation-based channel selection is implemented as explained in section 2.10:

1. A baseline model is selected and trained on the training data.
2. The baseline performance is calculated by evaluating the baseline model on the test data.
3. For each available channel k for all epochs in the test data, the data within is randomly shuffled (permuted) to corrupt the channel. The baseline model is evaluated on the test data with a corrupted channel and channel importance k is defined as the difference between the baseline performance and the new evaluation.

In our work we have used kappa to measure the performance of both the baseline and after channel permutation. The importance of channel k is the difference in kappa performance: A positive value indicates that the performance after permutation decreases, indicating that the channel is important. A negative value indicates that the channel is detrimental to the performance. The algorithm returns a vector of channel importances for all available channels. To select an n -channel subset, the channels are ordered by importance and the top n channels selected.

The other channel selection method is to utilize the PSD ratios between DE and NE. Channel subsets are then selected based on which channels have the highest difference. Each channel is given an importance based on the highest PSD ratio difference from 1 across the four frequency bands, calculated as

$$Importance = |Ratio - 1| \quad (4.3)$$

The ratio offset by 1 is to also capture when the ratio is low. The n channels with the highest importance are used for an n -channel subset. In the case of no available ratio in a particular channel, the PSD ratio is set to 1.

4.7.1 The ZW Dataset

For the NREM and REM dataset of ZW, permutation-based channel selection has been applied on raw or MEMD-filtered data using two methods; 1) PCA, CSP and KNN, 2) EEGNeX. The importance of each channel, based on the performance delta is presented in a topographic map. The important channel subsets of 29 and 10 channels are extracted from the permutation-based channel selection and further used in classification of DE vs NE.

For the PSD-based channel selection, MEMD-filtered data of 73 second recordings are used as input, both in the NREM and REM dataset of ZW. The 29 and 10 most important channels in terms of highest PSD ratio are then used in classification of DE vs NE. For NREM, PSD as a feature in combination with XBG is used, while for REM, PSD and KNN is used.

For the REM dataset of ZW, the approach has also been further developed, by removing the first IMF of all signals and then reconstruct the MEMD-filtered signal using the three closest IMFs. This is

performed with the purpose of removing the high-frequency components, as the first IMF contain any high-frequent noise present.

4.7.2 The Tononi Dataset

The data from Tononi has been tested on channel subsets of 50, 30, 20, and 10 channels to investigate the performance on fewer channels. This has been done for the raw data as input to EEGNeX and to the PCA, CSP, and ML pipeline. Channel subsets have been defined using permutation-based channel selection with the respective method as base. The channel importances are further represented in a topographic map.

A channel subset defined using the PSD ratio has been used as input for the MEMD, PSD, and ML pipeline. The last 73 seconds from 36 recordings, with all 256 channels, were used as input to the MEMD algorithm. We have defined subsets of 29 and 10 channels using the PSD ratio of these signals. This experiment is performed in the same way as for the ZW dataset to compare the two results using the same method, albeit with different channels as input to MEMD. The 29 and 10 channel subsets from PSD ratio are compared to two subsets of 29 and 10 random channels to investigate whether the PSD ratio-based channel selection improves the performance.

Permutation-based channel selection was used to identify a 10 channel subset used for experimenting with four different methods to increase the performance on a small subset. The base model is PSD and XGB with MEMD-filtered data. The subset was selected from the raw data and used as input to MEMD. The first and second method was with the regular IMF selection, by selecting the three closest IMFs. The three IMFs were tested on both PSD and XGB and CSP and XGB. For the third method the IMF selection was modified to ignore the first IMF, as this one typically contains noise, tested on PSD and XGB. The fourth method was made with the NA-MEMD algorithm using 4 noise channels and selecting the top three IMFs, including the first. The data was tested on PSD and XGB. These methods are compared to selecting the same channel subset after MEMD-filtering of all 256 channels before classifying using PSD and XGB.

Chapter 5

Results

Our experimental results are divided into three main categories: *ZW NREM data*, *ZW REM data*, and *Tononi Data*. The results obtained with the different methods are presented; Classification with all channels, PSD analysis and channel importance, and classification using channel subsets.

5.1 Classification Using the ZW Dataset During NREM Sleep

Here, we present the classification performed with the ZW dataset for DE vs NE when subjects were awakened during NREM sleep.

In the first approach to classifying DE vs NE in NREM sleep, we used 58 EEG channels. The techniques used are 1) PCA, CSP, and ML, 2) PSD and ML, and 3) EEGNeX. We present the classification performance obtained using raw data and balancing the dataset with artificial data augmentation. We also present results with and without MEMD-filtered data as input to the classifiers. The low-passed filtered data is referred to as the raw data throughout the section.

To understand which EEG channels are more important for the classification task, we present a topographic map obtained with permutation-based channel selection.

We also show a PSD-based quantitative analysis of DE and NE brain activity during NREM sleep.

We repeated the classification process using channel subsets selected with the permutation-based channel selection or the PSD-based analysis.

5.1.1 Classification Using 58 Channels

Figure 5.1 shows the classification performance for DE vs NE using the ZW dataset during NREM sleep with the 58 channels and different feature extraction techniques in combination with ML algorithms. There are two configurations represented for both figures in fig. 5.1, 1) the original NREM data, 2) the original NREM concatenated with augmented NE samples. The number of augmented samples are added to the train set used during training of the model, and the amount of augmented samples vary based on the amount of needed segments to obtain a balanced train set, as explained in eq. (4.1) in section 4.5.1.

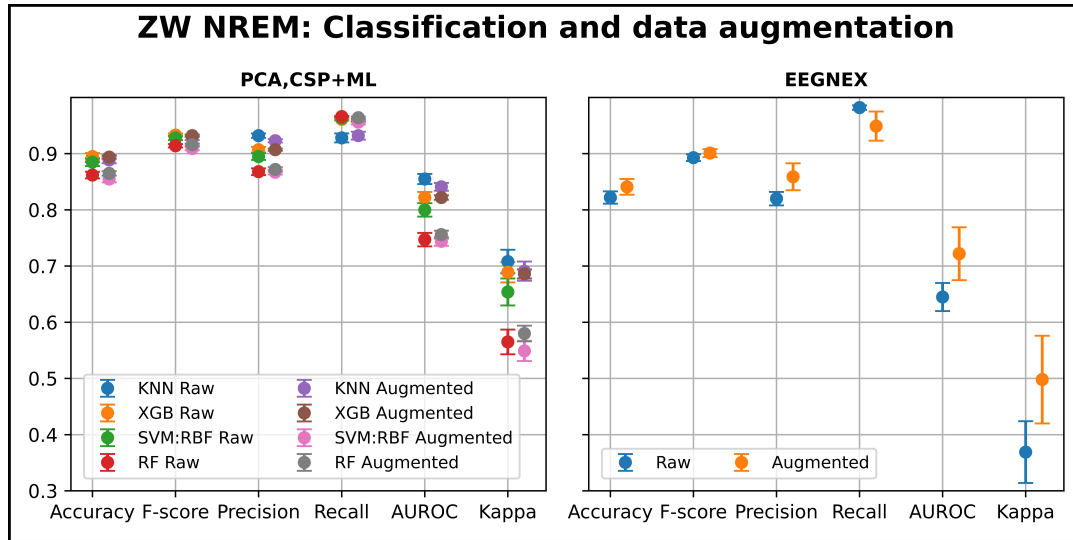


Figure 5.1: Classification performance for DE vs NE with and without data augmentation, using different ML algorithms (left) and EEGNeX (right).

PCA and CSP for feature extraction in combination with KNN classifier using the original data, gets the highest classification performance with 0.89 accuracy, 0.93 F-score, 0.93 precision, 0.93 recall, 0.86 AUROC and 0.71 kappa. The lowest performance was obtained with the SVM (0.89 accuracy, 0.8 AUROC and 0.65 kappa) and RF (0.86 accuracy, 0.75 AUROC and 0.57 kappa) classifiers.

When incorporating artificial data samples in the training data, there is no improvement, except for the RF classifier, showing an increase in all metrics, except for recall, which decreases 0.002, while the other metrics increase 0.002-0.004 for accuracy, F-score and precision, 0.009 for AUROC and 0.015 for kappa. For all the other classifiers, KNN, XGB and SVM, the performance decreases. For KNN the performance decreases between 0.002 and 0.017, while for recall it increases with 0.004. For XGB, precision and AUROC remain the same, while kappa have a maximum decrease of 0.003. For SVM, performance decreases between 0.007 and 0.105 for all metrics.

For EEGNeX all metrics, except recall, have higher performances when using data augmentation. The increase is less than 0.04 for accuracy, F-score and precision, 0.07 for AUROC, and 0.13 for kappa. As kappa and AUROC are metrics that are heavily dependent on class balance, it could explain the greater increase for these metrics. Even though the performance increases, it is still lower than using PCA, CSP and KNN with raw data.

The standard deviation of the metrics when using EEGNeX is higher than for the metrics when using classical ML. We can also see for EEGNeX that the standard deviation of the metrics increases when applying data augmentation, which indicates a lower robustness than classical ML algorithms.

Another approach of classifying DE and NE in NREM has been done using 58 channels and MEMD-filtered data as input to the classifiers, depicted in fig. 5.2 and fig. 5.3. The techniques used are 1) PCA, CSP for feature extraction and different ML algorithms, 2) PSD for feature extraction and different ML algorithms, and 3) EEGNeX.

For PCA, CSP and ML, the use of KNN classifier gives the highest performance, using MEMD-filtered data (fig. 5.2), with 0.94 accuracy, 0.96 F-score, 0.95 precision, 0.97 recall, 0.91 AUROC and 0.84 kappa.

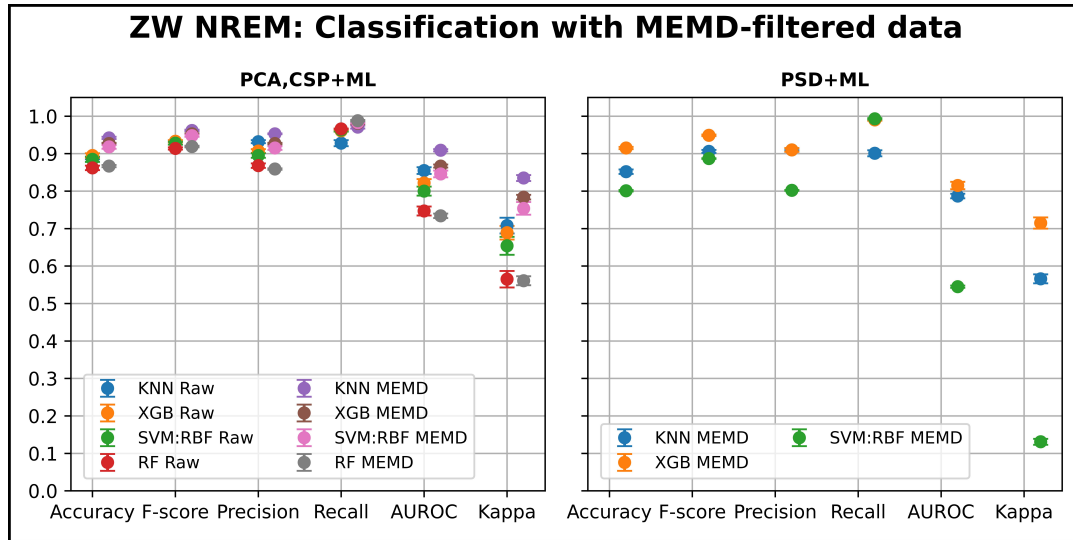


Figure 5.2: Classification performance for DE vs NE with raw or MEMD-filtered data, using PCA, CSP and different ML algorithms (left) and PSD and different ML algorithms (right).

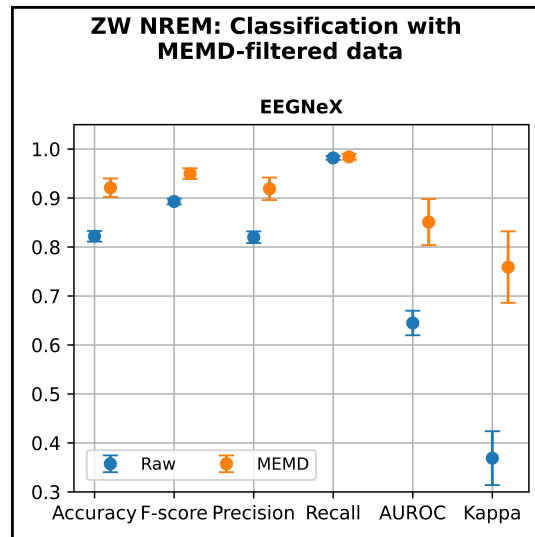


Figure 5.3: Classification performances for DE vs NE with with raw or MEMD-filtered data, using EEGNeX.

When using raw data, the metrics decrease with 0.05 accuracy, 0.03 F-score, 0.02 precision, 0.04 recall, 0.05 AUROC and 0.13 kappa. The MEMD-filtered data as input increases the performance for all classifiers, except for the precision, AUROC, and kappa metric of RF.

The subfigure on the right in fig. 5.8 shows the results of using PSD for feature extraction and different ML algorithms as classifiers. PSD in combination with XGB yields the highest performance with 0.92 accuracy, 0.95 F-score, 0.91 precision, 0.99 recall, 0.82 AUROC and 0.72 kappa. The SVM classifier gives the lowest performance, with a kappa of 0.13.

All the metrics increase when using MEMD-filtered data with EEGNeX; 0.099 accuracy, 0.06 F-score, 0.099 precision, 0.002 recall, 0.21 AUROC and 0.39 kappa (fig. 5.3). The standard deviation is also higher for EEGNeX when using MEMD-filtered data, than raw data. However, the increase when using the MEMD-data is still high enough to be a better option than using raw data, as the minimum kappa

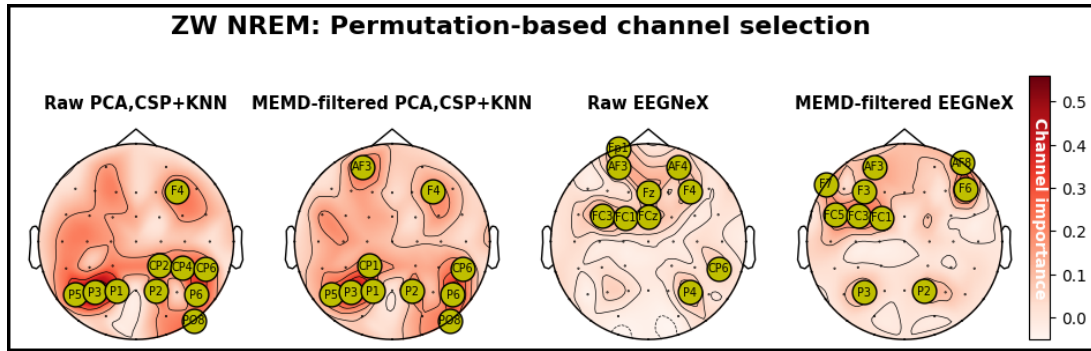


Figure 5.4: Topographic map showing the channel importances for NREM data, using raw or MEMD-filtered data, extracted using permutation-based channel selection. The 10 most important channels are indicated with a yellow circle.

value when using MEMD-filtered data (0.686) is still higher than the maximum kappa value when using raw data (0.424).

Overall, the KNN classifier in combination with PCA and CSP, using MEMD-filtered data outperforms the other methods using feature extraction and ML classifiers and EEGNeX.

5.1.2 Channel Importances and PSD Analysis

The topographic map of channel importances obtained from permutation-based channel selection for NREM data, using raw or MEMD-filtered data with two techniques; PCA, CSP and KNN or EEGNeX, is shown in fig. 5.4. When using PCA, CSP, and KNN with raw data, the most important channels, based on the performance delta of kappa, show predominant activity in the parietal lobe, with some distributed across the frontal and parieto-occipital regions. This indicates that these respective areas are significant for classifying DE and NE. When applying MEMD-filtered data to the feature extraction techniques (PCA and CSP) and subsequently to the KNN classifier, two of the channels are moved to the left parietal and frontal regions, while the rest remains.

Looking at the topographic maps of EEGNeX (fig. 5.4), also created with permutation-based channel selection, the most important channels are rather predominantly located in the frontal and central regions, while two of the channels lay in the parietal region. Using MEMD-filtered data with EEGNeX emphasizes the left frontal area, meaning that dreams in NREM sleep could contain elements of logical thinking.

The important regions marked in the topographic maps of PCA, CSP and KNN compared to the one of EEGNeX, exhibits significant differences. The maximum performance delta is also higher for KNN (0.56) than for EEGNeX (0.33). Therefore, the channel subsets of KNN is more important than for EEGNeX, as the removal of channels effects the performance significantly.

The PSD ratio (DE/NE) based channel importances, calculated on the NREM data using the MEMD-filtered data is shown in fig. 5.5. The topographic map of each frequency band is presented, revealing different patterns of activity for DE and NE.

In delta band, more activity is observed for NE compared to DE, with only a small part in the central region showing increased activity for DE. For theta band, the DE activity is prominent in the parietal,

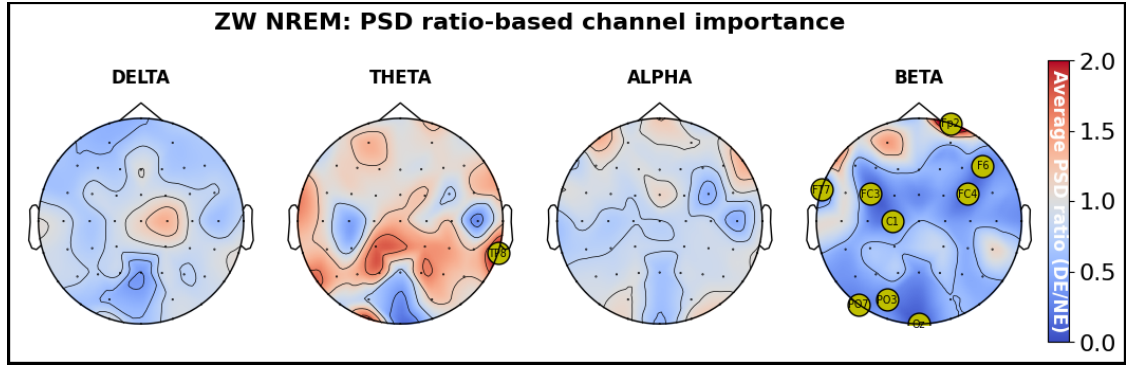


Figure 5.5: Topographic map showing average PSD ratios (DE/NE) for MEMD-filtered NREM data.

temporal and frontal region, showing more activity for DE than in delta band. For the alpha band, most of the activity is centered around neutral, not highlighting any particular regions for either DE or NE. Looking at the map in beta band, the dominating activity of DE is limited to a small area on the right frontal lobe. The frontal, central and occipital regions are however dominated by NE activity and this activity is also visible in the location of the 10 channels with highest PSD ratio.

While the majority of the important channels lie in the beta band, one channel in the theta band captures the dominant activity of DE in the right parietal area. Since most channels are in the beta band, the 10 most important channels extracted from the PSD analysis are more representative of NE activity.

Frequency Band	No. of IMFs Placed in Band	No. of IMFs with Freq of Max amplitude < 0.5Hz
Delta	20508	18679
Theta	474	0
Alpha	888	0
Beta	2142	0

Table 5.1: Summary of IMFs in DE for different frequency bands in NREM ZW

Frequency Band	No. of IMFs Placed in Band	No. of IMFs with Freq of Max Amplitude < 0.5Hz
Delta	6534	6041
Theta	117	0
Alpha	283	0
Beta	548	0

Table 5.2: Summary of IMFs in NE for different frequency bands in NREM ZW

Looking at 5.1 and 5.2, there are a clear majority of IMFs placed in the delta band, for both DE and NE. The amount of IMFs placed in delta band with a frequency below 0.5 Hz are illustrated, of where 91% of IMFs for DE and 92% for NE. The frequency of the IMFs are decided calculating the frequency of the maximum amplitude in the PSD of the corresponding IMF.

5.1.3 Classification Using Channel Subsets

Figure 5.6 shows the classification performances of applying PCA, CSP and KNN, and EEGNeX with permutation-based channel selection, on raw or MEMD-filtered data. For each of the methods, both a

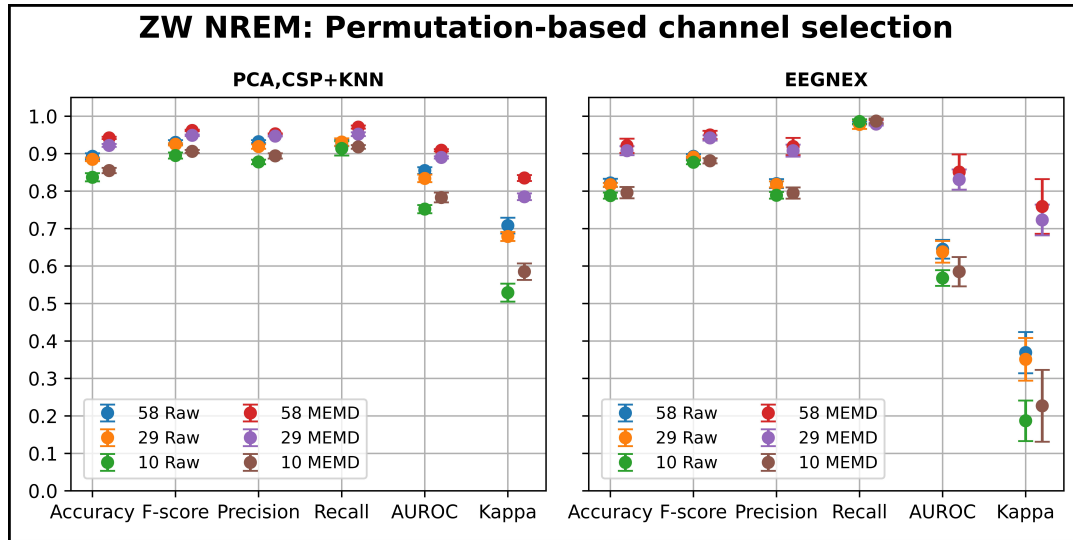


Figure 5.6: Classification performances for DE vs NE, with raw or MEMD-filtered data, using PCA, CSP and KNN (left) and EEGNeX (right). Permutation-based channel selection is used for choosing the channel subsets.

channel subset of 29 and 10 are shown. The results of 58 channels are shown for comparison reasons. Overall, using permutation-based channel selection, the channel subsets when using PCA, CSP and KNN achieves the highest performance with 0.92 accuracy, 0.95 F-score, 0.95 precision, 0.95 recall, 0.89 AUROC and 0.79 kappa when using 29 channels. For 10 channels the metric scores are 0.86 in accuracy, 0.91 F-score, 0.89 precision, 0.92 recall, 0.78 AUROC and 0.59 in kappa.

Looking at the scores when applying EEGNeX, the values show higher standard deviation when using MEMD-filtered data than raw data.

In general for both figures in fig. 5.6, the AUROC and kappa value drop more than the other metrics, when the number of channels in the subset is decreased, both for the raw and the MEMD-filtered data. As AUROC and kappa are metrics dependent on class balance, the wide range of the metric scores could be explained by this. However, they decrease less for the MEMD-filtered data, meaning that the values are not solely dependent on the balance of the data, but also on the quality of the data. Nonetheless, while data quality plays a significant role, achieving a balanced dataset is essential for reaching its maximum potential. This implies that while enhancing data quality may lead to an increase in metric scores, a balanced dataset remains important for attaining the optimal value.

Figure 5.7 show the classification performances of using PSD for feature extraction and XGB as classifier, with all 58 channels and the subsets of channels chosen through the PSD ratio-based analysis. The XGB classifier is shown based on the results obtained from fig. 5.2. Subsets of 29 and 10 are shown giving 0.9 accuracy, 0.94 F-score, 0.9 precision, 0.99 recall, 0.79 AUROC and 0.67 kappa for 29 channels, and 0.86 accuracy, 0.92 F-score, 0.86 precision, 0.97 recall, 0.7 AUROC and 0.49 kappa for 10 channels. Also, here, as observed in fig. 5.6, the AUROC and kappa metrics decreases more than the other metrics when the channel subsets are reduced. We also see that the 10 best channels yield higher performance than 10 random channels.

Overall, comparing the channel selection methods, the technique of applying PCA, CSP and KNN

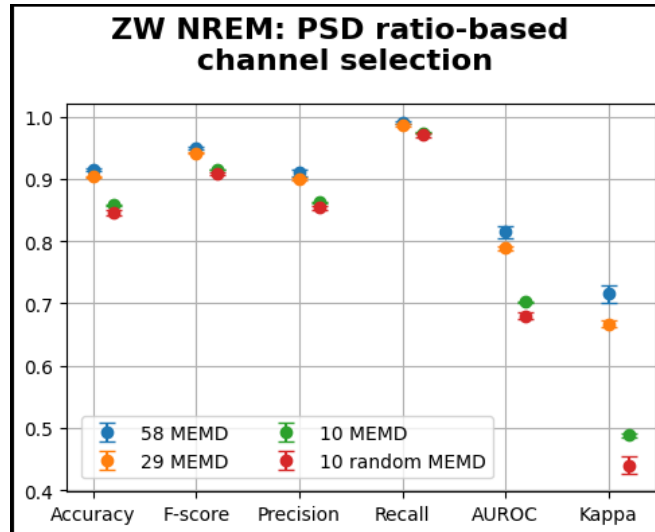


Figure 5.7: Classification performances with MEMD-filtered data, using PSD and XGB. The channels are chosen by the highest average PSD ratio (PSD of DE divided by PSD of NE for each channel), across subjects.

with permutation-based channel selection gets the highest performance.

5.2 Classification Using the ZW Dataset During REM Sleep

In this section, the classification performances of DE vs NE with REM recordings in the ZW dataset are shown.

Our first approach for classifying DE vs NE in REM sleep was done using 1) PCA, CSP, and ML, 2) PSD and ML, and 3) EEGNeX. The performances are presented using 58 channels with raw or MEMD-filtered data. Also here, the low-passed filter data is referred to as the raw data.

Furthermore, to identify the important channels used for the classification, topographic maps based on permutation-based channel selection were created.

The topographic maps created based on a PSD-based quantitative analysis of DE and NE during REM sleep are also presented.

The channel subsets extracted from the permutation-based channel selection and the PSD-based analysis are further used in classification tasks for DE vs NE.

5.2.1 Classification Using 58 Channels

The classification performances of using the feature extraction techniques, PCA, CSP and PSD, in combination with different with ML algorithms, with raw or MEMD-filtered are shown in fig. 5.8 and fig. 5.9.

The highest performances with an accuracy of 0.998, F-score 0.998, precision 0.998, recall 0.998, AUROC 0.998 and kappa 0.995, are achieved using MEMD-filtered data with PCA and CSP for feature extraction, along with the KNN classifier. However, performance metrics using PSD in combination with KNN on MEMD-filtered data (0.97 accuracy, 0.96 AUROC and 0.92 kappa) and EEGNeX on raw

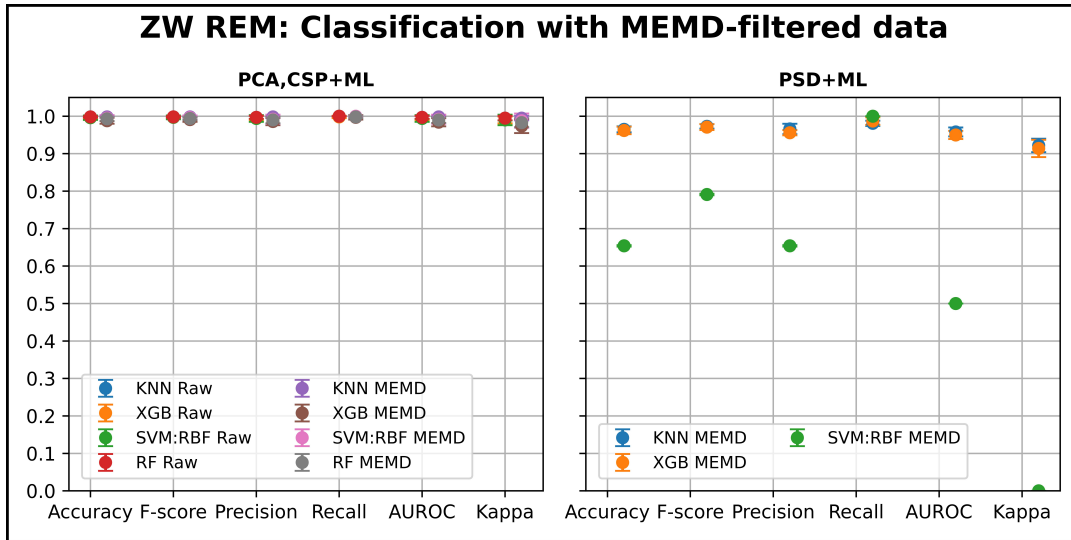


Figure 5.8: Classification performances with raw or MEMD-filtered data, using PCA, CSP and different ML algorithms (left) and PSD and different ML algorithms (right).

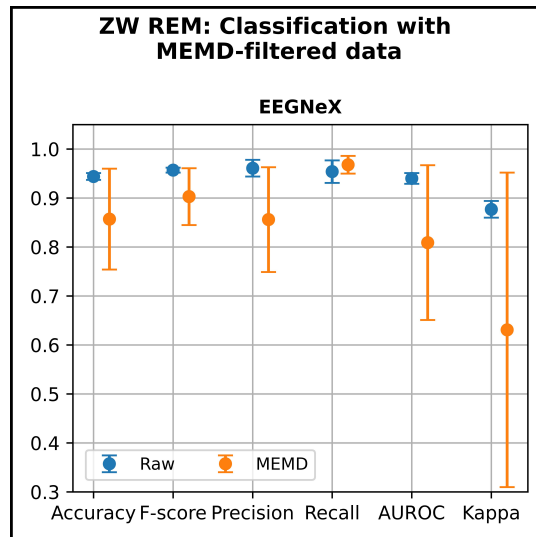


Figure 5.9: classification performances with raw or MEMD-filtered data, using EEGNeX.

data (0.86 accuracy, 0.81 AUROC and 0.63 kappa) also yield high results. This suggests that dreams during REM sleep in this dataset are prominent and easily detectable.

In the case of REM sleep, applying MEMD-filtered data instead of raw data do not increase the results significantly, as it did for NREM sleep (fig. 5.2, fig. 5.3). For PCA, CSP and KNN there are no differences for accuracy, F-score and kappa, and only 0.001 increase in precision and AUROC, while recall decreases with 0.002. For EEGNeX the metric scores decrease significantly with 0.087 accuracy, 0.054 F-score, 0.105 precision, 0.131 AUROC and 0.245 kappa. For recall, it increases with 0.014. Also, it is a notable increase in the standard deviation of EEGNeX when applying MEMD-filtered data, compared to raw data. This indicates that EEGNeX is less robust compared to the classical ML algorithms.

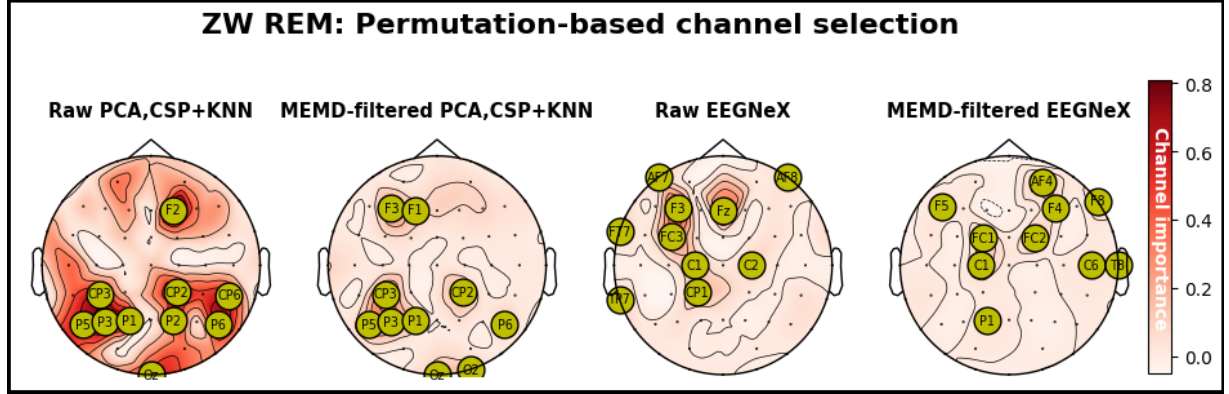


Figure 5.10: Topographic map showing the channel importances using raw or MEMD-filtered data, extracted using permutation-based channel selection. The 10 most important channels are indicated with a yellow circle.

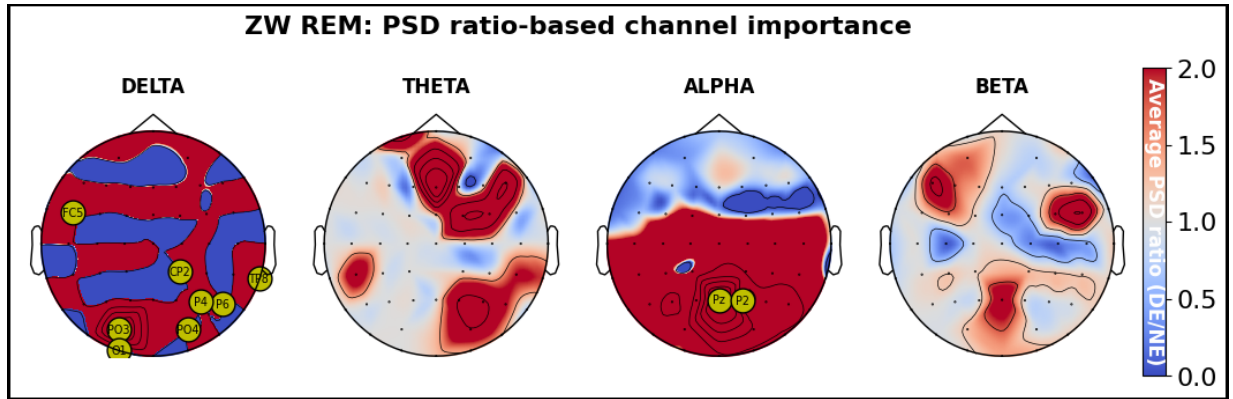


Figure 5.11: Topographic map showing average PSD ratios (DE/NE) using MEMD-filtered data.

5.2.2 Channel Importances and PSD Analysis

From the topographic maps derived from permutation-based channel selection, fig. 5.10, we observe significance in the parietal, central, and frontal regions for both PCA, CSP and KNN, and EEGNeX. However, for PCA, CSP and KNN there are also some channels in the occipital, and the majority of channels lay on the parietal, while for EEGNeX, the channels are mainly located in the frontal and central. The highest classification scores occur when applying MEMD-filtered data to PCA, CSP, and KNN, establishing the reliability of the corresponding map. However, the metric scores of the other experiments also prove to be high, indicating that, based on permutation-based channel selection, important channels are distributed throughout the brain, all contributing to the identification of dreams during REM for this dataset. This diversity is also reflected in the performance delta. Specifically, the maximum performance deltas using raw data (0.8 for PCA, CSP, and KNN, 0.38 for EEGNeX) are higher than when using MEMD-filtered data (0.3 for PCA, CSP, and KNN, 0.1 for EEGNeX). This implies that the removal of channels has a smaller impact on classification scores when using MEMD-filtered data. As the performance using MEMD-filtered data is more representative, we could explain the decreased importance of each channel by the fact that dream activity is distributed throughout the entire brain. Thus, the majority of channels effectively capture the dream activity, making the removal of any single channel less consequential.

Looking at the plots depicting the average power ratio of the different bands in REM sleep, fig. 5.11, we see a lot of activity all over the brain for the delta band, with an increased area of DE activity in the parieto-occipital. For theta, the frontal and parietal regions are standing out. For the alpha band, we see increased activity over the central, parietal, occipital and some in the right frontal regions for DE. The frontal and parietal areas are highlighted for DE in beta band, also showing some areas in the central for increased NE activity.

The 10 channels with the highest PSD ratio are located in the delta and alpha band. In delta they primarily lie in the parietal and parieto-occipital regions, while a few are located in the border of frontal and central, border of central and parietal and the temporal regions. In alpha, they are located in the parietal. This implies that these areas have the highest PSD ratio, and since the majority of them are located on red areas, it suggests that these are the areas with most increased DE activity compared to NE during REM sleep.

Figure 5.26 shows the PSD of DE activity for each frequency band, while fig. 5.27 shows the PSD activity of NE. Both topographic maps are created using MEMD-filtered data. Overall, we can see that the values of DE are in general higher and more prominent than NE.

In order to investigate the high values of ratio in for instance delta band (60 000), the individual maps of DE and NE are shown. The PSD ratio-based topographic map in fig. 5.11 could give the wrong picture of the activity by showing high activity of DE, when in reality the high values are not caused by high values of DE, but low values of NE, close to zero. This is prominent for channel PO3 in delta band, showing a value of 60000, but from the individual map in fig. 5.27, we see that channel PO3 is almost zero for NE, causing channel PO3 to be more prominent in the PSD ratio map, instead of channel TP8 which has a higher PSD of DE. From fig. 5.11, channel PO3 may therefore serve as an important channel for the activity, when in reality, channel TP8 has higher activity of DE. This could indicate that the individual maps are a better representation of the frequency map than the power ratio.

Looking at the topographic maps of DE and NE, it is more activity of DE than NE in delta. The DE activity is similar to the PSD ratio-based topographic map of delta, covering the frontal, central, occipital and temporal regions. However, it is only a small area in the central area with increased NE activity.

For theta, the DE activity is prominent in the frontal, in the border to the central and the parietal regions. While for NE, there are only a few spots, in the frontal and central.

For alpha, the map of DE resembles the patterns as in the PSD ratio-based map in fig. 5.11, showing high activity of DE in the parietal, central and frontal areas. However, observing high activity of NE in the frontal area also explains the absence of increased DE activity in fig. 5.11 for alpha.

For beta, it is mainly one channel, for both DE and NE, dominating the activity, shadowing the rest of the activity.

The topographic map of PSD ratio after removing the first IMF, before constructing the MEMD-filtered signal with the three closest IMFs are shown in fig. 5.12. Removing the first IMF, we see that it is less prominent activity in beta and alpha, compared to fig. 5.11. This indicates that some of the first IMFs which were removed, belonged to the beta and alpha band.

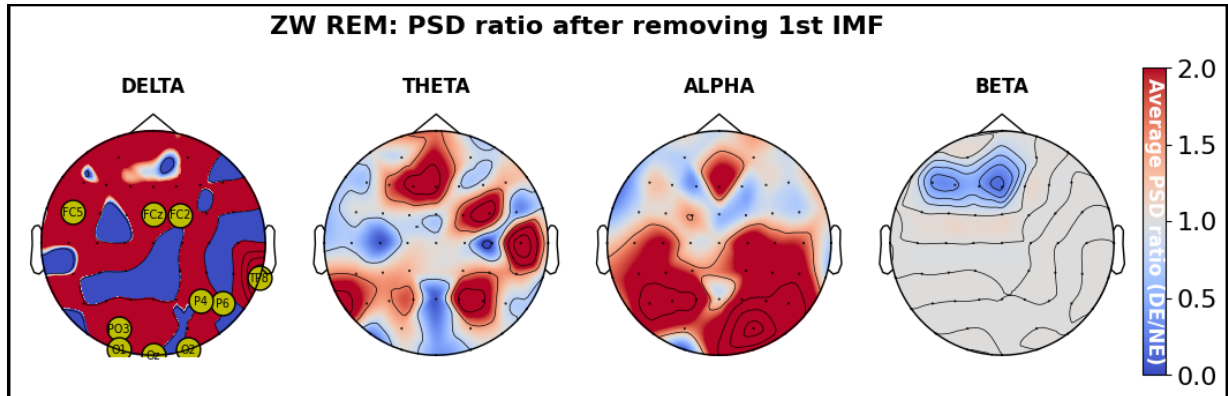


Figure 5.12: Average PSD ratio (DE/NE) with common scale for all frequency bands, after removing the first IMF before constructing the MEMD-filtered signal.

Otherwise, looking at fig. 5.12 there are increased activity of DE across the whole brain, as for fig. 5.11. However, in delta band, the DE activity is now more prominent in the frontal and occipital regions. This is also visible through the 10 channels chosen based on the PSD ratios, as some of them have moved from being located in the parietal to being placed in the border of the frontal and central, and in the occipital. The patterns of increased DE activity also resembles the individual activity of DE in section 5.3.3. Without removing the first IMF, the theta activity in the front is rather bigger, while when removing the first IMF, the activity of DE compared to NE is more distributed and covers the frontal, parietal and parieto-occipital regions. For alpha, it is still increased activity of DE in the parietal, central and occipital, but we see that the small area in the frontal is now more increased and prominent after removing the first IMF. For beta, it is only prominent activity of NE compared to DE, while all the activity of DE that was visible without removing the first IMF is now vanished.

Frequency Band	No. of IMFs Placed in Band	No. of IMFs with Freq of Max Amplitude < 0.5Hz
Delta	2356	1827
Theta	89	0
Alpha	202	0
Beta	311	0

Table 5.3: Summary of IMFs in DE for different frequency bands in REM ZW

Frequency Band	No. of IMFs Placed in Band	No. of IMFs with Freq of Max Amplitude < 0.5Hz
Delta	1111	1042
Theta	28	0
Alpha	114	0
Beta	313	0

Table 5.4: Summary of IMFs in NE for different frequency bands in REM ZW

The 5.3 and 5.4 show the distribution of IMFs placed in the respective frequency bands. The frequency of the IMFs are determined based on the frequency of the maximum amplitude in the PSD of the corresponding IMF. For both DE and NE, a big proportion of the IMFs placed in delta have a frequency of less than 0.5 Hz; 78% of the IMFs placed in delta for DE and 94% for NE.

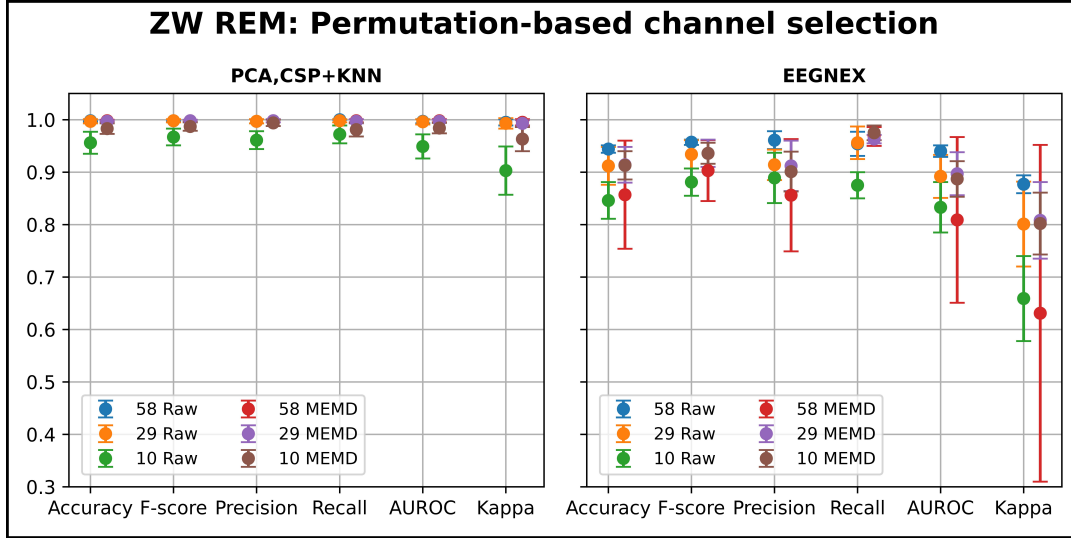


Figure 5.13: Classification performances with raw or MEMD-filtered data, using PCA, CSP and KNN (left) and EEGNeX (right). Permutation-based channel selection is used for choosing the channel subsets.

5.2.3 Classification Using Channel Subsets

Figure 5.13 shows the classification performances of using channel subsets of 29 and 10 channels, chosen from permutation-based channel selection. The methods use PCA, CSP and KNN, since KNN achieved highest performance in fig. 5.8, or EEGNeX on raw or MEMD-filtered. The best scores when using 29 and channels are obtained using PCA, CSP and KNN on MEMD-filtered data with 0.997 accuracy, recall and AUROC, 0.998 F-score and precision, and 0.993 kappa for 29 channels. For 10 channels, the best scores are 0.98 accuracy, 0.99 F-score, precision, 0.98 recall, AUROC and 0.96 kappa, showing a maximum decrease of only 0.03 from 58 to 10 channels in the metric values. We can also see that the difference in performance increases when using MEMD-filtered data compared to raw data. Therefore, the effect of MEMD-filtered data becomes prominent when decreasing the number of channels.

For EEGNeX on MEMD-filtered data, the performances are also high with 0.91 accuracy, 0.94 F-score, 0.91 precision, 0.96 recall, 0.9 AUROC and 0.81 kappa for 29 channels and less than 0.01 decrease for all metrics when using 10 channels. Recall even increases with 0.01 when reducing the channel subset from 29 to 10. Using 58 channels, the metric scores are lower, for all metrics, compared to 10 channels. We can also see that the standard deviation decreases significantly when reducing the channel subset.

Overall, the classification performances for channel subsets of 29 and 10 channels are highest when using permutation-based channel selection on PCA, CSP and KNN.

Using the 10 channels obtained from the PSD-based analysis in classifying DE vs NE with PSD as a feature and KNN classifier yields lower values than when using the permutation-based channel selection on PCA, CSP and KNN on MEMD-filtered data. The results also show that the 10 best channels based on the PSD analysis give lower performance than when using 10 random channels, meaning that the channel subset is not optimal.

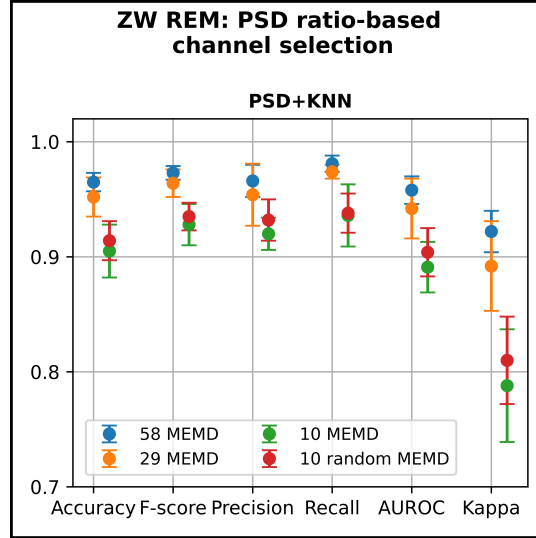


Figure 5.14: Classification performances with MEMD-filtered data, using PSD and KNN. the channels are chosen by the highest average PSD ratio (PSD of DE divided by PSD of NE for each channel), across subjects.

5.3 Classification Using the Tononi Dataset

The Tononi dataset consists of high-density EEG-recordings with 256 channels, collected from sleep stage N2. We present classification performances for all 256 EEG channels and channel subsets, together with topographic maps showing PSD ratios and channel importances calculated with permutation-based channel selection. The results are shown for DE and NE classification.

Three classification pipelines are used with all channels: 1) PCA, CSP, and ML algorithms, 2) DL with EEGNeX, and 3) PSD and ML algorithms. Pipeline 1) and 3) have been performed using both raw and MEMD-filtered data, whereas 2) is performed only on raw data. MEMD has been applied on 2-second-long input segments for the classification tasks, unless stated otherwise.

5.3.1 Classification Using 256 Channels

Figure 5.15 shows the classification performances using PCA, CSP, and the four ML algorithms KNN, XGB, SVM, and RF with both raw and MEMD-filtered data. The results show that with PCA and CSP as feature extraction the performance is high across all metrics. For raw data the best classifier is KNN, obtaining 0.978 accuracy, 0.978 F-score, 0.978 precision, 0.979 recall, 0.978 AUROC, and 0.956 kappa - indicating a near perfect performance. XGB and SVM perform equally, with the performance standard deviation overlapping the other across all metrics. RF scores lower across most metrics, but still achieves a high performance.

Applying the MEMD-filter lifts the performance for SVM, RF, and XGB, but also increases the standard deviation for all classifiers. The classifier with the highest accuracy, F-score, AUROC, and kappa is SVM, with the mean of KNN being within the standard deviation of SVM for these for metrics. With SVM we obtain 0.979 accuracy, 0.980 F-score, 0.976 precision, 0.984 recall, 0.978 AUROC, and 0.957 kappa. This performance is nearly equivalent to that of PCA, CSP, and KNN on raw data, with both similar mean performance and standard deviation.

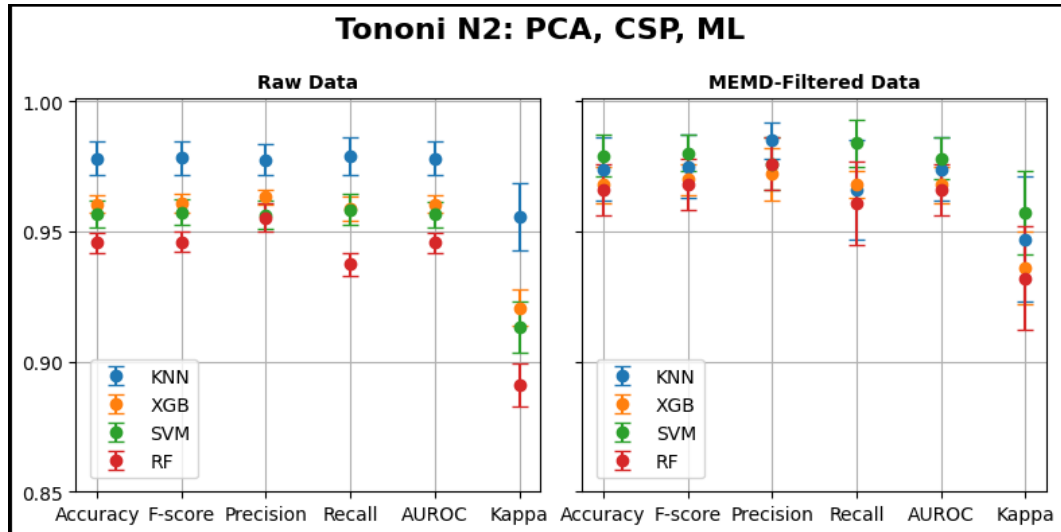


Figure 5.15: Classification performance for DE vs NE with raw or MEMD-filtered 2-second-long input segments into PCA, CSP, and ML.

The role of PCA as a part of the feature extraction between MEMD-filtering and the CSP was tested and the results are shown in fig. 5.16. Here we perform classification with CSP and ML algorithms on MEMD-filtered data. The highest performance is made with SVM, as before, achieving 0.978 accuracy, 0.979 F-score, 0.977 precision, 0.982 recall, 0.977 AUROC, and 0.955 kappa. This performance is almost identical to the case of adding PCA in the feature extraction, meaning that PCA neither increases nor reduces the performance. However, PCA was found to be useful during experimenting with channel subsets, where it solved a problem of data rank deficiency.

The results from the PCA, CSP and ML pipeline show that the classification task can be achieved with high performance across multiple classifiers. The CSP feature extraction separates the classes in a way that allows the classifiers to differentiate between DE and NE, using both raw and MEMD-filtered data.

In fig. 5.17 we see the classification performances for EEGNeX. With this technique we obtain 0.807 accuracy, 0.826 F-score, 0.784 precision, 0.878 recall, 0.804 AUROC, and 0.612 kappa. The results indicate a classification performance well above chance, but with room for improvement.

The classification performances using PSD as the feature are presented in fig. 5.18, from raw and MEMD-filtered data. KNN, XGB, SVM, and RF are used as classifiers. The performance of the classifiers vary from nearly perfect to nearly random, with XGB as the best performer for both raw and MEMD-filtered data. For raw data XGB obtains 0.974 accuracy, 0.975 F-score, 0.974 AUROC, and 0.949 kappa, and for MEMD-filtered data 0.993 accuracy, 0.994 F-score, 0.993 AUROC, and 0.987 kappa. Using PSD and XGB proves to be a powerful pipeline, but the difference in performance shows that the choice of classifier is important when using PSD as the feature.

The MEMD-filtering improves the results for all classifiers. For XGB kappa increases with 0.038, for KNN with 0.126, for RF with 0.16, and for SVM with 0.049. Although the performance increase is lowest for XGB, this performance was already only 0.053 in kappa from a perfect score.

Class DEWR was included in either NE or DE to investigate if classification could indicate which of

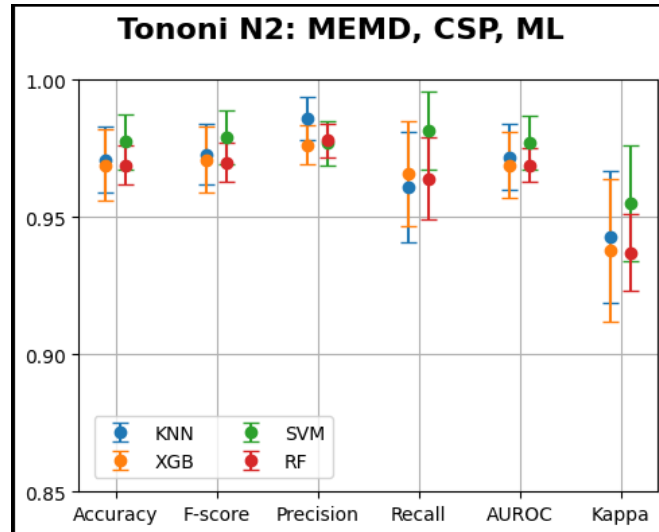


Figure 5.16: Classification performance for DE vs NE with MEMD-filtered 2-second-long input segments into CSP, and ML.

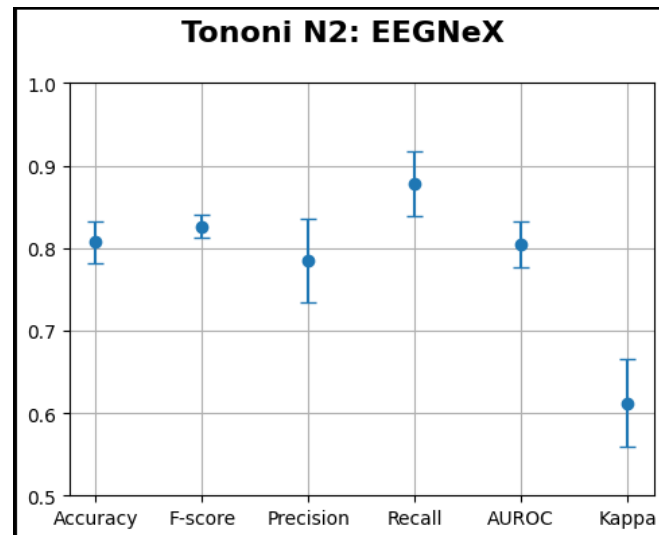


Figure 5.17: Classification performance for DE vs NE with raw data using EEGNeX.

the classes DEWR is more similar to. The results are seen in fig. 5.19, where for PCA and CSP, and PSD only the top performing classifier is chosen. The results for all classifiers are seen in appendix A. For PCA and CSP this is KNN, whereas for PSD this is XGB. For EEGNeX the performance is higher for AUROC and kappa when combining DEWR with NE, achieving scores of 0.808 AUROC and 0.627 kappa - compared to 0.787 AUROC and 0.576 kappa when DEWR is included with DE. For PCA and CSP, and PSD the classification scores are higher in general, but also more similar. Out of accuracy, F-score, AUROC, and kappa, F-score is the only metric where the results for both class combination are not within the standard deviation of each other. F-score is higher for DEWR combined with DE for both combinations; For PCA and CSP the F-score is 0.976 versus 0.958, and for PSD 0.980 versus 0.970. The results show that a high performance is feasible for both class combinations.

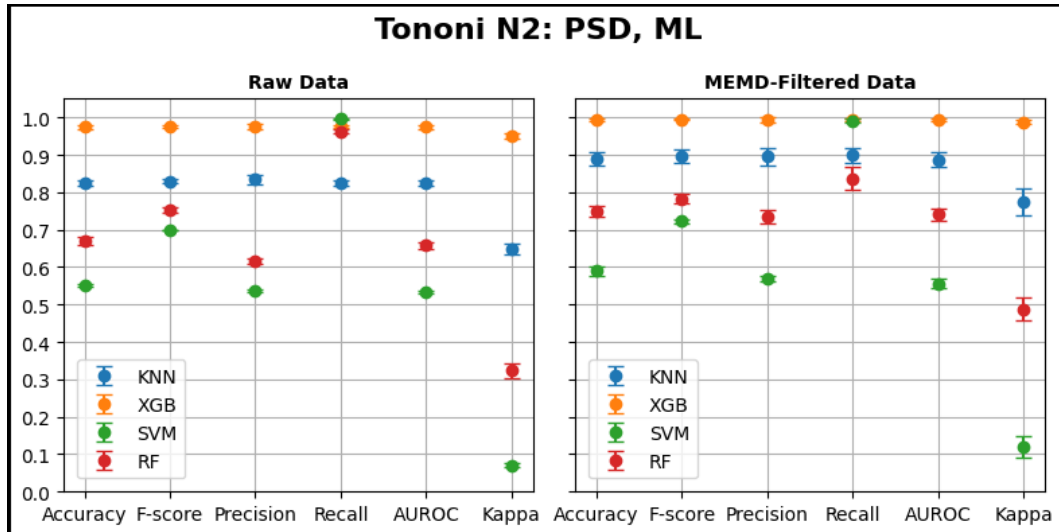


Figure 5.18: Classification performance for DE vs NE with raw or MEMD-filtered 2-second-long input segments into PSD and ML.

5.3.2 Channel Importances and PSD Analysis

In fig. 5.20 we present three topographic maps of channel importances, found using permutation-based channel selection with three different base models: 1) PCA, CSP, and XGB, 2) EEGNeX, and 3) MEMD, PSD, and XGB. The color scale, shown on the right in the figure, ranges from plus/minus maximum importance (0.044) across all models. The ten most important channels are depicted by the yellow dots in each figure.

Using PCA, CSP, and XGB we obtain the highest importances in the temporal areas of the brain, both left and right. The central, frontal, and occipital regions are rated as unimportant. The maximum and minimum importance are respectively 0.044 and -0.01.

With EEGNeX the topographic map is more scattered, with the most important channels appearing in the left frontal, central, and one in the right occipital area. The maximum and minimum importance are less than for PCA, CSP, and XGB - with values ranging between 0.012 and -0.015. There does not appear to be any particular region which stands out as more important, as the red and blue areas in the figure are scattered around each other.

With MEMD, PSD, and XGB the importance of each channel is very low. Three channels have an importance of 0.01, eight channels have an importance of 0.005, and two channels have a negative importance of -0.005. The important channels are mostly located in the central region, apart from three channels located in the central and right frontal region.

Figure 5.21 is created by MEMD-filtering 2-second-long segments, calculating the PSD for each IMF, and finding the grand average PSD value for each IMf in each frequency band. For each class we then have a single average PSD value for each frequency band, from where the ratio is found by dividing the PSD for class DE by the PSD for class NE. The ten channels with the highest PSD ratio are marked with a yellow dot in the frequency band from where it is found. The scale, located on the right in the figure, ranges from ratio 0 to 2 - however, multiple channels have much higher ratios. This range is chosen to evaluate when the PSD of DE is higher than NE and vice versa.

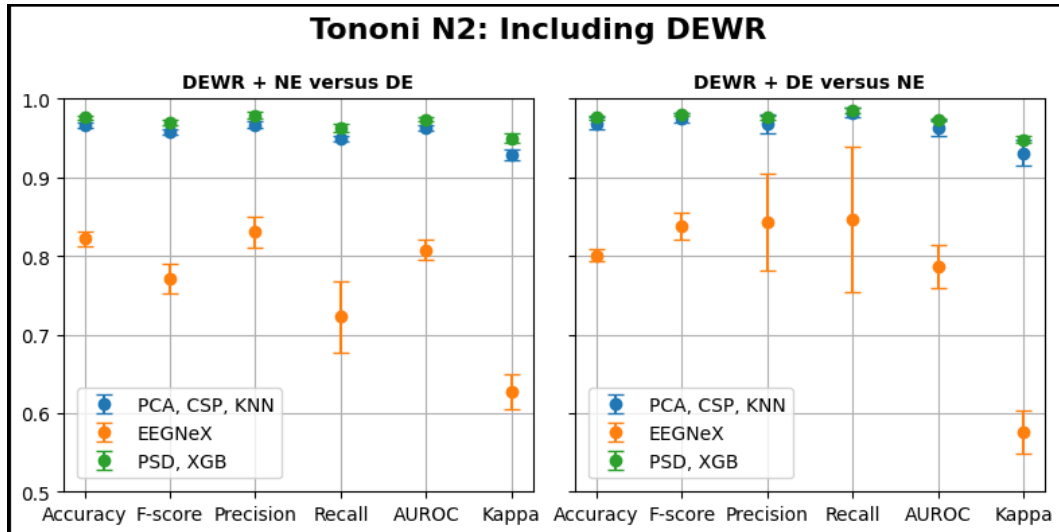


Figure 5.19: Classification performance after including DEWR in either NE or DE on raw data. Different feature extraction techniques in combination with machine learning algorithms and the EEGNeX model are used.

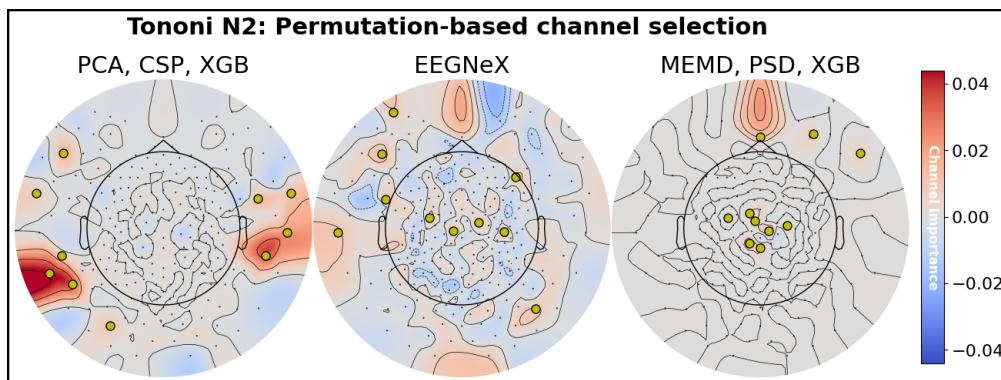


Figure 5.20: Permutation-based channel importance with raw or MEMD-filtered data from three base models. The yellow dots show the ten most important channels for each base model.

The topographic maps of the PSD ratios do not appear to depict any region in particular where the difference between DE and NE is clear. Instead, the ratios appear to jump from high to low from one channel to another. The beta band is an exception, where the PSD of DE is higher than for NE across most of the brain.

Of the ten channels with the highest PSD ratio, six are found in the delta band, two in the theta band, and two in the alpha band. Several of these channels are located on local peaks with much lower ratios around, which does not indicate that the region around the channel is important in itself. Eight of the ten channels are located in the central region, however, which could highlight this area as an important area when it comes to separating DE and NE.

In fig. 5.22 we have used 73-second-long input segments to the MEMD-filter before calculating the PSD ratio between the classes, coinciding with the data length in ZW NREM. The topographic map shows that the PSD ratios go from high to low from one channel to the other, which does not indicate any special region of importance. The left hemisphere in the alpha band shows a more unified low ratio, as well as multiple areas in the beta band.

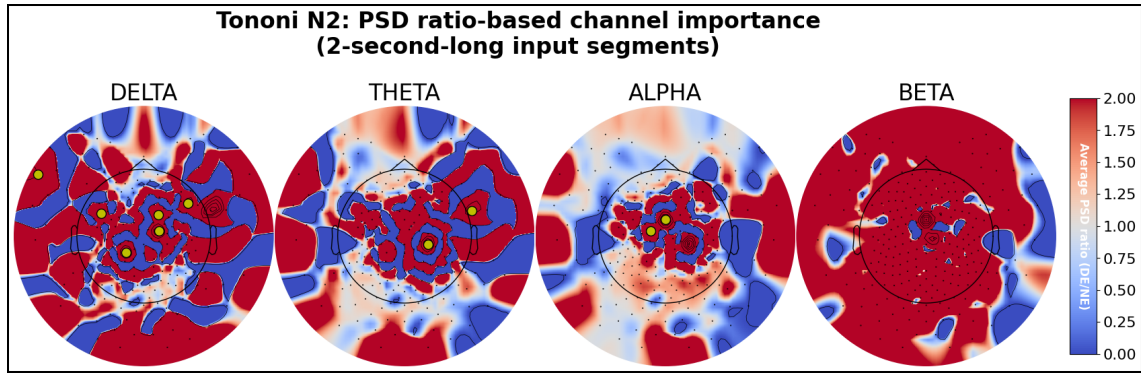


Figure 5.21: PSD ratio with MEMD-filtered 2-second-long segments. The yellow dots show the ten channels with highest ratio, divided into the corresponding frequency band where the ratio is highest.

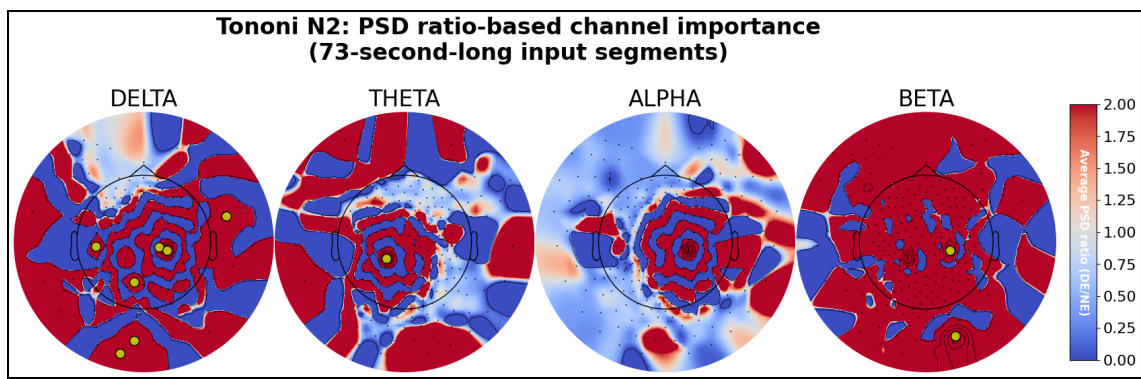


Figure 5.22: PSD ratio with MEMD-filtered 73-second-long segments. The yellow dots show the ten channels with highest ratio, divided into the corresponding frequency band where the ratio is highest.

This topographic map differs from the previous, where 2-second-long input segments were used, by visual inspection. As the previous, this topographic map also does not depict any clear regions of importance in the bands. For both maps the highest PSD ratios are mostly located in the delta band, with one single channel in the theta band and two channels in the beta band.

5.3.3 Classification Using Channel Subsets

In fig. 5.23 we present the results on four channel subsets, consisting of 50, 30, 20, and 10 channels, using EEGNeX and PCA, CSP, and XGB for classification. XGB is chosen from the ML classifiers as this showed the highest score for the 30, 20, and 10 channel subsets. The performances from all classifiers are seen in appendix A. The channel subsets are chosen by permutation-based channel selection, as shown in fig. 5.20.

The performance on channel subsets using PCA, CSP, and XGB is in general high. For 50, 30, and 20 channel subsets the AUROC is respectively 0.925, 0.907, and 0.891 and kappa 0.851, 0.813, and 0.782. Reducing the number of channels does reduce the performance, albeit only slightly. With 10 channels, the kappa drops to 0.484. The same performance decrease is seen in the other metrics as well.

With EEGNeX the classification results are very similar for all three subsets, with the 50 channel subset

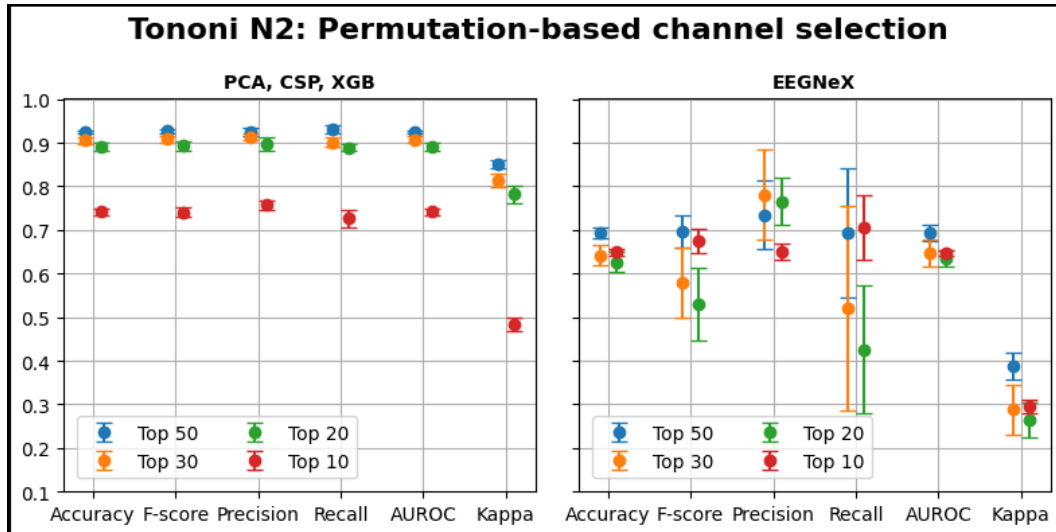


Figure 5.23: Classification performance subsets of 50, 30, 20, and 10 channels for DE vs NE on raw data. Channels selected by permutation-based channel selection.

achieving a slightly higher performance across accuracy, AUROC and kappa. The accuracy, AUROC and kappa for the 30, 20, and 10 channel subsets are within the standard deviation of each other, achieving respectively 0.640 AUROC and 0.288 kappa, 0.634 AUROC and 0.264 kappa, and 0.647 AUROC and 0.295 kappa. In general the classification performance is low for all subsets.

As when using all channels, the performance on channel subsets is higher and with less standard deviation when classifying with PCA, CSP, and XGB compared to EEGNeX. For both classifications the performance decreases compared to using all channels.

The results in fig. 5.24 show the classification performance using MEMD, PSD and XGB on four channel subsets. MEMD filtering is made using all available channels, before picking the channel subset. The top 29 and 10 channels are chosen from the PSD ratio, shown in fig. 5.22, as the channels with the highest ratio. The 29 and 10 random channels are shown for comparison reasons, to evaluate the impact of using PSD ratio for channel selection.

By selecting the top 29 channels we obtain mean metrics of 0.935 accuracy, 0.923 F-score, 0.93 AUROC, and 0.866 kappa. With the top 10 channels we obtain 0.876 accuracy, 0.854 F-score, 0.87 AUROC, and 0.748 kappa. The performance decreases when the number of channels decreases, but the performance is high in both cases.

The random channel subsets yields a similar performance for 29 channels, but a decrease for 10 channels. The mean values for 29 random channels are virtually identical to those of the top channels, albeit with a lower standard deviation. For the 10 channels the performance decreases to 0.824 accuracy, 0.797 F-score, 0.820 AUROC, and 0.642 kappa.

In fig. 5.25 we have tested variations of MEMD in order to increase the performance on a 10 channel subset. The subset was selected prior to MEMD filtering. The subset is found using permutation-based channel selection with MEMD, PSD, and XGB as the base model, locations shown in fig. 5.20. The recordings were divided into 2-second-long segments before MEMD-filtering.

The baseline performance is with MEMD(1), PSD, and XGB as this method gave the best results when

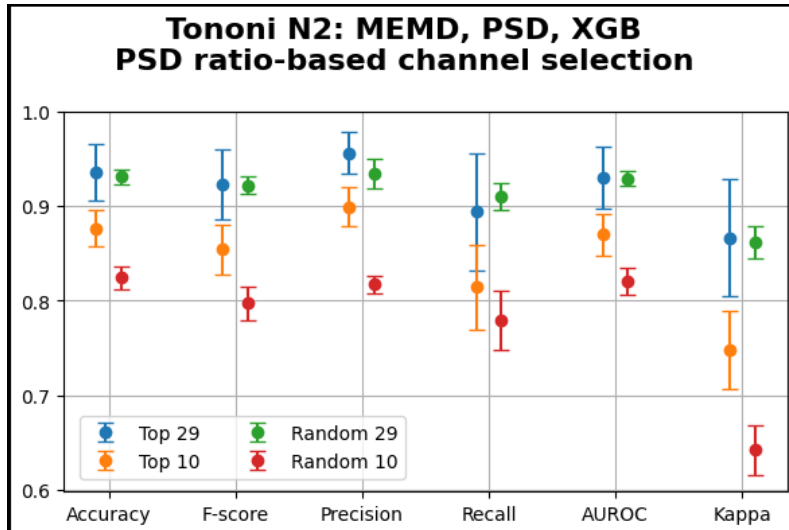


Figure 5.24: Classification performance with MEMD-filtered 73-second-long input segments into PSD and XGB on 29 and 10 channels. The top channels are chosen by the highest average PSD ratio.

using all channels. Here we use the method described in section 4.2, where all IMFs are regarded when choosing the three closest. This gives 0.620 AUROC and 0.241 kappa. For MEMD(2) the method is updated to disregard the first IMF before choosing the three closest, to explore whether this could reduce noise and yield a higher performance. This method yields 0.606 AUROC and 0.212 kappa.

Another alternative method is using PCA and CSP for feature extraction after MEMD-filtering, instead of PSD. With this method we obtain 0.579 AUROC and 0.158 kappa.

The best results come from using NA-MEMD, a modified MEMD-version which uses artificial channels filled with white noise. This gives the highest performance on the subset with 0.711 AUROC and 0.422 kappa, an increase from the baseline method of 0.091 in AUROC and 0.181 in kappa.

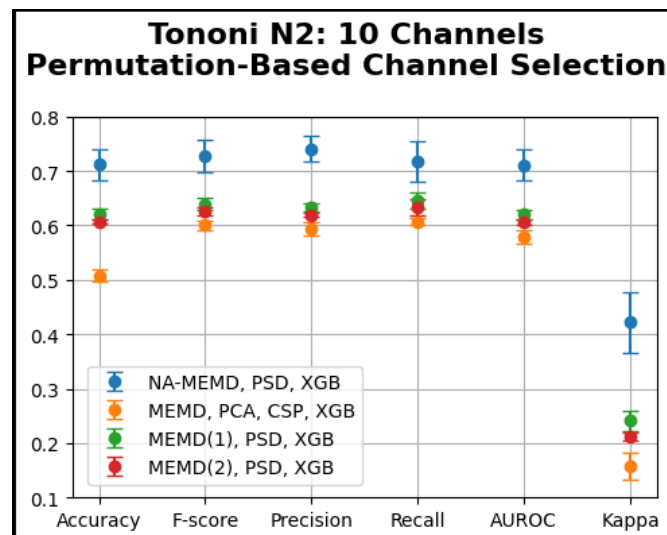


Figure 5.25: Classification performance using MEMD-variations on 10 channels with NREM data from Tononi, using 2-second-long input segments. MEMD(1) refers to choosing the three closest of all IMFs. MEMD(2) refers to disregarding the first IMF when choosing the three closest IMFs.

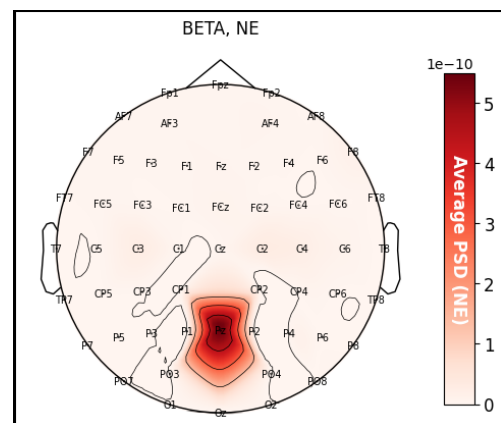
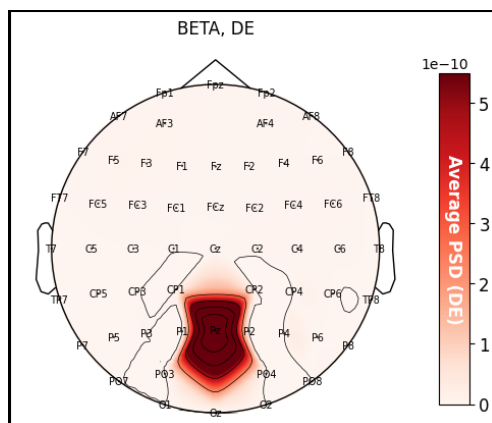
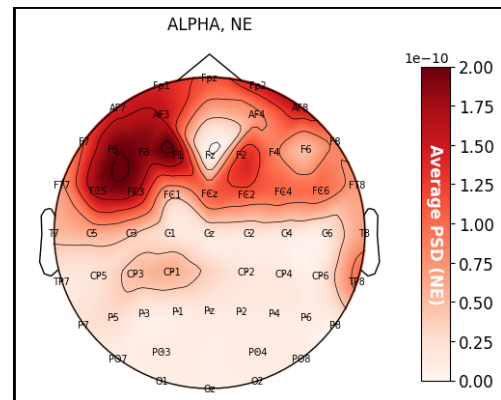
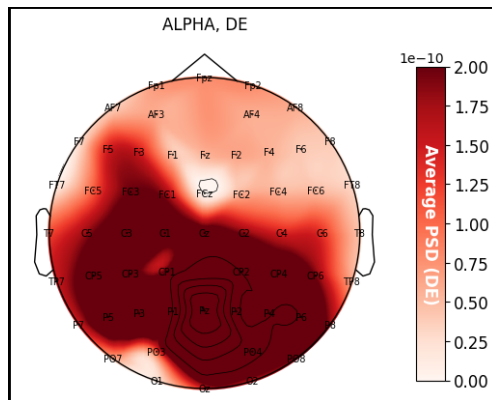
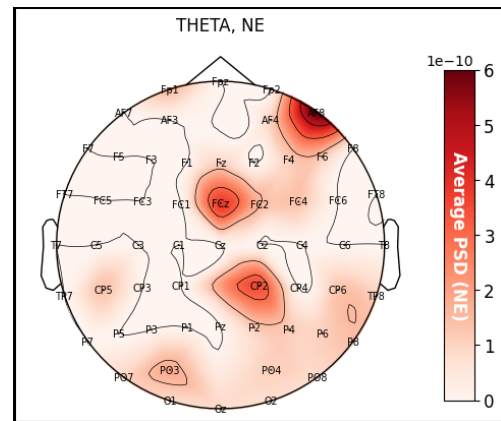
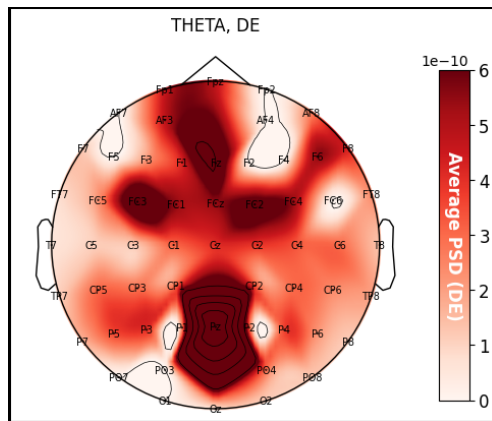
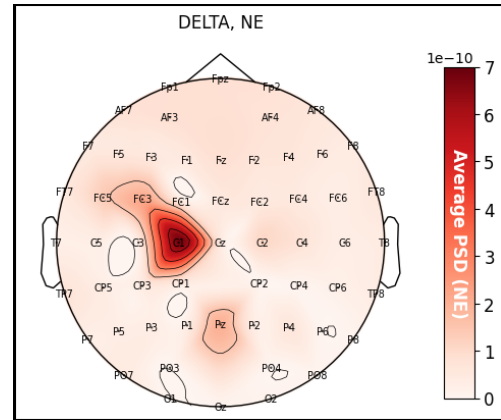
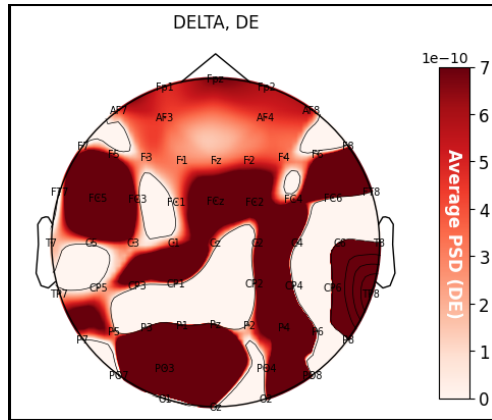


Figure 5.26: Average PSD of DE in REM sleep, for each individual frequency band.

Figure 5.27: Average PSD of NE in REM sleep, for each individual frequency band.

Chapter 6

Discussion

Our work has shown that detecting when a subject is dreaming using EEG signals and ML is a feasible approach. We analyzed two different datasets, in both obtaining promising results for dream recognition during both, REM and NREM.

6.1 Classification Methods

Three different methods for feature extraction are tested: PCA and CSP, PSD, and the CNN structure within EEGNeX. From the results we observe that PCA and CSP in general give higher scores across the classification models, whereas PSD is more model dependent. Both PCA and CSP, and PSD have low standard deviations, indicating the robustness of both methods. For raw data, EEGNeX is outperformed by the other methods in all the datasets.

Both PCA and CSP, and PSD for feature extraction yield very good results. CSP is a commonly used tool for EEG classification tasks, and the high performance obtained here coincides with previous research [108, 109, 112]. PSD is most powerful when combined with XGB or KNN, XGB especially for the high-density EEG data from Tononi. PSD is a simpler feature than PCA and CSP, which involves more complex computations. This is an advantage for potential BCI applications, where computational resources are limited. The solid performance of PSD as a feature extraction method indicates that DE and NE have clear differences in PSD representation. This is consistent with the literature review, where the distinction between DE and NE has been observed using PSD representations [82, 84–86].

6.2 The Effect of MEMD Filtering

The use of MEMD-filtered data has in many experiments increased the performance significantly and proven to be beneficial for the identification of DE in NREM and REM for both datasets, when using ML algorithms and EEGNeX.

The enhancement in performance using MEMD-filtered data could be due to the alignment capability of MEMD, which effectively identifies important signals and removes unnecessary frequency components in an adaptive manner. This is evident in the selection of three IMFs to represent the original signal, as many of the classification tasks using MEMD-filtered signals either outperform or

is as good as experiments based on raw data. These findings indicate that the MEMD algorithm effectively removes unnecessary frequency components and creates signals with valuable information for the classification process.

The results of MEMD-filtering also provide valuable insights into the data. Our findings indicate that the AUROC and kappa metrics often exhibit lower values compared to other metrics. This has previously been explained by the imbalance in the datasets. While metrics such as kappa is dependent on a balanced dataset to achieve optimal performance, the improvement observed when using MEMD-filtered data suggests that the quality of the data has also been a limiting factor in achieving higher scores.

However, the use of MEMD-filtered data has not had the same effect for;

1. PCA, CSP and RF using ZW NREM data
2. PCA, CSP in combination with SVM, XGB or RF using ZW REM data
3. EEGNeX using ZW REM data
4. PCA, CSP and KNN using the Tononi dataset
5. Applying MEMD after choosing a channel subset of 10 channels and then using PCA, CSP and XGB, when using the Tononi dataset

The decrease or no improvement in performance for these experiments could be explained by already significantly high performances, meaning that when removing IMFs from the original signal to construct a signal only containing three IMFs, it is removing some of the components important for the classification performances. Also, for the experiments with both high performances and the ones with room for improvement, the decrease in performance when applying MEMD-filtered data could appear because of too few IMFs chosen in the reconstruction process of the filtered signal. Exploring different numbers of IMFs for the reconstruction of the filtered signal may give the effect of removing unwanted components using MEMD, while still keeping the components important for the classification performance.

Furthermore, the use of MEMD-filtered data has given an increase in the standard deviation of the metrics when using EEGNeX, both for NREM and REM data from ZW. Observing this increase for EEGNeX, yet not for the ML methods, may indicate that the feature extraction part of EEGNeX is more dependent on all the IMFs and becomes less robust when removing frequency components and constructing a signal only containing three IMFs. The wide spread of performance also show that the model can not be generalized using MEMD-filtered data. It is also important to consider that the NREM and REM datasets do not have the same conditions, such as the segment size and balance in dataset, giving a potential problem when comparing the datasets.

6.2.1 Data Augmentation With MEMD

Applying data augmentation to overcome the challenge of class imbalance in the NREM dataset of ZW reflects in better results for EEGNeX, but not for the classical ML algorithms, except for the RF classifier. The results suggest that the augmented segments following our procedure may not contribute enough additional valuable information or features for classical ML algorithms or EEGNeX.

Although the performance of EEGNeX improves, it is still lower than the results of PCA, CSP and ML. For some instances, performance even decreased and standard deviation increased, indicating potential loss of important information during the splitting and reconstruction process of the original signals into IMFs. The results indicate that the method applied for data augmentation is not contributing enough, however other methods for data augmentation could help. A sanity check to see the effect of the method applied in our work could be to train the model with balanced classes, using the number of instances as the class with less instances. Then, randomly remove instances and create them artificially, and then look at the performance to see the effect of the method.

Uncertainty regarding potential data leakage during the creation of augmented segments is present. The process involves generating augmented segments from a subset of segments within the original dataset, which raises the possibility that the training set may contain IMFs from the test set. However, it's important to note that the augmented segments in the training data are not direct copies of segments in the test set. Instead, they are components extracted from segments and then reconstructed using components from other segments. While the augmented segments may share some characteristics with segments in the test set, they are not exact replicas. Nonetheless, the presence of shared characteristics could potentially lead to data leakage and affect the integrity of the model evaluation process.

For future work, it would be better to perform the process of data augmentation using the same idea of 5-fold cross validation. Meaning, setting aside a test set, then creating 10 000 augmented samples from the train set, and then evaluate the model with the designated test set. Then, repeat this process five times. In this way, ensuring that all augmented segments originate from the train set and thereby, completely eliminating any concerns regarding data leakage and mitigating the consequences of a poor train-test split.

As the results of MEMD data augmentation have not been proven to increase the metrics significantly, it has not been used further in the experiments.

6.3 Classification, Channel Importances, and PSD Analysis with NREM Data

Both datasets used in this thesis contain NREM data, allowing for comparison between the two. For both datasets we have conducted a PSD analysis, used permutation-based channel selection to find the important channels, and used classification between DE and NE. Here, we discuss the similarities and differences between our findings for the two datasets.

6.3.1 Classification Results

From the results, we observe that we are able to obtain high classification performance with the right pipeline. For the ZW NREM data the highest performance is using MEMD, PCA, CSP and KNN, whereas for the Tononi data MEMD, PSD, and XGB shows the highest performance. The results indicate that we are able to classify DE during NREM sleep for both datasets.

For the Tononi dataset the best classification method is using PSD and XGB with the MEMD-filtered data, achieving a near perfect classification across all metrics (0.993 accuracy, 0.994 F-score, 0.993

AUROC, and 0.987 kappa). The performance without MEMD is also near perfect (0.974 accuracy, 0.975 F-score, 0.974 AUROC, and 0.949 kappa), scoring within the second-best classification method with PCA, CSP, and KNN, indicating that PSD is a very powerful tool for classification. The same performance using PSD is not seen for ZW NREM, where it is outperformed by MEMD, PCA, CSP, and KNN. This could show that PSD is more dependent on having a high number of channels than PCA and CSP.

Channel subsets were defined using both permutation-based channel selection and PSD ratio for both datasets, achieving solid classification results. When using raw data the classifications results are generally higher for the Tononi dataset, with PCA, CSP, and XGB achieving 0.891 AUROC and 0.782 kappa using 20 channels, selected by permutation-based channel selection. This is a higher score than PCA, CSP, and KNN for the ZW NREM data using all 58 channels, which achieved 0.861 AUROC and 0.71 kappa. This could indicate that the raw data from Tononi is better suited for DE classification in NREM sleep.

The classification scores are also higher for the Tononi dataset after applying MEMD, using 73-seconds-long segments and all available channels as input. For both 29 and 10 channels, selected from the PSD ratio and randomly selected, the Tononi dataset achieves a higher score compared to the ZW NREM dataset. The reason could be that the MEMD-filtered data from Tononi contains information from more channels than for ZW NREM, as Tononi contains 256 channels and ZW NREM 58 channels.

For the Tononi dataset we have three available classes: DE, NE, and DEWR. DEWR can be considered a mid-state between DE and NE, as the dream content is not recalled. DEWR was combined with both classes, using all three classification methods, to investigate any change in performance. If the classification proved to be higher for one combination than for the other, this could have indicated which class DEWR is more similar to. Instead, high classification performances are obtained from both combinations - indicating that DEWR fits well with both classes, even when classified against the other. A potential reason for this could be that DEWR contains artifacts from both DE and NE, from which the classifiers focus on the artifacts containing the respective class it is combined with. Further investigation of DEWR is required to learn which class it resembles the most. Using a classifier trained for DE versus NE to classify DEWR could be a revealing approach for further work.

6.3.2 Permutation-Based Channel Selection

If some regions in the brain exhibit greater difference between DE and NE, we hypothesize that the channels located in these regions have a higher importance. The topographic maps of the channel importance for ZW NREM do highlight certain areas important for classifying between DE and NE during NREM sleep. For the Tononi dataset, such regions are not found.

Comparing the two topographic maps we observe that the channel importances for the Tononi dataset are much lower than for ZW NREM. The reason for this could be due to the high channel density in the Tononi data. The electrical signals in the brain propagate through the skull to the scalp, where the EEG measurements are made. Recordings made close to each other on the scalp would measure similar signals, given an equal impedance. The importance of each channel is therefore reduced when the channel density increases.

The ZW NREM topographic maps indicate especially two important regions: The parieto-occipital and the frontal. Although both are present for both the PCA, CSP, and KNN pipeline and EEGNeX, the two models differ in where the most important channels are located. PCA, CSP, and KNN highlight the parietal, whereas EEGNeX highlights the frontal. This indicates that both regions are important areas for DE versus NE classification.

From the Tononi topographic we observe different results. The PCA, CSP, and XGB pipeline indicates the lateral areas as important, whereas EEGNeX and MEMD, PSD, and XGB indicates especially the central area. The parieto-occipital and the frontal region are not highlighted as especially important, contrary to the findings from ZW NREM.

6.3.3 PSD Analysis

Both 2-seconds-long and 73-seconds-long input segments PSD topographic maps have been created for the Tononi dataset. In this discussion we compare the topographic map using 73-seconds-long input segments, which is the same length as used for the ZW NREM topographic map.

The PSD ratio topographic maps for ZW NREM and Tononi show very different characteristics. The ratios appear much smoother for ZW NREM than they do for Tononi, with a majority of the values being within zero and two. In the ZW NREM topographic map we observe patterns across all frequency bands. In the Tononi topographic map there is no clear distinction of brain regions with more activity, as both high- and low-PSD ratios appear to be scattered around the maps - with a notable exception in the beta band and the left hemisphere of the alpha band. The topographic maps are contradictory and together do not present a clear distinction of DE and NE brain activity. Especially the beta band is contradicting, where the ZW NREM map shows a low PSD ratio across most of the brain, whereas the Tononi map shows a generally high PSD ratio.

The method for creating the PSD ratio for ZW NREM and Tononi differ in one aspect. An analysis of the power spectra of all sets of IMFs was performed when creating the ZW NREM topographic map to identify potential outliers with irregularly high amplitudes. These outliers were removed before calculating the grand average PSD of the IMFs. This was not done for the Tononi topographic map. EEG measurements are dependent on two variables: The current produced by brain activity and the impedance between the electrode and the scalp. The current, which represents the brain activity, is what we want to measure using EEG. The impedance is a variable controlled by the experimental setup which varies depending on how well the electrode sits on the scalp. Maintaining a low impedance throughout eight hours of sleep on a sleeping subject is a difficult task, as factors such as the subject moving could alter it. When the impedance increases, so does the measured voltage, resulting in signals with a higher amplitude. As PSD is calculated as the square of the FFT magnitude, a signal with irregularly high amplitude would skew the average towards higher values.

The proposed reason for why the PSD ratios differ from channel to channel for the Tononi dataset is that the impedance in some of the channels are higher than for others due to factors in the experimental setup. This is more probable than a sudden current spike between one channel and another, as the current on the scalp is a result of currents in the brain propagating through the skull. For both datasets, and especially for the Tononi dataset, the channel distribution is very dense and should not show big sudden differences in similar areas. The removal of high amplitude signals could therefore

be an important step for correct assessment of the PSD activity. The topographic map for the ZW NREM dataset is therefore believed to be a better representation of the brain activity.

Previous research using PSD analysis on DE versus NE during NREM sleep is presented in the literature review. Three studies have shown a decrease in delta wave activity to be associated with DE [82, 85, 86], in respectively the parieto-occipital region and over the left hemisphere. Increased activity above 4 Hz and decreased activity below 1 Hz is also reported [84]. For the ZW NREM delta band we observe a general reduction in PSD for DE compared to NE, especially in the parieto-occipital region, the left hemisphere, and the anterior of the frontal lobe - similar to previous research. However, another study did not find differences in the spectral power of EEG [87], a result more similar to the results in the delta, theta and alpha bands for the Tononi dataset.

The increase in theta band activity for DE can be related to the nature of NREM dreams. Although NREM reports are most often thought-like and conceptual, characteristics more common from REM reports can also appear in NREM dreams. An increase in theta band activity is related to both daydreaming and movement in awake subjects [21, 113], which could explain the same increase during NREM dreams. The increase in theta activity is seen in the parieto-occipital and temporal lobes, regions related to, respectively, language, spatial recognition, and sensorimotor planning and vision, hearing, and declarative memory. Especially language and declarative memory could be activated during the daydreaming-like NREM reports, whereas spatial recognition, sensorimotor planning, vision, and hearing could be related to more vivid reports.

The nature of NREM dream reports does not resonate well with activities related to beta brain waves, which are conscious thought, logical thinking, and focus [21]. Still, previous research has found an increase in high-power activity in the parieto-occipital region and a greater brain activation related to DE [82, 86]. This is similar to the results seen for the Tononi data, where the PSD ratio shows an increase in PSD related to DE. Earlier we argued that the PSD ratio topographic map for ZW NREM is a better representation of brain activity. However, this shows a decrease in overall high-power activity related to DE - contradictory to previous research. Further work is necessary to investigate why this is the case.

6.3.4 Combined Analysis

The topographic maps of both channel importance and PSD ratio appear to partially coincide with each other within each dataset. For the Tononi dataset, neither has highlighted clear areas. For the ZW NREM dataset especially the parieto-occipital region is marked as an area of importance, as it shows both an increased activity in the theta band, decreased activity in the delta and beta band, and an increased channel importance. The two topographic maps do not show the exact same results, however. The permutation-based channel selection shows which channels are more important for the classification, and the PSD ratio shows where the difference in activity is. Although the activity is presumed to be connected to the classification, they are not necessarily equal.

6.4 Classification, Channel Importances, and PSD Analysis with REM Data

6.4.1 Classification Performances

The use of high-density EEG data and the methodologies used in this research has been proven to be more effective in detecting dreams during REM compared to previous research [2, 88]. Direct comparison is challenging due to the use of different datasets, especially as the previous studies used a dataset collected with six EEG channels located over the frontal lobe. In the project report an argument for increasing the number of channels was made, especially due to the low number and location of the EEG channels during data collection [2]. The results from using the ZW dataset shows that an increased number of channels does increase the performance.

In our research, the best performance is achieved through feature extraction and classical ML techniques. In the previous study, however, EEGNet outperformed such methods [88]. Using DWT for feature extraction in combination with various ML algorithms (gradient boosting (GB), KNN, RF, etc.) were outperformed by EEGNet. This divergence in results could be explained by differences in the datasets. Having six channels located over the frontal lobe would not allow capturing the same source signals as a uniform distribution over the head using 58 channels. Moreover, noise levels and other factors vary. Consequently, different feature extraction techniques and ML algorithms may yield varying effects depending on the characteristics of the data. Nonetheless, using CSP instead of DWT for feature extraction could potentially enhance performance, regardless of dataset variations, as certain techniques could prove to be powerful, generally for all EEG data.

6.4.2 Permutation-Based Channel Selection

Using permutation-based channel selection, the topographic maps in fig. 5.10 show important channels for the classification of DE versus NE. Given the highest classification performance, the topographic map using MEMD-filtered data with PCA, CSP and KNN is chosen as the most representative topographic map for analysing important channel subsets in ZW NREM.

The topographic map shows important channels spread in many of the regions, but the majority laying in the left part of the parietal region. The observation of importance in the parietal, occipital, frontal, and central regions is also consistent with the reported findings of frequency changes in the parieto-occipital, medial, and lateral frontoparietal regions for dreams in REM [82].

6.4.3 PSD-Based Analysis

The topographic maps showing the activity in REM sleep, fig. 5.11 show a lot of mixed frequency and a broad specter of regions covered with high activity. According to the maps, frequencies in the delta and alpha bands are most prominent when dreaming in REM sleep. Previous work investigating delta and beta waves indicates that REM sleep is spatially and temporally heterogenous, similar to the finding of the PSD ratio topographic maps presented in fig. 5.11 and fig. 5.12 [92].

Previous work has been done on source level, while we use recordings from the scalp level [82]. Therefore, the findings may vary and are not directly comparable. With this in mind; our results show an increase in the power in delta in DE, compared to NE in the parieto-occipital region and in parts of

the central regions. This is contradictory to previous work, stating a decrease in low frequencies between (1-4 Hz) in the parieto-occipital region. The disagreement of the findings is also present for high frequencies (20-50 Hz) as the DE and NE activity of beta in fig. 5.26 and fig. 5.27 do not show increased power in frontal and lateral regions for DE compared to NE.

Looking at the topographic map for the delta band in fig. 5.26 and fig. 5.12, there are a lot of prominent increased activity of DE. Table 5.3 and table 5.4 showing the number of IMFs placed in the frequency bands also show that the majority of IMFs are placed in the delta band. Although delta waves are markers of NREM sleep, recent studies have shown that local low-frequency oscillation below 10 Hz could also occur during REM sleep [27]. However, local decreases in slow wave activity in the posterior parts of the brain are associated with the occurrence of dreaming, which do not corresponds to the topographic maps of our result, showing a lot of increased activity of DE in delta [27].

Other studies have identified two types of delta waves in REM sleep. The sawtooth delta bursts with frequencies of 2.5-3 Hz, resembling PGO waves, which are persistent in REM sleep, could explain the increased activity in delta [92, 94]. The occurrence of the bursts are associated with occipital-temporal and frontal-central regions when modeled at source level [92]. This is also visible in fig. 5.11 and fig. 5.12 showing increased activity of delta in DE in the occipital-temporal and frontal-central regions. The work of Bernardi also show delta waves below 2 Hz present in the medial-occipital region, which is also the present in our results for delta band when removing the first IMF [92].

When creating the topographic maps, we observe a significant number of empty values (NaN-values) for theta band. Meaning that, when assigning the IMFs into their respective frequency band, there are no frequency components assigned to theta. This suggests fewer theta waves in dreams during REM sleep compared to the other bands. However, this contradicts findings from other studies that highlight the dominance of theta and beta waves in REM sleep [93]. The study also identified theta waves in the frontal lobe, which is however consistent with our results, as the activity of DE in theta are present and dominating in the frontal, yet also the parietal lobe. The previous study only state the findings of REM-sleep and do not consider whether the subject experienced any dreams, therefore the results of the previous study could be a pinpoint for what to expect, but not necessarily the same as our results.

Upon suspicion that noise in the data is present, we eliminated the first IMF before constructing the MEMD-filtered signal. As explained earlier, this would eliminate the high-frequent components present in the signal, here assumed to be noise. In contrast to the original approach, we see that the map after the removal of the first IMF has less increased frequency components in the alpha and beta band, both for DE and NE. This makes sense, because of the removal of high frequency components.

However, removing the first IMF, assuming it is noise, could also potentially remove relevant information from the signal. We clearly see that the alpha and beta waves are removed when performing the procedure of removing the first IMF. This could be problematic, as it is not unusual to find high-frequency waves in dreams during REM sleep [93]. Studies have shown the dominance of beta waves in REM sleep, which could explain the activity present in the topographic maps in fig. 5.26. There is therefore a possibility that the first IMF is an important component for the discrimination of DE and NE in REM sleep. For further work, it would therefore be more beneficial to analyse the data even further and see if it is noisy and explore other methods for removing the noise, if present.

Similarly to the observation in the NREM data using Tononi dataset, the PSD as a feature for the classification yields high performance, while the classification using channel subsets from the PSD ratio are not better than random. This indicates that the PSD ratio is not giving the optimal subset of channels and should not be relied on for identifying the important channels for distinguishing DE from NE in classification tasks.

6.4.4 Combined Analysis

The classification performances for REM sleep indicate that dreams in REM are prominent and easily detectable. Regarding the identifications of important brain regions, it is not possible to see the exact same patterns in both topographic maps, but there are clearly some similarities. The channel importances extracted from the permutation-based channel analysis show importance in the frontal, occipital, and central regions, with more importance in the left parietal region. PSD-based analysis also shows activity across many regions, with most channels in the parietal region and some in the frontal area. Although the 10 channels in the PSD analysis are not optimal, it is increased DE activity in the parietal region across all frequency bands, especially for delta and alpha, in the frontal regions for delta and theta, and in the occipital region for delta and alpha. Although increased PSD activity does not necessarily indicate areas of importance in terms of classification performance, there could be a correlation between the areas highlighted by PSD analysis and the permutation-based channel selection. If so, it could indicate that the frontal, parietal, and occipital regions are important areas for detecting dreams during REM sleep.

Our results observing more importance in the frontal region also correlate with the increase in activity in the motor cortex during REM sleep, as the motor cortex is located in the frontal lobe.

According to previous work performed on the ZW dataset, no EEG predictors of dreaming in REM sleep have been identified [84], contradicting the observation that the frontal, parietal and occipital regions are important in our results. The differences could be due to the application of different methods. We used MEMD-filtered data in combination with feature extraction techniques and ML, and PSD-based analysis, whereas the previous work has only used PSD-based analysis.

6.5 NREM versus REM

6.5.1 Classification Performances

The best classification performance for DE versus NE is near perfect for both NREM and REM. For NREM the importance of finding the right classification method is shown by the performance difference for the various methods. For REM the classification performance is in general high across all methods. DE during REM sleep therefore appears to be more easily identified than during NREM sleep. Dream reports from REM sleep are reported to be more vivid, contain more sensory stimuli and are more emotionally charged than NREM sleep, where reports are more thought-like and conceptual. These differences could indicate that the difference between DE and NE is higher during REM sleep, explaining why the classification is easier during REM.

For the ZW dataset, the optimal segment length for classification found by testing is different between NREM and REM, being respectively 0.5 and 2 seconds. The reason could be that the processes iden-

tifying REM are dependent on longer segments, with NREM characteristics being shorter in time.

6.5.2 Permutation-Based Channel Selection

To observe the difference in important regions between NREM and REM sleep we compare the topographic maps of channel importances using PCA, CSP, and KNN on MEMD-filtered data from the ZW dataset, as these have shown the most clear indications of important regions (fig. 5.4, fig. 5.10). Both depict a high channel importance in the parietal lobe, especially on the left side. We also observe that the frontal region is important, specifically for REM on the left side. The occipital lobe is more important during REM than during NREM, indicated by the presence of two channels in this region for REM sleep, but not for NREM sleep.

The literature has shown that brain activity for specific experiences is similar during both REM sleep dreaming and wakefulness [82, 90, 91]. As such, the important areas observed in the topographic maps could be related to specific dream content correlated with the function of the area. Perceiving and speech in dreams have been shown to indicate increased high-frequency activity in a left posterior temporal region. These are common contents during DEs, which could indicate a general importance of this region. Thinking, another common content, is associated with the frontal brain regions.

The topographic maps for NREM and REM sleep indicate an area in the posterior left hemisphere, although over the parietal region. Topographic maps indicate importance on the scalp level, while brain activity is happening inside the brain. The electrodes placed in this area could therefore capture the activity corresponding to perceiving and speech, which explains the high importance. In addition, the importance of the frontal areas could be explained by the association with this area, with thinking.

6.5.3 PSD-Based Analysis

Considering the topographic maps of NREM ZW and Tononi, and REM ZW obtained from the PSD-based analysis, there are overall no distinct similarities observable. Analyzing and comparing these topographic maps is not straightforward because of the lack of removal of possible outliers in NREM Tononi and REM ZW, as well as artifacts in all datasets. This makes the analysis biased towards possible bad epochs and harder to identify distinct similarities. However, despite these challenges, there are still some observations that can be referred to.

The work of Siclari has previously highlighted the parieto-occipital region as a potential neural correlate of dreaming in NREM and REM sleep [82]. Considering the important channels highlighted in the maps, and also taking into account the credibility of these channels based on previous observations regarding their performance in classification tasks, we see that many of the channels for REM sleep are indeed placed in the parieto-occipital region. For ZW NREM and Tononi NREM, this region is not as prominently represented; however, there are three channels located in this area, for both datasets.

When examining the PSD ratio independently of the important channels, we observe increased DE activity compared to NE in the delta, theta, and alpha bands for ZW REM, in theta for ZW NREM, and in all bands for Tononi NREM. Therefore, our results show some correlation with Siclari's work, suggesting that the parieto-occipital region may indeed play a role in dreaming. However, due to

the issues with the topographic maps, such as the possible presence of outliers and artifacts, it is not possible to state this agreement definitively.

Another correlation between NREM and REM sleep found in previous research is the presence of slow delta waves below 2 Hz in the medial-occipital region [92]. This correlation is not prominently observed on any of the topographic maps we generated. However, the important channels identified in the Tononi NREM data, located in the medial-occipital region, might indicate slow wave activity. Furthermore, the increased area of DE in the delta band in the REM ZW data supports this observation. In contrast, for NREM ZW, there are no clear indications of a slow wave cluster, suggesting that this phenomenon might be dataset-specific or influenced by the presence of artifacts and noise in the data. Also, the observations from the previous work were done using source modeling, unlike using scalp modeling, like in this project. This may give differences in observed patterns.

6.6 Limitations

6.6.1 Low-Frequency Components in the ZW Dataset

Looking at the values of the frequency of the IMFs (table 5.1, table 5.2, table 5.3, table 5.4), the majority of components in the delta band have a frequency of $<0.5\text{Hz}$. These low frequency components could be unwanted components due to body movements creating artifacts below 1 Hz [114]. To identify whether the frequencies are caused by body movements, a suggestion would be to investigate whether the rate of body movement is stronger during N1 and REM than for N2 and N3 [115].

The low frequency components could also be caused by respiratory artifacts, characterized by periodic slow-wave activity with a frequency that correlates with the respiratory rate. The normal respiratory frequency ranges from 0.1 to 0.5 Hz, which means that respiratory movements may be captured by the EEG electrodes, causing unwanted low-frequency components, mistaken as delta waves. The respiratory artifacts are also often maximal in the posterior head region, which is correlating with the PSD activity of DE in fig. 5.26, in REM sleep. Consequently, the activity of DE could be misinterpreted as delta activity, while it is actually low-frequency components that occur due to respiratory artifacts [116, 117]. For future work, it is therefore important to identify if the low-frequency components are artifacts and thereby unwanted, and then remove these components before further analysis is performed.

Performing PSD analysis has provided valuable insight into the data in both datasets. Observing low-frequency components in both ZW NREM and REM suggests a need to further explore and identify potential outliers, artifacts, and noise that should be removed. Similarly, the difficulties encountered in creating a presentable topographic map for the Tononi data highlight the importance of thoroughly investigating the datasets before conducting any classification. Therefore, future work should include a detailed frequency analysis to gain deeper insight and remove outliers and artifacts. This may, in turn, enhance both the performance and the understanding of the data.

6.6.2 Segmenting Before Train-Test Data Split

Using segmentation to obtain more data samples can be performed in two ways, segmenting before splitting the dataset into train and test sets, or segmenting after splitting into train and test. In our

work, we have segmented the data before splitting into train and test sets. Further on, feature extraction methods are applied on each segment individually. This may cause possible data leakage, as segments belonging to the same recording are likely to be in the train and test set. Although there are no exact replicas of the same segment in both the train and the test set, there might be characteristics in the segments belonging to the same recording, such as noise. This may in turn reflect in data leakage. To rule out the possibility of data leakage, future experiments should consider segmenting the recordings after splitting into train and test. In this way, we can ensure that there are no/minor differences or if the performances are indeed effected by data leakage.

In addition, we should ensure the same amount of instances from all the subjects in train and test, to ensure the trained model has a similar proportion of information from all the subjects.

6.6.3 Segmenting and Channel Selection Before MEMD Filtering

Applying MEMD before segmentation may give data leakage.

For practical use of a system that applies MEMD to the data before using it as input to a classifier, it is better to segment the data before applying MEMD. This would also rule out the possibility of data leakage. When interpolating the maximum and minimum points of the original signal during IMF extraction, new artificial data are created to construct the envelopes. Then, when segmenting the data, they may contain characteristics from the previous artificial data point. In addition, when splitting the data into train and test, there might be segments in the test that contain characteristics from the training set.

MEMD learns from all channels, before choosing a subset.

In our methods we have applied MEMD to all channels and then chosen a subset of channels. When applying MEMD to all channels to extract the MEMD-filtered signal, the algorithm uses all the channels to identify common scales and extract the most relevant signals. Meaning, when constructing MEMD-filtered data using all channels and then performing classification using channel subsets, the performance will be biased due to containing more information. However, if the construction of MEMD-filtered was to be done on only a subset of channels and then be used for classification, it will only contain the information across the channel subset. We could therefore expect a degraded performance.

A possible solution for this problem, enhancing the classification performance, could be to use NA-MEMD. From fig. 5.25 we can see an increase of performance applying NA-MEMD compared to only using MEMD. Exploring the NA-MEMD with different numbers of noise channels and other datasets would be a suggestion for future work, working towards a system using only a subset of channels, such as in BCI applications. It is also important to identify whether the performance decreases when applying MEMD to only a subset of channels to obtain the MEMD-filtered data, and if so, try to improve it.

6.6.4 Recall of DE, Not Necessarily DE

The findings from Siclari [82] received criticism from Ruby [89], who claimed that neural correlates of dreaming had not been identified. Ruby pointed out the possibility of having a DE without any recall,

defining it as an NE, leading to NE containing both DE and NE in an unknown proportion. The extent to which this happens is not possible to measure, as the only way to get information about dreams are through dream reports from subjects. It should still be noted, as it is a possible limitation on what is really measured: Recalls of DEs, not necessarily DEs in themselves.

6.7 Future Work

6.7.1 Single-Subject Classification

From our results, we have observed that classifying DE is possible with high accuracy across multiple subjects. In a real-world application the objective is to use a pre-trained model on a new subject to say whether the subject is having a DE or NE at a given time. This requires a pre-trained classification pipeline capable of performing the classification on new, unseen subjects. The next step in the development of a dream decoder is therefore to investigate these scenarios to make this a reality. If this classification is not doable with acceptable performance, a possible solution could be to apply transfer learning. Recordings from the new subject would be used to tune the pre-trained model to gain knowledge about the individual characteristics of this subject, before applying the model to new recordings.

6.7.2 Optimal Channel Selection Algorithm

Neither permutation-based nor PSD ratio-based channel selection returns optimal sets of channels. Instead, the first indicates the importance of each channel by itself, whereas the latter indicates the channels with the highest difference between the classes. Both are useful for indicating the regions of importance, but the channel subsets we have chosen in our work are not necessarily optimal. For the PSD ratio-based approach, we observe that for the top 10 channels for the ZW REM dataset and for the top 29 channels for the Tononi dataset, the resulting classification scores are similar to random subsets.

Using another algorithm for channel selection could yield better results. Previous research has shown the NSGA-II algorithm to be suitable for channel selection on EEG data [88]. This is computationally expensive, however, especially for high-density EEG data, as the models would have to be re-trained for every tried combination. This would especially be an issue when using the Tononi dataset. A potential solution could be to combine both methods: Use permutation-based or PSD ratio-based channel selection to find a larger channel subset to be used for NSGA-II.

6.7.3 Working Towards a Brain Computer Interface System

Our work has presented some important steps towards designing a dream decoder. The research presented here is aiming towards a wearable BCI system with the most important channels. In order to achieve this, it is especially important to identify the important and optimal channels for classifying DE versus NE. We also need to take into account the discussion regarding the MEMD-filtering after choosing the subsets and segmenting the data before applying MEMD to the data.

Chapter 7

Conclusion

Our results are promising for identifying dreams across datasets and sleep stages, using EEG data as input to feature extraction techniques in combination with ML algorithms. In NREM sleep, we obtain high performance for both ZW and Tononi, using PCA, CSP and KNN in ZW, and PSD and XGB in Tononi. For both datasets, the best performances are obtained using the MEMD-filtered data as input to the classifiers. For REM sleep, our best results are achieved using PCA, CSP and KNN, also with MEMD-filtered data.

The identification of important brain regions, has on the other hand been a more challenging task. The two approaches for identifying important brain regions, permutation-based channel selection and PSD ratio-based analysis, have not given consistent results. While the PSD ratio-based analysis has been difficult to interpret due to no clear patterns correlating with previous work, the topographic maps extracted using permutation-based channel selection has given reasonable results for ZW NREM and ZW REM. They indicate areas over the parietal and occipital lobe as important. The channel subsets extracted from permutation-based channel selection has also given high performances in classification tasks. For the Tononi data, the topographic maps generated are less clear, but the classification performances using the channel subsets generated from the permutation-based channel selection are high. The results of classification using channel subsets across datasets and sleep stages have therefore shown similar performances, to when all available channels are used. This is a promising result towards the identification of important brain regions involved when dreaming.

Bibliography

- [1] Mithila Packiyathan. Towards a dream decoder: A study identifying dream emotions during rem sleep with machine learning and eeg signals. TTK4550 Specialization Project at NTNU, 2023.
- [2] André Torvestad. Making a dream decoder: Eeg-based classification of rem sleep, dream experiences, and dream emotional content using deep learning algorithms. TTK4550 Specialization Project at NTNU, 2023.
- [3] GBD 2019 Mental Disorders Collaborators et al. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet Psychiatry*, 9(2):137–150, 2022.
- [4] Marine Ambar Akkaoui, Michel Lejoyeux, Marie-Pia d'Ortho, and Pierre A. Geoffroy. Nightmares in Patients with Major Depressive Disorder, Bipolar Disorder, and Psychotic Disorders: A Systematic Review. *Journal of Clinical Medicine*, 9(12):3990, December 2020. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.
- [5] Rebecca L. Campbell and Anne Germain. Nightmares and Posttraumatic Stress Disorder (PTSD). *Current Sleep Medicine Reports*, 2(2):74–80, June 2016.
- [6] Tore A. Nielsen, Luc Laberge, Jean Paquet, Richard E. Tremblay, Frank Vitaro, and Jacques Montplaisir. Development of Disturbing Dreams During Adolescence and Their Relation to Anxiety Symptoms. *Sleep*, 23(6):1–10, 09 2000.
- [7] Anton Rimsh and Reinhard Pietrowsky. Analysis of dream contents of patients with anxiety disorders and their comparison with dreams of healthy participants. *Dreaming*, 31(4):303–319, 2021. Place: US Publisher: Educational Publishing Foundation.
- [8] Elizaveta Solomonova, Claudia Picard-Deland, Iris L. Rapoport, Marie-Hélène Pennestri, Maysa Saad, Tetyana Kendzerska, Samuel Paul Louis Veissiere, Roger Godbout, Jodi D. Edwards, Lena Quilty, and Rebecca Robillard. Stuck in a lockdown: Dreams, bad dreams, nightmares, and their relationship to stress, depression and anxiety during the COVID-19 pandemic. *PLOS ONE*, 16(11):e0259040, November 2021. Publisher: Public Library of Science.
- [9] Guido Giovanardi, Alice Fiorini Bincoletto, Roberto Baiocco, Martina Ferrari, Daniela Gentile, Mauro Siri, Annalisa Tanzilli, and Vittorio Lingiardi. Lockdown dreams: Dream content and emotions during the COVID-19 pandemic in an italian sample. *Psychoanalytic Psychology*, 39(2):111–126, 2022. Place: US Publisher: Educational Publishing Foundation.

- [10] Marianne Leuzinger-Bohleber, Marielle Donié, Julia Wichelmann, Gilles Ambresin, and Tamara Fischmann. Changes in dreams - the development of a dream-transformation scale in psychoanalyses with chronically depressed, early traumatized patients. *The Scandinavian Psychoanalytic Review*, 46(1-2):82–93, July 2023. Publisher: Routledge.
- [11] Robert Stickgold and Erin J. Wamsley. Why We Dream. In *Principles and Practice of Sleep Medicine*, pages 509–514.e4. Elsevier, January 2017.
- [12] Francesca Siclari and Giulio Tononi. Chapter 7 - Sleep and Dreaming. In Steven Laureys, Olivia Gosseries, and Giulio Tononi, editors, *The Neurology of Consciousness (Second Edition)*, pages 107–128. Academic Press, San Diego, January 2016.
- [13] Eric R. Kandel, James H. Schwartz, Thomas M. Jessell, Steven A. Siegelbaum, and A. J. Hudspeth, editors. *Principles of Neural Science, Fifth Edition*. McGraw-Hill Education / Medical, New York, 5th edition edition, October 2012.
- [14] Karoline Seljevoll Herleiksplass, Luis Alfredo Moctezuma, Junya Furuki, Yoko Suzuki, Takashi Abe, and Marta Molinas. Automatic Sleep-Wake Scoring with Optimally Selected EEG Channels from High-Density EEG. In Feng Liu, Yu Zhang, Hongzhi Kuai, Emily P. Stephen, and Hongjun Wang, editors, *Brain Informatics*, pages 420–431, Cham, 2023. Springer Nature Switzerland.
- [15] Embla C. S. Neverlien, Rose Lu, Mohit Kumar, and Marta Molinas. Decoding emotions from eeg responses elicited by videos using machine learning techniques on two datasets. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–4, 2023.
- [16] Jerry Tang, Amanda LeBel, Shailee Jain, and Alexander G. Huth. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience*, 26(5):858–866, May 2023. Publisher: Nature Publishing Group.
- [17] William Wong, Thomas Andrillon, Nicolas Decat, Rubén Herzog, Valdas Noreika, Katja Valli, Jennifer Windt, and Naotsugu Tsuchiya. The DREAM database. 2023. Monash University. Dataset. <https://doi.org/10.26180/22133105.v4>.
- [18] Khalid H. Jawabri and Sandeep Sharma. Physiology, Cerebral Cortex Functions. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), 2024.
- [19] Michael C. Corballis. Left Brain, Right Brain: Facts and Fantasies. *PLoS Biology*, 12(1):e1001767, January 2014.
- [20] Karen R Konkoly, Kristoffer Appel, Emma Chabani, Anastasia Mangiaruga, Jarrod Gott, Remington Mallett, Bruce Caughran, Sarah Witkowski, Nathan W Whitmore, Christopher Y Mazurek, et al. Real-time dialogue between experimenters and dreamers during rem sleep. *Current Biology*, 31(7):1417–1427, 2021.
- [21] Priyanka A. Abhang, Bharti W. Gawali, and Suresh C. Mehrotra. Chapter 3 - Technical Aspects of Brain Rhythms and Speech Parameters. In Priyanka A. Abhang, Bharti W. Gawali, and

- Suresh C. Mehrotra, editors, *Introduction to EEG- and Speech-Based Emotion Recognition*, pages 51–79. Academic Press, January 2016.
- [22] Giulio Tononi and Chiara Cirelli. Sleep function and synaptic homeostasis. *Sleep Medicine Reviews*, 10(1):49–62, 2006.
- [23] Giulio Tononi and Chiara Cirelli. Sleep and the Price of Plasticity: From Synaptic and Cellular Homeostasis to Memory Consolidation and Integration. *Neuron*, 81(1):12–34, January 2014.
- [24] Faith S. Luyster, Patrick J. Strollo, Phyllis C. Zee, and James K. Walsh. Sleep: A Health Imperative. *Sleep*, 35(6):727–734, June 2012.
- [25] Ahmed Bahammam, Divinagracia Gacuan, Smitha George, Karen Lorraine Acosta, Seithikurippu R. Pandi-Perumal, and Ravi Gupta. POLYSOMNOGRAPHY I: PROCEDURE AND TECHNOLOGY. In *Synopsis of Sleep Medicine*, pages 443–456. Apple Academic Press, September 2016.
- [26] B. Richard B, A. Claude L, H. Susan M, and et al. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications Version 2.5*. American Academy of Sleep Medicine, 2018.
- [27] Francesca Siclari and Giulio Tononi. Local aspects of sleep and wakefulness. *Current Opinion in Neurobiology*, 44:222–227, June 2017.
- [28] Robert A. Wilson and Frank C. Keil. *The MIT Encyclopedia of the Cognitive Sciences (MITECS)*. The MIT Press, April 1999.
- [29] Carlo Cipolli, Michele Ferrara, Luigi De Gennaro, and Giuseppe Plazzi. Beyond the neuropsychology of dreaming: Insights into the neural basis of dreaming with new techniques of sleep recording and analysis. *Sleep Medicine Reviews*, 35:8–20, October 2017.
- [30] J. A. Hobson and R. W. McCarley. The brain as a dream state generator: an activation-synthesis hypothesis of the dream process. *The American Journal of Psychiatry*, 134(12):1335–1348, December 1977.
- [31] T. Okuma. On the psychophysiology of dreaming: a sensory image-free association hypothesis of the dream process. *The Japanese Journal of Psychiatry and Neurology*, 46(1):7–22, March 1992.
- [32] J. Allan Hobson, Edward F. Pace-Schott, and Robert Stickgold. Dreaming and the brain: Toward a cognitive neuroscience of conscious states. *Behavioral and Brain Sciences*, 23(6):793–842, December 2000. Publisher: Cambridge University Press.
- [33] Tore A. Nielsen. A review of mentation in REM and NREM sleep: “Covert” REM sleep as a possible reconciliation of two opposing models. *Behavioral and Brain Sciences*, 23(6):851–866, December 2000. Publisher: Cambridge University Press.
- [34] Mark Solms. Dreaming and REM sleep are controlled by different brain mechanisms. *Behavioral and Brain Sciences*, 23(6):843–850, December 2000. Publisher: Cambridge University Press.

- [35] J. Satheesh Kumar and P. Bhuvaneswari. Analysis of Electroencephalography (EEG) Signals and Its Categorization—A Study. *Procedia Engineering*, 38:2525–2536, January 2012.
- [36] Richard Caton. Electrical currents of the brain. *The Journal of Nervous and Mental Disease*, 2(4):610, 1875.
- [37] Hans Berger. Über das elektroenkephalogramm des menschen. *Archiv für psychiatrie und nervenkrankheiten*, 87(1):527–570, 1929.
- [38] Andrea Biasiucci, Benedetta Franceschiello, and Micah M. Murray. Electroencephalography. *Current Biology*, 29(3):R80–R85, February 2019.
- [39] Jeffrey W. Britton, Lauren C. Frey, Jennifer L. Hopp, Pearce Korb, Mohamad Z. Koubeissi, William E. Lievens, Elia M. Pestana-Knight, and Erik K. St. Louis. *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*. American Epilepsy Society, Chicago, 2016.
- [40] Fernando Lopes da Silva. EEG: Origin and Measurement. In Christoph Mulert and Louis Lemieux, editors, *EEG - fMRI: Physiological Basis, Technique, and Applications*, pages 23–48. Springer International Publishing, Cham, 2022.
- [41] Michal Teplan. Fundamental of EEG Measurement. *MEASUREMENT SCIENCE REVIEW*, 2, January 2002.
- [42] Aamir Saeed Malik and Hafeez Ullah Amin. Chapter 1 - Designing an EEG Experiment. In Aamir Saeed Malik and Hafeez Ullah Amin, editors, *Designing EEG Experiments for Studying the Brain*, pages 1–30. Academic Press, January 2017.
- [43] Shailaja Kotte and J. R. K. Kumar Dabbakuti. Methods for removal of artifacts from EEG signal: A review. *Journal of Physics: Conference Series*, 1706(1):012093, December 2020. Publisher: IOP Publishing.
- [44] Md Kafiul Islam, Amir Rastegarnia, and Zhi Yang. Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology*, 46(4):287–305, November 2016.
- [45] Sung-Phil Kim. Preprocessing of EEG. In Chang-Hwan Im, editor, *Computational EEG Analysis: Methods and Applications*, Biological and Medical Physics, Biomedical Engineering, pages 15–33. Springer, Singapore, 2018.
- [46] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis | Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, March 1998.
- [47] Josep Dinarès-Ferran, Rupert Ortner, Christoph Guger, and Jordi Solé-Casals. A New Method to Generate Artificial Frames Using the Empirical Mode Decomposition for an EEG-Based Motor Imagery BCI. *Frontiers in Neuroscience*, 12, 2018.

- [48] N. Rehman and D. P. Mandic. Multivariate empirical mode decomposition. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2117):1291–1302, December 2009. Publisher: Royal Society.
- [49] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454(1971):903–995, 1998.
- [50] José Fernando Adrán Otero, Karmele López-de Ipina, Oscar Solans Caballer, Pere Marti-Puig, José Ignacio Sánchez-Méndez, Jon Iradi, Alberto Bergareche, and Jordi Solé-Casals. EMD-based data augmentation method applied to handwriting data for the diagnosis of Essential Tremor using LSTM networks. *Scientific Reports*, 12(1):12819, July 2022. Number: 1 Publisher: Nature Publishing Group.
- [51] Naveed Rehman, Cheolsoo Park, Norden Huang, and Danilo Mandic. EMD via MEMD: multivariate noise-aided computation of standard EMD. *Advances in Adaptive Data Analysis*, 05, June 2013.
- [52] Isabelle Guyon and André Elisseeff. An Introduction to Feature Extraction. In Isabelle Guyon, Masoud Nikravesh, Steve Gunn, and Lotfi A. Zadeh, editors, *Feature Extraction: Foundations and Applications*, pages 1–25. Springer, Berlin, Heidelberg, 2006.
- [53] Anupreet Kaur Singh and Sridhar Krishnan. Trends in EEG signal feature extraction applications. *Frontiers in Artificial Intelligence*, 5:1072801, January 2023.
- [54] Dharmendra Pathak, Ramgopal Kashyap, and Surendra Rahamatkar. Chapter 10 - A study of deep learning approach for the classification of electroencephalogram (EEG) brain signals. In Rajiv Pandey, Sunil Kumar Khatri, Neeraj kumar Singh, and Parul Verma, editors, *Artificial Intelligence and Machine Learning for EDGE Computing*, pages 133–144. Academic Press, January 2022.
- [55] JOHN Dempster. CHAPTER SIX - Signal Analysis and Measurement. In JOHN Dempster, editor, *The Laboratory Computer*, Biological Techniques Series, pages 136–171. Academic Press, London, January 2001.
- [56] F Lotte, L Bougrain, A Cichocki, M Clerc, M Congedo, A Rakotomamonjy, and F Yger. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, 15(3):031005, June 2018.
- [57] Hamdi Altaheri, Ghulam Muhammad, Mansour Alsulaiman, Syed Umar Amin, Ghadir Ali Altuwaijri, Wadood Abdul, Mohamed A. Bencherif, and Mohammed Faisal. Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: a review. *Neural Computing and Applications*, 35(20):14681–14722, July 2023.
- [58] W.T. Cochran, J.W. Cooley, D.L. Favin, H.D. Helms, R.A. Kaenel, W.W. Lang, G.C. Maling, D.E. Nelson, C.M. Rader, and P.D. Welch. What is the fast Fourier transform? *Proceedings of the IEEE*, 55(10):1664–1674, 1967.

- [59] Andrzej Maćkiewicz and Waldemar Ratajczak. Principal components analysis (PCA). *Computers & Geosciences*, 19(3):303–342, March 1993.
- [60] I.T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag, New York, 2002.
- [61] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-robert Muller. Optimizing Spatial filters for Robust EEG Single-Trial Analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008.
- [62] Benjamin Blankertz, Guido Dornhege, Matthias Krauledat, Klaus-Robert Müller, and Gabriel Curio. The Berlin brain-computer interface: Report from the feedback sessions. Technical report, Fraunhofer FIRST, 2005.
- [63] Thomas W. Edgar and David O. Manz. Chapter 6 - Machine Learning. In Thomas W. Edgar and David O. Manz, editors, *Research Methods for Cyber Security*, pages 153–173. Syngress, January 2017.
- [64] Stuart Russell. Chapter 4 - Machine Learning. In Margaret A. Boden, editor, *Artificial Intelligence*, Handbook of Perception and Cognition, pages 89–133. Academic Press, San Diego, January 1996.
- [65] Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.
- [66] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [67] Abdulhamit Subasi, Siba Smarak Panigrahi, Bhalchandra Sunil Patil, M. Abdullah Canbaz, and Riku Klén. Chapter 8 - Advanced pattern recognition tools for disease diagnosis. In Akash Kumar Bhoi, Victor Hugo Costa de Albuquerque, Samarendra Nath Sur, and Paolo Barsocchi, editors, *5G IoT and Edge Computing for Smart Healthcare*, Intelligent Data-Centric Systems, pages 195–229. Academic Press, January 2022.
- [68] Siddharth Misra and Hao Li. Chapter 9 - Noninvasive fracture characterization based on the classification of sonic wave travel times. In Siddharth Misra, Hao Li, and Jiabo He, editors, *Machine Learning for Subsurface Characterization*, pages 243–287. Gulf Professional Publishing, January 2020.
- [69] Y. S. Park and S. Lek. Chapter 7 - Artificial Neural Networks: Multilayer Perceptron for Ecological Modeling. In Sven Erik Jørgensen, editor, *Developments in Environmental Modelling*, volume 28 of *Ecological Model Types*, pages 123–140. Elsevier, January 2016.
- [70] Iqbal H. Sarker. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, 2(6):420, August 2021.
- [71] Björn Lindqvist, Sina Sharif Mansouri, Christoforos Kanellakis, and Vignesh Kottayam Viswanathan. Chapter 12 - ARW deployment for subterranean environments. In

- George Nikolakopoulos, Sina Sharif Mansouri, and Christoforos Kanellakis, editors, *Aerial Robotic Workers*, pages 213–243. Butterworth-Heinemann, January 2023.
- [72] Yasen Jiao and Pufeng Du. Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quantitative Biology*, 4(4):320–330, December 2016.
- [73] J. Richard Landis and Gary G. Koch. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174, 1977. Publisher: [Wiley, International Biometric Society].
- [74] Chao He, Jialu Liu, Yuesheng Zhu, and Wencai Du. Data Augmentation for Deep Neural Networks Model in EEG Classification Task: A Review. *Frontiers in Human Neuroscience*, 15, 2021.
- [75] Alhassan Mumuni and Fuseini Mumuni. Data augmentation: A comprehensive survey of modern approaches. *Array*, 16:100258, December 2022.
- [76] Turkey Alotaiby, Fathi E. Abd El-Samie, Saleh A. Alshebeili, and Ishtiaq Ahmad. A review of channel selection algorithms for EEG signal processing. *EURASIP Journal on Advances in Signal Processing*, 2015(1):66, August 2015.
- [77] Abdullah, Ibrahima Faye, and Md Rafiqul Islam. EEG Channel Selection Techniques in Motor Imagery Applications: A Review and New Perspectives. *Bioengineering*, 9(12):726, November 2022.
- [78] Hiromasa Kaneko. Cross-validated permutation feature importance considering correlation between features. *Analytical Science Advances*, 3(9-10):278–287, 2022.
- [79] Luis Alfredo Moctezuma, Yoko Suzuki, Junya Furuki, Marta Molinas, and Takashi Abe. Enhancing sleep stage classification with 2-class stratification and permutation-based channel selection. In *45th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society*, pages 1–4. IEEE, 2024.
- [80] Thomas Bartz-Beielstein, Jürgen Branke, Jörn Mehnen, and Olaf Mersmann. Evolutionary Algorithms. *WIREs Data Mining and Knowledge Discovery*, 4(3):178–195, 2014.
- [81] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, April 2002. Conference Name: IEEE Transactions on Evolutionary Computation.
- [82] 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, June 2017. Number: 6 Publisher: Nature Publishing Group.
- [83] Valdas Noreika, Katja Valli, Hetti Lahtela, and Antti Revonsuo. Early-night serial awakenings as a new paradigm for studies on NREM dreaming. *International Journal of Psychophysiology*, 74(1):14–18, October 2009.

- [84] Jing Zhang and Erin Wamsley. EEG predictors of dreaming outside of REM sleep. *Psychophysiology*, 56, 03 2019.
- [85] Francesca Siclari, Giulio Bernardi, Jacinthe Cataldi, and Giulio Tononi. Dreaming in NREM Sleep: A High-Density EEG Study of Slow Waves and Spindles. *Journal of Neuroscience*, 38(43):9175–9185, October 2018. Publisher: Society for Neuroscience Section: Research Articles.
- [86] Serena Scarpelli, Aurora D’Atri, Anastasia Mangiaruga, Cristina Marzano, Maurizio Gorgoni, Cinzia Schiappa, Michele Ferrara, and Luigi De Gennaro. Predicting Dream Recall: EEG Activation During NREM Sleep or Shared Mechanisms with Wakefulness? *Brain Topography*, 30(5):629–638, September 2017.
- [87] William Wong, Valdas Noreika, Levente Móró, Antti Revonsuo, Jennifer Windt, Katja Valli, and Naotsugu Tsuchiya. The Dream Catcher experiment: blinded analyses failed to detect markers of dreaming consciousness in EEG spectral power. *Neuroscience of Consciousness*, 2020(1):niaa006, January 2020.
- [88] Luis Alfredo Moctezuma, Felix Ipanaque, Marta Molinas, and Takashi Abe. Dream emotions identified without awakenings by machine and deep learning from electroencephalographic signals in rem sleep. In *2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRaine)*, pages 444–449, 2023.
- [89] Perrine Marie Ruby. The Neural Correlates of Dreaming Have Not Been Identified Yet. Commentary on “The Neural Correlates of Dreaming. Nat Neurosci. 2017”. *Frontiers in Neuroscience*, 14, 2020.
- [90] T. Horikawa, M. Tamaki, Y. Miyawaki, and Y. Kamitani. Neural Decoding of Visual Imagery During Sleep. *Science*, May 2013. Publisher: American Association for the Advancement of Science.
- [91] Martin Dresler, Stefan P. Koch, Renate Wehrle, Victor I. Spoormaker, Florian Holsboer, Axel Steiger, Philipp G. Sämann, Hellmuth Obrig, and Michael Czisch. Dreamed movement elicits activation in the sensorimotor cortex. *Current biology: CB*, 21(21):1833–1837, November 2011.
- [92] Giulio Bernardi, Monica Betta, Emiliano Ricciardi, Pietro Pietrini, Giulio Tononi, and Francesca Siclari. Regional Delta Waves In Human Rapid Eye Movement Sleep. *Journal of Neuroscience*, 39(14):2686–2697, April 2019. Publisher: Society for Neuroscience Section: Research Articles.
- [93] Sujith Vijayan, Kyle Q Lepage, Nancy J Kopell, and Sydney S Cash. Frontal beta-theta network during REM sleep. *eLife*, 6:e18894, January 2017.
- [94] Jin-Xian Gao, Guizhong Yan, Xin-Xuan Li, Jun-Fan Xie, Karen Spruyt, Yu-Feng Shao, and Yi-Ping Hou. The ponto-geniculo-occipital (pgo) waves in dreaming: An overview. *Brain Sciences*, 13(9), 2023.
- [95] Erin Wamsley, Jing Zhang, and Megan Collins. Zhang & Wamsley 2019 Final. 2023. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.22226692.v1>.

- [96] Francesca Siclari, Benjamin Baird, Lampros Perogamvros, Melanie Boly, Joshua LaRocque, and Giulio Tononi. Tononi Serial Awakenings. 2023. Monash University. Dataset. <https://doi.org/10.26180/23306054.v2>.
- [97] Xia Chen, Xiangbin Teng, Han Chen, Yafeng Pan, and Philipp Geyer. Toward reliable signals decoding for electroencephalogram: A benchmark study to EEGNeX. *Biomedical Signal Processing and Control*, 87:105475, January 2023.
- [98] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.
- [99] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010.
- [100] Charles R. Harris et al. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.
- [101] Pauli Virtanen et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.
- [102] Alexandre Gramfort et al. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267):1–13, 2013.
- [103] F. Pedregosa et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [104] Martín Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [105] François Chollet et al. Keras, 2015. Software available from keras.io.
- [106] Magnus Sjölander, Magnus Jahre, Gunnar Tufte, and Nico Reissmann. EPIC: an energy-efficient, high-performance GPGPU computing research infrastructure. *CoRR*, abs/1912.05848, 2019.
- [107] Tang Yan, Tang Jingtian, and Gong Andong. Multi-Class EEG Classification for Brain Computer Interface based on CSP. In *2008 International Conference on BioMedical Engineering and Informatics*, volume 2, pages 469–472, May 2008. ISSN: 1948-2922.
- [108] Mary Judith Antony, Baghavathi Priya Sankaralingam, Rakesh Kumar Mahendran, Akber Abid Gardezi, Muhammad Shafiq, Jin-Ghoo Choi, and Habib Hamam. Classification of EEG Using Adaptive SVM Classifier with CSP and Online Recursive Independent Component Analysis. *Sensors*, 22(19):7596, January 2022. Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.
- [109] William Yaputra Budiman, Handayani Tjandrasa, and Dini Adni Navastara. Classification of EEG signals using Common Spatial Pattern-Principle Component Analysis and Interval Type-2

- Fuzzy Logic System. In *2016 International Conference on Information & Communication Technology and Systems (ICTS)*, pages 85–89, October 2016. ISSN: 2338-185X.
- [110] James Bergstra, Daniel Yamins, and David Cox. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28(1) of *Proceedings of Machine Learning Research*, pages 115–123, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.
- [111] Håkon Stenwig, Andres Soler, Junya Furuki, YoRecalling Dreams: An Examination of Daily Variation Suzuki, Individual Differencesko, Takashi Abe, and Marta Molinas. Automatic Sleep Stage Classification with Optimized Selection of EEG Channels. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1708–1715, December 2022.
- [112] Neha Hooda and Neelesh Kumar. Cognitive Imagery Classification of EEG Signals using CSP-based Feature Selection Method. *IETE Technical Review*, 37(3):315–326, May 2020. Publisher: Taylor & Francis.
- [113] Zahra M. Aghajan, Peter Schuette, Tony A. Fields, Michelle E. Tran, Sameed M. Siddiqui, Nicholas R. Hasulak, Thomas K. Tcheng, Dawn Eliashiv, Emily A. Mankin, John Stern, Itzhak Fried, and Nanthia Suthana. Theta Oscillations in the Human Medial Temporal Lobe during Real-World Ambulatory Movement. *Current Biology*, 27(24):3743–3751.e3, December 2017.
- [114] Gavin R. Philips, Mehrnaz Kh. Hazrati, Janis J. Daly, and Jose C. Principe. Addressing low frequency movement artifacts in EEG signal recorded during center-out reaching tasks. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6497–6500, August 2014. ISSN: 1558-4615.
- [115] Johanna Wilde-Frenz and Hartmut Schulz. Rate and Distribution of Body Movements during Sleep in Humans. *Perceptual and Motor Skills*, 56(1):275–283, February 1983. Publisher: SAGE Publications Inc.
- [116] Jeffrey W. Britton. Adult EEG: Artifacts and the EEG. In Devon I. Rubin and Jasper R. Daube, editors, *Clinical Neurophysiology*, page 0. Oxford University Press, March 2016.
- [117] Delaram Jarchi, Dario Salvi, Lionel Tarassenko, and David A. Clifton. Validation of Instantaneous Respiratory Rate Using Reflectance PPG from Different Body Positions. *Sensors (Basel, Switzerland)*, 18(11):3705, October 2018.

Appendix A

Additional Classification Results

This appendix contains additional classification results not presented in Results, when only the best classifier for a given task is presented.

A.1 PSD and ML Classification: Including DEWR in Tononi Data

The following tables contain classification results for all classifiers tested (KNN, XGB, and SVM) when including DEWR into either the DE or the NE class for the Tononi dataset. Results are presented for using PCA and CSP, and PSD as feature extraction methods.

Metric	KNN	XGB	SVM	RF
Accuracy	0.979 ± 0.002	0.966 ± 0.003	0.960 ± 0.004	0.943 ± 0.006
F-score	0.975 ± 0.003	0.958 ± 0.004	0.950 ± 0.005	0.928 ± 0.009
Precision	0.972 ± 0.005	0.967 ± 0.004	0.968 ± 0.009	0.960 ± 0.007
Recall	0.977 ± 0.003	0.949 ± 0.004	0.934 ± 0.007	0.898 ± 0.010
AUROC	0.979 ± 0.002	0.963 ± 0.003	0.956 ± 0.004	0.936 ± 0.007
Kappa	0.957 ± 0.004	0.929 ± 0.007	0.917 ± 0.008	0.880 ± 0.014

Table A.1: Performance metrics for PCA, CSP and classifiers KNN, XGB, RBF SVM, and RF on low-pass filtered data from Tononi Serial Awakenings. Class combination: DE vs NE and DEWR.

Metric	KNN	XGB	SVM	RF
Accuracy	0.983 ± 0.004	0.969 ± 0.008	0.964 ± 0.005	0.951 ± 0.002
F-score	0.986 ± 0.003	0.976 ± 0.006	0.972 ± 0.004	0.963 ± 0.001
Precision	0.983 ± 0.006	0.969 ± 0.012	0.962 ± 0.010	0.940 ± 0.009
Recall	0.990 ± 0.004	0.982 ± 0.006	0.982 ± 0.003	0.987 ± 0.008
AUROC	0.980 ± 0.005	0.963 ± 0.011	0.956 ± 0.008	0.937 ± 0.005
Kappa	0.962 ± 0.010	0.931 ± 0.017	0.921 ± 0.011	0.891 ± 0.004

Table A.2: Performance metrics for PCA, CSP and classifiers KNN, XGB, RBF SVM, and RF on low-pass filtered data from Tononi Serial Awakenings. Class combination: DE and DEWR vs NE.

Metric	KNN	XGB	SVM
Accuracy	0.824 ± 0.003	0.976 ± 0.003	0.599 ± 0.002
F-score	0.789 ± 0.004	0.970 ± 0.004	0.091 ± 0.005
Precision	0.794 ± 0.008	0.979 ± 0.004	0.872 ± 0.055
Recall	0.784 ± 0.013	0.963 ± 0.005	0.048 ± 0.003
AUROC	0.819 ± 0.004	0.974 ± 0.003	0.521 ± 0.002
Kappa	0.639 ± 0.007	0.950 ± 0.006	0.049 ± 0.005

Table A.3: Performance metrics for PSD and classifiers KNN, XGB, SVM, and RF on low-pass filtered data from Tononi Serial Awakenings. Class combination: DE vs NE and DEWR.

Metric	KNN	XGB	SVM
Accuracy	0.837 ± 0.004	0.976 ± 0.002	0.636 ± 0.003
F-score	0.868 ± 0.004	0.980 ± 0.002	0.770 ± 0.001
Precision	0.864 ± 0.003	0.976 ± 0.003	0.629 ± 0.002
Recall	0.872 ± 0.008	0.985 ± 0.003	0.996 ± 0.001
AUROC	0.827 ± 0.004	0.973 ± 0.002	0.531 ± 0.004
Kappa	0.657 ± 0.008	0.948 ± 0.004	0.076 ± 0.009

Table A.4: Performance metrics for PSD and classifiers KNN, XGB, and RBF SVM on low-pass filtered data from Tononi Serial Awakenings. Class combination: DE and DEWR vs NE.

A.2 Channel Subset Classification with Tononi Data

This section contains the results from classifying on subsets of 50, 30, 20, and 10 channels from the Tononi dataset, selected by permutation-based channel selection. For each subset we present results from PCA and CSP as feature extraction for the classifiers KNN, XGB, SVM, and RF, and EEGNeX.

Metric	KNN	XGB	SVM	RF	EEGNeX
Accuracy	0.932 ± 0.01	0.925 ± 0.004	0.871 ± 0.013	0.906 ± 0.011	0.694 ± 0.012
F-score	0.933 ± 0.009	0.927 ± 0.004	0.874 ± 0.013	0.908 ± 0.011	0.696 ± 0.037
Precision	0.929 ± 0.012	0.924 ± 0.009	0.867 ± 0.016	0.908 ± 0.016	0.735 ± 0.079
Recall	0.938 ± 0.013	0.931 ± 0.009	0.882 ± 0.02	0.907 ± 0.015	0.694 ± 0.149
AUROC	0.932 ± 0.01	0.925 ± 0.004	0.871 ± 0.013	0.906 ± 0.011	0.694 ± 0.018
Kappa	0.863 ± 0.019	0.851 ± 0.008	0.742 ± 0.027	0.812 ± 0.022	0.387 ± 0.031

Table A.5: Performance metrics for PCA, CSP and classifiers KNN, XGB, RBF SVM, and RF, and EEGNeX on subset of low-pass filtered data from Tononi Serial Awakenings. Channel subset: 50 channels chosen by permutation-based channel selection.

Metric	KNN	XGB	SVM	RF	EEGNeX
Accuracy	0.906 ± 0.01	0.906 ± 0.008	0.851 ± 0.018	0.89 ± 0.015	0.641 ± 0.024
F-score	0.905 ± 0.011	0.908 ± 0.008	0.849 ± 0.021	0.891 ± 0.015	0.58 ± 0.08
Precision	0.938 ± 0.011	0.914 ± 0.008	0.874 ± 0.013	0.907 ± 0.015	0.78 ± 0.103
Recall	0.873 ± 0.015	0.901 ± 0.01	0.827 ± 0.034	0.875 ± 0.019	0.521 ± 0.235
AUROC	0.907 ± 0.011	0.907 ± 0.008	0.851 ± 0.018	0.891 ± 0.015	0.646 ± 0.03
Kappa	0.812 ± 0.021	0.813 ± 0.016	0.702 ± 0.037	0.781 ± 0.03	0.288 ± 0.057

Table A.6: Performance metrics for PCA, CSP and classifiers KNN, XGB, RBF SVM, and RF, and EEGNeX on subset of low-pass filtered data from Tononi Serial Awakenings. Channel subset: 30 channels chosen by permutation-based channel selection.

Metric	KNN	XGB	SVM	RF	EEGNeX
Accuracy	0.882 ± 0.011	0.891 ± 0.01	0.836 ± 0.012	0.884 ± 0.009	0.626 ± 0.024
F-score	0.882 ± 0.012	0.893 ± 0.01	0.834 ± 0.013	0.885 ± 0.01	0.529 ± 0.083
Precision	0.897 ± 0.011	0.897 ± 0.014	0.864 ± 0.013	0.9 ± 0.012	0.766 ± 0.053
Recall	0.868 ± 0.016	0.889 ± 0.007	0.805 ± 0.014	0.87 ± 0.009	0.426 ± 0.147
AUROC	0.882 ± 0.011	0.891 ± 0.01	0.837 ± 0.012	0.885 ± 0.01	0.634 ± 0.019
Kappa	0.764 ± 0.023	0.782 ± 0.02	0.672 ± 0.024	0.769 ± 0.019	0.264 ± 0.04

Table A.7: Performance metrics for PCA, CSP and classifiers KNN, XGB, RBF SVM, and RF, and EEGNeX on subset of low-pass filtered data from Tononi Serial Awakenings. Channel subset: 20 channels chosen by permutation-based channel selection.

Metric	KNN	XGB	SVM	RF	EEGNeX
Accuracy	0.739 ± 0.014	0.742 ± 0.008	0.702 ± 0.02	0.744 ± 0.011	0.649 ± 0.008
F-score	0.739 ± 0.014	0.741 ± 0.01	0.664 ± 0.021	0.739 ± 0.012	0.675 ± 0.026
Precision	0.752 ± 0.017	0.757 ± 0.011	0.782 ± 0.035	0.771 ± 0.014	0.651 ± 0.019
Recall	0.728 ± 0.022	0.727 ± 0.02	0.579 ± 0.023	0.71 ± 0.016	0.707 ± 0.074
AUROC	0.739 ± 0.014	0.742 ± 0.008	0.704 ± 0.02	0.745 ± 0.011	0.647 ± 0.007
Kappa	0.477 ± 0.028	0.484 ± 0.015	0.407 ± 0.04	0.489 ± 0.023	0.295 ± 0.015

Table A.8: Performance metrics for PCA, CSP and classifiers KNN, XGB, RBF SVM, and RF, and EEGNeX on subset of low-pass filtered data from Tononi Serial Awakenings. Channel subset: 10 channels chosen by permutation-based channel selection.

Appendix B

**ESANN2024 Conference Paper: *Unveiling
Dreams: Moving Towards Automatic
Dream Decoding via qualitative EEG
Analysis and Machine Learning***

Unveiling Dreams: Moving Towards Automatic Dream Decoding via qualitative EEG Analysis and Machine Learning

André Torvestad¹, Mithila Packiyathan¹,
Luis Alfredo Moctezuma^{2*} and Marta Molinas¹

1- Department of Engineering Cybernetics,
Norwegian University of Science and Technology. Trondheim, Norway

2- International Institute for Integrative Sleep Medicine,
University of Tsukuba. Tsukuba, Ibaraki, Japan.

Abstract. Equipping brain-computer interfaces with dream decoding capabilities could be vital in healthcare applications. We used high-density electroencephalogram data from non-rapid eye movement sleep to conduct qualitative analysis employing multivariate empirical mode decomposition and power spectral density (PSD) for preprocessing and machine learning algorithms to distinguish between a dream experience and no experience. Qualitative analysis shows differences between the two classes, especially in the theta and beta bands. We achieve a classification performance of 0.915 in accuracy, 0.851 in AUROC, and 0.715 in kappa with PSD features and extreme gradient boosting classifier.

1 Introduction

Throughout human history, understanding the nature of dreams has been long-sought. In the book of Genesis the Pharaoh sent for all the wise men in Egypt to help him understand the meaning of his dreams [1]. Today, we are on the path of making the Pharaohs long-overdue dream decoder a reality. The goal has transcended deciphering messages from the gods, it now aims to deepens our comprehension of individuals' mental and emotional states. Research has shown a correlation between dream content and mental health conditions, suggesting that analyzing dream content could be beneficial for the early diagnosis and treatment of various mental disorders [2, 3].

Previous research has shown differences in EEG patterns between dream experiences (DE) and no experiences (NE) during non-rapid eye movement (NREM) sleep using power spectral density (PSD) [4–7]. Our work aims to further contribute to the understanding of dreams during NREM sleep from EEG-based analysis and classification, as a step towards the realization of a brain-computer interface (BCI) with dream decoding capabilities.

Here, we present a set of experiments for dream experience and no experience classification based on high-density EEG recordings during NREM sleep. We present a qualitative analysis of sleep EEG with dream experiences versus no

*This work was partially supported by the Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship for Research in Japan: Fellowship ID P22716

experience. To facilitate integration into BCIs, we present classification results utilizing channel subsets derived from qualitative analysis.

2 Materials and Methods

2.1 EEG Dataset

The dataset consists of EEG recordings from 28 participants recorded using a 58-electrode cap placed according to the international 10-10 system, a sampling rate of 400 Hz, and a high-pass filter set at 0.1 Hz [5].

Participants were instructed to, whenever awakened, provide a dream report. The reports were verbal descriptions of everything that was going through their mind just before awakening or indicating if they could not recall anything. The awakenings were made during sleep onset, N2 sleep, and REM sleep. Most of the recordings were made on sleep onset during the first hour of the night, which came from sleep stages N1 and N2. One N2 and one REM sleep dream reports were made at least one hour after the last sleep onset report. The N2 and REM reports were separated by a minimum of thirty minutes.

For our experiment, we considered whether the participant reported a DE or NE, and the data collected during the sleep stages N1 and N2. A total of 181 NREM sleep epochs are considered, with 24% labeled NE and 76% as DE. Each epoch lasts 73 seconds before awakening.

2.2 Preprocessing, power spectral density, and classification

A low-pass filter with cutoff frequency of 45 Hz was applied to avoid high-frequency components in the data (referred to as low-pass filtered). Multivariate empirical mode decomposition (MEMD) was applied to the low-pass filtered data to extract the intrinsic mode functions (IMFs), then the three closest IMFs to the low-pass filtered signal, calculated by Euclidean distance, were selected to reconstruct the EEG signals (referred to as MEMD-filtered).

To increase the number of instances for the classifiers, the IMFs were divided into 0.5-second segments before calculating the PSD; This segment size was selected because it showed the best classification performance.

PSD was calculated from the IMFs for two purposes: 1) To create a topographic map of the average PSD ratio between the DE and NE reports. 2) To be used as features for extreme gradient boosting (XGBoost). The convolutional neural network (CNN) EEGNeX was also tested on MEMD-filtered data for comparison [8].

2.3 EEG Channel Selection

After PSD calculation, the 58 EEG channels were ordered based on their PSD values, from highest to lowest. The importance of each channel is calculated as $\frac{PSD(DE)}{PSD(NE)} - 1$. With this, we utilize the distinctions between the classes that have been observed in the topographic maps to define the channel importance both when DE is higher than NE and vice versa.

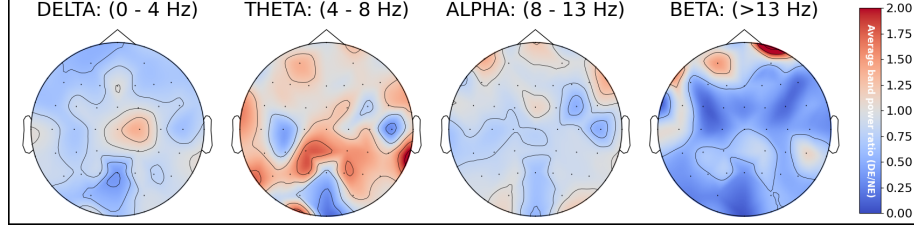


Fig. 1: The average PSD ratio (DE over NE) across all recordings in each channel for IMFs in different frequency bands.

Channel subsets, ranging from one to thirty channels, were created based on channel importance. The n -th subset contains the n most important channels. Classification was performed in each channel subset and evaluated with the kappa metric. The kappa score evaluates performance taking into account expected performance. As expected performance is higher for an imbalanced data set, we chose the kappa metric to evaluate performance on the channel subsets. The kappa score can be interpreted as the agreement between the true data labels and the predicted labels; 0.0-0.2 indicates slight agreement, 0.21-0.40 poor agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1.00 almost perfect agreement [9].

3 Results

3.1 PSD Ratio between DE and NE with a Topographic Map

The topographic map was created to visualize the difference in PSD between DE and NE in different frequency bands. We iterate through all IMFs within each class and sort them into one of four groups: delta (0 - 4 Hz), theta (4 - 8 Hz), alpha (8 - 13 Hz) or beta (>13 Hz). The sorting is based on the frequency corresponding to the maximum PSD of the IMF. The grand average of IMFs was calculated for each group in each channel, and then the PSD ratio as the grand average PSD for DE over NE.

The resulting topographic maps are shown in fig. 1. The PSD of DE is higher in the theta band, especially in the parietal and temporal lobes. This is also seen in the frontal lobe in the beta band and in a small central area in the delta. The PSD of NE is generally higher in both the delta and the beta band. This is most prominent in the beta band, which shows the lowest ratio values. The alpha band shows some differences, without any areas being prominent.

3.2 Classification Performance for DE and NE

Three classification configurations are presented: 1) PSD and XGBoost on MEMD-filtered data, 2) EEGNeX on MEMD-filtered data, and 3) XGBoost and PSD on low-pass filtered data. All configurations are validated with five-fold cross-validation, each with a train-test split ratio of 80-20 for each fold.

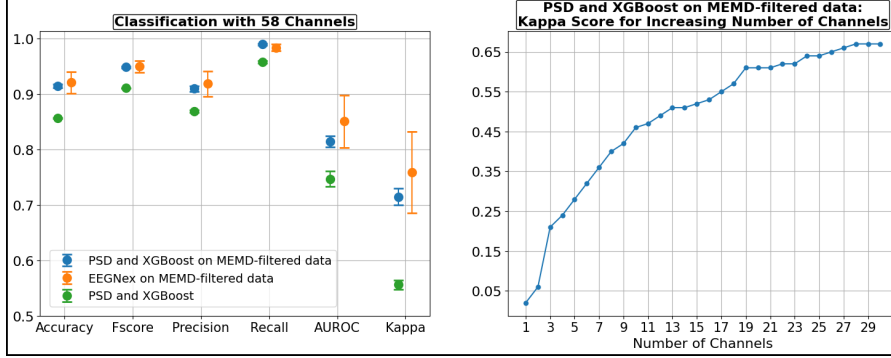


Fig. 2: Classification performance under different configurations: *Left*) XGBoost and PSD on MEMD-filtered data. EEGNeX on MEMD-filtered data. XGBoost and PSD on low-pass filtered data. *Right*) The kappa score evaluated on channel subsets with an increasing number of channels, sorted by highest PSD ratio difference from 1 across all frequency bands.

The classifications were evaluated using accuracy, F-score, precision, recall, area under the receiver operating characteristic curve (AUROC), and kappa.

We also present the kappa score from XGBoost and PSD in MEMD-filtered data on channel subsets with an increasing number of channels, up to thirty channels. The classification performance for each channel subset is validated with a five-fold cross-validation, using a train-test split ratio of 80-20.

Using PSD and XGBoost on MEMD-filtered data results in 0.915 accuracy, 0.949 F-score, 0.910 precision, 0.990 recall, 0.815 AUROC, and 0.715 kappa. With EEGNeX on MEMD-filtered data we obtain 0.921 accuracy, 0.950 F-score, 0.919 precision, 0.984 recall, 0.851 AUROC, and 0.759 kappa. Although the metrics are higher for EEGNeX, the standard deviation is also higher. As visualized in fig. 2, the performance of XGBoost is within the standard deviation of EEGNeX for all metrics, with a smaller standard deviation. Omitting MEMD-filtering with PSD and XGBoost, the performance drops to 0.857 accuracy, 0.911 F-score, 0.869 precision, 0.958 recall, 0.747 AUROC, and 0.556 kappa, indicating the importance of MEMD-filtering.

The kappa score with an increasing number of EEG channels shows that the performance increases rapidly in the beginning, before gradually saturating. We obtain a kappa score of 0.46 with the top 10 most important EEG channels, 0.61 with the top 20, and 0.67 with the top 30.

4 Discussion

From our PSD analysis, we observed higher brain activity while dreaming in the theta frequency band, especially over the parietal and temporal lobes. Theta waves are, in awake subjects, related to daydreaming and movement, as well as hyperactivity and impulsivity [10, 11]. The parietal lobe is related to language,

spatial recognition, and sensorimotor planning, whereas the temporal lobe is related to vision, hearing, and declarative memory [12]. Dream reports from NREM sleep tend to be more thought-like and less bizarre than dreams from REM sleep - more similar to daydreams [13]. Therefore, they resonate well with the observed increase in activity.

The PSD ratio is in general low in the beta frequency band, except for an increase in the frontal lobe and a moderate value in the central-parietal and right temporal lobes. Beta waves are not expected to be prominent during NREM sleep, as they are related to conscious thought, logical thinking, and focus [10]. The reduction of beta waves for DE could indicate a decrease in these processes.

In the delta band, there is, apart from in a spot in the center-right parietal lobe, higher PSD for NE. Similarly to our topographic maps, previous research has found a decrease in PSD for DE during NREM sleep at the source level within the delta frequency band in the parieto-occipital region and in the left brain hemisphere [4, 6].

The observed brain activity linked to DE has been inconsistent in literature [5], having shown both increased and decreased activity in similar areas. We add to this discussion using MEMD-filtered data and show that the observed results can be used for classification.

PSD proves its worth as both a classification feature and as a way to visually understand the differences between DE and NE. PSD combined with XGBoost scores within the standard deviation of the more complex EEGNeX, with both methods achieving accuracy and F-score above 0.9 and kappa above 0.7. PSD also makes it possible to visualize the difference between classes when paired with MEMD.

With PSD and XGBoost, MEMD-filtering improves the performance with 5.8% accuracy, 3.8% F-score, and 15.9% kappa. By selecting the three closest IMFs, and disregarding the others, we removed frequency components from the low-pass filtered data that were possibly negatively affecting the classification performance.

MEMD-filtering is an expensive filtering; we combined it with a relatively simple feature such as PSD to avoid additional expensive computation. Future work should continue to explore simple features paired with MEMD to further investigate the distinction between DE and NE.

To incorporate dream-decoding capabilities in a BCI, high-performance with a reduced channel subset is required. To achieve a substantial agreement, i.e. kappa score between 0.61-0.8, 19 channels are needed.

Dream reports collected from REM sleep differ from those collected from NREM sleep [13]. We analyze the differences between DE and NE in NREM sleep, and therefore an analysis of REM dream reports is necessary to further understand the nature of dreams.

References

- [1] King James Bible Online. Genesis 41:8, n.d. Available at: <https://www.kingjamesbibleonline.org/Genesis-41-8> (Accessed: 05-05-2024).

- [2] Tore A. Nielsen, Luc Laberge, Jean Paquet, Richard E. Tremblay, Frank Vitaro, and Jacques Montplaisir. Development of Disturbing Dreams During Adolescence and Their Relation to Anxiety Symptoms. *Sleep*, 23(6):1–10, September 2000.
- [3] Marine Ambar Akkaoui, Michel Lejoyeux, Marie-Pia d’Ortho, and Pierre A. Geoffroy. Nightmares in Patients with Major Depressive Disorder, Bipolar Disorder, and Psychotic Disorders: A Systematic Review. *Journal of Clinical Medicine*, 9(12):3990, December 2020.
- [4] 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, June 2017. Number: 6 Publisher: Nature Publishing Group.
- [5] Jing Zhang and Erin J. Wamsley. EEG predictors of dreaming outside of REM sleep. *Psychophysiology*, 56(7):e13368, 2019.
- [6] Serena Scarpelli, Aurora D’Atri, Anastasia Mangiaruga, Cristina Marzano, Maurizio Gorgoni, Cinzia Schiappa, Michele Ferrara, and Luigi De Gennaro. Predicting Dream Recall: EEG Activation During NREM Sleep or Shared Mechanisms with Wakefulness? *Brain Topography*, 30(5):629–638, September 2017.
- [7] Luis Alfredo Moctezuma, Felix Ipanaque, Marta Molinas, and Takashi Abe. Dream emotions identified without awakenings by machine and deep learning from electroencephalographic signals in REM sleep. In *2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE)*, pages 444–449. IEEE, 2023.
- [8] Xia Chen, Xiangbin Teng, Han Chen, Yafeng Pan, and Philipp Geyer. Toward reliable signals decoding for electroencephalogram: A benchmark study to EEGNeX. *Biomedical Signal Processing and Control*, 87:105475, January 2023.
- [9] J. Richard Landis and Gary G. Koch. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174, 1977. Publisher: [Wiley, International Biometric Society].
- [10] Priyanka A. Abhang, Bharti W. Gawali, and Suresh C. Mehrotra. Chapter 3 - Technical Aspects of Brain Rhythms and Speech Parameters. In Priyanka A. Abhang, Bharti W. Gawali, and Suresh C. Mehrotra, editors, *Introduction to EEG- and Speech-Based Emotion Recognition*, pages 51–79. Academic Press, January 2016.
- [11] Zahra M. Aghajan, Peter Schuette, Tony A. Fields, Michelle E. Tran, Sameed M. Siddiqui, Nicholas R. Hasulak, Thomas K. Tcheng, Dawn Eliashiv, Emily A. Mankin, John Stern, Itzhak Fried, and Nanthia Suthana. Theta Oscillations in the Human Medial Temporal Lobe during Real-World Ambulatory Movement. *Current Biology*, 27(24):3743–3751.e3, December 2017.
- [12] Khalid H. Jawabri and Sandeep Sharma. Physiology, Cerebral Cortex Functions. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), 2024.
- [13] Robert Stickgold and Erin J. Wamsley. Chapter 48 - why we dream. In Meir Kryger, Thomas Roth, and William C. Dement, editors, *Principles and Practice of Sleep Medicine (Sixth Edition)*, pages 509–514.e4. Elsevier, sixth edition edition, 2017.

Appendix C

**ESANN2024 Conference Paper: *Exploring
High- and Low-Density
Electroencephalography for a Dream
Decoding Brain-Computer Interface***

Exploring High- and Low-Density Electroencephalography for a Dream Decoding Brain-Computer Interface

Mithila Packiyathan¹, André Torvestad¹,
Luis Alfredo Moctezuma^{2*} and Marta Molinas¹

1- Department of Engineering Cybernetics,
Norwegian University of Science and Technology. Trondheim, Norway

2- International Institute for Integrative Sleep Medicine,
University of Tsukuba. Tsukuba, Ibaraki, Japan

Abstract. A high-performance real-time brain-computer interface system capable of identifying dreams has potential for healthcare applications. To address this, we use electroencephalogram (EEG) data from non-rapid eye movement sleep to classify *dream experience* and *no-experience*. Using 58 EEG channels, we achieve an accuracy of 0.94, an AUROC of 0.91, and a kappa score of 0.84, accomplished by first filtering the data through multivariate empirical mode decomposition followed by a combination of principal component analysis and common spatial patterns for feature extraction and K-nearest neighbors classifier. Interestingly, comparable results are obtained using 29 or 10 EEG channels selected by permutation-based channel selection.

1 Introduction

Neurological disorders and diseases often manifest during sleep before symptoms occur during waking conditions [1, 2]. Given that the content of dreams can be influenced by one’s mental health, research suggests that dreams could offer insight into the mental and emotional well-being of an individual [3, 4]. In light of this, the development of a tool capable of determining if a subject is dreaming and analyzing its emotional content could prove essential in healthcare facilities. Such a dream decoder could help psychologists diagnose and monitor patients with conditions such as depression, anxiety disorders, and various neurological disorders [4–6].

A brain-computer interface (BCI) is one way to build a dream decoder. Artificial intelligence (AI) and signal processing algorithms have contributed significantly to the advancement of BCI systems in general. However, achieving robustness remains a challenge, primarily due to the difficulty of obtaining a good trade-off between high performance and real-time speed, as well as the use of low-density electroencephalogram (EEG) systems [7, 8]. Currently, there are no BCI solutions for dream decoding in real-life applications, as research has focused mainly on analyzing dreams through the power spectrum, and only recently using machine learning (ML) techniques [5, 9–11].

*This work was partially supported by the Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship for Research in Japan: Fellowship ID P22716

Moving towards the development of a BCI system for dream decoding using EEG signals, we propose the use of multivariate empirical mode decomposition (MEMD) as a filter to then 1) extract features using principal component analysis (PCA) and common spatial patterns (CSP) and use them as input to the K-nearest neighbors (KNN), or 2) use a convolutional neural network (CNN) called EEGNeX, both to classify *dream experiences* and *no-experiences* during non-rapid eye movement (NREM) sleep. To identify the most important EEG channels and minimize their numbers while maintaining classification performance, we employ permutation-based channel selection. This technique also sheds light on the regions of the brain that are important to identify dream experiences.

2 Materials and Methods

2.1 EEG dataset

The dataset comprises high-density EEG recordings collected in 28 participants using a 58-electrode cap positioned according to the 10-10 international system with a sampling rate of 400 Hz and configured with a high-pass filter set at 0.1 Hz [9].

Participants were instructed to provide dream reports upon awakening. They were encouraged to verbally give all details of their experience, regardless of whether they had a dream experience or not. The study defined dreams as any mental experience recalled from sleep and the reports were judged by two independent raters, who labeled the reports as *dream experience* or *no-experience*. The awakenings were performed during sleep onset, N2 and rapid eye movement (REM) sleep stage. The awakenings during sleep onset were performed by awakening the subjects after 30, 60 or 90 seconds of elapsed sleep during the first hours of sleep, collecting up to 10 reports per subject. At least one hour after the last sleep onset report, the subjects were awakened when N2 was detected and then at least 30 minutes later when REM sleep was detected.

Our experiments consider the data collected during the N1 and N2 sleep stages, here referred to as NREM data. In total, we used 181 NREM epochs: 76% belonging to *dream experience* and 24% to *no-experience*.

2.2 Preprocessing, feature extraction and classification

To minimize high-frequency noise in the signals, a low-pass filter with a cutoff frequency of 45 Hz, experimentally selected, was applied to the EEG data. The EEG channels are referenced using the average between the left and right mastoids $\frac{(A1+A2)}{2}$.

We considered two experiment configurations: 1) the low-pass filtered data, referred to as the raw data. 2) The application of MEMD to low-pass-filtered recordings, choosing the three intrinsic mode functions (IMFs) closest to the original signal, in terms of Euclidean distance, and combining them to reconstruct a MEMD-filtered signal. For both configurations, the epochs

were segmented into 0.5-second instances, selected experimentally based on the classification performance using different segment sizes (0.5, 1, 2, 5 and 10 second segments).

For the classification process, two methods were used. The first method involved PCA and CSP for feature extraction and the KNN classifier. The number of neighbors in KNN was selected experimentally, checking the optimal value when $k = \{1, 10\}$. The second approach used EEGNeX, a CNN that outperforms existing state-of-the-art methods in EEG-based BCI classification tasks [12].

We used 5-fold cross-validation with a train-test split ratio of 80-20. The models were evaluated using accuracy, Fscore, precision, recall, area under the receiver operating characteristic curve (AUROC), and kappa.

To analyze which channels are more relevant and thus decrease the number of required EEG channels for the classification task, we used permutation-based EEG channel selection. For each channel, we randomly shuffled the EEG data within the channel, changing the information that was used during the training of the baseline model. The subdataset containing the shuffled channel was then evaluated using the baseline model. If the kappa score decreased when changing the data for a particular channel, it meant that the channel was important to obtain the baseline classification performance. Then, this process was repeated for all 58 channels. In this way, the algorithm identified the most important channels based on a comparison of the kappa score of the baseline model performance and the other performances; the performance delta.

3 Results

3.1 Classification performance under different experiment configurations

For the first approach to classifying dream experiences during NREM, we used 58 EEG channels with raw data and MEMD-filtered data. The process used KNN and the EEGNeX models.

To see the performance development when the number of channels is reduced, the classification task was also performed using 10 and 29 EEG channels.

The results of the classification using PCA, CSP, and KNN, and EEGNeX with both experiment configurations as input are presented in Fig. 1. The important brain regions obtained using the permutation-based channel selection are presented in the topographic maps in Fig. 2 where the 10 most important EEG channels are indicated with a yellow circle.

Fig. 1 shows a high classification performance for both KNN and EEGNeX; however, the standard deviation for the AUROC and kappa is higher for EEGNeX, showing less robustness than the KNN model. Furthermore, using MEMD-filtered data with 29 channels gives a higher performance than the raw data with 58 channels.

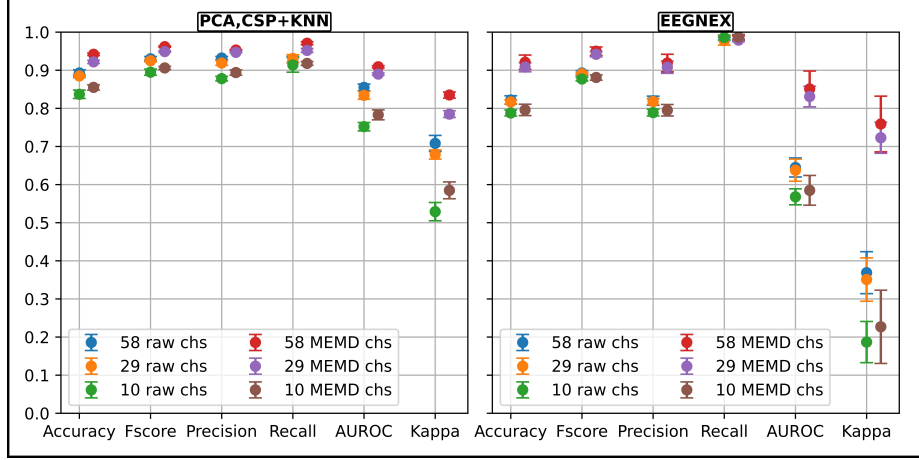


Fig. 1: Average classification performance of *dream experiences* and *no-experiences* using PCA, CSP and KNN (left), and EEGNeX (right), with both the raw data and the MEMD-filtered data.

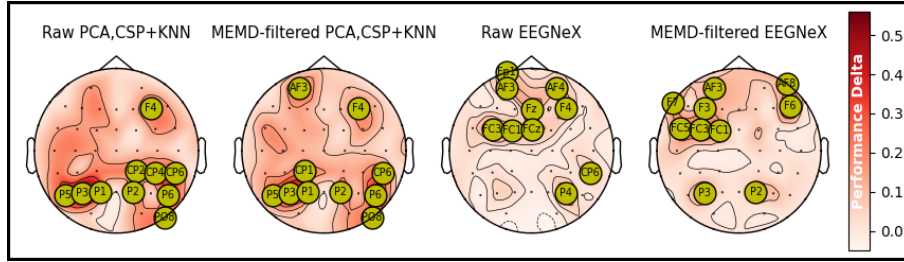


Fig. 2: Channel importance derived from classification experiments presented in Fig. 1. The 10 most important channels are indicated with yellow circles.

4 Discussion

We have presented experimental results using EEG signals for automatic classification of *dream experience* or *no-experience* during NREM.

In general, our results show that using MEMD-filtered data, instead of raw data, helps to obtain higher performance to classify when a subject is dreaming, both when using KNN or EEGNeX. Using KNN, performance increases 0.13 in kappa, 0.04-0.05 in accuracy, recall, and AUROC and 0.02-0.03 in Fscore and precision. For EEGNeX, 0.39 in kappa, 0.21 in AUROC, 0.1 in accuracy and precision, 0.06 in Fscore, and 0.002 in recall.

In the best case, we obtain up to 0.94 accuracy, 0.96 Fscore, 0.95 precision, 0.97 recall, 0.91 AUROC, and 0.84 kappa when using MEMD-filtered data and PCA, CSP for feature extraction and KNN classifier. This indicates that feature extraction techniques and the classical ML algorithm are better able to capture the features of EEG signals with dream content during NREM than EEGNeX.

Original signals are represented by three IMFs, which indicates that the removal of unnecessary frequency components makes the MEMD-filtered signal contain more important and valuable information for the classification process [13].

From the topographic maps in Fig. 2, we see the important regions of the dream classification experiments employing PCA, CSP and KNN. The 10 most important channels are located primarily in the parietal lobe, while some are distributed throughout the frontal and parieto-occipital regions. The parietal region is responsible for creating a spatial understanding of the surrounding environment and our location in it, and it is also where most sensory information is interpreted [14]. However, aspects related to prospective memory, future planning, and decision-making occur in the frontal lobe [14]. The map highlighting the parietal and frontal areas indicates that these are key areas to distinguish between *dream experiences* and *no-experiences*. When applying the MEMD-filtered data, the left parietal and front regions are emphasized.

For EEGNeX, the highlighted channels are predominantly located in the frontal, central, and parietal regions. The frontal region emerges as the most important area, suggesting that CNN captures activity in this area rather than in the parietal region, as observed with PCA, CSP, and KNN. The MEMD-filtered EEGNeX map is of more importance in the left frontal region. This could mean that dreams incorporate elements of logical reasoning, as research suggests that the left brain is associated with logical thinking [6, 15].

The topographic maps of PCA, CSP, and KNN exhibit notable differences compared to those of EEGNeX, both in important regions and in the performance delta. The maximum performance delta is lower when using EEGNeX (0.33) than when using KNN (0.56). That means that the subset of channels used is not as relevant for EEGNeX as for KNN, where the removal of specific channels greatly decreases performance.

Classification using PCA, CSP and KNN with MEMD-filtered data yields the highest performance; therefore, the associated topographic map is more reliable. The observations of the topographic map also align with the reported findings of frequency changes in the parieto-occipital, medial, and lateral frontoparietal regions while dreaming in NREM [5].

The classification using 10 channels instead of 58 results in less than 0.06 of a decrease in Fscore, precision, and recall, 0.09 of accuracy, 0.13 in AUROC and 0.25 in kappa, when using PCA, CSP, and KNN with MEMD-filtered data. This result holds promise as a first step towards the development of a BCI system for dream decoding applications. However, since the analysis is limited to NREM data, the findings for REM data may vary and need further investigation.

References

- [1] Geert Mayer, Svenja Happe, Stefan Evers, Wiebke Hermann, Sabine Jansen, Ulf Kallweit, Maria-Lucia Muntean, Dieter Pöhlau, Dieter Riemann, Michael Saletu, Melanie Schichl, Wolfgang J. Schmitt, Friederike Sixel-Döring, and Peter Young. Insomnia in neurological diseases. *Neurological Research and Practice*, 3(1):15, March 2021.

- [2] Sangeeta Jethwa, Ronit M Pressler, Didem Kaya, and Alexandre N Datta. Sleep architecture in neonatal and infantile onset epilepsies in the first six months of life: A scoping review. *European Journal of Paediatric Neurology*, 2022.
- [3] Marine Ambar Akkaoui, Michel Lejoyeux, Marie-Pia d'Ortho, and Pierre A. Geoffroy. Nightmares in Patients with Major Depressive Disorder, Bipolar Disorder, and Psychotic Disorders: A Systematic Review. *Journal of Clinical Medicine*, 9(12):3990, December 2020. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.
- [4] Anton Rimsh and Reinhard Pietrowsky. Analysis of dream contents of patients with anxiety disorders and their comparison with dreams of healthy participants. *Dreaming*, 31(4):303–319, 2021. Place: US Publisher: Educational Publishing Foundation.
- [5] 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, June 2017. Number: 6 Publisher: Nature Publishing Group.
- [6] Karen R Konkoly, Kristoffer Appel, Emma Chabani, Anastasia Mangiaruga, Jarrod Gott, Remington Mallett, Bruce Caughran, Sarah Witkowski, Nathan W Whitmore, Christopher Y Mazurek, et al. Real-time dialogue between experimenters and dreamers during rem sleep. *Current Biology*, 31(7):1417–1427, 2021.
- [7] Priyanka A. Abhang, Bharti W. Gawali, and Suresh C. Mehrotra. Chapter 8 - Brain–Computer Interface Systems and Their Applications. In Priyanka A. Abhang, Bharti W. Gawali, and Suresh C. Mehrotra, editors, *Introduction to EEG- and Speech-Based Emotion Recognition*, pages 165–177. Academic Press, January 2016.
- [8] Simanto Saha, Khondaker A. Mamun, Khawza Ahmed, Raqibul Mostafa, Ganesh R. Naik, Sam Darvishi, Ahsan H. Khandoker, and Mathias Baumert. Progress in Brain Computer Interface: Challenges and Opportunities. *Frontiers in Systems Neuroscience*, 15, February 2021. Publisher: Frontiers.
- [9] Jing Zhang and Erin J Wamsley. EEG predictors of dreaming outside of REM sleep. *Psychophysiology*, 56(7):e13368, 2019.
- [10] 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, June 2019. Publisher: Society for Neuroscience Section: Research Articles.
- [11] Luis Alfredo Moctezuma, Felix Ipanaque, Marta Molinas, and Takashi Abe. Dream emotions identified without awakenings by machine and deep learning from electroencephalographic signals in REM sleep. In *2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE)*, pages 444–449. IEEE, 2023.
- [12] Xia Chen, Xiangbin Teng, Han Chen, Yafeng Pan, and Philipp Geyer. Toward reliable signals decoding for electroencephalogram: A benchmark study to EEGNeX. *Biomedical Signal Processing and Control*, 87:105475, January 2023.
- [13] N. Rehman and D. P. Mandic. Multivariate empirical mode decomposition. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2117):1291–1302, May 2010.
- [14] Khalid H. Jawabri and Sandeep Sharma. Physiology, Cerebral Cortex Functions. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), 2024.
- [15] Michael C. Corballis. Left Brain, Right Brain: Facts and Fantasies. *PLoS Biology*, 12(1):e1001767, January 2014.