

Doctoral thesis

Doctoral theses at NTNU, 2021:7

Luis Alfredo Moctezuma

Towards Universal EEG systems with minimum channel count based on Machine Learning and Computational Intelligence

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical
Engineering
Department of Engineering Cybernetics



Norwegian University of
Science and Technology

Luis Alfredo Moctezuma

Towards Universal EEG systems with minimum channel count based on Machine Learning and Computational Intelligence

Thesis for the Degree of Philosophiae Doctor

Trondheim, August 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics



Norwegian University of
Science and Technology

NTNU

Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

© Luis Alfredo Moctezuma

ISBN 978-82-471-9693-9 (printed ver.)
ISBN 978-82-471-9970-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2021:7

Printed by NTNU Grafisk senter

To my family

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of Philosophiae Doctor (Ph.D.) at the Norwegian University of Science and Technology (NTNU). The research was conducted at the Department of Engineering Cybernetics (ITK) from June 2018 to August 2021.

During this time, I had the opportunity to attend conferences in various countries and collaborate with other universities, as well as work with Master's and Ph.D. students.

My first words of gratitude are for Professor Marta Molinas for sharing her time and passion for research with me during these years. Thank you for giving me the freedom to follow my ideas and for supporting them.

I would also like to thank Andres F. Soler, Erwin Habibzadeh, Chen Zhang, Alejandro A. Torres, and Pablo Muñoz for sharing their time and ideas. Thank you to all the staff of NTNU. Your work was essential throughout my studies at the university.

Thank you to all the anonymous reviewers of my conferences and journal papers. Their comments were truly useful and they helped me to raise the level of my work.

Mis ultimas palabras de gratitud son para mi esposa Laura Encarnación, gracias por soportarme y apoyarme siempre, te amo. Gracias a mi mamá y a mi papá por darme la vida y por guiarme siempre, sé que no ha sido fácil y que siempre han dado todo por mí y por mis hermanos.

LUIS ALFREDO MOCTEZUMA

AUGUST 2021, TRONDHEIM NORWAY

Abstract

The aim of this thesis is to move one step forward towards the concept of electroencephalographic (EEG) systems that can achieve the same objectives as high-density EEG with a minimum required number of channels. This requires EEG signal analysis, computational intelligence, and optimization techniques that can systematically identify the minimum number of channels that fulfills the objectives currently achieved with high-density EEG systems. Achieving this goal will pave the way towards the hardware-software realization of user-centric, easy-to-use, readily affordable EEG systems for universal applications. Enabling portability while ensuring performance of comparable or higher quality than that of high-density EEG will expand the accessibility of EEG to non-traditional users and personal applications moving EEG out of the lab. The application horizon will be expanded from experimental research to clinical use, to the gaming industry, intelligence and security sectors, education and daily use by people for self-knowledge.

The methods proposed in the thesis comprise the combination of feature extraction techniques and channel selection algorithms with optimization techniques that allow extracting the most essential information from a minimum set of required EEG channels that were tested in two cases-studies: **Epileptic seizure classification**, and **EEG-based biometric systems**. The Discrete Wavelet Transform (DWT) and Empirical Mode Decomposition (EMD) were used to decompose EEG signals into different frequency bands and then four features were computed for each sub-band, the *Teager* and *Instantaneous* energies and the *Higuchi* and *Petrosian* fractal dimensions.

For the optimization stage, non-dominated sorting genetic algorithms (NSGA) were used for channel selection, using binary values to represent the channels in

the chromosomes, 1 if the channel is used in the classification and optimization process, and 0 if not. Additional genes to represent important parameters for the classifiers were added using integer and decimal values.

For Case-study 1, NSGA-III selected one or two channels from a set of 22 for epileptic seizure classification, obtaining an accuracy of up to 0.98 and 1.00, respectively, using EMD/DWT-based features.

For Case-study 2, a task-independent, resting-state-based biometric system using Local Outlier Factor (LOF)- and DWT-based features showed a True Acceptance Rate (TAR) of up to 0.993 ± 0.01 and a True Rejection Rate (TRR) of up to 0.941 ± 0.002 using only three channels selected by NSGA-III from a set of 64.

The results presented herein can be considered to be a first proof-of-concept, showing that it is possible to reduce the number of required EEG channels for classification tasks and opens the way to explore these methods on other neuroparadigms. This will lead to reduced real-time computational costs for EEG signal processing, removing task-irrelevant and redundant information, as well as reducing the preparation time for use of the EEG headsets.

The results of such a reduction in the number of required EEG channels will make possible a low-power hardware design, expanding the range of EEG-based applications from clinical diagnosis and research to health-care, to non-medical applications that can improve our understanding of cognitive processes, learning and education and to the discovery of current hidden/unknown properties behind ordinary human activity and ailments.

Contents

Abstract	i
List of Abbreviations	vii
List of Tables	xi
List of Figures	xiii
1 Introduction	1
1.1 Motivations for the research and knowledge gaps	1
1.2 Research Questions and Objectives	3
1.3 Contributions	5
1.4 Structure of the thesis	8
2 Fundamentals of Electroencephalography, evolution, and open challenges	11
2.1 Electroencephalography	11
2.1.1 Mechanisms of EEG generation	12
2.1.2 Normal and abnormal EEG	12
2.1.3 EEG signal acquisition	16
2.1.4 A brief comparison with other brain signal acquisition methods	17
2.1.5 International EEG electrode placement systems	18
2.1.6 Consumer-grade low-density EEG headsets	19
2.1.7 Using brain signals for control purposes	21
2.2 EEG paradigms	23
2.2.1 Event-related potentials and P300	23
2.2.2 Resting-state	24
2.3 Current and future trends in EEG	26

3	Materials and Methods	29
3.1	Improving the signal-to-noise ratio	29
3.2	Data analysis	31
3.2.1	Empirical Mode Decomposition	31
3.2.2	Discrete Wavelet Transform	34
3.3	Data features	37
3.3.1	Energy distribution	37
3.3.2	Fractal dimension	39
3.4	Computational intelligence methods for classification	42
3.4.1	Multi-class classification	42
3.4.2	One-class classification	43
3.4.3	Evaluation of classifier performance	47
3.5	Channel reduction and selection	48
3.5.1	Greedy algorithms	49
3.5.2	Multi-objective optimization methods	50
3.6	Description of datasets used in the thesis	53
3.6.1	CHB-MIT	53
3.6.2	EEGMMIDB	54
3.6.3	P300-speller	56
3.7	Methods proposed in the thesis	57
3.7.1	Pre-processing, feature extraction and classification	57
3.7.2	General overview of the proposed method	59
3.8	Hardware and software tools used in the thesis	61
4	Case study 1: Channel count optimization for Epileptic seizure classification	63
4.1	Introduction	63
4.2	State-of-the-art	64
4.3	Definition of the problem to optimize	66
4.4	Channel selection for Epileptic-seizure classification with EMD-based features	68
4.5	Channel selection for Epileptic-seizure classification with DWT-based features	74
4.6	Discussion	76

5	Case study 2: Channel count optimization for EEG-based biometric systems	83
5.1	Introduction	83
5.2	State-of-the-art	85
5.3	First approach using a two-stage classification process	87
5.3.1	Defining the problem to optimize	89
5.3.2	Solving the four-objective optimization problem using NSGA-II with subjects 1-13 as non-intruders and 14-26 as intruders.	90
5.3.3	Solving the four-objective optimization problem using NSGA-II with subjects 14-26 as non-intruders and subjects 1-13 as intruders.	91
5.3.4	NSGA-III for solving the four-objective optimization problem.	95
5.3.5	Testing the proposal in 10 random subdivisions of subjects using NSGA-II and NSGA-III.	96
5.4	Discussion	99
5.5	Second approach, using a one-stage one-class algorithm	101
5.5.1	Defining the problem to optimize	103
5.5.2	Channel selection using NSGA-III and OCSVM for EEG signals for the resting-state with the eyes open	104
5.5.3	Channel selection using NSGA-III and LOF for EEG signals for the resting-state with the eyes open	107
5.5.4	Channel selection using NSGA-III and LOF for EEG signals for the resting-state with the eyes closed	111
5.6	Discussion	115
6	Conclusions and future work	123
6.1	Summary of findings	123
6.1.1	Feature extraction and channel count optimization for epileptic seizure classification	123
6.1.2	Channel count optimization for EEG-based biometric systems	124
6.2	Conclusion of the thesis contributions	125
6.3	Future work	127

References

131

List of Abbreviations

2D	Two-dimensional.
3D	Three-dimensional.
ABC	Artificial bee colony.
AEMD	Adaptive Empirical Mode Decomposition.
BCI	Brain-Computer Interfaces.
BFPA	Binary flower pollination algorithm.
BSS	Blind source separation.
CAR	Common Average Reference.
CNN	Convolutional neural network.
CNN-GRU	Convolutional neural network gated recurrent units.
CRR	Correct recognition rate.
CT	Computerized tomography.
DMD	Dynamic mode decomposition.
DT	Decision tree.
DWT	Discrete Wavelet Transform.

Ear-EEG	In-the-ear Electroencephalography.
ECG	Electrocardiograph.
EEG	Electroencephalography.
EEGMMIDB	Motor movement/imagery dataset.
EEMD	Ensemble Empirical Mode Decomposition.
EMD	Empirical Mode Decomposition.
EMG	Electromyography.
EWT	Empirical wavelet transform.
FAR	False acceptance rate.
fMRI	Functional magnetic resonance imaging.
FN	False negatives.
FP	False positives.
FT	Fourier transform.
GA	Genetic algorithms.
GNMM	Genetic neural mathematics method.
HTER	Half total error rate.
ICA	Independent component analysis.
iEEG	Intracranial Electroencephalography.
IMFs	Intrinsic Mode Functions.
KNN	k-nearest neighbors.
LAP	Laplacian Filter.

LDA	Linear discriminant analysis.
LOF	Local Outlier Factor.
LRD	Local reachability density.
LS-SVM	Least-square support vector machine.
MEG	Magnetoencephalography.
MEMD	Multivariate Empirical Mode Decomposition.
MI	Mutual information.
MOEA/D	Multi-objective evolutionary algorithms based on decomposition.
MOOP	Multi-objective optimization problem.
MRI	magnetic resonance imaging.
NB	Naive Bayes.
NN	Neural networks.
NSGA	Non-dominated sorting genetic algorithm.
OCC	One-class classification.
OCSVM	One-class support vector machine.
PCA	Principal component analysis.
PET	Positron emitted tomography.
PSR	Phase space representation.
RBF	Radial basis function.
RF	Random Forest.
RSNs	Resting-state networks.

SVM	Support vector machine.
TAR	True Acceptance Rate.
TIRDA	Temporal intermittent rhythmic delta activity.
TLE	Temporal-lobe epilepsy.
TN	True negatives.
ToC	Third-order cumulant.
TP	True positives.
TRR	True Rejection Rate.

List of Tables

3.1	Details of the epileptic-seizure data presented in [218].	55
4.1	Accuracy obtained using EMD for feature extraction with NSGA-II and NSGA-III for EEG channel selection (subjects 1-12).	71
4.2	Accuracy obtained using EMD for feature extraction with NSGA-II and NSGA-III for EEG channel selection (subjects 13-24).	72
4.3	Accuracy obtained using DWT for feature extraction with NSGA-II and NSGA-III for EEG channel selection (subjects 1-12).	75
4.4	Accuracy obtained using DWT for feature extraction with NSGA-II and NSGA-III for EEG channel selection (subjects 13-24).	76
4.5	Comparison of relevant existing methods for epileptic-seizure classification using the CHB-MIT Scalp EEG dataset presented in [218].	79
4.6	Comparison of several relevant existing methods for epileptic-seizure classification using different datasets.	80
5.1	TAR, TRR, and accuracy for subject identification and authentication with EEG data from all channels using different nu and gamma values for one-class SVM.	88
5.2	TAR, TRR, and accuracy values obtained for the Pareto-front for four objectives solved with NSGA-II using subjects 1-13 as non-intruders.	93
5.3	TAR, TRR, and accuracy values obtained for the first 30 EEG channels in the Pareto-front for four objectives solved with NSGA-II using subjects 14-26 as non-intruders.	94

5.4 TAR, TRR, and accuracy values obtained in the Pareto-front when using 7-15 EEG channels with four objectives solved with NSGA-III using subjects 1-13 as non-intruders and 14-26 as intruders and vice-versa. 96

5.5 Mean TAR, TRR, and accuracy values obtained in the Pareto-front when using 7-15 EEG channels validated in 10 random subdivisions of all the subjects, using 50% as intruders and 50% as non-intruders. 98

5.6 Average TARs and TRRs for subject detection with EEG data from 64 channels and 109 subjects using different parameters for OCSVM and LOF, with EMD- and DWT-based features. 102

5.7 TARs and TRRs obtained for the first five EEG channels in the Pareto-front for three objectives solved with NSGA-III using EMD- and DWT-based features with OCSVM. 105

5.8 TARs and TRRs obtained for the first seven EEG channels in the Pareto-front for three objectives solved with NSGA-III using EMD-based and DWT-based features and LOF. 110

5.9 TARs and TRRs obtained with LOF for the first seven EEG channels in the Pareto-front for three objectives solved with NSGA-III using EMD- or DWT-based features and the resting-state with the eyes closed. 114

List of Figures

1.1	Flowchart of contributions of papers to each Research Question.	5
1.2	General overview of the methodology and contributions to the thesis.	10
2.1	EEG electrode placement methods: bipolar (a) and monopolar (b).	16
2.2	The original figure illustrating the international 10-20 system. Note that the electrodes are erroneously located inside the skull on the surface of the cortex [2].	19
2.3	Timeline of the evolution of EEG systems and relevant consumer-grade wearable EEG headsets.	20
2.4	FlexEEG concept. FlexEEG moves from X_1 to X_2 to capture sources S_1 and S_2 [58].	22
2.5	Schematic representation of certain ERP components after the onset of a visual stimulus [72].	24
2.6	Topography of four microstate maps from [92]. Map areas of opposite polarity are coded in red and blue using a linear color scale. The left ear is to the left and the nose is at the top	26
3.1	Stages of the methodology followed in the thesis.	30
3.2	IMFs plus residue (Sub-fig. 3.2a) obtained from the synthetic signal presented in sub-fig. 3.2b, as well as the reconstructed signal using all the IMFs (Sub-fig. 3.2c) and three IMFs selected using the Minkowski distance plus the residue (Sub-fig. 3.2d).	35

3.3	Details and approximation coefficients extracted from the original signal using DWT with four levels of decomposition and the mother wavelet biorthogonal 1.3.	38
3.4	Teager and Instantaneous energy distribution of EMD and DWT sub-bands from Figs. 3.2 and 3.3.	40
3.5	Higuchi and Petrosian fractal dimension of EMD and DWT sub-bands from Figs. 3.2 and 3.3.	41
3.6	Decision boundaries in OCSVM for a random dataset with outliers	45
3.7	Decision boundaries with LOF for a random dataset with outliers	46
3.8	An illustrative example of the NSGA-II procedure [211].	52
3.9	Reference points of NSGA-III in a three-objective optimization problem.	53
3.10	Example of the raw EEG data of <i>C3-P3</i> , <i>T7-FT9</i> and <i>C4-P4</i> channels from the third instance of Patient 1 of the CHB-MIT dataset. . . .	54
3.11	Example of the raw EEG data of <i>F5</i> , <i>T8</i> and <i>T10</i> channels of the first instance of subject 1 of the EEGMMIDB dataset.	56
3.12	Protocol design for recording positive or negative feedback-related responses in the P300-speller dataset [220].	57
3.13	Example of the raw EEG data of <i>P7</i> , <i>P8</i> and <i>T8</i> channels of the first instance of subject 1 of the P300-speller dataset.	58
3.14	Flowchart summarizing feature extraction using DWT.	59
3.15	Flowchart summarizing the feature extraction procedure using EMD.	59
3.16	Flowchart of the procedure followed for EEG signal classification.	59
3.17	Example of chromosome representation and flowchart of the optimization process for parameter optimization and EEG channel selection using NSGA-III.	60
4.1	Complete process for EEG channel selection using NSGA-II or NSGA-III for epileptic-seizure classification.	67
4.2	EEG Channel Selection for epileptic seizure classification of patient 1 using EMD-based features. Comparison between NSGA-II and the backward-elimination algorithm.	69
4.3	Four EEG Channel subsets selected by NSGA-II (a) and backward-elimination (b) for epileptic-seizure classification in patient 1. . .	70

4.4	EEG Channel selection for epileptic-seizure classification of patient 19 using EMD-based features. Comparison between NSGA-III and the backward-elimination algorithm.	73
4.5	Comparison of the most used classifiers by NSGA-II (left) and NSGA-III (right) for the 24 patients using EMD-based feature extraction.	73
4.6	Comparison of the most-used classifiers by NSGA-II (left) and NSGA-III (right) for the 24 patients using DWT-based feature extraction.	77
5.1	Flowchart of the first approach for intruder detection and subject identification.	88
5.2	Example of the complete process for EEG channel selection using NSGA-II, including the chromosome representation using 56 genes for the EEG channels and eight for the <i>nu</i> and <i>gamma</i> parameters.	90
5.3	Four different views of the results obtained with NSGA-II using subjects 1-13 as non-intruders and 14-26 as intruders.	92
5.4	Relevant EEG channel subsets in the Pareto-front for four objectives using NSGA-II, considering subjects 14-26 as intruders in the previous experiment and subjects 1-13 as intruders in the current experiment.	95
5.5	Relevant EEG channel subsets in the Pareto-front for four objectives using NSGA-III, considering subjects 14-26 as intruders in the previous experiment and subjects 1-13 as intruders in current experiment.	97
5.6	TARs and TRRs obtained using various numbers of neighbors with the LOF <i>k-d tree</i> algorithm and DWT-based features.	103
5.7	Chromosome representation and flowchart of the optimization process for EEG channel selection using NSGA-III and LOF.	104
5.8	Frontal and aerial view of the TARs and TRRs obtained in the channel-selection process using EMD-based features (a) and DWT-based features (b) with OCSVM.	106

5.9	Set of one to five channels found during the optimization process for creating the biometric system with OCSVM using EMD-based features (<i>a</i>) or DWT-based features(<i>b</i>) and the resting-state with the eyes open.	108
5.10	Frontal and aerial view of the TARs and TRRs obtained in the channel-selection process using EMD-based features (a), and DWT-based features (b) with LOF.	109
5.11	Average distribution of the algorithms and number of neighbors used in the optimization process with EMD-based features (a) and DWT-based features (b).	110
5.12	Average distribution of the algorithms and number of neighbors used for the results in the Pareto-front of the optimization process with EMD-based features (a) and DWT-based features (b).	111
5.13	Set of one to seven channels found during the optimization process for creating the biometric system with LOF and EMD-based features (<i>a</i>) or DWT-based features(<i>b</i>) for the resting-state with the eyes open.	112
5.14	Frontal and aerial view of the TARs and TRRs obtained in the channel-selection process using EMD- (a) and DWT-based features (b) for the resting-state with the eyes closed, using LOF.	113
5.15	Average distribution of the algorithms and number of neighbors used in the optimization process with EMD-based features (<i>a</i>) and DWT-based features (<i>b</i>) using EEG signals for the resting-state with the eyes closed.	114
5.16	Average distribution of the algorithms and number of neighbors used for the results in the Pareto-front of the optimization process with EMD-based features (<i>a</i>) and DWT-based features (<i>b</i>) using EEG signals for the resting-state with the eyes closed.	115
5.17	Set of one to seven channels found during the optimization process for creating the biometric system with LOF using EMD-based features (<i>a</i>) or DWT-based features(<i>b</i>) and the resting-state with the eyes closed.	116

Chapter 1

Introduction

The objective of this thesis is to move one step forward towards a concept of electroencephalographic (EEG) systems, with a minimum number of channels, that can contribute to the realization of low-cost real-time applications, thus enabling the portability of EEG headsets while retaining quality comparable to, or higher than, that of high-density EEG-based systems. This requires EEG signal analysis, computational intelligence, and optimization techniques that can systematically identify a minimum number of EEG channels that fulfill the objectives currently achieved using high-density EEG systems. To this end, the thesis proposes to systematically apply greedy algorithms and multi-objective optimization methods for which targeted algorithms were developed and implemented to solve the problem of channel selection and parameter optimization.

*This Ph.D. research is part of a larger project, **David and Goliath: single-channel EEG unravels its power through adaptive signal analysis**, which aims to identify an optimal minimum EEG channel count for wearable EEG solutions for universal applications. This thesis contributes to this goal by achieving one of the three objectives of **David and Goliath: Optimization-based channel reduction**.*

This Chapter provides an overview of the main contributions of the thesis, including an overview of the publications associated with the work.

1.1 Motivations for the research and knowledge gaps

Consumer-wearable EEG technologies have experienced steady growth, with a growing number of devices with a reduced number of EEG channels available for personal uses, such as meditation, relaxation training, motor imagery, and

the control of moving objects [1]. As a result, people today can measure their own brain signals outside medical laboratories due to the proliferation of low-cost wireless headset EEG devices with varying numbers and configurations of EEG channels, with dry or wet electrodes, using the 10-5, 10-10, or 10-20 international system [2–5].

There are a number of critical open issues (i.e., real-time use, quality of recordings, portability, ease-of-use, and user orientation) that are as yet unexplored [6]. One of the unexplored aspects that can influence these issues is electrode placement, which in most EEG devices is fixed and inflexible, depending on the targeted application/s. For real-time applications, high-quality/high-density EEG devices are computationally costly and the applications are very limited. The existing wireless portable devices, with fixed electrode placement, also have limitations. Depending on the related task, neuro-paradigm used, and age and sex of the subject, the most relevant features of brain signals may be obtained at locations different from those of the electrodes in the scalp [7–10].

Most EEG devices available on the market were designed for a set of related tasks and neuro-paradigms and in general, are found to be reliable only within the context of such tasks and neuro-paradigms. The accuracy and reliability of these systems for prolonged and repeated measurements have not been well-established and a rigorous comparative investigation of the different portable solutions is not yet available. Most importantly, it is not clear whether the limited number of channels and their fixed localization can provide sufficient data and anatomical coverage to obtain the neural signatures necessary for the given tasks, as these concepts are not supported by openly available research. They are based on proprietary technology backed by protected research or IP not available to the public. Essentially, this is because both electrode localization and the number of electrodes are task-dependent [1, 7, 11]. Moreover, these commercial solutions are intended to only support the tasks/paradigms for which they were designed.

The current state-of-the-art consists of methods to decompose and extract information from brain signals using wet or dry EEG electrodes. However, the behavior of brain signals varies depending on the neuro-paradigm, the technology of the device, and the specific characteristics of the subject (culture, age, IQ/cognition level, sex, etc.) [7]. In addition, because of the non-stationary/non-

linear nature of brain signals, it is necessary to create a method with multiple sub-steps to extract the most essential features that can help identify the targeted tasks (e.g., event detection and classification). If such advances are plausible, the performance of Brain-Computer Interfaces (BCI) can increase and applications will span-new areas of research, from medical applications to industrial security systems.

The major motivations and objectives behind the reported research work in this thesis are based on the following knowledge gaps that were identified based on the literature review in Chapter 3, 4, and 5.

- **Knowledge gap 1:** High-density EEG is challenged by high computational cost, immobility of the equipment, and the use of inconvenient conductive gels. Several studies have explored reducing the number of electrodes required for a certain task and electrode placement towards real-time EEG signal processing. Most were based on *a priori* or empirical knowledge. Consolidated studies based on systematic searches aiming to reduce the EEG channel count required for a given task are not currently available. Such an approach can be achieved by applying systematic search algorithms and optimization techniques for identifying the most relevant electrode position/placement for a given paradigm.
- **Knowledge gap 2:** There is currently insufficient knowledge of feature extraction for better representation of low-density EEG signals that can also reduce the computational cost. Most research on feature extraction has been based on high-density EEG.
- **Knowledge gap 3:** There are several proposed methods for feature extraction and classification in the state-of-the-art, but they are used for specific tasks and the results may vary for different tasks. In other words, the methods are neither generalized nor replicable for different applications.

1.2 Research Questions and Objectives

The objective of this thesis is the analysis of EEG signals with high-density and low-density channel arrays to compare their performance in two case studies: **Epileptic seizure classification** and **EEG-based biometric systems**. For this

objective, it was necessary to create various algorithms for channel reduction and selection to ensure a reliable method to extract the most relevant information from the raw EEG signals.

The data used in the experiments were extracted from public repositories to ensure the quality of the analysis. The stages of the methodology include noise removal, feature extraction, optimization techniques, which were all explored and combined to effectively represent large raw EEG signals for classification tasks. These steps aim to improve the quality and response time of the machine-learning based models.

Based on the analysis of the knowledge gaps presented, the thesis concentrated on the following three Research Questions:

- **Research Question 1: Channel Dimensionality Reduction** Can the number of EEG channels required for classification tasks be reduced while increasing, or at least maintaining, the accuracy relative to the use of high-density EEG?
- **Research Question 2: Data Dimensionality Reduction** Can a few useful features be sufficient to effectively represent large raw EEG signals for classification and thus accelerate the computational performance of the used methods for classifying different tasks?
- **Research Question 3: Generalizing the Methodology** Can the same process of feature extraction, classification, and channel selection be generalized or at least used (expand the methodology) for different problems related to the classification of EEG signals (i.e., task-dependent and task-independent)?

Testing state-of-the-art methods on certain specific problems and conditions will make it possible to propose new methods to tackle the feature extraction and dimensionality-reduction problem associated with EEG signals. Then, if the number of required channels can be reduced, it will be possible to draw certain conclusions and entertain the possibility of a new type of EEG headset. During this process, it will be necessary to repeat the methodology for different task-dependent and task-independent neuro-paradigms using EEG signals and analyze their behavior, trying to draw more general conclusions.

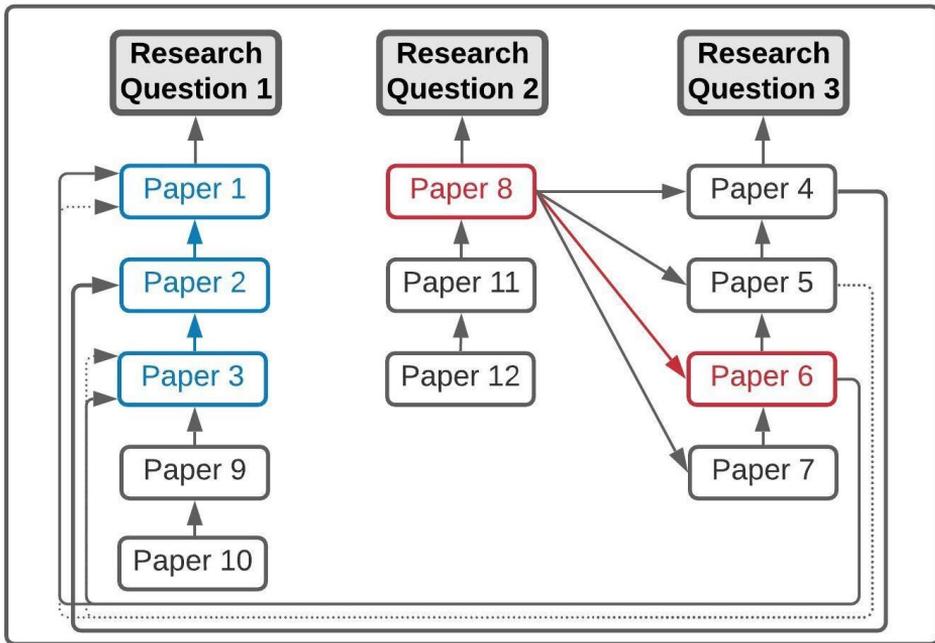


Figure 1.1: Flowchart of contributions of papers to each Research Question.

1.3 Contributions

Fig. 1.1, presents a flowchart of the contributions to the thesis for each research question. Paper 8 presented the first approach using a feature extraction process based on the Empirical Mode Decomposition (EMD), which was later compared to the second approach of the thesis, consisting of features based on the Discrete Wavelet Transform (DWT), introduced in Paper 6. This connection is indicated by the red rectangles and arrows. The method presented in paper 8 was used in most of the subsequently published papers, indicated by the arrows connecting the papers that contributed to Research Question 3. All the papers presented in Fig. 1.1 contributed to the achievement of the objectives, but papers 1, 2, and 3 presented the final contributions, as they presented the use of greedy and non-dominated sorting genetic algorithm (NSGA)-based algorithms for channel selection and parameter optimization, and are the most relevant contributions to this thesis.

The following articles and conference papers were published during the Ph.D. and are directly related to the thesis:

Journal articles

1. Moctezuma, Luis Alfredo, Marta Molinas. "Towards a minimal EEG channel array for a biometric system using resting-state and a genetic algorithm for channel selection". *Scientific Reports* (2020). DOI: [10.1038/s41598-020-72051-1](https://doi.org/10.1038/s41598-020-72051-1)
2. Moctezuma, Luis Alfredo, Marta Molinas. "EEG Channel-selection method for epileptic-seizure classification based on multi-objective optimization". *Frontiers in neuroscience* (2020). DOI: [10.3389/fnins.2020.00593](https://doi.org/10.3389/fnins.2020.00593)
3. Moctezuma, Luis Alfredo, Marta Molinas. "Multi-objective optimization for EEG channel selection and accurate intruder detection in an EEG-based subject identification system". *Scientific Reports* (2020). DOI: [10.1038/s41598-020-62712-6](https://doi.org/10.1038/s41598-020-62712-6)
4. Moctezuma, Luis Alfredo, Marta Molinas. "Classification of low-density EEG epileptic seizures by energy and fractal features based on EMD". *Journal of Biomedical Research* (2019). DOI: [10.7555/JBR.33.20190009](https://doi.org/10.7555/JBR.33.20190009)

Peer-reviewed Conferences

5. Moctezuma, Luis Alfredo, and Marta Molinas. "Event-related potential from EEG for a two-step Identity Authentication System". *IEEE international conference on industrial informatics, indin'19* (2019):. DOI: [10.1109/INDIN41052.2019.8972231](https://doi.org/10.1109/INDIN41052.2019.8972231)
6. Moctezuma, Luis Alfredo, and Marta Molinas. "Subject identification from low-density EEG-recordings of resting-states: A study of feature extraction and classification". In *Future of Information and Communication Conference (FICC)*, 2019:.. DOI: [10.1007/978-3-030-12385-7_57](https://doi.org/10.1007/978-3-030-12385-7_57)
7. Moctezuma, Luis Alfredo, and Marta Molinas. "Sex differences observed in a study of EEG of linguistic activity and resting-state: Exploring optimal EEG channel configurations". In *the 7th International Winter Conference on Brain-Computer Interface*, 2019. DOI: [10.1109/TWW-BCI.2019.8737312](https://doi.org/10.1109/TWW-BCI.2019.8737312)
8. Moctezuma, Luis Alfredo, and Marta Molinas. "EEG-based Subjects Identification based on Biometrics of Imagined Speech using EMD". In *International Conference on Brain Informatics*. Springer, Cham, 2018:.. DOI: [10.1007/978-3-030-05587-5_43](https://doi.org/10.1007/978-3-030-05587-5_43)

Peer-reviewed abstracts

9. Soler-Guevara, Andres Felipe, **Luis Alfredo Moctezuma**, Eduardo Giraldo, Marta Molinas. "EEG channel-selection method based on NSGA-II for source localization". The 4th HBP Student Conference on Interdisciplinary Brain Research (2020):.
10. Moctezuma, Luis Alfredo, Andres Felipe Soler, Erwin H. T. Shad, Marta Molinas, Alejandro A. Torres-Garcia. "David versus Goliath: Low-density EEG unravels its power through adaptive signal analysis - FlexEEG". The 4th HBP Student Conference on Interdisciplinary Brain Research (2020):.

Book Chapters

11. Moctezuma, Luis Alfredo, and Marta Molinas. "EEG-based subject identification with multi-class classification". In Biosignal Processing and Classification using Computational Learning and Intelligence (2020). (In press)
12. Torres-Garcia Alejandro A., Omar Mendoza-Montoya, Marta Molinas, Mauricio Antelis, **Luis Alfredo Moctezuma**. "Pre-processing and Feature Extraction". In Biosignal Processing and Classification using Computational Learning and Intelligence (2020). (In press)

Other contributions

Contributions written during the Ph.D. but not directly related to the thesis:

Peer-reviewed Conferences

13. Alejandro A. Torres-Garcia, **Luis Alfredo Moctezuma** and Marta Molinas. "Assessing the impact of idle state type on the identification of RGB color exposure for BCI". In 13th International Joint Conference on Biomedical Engineering Systems and Technologies (2020):. [10.5220/0008923101870194](https://doi.org/10.5220/0008923101870194)
14. Torres-Garcia Alejandro A., **Luis Alfredo Moctezuma**, Sara Asly and Marta Molinas. "Discriminating between color exposure and idle state using EEG signals for BCI application". In 7-th edition of the International Conference on e-Health and Bioengineering (2019):. DOI: [10.1109/EHB47216.2019.8969919](https://doi.org/10.1109/EHB47216.2019.8969919)

15. Asly, Sara, **Luis Alfredo Moctezuma**, Monika Gilde, Marta Molinas. “Towards EEG-based signals classification of RGB color-based stimuli”. In 8th Graz Brain-Computer Interface Conference 2019 (2019):. DOI: [10.3217/978-3-85125-682-6-61](https://doi.org/10.3217/978-3-85125-682-6-61)
16. Moctezuma, Luis Alfredo, Marta Molinas, AA Torres Garcia, Luis Villaseñor Pineda, and Maya Carrillo. “Towards an API for EEG-based imagined speech classification”. In International Conference on Time Series and Forecasting. 2018:. Proceedings at itise.ugr.es/ITISE2018_Papers_Vol_3.pdf

Peer-reviewed abstracts

17. Torres-Garcia Alejandro A., Marta Molinas, **Luis Alfredo Moctezuma**. “Towards a BCI based on Color Exposure Recognition”. The 4th HBP Student Conference on Interdisciplinary Brain Research (2020):.

1.4 Structure of the thesis

Chapter 1 introduces the work in this thesis and the knowledge gaps and research motivations are listed. The contributions to the thesis are presented in a flowchart, showing how the published papers are connected to the defined research questions. Finally, a list of the results published separately in journals, conference papers, and abstracts is presented, including contributions directly related to the thesis, as well as published results not directly related to the objective of the thesis.

In Chapter 2, the fundamentals of EEG, a brief history of EEG and EEG signal analysis, international EEG standards, and the two paradigms of interest for this thesis are presented, which are event-related potentials (ERPs) and the resting-state.

Chapter 3 presents the fundamentals of the methods used for EEG signal analysis, which include EMD and DWT and the reasons for choosing them in this study. This is followed by a presentation of how the energy distribution and fractal dimension feature functions in the context of feature extraction. Then, the multi-class and one-class classifiers tested and the metrics for evaluating performance are presented. A description of NSGA and how it is used for solving multi-objective optimization problems is provided in this Chapter.

The description of the datasets used in the two investigated scenarios are also presented in Chapter 3, in which a general flowchart of the proposed methodology

for feature extraction, classification, and optimization process handled by NSGA algorithms is presented and explained.

Chapter 4 presents Case-study 1, which is focused on validation of the methods for channel count minimization in a case of epileptic seizure classification using multi-class classification. Two different approaches for representing the epileptic-seizure and seizure-free EEG signals are presented. The first approach is based on DWT and the second EMD. Using these two approaches, the EEG data is decomposed into different frequency sub-bands and then a set of four features per sub-band is calculated. Once this is carried out, a multi-objective optimization process is organized and solved using NSGA-II and NSGA-III. The objective of the optimization process is to increase the accuracy of the machine-learning models for classification of epileptic seizures and seizure-free periods while decreasing the number of required EEG channels. Finally, a discussion about the results obtained is presented and they are compared with those of other approaches using the same datasets and other datasets.

Case-study 2, which consists of a proposal for a biometric system with minimal channel count, is presented in Chapter 5. Two different approaches are presented, a two-stage approach consisting of a multi-class classification layer and then a one-class classifier, and a second approach using only one-class classifiers. The experiments are compared using different methods for feature extraction and NSGA-II or NSGA-III for solving the optimization process. As in Chapter 4, the work in Chapter 5 also has the objective of minimizing or reducing the number of required EEG channels while increasing or maintaining classification accuracy, which in this case consist of increasing the True Acceptance Rate (TAR) of the subjects with access and the True Rejection Rate (TRR) of intruders.

Finally, Chapter 6 presents the conclusions of the thesis and identifies opportunities for further work.

Fig. 1.2, presents an overview of the methods proposed and used to achieve the objectives of the thesis. As will be explained later, all the EEG datasets used are freely available to the public at no cost, but the number of subjects, the number of channels, etc., were considered to select them (**a**). In the feature extraction stage (**b**), two methods were used to decompose the EEG signals into different frequency bands and then a set of four features were calculated to obtain a single

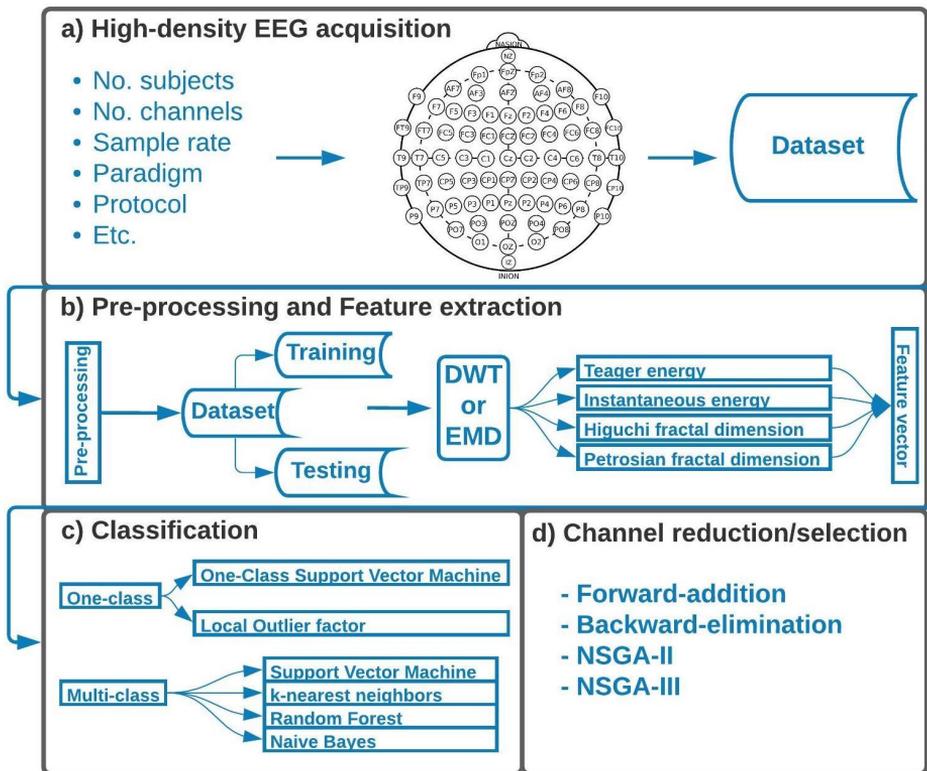


Figure 1.2: General overview of the methodology and contributions to the thesis.

feature vector for each instance. Then, depending on the case study, one-class or multi-class classifiers were developed and validated. In each case, different methods were used to compare their performance (**c**). During this work, four different methods for channel reduction and selection were developed. This stage in the methodology (**d**) is the main focus of the thesis and, therefore, is where the main contributions of the thesis can be found.

Chapter 2

Fundamentals of Electroencephalography, evolution, and open challenges

This Chapter presents the main concepts related to EEG signals, signal analysis, the evolution of EEG technology, the two paradigms of interest for this thesis, and open challenges related to applications such as brain-computer interfaces, neurofeedback, ambulatory EEG, etc.

2.1 Electroencephalography

EEG is an electrophysiological monitoring method that measures the electrical activity generated by the synchronized activity of thousands of neurons of the brain via intracranial electrodes or electrodes placed on the scalp surface, i.e., using invasive or non-invasive methods. The first known neurophysiological recordings were made by Richard Caton in 1875, when he presented his findings on the electrical phenomena of the exposed cerebral hemispheres of rabbits and monkeys [12, 13]. In 1890, Adolf Beck published an investigation on the spontaneous electrical activity of the brain of rabbits and dogs, which included rhythmic oscillations altered by light [14, 15]. Later, in 1924, Hans Berger recorded the first human EEG [13, 16].

Hans Berger described EEG in 1929 with the promise that it would be a technique that provides a “window into the brain” [16]. Recent progress in EEG sensors and methods for signal analysis have made this window more transparent

but the analytic potential and potential applications of EEG have not yet been fully exploited [17].

2.1.1 Mechanisms of EEG generation

Most of the electrical activity recorded in an EEG is generated by groups of well-aligned cortical pyramidal neurons that fire together and are oriented perpendicular to the surface of the brain, as well as near the scalp where the recording electrodes are placed. Each scalp electrode collects an estimated synchronous cortical activity of at least 6cm^2 [18].

The neural/electrical activity detectable by EEG is the sum of the excitatory and inhibitory postsynaptic potentials from thousands of pyramidal cells firing synchronously near each recording electrode. If the cells do not have a similar spatial orientation, their ions do not line up and thus do not create detectable waves. This summed activity can be represented as a field with positive and negative poles (dipole). The dipole vector is parallel to the orientation of the pyramidal cells that generate the activity [18, 19]. Negative dipoles are mostly detected when they are perpendicular and pointed directly at a recording electrode. The positive end of the dipole is subcortical and thus can be recorded only with deep electrodes (e.g., by intracranial EEG) [20].

Conventional scalp EEG is unable to record spontaneous changes in local field potential arising from neuronal action potentials. Because voltage fields fall off with the square of distance, activity from deep sources is more difficult to detect than currents near the skull [18, 20].

Cerebral voltages must traverse the brain, cerebrospinal fluid, meninges, skull, and skin prior to reaching the recording site where they can be detected. Cortical synaptic action generates electrical signals that change in the 10- to 100-millisecond range. EEG and magnetoencephalography (MEG) are the only widely available technologies with sufficient temporal resolution to follow such rapid dynamic changes.

2.1.2 Normal and abnormal EEG

The electrical activity measured by EEG is caused by the activation of neurons, but if these neurons are activated abnormally, sudden impulses can occur, which are defined as seizures. An EEG waveform is normal when the EEG recording

does not show unusual seizures. The waveform exhibits unusual characteristics, such as frequent, long, or continuous seizures, when the subject is affected by a tumor or brain disorder [18, 21].

Abnormal activity can be separated into epileptiform and non-epileptic activity. Focal abnormal non-epileptiform activity can occur in areas of the brain where there is focal damage to the cortex or white matter. It consists of an increase in slow-frequency rhythms and/or a loss of normal higher frequency rhythms [21, 22].

EEG waveforms are generally classified according to their frequency, amplitude, and shape, but the most familiar classification uses the EEG waveform frequency. This EEG waveform information is dependent on the subject's age and state of alertness and location of the electrodes on the scalp.

2.1.2.1 EEG frequency bands

The frequency of the EEG waveforms is important because the predominant frequencies vary according to the subject's condition. Frequency bands are typically within the range of 0.5 to 32 Hz. However, these frequency bands may vary slightly depending on the laboratory/headset and can be broken down into more limited components as required by the research or clinical question.

There are five commonly used frequency bands that are examined by spectral analysis; alpha, beta, theta, delta, and gamma. However, there is no consensus in the literature on what the ranges should be. For example, the values for the upper end of alpha and the lower end of beta include 12, 13, 14, and 15 Hz [18, 23]. Frequencies above 25 Hz are not commonly found on scalp EEG, but can be seen arising directly from the cortical surface during intracranial recordings; these frequencies are called gamma and are divided into low (25 – 70Hz) and high gamma (> 70Hz) [18, 24, 25]. Below, a brief overview of the five main frequency bands, including important points and frequency ranges, is presented.

- **Delta:** frequency range of 0.5-4 Hz. This activity is positively associated with the homeostatic sleep drive in such a way that it increases concomitantly with increasing time spent awake [26]. It tends to have the highest amplitude and the slowest waves. It is seen normally in adults in slow-wave sleep. Temporal intermittent rhythmic delta activity (TIRDA)

is frequently seen in individuals who have temporal lobe epilepsy [27].

- **Theta:** frequency range of 4-8 Hz. This activity is similar to delta activity and is positively associated with the homeostatic sleep drive [26]. It has been associated with reports of relaxed, meditative, and creative states. Excess theta activity for age represents abnormal activity, and focal theta activity during awake states is suggestive of focal cerebral dysfunction [28].
- **Alpha:** frequency range of 8-12 Hz. This activity is positively associated with relaxed wakefulness and drowsiness associated with the onset of sleep, and is also present during REM sleep [29–31]. Hans Berger named the first rhythmic EEG activity he observed the “alpha wave”. Deceleration of the background alpha rhythm is considered to be a sign of generalized brain dysfunction [32]. The amplitude of the alpha rhythm varies between individuals, as well as at different times in the same individual [31]. It is best seen with the eyes closed and during mental relaxation and is attenuated by eye-opening and mental effort.
- **Beta:** frequency range of 13-30 Hz. This activity is the dominant rhythm of subjects who are alert or anxious or who have their eyes-open. It is the most frequently seen rhythm in normal adults and children and is associated with physiological arousal and psychological stress [33]. This activity is closely linked to motor behavior and is generally attenuated during active movement [34]. The amplitude of beta activity is typically 10-20 μV , rarely increasing above 30 μV .
- **Gamma:** frequency range of approximately 30-100 Hz, consisting of ripples (80 to 200 Hz) and fast ripples (200 to 500 Hz). Ultra-fast EEG activity correlates with cognitive states and ERPs. It has been attributed to sensory perception that integrates different areas. There has been extensive research on high-frequency oscillations, particularly in relation to epilepsy [24, 25, 35]. Epileptic foci are known to generate very high-frequency episodes of activity. Intracranial depth recordings of the epileptic hippocampus have reported ultra-fast frequency bursts or fast waves, which probably correlate with the local epileptogenicity of brain tissue

[35]. Subdural recordings during presurgical evaluation of epilepsy have demonstrated that activity bursts at a relatively lower frequency range (60 to 100 Hz) may likewise indicate the location of an epileptic focus [28, 35].

2.1.2.2 Artifacts

Electrical signals detected on the scalp by an EEG sensor, but which are non-cerebral in origin, are called artifacts. Artifacts originate from both physiological and non-physiological sources, of which physiological artifacts arise from a variety of bodily activities and non-physiological artifacts from outside the human body [36–38].

The most highly studied artifacts include **eye-induced artifacts**, which include eye blinks, eye movements, and extra-ocular muscle activity, **electrocardiograph (ECG) artifacts**, which are related to heart beat (cardiac electrical activity), **electromyography (EMG)-induced artifacts**, which are related to muscle activation, and **glossokinetic artifacts** from tongue movement. Respiration can also cause artifacts by introducing rhythmic activity that is synchronized with the respiratory movements of the body. Skin responses, such as sweating, can alter the impedance of the electrodes and cause artifacts in EEG signals [18, 37, 39].

Certain artifacts are essential for understanding brain function but many are not and limit the interpretation of the EEG. Artifact removal is the process of identifying and removing artifacts from brain signals. This can be accomplished by applying frequency-band and spatial filters but artifacts can overlap with the signal of interest in the spectral domain. An artifact-removal method should be able to remove the artifacts while keeping the related neurological phenomenon intact. The first step in managing artifacts is to prevent them from occurring by issuing proper instructions to users. For example, users are instructed to avoid blinking or moving their body during data collection. Some of the common methods for removing artifacts in EEG signals are linear filtering, linear combination and regression, blind source separation (BSS), independent component analysis (ICA), and principal component analysis (PCA) [37–40].

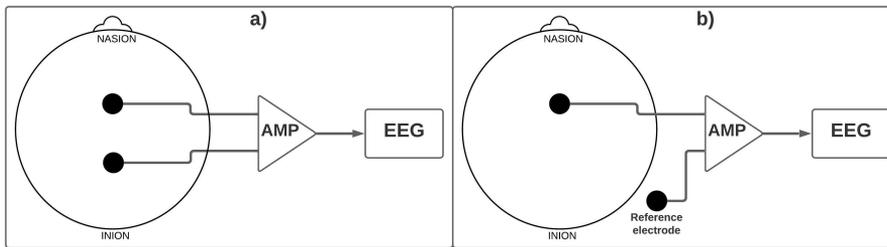


Figure 2.1: EEG electrode placement methods: bipolar (a) and monopolar (b).

2.1.3 EEG signal acquisition

EEG uses the principle of differential amplification, or recording of voltage differences between different points using a pair of electrodes that compares an active scanning electrode site with another neighboring or distant reference electrode. This can be accomplished using monopolar or bipolar recordings, in which measuring differences in electrical potential generates detectable EEG waveforms [41, 42].

The difference between monopolar and bipolar recordings is the location of the electrodes. In bipolar recordings, the electrodes are both placed on the scalp, i.e., in the area of interest, whereas in the monopolar electrode placement method, one of the measurement electrodes is placed on the scalp and the other is located away from the area of interest (see Fig. 2.1).

In both cases, the amplifier captures the difference between the respective activity at each site. Both are in fact bipolar recordings, in the sense that there are two inputs to the amplifier. When the second electrode is placed on an EEG neutral site, the recording is considered to be monopolar (also known as referential), because only one site is believed to be capturing the EEG data. If both electrodes are placed over sites that capture active EEG data, the recording is called bipolar (also called sequential or differential) [42].

There are several reasons why monopolar recordings are recommended for surface EEG recordings. One reason is, because the bipolar or differential amplifier rejects everything that is common to both electrodes, it will reject any common EEG activity, which is far less present in monopolar recordings. Another reason is that a bipolar recording can be derived from a monopolar recording using simple arithmetic, whereas a bipolar recording can never be transformed into a

monopolar one [43].

2.1.4 A brief comparison with other brain signal acquisition methods

There are several brain-imaging methods available for neuroscientists and researchers. These imaging modalities can be divided into structural and functional imaging techniques. They all allow the study of brain structures and their function but differ in the spatial and temporal resolution at which connectivity is captured. Structural imaging provides details on the morphology and structure of tissues, whereas functional imaging reveals physiological activities, such as changes in metabolism, blood flow, regional chemical composition, and absorption.

Non-invasive EEG and MEG reflect the average activity of dendritic currents in a large population of cells. The temporal resolution of EEG and MEG for measuring changes in neuronal activity is very good, typically on the order of milliseconds, but the spatial resolution for determining the precise position of active sources in the brain is poor relative to modern imaging methods, such as computerized tomography (CT), positron emitted tomography (PET), and magnetic resonance imaging (MRI) [17, 44].

Despite its limited spatial resolution, EEG is still a valuable tool for research and diagnosis. It is one of the few mobile techniques available and offers millisecond-range temporal resolution that is not possible with CT, PET, or MRI. The poor spatial resolution, particularly for sources deeper in the brain, is due to the spatial mixing of electrical activity generated by different cortical areas and the passive conductance of these signals through brain tissue, cerebrospinal fluid, bone, and skin/scalp [17, 19, 44]. Additionally, these measurements are very susceptible to artifacts arising from muscle and eye movements. Invasive versions of EEG improve spatial resolution by placing subdural and/or deep electrodes for a more direct recording of spontaneous or evoked neural activity.

Functional magnetic resonance imaging (fMRI) measures changes in blood hemoglobin concentrations associated with neural activity, based on the differential magnetic properties of oxygenated and deoxygenated hemoglobin. fMRI has much better spatial resolution than EEG and MEG, but the temporal resolution is poor, which puts an upper bound on the bit rate for fMRI in BCI applications. Recently, an approach was presented that uses intracranial EEG

(iEEG) that can collect as much data as fMRI, but using a portable device inside a backpack [45]. This will allow the study of brain function of subjects while they are interacting with others, rather than inside an fMRI machine.

Since the inception of EEG, various standards and guidelines have been proposed for electrode placement to ensure signal integrity and repeatability of recordings, as described below.

2.1.5 International EEG electrode placement systems

H.H. Jasper studied possible methods to standardize electrode placement, resulting in the definition of the 10-20 international system, which consists of 21 electrodes placed at distances of 10% and 20% along certain contours over the scalp, as illustrated in Fig. 2.2 [2]. Since then, the 10-20 international system has become the standard for the study of EEG and ERPs in both clinical and non-clinical settings. Later, the extended 10-20 or 10-10 system was proposed to extend the number of channels from 21 up to 74. These systems simply extend the number of electrodes by placing them at every 10% along the medial-lateral contours and by introducing new contours in between the existing ones [46].

The extended 10-20 or 10-10 system have been accepted and endorsed as the standard of the American Electroencephalographic Society and the International Federation of Societies for Electroencephalography and Clinical Neurophysiology [4, 5]. There is a proposed extension to accommodate a larger number of electrodes, known as the 10-5 system, which includes the 10-20 system and 10-10 system locations, enabling the use of up to over 300 electrode locations [3].

In all cases, the electrode names consist of one or more letters and a number, with the electrodes on the left being odd numbered and the electrodes on the right even numbered. The electrodes at the center, or midline, are designated by the letter **z**, indicating that the electrode is neither even nor odd. The electrodes at the midline have the smallest numbers and the numbers increase towards the side, where the letter indicates the location on the head, which are **Fp: frontal pole**, **F: frontal**, **C: central**, **T: temporal**, **P: parietal**, **O: occipital**. Additionally, combinations of two letters indicate intermediate locations, i.e., **FC: in between frontal and central electrode locations**, **PO: in between parietal and occipital electrode locations**.

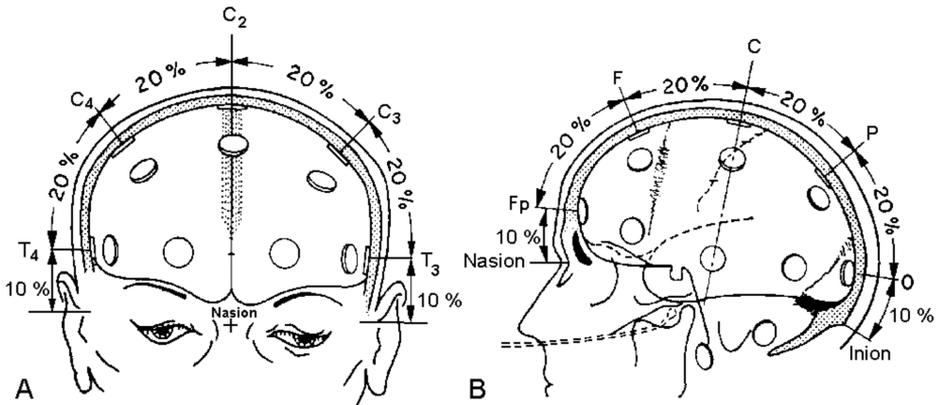


Figure 2.2: The original figure illustrating the international 10-20 system. Note that the electrodes are erroneously located inside the skull on the surface of the cortex [2].

2.1.6 Consumer-grade low-density EEG headsets

High-density EEG uses a dense array of EEG channels, in which the number of electrodes can vary from 32 to 256 or more [47–49]. However, there is no fixed number of channels that defines a low-density EEG headset. The 21 channels from the 10-20 international system is considered to be low-density and in some studies, the authors considered low-density EEG to consist of arrays with 25 channels [50] and others when using arrays of 32, 16, or 8 channels [51]. In this context, EEG can be considered low-density when less than 32 channels are used.

There is currently a wide range of consumer-grade EEG headsets available that follow the 10-20, 10-10, or 10-5 system [52, 53]. A review published in 2015 provides information about the headsets Emotiv, NeuroSky, interaXon (Muse), and OpenBCI, which are mainly used for cognitive studies, BCI research, education, and gaming [52]. Interestingly, Emotiv products are popular for cognitive studies and gaming, NeuroSky dominates the educational field, and published BCI research has only used Emotiv and OpenBCI headsets. In [54] there is a review of various BCI applications and cognitive neuroscience research using Emotiv up to 2019, showing that most of the research has come from the United States, India, China, Poland, and Pakistan. Fig. 2.3 presents a timeline of the evolution of EEG systems since the time of Hans Berger and several relevant consumer-grade EEG headsets.

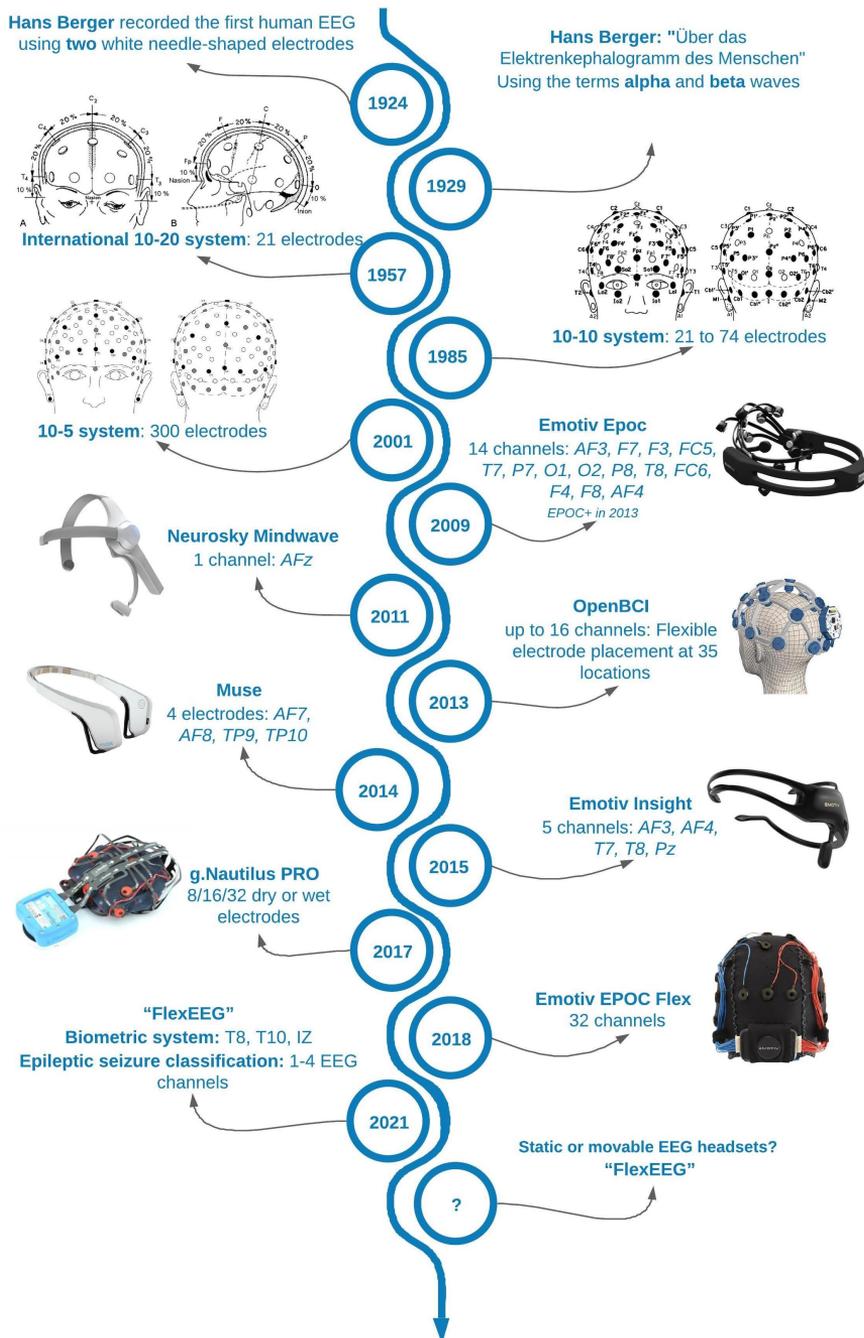


Figure 2.3: Timeline of the evolution of EEG systems and relevant consumer-grade wearable EEG headsets.

Fig. 2.3 shows the starting point for recording human EEG signals, using two white needle-shaped electrodes, which was performed by Hans Berger in 1924 and reported in 1929. High-density EEG was the starting point for analysis for certain applications, initiating the publication of international standards, starting with the international 10-20 system, and subsequent standards by placing electrodes in the middle and around this first system.

Fig. 2.3 also presents the set of channels found in this thesis, which will be later described in Chapters 4 and 5. As explained in Chapter 1, the thesis focused on two main applications: **Epileptic seizure classification**, and **EEG-based biometric systems**, finding that a set of 1-3 EEG channels can be used for epileptic seizure classification, and 1-4 EEG channels for creating EEG-based biometric systems.

Various consumer-grade wearable EEG headsets using dry or wet electrodes have gradually emerged, featuring different channel configurations or even flexible solutions, such as for the openBCI. Indeed, there is evidence that it is possible to obtain similar results to that of medical grade equipment using the openBCI with dry electrodes [55]. However, work is still needed to improve the recording quality and increase the sample rate, which is limited to 250Hz for the openBCI for a maximum of eight channels or 125Hz if more are used.

There are various areas of application for which the creation of new EEG headsets could be interesting but the idea of comparing the use of static versus movable EEG electrodes for a single headset for different applications needs further exploration, as discussed in [56–58]. Recently, a research project entitled **FlexEEG** was presented, which aims to achieve real-time BCI with brain mapping capabilities [58]. The FlexEEG concept is different from the standard high-density EEG in that it involves dynamically scanning the human scalp to achieve the minimum required recordings, rather than having electrodes attached to the scalp, as illustrated in Fig. 2.4. The work in this thesis can contribute to the realization of such a low-density EEG array by providing the software that can identify the minimum EEG channel count required for a given neuro-paradigm.

2.1.7 Using brain signals for control purposes

Technological progress has allowed the analysis of EEG to move from pure visual inspection of amplitude and frequency modulation to a more rigorous and automatic exploration of the temporal and spatial features of the recorded

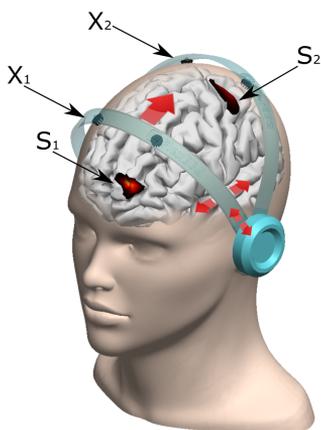


Figure 2.4: FlexEEG concept. FlexEEG moves from X_1 to X_2 to capture sources S_1 and S_2 [58].

signals.

As a result, EEG is accepted as a powerful tool to capture brain function and has been shown to be valuable in clinical diagnosis, i.e., the identification of epilepsy and sleep and mental disorders, the evaluation of various dysfunctions, etcetera [17, 44].

Since the first proposal to use EEG signals to control external devices (i.e., prosthetic arms) [59], efforts to improve the interpretation of brain signals through EEG signals, and thus establish more robust control over external devices, have rapidly increased [60, 61].

The assumption that invasive methods can provide better performance has not been completely supported by the results of several studies [62–66], which have shown that the control of movement obtained with scalp-recorded sensorimotor rhythms falls in the same range in terms of speed and precision as the control obtained with invasive methods [63].

Recently, several approaches using invasive methods have been presented that allow subjects to control a prosthetic limb with 10° of freedom (three-dimensional (3D) translation, 3D orientation, four-dimensional hand shaping) [67]. However, this required two 96-channel intracortical electrode arrays implanted in the subject's left motor cortex.

The processes followed for invasive and non-invasive methods, assumptions,

and results obtained in each case are too different to allow a good comparison of invasive and non-invasive methods. For example, current non-invasive studies suggest that a spelling protocol that uses a goal-selection approach (such as P300-speller) may be faster and more reliable than a spelling protocol that uses a process-control approach [60, 61, 68].

The most appropriate protocol and paradigm need to be selected following careful analysis, according to the purpose of the BCI. In addition there are numerous different paradigms available, such as motor imagery paradigms, external stimulation paradigms (i.e., P300), error-related potential, etcetera [69].

Then, it is necessary to create a training set using the selected paradigm, which can be task-dependent or task-independent during the resting-state, and collect the EEG data for creating the models using mathematical methods. The EEG data are then collected while the subject performs the same task (or during the resting-state), the created model used to predict the task, and the predicted task used for BCI control.

2.2 EEG paradigms

Paradigm selection is important and must be associated with the purpose of the EEG-based control application or EEG-based controller or BCI. Below, one important paradigm and several relevant aspects about the resting-state, which are referred to throughout the thesis, are described.

2.2.1 Event-related potentials and P300

ERPs are very small voltages that appear on the scalp as a response of the human brain to specific events or stimuli that are time- and phase-locked. These have been used to evaluate brain function and the response to stimuli. These signals include both spontaneous electrical activity of the cerebral network and the cortical response to external or internal events.

ERPs produce several well-known patterns (see Fig. 2.5). One of the most extensively studied and used for BCIs is the P300 peak, also known as *P3* [69–71]. The *P300* component is elicited in response to infrequent events using what is known as an *oddball paradigm*. It consists of a positive peak in the ERP ranging from 5 to 10 μV in amplitude with a latency between 220 to 500 ms after onset of the stimulus, and is most significant at central-parietal scalp and midline skull

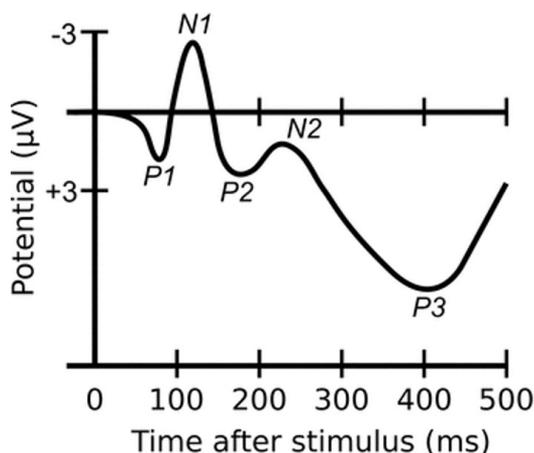


Figure 2.5: Schematic representation of certain ERP components after the onset of a visual stimulus [72].

locations, i.e., Pz, Cz, and Fz in the 10-20 international system. Normally, hundreds of ERPs are generated, collected, and averaged to visually distinguish the P300 peak from the background activity, thus cancelling the influence of noise.

The P300-speller paradigm was developed with the initial aim to restore communication to locked-in state patients [73] and normally consists of a $N \times N$ matrix of characters that is presented to the subject in random sequences of intensified columns and rows (Flashed), thus constituting an oddball paradigm [70, 73].

An important advantage of P300 for a BCI is that most subjects can use it with very high accuracy and it can be calibrated in a few minutes, which means that subjects can use BCI systems to control devices quickly. However, disadvantages of this paradigm are that it may produce fatigue and that subjects with visual impairment are not able to use BCIs based on this paradigm [73–76].

2.2.2 Resting-state

The resting-state, also called resting-state activity, is typically used to analyze problems relative to the subject's internal state of mind. A stable resting-state does not necessarily exist, because spontaneous changes in regional neuronal firing occur even when the organism is apparently in resting-state [77].

In addition, spontaneous activation can change local blood flow and cause

low-frequency blood oxygenation level-dependent signal fluctuations [78]. In other words, the brain is never truly at rest [79] and the term only refers to the absence of goal-directed neuronal action with the integration of information of the external environment and the subject's internal state, as well as when the subject is not actively engaged in sensory or cognitive processing.

Brain activity can be studied in the resting-state in children or patients who would otherwise be unable to complete long experiments or perform complex cognitive tasks and the simplicity of the procedure for collecting EEG signals has also facilitated the replication of experiments and comparison of results.

The resting-state is typically used to analyze clinical or psychological problems [80–82] and for most cases of real-time implementation of BCI approaches, as it is necessary to differentiate between the tasks associated with the paradigm and the resting-state [83]. The resting-state can also be used for various EEG-based systems [83–87].

Most resting-state features from EEG consist of ongoing amplitude-modulated oscillations in the approximate frequency range of 0.5-70 Hz [88]. There is evidence that the alpha frequency band of the multi-channel resting-state in EEG signals can be parsed into a set of discrete states, called *microstates*, which are defined by topographies of electrical potentials, and remain stable for 80–120 ms before rapidly transitioning to a different microstate [89, 90].

Resting-state EEG microstates reflect neural activity in a task-negative state, which is considered to be primarily involved in involuntary actions. Brain regions exhibiting functional connectivity are organized into discrete networks associated with distinct functions. Among them are a host of so-called resting-state networks (RSNs), which represent functionally connected areas that are active in the task-negative state [90]. One such network is the **default-mode network, which is active in the task-negative state** but becomes deactivated in a wide array of cognitive tasks [91].

Interestingly, only four predominant topographies occur during the resting-state and all can be reliably identified in healthy individuals throughout their life span and explain most global topographical variance [92, 93], as shown in Fig. 2.6. However, several studies have been published that show more than four microstates [94]. This can all influence the selection of the most relevant channels

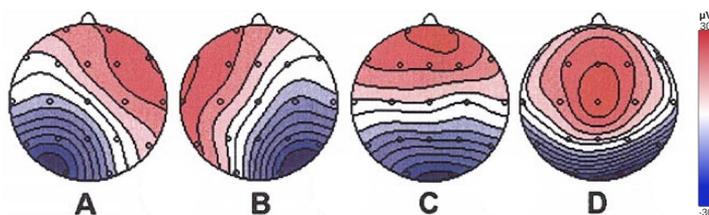


Figure 2.6: Topography of four microstate maps from [92]. Map areas of opposite polarity are coded in red and blue using a linear color scale. The left ear is to the left and the nose is at the top

for extracting information in BCI applications.

Fig. 2.6 presents the eyes-closed resting-state EEG microstates from [92], which consist of four classes of microstates: **class A**, with a left occipital to right frontal orientation; **class B**, from right occipital to left frontal orientation; **class C**, with a symmetrical occipital to prefrontal orientation; and **class D**, also symmetrical, but with a fronto-central to occipital axis. The resting-state microstates are shown to move around the sensorimotor areas of the brain, as a way of sensing the brain through the most important senses of the human body.

A review compared the four microstate maps determined in various independent studies using a varying number of electrodes, participants, filter settings, etcetera [95]. The four presented microstate maps were distinct in the studies but highly reproducible, with the **class A** and **class B** similarities being clearer.

As will be shown in Chapter 5, the channel distribution found during the followed optimization process showed a similar channel distribution as the four topographies of the resting-state microstates presented in Fig. 2.6.

2.3 Current and future trends in EEG

There is a growing interest in the use of EEG in medical ambulatory and non-medical and wearable applications, such as entertainment, day-to-day mobile EEG, sports, neuro-assisted learning, and brain-computer interfaces. This will require the implementation of miniaturized, user-centric, wireless EEG acquisition systems with ultra-low power dissipation that is robust to motion artifacts. However, currently available mobile EEG systems are still quite bulky and use structures with a large number of fixed electrodes, which are not comfortable for day-to-day

mobile EEG monitoring.

There are many fronts on which these requirements can be addressed. Two central research points in terms of EEG electrodes are the creation of newer electrode technologies and lower-power consumption electronics. To increase the battery lifetime of wearable EEG devices, research is also being carried out on data reduction approaches. For example, in the diagnosis of epilepsy, data reduction techniques have been used to extend the battery life of wearable EEG devices through intelligent selection and solely transmission of EEG data relevant for diagnosis [96].

There is a trend towards applying combined sets of features that can produce better performance for classification rather than using features independently [97]. Future directions should combine machine learning and traditional approaches for effective automatic artifact removal [98]. One of the main concerns regarding EEG and BCIs is that almost all published experiments have been performed in a controlled laboratory, whereas the need is towards improving artifact removal in daily-life EEG-BCI, which is also important for the use of dry electrodes, for which more research is clearly needed [99, 100]. When designing new EEG headsets, it is important to thoroughly examine the basic criteria of the system, environmental aspects, situation, and target users/applications [98, 101].

For certain applications and environments, the trend is towards higher sample rates and more recording channels. However, for low-power, easy-to-use portable systems, the channel count needs to be minimized without affecting the accuracy of manual/visual inspection and machine learning based applications [99].

The integration of brain monitoring based on EEG into everyday life has been hindered by the limited portability and long setup time of current wearable systems, as well as the invasiveness of implanted systems. There is a current trend towards exploring the potential of recording EEGs in the ear canal for brain monitoring, which is known as in-the-ear EEG (Ear-EEG) [102, 103]. Ear-EEG has been presented as a system that promises a number of advantages, including fixed electrode position, user comfort, robustness to electromagnetic interference, and ease of use, and that can be used for long-term monitoring [102].

Research efforts are ongoing to make EEG devices smaller, more portable, and easier to use. The so-called *wearable EEG* is based on the creation of low-power

wireless collection electronics and dry electrodes that do not require a conductive gel for use [104, 105]. Wearable EEG aims to provide small EEG devices that are present only on the head and can record for days, weeks, or months, as promised by ear-EEG [100, 102].

In general, wearable EEG is envisioned as the evolution of ambulatory EEG units from the bulky, limited-life devices available today to small devices. Such miniaturized devices will enable long-term monitoring of diseases, such as epilepsy and various mental disorders, as well as improve end-user acceptance of BCI systems [100, 102, 105].

Future wearable EEG systems should be unobtrusive, lightweight, discrete, and durable, which can be achieved by eliminating the large ambulatory EEG recording units and wires that attach them to the electrodes. These will be replaced by microchips containing the necessary amplifiers, quantizers, and wireless transmitters, which are mounted on top of the electrodes. EEG data will then be transmitted wirelessly to a suitable mobile phone or similar device, which people often keep a short distance from themselves [104, 105].

In some cases, such as epilepsy diagnosis, wireless transmission of EEG data is not strictly necessary, as data analysis is normally performed after data collection, but wireless transmission will be necessary for future applications in predicting epileptic-seizures and their automatic treatment. Even wireless connections between electrodes is desirable to enable miniaturization [100, 104, 105].

Chapter 3

Materials and Methods

This chapter introduces the concepts that provide the basis for the thesis contributions and a summary of the datasets used, as well as a flowchart describing the proposed methods for feature extraction and classification. The proposed methods for channel-count optimization used in the cases studied are presented.

As introduced in Chapter 1, a comprehensive view of the necessary methods and tools used to achieve the objectives of the thesis, is presented. Fig. 3.1 presents the stages followed, which includes the EEG datasets (a)), pre-processing and feature extraction (b)), the classifiers used (c)), and the various methods for channel reduction and selection (d)). Each necessary step is presented and explained below for the datasets used, which are presented in Section 3.6.

3.1 Improving the signal-to-noise ratio

As introduced in Section 2.1.2.2, EEG signals can be contaminated by various sources of artifacts or noise produced by body movement, EMG, ECG, eye movements, sweating, power lines, impedance fluctuations, cable movements, etcetera [106]. Therefore, an important step before analyzing EEG signals is to enhance the signal-to-noise ratio, for which there are several spatial filtering techniques [38, 107–109]. Among the simplest and most used methods are the Common Average Reference (CAR) and Laplacian Filter (LAP) [110–112].

In this thesis, the signal-to-noise ratio from the EEG signal was improved using the CAR method, which removes simultaneously-recorded common information from all electrodes. CAR can be computed for an EEG channel V_i^{CAR} , where i is the number of the channel, as follows:

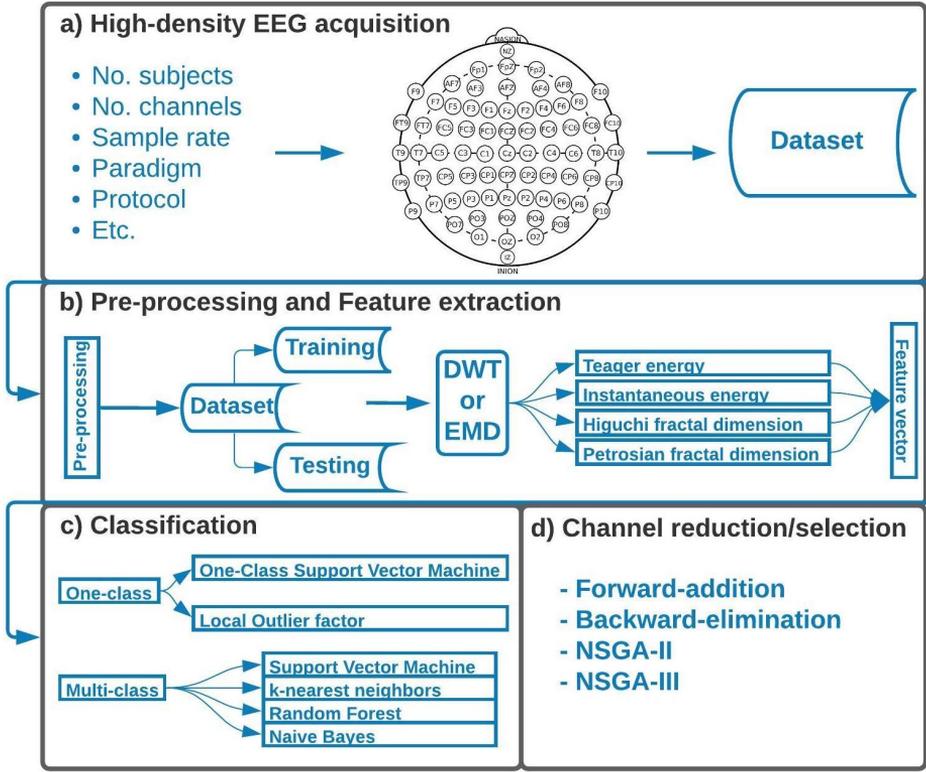


Figure 3.1: Stages of the methodology followed in the thesis.

$$V_i^{CAR} = V_i^{ER} - \frac{1}{n} \sum_{j=1}^n V_j^{ER} \quad (3.1)$$

where

- V_i^{ER} is the potential between the i^{th} electrode and the reference, and n is the number of electrodes.

After removing the noise from the EEG signals, it can be processed using data transformation techniques, such as EMD or DWT, to decompose the signals into different frequency bands and thus extract relevant features from each sub-band, as explained below.

3.2 Data analysis

Data analysis helps to provide information hidden in the data. It refers to the process of manipulating and transforming/converting data from one format, structure, or domain to another. For example, data analysis techniques can be used to convert a signal from the time-amplitude to time-frequency or amplitude-frequency domain, and vice-versa. This process can increase the value and efficiency of analytical or feature extraction procedures. When working with noisy raw data, the extraction of a handful of fundamental features (mean, variance, slope, etc.) is not generally sufficient, but valuable information can be extracted by manipulating or transforming the data. When working with EEG signals, feature extraction techniques can be time-based, frequency-based, or time-frequency-based. Time-frequency-based features are used more frequently as they can simultaneously provide information about the time and frequency of the EEG signals. EMD and DWT are the most popular and useful feature extraction techniques [113–115].

3.2.1 Empirical Mode Decomposition

EMD is an adaptive data analysis method used for decomposing non-linear and non-stationary signals, which may be mono-component or multi-component, into a finite number of amplitude and frequency-modulated zero-mean signals without leaving the time domain, called *Intrinsic Mode Functions* (IMFs), which satisfy two conditions [116]:

1. The number of extrema and the number of zero crossings must be either equal or differ at most by one.
2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The method decomposes a signal into oscillatory components by applying a process called *sifting*, making EMD a data-driven method that does not depend on any *a priori* defined system. This process removes riding waves and makes the wave-profile more symmetrical [116, 117]. EMD decomposes a time-series $x(t)$ into IMFs $x_i(t)$ and a residue, such that the signal can be represented and reconstructed as shown in Eq. 3.2 and summarized, as shown in algorithm 1:

$$x(t) = \sum_{i=1}^n x_i(t) + residue \quad (3.2)$$

An important aspect presented in algorithm 1 is whether a given sample is or is not an upper or lower extrema, since it must be based on the relationship of the actual sample with its left and right neighbours. The envelopes will be different depending on the accuracy of the method for finding these upper and lower extrema points, as the sifting process is implemented by connecting all of the local minima or maxima by a cubic spline line to extract the IMFs. Additionally, it may lead to minor deviations from the true mean envelope depending on the spline used for the interpolation, producing different IMFs. According to [118], the natural spline is the most reasonable one to select.

During the interpolation process, at least one extrema on each side must be free, unless the first and last points were simultaneously considered as the maximum and minimum. This is known as an end effect and can be solved by using mirror continuation [119–122]. However, the requirement for this approach is that the mirror be placed at the extrema point, but if the signal cannot determine whether the endpoint is the extrema point, then it amputates part of the data to place the mirror at the extrema point. The authors in [122] proposed a combination based on support vector machine (SVM) and EMD mirror extension methods to predict the extrema points near the end of the signal and thus solve the EMD end-effect problem. Briefly, an SVM model is used to extend the two ends of the original data to obtain local extrema points, then the image in the mirror is mapped to a ring signal with no endpoints by mirror extension. The stopping criterion is another important part of EMD, as it determines the number of sifting steps to produce an IMF, and the sifting process has to be repeated as many times as necessary to eliminate all riding waves. Generally, it is critically important in the successful implementation of EMD.

Mode mixing is another well-known problem encountered during the sifting process and happens when EMD tries to extract mono-components from a multi-component signal. In such cases, the sifting process only identifies modes that clearly contribute their own maxima and minima. Otherwise, EMD will not be able to separate the mode in a single IMF and the mode will remain mixed in another

Algorithm 1 The sifting process for a signal $x(t)$

```

1: Data: signal =  $x(t)$ 
2: Result: IMFs
3: sifting = True
4: while sifting = True do
5:   Identify all upper extrema in  $x(t)$ 
6:   Interpolate the local maxima to form an upper envelope  $u(x)$ .
7:   Identify all lower extrema of  $x(t)$ 
8:   Interpolate the local minima to form an lower envelope  $l(x)$ 
9:   Calculate the mean envelope:
      
$$m(t) = \frac{u(x)+l(x)}{2}$$

10:  Extract the mean from the signal:
      
$$h(t) = x(t) - m(t)$$

11:  if  $h(t)$  satisfies the two IMF conditions then
12:     $h(t)$  is an IMF                                { Add  $h(t)$  to IMFs }
13:    sifting = False                                  { Stop sifting }
14:  else
15:     $x(t) = h(t)$ 
16:    sifting = True                                    { Keep sifting }
17:  end if
18:  if  $x(t)$  is not monotonic then
19:    Continue
20:  else
21:    Break
22:  end if
23: end while

```

IMF or split between several IMFs [123, 124]. Data affected by the presence of intermittence and noise can also produce the mode-mixing problem.

There are EMD-based methods for noise removal, solving end effects, and the mode-mixing problem. For example, Ensemble EMD (EEMD) defines true IMFs as the mean of an ensemble of trials [124]. However, EEMD is not recommended for real-time applications due to the computational cost [125].

3.2.1.1 IMF selection

Depending on the parameters selected for the EMD method (spline for the interpolation, the method for solving the end-effect problem, etc.) and because the numerical procedure is susceptible to errors, some IMFs that contain limited

information may appear in the decomposition [126].

There are several approaches for selecting the IMFs that contain the most relevant information about the signal, i.e., using energy-based techniques or using a threshold or distance [127–129]. For illustrative purposes, an example employing the Minkowski (Euclidean) distance (d_{mink}) is presented, which is defined as follows.

$$d_{mink} = \left(\sum_{i=1}^n |x_i - y_i|^2 \right)^{1/2} \quad (3.3)$$

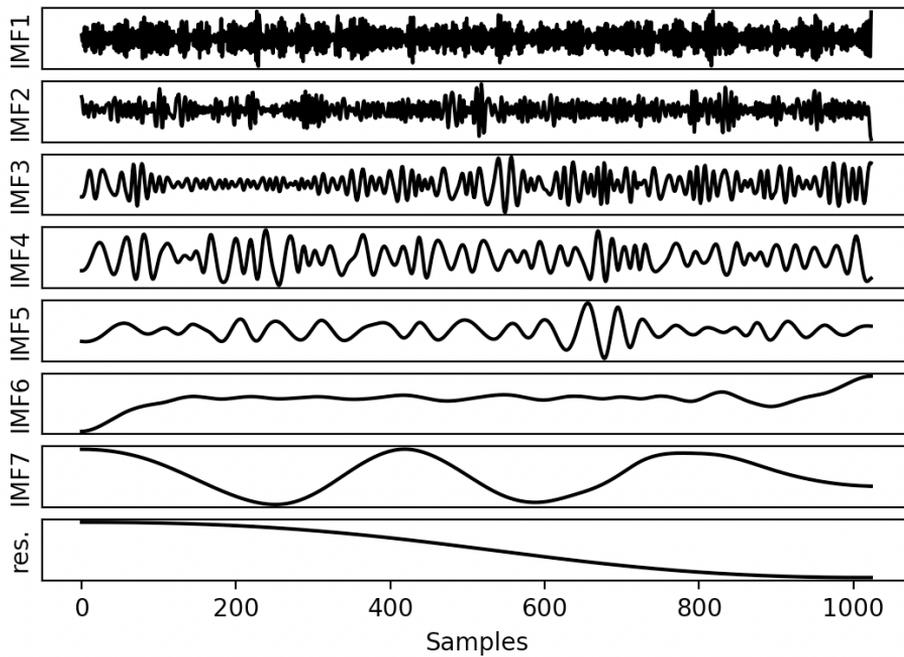
where x_i and y_i are the i -th respective samples of the observed signal and the extracted IMF. According to [128], the redundant IMFs have a shape and frequency content different from those of the original signal, which means that when an IMF is not appropriate, the d_{mink} presents a maximum value.

Fig. 3.2, presents an example using a synthetic signal generated by $x(t) = \sin(3\pi * t) + \sin(\pi * t) + \text{white_noise}$, which can be compared to the IMF selection methods presented by [127, 129]. For the example presented, it was considered to be a trial of two seconds with a sample rate of 512 Hz and, for illustrative purposes, only the first three most relevant IMFs, according to the Minkowski distance, were selected (the closest three IMFs). However, this number may vary depending on the nature of the data, sample rate, trial-duration, and other factors.

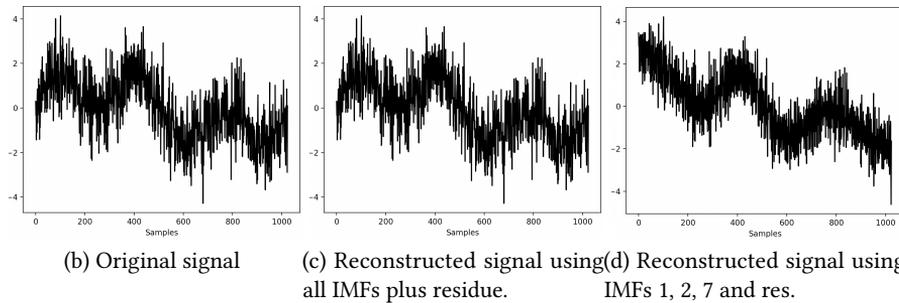
Fig. 3.2 shows that the original signal can be reconstructed by using all the obtained IMFs, but also if only the three closest IMFs and the residue are used. This means that EMD can decompose a signal into different components and also capture the most relevant information in different IMFs. This may be important for certain applications and depending on the nature of the signal, as the use of a large dataset can increase the computational cost. Therefore, using only the most relevant IMFs, it is possible to extract the main components (relevant information) from the signal and analyze it further.

3.2.2 Discrete Wavelet Transform

A wavelet is a brief rapidly decaying wave-like oscillation with an amplitude that begins at zero, increases, and decreases back to zero, and has a finite duration. The wavelet transform (WT) replaces the sine and cosine functions of Fourier transform



(a) IMFs and residue (res.) extracted from the original signal using EMD.



(b) Original signal

(c) Reconstructed signal using all IMFs plus residue.

(d) Reconstructed signal using IMFs 1, 2, 7 and res.

Figure 3.2: IMFs plus residue (Sub-fig. 3.2a) obtained from the synthetic signal presented in sub-fig. 3.2b, as well as the reconstructed signal using all the IMFs (Sub-fig. 3.2c) and three IMFs selected using the Minkowski distance plus the residue (Sub-fig. 3.2d).

(FT) by translations and dilations of a wavelet. It is basically a mathematical technique in which a particular signal is analyzed in the time domain using different versions of a translated and dilated basis function called a *mother wavelet*. WT is suitable for analyzing irregular data patterns, such as non-stationary signals,

and it provides well-defined frequency and time resolution for both low and high frequencies.

There are two important parameters used in the transformation: scaling and shifting. A stretched wavelet, which is produced with large-scale factors, helps to capture the slowly varying changes (low frequencies), whereas a compressed wavelet, produced with small-scale factors, helps to capture the abrupt changes (high frequencies). The wavelet has to be shifted to align with the desired feature. Shifting a wavelet means delaying or advancing the onset of the wavelet along with the signal. In general, WT is represented in Eq. 3.4.

$$\psi_{a,b} = \frac{1}{\sqrt{|a|}} \psi \frac{t-b}{a} \quad (3.4)$$

where

- **a** and **b** are the scaling and shifting parameters, respectively.
- ψ is the mother wavelet
- For a given scaling parameter **a**, the wavelet is translated by varying the parameter **b**.

Selecting an appropriate mother wavelet is crucial for analyzing the signals, as it will affect the outcome and various wavelets applied on the signal may produce different results. It is common to select a mother wavelet that is similar in shape to the original raw signal, but it can be selected experimentally.

DWT provides a time-frequency representation of a signal and decomposes a signal in the time domain into shifted and scaled versions of a mother wavelet. DWT provides sufficient information of the original signal with a significant reduction in computation time by passing the signal through a series of low-pass and high-pass filter pairs. The DWT is presented in Eq. 3.5.

$$DWT_{j,k} = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{|2^j|}} \psi \frac{t-2^j k}{2^j} dt \quad (3.5)$$

where

- **j** and **k** are the scaling and shifting parameters, respectively.

- ψ is the mother wavelet
- 2^j and $2^j k$ replace a and b from Eq. 3.4, respectively.

Additionally, it is necessary to pre-define two parameters, the decomposition level and the mother wavelet. The outputs provide the level 1 high-frequency part, called detail coefficients (D1), and the level 1 low-frequency part, called approximation coefficients (A1). Subsequently, the low-pass portion is fed into a new set of filters and the process is repeated until the signal is decomposed to a pre-defined level. Briefly, the wavelet decomposition of a signal $x(t)$ in the j decomposition level has the structure $[A_j, D_j, D_{j-1}, \dots, D_1]$. It should be noted that at every level, half of the samples can be removed according to the Nyquist theorem [130].

Fig. 3.3, presents an example using a synthetic signal generated by $x(t) = \sin(3\pi * t) + \sin(\pi * t) + \text{white_noise}$, using four levels of decomposition and the mother wavelet biorthogonal 1.3. As in the example presented in 3.2.1, it was considered to be a trial of two seconds with a sample rate of 512 Hz.

3.3 Data features

A feature is an individual measurable property or characteristic of a phenomenon being observed. **They can be mainly divided into two types, fundamental and complex.** Fundamental features, also know as time-domain features, are explicitly present in the acquired data and can be directly used, i.e., mean, median, variance, standard deviation, amplitude, kurtosis, skew, etc. Complex features are generated by manipulation or transformation of the data (transformations using methods such as EMD or DWT), and after a certain amount of transformation of the data, **it is necessary to extract certain relevant patterns, which also helps in dimensionality reduction.** Choosing informative, discriminating, and independent features is a crucial step for effective training of algorithms in pattern recognition, classification, and regression. Below, a set of energy and fractal features relevant to this thesis is introduced.

3.3.1 Energy distribution

The energy E_s of a discrete signal (n) is defined as the area under the squared magnitude of the signal, and is calculated as in Eq. 3.6.

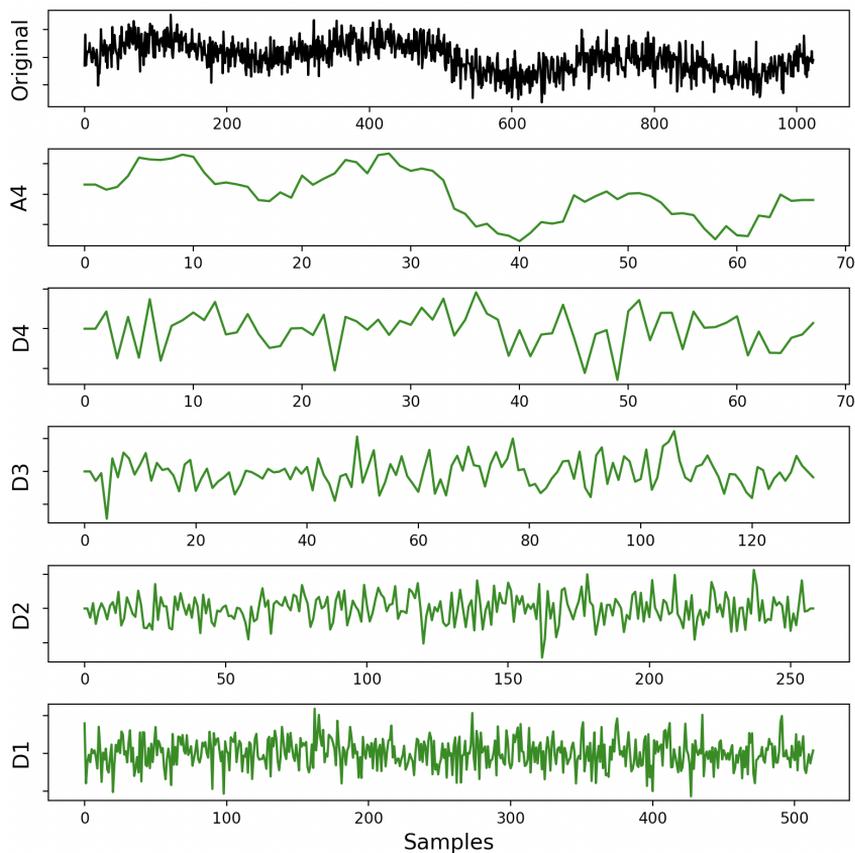


Figure 3.3: Details and approximation coefficients extracted from the original signal using DWT with four levels of decomposition and the mother wavelet biorthogonal 1.3.

$$E_s = \langle x(n), x(n) \rangle = \sum_{n=-\infty}^{\infty} |x(n)|^2 \quad (3.6)$$

There are several approaches for computing the energy distribution, which has been used for feature extraction in various signal processing applications, including those for audio and EEG signals [131–133]. In EEG, the features to represent the energy distribution can be computed to reduce the computational cost and obtain a better representation of the obtained sub-bands by transformation using EMD or DWT.

As shown below, let $w_j(r)$ denote the coefficient of one of the sub-bands (level

of decomposition or IMF) at position r , with N as the length of the sub-band.

The *instantaneous energy* gives the energy distribution in log base 10 of a time series [133], and can be computed in Eq. 3.7:

$$f_j = \log_{10} \left(\frac{1}{N_j} \sum_{r=1}^{N_j} (w_j(r))^2 \right) \quad (3.7)$$

The Teager energy is a robust parameter, as it attenuates auditory noise [131–133]. This log base 10 energy operator reflects variations in both amplitude and frequency of the signal, which is computed as in Eq. 3.8:

$$f_j = \log_{10} \left(\frac{1}{N_j} \sum_{r=1}^{N_j-1} \left| (w_j(r))^2 - w_j(r-1) * w_j(r+1) \right| \right) \quad (3.8)$$

There are more approaches for computing different values of energy features, but these two parameters have proven to be robust for representing the sub-bands of EEG signals [87, 132–135].

Fig. 3.4, presents the average value and standard deviation of the Teager and instantaneous energy distribution of the IMFs from EMD and the levels of decomposition using DWT from Figs. 3.2 and 3.3.

3.3.2 Fractal dimension

A fractal is an irregular geometric object that exhibits similar patterns at increasingly small scales called self-similarity. A fractal dimension is a ratio providing a statistical index of complexity comparing how details in a pattern change with the scale at which it is measured. It is used to measure the *roughness* of a signal, i.e., a *mild* or *wild* randomness, and the complexity of an EEG signal can be directly evaluated by its fractal dimension [136].

There are several self-similarity features from fractal geometry that are useful in describing the complexity of an EEG signal and they have been shown to be highly insensitive to noise [137]. Some have been used to directly characterize EEG signals from raw data or using various methods to extract the information [87, 136, 138]. In particular, Higuchi and Petrosian fractal dimensions have been used to characterize non-linear and non-stationary data [87, 137–141].

The **Higuchi fractal dimension** algorithm approximates the mean length of the curve using segments of k samples and estimates the dimension of a

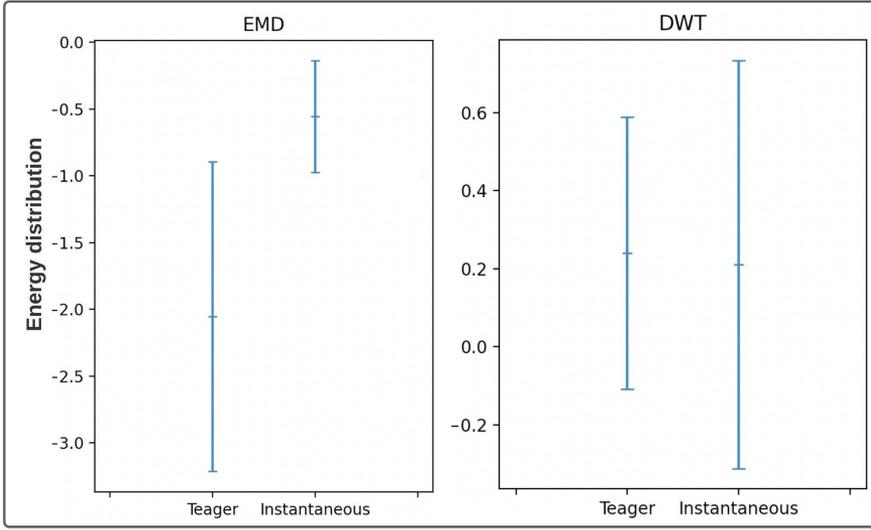


Figure 3.4: Teager and Instantaneous energy distribution of EMD and DWT sub-bands from Figs. 3.2 and 3.3.

time-varying signal directly in the time domain [142]. Consider a finite set of observations taken at a regular interval: $X(1), X(2), X(3), \dots, X(N)$. From this series, a new one X_k^m must be constructed,

$$X_k^m : X(m), X(m+k), X(m+2k), \dots, X\left(m + \left(\frac{N-m}{k}\right)k\right) \quad (3.9)$$

Where $m = 1, 2, \dots, k$, m indicates the initial time, and k the interval time. Then, the length of the curve associated with each time series X_k^m can be computed as follows:

$$L_m(k) = \frac{1}{k} \left(\sum_{i=1}^{\frac{N-m}{k}} \left(X(m+ik) - X(m+(i-1)k) \right) \right) \left(\frac{N-1}{\left(\frac{N-m}{k}\right)k} \right) \quad (3.10)$$

Higuchi takes the mean length of the curve for each k , as the average value of $L_m(k)$, for $m = 1, 2, \dots, k$ and $k = 1, 2, \dots, k_{max}$, which is calculated as:

$$L(k) = \frac{1}{k} \sum_{m=1}^k (L_m(k)) \quad (3.11)$$

The Higuchi fractal dimension depends only on the free parameter k_{max} , which represents the maximum number of scales to explore in the process of

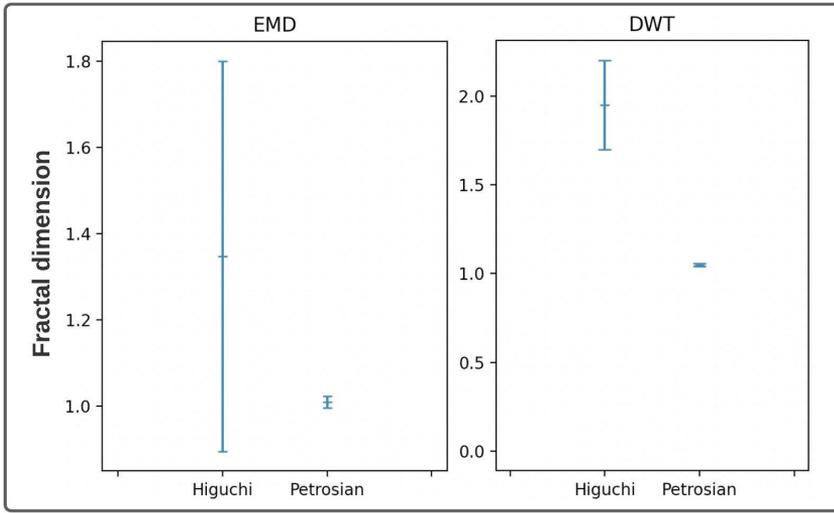


Figure 3.5: Higuchi and Petrosian fractal dimension of EMD and DWT sub-bands from Figs. 3.2 and 3.3.

calculation. In this thesis, it was set at $k_{max} = 10$, but different values have been used when working with brain signals [143–145].

The **Petrosian fractal dimension** can be used to provide a rapid computation of the fractal dimension of a signal by translating the series into a binary sequence [146].

$$FD_{Petrosian} = \frac{\log_{10} n}{\log_{10} n + \log_{10} \left(\frac{n}{n+0.4N_{\nabla}} \right)} \quad (3.12)$$

Where n is the length of the sequence and N_{∇} is the number of sign changes in the binary sequence.

Fig. 3.5, presents the Higuchi and Petrosian fractal dimension of the IMFs from EMD and the levels of decomposition using DWT from Figs. 3.2 and 3.3. It presents the average value and the standard deviation of the fractal dimension values from all the IMFs or levels of decomposition. Using this process, a visual comparison between the fractal features of EEG signals from different classes is easy to interpret, as presented in [141]. However, for the interest of this thesis, this process will be accomplished using machine learning algorithms, as explained later.

3.4 Computational intelligence methods for classification

Machine learning is a well-known research area defined as computational methods using experience to improve performance or to make accurate predictions. Supervised learning is the task of learning or inferring a function from labeled training data of a set of training examples [147].

Deep learning algorithms have been shown to be successful in image processing and other fields, but have not shown convincing or consistent improvement when using EEG data over the most advanced current methods. In addition, its performance depends on the use of a large number of instances, something that is not common when using EEG data [148–151]. Below, a set of methods that have been shown to be effective with little training data is described [148, 152–155].

3.4.1 Multi-class classification

Machine learning gives computers the ability to learn from experience by using supervised or unsupervised learning [156]. Using machine learning, it is possible to train models for predicting the labels or classes of new inputs. Considering X as the sample space and Y as the target space, the goal is to construct a function that predicts Y from X . There are several approaches using supervised learning of interest for this thesis, which are described below:

- **Support Vector Machine or SVM:** This approach uses hyperplanes to separate classes of data by maximizing the margins, which are the distances between the nearest training points from different classes. The hyperplane is defined by vectors called support vectors. SVM has the advantage of transforming nonlinear data to higher-dimensional space for easier separation using the kernel trick and is therefore flexible in representing complex functions while providing a global solution. There is a linear kernel and there are nonlinear kernels, such as the radial basis function (RBF), sigmoid, and polynomial. The classification complexity does not depend on the dimensionality of the feature space and the sensitivity to the number of features is relatively low [157], as the necessary time to create a model is $O(N^3)$, where N is the length of the feature vector and $O(1) + O(N)$ is required to predict the class of a new instance using the created model [158].

- **k-nearest neighbors (KNN):** This algorithm does not attempt to construct a general internal model. Instead, it stores instances of the training data, so no learning is required. The k data points most similar to a new data point from the training dataset are localized [159, 160]. A prediction is then obtained by majority voting applied over the k -nearest data points. The learning is based on the k -nearest neighbors, where k is an integer value that must be specified and the optimal choice of the k value is highly data-dependent. A large k suppresses the effect of noise but makes the classification boundaries less distinct [161].
- **Random Forest (RF):** This is an ensemble learning algorithm, meaning it generates classifiers and aggregates their results. It consists of several decision trees (DT), each giving a prediction, and the class with most votes becomes the models' prediction. Each node is split using the best subset of predictors randomly chosen at that node. RF has been shown to outperform SVM and KNN and is robust against over-fitting [162]. Two parameters must be defined for RF, the number of trees in the forest and the number of variables in the random subset at each node, but it is not very sensitive to such values [163].
- **Naive Bayes (NB):** This is a probabilistic classifier based on Bayes' Theorem. The simple form of the calculation for Bayes Theorem is as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3.13)$$

where $P(A|B)$ is the probability of interest. Bayes Theorem assumes that each input variable depends on all other variables, which causes complexity in the calculation. Removing the assumption of dependency and considering each input variable to be independent from each other simplifies the calculation. An advantage of NB is fast computing when making decisions and it does not require large amounts of data before learning can begin [164].

3.4.2 One-class classification

A one-class classification (OCC) algorithm consists of identifying objects of a specific class among all objects by learning from a training set that contains only

the objects of the target class. This task can be more challenging than a multi-class classification problem, as it is assumed that information for only one of the classes is available, and the boundary between normal and abnormal data has to be estimated solely from normal data in such a way that as many target objects as possible are accepted while minimizing the possibility of accepting outliers [165].

3.4.2.1 One-class Support Vector Machine

In SVM, the input data is represented in an N -dimensional space, where N is the number of features. The algorithm seeks to find a decision boundary or a hyperplane that can separate the data points into classes. The distances from each point to the decision boundary are called *support vectors*. The algorithm searches for the decision boundary with *maximised margins*, that is the boundary that maximizes the sum of the support vectors. In one-class SVM (OCSVM), which is an unsupervised algorithm, this translates to identifying the smallest *hypersphere* (with radius r , and center c) that consists of all data points belonging to the class. The model infers the properties of the training set, and from these properties it can predict which trials from a test set are different from the training set.

OCSVM learns a decision function for outlier detection, classifying new data as similar to or different from that of the training set. As in SVM, different kernels can be used and certain important parameters require fitting, including the ν and γ parameters. The ν parameter is an upper bound on the fraction of training errors and a lower bound of the fraction of support vectors that should be in the interval $[0, 1]$. γ defines how much influence a single training example has: the larger the γ , the closer other examples must be to be affected and the interval must be greater than 0; normally it is $1/\text{no_features}$.

A grid search can be used to adjust the parameters by cross-validation, which has been shown to be powerful and able to significantly improve the results. However, it is a very slow process [166]. These parameters differ depending on the size of the feature vector and it is necessary to re-compute them each time. To illustrate this point, Fig. 3.6 presents an example of two different decision boundaries in OCSVM obtained by using different ν and γ parameters with a random dataset of 100 trials for training (two features per trial), 30 new regular trials, and 30 new abnormal trials. The results obtained clearly show that OCSVM can be sensitive to these values and they must be fitted correctly to obtain

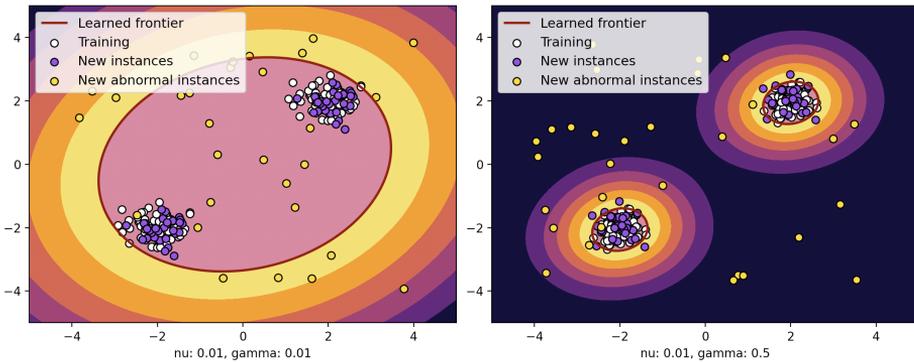


Figure 3.6: Example of two different decision boundaries in OCSVM and a random dataset with outliers.

generalized results. They also show that the learned frontier better fits the training set when the recommended gamma parameter ($1/\text{no_features}$) is used.

3.4.2.2 Local Outlier Factor

Local Outlier Factor (LOF) is a density-based unsupervised outlier detection algorithm that defines the degree of being an outlier by calculating the local deviation of a given data point with respect to its surrounding neighborhood. The score assigned to each data point is called the *local outlier factor* [167]. It is based on a concept of local density given by the distance of the k -nearest neighbors. Comparing the local density of a data point with the local densities of its k neighbors, it is possible to identify regions with similar density and outliers, which have lower density: the lower the density of a data point, the more likely it is to be identified as an outlier. A small k has a more local focus, and a large k can miss local outliers. *Brute force*, *ball tree*, or *k-d tree* algorithms can be used to compute the nearest neighbors.

The k -distance is the distance of a point to its k^{th} neighbor and the reachability distance is the maximum of the distance of two points (i.e., $\text{distance}(a, b)$) and the k -distance of the second point (i.e., $k_distance(b)$), as presented in Eq. 3.14.

$$\text{reach_dist}(a, b) = \max\{k_distance(b), \text{distance}(a, b)\} \quad (3.14)$$

The reachability distance of a to all its k nearest neighbors has to be calculated

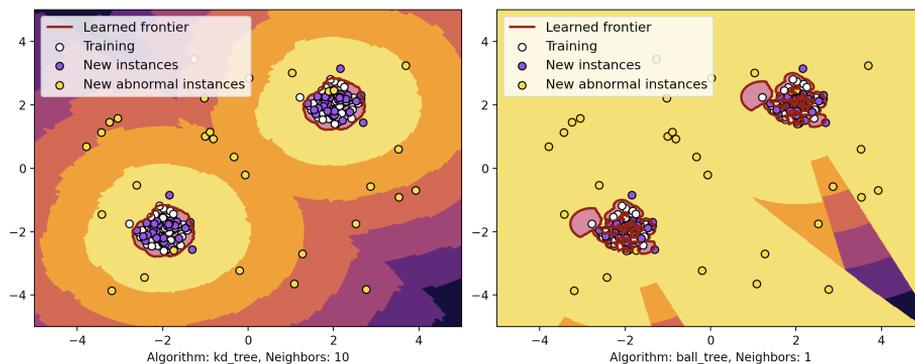


Figure 3.7: Example of two different decision boundaries using LOF and a random dataset with outliers.

and then the average of that number obtained. Thus, the local reachability density (LRD) can be calculated, which is the inverse of the obtained average, as presented in Eq. 3.15. The *LRD* indicates the distance that must be traveled from a point to reach the next point (or cluster of points): the lower it is, the less dense it is, and the longer the distance.

$$LRD(a) = 1 \left/ \left(\frac{\sum_{b \in N_k(a)} reach_dist_k(a, b)}{|N_k(a)|} \right) \right. \quad (3.15)$$

The *LRD* of each point is then compared to the *LRD* of its k neighbors. The LOF is the average ratio of the *LRD*s of the k neighbors of a to the *LRD* of a , as shown in Eq. 3.16.

$$LOF_k(a) := \frac{\sum_{b \in N_k(a)} \frac{LRD_k(b)}{LRD_k(a)}}{|N_k(a)|} \quad (3.16)$$

A ratio < 1 indicates a denser region, which means that the point is an inlier, whereas a ratio > 1 indicates that the point is an outlier. Fig. 3.7 presents an example of two different decision boundaries of the LOF obtained by using different algorithms and numbers of neighbors with a random dataset of 100 trials for training (two features per trial), 30 new regular trials, and 30 new abnormal trials.

3.4.3 Evaluation of classifier performance

Evaluating a classifier's performance, which is performed during the learning process, provides information about how good or bad the followed method is, compares the results with other proposals, and generalizes the results [168]. There are several parameters that can be calculated, depending on the approaches followed, i.e., some for multi-class classification and others for one-class classification approaches. Relevant metrics for the validation of the proposals are presented below.

3.4.3.1 K-fold cross-validation

This method splits a dataset into k folds. One is then used as the test set and the rest as the training set. The number of trials per class must be the same or similar in each fold. The model is trained using the training set and scored using the test set. Then, the process is repeated until each unique group has been used as the test set. Thus, every data point is used $k - 1$ times as part of the training set and one time as a test set. Through cross-validation, an unbiased evaluation of the model can be obtained without reducing the training dataset.

The choice of k is usually 5 or 10, but the bias is smaller for $k = 10$ than $k = 5$. However, there is no general rule. As k gets larger, the difference in size between the training set and the re-sampling subsets gets smaller. The most common value used for cross-validation is $k = 10$ [168, 169].

3.4.3.2 Evaluation metrics

For evaluation and analysis of the results, a confusion matrix is generally used, which in a multi-class problem is a $m \times m$ matrix, where m is the number of classes in the dataset. The columns in the matrix are the *true classes* and the rows the *predicted classes*.

For example, in a two-class classification problem, lets say A and B , it is obtained 1) true positives (TP), cases in which the classifier correctly predicted instances from A , 2) true negatives (TN), cases in which the classifier correctly predicted instances from B , 3) false positives (FP), cases in which the classifier erroneously predicted instances from B in A , and 4) false negatives (FN), cases in which the classifier erroneously predicted instances from A in B . With such a confusion matrix, the accuracy, specificity, and sensitivity can be computed, as

presented in Eq. 3.17, 3.18, and 3.19.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.17)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3.18)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3.19)$$

An important aspect to consider when evaluating the models is to verify whether the models are over-fitted or under-fitted. A low variance error is obtained when the error using the training set is low but high when validating the model with the test set. This indicates that the model is over-fitted and that it has been too highly adjusted to the training set, adopting its variability. A solution to avoid over-fitting may be to add more training data or adjust the classifier parameters. Another problem is called bias-error, which is when the error of the model with both the training set and testing set is high, indicating that the model is not able to adjust to the dataset or is under-fitted. Depending on the nature of the dataset and the classifier, this problem can be avoided by considering longer training times, lower learning rates, more layers, etcetera [170].

For one-class problems, there are several metrics that can be computed. Particularly for biometric systems, the *true acceptance rate*, or TAR, and *true rejection rate*, or TRR, are important and among the most widely used metrics for evaluating models. The TAR is the percentage of times the system correctly verifies a true claim of identity and the TRR the percentage of times it correctly rejects the subjects that are not in the system.

3.5 Channel reduction and selection

While a laboratory setting and research-grade EEG equipment ensure a controlled environment and high-quality multiple-channel EEG recording, there are applications, situations, and populations for which this is not suitable. Conventional EEG is challenged by a high computational cost, high-density, immobility of the equipment, and the use of inconvenient conductive gels.

The main objectives for channel reduction and selection are to **1)** reduce the

computational cost for EEG signal processing, 2) reduce the over-fitting that can occur due to the use of unnecessary channels and improve the classification accuracy, since a large number of channels can contain redundant or useless information, 3) identify the brain areas that generate task-dependent activity, and 4) reduce preparation time. All of these objectives can be achieved by selecting the most relevant channels and removing task-irrelevant and redundant channels, thus extracting the most relevant features [171, 172].

An important point is that selection of a low number of channels can result in a low-power hardware design. This would allow expansion of the range of applications of EEG signals from clinical diagnosis and research to healthcare, a better understanding of cognitive processes, learning and education, and currently hidden/unknown properties behind ordinary human activity and ailments (i.e., resting-state, walking, sleeping, complex cognitive activity, chronic pain, insomnia, etc.) [173].

Various channel reduction and selection methods have been tested for extracting channel subsets, ranging from algorithms, such as filtering, wrapper, embedded, and hybrid methods [171, 172, 174–189] to the use of genetic algorithms, such as the simple GA, steady-state genetic algorithm, genetic neural mathematics method (GNMM), artificial bee colony (ABC) algorithm, and NSGA-based algorithms [87, 138, 190–201]. These methods have been generally tested in motor imagery, but a unique set of channels for this task has not been found [172, 174, 176, 179, 188, 196, 198, 199].

In a low-density device, the channel selection approach can be possibly used to modify the channel's position or at least activate the relevant sensors in real-time and, thus, increase classification accuracy and reduce processing time. Two greedy and one multi-objective optimization algorithm of interest for this thesis are presented next.

3.5.1 Greedy algorithms

A greedy algorithm makes the optimal decision at each stage (local optimal or local maximum) and generally does not produce an optimal solution, but this strategy approximates a globally optimal solution in a short period of time [202].

An easy and rapid way to evaluate the most relevant parameters or features for obtaining the best results in a problem is the use of greedy algorithms [202]. The

idea of using greedy algorithms for channel selection is to obtain all combinations, removing 1 channel at a time, and selection of the subset with the best results, which represents the local maximum. The procedure is then repeated using the obtained subset while the length of the subset is still greater than 1 channel.

The same process can be applied but first after selecting the single channel with the best results. The process is then repeated trying to add another channel and selecting the subset of two channels with the best results. The process is repeated, adding additional channels until all the channels have been added to the subset. This method provides a general idea of the channels with the most useful information for the classifiers.

These methods are known in combinatorial optimization and artificial intelligence as *backward-elimination* and *forward-addition* algorithms and have been used in feature subset and channel selection [173, 203–206]. Both methods provide an optimal solution at each step, but neither is able to predict complex iterations between channels or features that may affect the performance of the classifier, which is why they are not considered to be a global solution.

3.5.2 Multi-objective optimization methods

An optimization problem consists of maximizing or minimizing a function by systematically choosing input values from a valid set and computing the value of the function, which can be limited to one or more restrictions, or it can be without any restriction. In an optimization problem, the model is feasible if it satisfies all the restrictions and it is optimal if it also produces the best value (minimum or maximum) for the objective function.

A Multi-objective optimization problem (MOOP) has two or more objective functions that are to be either minimized or maximized. As in a single-objective optimization problem, a MOOP may contain a set of constraints, which any feasible solution must satisfy [207]. Eq. 3.20 presents a MOOP in its general form.

$$\begin{aligned}
& \text{Minimize/Maximize} && f_m(x), && m = 1, 2, \dots, M \\
& \text{subject to} && g_j(x) \geq 0, && j = 1, 2, \dots, J \\
& && h_k(x) = 0, && k = 1, 2, \dots, K \\
& && x_i^{(L)} \leq x_i \leq x_i^{(U)}, && i = 1, 2, \dots, n
\end{aligned} \tag{3.20}$$

As a result of the optimization process, a set of solutions is obtained, where a solution $x \in \mathbf{R}^n$ is a vector with n decision variables, $\mathbf{x} = [x_1, x_2, \dots, x_n]$. The objective functions constitute a multi-dimensional space called the objective space, or $Z \subset R^M$. For each solution x in the decision variable space, there is a point $\mathbf{z} \subset R^M$ in the objective space, denoted by $f(x) = \mathbf{z} = [z_1, z_2, \dots, z_M]$.

3.5.2.1 Non-dominated sorting genetic algorithms (NSGA)

Genetic algorithms (GAs) mimic Darwinian evolution and use biologically inspired operators. Their population is comprised of a set of candidate solutions, each with chromosomes that can be mutated and altered. GAs are normally used to solve complex optimization and search problems [208].

GAs normally consists of **1)** population initialization, **2)** fitness function calculation, **3)** crossover, **4)** mutation, **5)** survivor selection, and **6)** termination criteria to return the best solutions. The population consists of a set of chromosomes that are possible solutions to the problem and each chromosome can have as many genes as variables in the problem. There are various proposed methods in the state-of-the art for each stage [208–211].

For the genetic representation of the solution domain, it is possible to define chromosomes using genes with binary values, i.e., 0 or 1, as well as those with integer or decimal values. For example, if the *gamma* parameter of OCSVM has to be optimized, it can be defined as a gene with decimal values in the interval $[0, 1]$.

The non-dominated sorting genetic algorithm, or NSGA [210], uses a non-dominated sorting ranking selection method to emphasize good candidates and a niche method to maintain stable sub-populations of good points (Pareto-front), where a non-dominated solution is a solution that is not dominated by any other solution. NSGA-II was used to solve certain problems related to computational complexity, the non-elitist approach, and the need to specify a sharing parameter

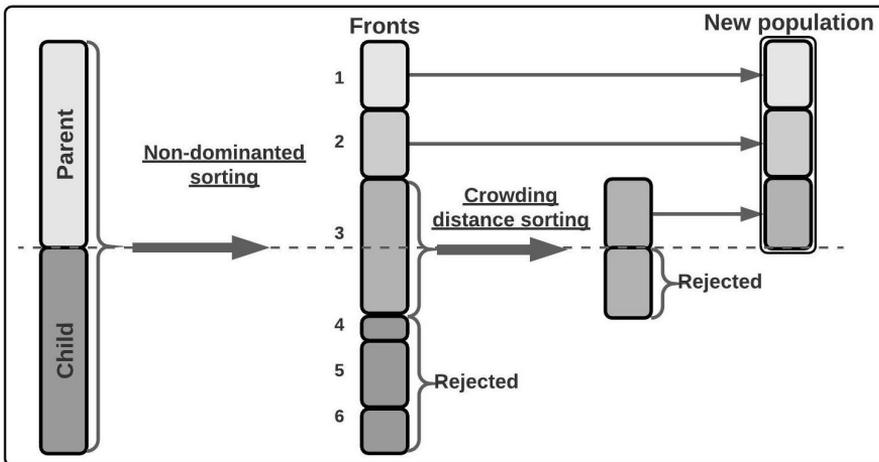


Figure 3.8: An illustrative example of the NSGA-II procedure [211].

to ensure diversity in a population presented in the first version. NSGA-II also reduced the computational cost from $O(MN^3)$ to $O(MN^2)$, where M is the number of objectives and N the population size. Additionally, the elitist approach was introduced by comparing the current population with the previously found best non-dominated solutions [211].

Fig. 3.8 presents the NSGA-II framework, in which parent and child populations are compared using the fitness function and organized using the non-dominated sorting algorithm for creating different fronts, from high to low importance. Then, the individuals in the first front are selected to be used in the next generation. There are situations in which a front has to be split (In Fig. 3.8, front 3) because not all individuals are allowed to survive. In this split front, solutions are selected based on crowding distance [211].

NSGA-III has been shown to efficiently solve 2- to 15-objective optimization problems [212]. NSGA-III follows the NSGA-II framework but uses a set of predefined reference points that emphasize population members that are non-dominated, yet close to the supplied set [212, 213]. The predefined set of reference points are used to ensure diversity in the obtained solutions. When using NSGA-III, the reference points are generally places on a normalized hyper-plane that is equally inclined to all objective axes and has an intersection with each. For example, in a three-objective optimization problem, the reference points are

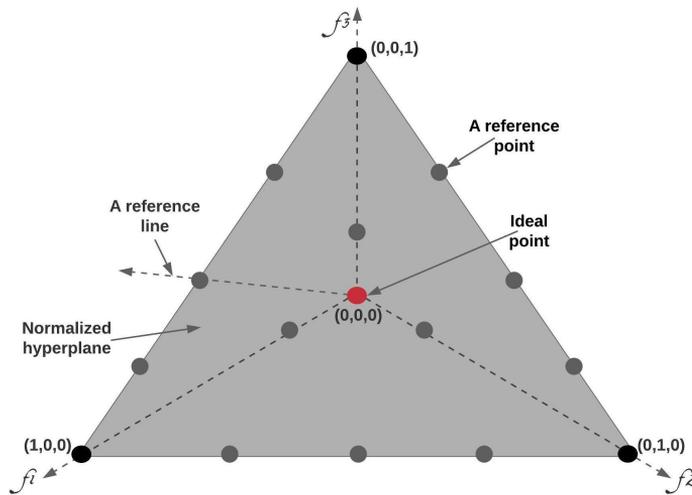


Figure 3.9: Reference points of NSGA-III in a three-objective optimization problem.

created on a triangle with apexes at $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ [213, 214], as shown in Fig. 3.9.

3.6 Description of datasets used in the thesis

3.6.1 CHB-MIT dataset

Most of the proposed methods for epileptic seizure classification in the state-of-the-art are tested on datasets from the PhysioNet [215] and EPILEPSIAE [216] projects and the TUH EEG Corpus [217], in which some of the datasets consist of private repositories or to which access is limited.

The EEG recordings used were obtained from pediatric patients with intractable seizures who were monitored for several days at the Boston Children's Hospital following the withdrawal of anti-seizure medication to characterize their seizures and assess their candidacy for surgical intervention. The dataset used comes from the PhysioNet project and is partially described in [215, 218] and can be found in the [CHB-MIT Scalp EEG Database](https://doi.org/10.13026/C2K01R) or doi.org/10.13026/C2K01R. The dataset consists of bipolar EEG signals from 24 patients that were recorded using 22 channels (FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, P8-O2, FZ-CZ, CZ-PZ, P7-T7, T7-FT9, FT9-FT10, FT10-T8, and T8-P8), with a sampling rate of 256 Hz, using the 10-20

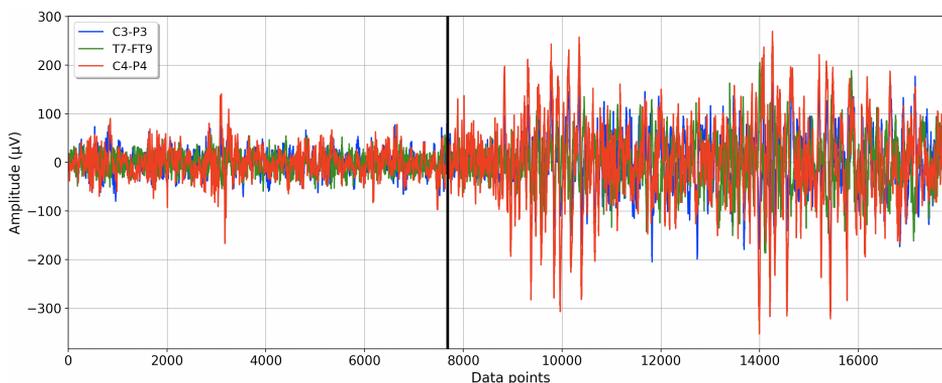


Figure 3.10: Example of the raw EEG data of *C3-P3*, *T7-FT9* and *C4-P4* channels from the third instance of Patient 1 of the CHB-MIT dataset.

international system. It should be noted that channels FT9 and FT10 are not part of the 10-20 international system.

The EEG data for each epileptic seizure and epileptic-free period is of six seconds and there are an average of 80 instances for each class for each patient. More details can be found in [135, 215, 218], and in the [CHB-MIT Scalp EEG Database](#).

Certain important details are shown in Table 3.1, including the duration (in seconds) of the EEG signal for each epileptic event. However, six-second segments of the epileptic seizures are also considered to compare the seizures between subjects with similar components.

Fig. 3.10 presents the raw EEG signal of an epileptic seizure and 30 seconds before onset (the onset is indicated by a vertical line in black) of the first instance of subject 1, showing the EEG data corresponded to *C3-P3*, *T7-FT9* and *C4-P4* channels.

3.6.2 EEGMMIDB dataset

This dataset consists of EEG signals of 109 subjects collected from 64 EEG channels, localized according to the 10-10 international system, with a sample rate of 160 Hz and a recorder using the BCI2000 system. The public motor movement/imagery dataset (EEGMMIDB) is part of the PhysioNet project [215].

Each subject performed two one-minute resting-state runs, one with the eyes

Table 3.1: Details of the epileptic-seizure data presented in [218].

Patient	Gender	Age	Seizures	Length in seconds			
				Average	Max	Min	Segments of 6 s
1	F	11	7	63.1	101	27	74
2	M	11	3	57.3	82	9	29
3	F	14	7	57.4	69	47	67
4	M	22	4	94.5	116	49	63
5	F	7	5	111.6	120	96	93
6	F	1.5	7	15.6	20	12	18
7	F	14.5	3	108.3	143	86	54
8	M	3.5	5	183.8	264	134	153
9	F	10	4	69.0	79	62	46
10	M	3	7	63.9	89	35	74
11	F	12	3	268.7	752	22	134
12	F	2	38	36.9	97	13	234
13	F	3	12	44.6	70	17	89
14	F	9	8	21.1	41	14	28
15	M	16	20	99.6	205	31	332
16	F	7	6	8.8	14	6	9
17	F	12	3	97.7	115	88	49
18	F	18	6	52.8	68	30	53
19	F	19	3	78.7	81	77	39
20	F	6	8	36.8	49	29	49
21	F	13	4	49.8	81	12	33
22	F	9	3	68.0	74	58	34
23	F	6	10	60.6	113	20	101
24	–	–	13	31.9	70	16	69
Sum			189				1925
Mean			7.9	74.2	121.4	41.3	
Max					752		
Min						6	

open and one with the eyes closed. Then, three two-minute runs were carried out for four different tasks: two motor movement tasks and two imagery tasks [219]. The four types of motor movement and imagery tasks were performed for opening and closing the left or right fist, imagining opening and closing the left or right fist, opening and closing both fists or both feet, and imagining opening and closing both fists or both feet according to the position of a target on the screen (Left, right, top, or bottom).

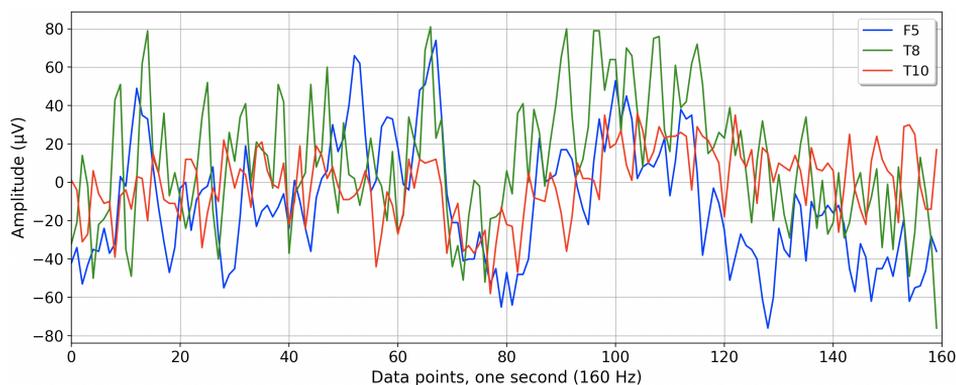


Figure 3.11: Example of the raw EEG data of *F5*, *T8* and *T10* channels of the first instance of subject 1 of the EEGMMIDB dataset.

For the experiments carried out in this thesis, only the two one-minute baseline runs were used to create instances of one second, obtaining 60 instances of one second in the resting-state with the eyes open and 60 instances of one second in the resting-state with the eyes closed for each subject.

Fig. 3.11 presents the raw EEG signal of resting-state with the eyes open of the first instance of subject 1, showing the EEG data corresponded to *F5*, *T8* and *T10* channels.

3.6.3 P300-speller dataset

This dataset consists of EEG signals from 26 subjects (24 right-handed and 2 left-handed), with an average age of 29.2 ± 5.5 years, from 56 passive Ag/AgCl EEG electrodes that were placed following the extended 10-20 international system. The EEG signals were all referenced to the nose and the ground electrode was placed on the shoulder, the impedance was kept below 10 k Ω . The EEG data was collected during five sessions and consist of 60 instances per session, with a sample rate of 600 Hz, that were down-sampled at 200 Hz [220].

The protocol used to record the EEG signals used the P300-speller paradigm (as is illustrated in Fig. 3.12) and introduced in [220]. Briefly, the target letter (the letter to be presented) is indicated by a green circle for one second. Then, letters and numbers (6 X 6 items, 36 possible items displayed on a matrix) are flashed in groups of six characters. Next, the display remains blank for a period of 2.5

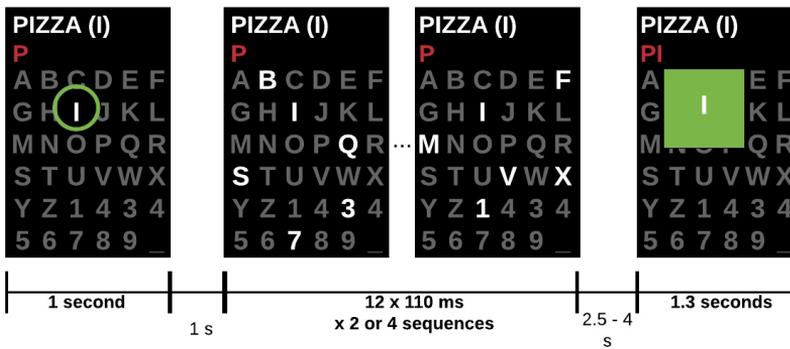


Figure 3.12: Protocol design for recording positive or negative feedback-related responses in the P300-speller dataset [220].

to 4 s, representing the resting-state. During this random period, the subjects are requested to remember the letter displayed. Then, the letter chosen by the implemented P300 classifier is displayed for 1.3 s. If the presented letter is the one that was previously presented, the subject sends a positive response; otherwise, the subject sends a negative response.

An example of a positive feedback-related response corresponding to the target letter *i* is shown in Fig. 3.12. For the experiments carried out, only the positive-feedback responses were used. Thus, the number of positive-feedback trials can be different between subjects and sessions. The minimum number of positive-feedback related responses was selected, which was 25 instances per session per subject. Fig. 3.13 presents the raw EEG signal of the first instance of subject 1, showing the EEG data corresponded to *P7*, *P8* and *T8* channels.

3.7 Methods proposed in the thesis

This section describes the general flowchart of the proposal presented in Fig. 3.1 but it may differ, depending on the dataset used and the application. Thus, more details are added for each case in the following Chapters.

3.7.1 Pre-processing, feature extraction and classification

The CAR method was applied to the EEG data and then EMD or DWT methods for decomposing the EEG signals into different sub-bands were applied. After decomposing the EEG signals, two energy values (Teager and instantaneous

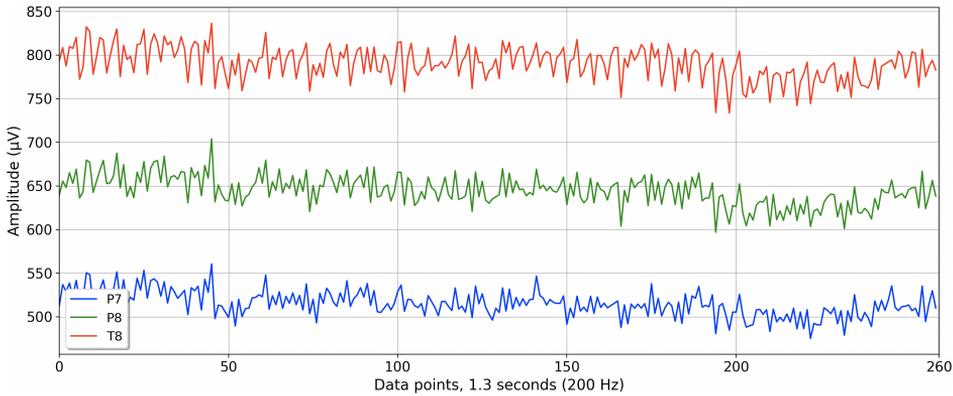


Figure 3.13: Example of the raw EEG data of *P7*, *P8* and *T8* channels of the first instance of subject 1 of the P300-speller dataset.

energy) and the two fractal dimension features (Higuchi and Petrosian fractal dimension) were computed for each sub-band.

EMD was tested using various numbers of IMFs but only the two closest IMFs were used based on the Minkowski/Euclidean distance because they have been shown to provide the same performance as that of using more. For DWT, the 2.2 mother function bi-orthogonal, with four levels of decomposition, was used based on the results obtained from previous studies [86, 87, 135, 138, 173, 221–223]. The process for extracting four features for each selected IMF returns eight features per channel or 20 features per channel when using DWT. The process is repeated for each channel used and then concatenated to obtain a single vector of features that represents the EEG signal for each instance. Figs. 3.14 and 3.15 present the flowchart of the process followed for DWT and EMD, respectively.

Different classifiers for creating the machine-learning models were tested using the obtained feature vectors for each instance, depending on the application and experiment. In general, the process can be summarized as in Fig. 3.16, in which the training and testing sets were separated after obtaining the features from the EEG dataset, whenever possible. The training set was used to create the machine-learning model using 10-fold cross validation and the model validated using the testing set, which was 20% of the dataset. Using this approach, the metrics can be obtained for evaluating the performance of the method in each experiment, consisting of the accuracy and standard deviation from the 10-fold

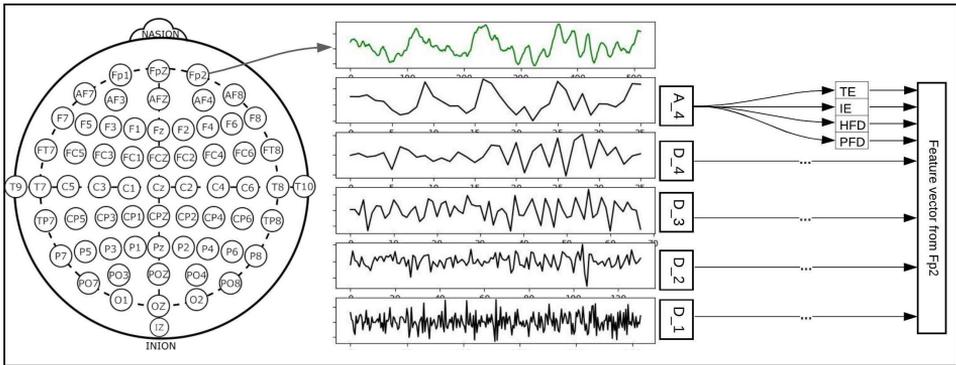


Figure 3.14: Flowchart summarizing feature extraction using DWT.

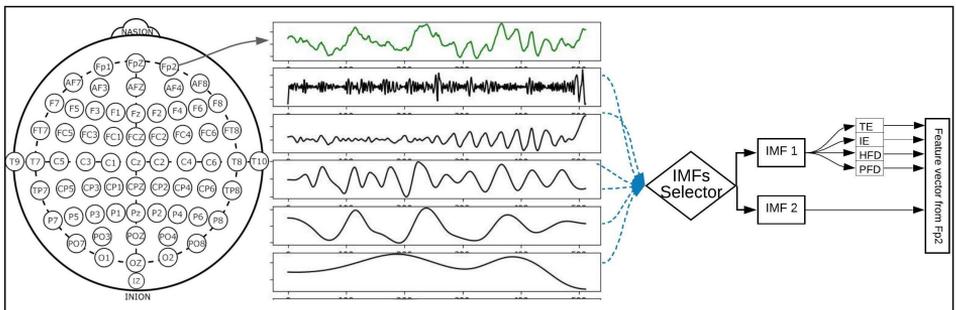


Figure 3.15: Flowchart summarizing the feature extraction procedure using EMD.

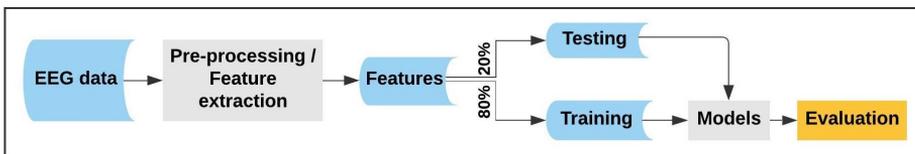


Figure 3.16: Flowchart of the procedure followed for EEG signal classification.

cross-validation, as well as the accuracy and standard deviation from the testing set.

3.7.2 General overview of the proposed method

The flowchart presented in Fig. 3.16 is for a single iteration of the method, but the purpose of the proposal is to repeat this process several times to reduce the number of necessary channels while increasing, or at least maintaining, the

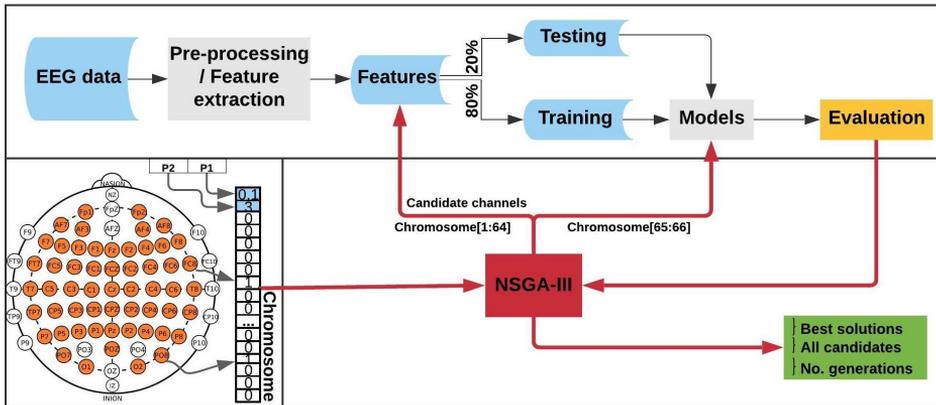


Figure 3.17: Example of chromosome representation and flowchart of the optimization process for parameter optimization and EEG channel selection using NSGA-III.

performance. Additionally, it is also necessary to optimize certain parameters for certain classifiers.

Fig. 3.17 presents an example of the process for feature extraction and classification, but the entire process can be handled by an optimization algorithm. In the example presented, the process is handled by NSGA-III using a chromosome representation with 64 EEG channels, **1** if the channel will be used and **0** if not, and two genes to optimize the parameters of the model (indicated as *P1* and *P2*), one with integer values (which can be, for example, from 0 to 5) and the other with decimal values (which can be from 0 to 1).

The parameters of the classifier can be tuned using simple methods, such as grid search [224], but they need to be tuned to the model under specific circumstances and for a specific number of channels. In this case, the best parameters for the models must be found and this can be accomplished by adding a gene for each parameter to the chromosomes generated by the genetic algorithms.

In the example, the process starts using the raw EEG signals, from which feature extraction is performed and the results organized and stored for iterative use. From this point on, the main process is handled by NSGA-III, which starts creating all possible candidates (chromosomes) for each population. Then, the first 64 genes are used to extract the sub-dataset for the channels, represented as 1 in

the chromosome, and the subset evaluated with the classifiers using genes 65 and 66 to define the classifier's parameters. The best results obtained and the number of EEG channels used is returned to NSGA-III to evaluate each chromosome in the current population. The process is repeated, creating different populations, until the termination criterion is reached.

The termination criterion for the optimization process is defined by the objective space tolerance, which is defined as 0.0001. This criterion is calculated every 5th generation. If optimization is not achieved, the process stops after a maximum number of generations. The definition of the problem to optimize, the number of objectives, the size of each population in each iteration, and the maximum number of generations are defined for each experimental configuration in Chapters 4 and 5.

3.8 Hardware and software tools used in the thesis

Free public EEG datasets, as well as tools and libraries for creating the code on python3 [225], were used. Implementation of the classifiers was based on the *scikit-learn* python library [226] and the NSGA algorithms on *pymoo* [227].

Other important python libraries used included *Dask* (for task distribution using parallel computing), *Scipy*, and *Numpy* [228–230]. For the implementation of EMD and DWT, the *PyWavelets* and *pyhht* libraries were used [231, 232].

Most of the experiments in which optimization with NSGA was used were carried out on the NTNU IDUN computing cluster [233]. The cluster has more than 70 nodes and 90 GPGPUs. Each node contains two Intel Xeon cores and at least 128 GB of main memory and is connected to an Infiniband network. Half of the nodes are equipped with two or more Nvidia Tesla P100 or V100 GPGPUs. Idun storage is provided by two storage arrays and a Lustre parallel distributed file system.

Chapter 4

Case study 1: Channel count optimization for Epileptic seizure classification

In this Chapter, the proposed method for feature extraction is implemented for representing epileptic seizures and seizure-free periods. Different classification algorithms are tested and compared using the obtained features. The main objective of this thesis, which is reduction of the number of required EEG channels, is assessed by implementing various channel-reduction and selection methods using greedy and multi-objective optimization algorithms.

This Chapter is based on the journal articles [135, 200] and mainly addresses the 1st and 2nd research questions and partially the 3rd.

4.1 Introduction

Epilepsy is a group of neurological disorders, characterized by recurrent epileptic seizures, that affects approximately 1% of the world's population of all ages, both sexes, and all races and ethnic backgrounds [234]. It consists of widespread electrical discharges of a set of neurons inside the brain [235]. Epileptic seizures are normally detected by continuous monitoring of EEG signals; the epileptiform can be categorized into ictal, interictal, and postictal periods. The identification of seizures by visual inspection can be time-consuming and lead to an incorrect interpretation of EEG signals, which can trigger under/over medication of patients [236].

Suitable methods and proper detection of epileptic seizures could facilitate the rapid treatment of patients and improve the diagnosis of epilepsy. Epileptic events are attributed to localized disturbances in various areas of the brain [237]. The epileptogenic focus for approximately 33% of epilepsy patients is located in the temporal lobe and their condition is referred to as temporal-lobe epilepsy (TLE) [238, 239].

4.2 State-of-the-art

Current state-of-the-art efforts attempt to improve the feature extraction stage for correct representation of the seizure and seizure-free periods using machine-learning methods. Several relevant studies using the same public dataset have been published, using various experimental setups. The research and applications for automatic classification and detection of epileptic seizures based on EEG, using supervised, semi-supervised, and deep-learning techniques, have increased during the last few years. However, comparisons between experiments, even using the same datasets, have shown conflicting results.

In one study [240], the authors used iEEG signals from only five subjects, with only 20 epileptic seizures for each. Thus, they had data for only 100 epileptic seizures and EEG signals from the epileptogenic zone during free intervals as seizure-free periods. They reported an accuracy of 99.6% from only one channel using a neural network. However, this approach is known to work better when using a large amount of data during the training process, as neural networks learn only by weight adjustment and require all the possibilities to be adequately trained. In another study, the authors used the same dataset and performed five levels of DWT and fuzzy approximate entropy for feature extraction [241].

The study presented by [242] used relative energy values and normalized variation coefficients from DWT in the feature extraction stage and then linear discriminant analysis (LDA) for classification. The method was evaluated on the data of five subjects of the CHB-MIT dataset, with 23, 24, or 26 channels, depending on the subject and the available data. In the classification process, they used approximately 80% of the data for training and the rest for testing, obtaining an accuracy of 0.91. Later [243] presented a method for feature extraction with even features from the intersection sequence of Poincaré section with phase space using LDA and naive Bayes classifiers. They used 23 channels from the CHB-MIT

dataset, obtaining accuracies of 0.93 using 25% of the data for training and 0.94 using 50%.

The signal curve length of the time-domain EEG signal and the mode powers of dynamic mode decomposition (DMD) were used by [244] for feature extraction using 18 channels of the CHB-MIT dataset, which were manually selected. They reported a sensitivity of 0.87 using approximately 50% of the data for training their models for epileptic-seizure classification.

An approach using EMD to decompose EEG signals into different IMFs and five features for each chosen IMF was presented in [135]. In the aforementioned study, the results of an approach based on channel reduction using the backward-elimination algorithm were presented, obtaining an average classification accuracy of 0.93 when five channels and 10-fold cross-validation were used.

The work presented in [245] used a multivariate extension of the empirical wavelet transform (EWT) to decompose the EEG signal into different oscillatory levels and compute three features for each level. The accuracies obtained ranged from 0.95 to 0.99 using five channels and various classifiers. This method selects the channel with the lowest standard deviation and then the remaining four channels with the highest mutual information (MI) with the previously chosen channel. A method based on 24 feature types and SVM classifiers was presented by [246]. The experiments were performed using the 22 available EEG channels of the TUH EEG Corpus [217] and the accuracy obtained was 0.994.

Several methods have been proposed using various values of entropy for feature extraction [247], EMD for decomposing the EEG signals [248], features based on Fourier-Bessel series expansion [249, 250], and the energy from sub-bands extracted using the Taylor-Fourier filter bank [251]. The proposals used machine learning classifiers [247–251] and neural networks [252]. However, these approaches were tested using the Bonn university EEG database, which consist of a single channel and is based on invasive seizure EEG signals [253].

Based on the previous presented studies, epileptic-seizure classification can still be improved by representing the seizure and seizure-free periods correctly to obtain better results using EEG signals. Certain state-of-the-art methods have been tested on small or single-channel (using iEEG) datasets, showing competitive accuracies for classifying epileptic seizures; however, the use of EEG signals

has only been assessed in experiments using all available channels or manually selected channel arrays.

The feature extraction process and classifier design are important for the classification and detection of epileptic seizures, but the use of only a few EEG channels (without using iEEG) will provide new areas of research and expand potential applications in and outside of hospitals and laboratories. This will require the use of robust EEG channel-selection procedures that will reduce the current limitations of portability, as well as the computational cost to obtain faster results, decreasing possible over-fitting that comes from using all available channels. Recent efforts and improved technology of dry EEG sensors have opened up new possibilities to develop new types of EEG systems [254, 255]. In this context, future efforts will be focused on low-cost portable devices for personal use, reducing the necessary number of EEG channels while maintaining or increasing the accuracy of machine-learning-based algorithms.

In this Chapter, two methods for feature extraction, four classifiers with various parameters, and two-channel selection methods to classify epileptic-seizure and seizure-free periods are analyzed. The process of selecting channels was considered as a multi-objective optimization problem, using the lowest possible number of EEG electrodes and obtaining the highest possible accuracy. The approach was tested on a well-known public dataset, described in Section 3.6.1 [215].

4.3 Definition of the problem to optimize

The problem that requires optimization is the selection of the most relevant and necessary EEG channels for epileptic-seizure classification while increasing or at least maintaining the accuracy of the classifiers. This requires organizing the dataset and a representation of the variables in the GA. NSGA-II and NSGA-III will be used to manage minimization of the objective functions and compare the results using different feature extraction methods and classifiers.

In general, a GA requires a genetic representation of the solution domain and a fitness function to evaluate the solutions domain, which in this case, was an array representing each channel (see Fig. 4.1) and the fitness function for the two-objective optimization problem defined as $[\mathbf{Acc}, \mathbf{No}]$, where \mathbf{Acc} was the classification accuracy obtained with the chromosome and \mathbf{No} the number of EEG channels used.

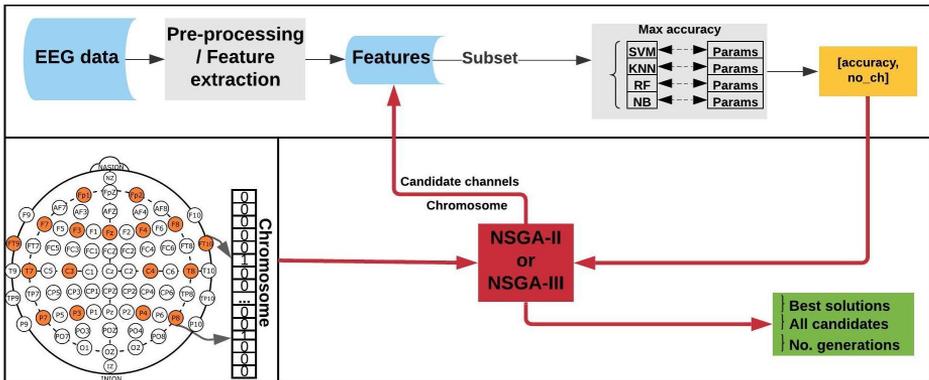


Figure 4.1: Complete process for EEG channel selection using NSGA-II or NSGA-III for epileptic-seizure classification.

Fig. 4.1 shows a binary representation for creation of the chromosomes, with each gene representing a channel, 1 if the channel is used for the classification process and 0 if not. All possible channels that can be used are colored, representing the search space, which is 22, as already mentioned in the description of the dataset in Section 3.6.1. It should be noted that channels FP1-F7, FP1-F3, T7-P7, T7-FT9, P7-T7, P7-O1, FP2-F4, and FP2-F8 were considered to be different, as the references for the channels are different and the dataset provides the EEG signals for each one separately.

All the best solutions found in the optimization process for epileptic-seizure classification were analyzed. There are certain applications that use EEG signals in which the automatic selection of the best solution may be important, especially for cross-subject analysis. Here, however, it was important to analyze all the results for each patient individually. With this assumption, the designer of a potential low-cost EEG headset can consider whether it is better to sacrifice accuracy or the number of EEG channels, depending on how easy or difficult it is to detect epileptic seizures for a given individual.

The problem to be optimized is defined by two unconstrained objectives: first, to maximize accuracy and second, to decrease the number of channels used for epileptic seizure classification. The termination criterion for the optimization process is defined by the objective space tolerance, which is defined as 0.0001. This criterion is calculated every 5th generation and if not achieved, the process stops

after a maximum of 500 generations. Fig. 4.1 shows the complete process, which consists of three main stages: feature extraction, classification, and optimization.

Classification experiments were performed using the characterized EEG signals for each patient separately, while reducing or selecting the EEG channels for creating models to detect epileptic seizures. For each patient, a carefully balanced dataset was created using epileptic-seizure and seizure-free segments of six-seconds (as explained in Section 3.6.1).

The process starts by using the raw EEG signals of one patient at a time, from which feature extraction is performed and the results organized and stored for iterative use (see Fig. 4.1). From this point on, the main process is handled by the NSGA, which starts creating all possible candidates (chromosomes) for each population, obtaining the corresponding subset of features for the channels represented as 1 in the chromosome and evaluating the subset with four different classifiers, with different parameters for each. The best accuracy obtained and the number of EEG channels used is returned to the NSGA to evaluate each chromosome in the current population. The process is repeated, creating different populations, until the termination criterion is reached.

In summary, the chromosome has 22 genes, each representing an EEG channel. Each population size in each iteration is defined as 20, which was selected experimentally. Four classifiers were tested for each possible solution, but only the highest accuracy was retained and the corresponding classifier used stored for analytical purposes.

4.4 Channel selection for Epileptic-seizure classification with EMD-based features

For this experiment, EMD-based feature extraction was used, followed by the greedy algorithm for channel reduction, and both NSGA-II and NSGA-III for channel selection. The process described in Fig. 4.1 was repeated for each patient using the above techniques.

For illustrative purposes, Fig. 4.2 presents the results obtained using NSGA-II for epileptic-seizure classification of patient 1.

Fig. 4.2 clearly shows that NSGA-II managed to cope with both objectives, whereas the opposite was true when using a lower number of channels, although the backward-elimination algorithm sometimes showed higher accuracy when

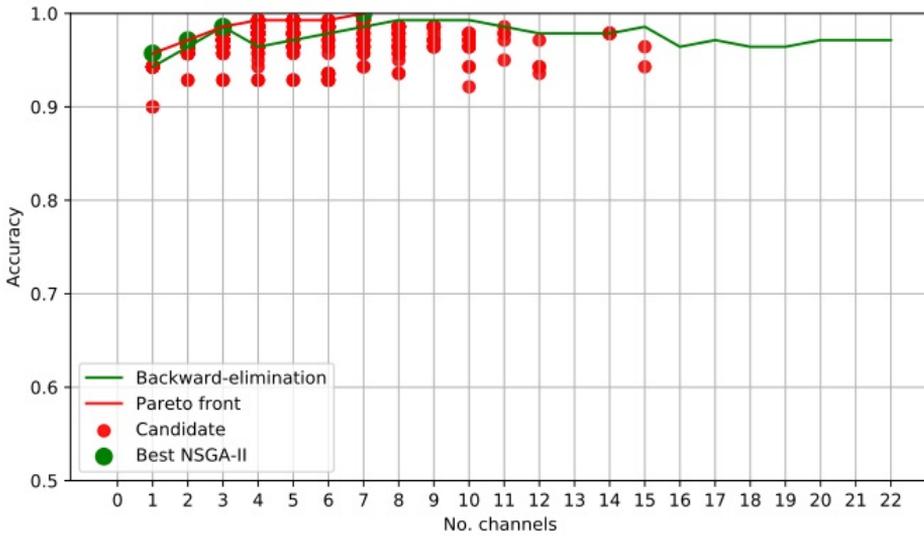


Figure 4.2: EEG Channel Selection for epileptic seizure classification of patient 1 using EMD-based features. Comparison between NSGA-II and the backward-elimination algorithm.

using a high number of channels.

In this case, the best results obtained using NSGA-II consisted of four subsets of channels, which did not necessarily overlap. This is because each chromosome was almost independent and may have come from different parents. The illustrative example presented in Fig. 4.3 shows the subsets of channels used for obtaining the highest accuracy.

Channel Cz was selected in the first four subsets shown using the NSGA-II method, but not when backward-elimination was used. The accuracy obtained by backward-elimination was notably lower than when NSGA-II was used, i.e., 0.964 and 0.993, respectively (see Fig. 4.2), which shows the feasibility of the method, as well as the importance of a robust method for channel selection.

Tables 4.1 and 4.2 show the accuracy obtained using each of the methods on data from all of the patients. Most of the best results were obtained when 10 channels were reduced to one (see Fig. 4.2). The tables show only the results for channels 1 to 10 for all patients, but the experiment was carried out with all channels. As an automatic termination criterion was used, the number of generations for each patient was different and is shown in the tables.

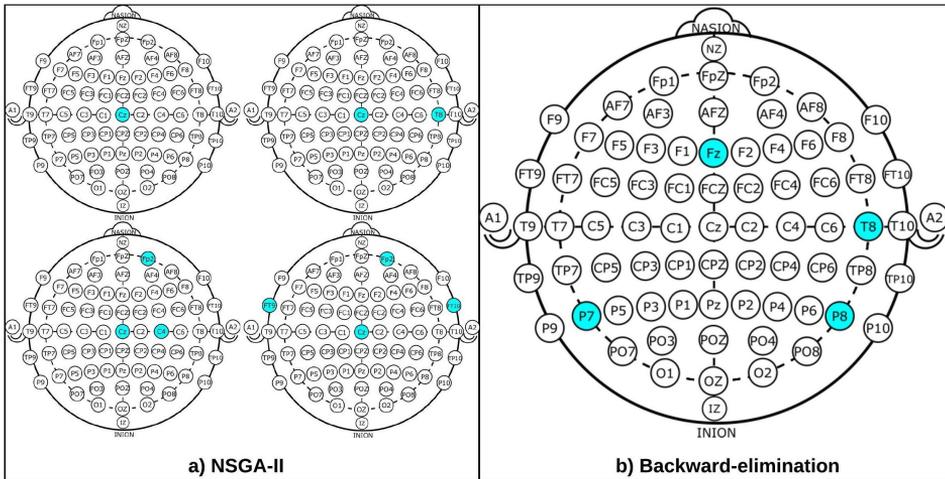


Figure 4.3: Four EEG Channel subsets selected by NSGA-II (a) and backward-elimination (b) for epileptic-seizure classification in patient 1.

Supplementary material in [200] provides data on the accuracy, specificity, and sensitivity for the first four EEG channels of Tables 4.1 and 4.2.

The results highlighted in gray are those for which the accuracy obtained was higher than when using backward-elimination. The average number of generations was 39 ± 12 for NSGA-II and 47 ± 13 for NSGA-III.

Patient 13 appears to be a possible special case, as similar accuracy was obtained with all methods. NSGA-II showed the highest accuracy when using three channels and NSGA-III when using five, reaching 0.813. The addition of more channels to detect epileptic seizures resulted in fluctuations in the accuracy but it did not increase.

Table 4.2 shows a number of empty cells when using NSGA-II and NSGA-III, meaning that the accuracy obtained was not part of the best solutions. This is best illustrated for the results obtained for patient 19 using the NSGA-III method (see Fig. 4.4). This case shows a clear example of how the method works, as the accuracy obtained using two channels was 0.975 but the addition of more channels only decreased the accuracy, except for the use of six channels. This is related to the small amount of information provided by the added channels.

As mentioned previously, the classifier used each time is that resulting in the highest accuracy using the subsets of EEG channels. The NSGA-based algorithms

Table 4.1: Accuracy obtained using EMD for feature extraction with NSGA-II and NSGA-III for EEG channel selection (subjects 1-12).

Id	Method	No. channels									
		1	2	3	4	5	6	7	8	9	10
1	B-E	0.943	0.964	0.986	0.964	0.971	0.979	0.986	0.993	0.993	0.993
	NSGA-II	0.979	0.979	0.986	0.993						
	NSGA-III	0.964	0.979					1.000			
2	B-E	0.815	0.899	0.921	0.921	0.961	0.976	0.969	0.985	0.985	0.985
	NSGA-II	0.866	0.921								
	NSGA-III	0.866									
3	B-E	0.796	0.888	0.912	0.920	0.960	0.976	0.969	0.985	0.985	0.985
	NSGA-II	0.911	0.943	0.958	0.975		0.976	0.975			
	NSGA-III	0.876	0.927	0.951	0.975	0.976					
4	B-E	0.832	0.940	0.948	0.977	0.976	0.985	0.977	0.986	0.986	0.986
	NSGA-II	0.914	0.946	0.955	0.977	0.992					
	NSGA-III	0.897	0.955	0.963			1.000				
5	B-E	0.972	0.978	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	NSGA-II	0.974	0.995	1.000							
	NSGA-III	0.970	0.995								
6	B-E	0.975	1.000	0.975	1.000	1.000	0.975	1.000	1.000	1.000	1.000
	NSGA-II	1.000	1.000								
	NSGA-III	1.000	1.000								
7	B-E	0.962	0.962	0.963	0.992	0.992	0.992	0.992	0.992	0.992	0.992
	NSGA-II	0.962	0.972	0.982	1.000						
	NSGA-III	0.962	0.972		1.000						
8	B-E	0.884	0.884	0.877	0.877	0.874	0.877	0.865	0.884	0.874	0.890
	NSGA-II	0.884	0.890	0.890	0.890						
	NSGA-III	0.884	0.884								
9	B-E	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	NSGA-II	1.000									
	NSGA-III	1.000									
10	B-E	0.993	0.993	0.993	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	NSGA-II	0.993	1.000								
	NSGA-III	0.993	1.000								
11	B-E	0.996	0.996	0.996	0.992	0.996	0.992	0.992	0.992	0.992	0.996
	NSGA-II	0.996	0.996								
	NSGA-III	0.996	0.996								
12	B-E	0.899	0.892	0.918	0.911	0.921	0.925	0.925	0.929	0.922	0.925
	NSGA-II	0.899	0.908	0.919	0.928	0.932	0.941				
	NSGA-III	0.899	0.912				0.942				

were clearly able to handle the complete process and the classifiers most used to obtain the highest accuracy are presented in Fig. 4.5. The results show the percentage of use of each classifier for each patient. For example, in the case of

Table 4.2: Accuracy obtained using EMD for feature extraction with NSGA-II and NSGA-III for EEG channel selection (subjects 13-24).

Id	Method	No. channels									
		1	2	3	4	5	6	7	8	9	10
13	B-E	0.775	0.777	0.775	0.806	0.788	0.726	0.749	0.782	0.782	0.733
	NSGA-II	0.775	0.777	0.798	0.806			0.813			
	NSGA-III	0.775	0.777			0.813					
14	B-E	0.925	0.933	0.942	0.942	0.942	0.967	0.967	0.983	0.983	0.983
	NSGA-II	0.933	0.967	0.983	0.983						
	NSGA-III	0.933	0.942	0.983							
15	B-E	0.971	0.969	0.978	0.981	0.985	0.986	0.986	0.988	0.988	0.988
	NSGA-II	0.981	0.981	0.988	0.988						
	NSGA-III	0.981	0.985	0.988							
16	B-E	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.800
	NSGA-II	0.900	0.900								
	NSGA-III	0.900	0.900								
17	B-E	0.940	0.980	0.980	0.990	1.000	1.000	1.000	1.000	1.000	1.000
	NSGA-II	0.980	0.990	1.000							
	NSGA-III	0.980		1.000							
18	B-E	0.790	0.852	0.832	0.862	0.853	0.882	0.892	0.910	0.900	0.900
	NSGA-II	0.803	0.852	0.870	0.900		0.910	0.920			
	NSGA-III	0.783	0.852	0.862	0.880	0.890	0.892				
19	B-E	0.913	0.908	0.925	0.925	0.950	0.963	0.975	0.975	0.988	0.988
	NSGA-II	0.921	0.946	0.950	0.963	0.975	0.988	1.000			
	NSGA-III	0.913	0.975				1.000				
20	B-E	0.948	0.970	0.957	0.957	0.970	0.980	0.990	0.990	0.968	0.980
	NSGA-II	0.980		0.990							
	NSGA-III	0.980		0.990							
21	B-E	0.879	0.933	0.888	0.888	0.908	0.938	0.904	0.942	0.933	0.908
	NSGA-II	0.888	0.950	0.954	0.967	0.970	0.983				
	NSGA-III	0.888	0.942	0.954	0.983						
22	B-E	0.971	0.971	0.983	0.983	0.983	0.983	0.983	0.983	0.983	0.983
	NSGA-II	0.983		0.983							
	NSGA-III	0.983									
23	B-E	0.938	0.940	0.938	0.955	0.962	0.955	0.962	0.962	0.962	0.962
	NSGA-II	0.938	0.948	0.962							
	NSGA-III	0.938	0.946				0.970				
24	B-E	0.975	0.975	0.992	0.992	0.992	0.992	0.992	0.992	0.992	0.992
	NSGA-II	0.975	0.992	0.992	1.000						
	NSGA-III	0.992			1.000						

NSGA-II for patient 1, the most highly used classifier was *RF*, which was used 54.59% of the time, then *SVM* with 33.72%, *KNN* with 7.35%, and *NB* with 4.34%.

SVM and *RF* were the most highly used classifiers to obtain the highest accuracy

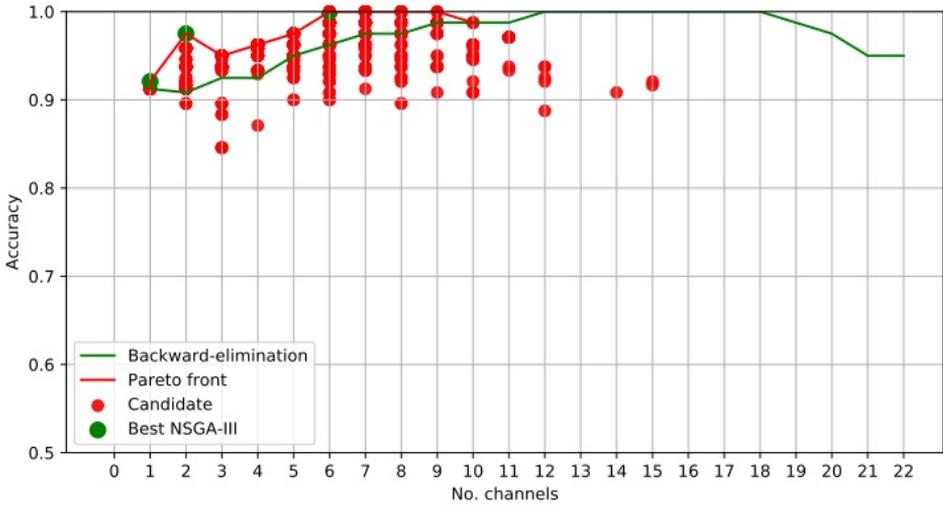


Figure 4.4: EEG Channel selection for epileptic-seizure classification of patient 19 using EMD-based features. Comparison between NSGA-III and the backward-elimination algorithm.

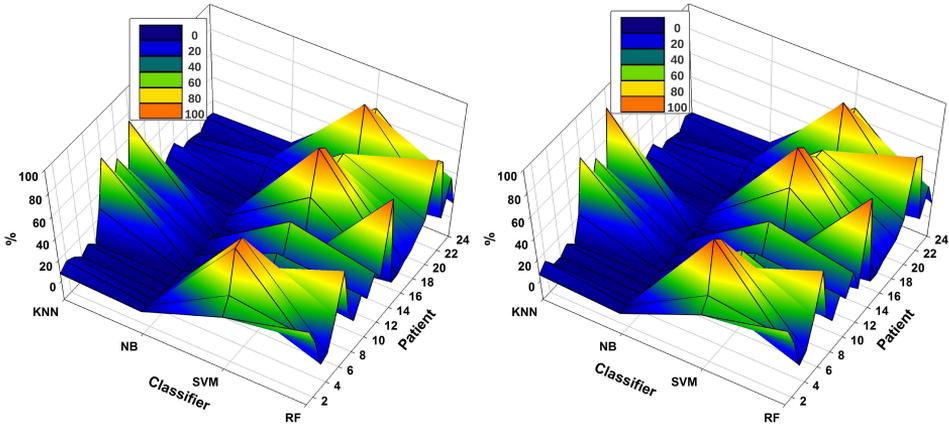


Figure 4.5: Comparison of the most used classifiers by NSGA-II (left) and NSGA-III (right) for the 24 patients using EMD-based feature extraction.

in all iterations of NSGA-II and NSGA-III (see Fig. 4.5). On the other hand, *NB* was used in all iterations but only returned the highest accuracy a few times. In general, *RF* was used $32.8\% \pm 24.2$ of the time for all patients, *SVM* $47.0\% \pm 27.9$, *NB* $3.1\% \pm 4.2$, and *KNN* $17.1\% \pm 20.5$. For NSGA-III, the *RF* classifier was used $32.0\% \pm 25.1$ of the

time, *SVM* 48.8%±28.6, *NB* 2.8%±3.6, and *KNN* 16.4%±21.7.

The analysis of the most highly used classifier in all generations and each chromosome is important because it allows discarding the use of some to decrease the computational cost and also because it shows that the classifier necessary to obtain the highest accuracy may differ, depending on the patient and the EEG channel subsets used.

4.5 Channel selection for Epileptic-seizure classification with DWT-based features

The experiment was repeated but now using DWT to extract the sub-bands and then compute the four features per sub-band, as described above. The experiments were repeated using NSGA-II and NSGA-III for the 24 patients. Additionally the accuracies obtained were also compared to those obtained using the backward-elimination algorithm. The results are summarized in Tables 4.3 and 4.4. Supplementary material in [200] provides the accuracy, specificity, and sensitivity for the first four EEG channels.

The results in Tables 4.3 and 4.4 show that an average of 36±7 generations was required for NSGA-II and 41±11 for NSGA-III. ***In general, the use of DWT for feature extraction resulted in more rapid EEG channel selection and better accuracy.***

In the case of patient 13, the use of DWT instead of EMD considerably improved epileptic-seizure classification, i.e., an improvement from 0.775 to 0.820 using one EEG channel and from 0.777 to 0.849 using two. In general, both methods showed high accuracy when the the EEG channels were selected using NSGA-based methods. The most-used classifiers when DWT was used for feature extraction were *SVM* and *KNN* for both NSGA-II and NSGA-III, as shown in a mesh plot of the most-used classifier for each patient (see Fig. 4.6). Specifically, for NSGA-II, *RF* was used an average of 20.5%±16.5 of the time for all patients, *SVM* 46.1%±23.5, *NB* 3.6%±3.8, and *KNN* 29.8%±23.1. When selecting the EEG channels using NSGA-III, the *RF* classifier was used an average of 22.1%±19.0 of the time, *SVM* 47.3%±24.5, *NB* 1.0%±1.4, and *KNN* 29.5%±23.3.

SVM was the most highly-used classifier in general, but *RF* and *KNN* were also highly used (see Fig. 4.6). These data also show that *KNN* was more highly used with DWT-based features than with EMD-based features (see Fig. 4.5). *NB*

Table 4.3: Accuracy obtained using DWT for feature extraction with NSGA-II and NSGA-III for EEG channel selection (subjects 1-12).

Id	Method	No. channels									
		1	2	3	4	5	6	7	8	9	10
1	B-E	0.950	0.993	0.993	0.993	1.000	0.993	0.993	0.993	1.000	1.000
	NSGA-II	0.986	1.000								
	NSGA-III	0.986		1.000							
2	B-E	0.983	0.992	0.992	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	NSGA-II	0.992	0.992	1.000							
	NSGA-III	0.992	0.992		1.000						
3	B-E	0.983	0.985	0.992	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	NSGA-II	0.983	0.992	1.000							
	NSGA-III	0.983		1.000							
4	B-E	0.952	0.966	0.975	0.983	0.976	0.983	0.983	0.983	0.976	0.983
	NSGA-II	1.00									
	NSGA-III	1.00									
5	B-E	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	NSGA-II	1.000									
	NSGA-III	1.000									
6	B-E	0.975	0.950	0.950	0.950	0.950	0.950	0.950	0.950	0.900	1.000
	NSGA-II	0.975	0.975	0.975							
	NSGA-III	0.975	0.975				1.000				
7	B-E	0.962	0.972	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.980
	NSGA-II	0.980	0.982	1.000							
	NSGA-III	0.980		1.000							
8	B-E	0.914	0.903	0.917	0.904	0.894	0.884	0.894	0.890	0.890	0.894
	NSGA-II	0.917	0.917								
	NSGA-III	0.971		0.917							
9	B-E	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	NSGA-II	1.000	1.000								
	NSGA-III	1.000									
10	B-E	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	NSGA-II	1.000									
	NSGA-III	1.000	1.000								
11	B-E	1.000	1.000	1.000	1.000	0.996	0.996	0.996	1.000	0.996	1.000
	NSGA-II	1.000									
	NSGA-III	1.000									
12	B-E	0.899	0.932	0.942	0.942	0.949	0.935	0.942	0.945	0.952	0.945
	NSGA-II	0.911	0.948	0.948	0.952						
	NSGA-III	0.911						0.952			

was the classifier with the lowest percentage of use for both approaches.

Table 4.4: Accuracy obtained using DWT for feature extraction with NSGA-II and NSGA-III for EEG channel selection (subjects 13-24).

Id	Method	No. channels									
		1	2	3	4	5	6	7	8	9	10
13	B-E	0.822	0.827	0.793	0.827	0.795	0.798	0.776	0.798	0.776	0.827
	NSGA-II	0.820	0.849		0.855		0.864				
	NSGA-III	0.820			0.850						
14	B-E	0.950	0.967	0.983	0.983	0.983	1.000	1.000	1.000	1.000	1.000
	NSGA-II	0.967	0.983	0.995							
	NSGA-III	0.967	0.983		1.000						
15	B-E	0.978	0.985	0.981	0.986	0.986	0.988	0.994	0.995	0.998	0.997
	NSGA-II	0.978	0.994	1.000							
	NSGA-III	0.978	0.994	0.998		1.000					
16	B-E	0.800	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	NSGA-II	1.000									
	NSGA-III	1.000									
17	B-E	0.930	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	NSGA-II	1.000									
	NSGA-III	1.000									
18	B-E	0.862	0.862	0.912	0.922	0.922	0.922	0.940	0.952	0.932	0.952
	NSGA-II	0.890	0.913	0.950			0.952				
	NSGA-III	0.862	0.913			0.952					
19	B-E	0.987	1.000	0.987	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	NSGA-II	0.988	1.000								
	NSGA-III	0.988	1.000								
20	B-E	1.000	1.000	1.000	1.000	1.000	0.990	0.990	0.990	1.000	0.990
	NSGA-II	1.000									
	NSGA-III	1.000									
21	B-E	0.921	0.950	0.938	0.967	0.983	0.966	0.966	0.966	0.966	0.966
	NSGA-II	0.925	0.950	0.971	0.983						
	NSGA-III	0.933	0.950			0.983					
22	B-E	0.983	0.983	0.983	0.983	0.983	0.983	0.983	0.983	0.983	0.983
	NSGA-II	0.995	0.998	1.000							
	NSGA-III	0.995	0.995								
23	B-E	0.938	0.946	0.953	0.961	0.961	0.962	0.955	0.962	0.969	0.969
	NSGA-II	0.939	0.961	0.969	0.970	0.970	0.977				
	NSGA-III	0.939	0.961					0.977			
24	B-E	0.975	0.975	0.975	0.975	0.975	0.983	0.975	0.983	0.975	0.983
	NSGA-II	0.985	0.992	1.000							
	NSGA-III	0.985	0.988		1.000						

4.6 Discussion

The EEG channel selection method for epileptic-seizure classification proved to be robust. For example, the accuracy for patient 1 with DWT-based features was

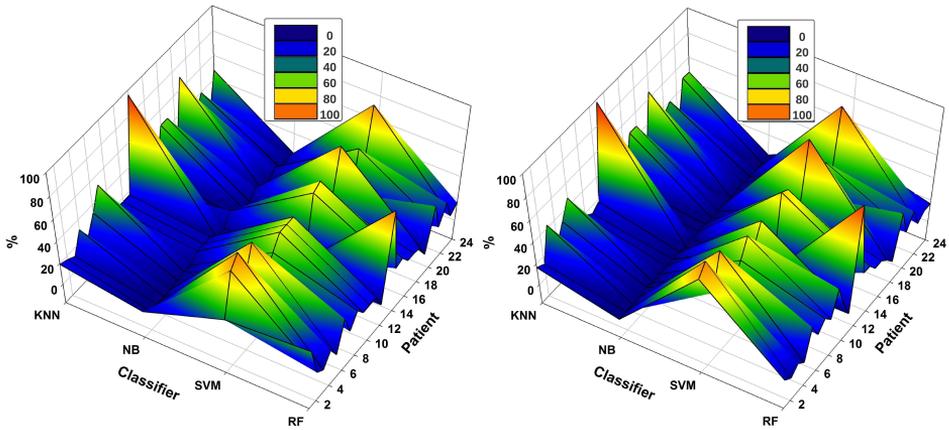


Figure 4.6: Comparison of the most-used classifiers by NSGA-II (left) and NSGA-III (right) for the 24 patients using DWT-based feature extraction.

0.97 using all EEG channels. The accuracy was even higher when using the EEG channels selected by NSGA-II or NSGA-III (1 or 2 channels): 0.98 for EMD and 1.00 for DWT.

For example, the results obtained with the data of patient 12 showed the highest accuracy using EMD to be 0.942 using six EEG channels selected by NSGA-III. ***The highest accuracy obtained using DWT-based features was 0.952 using four EEG channels.*** An important feature of the classification of the epileptic seizures of this patient is that most of the highest accuracy values were obtained using the KNN classifier (see Figs. 4.5 and 4.6), i.e., an average of 73% and 84% using EMD-based features and an average of 96% and 98% using DWT-based features, for NSGA-II and NSGA-III, respectively.

Examination of the number of epileptic seizures described in the database [215] showed this patient to have had 38 epileptic seizures and after segmentation (six-second segments), 234 instances of epileptic seizures and 234 seizure-free periods were obtained. This amount of data was one of the highest of the patients used for this study. However for patient 15, for whom there was a similar amount of data, the highest accuracy values were obtained using SVM. Thus, it is not possible to argue that this is due to the amount of data. Therefore, future work will also analyze more parameters related to the classifier (i.e., number of neighbors for KNN and kernel, as well as kernel parameters for SVM) and how accuracy is

affected by the number of seizure periods/trials and then, a possible relationship between the feature extraction method, the classifier and classifier's parameters, and more factors (sample rate, wet or dry electrodes, EEG device, etc.) that can affect a solid conclusion will be determined.

As shown in Figs. 4.5 and 4.6, *SVM* was generally the most highly-used classifier but *KNN* was also highly used, independently of the feature extraction method and whether *NSGA-II* or *NSGA-III* was used for channel selection. These data also show that *KNN* was more highly used with *DWT*-based features than *EMD*-based features. *NB* was the classifier with the lowest percentage of use for both approaches. For future steps, these findings will be considered and used for testing other important parameters related to each classifier to reduce the computation cost, instead of testing *NB* again.

In general, the results presented in this Section show that this approach is able to classify epileptic seizure and seizure-free periods with an average accuracy of up to 0.97 ± 0.05 using only one EEG electrode. This result was obtained using *DWT*-based features. The use of two or more channels can increase the accuracy to 0.98 and 0.99, especially when the EEG channels are selected by *NSGA-III* (see Table 4.5).

In the state-of-the-art, there are several relevant studies in which the authors present various methods for feature extraction and classification using the same dataset under different experiment setups. Table 4.5 presents a general overview of such studies for analysis and comparison.

Table 4.5 shows the state-of-the-art and classification accuracy of approaches using *EMD*-based or *DWT*-based features, as well as *NSGA-II* or *NSGA-III*. It should be noted that the results are not directly comparable to those from previous studies as a lower number of EEG channels were used, found by *NSGA*-based algorithms, and the experiments were based on 24 subjects and used different experimental setups. It should be noted that the average values presented in the results were obtained from Tables 4.1, 4.2, 4.3, and 4.4, which correspond to the results obtained in the Pareto-front for each subject in the dataset. In addition, the average accuracy was affected for some subjects when using two or three channels, for whom the highest accuracy values were not obtained with this number of EEG channels (see Tables 4.1, 4.2, 4.3, and 4.4), i.e., using *EMD*-based features, the

Table 4.5: Comparison of relevant existing methods for epileptic-seizure classification using the [CHB-MIT Scalp EEG dataset](#) presented in [218].

Ref.	Method	Subjects, channels	Evaluation
[256]	Energy and coefficient of variation extracted from DWT, interquartile range, median absolute deviation from raw signal.	23, 23	accuracy of 0.80 using 80% for training.
[242]	Relative values of energy and normalized coefficients of variation from DWT.	5, (23, 24 or 26)	accuracy of 0.91 using 80% for training.
[243]	Seven features from the intersection sequence of Poincaré section with phase space.	23, 23	accuracy values of 0.93 and 0.94 using 25% and 50% for training, respectively.
[245]	Three features extracted from different oscillatory levels using multivariate extension of EWT. The channel with the lowest standard deviation was selected and the four channels with higher mutual information then added.	23, 5	accuracy of 0.99 using 10-fold cross-validation.
[244]	Signal curve length of the time-domain EEG signal and the mode powers of the dynamic mode decomposition.	12, 18	sensitivity of 0.87 using 50% for training.
[135]	<i>Teager</i> and <i>instantaneous</i> energy, <i>Higuchi</i> and <i>Petrosian</i> fractal dimension, and DFA from 2 IMFs based on the EMD. Channels selected using the backward-elimination algorithm.	24, 5	average accuracy of 0.93 using 10-fold cross-validation.
Proposed method using EMD-based features	<i>Teager</i> and <i>instantaneous</i> energy, and <i>Higuchi</i> and <i>Petrosian</i> fractal dimension from 2 IMFs based on EMD.	24, 1-3	average accuracy values of 0.93±0.06 , 0.95±0.06 , and 0.95±0.05 using 10-fold cross-validation for 1, 2, 3, and 4 channels selected by NSGA-II.
		24, 1-3 channels	average accuracy values of 0.93±0.06 , 0.94±0.06 , and 0.96±0.04 using 10-fold cross-validation for 1, 2, and 3 channels selected by NSGA-III.
Proposed method using DWT-based features	<i>Teager</i> and <i>instantaneous</i> energy and <i>Higuchi</i> and <i>Petrosian</i> fractal dimension from 4 decomposition levels of the DWT.	24, 1-3	average accuracy values of 0.97±0.05 , 0.97±0.04 , and 0.98±0.02 using 10-fold cross-validation for 1, 2 and 3, channels selected by NSGA-II.
		24, 1-3	average accuracy values of 0.97±0.05 , 0.98±0.03 , and 0.99±0.01 using 10-fold cross-validation for 1, 2, and 3 channels selected by NSGA-III.

accuracy for the Pareto-front for NSGA-III was 0.992 with one channel, and 1.00 using four EEG channels, but there was no information for the combination with

two or three channels for obtaining the accuracy in the Pareto-front.

Table 4.6: Comparison of several relevant existing methods for epileptic-seizure classification using different datasets.

Ref.	Method	Subjects, channels	Evaluation
[257]	Features based on approximate entropy and classification using Elman and probabilistic neural networks.	5, 1	accuracy of 1.000.
[258]	Five levels of decomposition by DWT and features using PCA, independent component analysis (ICA), and LDA. The classification used SVM.	5, 1	accuracy values of 0.987, 0.995, and 1.000 using features based on PCA, ICA, and LDA, respectively.
[247]	Entropy-Fuzzy Classifier with three classes, normal vs. pre-ictal vs. epileptic.	5, 1	accuracy of 0.981.
[248]	Features based on two-dimensional (2D) and 3D phase space representation (PSRs) of IMFs from EMD, and least-square SVM (<i>LS-SVM</i>) classifier.	5, 1	accuracy of 0.986.
[246]	Using the TUH EEG corpus, they used 10-second segments with a sample rate of 250 Hz and computed 24 features per channel. Six different classifiers were compared: SVM, NB, KNN, RF, gradient boosting, and logistic regression.	43, 22	accuracy of 0.994 using SVM.
[249]	Features based on Fourier-Bessel series expansion and classified using <i>LS-SVM</i>	5, 1	accuracy of 0.990 in the best case.
[252]	Third-order cumulant (ToC) and neural network with softmax classifier.	5, 1	accuracy of 1.000.
[251]	Energy features from sub-bands extracted using the Taylor-Fourier filter bank and <i>LS-SVM</i> .	5, 1	accuracy of 0.948.
[185]	Wavelet coefficients from sub-bands obtained using DWT with 7 levels of decomposition using iEEG from 10 patients of the Flint Hills Scientific dataset.	10, 3	sensitivity of 0.96.

It is important to mention that in the work presented in [246–249, 251, 252, 257, 258], no methods of channel selection were used, as the dataset used consisted of only one or two EEG channels and the study [185] used methods based on variance or entropy to select the channels before the classification process.

Most of the studies presented in Table 4.6 were based on invasive EEG, which

provides better signal quality [253]. Therefore, their performance should be re-tested on non-invasive EEG signals for continuous monitoring. **Note that in the presented work, the SVM classifier was the most widely used and provided the highest accuracy values relative to the other classifiers and neural networks, consistent with the results obtained in this thesis.**

According to the results in this thesis, NSGA-III is able to find the most relevant EEG channel combinations using DWT-based features to achieve an average accuracy of up to 0.99 using only three channels. Looking towards improving the general performance of this approach and testing it using additional public epileptic-seizure datasets, new experiments will be performed considering more than two objective functions in the problem and verify whether NSGA-III is still the best method for solving this problem [212, 213].

Results have shown that the best accuracy can be reached using one to three channels for certain subjects and four or more for others. Thus, testing different methods in an attempt to improve the channel-selection process and decrease the complexity is proposed for future studies. This can be achieved by testing and comparing methods such as that presented by [245], which selects a channel with the lowest SD and then four channels with the highest MI with the previously chosen channel, as well as other optimization approaches [87, 138, 190–201].

Epileptic-seizure classification using EEG signals is important for evaluating the state of the brain. Following the evolution of the signals through continuous monitoring will enable prediction with a low number of EEG channels, making it easier to use and thus allowing long-term monitoring using a possibly personalized portable EEG device [259, 260]. However, there are several challenges that need to be addressed before implementation in real life.

Because epilepsy can cause a variety of other neurological disorders (i.e., depression, anxiety, etc.) such confounders should be additionally studied to better distinguish between an epileptic seizure and seizure-free periods. Thus, future efforts will also include the study of epilepsy-related disorders and how they can be recognized on EEG signals. A possible portable low-density EEG device will facilitate monitoring in daily life, which will allow healthcare professionals more confident management of seizures, not only in the hospital or laboratory but also in conjunction with the recent progress in telehealth and telemedicine

[261–264].

From the results presented in this Chapter, it is clear that EMD-based or DWT-based features can be useful for epileptic-seizure classification. Using these approaches, a possible subject-tailored method can consider the addition of another gene in the chromosome for the optimization process and thus select the most useful method for detecting epileptic seizures for that subject. This will be tested in future studies based on the findings here, as well as different chromosome representations for solving all possible problems related to parameter optimization at the same time.

The computational complexity of the method used for channel selection is $O(MN^2)$. However, the study of the most relevant channels is important and it must be performed for analysis and, as presented here, to verify whether epileptic seizures can be detected using a few non-invasive EEG channels. The limitations of the methods used for feature extraction are related to the well-known problems of EMD, such as the selection of the best spline, the end effect, and the mode mixing problem [116, 126, 128].

For DWT, the main problems are related to parameter selection, such as the number of levels of decomposition and the mother function. Some of these limitations have already been considered in the literature or can be solved by using recent progress in code optimization [227, 228, 265]. Future efforts for classification will focus on testing and comparing shallow convolutional neural networks and Riemannian classifiers, as they have been shown to provide high accuracy values for EEG-signal classification [148, 266, 267].

Future efforts will concentrate on testing the methods used for epileptic-seizure classification, the epileptic seizure prediction problem, testing methods for feature extraction and classification, and testing whether the methods for channel selection can find the most relevant subsets for this task and seizure onset detection [171, 175, 184, 185].

Chapter 5

Case study 2: Channel count optimization for EEG-based biometric systems

This Chapter presents two approaches for creating EEG-based biometric systems using various methods for channel selection and implementing them for feature extraction and classification. This is tested in experiments using multi-class classification, as well as one-class classification

This Chapter is based on the journal articles [87, 138, 223] and addresses the 1st, 2nd, and 3rd Research Questions.

5.1 Introduction

Security systems are used by organizations to protect places or information for which privileges are needed or require access authorization, as well as to deny unauthorized access to facilities, equipment, or resources and protect against espionage, theft, or even terrorist attacks. Various safety measures have long been proposed, ranging from the use of generic systems (security guards, closed-circuit television, smart cards, proximity readers, and RFID) to that of biometric identifiers (fingerprints, palmprints, retinal scans, etc.) [268, 269].

Biometric recognition refers to the automatic recognition of individuals based on their physiological and/or behavioral features [268]. A biometric system is a pattern recognition system that operates by acquiring biometric data from subjects, extracting a set of features, and comparing this set of features against a template

set in the database. Biometric systems have advantages over generic systems, as it is more difficult to steal, compromise, or duplicate the key. However, biometric systems are vulnerable to a variety of attacks aimed at undermining the integrity of the authentication process [269]. For example, an intruder may fraudulently obtain the latent fingerprints of an user and later used them to construct a digital or physical artifact of the user's finger [270]. This is possible because authentication systems cannot discriminate between an intruder who fraudulently obtains access privileges and authorized users.

Due to the increasing threat of bypassing the authentication and authorization process of current traditional/biometric security systems [269], there is a growing interest in exploring new biometric measures. In this context, the use of brain signals to create biometric markers using various neuro-paradigms has emerged as a robust alternative to the above-mentioned vulnerabilities.

Brain signals can be used as a basis for the design of biometric markers, as any human physiological and/or behavioral characteristic can be used as a biometric feature, as long as it satisfies the following requirements: *universality, permanence, collectability, performance, acceptability, and circumvention* [268]. Brain signals are highly reliable and secure because biometric markers obtained from EEG-recordings of human brain activity are almost impossible to duplicate, as the brain is highly individual [271].

An authentication system may include a stage in which the data is used in a multi-class model with all the subjects in the dataset to identify a specific subject. It may also include a verification step to compare the data from the claimed subject with that of the true subject, alone in the dataset, to detect whether the subject is an intruder or not. The order of these stages may differ depending on the approach. The number of EEG-based biometric systems has been steadily growing using various approaches to solve problems related to the authentication and verification stages.

A research-grade EEG device guarantees a controlled environment and high-quality multi-channel EEG recording, but this is offset by the high computational cost, non-portability of the equipment, and use of inconvenient conductive gels. The development of dry EEG sensors has created new possibilities for the development of new types of portable EEG systems. An important step towards

this goal is a reduction in the number of required EEG channels while increasing, or at least maintaining, the same performance as high-density EEG.

5.2 State-of-the-art

Depending on whether the paradigm is task-dependent or task-independent, certain EEG channels provide only redundant or sub-optimal information. Several techniques have been studied with the aim of developing low-density EEG-based systems with high performance, i.e., pre-processing and feature extraction, channel selection, and paradigms to stimulate brain signals. For EEG-based biometric systems, several approaches have been presented using various paradigms to stimulate and record the EEG signals, i.e., imagined speech [222, 223, 272], resting-state [85, 173, 273–277], and ERPs [138, 206].

In general, resting-state potentials and ERPs have been shown to be good candidates for a new biometric system for which there are several different state-of-the-art approaches [206, 273, 276–278], with the localization of the relevant channels differing, depending on the paradigm.

An important element is dimensionality reduction, which can be tackled through channel selection and feature extraction. Several approaches can be used to accomplish this task, including those based on methods such as PCA, DWT, EMD, and even approaches using raw data as input for different configurations of neural networks (NN) [138, 206, 222, 223, 279–283].

Several approaches have been proposed for the creation of biometric systems following various experiment configurations with various paradigms and methods for feature extraction and classification using the EEGMMIDB dataset (see Section 3.6.2), using various configurations of neural networks [280, 284–286], other supervised and unsupervised techniques [274, 278, 287–296], and methods for EEG channel selection [201, 275, 297].

One approach used a subset of eight pre-selected channels [297] and EEG data from a task for training and then that from another task for testing. The selection of the channels was justified based on their stability across various mental tasks, and the results presented were evaluated using the half total error rate (HTER), which was 14.69%. Another approach used various tasks from the EEGMMIDB and channel selection, using the binary flower pollination algorithm (BFPA), and reported accuracy values of up to 0.87 using supervised learning and

approximately 32 EEG channels [201]. However, the analysis considered only non-intruders when using multi-class classification, and therefore the addition of more stages for detecting the intruders is necessary.

Other approaches use instances of different length with the same dataset, such as instances of 10 or 12 seconds [274, 290]. Resting-state instances of 10 seconds have been validated with the leave one-out framework, consisting of five instances of 10 seconds for training and one instance for validating the model [290], resulting in a correct recognition rate (CRR) of 0.997 for the resting-state with the eyes-open and 0.986 with the eyes-closed, all using 64 EEG channels.

An approach with one-second EEG signals from the FP1 and FP2 channels and a 256-Hz sample rate during the resting state has been proposed for a biometric system, extracting features directly from the raw data and using Fisher's discriminant analysis [276], obtaining a TAR of up to 0.966 and a false acceptance rate (FAR) of 0.034. Another approach used two-second EEG signals from the FP1 and FP2 channels, with a 2048-Hz sample rate, and the authors used a set of classifiers to perform multi-class classification [273]. They obtained an accuracy of 0.93 and a false positive identification rate of 0.165. Another approach presented the results of a study using the Cz EEG channel, which was manually selected, on 20 subjects during the resting-state [277], obtaining a TAR of 1.0 and TRR of over 0.8. None of these studies attempted to systematically select the minimal number of optimal channels to perform the task.

Deep-learning algorithms have shown success in image processing and other fields but have not shown convincing and consistent improvement over the most advanced current methods for EEG data [148, 282]. However, several new approaches have been recently presented that show high accuracy. For example, an approach using convolutional neural network (CNN) gated recurrent units (CNN-GRU) was presented in [281], and the authors evaluated the proposed method in a public dataset called DEAP, which consists of EEG signals from 32 subjects recorded from 32 channels using different emotions as a paradigm [298]. Their experiments were performed using 10-second segments of EEG signals and they reported a mean CRR of up to 0.999 with 32 channels using CNN-GRU and 0.991 with five channels that were selected using one-way repeated measures ANOVA with Bonferroni pairwise comparison (post-hoc). The findings of this

work are interesting and the accuracy values obtained high. However, deep-learning approaches require a large amount of data and the length of the signal segments and the paradigm followed are not standard. Furthermore, for a real-time application, the collection of a large number of instances and instances during long periods can be exhausting, making such an approach noncompetitive with current biometric systems in the industry (i.e., fingerprints, face recognition, etc.).

The amount of data and time required for training NN are the main concerns for effective deployment and adoption of EEG-based biometric systems in real-life scenarios. In the literature, researchers have reported results using from simple NN structures (i.e., a single hidden layer) to more complex networks (recurrent and CNN), but this requires the improvement of computational power, with faster CPUs and the use of GPUs [148, 278, 281, 294–296]. The large amount of data required by deep-learning approaches can be overcome using an approach based on simple data augmentation techniques by creating overlapped time windows [284].

Other related proposals using neural networks have been presented and compared to the state-of-the-art [278, 294–296], amongst which some of the most relevant studies used approximately 100 subjects and mostly 64 channels for testing their approaches [279, 280, 284, 299]. However, there is no defined method for channel selection, since the process for selecting the most relevant channels requires repetition of the classification process several times and it is well known that deep-learning approaches are computationally costly [148, 296].

5.3 First approach using a two-stage classification process

In this approach, the P300-speller dataset described in Section 3.6.3 and a two-stage approach for the entire process, illustrated in Fig. 5.1, were used. An OCSVM model was created with the aim to train the model to recognize subjects that are already in the system and to reject those who are not (Intruders). In the first part of this experiment, the model was trained using subjects with IDs 1-13 (non-intruder) and only EEG signals from session one, using 30 instances and all EEG channels (56 channels). Then the EEG signals from all the subjects of session two were used, considering subjects 14-26 as intruders, to validate the model (see Fig. 5.1). The results were evaluated using the TAR, TRR, and accuracy of multi-class classification (see Table 5.1).

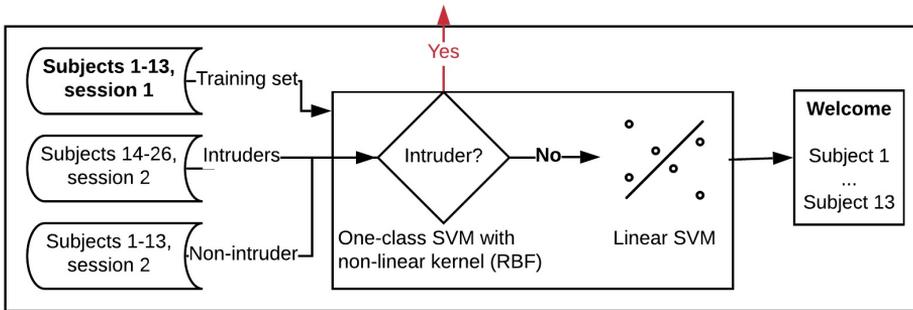


Figure 5.1: Flowchart of the first approach for intruder detection and subject identification.

Table 5.1: TAR, TRR, and accuracy for subject identification and authentication with EEG data from all channels using different ν and γ values for one-class SVM.

	Subjects	ν	γ	TAR	TRR	Accuracy
Non-intruders	1 - 13	0.01	0.01	0.923	-	0.98 ± 0.2
Intruders	14 - 26			-	0.083	
Non-intruders	1-13	0.10	0.10	0.545	-	-
Intruders	14 - 26			-	0.449	
Non-intruders	14 - 26	0.01	0.01	0.951	-	1.00 ± 0.0
Intruders	1 - 13			-	0.212	
Non-intruders	14 - 26	0.10	0.10	0.495	-	-
Intruders	1-13			-	0.551	

Table 5.1 presents an example of the results using subjects 1-13 as non-intruders and subjects 14-26 as intruders. The results show that approximately 90% of the subjects were correctly accepted but also that only approximately 8% of the intruders were correctly rejected. However, changing the ν and γ parameters for the SVM RBF changed the TAR and TRR to approximately 50% in both cases.

Given that all subjects with access (subjects 1-13) passed the first layer, a multi-class classifier was created for subject identification. An SVM with a linear kernel was defined and used because of the results obtained in previous studies and also because it was found experimentally to be the best solution. The flowchart of the

complete method is presented in Fig. 5.1. The accuracy obtained following 10-fold cross-validation was **0.98**, with a standard deviation of **0.02** (see Table 5.1).

This approach was used because the aim was to find the best configuration for the entire process. Creating a model using only the subjects with correct permission who passed the first layer would have affected the results and therefore would not have been valid.

5.3.1 Defining the problem to optimize

Once the non-intruder and intruder subsets were defined, the signals were pre-processed and the features extracted. They can be used as input for the authentication system, which can be distributed as presented in Fig. 5.1. However, the use of a more complex system is required to fit certain important parameters and select the most relevant EEG channels, which in this case was analyzed as an optimization problem.

The problem to be optimized is defined by four unconstrained objectives: 1) *Reduce the number of EEG channels*, 2) *maximize the accuracy of the multi-class classification*, 3) *maximize the number of accepted subjects with access*, and 4) *maximize the number of intruders rejected*. Each population size in each iteration is defined as 30, which was selected experimentally. The termination criterion for the optimization process is defined by the objective space tolerance, which is defined as 0.0001. This criterion is calculated every 10th generation. If optimization is not achieved, the process stops after a maximum of 500 generations.

The chromosome created to represent the search space in the scalp for this first approach is presented in Fig. 5.2, in which genes 1-56 represent the EEG channels and the *nu* parameter is calculated using genes 57-60 and the *gamma* parameter calculated using genes 61-64. When calculating the *nu* and *gamma* parameters, the binary representation is converted into a decimal value, which represents the position in a vector with the possible values for the parameter. Thus possible values were defined experimentally, which in a key-value array are {0 : 0.000001, 1 : 0.0001, 2 : 0.0005, 3 : 0.001, 4 : 0.005, 5 : 0.01, 6 : 0.1, 7 : 0.2, 8 : 0.3, 9 : 0.4, 10 : 0.5, 11 : 0.6, 12 : 0.7, 13 : 0.8, 14 : 0.9, 15 : 1.0}, for both *nu* and *gamma*. The complete process is illustrated in Fig. 5.2.

Eight features per EEG channel were extracted for all subjects and each instance following the previously explained method and that shown in the

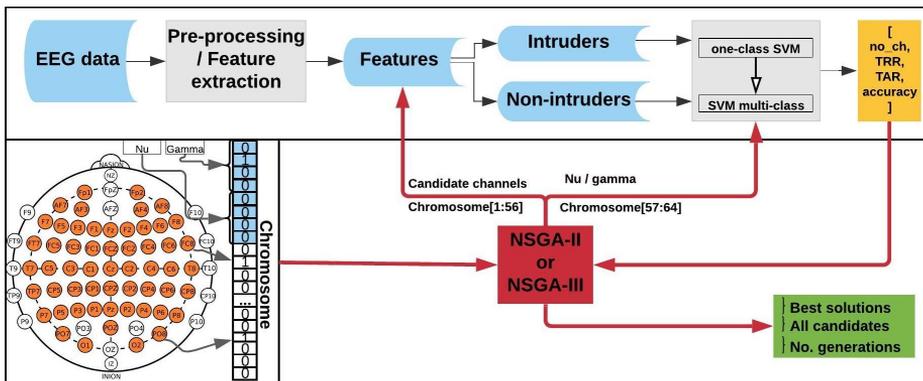


Figure 5.2: Example of the complete process for EEG channel selection using NSGA-II, including the chromosome representation using 56 genes for the EEG channels and eight for the nu and $gamma$ parameters.

flowchart presented in Fig. 3.15, in which the results are organized and stored for iterative use, as shown in Fig. 5.2. The entire process is then handled by NSGA-II or NSGA-III, which starts creating all possible candidates using a binary chromosome representation for which the corresponding subset of features for the channels is obtained, represented as 1 for genes 1-56 of the chromosome, the nu parameter calculated using genes 57-60, and the $gamma$ parameter calculated using genes 61-64.

Then, the obtained classification accuracy, number of accepted subjects with access, number of rejected subjects, and number of EEG channels used are returned to NSGA-II or NSGA-III to evaluate each chromosome in the current population. The process is repeated, creating different populations by the NSGA until the termination criterion is reached.

5.3.2 Solving the four-objective optimization problem using NSGA-II with subjects 1-13 as non-intruders and 14-26 as intruders.

This Section presents experiments that simultaneously considered all the problems to investigate whether there is a particular combination that can solve the optimization problem defined in the *Methods* Section using NSGA-II.

The experiment consisted of finding the best nu and $gamma$ for the SVM with

the RBF kernel to increase the TAR, TRR, and accuracy of subject identification or maintain them as high as possible from previous configurations, while using the lowest number of EEG channels. Briefly, NSGA-II was used for channel selection using the first 56 genes in a chromosome to represent the EEG channels and then four genes each to select the best *nu* and *gamma* parameters, obtaining thus a chromosome of 64 genes.

Several plots of the results obtained considering the four objectives are presented in Fig. 5.3 to illustrate the importance of the optimization process (see Sub-figs. 5.3a, 5.3b, 5.3c and 5.3d), as only 11.11% of the possible channel combinations resulted in a TAR and TRR between 0.9 and 1.0 (see Sub-fig. 5.3e). The classification accuracy according to the number of channels used and in relation to the Pareto-front are shown in Sub-figs. 5.3d and 5.3f.

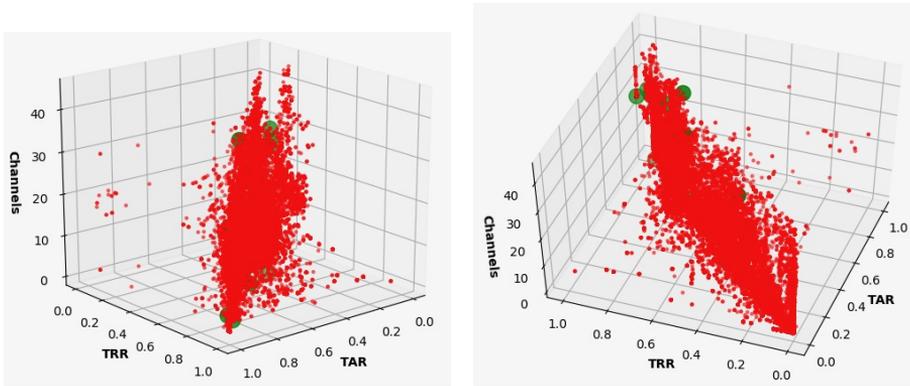
The results for the Pareto-front for all objectives are presented in Table 5.2. NSGA-II found a two-channel combination for which a TAR of 0.91, TRR of 0.88, and an accuracy of 0.78 for subject identification were obtained. NSGA-II also found a 12-channel combination for which the accuracy of subject identification was 0.93, the TAR 0.93, and the TRR 0.95. This result shows that it is possible to reduce the number of channels from 23, 24, etcetera (which gave similar accuracy values) by almost half using this approach.

5.3.3 Solving the four-objective optimization problem using NSGA-II with subjects 14-26 as non-intruders and subjects 1-13 as intruders.

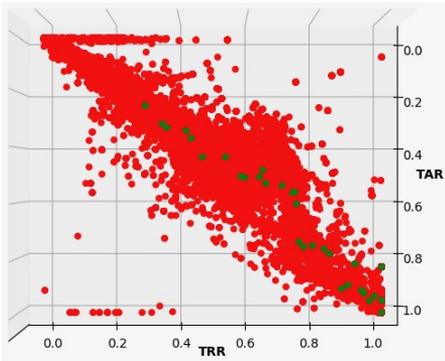
With the aim of searching for more global results, the previous experiment was repeated using the same configuration but now considering subjects 14-26 as non-intruders and subjects 1-13 as intruders. The results obtained for the four objectives are presented in Table 5.3.

As in the previous experiment, an accuracy of up to 0.83 for subject identification was obtained, with both a TAR and TRR of 1.00, using just a three-channel combination (see Table 5.3). Increasing the classification accuracy for subject identification, while maintaining the same TAR and TRR, required 16 EEG channels, in contrast to the previous experiment for which the optimal number of EEG channels was 12.

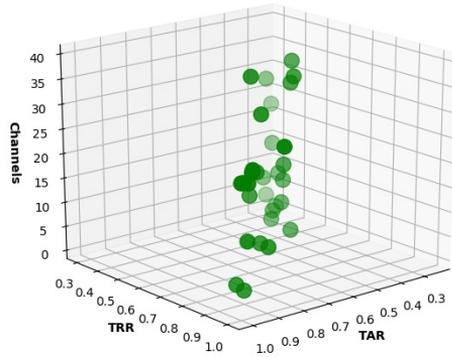
Table 5.3 presents the results obtained in the Pareto-front for the first 30 EEG



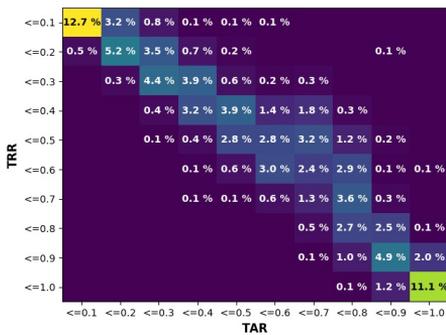
(a) First view of the candidates and the Pareto-front. (b) Second view of the candidates and the Pareto-front.



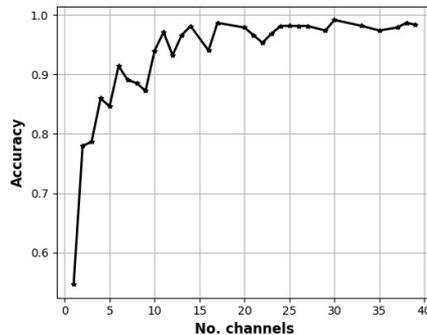
(c) Aerial view.



(d) Points in the Pareto-front.



(e) Distribution of the results obtained.



(f) Classification accuracy for the combination in the Pareto-front.

Figure 5.3: Four different views of the results obtained with NSGA-II using subjects 1-13 as non-intruders and 14-26 as intruders.

Table 5.2: TAR, TRR, and accuracy values obtained for the Pareto-front for four objectives solved with NSGA-II using subjects 1-13 as non-intruders.

No. channels	Accuracy	TAR	TRR	<i>nu</i>	<i>gamma</i>
1	0.55	0.90	0.90		
2	0.78	0.91	0.88	0.0001	0.9
3	0.79	0.34	0.42		
4	0.86	0.31	0.35		
5	0.85	0.50	0.58		
6	0.91	0.56	0.74		
7	0.89	0.51	0.60		
8	0.89	0.79	0.85	0.0010	0.9
9	0.87	0.82	0.92	0.0001	0.2
10	0.94	0.53	0.66		
11	0.97	0.43	0.47		
12	0.93	0.93	0.95	0.0001	0.9
13	0.97	0.43	0.54		
14	0.98	0.51	0.64		
16	0.94	0.76	0.77		
17	0.99	0.37	0.44		
20	0.98	0.61	0.75		
21	0.97	0.76	0.80		
22	0.95	0.25	0.30		
23	0.97	0.92	0.94		
24	0.98	0.96	0.96		
25	0.98	1.00	1.00		
26	0.98	0.94	0.98		
27	0.98	0.96	1.00		
29	0.97	0.93	0.96		
30	0.99	0.83	1.00		

channels, indicating the accuracy values obtained and the TAR and TRR, as well as the *nu* and *gamma* values used for creating the one-class classifiers to obtain the TAR and TRR results. The most relevant accuracy values, TAR, and TRR and the corresponding number of channels used are marked in gray; the *nu* and *gamma* values used to obtain these results were also added to determine whether there are similarities between these cases.

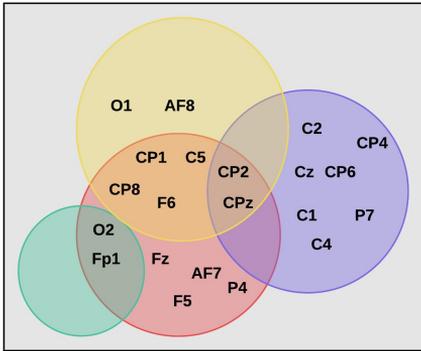
The channel combinations for this and the previous experiments were independent. Venn diagrams were generated to compare the channels used in the Pareto-front between this and the previous experiment to detect a possible pattern or a more relevant area (see Fig. 5.4). The EEG channels used to obtain the results marked in gray in Table 5.2 and the channel localization in Sub-fig. 5.4c

Table 5.3: TAR, TRR, and accuracy values obtained for the first 30 EEG channels in the Pareto-front for four objectives solved with NSGA-II using subjects 14-26 as non-intruders.

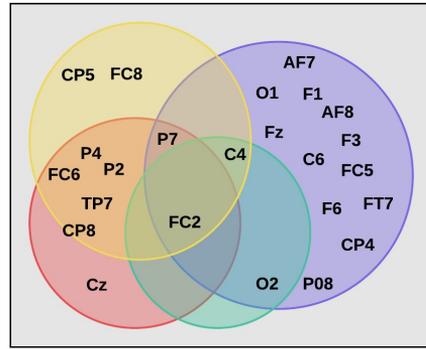
No. channels	Accuracy	TAR	TRR	nu	$gamma$
1	0.53	0.70	0.70		
2	0.62	0.31	0.31		
3	0.83	1.00	1.00	0.00001	0.6
4	0.87	0.41	0.37		
5	0.88	0.49	0.49		
6	0.96	0.81	0.73		
7	0.96	0.74	0.78		
8	0.91	0.88	0.89	0.3000	0.8
9	0.97	0.52	0.54		
10	0.97	0.90	0.91	0.0005	0.6
11	0.96	0.83	0.88		
12	0.97	0.55	0.56		
13	0.98	0.40	0.52		
14	0.98	0.80	0.84		
15	0.98	0.50	0.56		
16	1.00	1.00	1.00	0.00001	0.6
17	0.99	0.73	0.65		
18	0.98	0.93	0.93		
19	0.99	0.38	0.59		
20	0.99	0.47	0.57		
21	0.98	0.74	0.71		
22	0.99	0.99	0.99		
23	0.98	0.76	0.72		
24	1.00	0.74	0.64		
25	1.00	0.99	0.99		
26	1.00	1.00	0.99		
27	1.00	1.00	1.00		
28	1.00	0.96	0.96		
29	1.00	0.95	0.97		
30	1.00	1.00	1.00		

are presented in Sub-fig. 5.4a. The results marked in gray in Table 5.3 are shown in Sub-fig. 5.4b and EEG channel localization in Sub-fig. 5.4d.

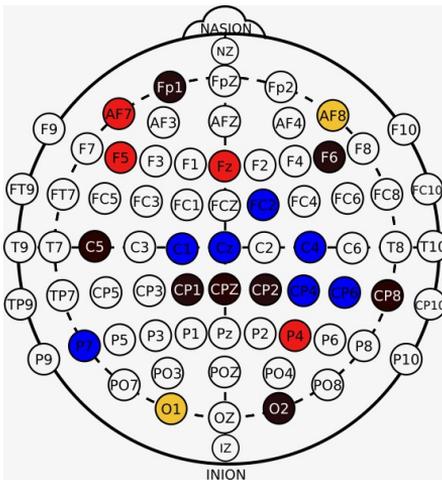
Fig. 5.4 shows certain channels within a black circle if they intersected with one or more subsets. For example, sub-fig.5.4c shows the CPZ channel in a black circle, which means that it was used in one or more subsets, as shown in sub-fig. 5.4a. It is important to highlight these channels for the discussion of the results and for the purpose of comparison with the following experiments in the thesis.



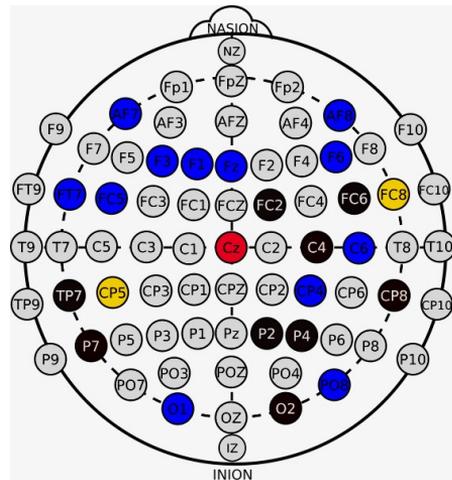
(a) Venn diagram of the subsets for 2, 8, 9, and 12 channels in the previous exp. presented in Table 5.2.



(b) Venn diagram of the subsets for 3, 8, 10, and 16 channels in the current experiment presented in Table 5.3.



(c) Channel subsets from Sub-fig. 5.4a.



(d) Channel subsets from Sub-fig. 5.4b.

Figure 5.4: Relevant EEG channel subsets in the Pareto-front for four objectives using NSGA-II, considering subjects 14-26 as intruders in the previous experiment and subjects 1-13 as intruders in the current experiment.

5.3.4 NSGA-III for solving the four-objective optimization problem.

The previous two experiments were repeated to solve the four-objective optimization problem with the same configuration, but now using NSGA-III. A comparison between the results obtained in the Pareto-front in the two

Table 5.4: TAR, TRR, and accuracy values obtained in the Pareto-front when using 7-15 EEG channels with four objectives solved with NSGA-III using subjects 1-13 as non-intruders and 14-26 as intruders and vice-versa.

S	Eval.	No. channels									
		7	8	9	10	11	12	13	14	15	
1-13	Accuracy	0.96	0.96	0.98	0.98	0.98	0.99	0.99	0.99	0.98	
	TAR	0.41	0.41	0.94	0.94	0.61	0.70	0.60	1.00	0.29	
	TRR	0.47	0.48	0.94	0.94	0.84	0.85	0.60	1.00	0.37	
	<i>nu</i>			0.0005	0.0001				0.0005		
	<i>gamma</i>			0.1	0.1				0.1		
14-26	Accuracy	0.98	0.97	0.98	0.97	0.99	0.98	1.00	1.00	0.99	
	TAR	0.95	0.93	0.90	0.93	0.95	0.94	0.93	0.94	0.72	
	TRR	0.93	0.93	0.91	0.94	0.95	0.92	0.93	0.95	0.83	
	<i>nu</i>	0.0100				0.0001			0.0001		
	<i>gamma</i>	0.7				0.9			0.9		

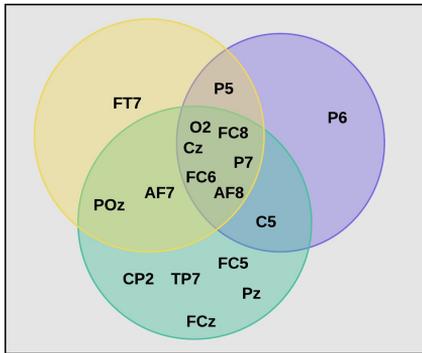
experiments, using subjects 1-13 for training (subjects 1-13 as non-intruders and 14-26 as intruders) and subjects 14-26 for training (subjects 14-26 as non-intruders and 1-13 as intruders), is shown in Table 5.4.

In this experiment, subsets with 9, 10, and 14 optimal EEG channels were found using subjects 1-13 as non-intruders and subsets with 7, 11, and 14 EEG channels using subjects 14-26 as non-intruders. As in the previous experiments, a comparison of several relevant subsets presented in Table 5.4 is presented in Fig. 5.5 for both cases, either using subjects 1-13 as non-intruders (see Sub-figs. 5.5a and 5.5c) or 14-26 as non-intruders (see Sub-figs. 5.5b and 5.5d).

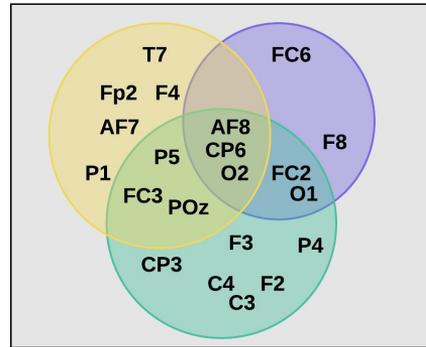
Fig. 5.5 presents a comparison between different subsets found by NSGA-III when using subjects 1-13 as non-intruders and when using them as intruders. This figure shows a lower number of channels in the interceptions, but it also shows that most of the EEG channels used for obtaining the best results presented in Table 5.4 were obtained using channels around the parietal and occipital areas, which is consistent with the paradigm used for collecting the EEG signals [300].

5.3.5 Testing the proposal in 10 random subdivisions of subjects using NSGA-II and NSGA-III.

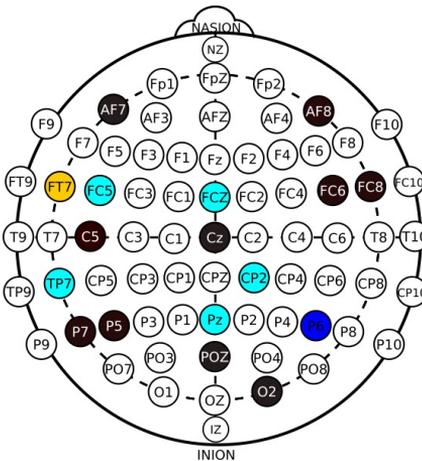
In the previous experiments, the results obtained were presented using different subsets manually selected with 50% of the subjects as non-intruders and 50% as intruders (i.e., subjects 1-13 as non-intruders and 14-26 as intruders, and



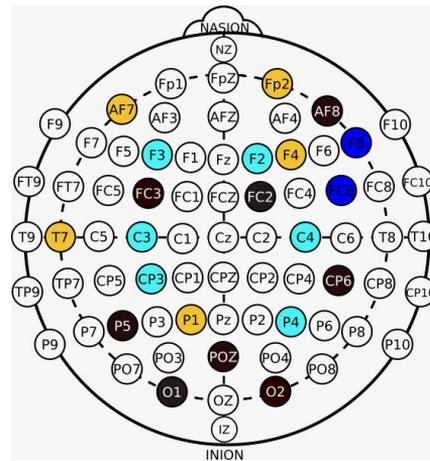
(a) Venn diagram for the subsets for 9, 10, and 14 channels using subjects 1-13 as non-intruders in the current experiment presented in Table 5.4.



(b) Venn diagram for the subsets for 7, 11, and 14 channels using subjects 14-26 as non-intruders in the current experiment presented in Table 5.4.



(c) Channel subsets from Sub-fig. 5.5a.



(d) Channel subsets from Sub-fig. 5.5b.

Figure 5.5: Relevant EEG channel subsets in the Pareto-front for four objectives using NSGA-III, considering subjects 14-26 as intruders in the previous experiment and subjects 1-13 as intruders in current experiment.

vice-versa.). The differences found when using NSGA-II or NSGA-III were also presented. However to provide a more general validation of the proposal, random subsets with 50% of the subjects as non-intruders and 50% as intruders were created and the optimization problem then solved by simultaneously considering the four objectives. This process was repeated 10 times, thus obtaining 10-fold

Table 5.5: Mean TAR, TRR, and accuracy values obtained in the Pareto-front when using 7-15 EEG channels validated in 10 random subdivisions of all the subjects, using 50% as intruders and 50% as non-intruders.

Method	Eval.	No. channels								
		7	8	9	10	11	12	13	14	15
NSGA-II	Acc.	0.96±0.02	0.96±0.01	0.97±0.02	0.98±0.02	1.00±0.00	0.99±0.01	1.00±0.00	1.00±0.00	0.99±0.01
	TAR	0.74±0.18	0.81±0.18	0.59±0.07	0.74±0.05	0.81±0.08	0.61±0.25	0.81±0.17	0.86±0.13	0.90±0.10
	TRR	0.85±0.14	0.79±0.10	0.68±0.16	0.87±0.13	0.69±0.18	0.89±0.10	0.88±0.12	0.90±0.09	0.94±0.06
NSGA-III	Acc.	0.97±0.03	0.97±0.01	0.97±0.02	0.98±0.02	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00
	TAR	0.72±0.14	0.81±0.12	0.64±0.14	0.79±0.07	0.86±0.08	0.78±0.15	0.82±0.17	0.86±0.13	0.92±0.08
	TRR	0.74±0.12	0.85±0.10	0.65±0.21	0.85±0.13	0.80±0.13	0.89±0.10	0.89±0.10	0.89±0.09	0.94±0.02

cross-validation of the proposed method. The experiment was repeated using both algorithms, NSGA-II and NSGA-III. The mean results and standard deviation are presented in Table 5.5.

The results presented in Table 5.5 show that the mean accuracy decreased in both cases when using NSGA-II or NSGA-III when considering 10 random partitions of the subjects as non-intruders or intruders. In addition, the standard deviation was $> 10\%$ in most cases when using less than 10 channels. This is because the number of channels for the best arrays, as well as the best channels, were not the same in each randomly created partition. For example, in the previous experiment presented in Table 5.4, the best results were clearly obtained using subjects 1-13 as non-intruders with nine EEG channels (i.e., an accuracy of 0.98 and a TAR of 0.94, and TRR of 0.94). However, when considering subjects 14-26 as non-intruders, the best results were obtained using seven channels (i.e., an accuracy of 0.98 and a TAR of 0.95 and TRR of 0.93).

For example, Table 5.5 shows that the accuracy values, TAR, and TRR were similar in both cases for both NSGA-II and NSGA-III when using eight EEG channels. However, the standard deviation was $> 10\%$ for the TAR and TRR, which means that the best results were not obtained with eight channels for certain subsets of subjects, i.e., sometimes with seven and sometimes with nine channels, as in the previous experiments. In summary, this new experiment shows the accuracy for subject identification to be consistently high (i.e., higher than 0.96 in all cases, as in the previous experiments presented), but the TAR and TRR can vary widely depending on which subset of subjects used as intruders or

non-intruders.

5.4 Discussion

EEG-based biometric systems have been presented as good candidate for use in authentication systems. In previous studies, various paradigms, i.e., resting-state potentials and ERPs, have been studied and compared using various types of electrodes, various numbers of channels, and varying channel localization [173, 206, 222, 223]. Several parameters are yet to be optimized. Thus, no industrial-level EEG-based biometric systems are currently available.

In the context of designing a portable EEG headset, applications for multi-task purposes and scenarios are being widely studied. NSGA-based algorithms were proposed for the optimization process, with the final objective of reducing the necessary number of EEG channels for subject identification. These algorithms depend upon several parameters that influence the performance and results. In addition, machine-learning algorithms also require the definition of several parameters, which were defined using eight genes of a created chromosome.

The new scheme introduced for subject identification and authentication shows that it can identify subjects by their EEG brain signals and distinguish between subjects who were part of the training dataset from those that are intruders. Using NSGA-II in the first experiments, channel subset combinations consisting of only two EEG channels were found, with which an accuracy of 0.78, a TAR of 0.91, and a TRR of 0.88 were obtained. However, 8, 9, or 12 channels were required to increase the value of the results for the objectives when they were simultaneously applied. NSGA-III found subsets with 7, 9, 10, or 11 EEG channels with an accuracy of up to 0.99 and both a TAR and TRR of 1.00.

Initially, the aim was to create a new fixed headset with a limited number of EEG channels, but as the results of this work show, it is not possible to argue that a certain “good” subset works better than others, as various factors are critical when choosing whether it is better to use a lower number of EEG channels or propose improvements at the classification stage. The proposed method shows that different channel subsets can provide high accuracy, TAR, and TRR values. However, deeper analysis and further experiments are required on a larger population.

P300 from ERPs have shown to be good candidates but they are not the gold

standard for this application, as there is not yet sufficient research evidence to support it. They were proposed in this work as candidates as it was shown that they exhibit strong signatures that are unique to the subject and the process does not require any training, which will be essential in a real-life application. In a real-life scenario, the biometric system can display something on a screen (an image, a weak flashlight beam aimed directly at the eyes, etc.), record the brain activity corresponding to the response to the presentation, and use it for the identification and authentication process.

The internal state of the subject, such as the resting state, could also be used as an alternative to obtain specific information on the subject, as previously discussed [173]. The EEG channel selection process is in itself informative because it can provide information about the most relevant areas in the brain for a certain neural task for a certain subject or group of subjects. This can be analyzed using *a-priori* information related to the paradigm, which can limit the search space and therefore the results.

The results presented in the first experiments show that most of the common channels in the subsets providing the highest accuracy, TAR, and TRR, come from the occipital and parietal areas, but certain channels in the frontal area were also important (FC2, FC3, FC6, FC8, F6, AF7, AF8, and Fp1). ***A final conclusion about the minimum number of necessary EEG channels for subject identification, taking into account the classification accuracy, TAR, and TRR, cannot be proposed solely based on the results of this work, as the minimum number of necessary channels will be different depending on various factors (i.e., the number of subjects, trials, sessions, feature extraction method, channel selection approach and their parameters, etc.).*** In addition, channel localization for the subsets differed between subjects and whether NSGA-II or NSGA-III methods were used, as clearly presented in Figs. 5.4 and 5.5. When 10 random subdivisions of the subjects were tested, the mean TAR and TRR decreased and the standard deviation increased. In addition, the *nu* and *gamma* values used were different in each subdivision, but the classification accuracy was maintained, similar to that of the first experiments presented.

The complexity of the analysis can be as high as that required. In the first experiments, a model with EEG signals from session 1 was trained and the

authentication and verification process was constructed using EEG signals from session 2. However, due to the plasticity of the brain, an analysis of sessions from different days/weeks/months is also necessary before a proof of concept, as well as an analysis of how this can affect the biometric approach. Another important aspect that requires further study is the scalability; it will be necessary to verify the number of subjects that can be added to this system while maintaining similar performance to that when using a small number of subjects.

Here, a first layer using the EEG data from all the subjects to search for a method to increase the TAR and TRR was created. Future studies will focus on all these relevant aspects, involving the optimization of multiple parameters related to the feature extraction and machine-learning methods by using discrete values to represent the chromosomes and not only as a binary sequence. Another important aspect to be further investigated is the use of larger datasets with $k - fold$ validation to verify whether a possible modification to the proposed approach can allow identification of a single optimal array of EEG channels for different randomly created subdivisions of subjects while consistently fulfilling all of the defined objectives and necessary parameters by optimization as in the experiments presented and discussed in this thesis.

5.5 Second approach, using a one-stage one-class algorithm

In this Section, EEG signals from 64 channels of 109 subjects and 60 instances of one second with a sample rate of 160 Hz that were recorded *during the resting-state, in which the eyes of the subject were open*, were used, as described in Section 3.6.2. EMD- or DWT-based features were used and the results evaluated using the TAR and TRR.

To ensure 10-fold cross-validation, the experiments were performed 10 times, randomly selecting 80% of the instances for training and 20% for testing, thus ensuring that the method can be generalized and that the results can be obtained even when using another subset of instances for training and testing. The models were created using OCSVM or LOF models. It should be noted that the channels and parameters were optimized for all the subjects at the same time but a single machine-learning model was created for each subject. In general, the results presented in Table 5.6 were obtained by creating a model for each of the 109 subjects in which the model of the subject was used to recognize the subject and

Table 5.6: Average TARs and TRRs for subject detection with EEG data from 64 channels and 109 subjects using different parameters for OCSVM and LOF, with EMD- and DWT-based features.

Method	Algorithm	No. neighbors	EMD-based features		DWT-based features	
			TAR	TRR	TAR	TRR
OCSVM			0.502±0.004	0.993±0.001	0.499±0.002	0.998±0.000
LOF	ball tree	1	1.000±0.000	0.923±0.005	1.000±0.000	0.979±0.002
LOF	ball tree	10	0.926±0.002	0.963±0.007	0.968±0.0038	0.989±0.012
LOF	kd tree	1	1.000±0.000	0.989±0.005	1.000±0.000	0.998±0.001
LOF	kd tree	10	0.926±0.001	0.955±0.006	0.923±0.001	0.988±0.002
LOF	brute	1	1.000±0.000	0.926±0.004	1.000±0.000	0.979±0.004
LOF	brute	10	0.927±0.001	0.939±0.007	0.924±0.003	0.989±0.002

reject the rest of the 108 who were not part of the model.

The results obtained with OCSVM showed the lowest TAR (see Table 5.6), meaning that the models created with OCSVM did not learn from the training set and thus rejected an average of approximately 50% of the instances, explaining why the TRR was high when using OCSVM. The results obtained with LOF, using three different algorithms and one or ten neighbors, are also shown in Table 5.6 for illustrative purposes. LOF using the *k-d tree* algorithm and one neighbor resulted in the highest TAR and TRR, meaning that it was possible to identify each subject and reject almost all the rest that did not correspond to the models.

Previous results have shown that the algorithm and number of neighbors used are important for increasing the TAR and TRR. The experiments were repeated using DWT-based features considering only LOF with the *k-d tree* and 1 to 10 neighbors to provide more information about this behavior. The average results obtained using 10-fold cross-validation are presented in Fig. 5.6.

The use of a higher number of neighbors resulted in a decrease in the TAR from 1.000 to 0.923 and an increase in the TRR or its remaining higher than 0.988 (see Fig. 5.6), meaning that the models were unable to learn about the features of each subject using a higher number of neighbors. ***This is relevant, as it shows the importance of selecting not only the best feature extraction method but also the LOF algorithm and the best number of neighbors.***

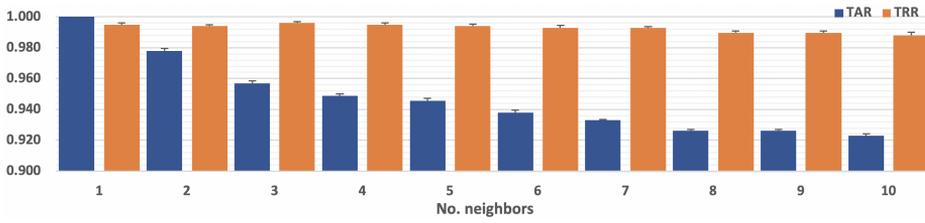


Figure 5.6: TARs and TRRs obtained using various numbers of neighbors with the LOF k - d tree algorithm and DWT-based features.

5.5.1 Defining the problem to optimize

After the pre-processing and feature extraction stages, a set of features were obtained for each EEG channel. These features can be used to create a model for each subject that can recognize it and reject the rest of the subjects. The approach is to create a model for each subject with 80% of the instances and use 20% for testing, as this dataset consists of only EEG data from one session, as described in Section 3.6.2. This requires that certain important parameters be fitted and that the most relevant EEG channels are selected.

Thus, the problem is defined as an optimization problem with three unconstrained objectives: **1) minimize the number of necessary EEG channels**, **2) maximize the TAR**, and **3) maximize the TRR**. The size of each population in each iteration is defined as 20, the termination criterion for the optimization process is defined by the objective space tolerance, which is defined as 0.0001. This criterion is calculated every 10th generation. If optimization is not achieved, the process stops after a maximum of 300 generations.

Sixty-four binary genes in a chromosome were created to represent the 64 EEG channels, as well as one gene with integer values for the algorithm (1: Ball tree, 2: k - d tree, 3: Brute force) and another with integer values for the number of neighbors (from 1 to 10, which were proposed experimentally), obtaining thus a chromosome of 66 genes. When using OCSVM in the optimization process, the same 64 genes were used for representing the EEG channels, as well as two genes with decimal values for the nu and $gamma$ parameters, similarly to the approach presented in Section 5.3. The chromosome created to represent the candidate channels in the search space and the flowchart of the complete optimization

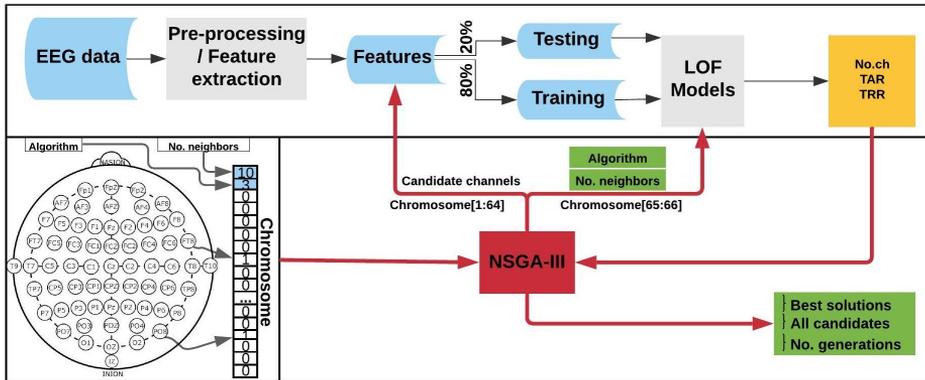


Figure 5.7: Chromosome representation and flowchart of the optimization process for EEG channel selection using NSGA-III and LOF.

process using LOF models is illustrated in Fig. 5.7.

As explained in the feature extraction method, eight features were extracted per channel when using EMD, and 16 when using DWT. The features were organized and stored for iterative use, depending on the channels marked as *1* in the chromosomes. For example, using EMD-based features, the classification process would be performed with only eight features from the channel indicated in the chromosome if the chromosome consists of only one gene. The entire process was then performed by NSGA-III, as shown in Fig. 5.7, which starts by creating 20 possible candidates for each generation.

The output for each chromosome for each generation is the number of channels used and the obtained TAR and TRR for the subset of channels in the chromosome. The results are returned to NSGA-III to evaluate each chromosome in the current population and the new generation of chromosomes is created based on the best candidates found. This process is repeated until the termination criterion or the maximum number of generations is reached.

5.5.2 Channel selection using NSGA-III and OCSVM for EEG signals for the resting-state with the eyes open

It was previously shown that the TAR and TRR of the models created using OCSVM can be improved by finding the best *nu* and *gamma* parameters [138]. The optimization process defined in the *Methods* Section was performed to provide

Table 5.7: TARs and TRRs obtained for the first five EEG channels in the Pareto-front for three objectives solved with NSGA-III using EMD- and DWT-based features with OCSVM.

No. channels	EMD-based features		DWT-based features	
	TAR	TRR	TAR	TRR
1	0.776 ± 0.138	0.851 ± 0.055	0.801 ± 0.063	0.905 ± 0.042
2	0.776 ± 0.092	0.911 ± 0.043	0.774 ± 0.066	0.958 ± 0.023
3	0.763 ± 0.150	0.969 ± 0.020	0.629 ± 0.180	0.959 ± 0.022
4	0.779 ± 0.144	0.966 ± 0.033	0.720 ± 0.069	0.980 ± 0.020
5	0.822 ± 0.028	0.969 ± 0.022	0.822 ± 0.028	0.981 ± 0.017

more information about the behavior of the OCSVM models using a larger dataset, attempting to improve the TAR and TRR while reducing the necessary number of EEG channels for subject identification.

For this experiment, EEG signals of the 109 subjects in the resting-state, with their eyes-open, were used, using 80% of the instances for training and 20% for testing. NSGA-III was used for the channel selection method using 64 binary genes in a chromosome to represent the EEG channels (1 if the channel is used, 0 if not) and two genes with decimal values (both from 0 to 1) to select the best nu and $gamma$ parameters, obtaining thus a chromosome of 66 genes.

The distribution of the results of one run obtained using EMD- and DWT-based features is shown in Fig. 5.8, as an example. The average and standard deviation of the results obtained using 10-fold cross-validation are presented in Table 5.7.

As mentioned previously, the optimization was performed 10 times for cross-validation. For certain runs, the Pareto-front contained only channel combinations with one to five channels and others with one to seven. The channels in common and other subsets can be further analyzed using these identified subsets. Thus, it may be possible to recommend a set of channels for a new possible headset (considering the best subset found and those that are the most appropriate for a new design.). However, it is first necessary to perform the analysis to choose the best paradigm or sub-task (i.e., resting-state with the eyes open or closed) for EEG data collection. For comparative purposes, the average TAR and TRR obtained using channel combinations of one to five channels in the Pareto-front of the 10

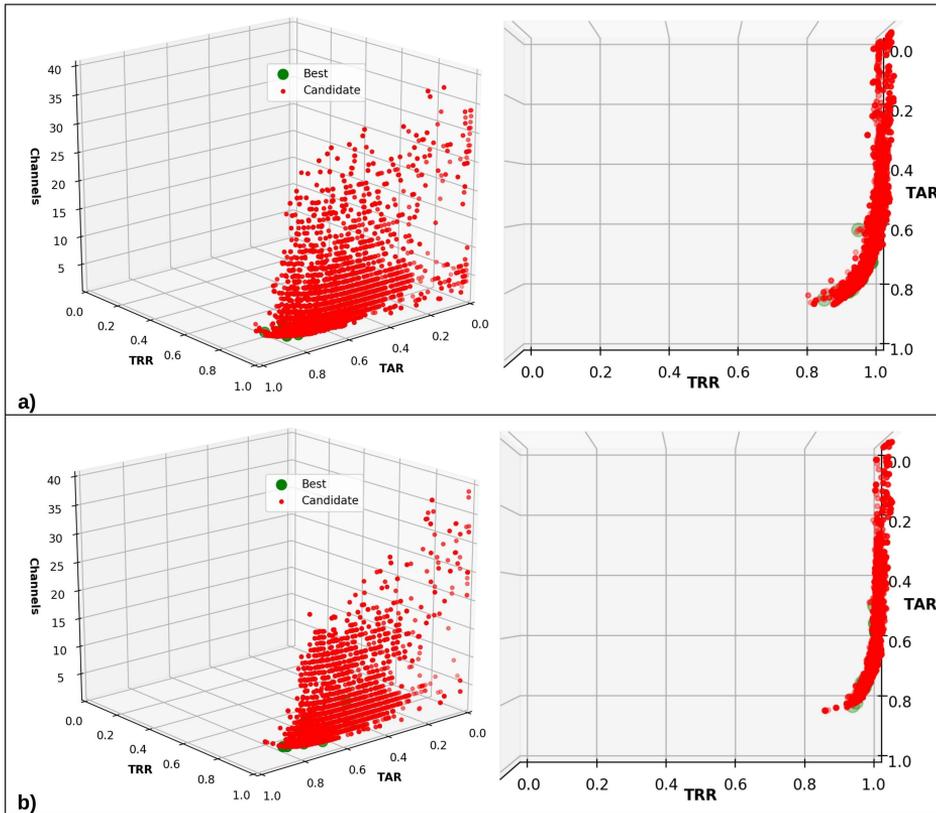


Figure 5.8: Frontal and aerial view of the TARs and TRRs obtained in the channel-selection process using EMD-based features (a) and DWT-based features (b) with OCSVM.

runs are presented.

A TAR of 0.822 ± 0.028 and a TRR of 0.969 ± 0.022 were obtained with only five channels using EMD-based features (see Table 5.7). The TAR and TRR were 0.822 ± 0.028 and 0.981 ± 0.017 , respectively, using DWT-based features and five channels with the optimization process.

As presented in Fig. 5.8, the candidates generated using EMD- or DWT-based features and OCSVM showed a clear tendency to reject all the subjects (which increased the TRR, since the models correctly rejected the intruders), even those in each model (which decreased the TAR), meaning that the models created for each subject did not learn from the provided features. TAR increased only if the

correct nu and $gamma$ parameters and channels were selected, which also varied in each run, as reflected by the standard deviations.

A set of channels used during the optimization process in the 10 runs is presented in Fig. 5.9. The set of channels identified when using EMD-based features is presented in *B*) and that when using DWT-based features in *a*). Each set of channels, from left to right, corresponds to the use of one to five channels, and, as mentioned earlier, the channels found by NSGA-III differed between runs for certain runs. The figure presents one set. Using EMD-based features, the channels found when using one to five channels differed, but those around *T10* and *T8* were consistent across most sets. When using DWT-based features, channel *IZ* clearly appeared in all sets, and channels *C4* and *T10* appeared in most.

5.5.3 Channel selection using NSGA-III and LOF for EEG signals for the resting-state with the eyes open

The optimization process was performed using the 109 subjects in the dataset, but now considering LOF for creating the models of each subject. NSGA-III was used for the channel-selection method using 64 binary genes in a chromosome to represent the EEG channels and two genes with integer values for the algorithm (1: ball tree, 2: k-d tree, 3: brute force) and the number of neighbors (From 1 to 10, which were proposed experimentally) to be used, obtaining thus a chromosome of 66 genes. The experiment was repeated 10 times for validation, each time using 80% of the instances of each subject for training and 20% for testing.

The results of the first run are presented in Fig. 5.10 as an example of the distribution of the TARs and TRRs during the optimization process and Table 5.8 presents the average results for both methods of feature extraction, EMD and DWT.

Using DWT-based features, it was possible to obtain an average TAR of up to 0.993 ± 0.001 and an average TRR of 0.941 ± 0.002 using only three EEG channels (see Table 5.8). The distribution of the results was very distinct and clear (see Fig. 5.10), indicating that similar TARs and TRRs can be obtained with different channel combinations using LOF and EMD- or DWT-based features.

The average distribution of the parameters used in the complete optimization process (for all generations and all chromosomes) is presented in Fig. 5.11, showing that the algorithm most often used by LOF was *ball tree* with three neighbors

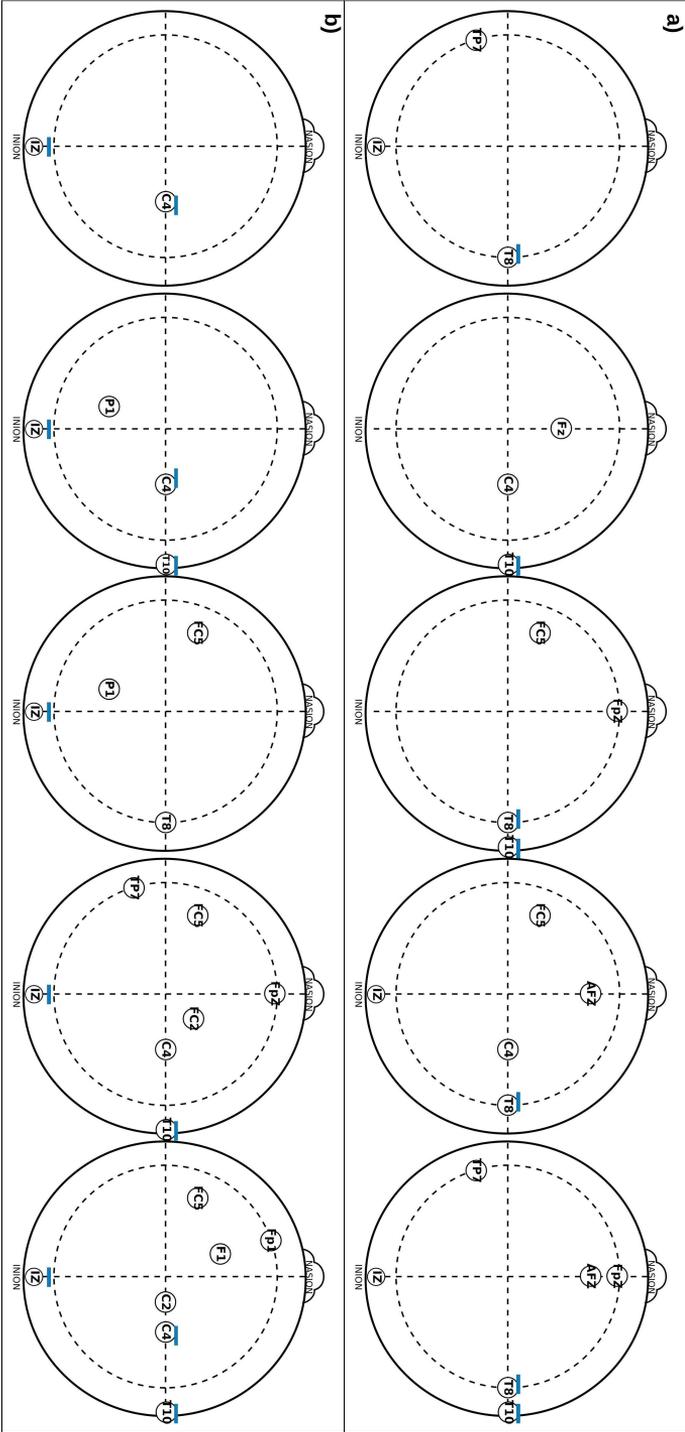


Figure 5.9: Set of one to five channels found during the optimization process for creating the biometric system with OCSVM using EMD-based features (a) or DWT-based features (b) and the resting-state with the eyes open.

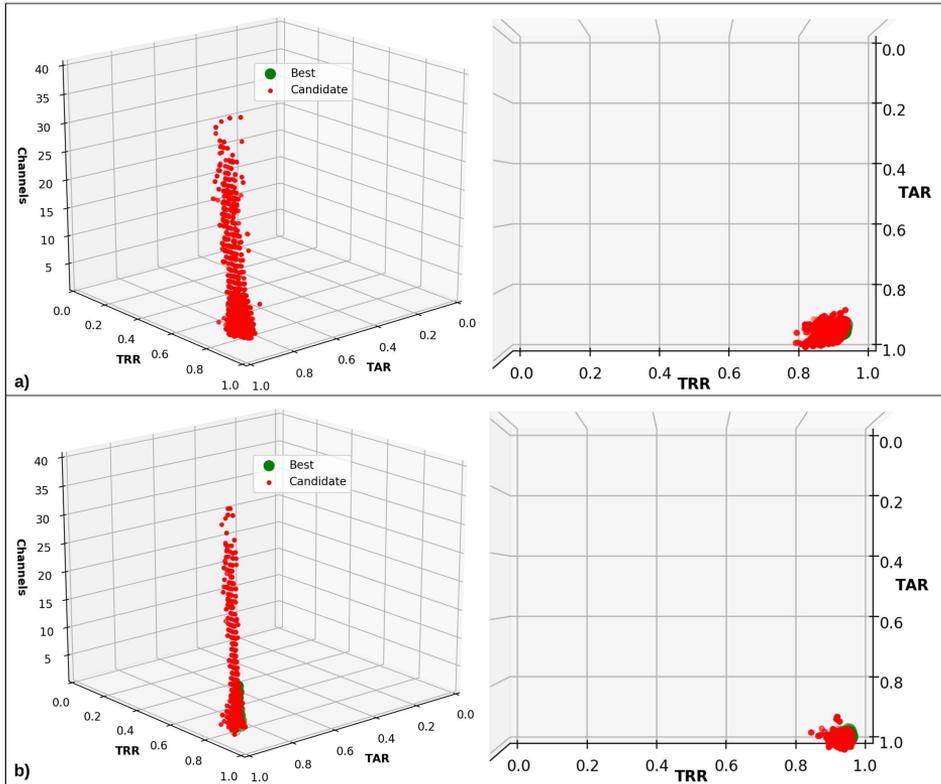


Figure 5.10: Frontal and aerial view of the TARs and TRRs obtained in the channel-selection process using EMD-based features (a)), and DWT-based features (b)) with LOF.

when using EMD-based features. The *ball tree* and *k-d tree* algorithms were used equally, with three neighbors, when DWT-based features were used. Analysis of only the parameters used for the results in the Pareto-front in the 10-fold cross-validation (for obtaining the results presented in Table 5.8) confirmed that the *ball tree* algorithm with three to four neighbors was the most often used for EMD-based features and the *ball tree* and *k-d tree* algorithms were used with only two neighbors for DWT-based features, as shown in Fig. 5.12.

Fig. 5.13 presents the set of channels of the 10 runs used to obtain the results presented in Table 5.8, which correspond to the use of one to seven channels using EMD-based features (a) in the figure) and DWT-based features (b) in the figure). In this case, the channels were almost the same using both methods and they did not

Table 5.8: TARs and TRRs obtained for the first seven EEG channels in the Pareto-front for three objectives solved with NSGA-III using EMD-based and DWT-based features and LOF.

No. channels	EMD-based features		DWT-based features	
	TAR	TRR	TAR	TRR
1	0.930 ± 0.005	0.904 ± 0.006	0.979 ± 0.001	0.888 ± 0.003
2	0.949 ± 0.002	0.909 ± 0.005	0.991 ± 0.001	0.922 ± 0.002
3	0.960 ± 0.003	0.909 ± 0.005	0.993 ± 0.001	0.941 ± 0.002
4	0.964 ± 0.005	0.918 ± 0.028	0.995 ± 0.011	0.949 ± 0.004
5	0.969 ± 0.008	0.926 ± 0.011	0.996 ± 0.006	0.952 ± 0.004
6	0.980 ± 0.003	0.938 ± 0.011	0.997 ± 0.006	0.957 ± 0.009
7	0.980 ± 0.004	0.940 ± 0.005	0.997 ± 0.001	0.957 ± 0.005

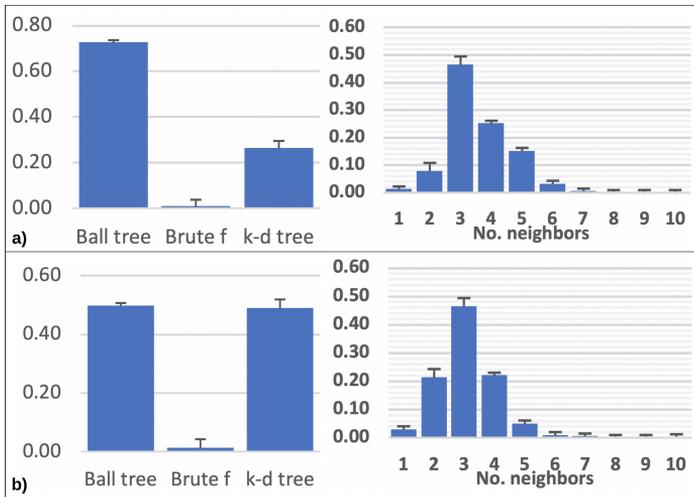


Figure 5.11: Average distribution of the algorithms and number of neighbors used in the optimization process with EMD-based features (a)) and DWT-based features (b)).

differ much when using one or three channels. Another important point is that channels *IZ*, *T8*, and *T10* were used in most cases for both EMD- and DWT-based features. The most relevant area was clearly centered around channels *C6*, *T8*, *T10* and *F5*.

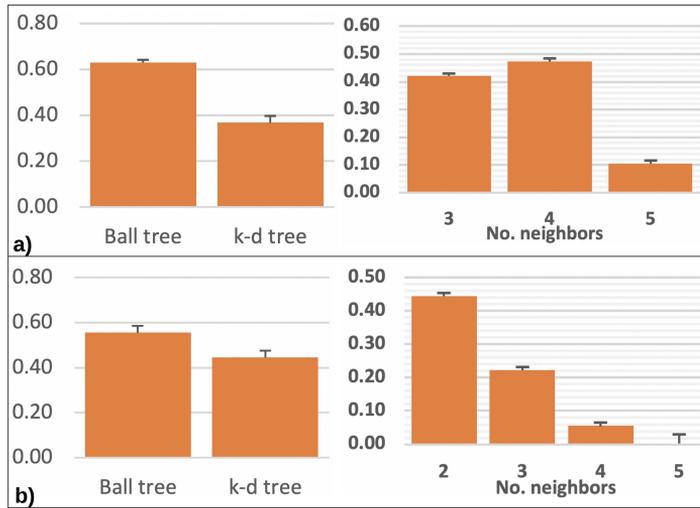


Figure 5.12: Average distribution of the algorithms and number of neighbors used for the results in the Pareto-front of the optimization process with EMD-based features (a) and DWT-based features (b).

5.5.4 Channel selection using NSGA-III and LOF for EEG signals for the resting-state with the eyes closed

Previous experiments using LOF resulted in higher TARs and TRRs with a lower number of EEG channels than when using OCSVM. The optimization process was repeated with EEG data from the 109 subjects but considering the resting-state with the eyes closed to provide additional information about the performance of LOF with EMD- and DWT-based features.

The chromosome representation was as in the previous experiment: 64 genes to represent the EEG channels and two additional genes with integer values for the different algorithms and number of neighbors. Each experiment was performed 10 times, randomly selecting 80% of the instances for training and 20% for testing, thus ensuring 10-fold cross-validation. The results obtained for runs using either EMD- or DWT-based features are presented in Fig. 5.14 for visualization and understanding of the behavior during the optimization process.

The average TAR and TRR in the Pareto-front for the first seven channels using EMD or DWT for feature extraction are presented in Table 5.9. The results show that subject identification was possible using the resting-state with the eyes

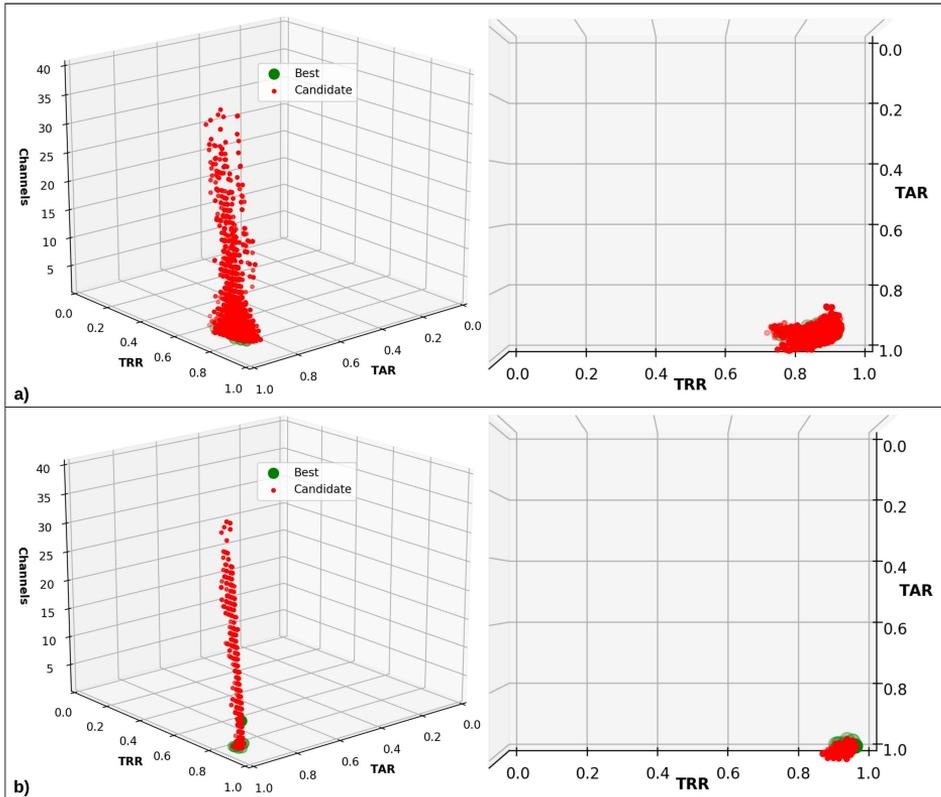


Figure 5.14: Frontal and aerial view of the TARs and TRRs obtained in the channel-selection process using EMD- (a)) and DWT-based features (b)) for the resting-state with the eyes closed, using LOF.

closed. The TAR and TRR were similar to those presented in Table 5.8 for the eyes open. The results were maintained throughout the 10 runs, especially when using DWT for feature extraction, as the standard deviation was 0.011 for the TAR and 0.009 for the TRR.

The average distribution of the parameters used during the entire optimization process is shown in Fig. 5.15. The *k-d tree* algorithm was the most used in both cases (using EMD or DWT) and the number of neighbors ranged from one to four, with a clear advantage of using two neighbors. The average parameters used for obtaining the results in the Pareto-front are presented in Fig. 5.16, confirming that the *k-d tree* algorithm was the most used and the number of neighbors still ranged

Table 5.9: TARs and TRRs obtained with LOF for the first seven EEG channels in the Pareto-front for three objectives solved with NSGA-III using EMD- or DWT-based features and the resting-state with the eyes closed.

No. channels	EMD-based features		DWT-based features	
	TAR	TRR	TAR	TRR
1	0.945 ± 0.005	0.888 ± 0.008	0.979 ± 0.001	0.881 ± 0.004
2	0.945 ± 0.005	0.918 ± 0.007	0.995 ± 0.001	0.935 ± 0.005
3	0.955 ± 0.005	0.918 ± 0.007	0.997 ± 0.002	0.950 ± 0.005
4	0.969 ± 0.003	0.926 ± 0.006	0.997 ± 0.002	0.950 ± 0.003
5	0.971 ± 0.002	0.933 ± 0.002	0.997 ± 0.002	0.951 ± 0.003
6	0.975 ± 0.001	0.945 ± 0.002	0.998 ± 0.000	0.953 ± 0.002
7	0.979 ± 0.002	0.955 ± 0.005	0.998 ± 0.000	0.955 ± 0.002

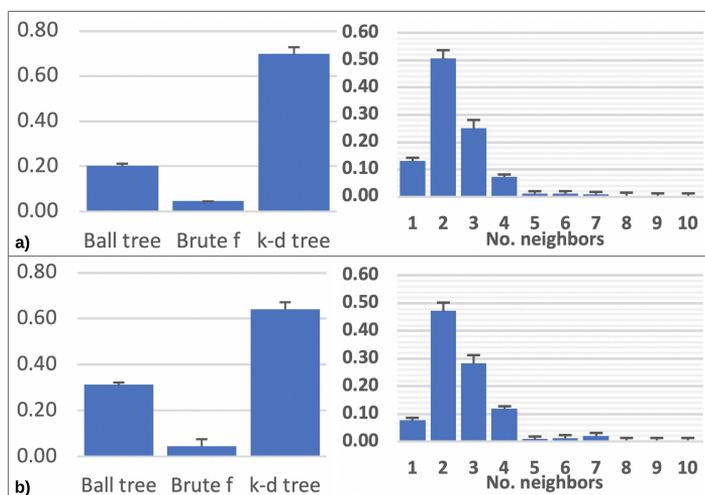


Figure 5.15: Average distribution of the algorithms and number of neighbors used in the optimization process with EMD-based features (*a*) and DWT-based features (*b*) using EEG signals for the resting-state with the eyes closed.

from one to four, with preferential use of only two neighbors.

As for the previous experiment using the resting-state with eyes open, Fig. 5.17 presents the set of channels found by the optimization process of the 10 runs used to create the models for the biometric system using the resting-state with the eyes closed and EMD-based features (*a* in the figure), as well as DWT-based

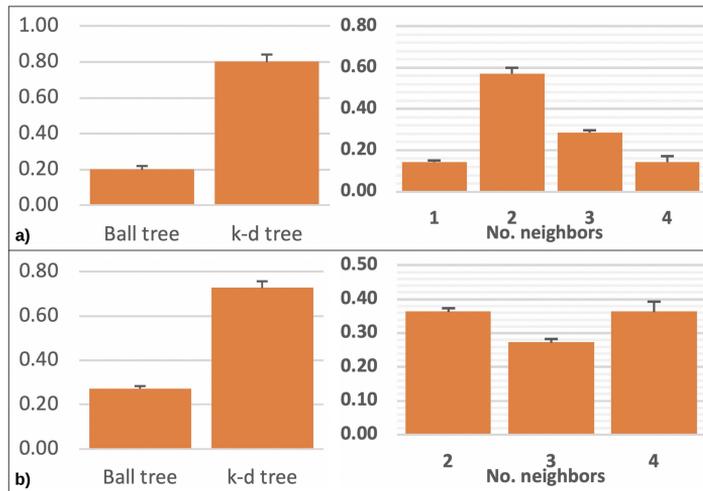


Figure 5.16: Average distribution of the algorithms and number of neighbors used for the results in the Pareto-front of the optimization process with EMD-based features (a) and DWT-based features (b) using EEG signals for the resting-state with the eyes closed.

features (b) in the figure). The results presented in 5.13 and 5.17 differed little, even between methods and the sets of different numbers of channels (In the sets created in the 10 runs with 1 to 7 channels). The most relevant area was still centered around channels *C6*, *T8*, *T10*, and *IZ*.

5.6 Discussion

This Chapter presented the application of EEG channel selection for biometric systems focused on the study and comparison of various task-dependent and task-independent paradigms, i.e., resting-state and ERPs, using various types of electrodes and various numbers of channels [173, 206, 222, 223]. The resting-state has been used in the state-of-the-art for this purpose as it does not require any training process for the subject. There are several approaches based on multi-class classification using machine-/deep-learning and one-class classification. ***Although most of the approaches can discriminate between the subjects in a database when using multi-class classification, they do not consider possible intruders.*** In the best case, one study presented a set of eight EEG channels selected beforehand [297]. Another used deep learning with a set of five

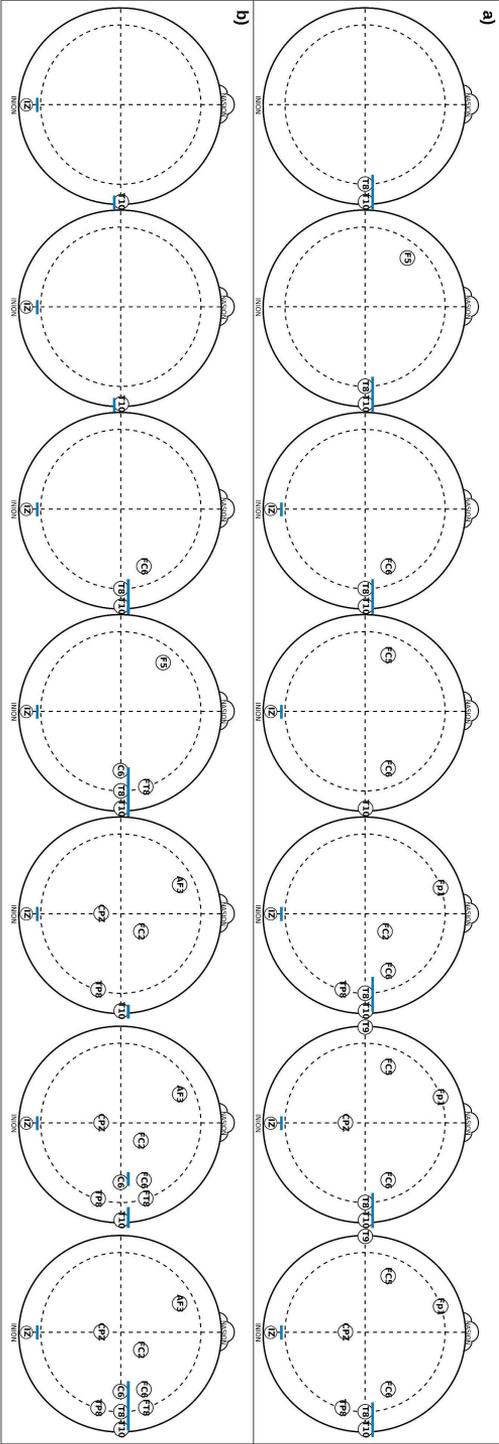


Figure 5.17: Set of one to seven channels found during the optimization process for creating the biometric system with LOF using EMD-based features (a) or DW-T-based features (b) and the resting-state with the eyes closed.

EEG channels, also selected beforehand, but they did not use the resting-state [281].

A method for channel selection was presented in Section 5.3 using a two-stage method tested on a dataset with 26 subjects for detecting intruders and then using multi-class classification to detect the name of the subject [138]. The stage for intruder detection was created using OCSVM with nu and $gamma$ parameters determined by a genetic algorithm that also selected the most relevant channels for the task. However, OCSVM was very sensitive to the nu and $gamma$ parameters.

Later, a new approach for an EEG-based biometric system was presented using brain signals recorded during the resting-state with the eyes open and the resting-state with the eyes closed using LOF and channels selected by NSGA-III. Briefly, a model using LOF with EMD-/DWT-based features was created for each subject that was able to reject the other 108 subjects in the dataset, ***confirming that the features extracted from each subject can help to discriminate between the subject in the model and the rest of the subjects, with good results, even with a low number of EEG channels and using 108 subjects as intruders.***

In this new approach, experiments using EEG signals for the resting-state with the eyes open and 64 EEG channels, with OCSVM and LOF using different parameters, were conducted. It was shown that a TAR of up to 1.000 ± 0.000 and a TRR of 0.998 ± 0.001 can be achieved using LOF and the k - d tree algorithm with only one neighbor, all using DWT-based features. Then, the experiment was repeated using 1 to 10 neighbors with DWT-based features, LOF, and the k - d tree algorithm, as they were the best parameters found in the previous experiment and also to show that a different number of neighbors affects the TAR and TRR.

It was also shown that OCSVM resulted in a TAR of 0.502 ± 0.004 and a TRR of 0.993 ± 0.001 , meaning that the models were unable to learn from any of the features of the subjects (EMD- or DWT-based). It was thus necessary to fit the best nu and $gamma$ parameters by using the multi-objective optimization process [138]. This resulted in substantially higher TAR and TRR values (see Fig. 5.8). In the best case, a TAR of up to 0.822 ± 0.028 and a TRR of 0.969 ± 0.22 using EMD-based features, and a TAR of 0.822 ± 0.28 and a TRR of 0.981 ± 0.017 using DWT-based features were obtained. However, the standard deviation was high.

The results presented with LOF when using the resting-state with the eyes

open show that a TAR of up to 0.993 ± 0.01 and a TRR of 0.941 ± 0.002 , with only three EEG channels and with only two EEG channels using DWT-based features, can be obtained. TAR and TRR values above 0.900 were obtained, which are higher than the best results obtained in the Pareto-front using EMD-based features. As shown in Fig. 5.10, the distribution of the TAR and TRR values was consistent when reducing the number of EEG channels during the optimization process, showing that the models created with LOF learned well from the features provided and that different channel combinations were used to obtain the best results, as presented in Table 5.8. In this case, the most highly used algorithm for the complete optimization process was *ball tree*, with three neighbors. Analysis of the parameters using DWT-based features and only the results obtained in the Pareto-front show the use of the *ball tree* and *k-d tree* algorithms to be highly similar using only two neighbors.

The use of EEG signals from the resting-state with the eyes closed and LOF confirmed that DWT-based features work better, with a TAR of up to 0.997 ± 0.002 and TRR of up to 0.950 ± 0.005 with only three EEG channels. The *k-d tree* algorithm with two to four neighbors was the most used for the complete optimization process, as well as the results obtained for the Pareto-front.

The use of OCSVM can provide good results if the appropriate parameters are chosen. Otherwise, the TAR can decrease substantially. This behavior needs to be further investigated using different feature extraction methods and compared to the results using different-sized datasets. On the other hand, LOF proved to be a robust classifier for creating an EEG-based biometric system, especially using DWT-based features with the *ball tree* or *k-d tree* algorithms and two to four neighbors. In the future, it will be evaluated to determine whether solving the problems related to EMD (best spline, end effects, mode mixing, etc.) can improve the results presented in this study.

Comparing the results presented in Figs. 5.9, 5.13 and 5.17, it is evident that the use of LOF allowed localization of the potentially most relevant area for choosing a possible set of channels, which will require further investigation in the future. ***It is noteworthy that the channel distribution did not substantially vary whether the eyes were open or closed in the resting state.***

The localization of most of the relevant channels, i.e., the channels that were

found in most of the sets, was mainly centered around channels *F5*, *T8*, *T10*, and *IZ*, and as shown in Fig. 5.13, it was clearer for the resting-state with the eyes open. In general, most of the channels are localized in the temporal and frontal areas, as well as around the inion, which may be associated with the previous task performed during the data collection. This is an aspect that must be tested using other datasets [301–303].

One of the purposes of this study was to prove that the resting-state can be used as a paradigm to create a biometric system in large datasets. A set of experiments was provided in which high-density EEG data was available for the training and testing stages, but for real-time implementation of a biometric system, only a few of the best channels will be selected for designing a new portable headset tailored for this purpose. With the set of experiments and the methods tested for classification and optimization, a proof-of-concept for a biometric system based on the resting-state was provided using a small number of electrodes using a pool with a large number of subjects (109 subjects) versus previous studies using smaller datasets.

However, the current results do show whether or not there is a unique subset of EEG channels or brain regions that works better for creating a biometric system using the resting-state. This study lays the groundwork for pursuing further research into the analysis of various public and private datasets to identify a unique subset of channels that can be used in the design of a new portable and easy-to-use EEG headset that can be tested in real-time, adding new subjects to the system and identifying them using only a few electrodes.

The progress in subject identification using EEG signals from various paradigms has been remarkable in the last several years, but one of the most relevant unsolved problems is the fact that the new approaches have all been tested and validated using EEG datasets recorded in well-controlled environments [296, 304]. Most of the studies using high-density EEG signals were recorded with medical-grade sensor systems (using a gel or saline solution for improving conductivity), which may increase the performance of the methods. However, ease-of-use will be essential for practical and portable devices and dry electrodes may offer certain opportunities [304, 305]. In general, analysis and validation in real-life scenarios is necessary. In this context, the best and fastest methods will

be studied in a more realistic way and the appropriate and necessary number of trials per subject will be considered [173].

For certain BCI applications, the problem of recognizing new instances from new sessions has been studied using EEG data from different sessions or adding new instances for calibration. In the case of session-to-session or subject-to-subject transfer, the learning problem has been studied using LDA and SVM, based on motor imagery or P300 paradigms [148, 306–309]. To adapt the EEG feature space and thus reduce session-to-session variability, a data-space adaptation method based on the Kullback-Leibler divergence criterion (also called relative entropy) can be used, aiming to minimize the distribution of differences from the training session to a different session [307]. There is evidence that for certain BCIs, it is possible to use background noise immediately before a new session to improve session-to-session variability using a regularized spatio-temporal filter [308].

The dataset used in the second approach consists of EEG signals from a single session (see Section 5.5), which limits the experimental configurations and does not allow evaluation of whether one can create models for each subject from a certain session and be able to recognize the subjects or reject them using data from another session. Future steps will be focused on tackling this problem by analyzing possible ways to use new correctly-classified instances to decrease session-to-session variability, data augmentation techniques, as well as using and comparing current progress in transfer learning using machine-/deep-learning methods to address this problem [282, 309].

Another point to be analyzed in future work is to develop new ways to extract and select the features to improve the TRR and TAR. ***This can be achieved using a big bag-of-features from the different sub-bands (possibly from both the EMD and DWT methods) and by adding additional GA genes to represent such features in the chromosomes and thus select the best features during the optimization process, at the same time as selection of the best channels.***

In general, the resting-state has been shown to be a good candidate but there is not yet sufficient research evidence using larger datasets and different stages. Future efforts will be focused on relevant parameters that can be extracted from the EEG signals of each subject and thus add information for the complete authentication and verification process, such as re-evaluating the accepted subject

using multi-class classification, detecting the age-range and sex of the subjects, etcetera [86].

This research has been focused towards a portable (non-invasive) wireless low-density EEG system for various applications that can help the subject-identification process by providing EEG information from different channel combinations using a movable sensor [57, 173]. Following the results found in this work and the proposed experiments, the possibility of a fixed or movable electrode version of a new EEG headset that incorporates the best results obtained in this thesis for subject identification and authentication will be evaluated.

Chapter 6

Conclusions and future work

In this Chapter, an overview of the achieved results in comparison with the objectives of the thesis formulated in Section 1.2 is provided and their implications for future work discussed.

6.1 Summary of findings

6.1.1 Feature extraction and channel count optimization for epileptic seizure classification

In the first paper related to this thesis [135], the backward-elimination algorithm was used to reduce the number of necessary EEG channels for epileptic seizure classification and was the basis for understanding the problem and the necessary parameters to be optimized for this task. Later, in Chapter 4 and [200] the method for channel selection was improved using NSGA-II and proved to be robust for epileptic-seizure classification.

It was shown that *SVM* was the most highly-used classifier, independently of whether the features were extracted using the EMD-based or DWT-based method or whether NSGA-II or NSGA-III were used for channel selection. The presented results show that *KNN* was also highly used but only when the features were extracted using the DWT.

The presented methods show that it is possible to classify between epileptic seizures and seizure-free instances using only one channel, obtaining accuracy values of up to 0.97 ± 0.05 using DWT-based features and selecting the channels using the NSGA-III algorithm. An important finding is that NSGA-III is able to find the most relevant EEG channels with features based on DWT, selecting

combinations with only two or three channels, obtaining accuracy values of up to 0.98 and 0.99, respectively.

The results discussed in Chapter 4 and, in general, the methods implemented for channel selection and feature extraction will enable the prediction of epileptic seizures with low-density EEG headsets for long-term monitoring in daily life, attaining the advantages related to channel selection described in Section 3.5.

6.1.2 Channel count optimization for EEG-based biometric systems

This thesis has argued that EEG-based biometric systems are a good candidate for use in authentication systems [87, 138, 173, 206, 222, 223]. The presented results have shown that it is possible to identify subjects by their brain signals using the methods proposed for feature extraction and classification. The most important aspect is that it is also possible to distinguish between subjects who were part of the trained dataset from those who are intruders.

The first approach presented consisted of a two-stage method tested in a dataset with 26 subjects. The first stage consisted of OCSVM, validating the results with the TAR and TRR, and the second stage used multi-class classification to identify the name of the subject. This set of experiments showed that OCSVM is sensitive to the nu and $gamma$ parameters.

NSGA-II found channel sets of two EEG channels to obtain accuracy values of up to 0.78, with a TAR of 0.91 and a TRR of 0.88. However, using NSGA-III, it was possible to find subsets with 7, 9, 10, or 11 EEG channels to obtain accuracy values of up to 0.99 and both a TAR and TRR of 1.00.

Several facts make it impossible to draw any final conclusions about the minimum number of necessary EEG channels for a new biometric system based on ERPs or P300, as the channel subsets differed depending on the number of instances per subject, the sessions available, and the method used for feature extraction. The sets of channels also differed depending on whether the NSGA-II or NSGA-III algorithm was used for channel selection.

When the biometric system was created using the resting-state, LOF for one-class classification, and the channels selected by NSGA-III, the results were more robust using EMD or DWT for feature extraction and a low number of EEG channels, as the models were able to reject 108 subjects.

The results obtained with EEG signals while the subjects had their eyes open show that it is possible to obtain a TAR of up to 0.993 ± 0.01 and a TRR of 0.941 ± 0.002 using two or three channels with DWT-based features.

From the results presented in Chapter 5, it is possible to argue that LOF proved to be a robust classifier for creating an EEG-based biometric system, especially using DWT-based features with the *ball tree* or *k-d tree* algorithms and two to four neighbors.

It is noteworthy that the subsets of channels selected by NSGA-III did not substantially differ whether the eyes were open or closed during the resting state, i.e., it is possible to find certain relevant areas, which in this case was centred around channels *F5*, *T8*, *T10*, and *IZ*.

It is not currently possible to argue that there is a unique set of channels that works better for extracting features to create a biometric system using the resting-state. This will need to be tested in a larger population and the influence of the main four micro-states during the resting-state verified [89, 90, 92–94].

6.2 Conclusion of the thesis contributions

The work presented in this thesis consisted of a method for decomposing EEG signals into different sub-bands using EMD or DWT, followed by the extraction of four features: the *Teager* and *instantaneous* energy distributions and the *Higuchi* and *Petrosian* fractal dimensions. With these features, the EEG signal segment corresponding to the resting-state, P300 response, or epileptic seizures, as well as seizure-free periods, are successfully represented. Thus, the proposed method has been presented as a robust method for extracting information from EEG signals and thus represents the events of interest in a compact form for creating a classifier model that can be used for classification in real-time. In this context, various classifiers were tested, either multi-class classifiers or one-class classifiers, depending on the case of the study.

Tailored experiments were performed using methods for channel reduction (using the backward-elimination and forward-addition greedy algorithms) and selection [86, 87, 135, 138, 173, 200, 206, 206, 223]. However, for the experiments presented in this thesis, the backward-elimination algorithm was only briefly used. Most of the experiments for channel selection were carried out using NSGA-based algorithms, especially NSGA-III.

In the first approaches using NSGA, certain important features for the classifiers were optimized by adding genes with only two possible values, 0 or 1. However, the possible values that can be generated by these combinations are reduced. Thus, the parameters to be optimized were later represented using decimal values. An example is the optimization of the *nu* and *gamma* parameters of OCSVM, in which both genes were defined using decimal values. However, in other cases, the range of possible values for the genes was defined as an interval to select the number of neighbors for the LOF classifier. Thus, the chromosome representation for the optimization process is reduced and the interpretation of the results made easier. In addition, the possible values of these genes better represent the problem.

A method that showed good performance was presented in two different case studies, thus contributing to the idea that a general method for EEG signal processing and feature extraction can be proposed. This thesis focused on **case study 1**, in which it was shown that the classification of epileptic seizures is possible, even when using a reduced array of EEG channels, and **case study 2**, in which various experiments were presented comparing methods and approaches for creating a biometric system using EEG signals.

The method for representing the EEG channels, as well as important parameters for the classifiers, were shown to be robust for selecting the most important source of information in the classification process. With these results, it appears to be possible to work with a small array of non-invasive EEG sensors for different classification problems using brain signals. This is important, as this could contribute to a reduction in the current size of EEG headsets and caps for portability, thus increasing the classification performance by using only the important information related to the task and widening the spectrum of applications using brain signals.

The results presented and the ideas discussed support the objective of channel selection presented in Section 3.5. Importantly, they will also help to reduce the preparation time for using an EEG headset and help to achieve a low-power hardware design.

Some of the proposed work has already been carried out on different EEG signal classification tasks. For example, a similar process was used in a Master's degree

theses [310–312] and the same process for feature extraction and classification of the response to RGB color exposure [313–315]. The process for channel selection using NSGA-II was also used for source localization, reducing the number of EEG channels from 231 to less than 10, while obtaining similar localization errors [316]. This shows that the method can be adapted to different problems with the same objective of reducing the number of necessary EEG channels for diverse BCI applications.

6.3 Future work

For the first case study, the multi-class classifier used was selected by first testing all the classifiers and performing iterations between a set of parameters, i.e., SVM was tested with the *linear*, *RBF*, *sigmoid* and *polynomial*. However, all possible parameters for the classifiers will be represented in the same chromosome representation in future work, as for the channels. Thus, a set of the best parameters for epileptic-seizure classification will be ensured, as for the case of EEG-based biometric systems.

As discussed in Chapter 5, the EEG-based biometric system can be modified to include more stages, in which, for example, the age of the subject, their sex, stress level, and other important descriptors can be identified [86]. By doing this, intruder detection will be easier to handle and the biometric system more robust to manage a larger number of subjects in the database.

Future studies will therefore be focused on: **1)** improving the proposal for the biometric system and validating it using a larger dataset with EEG signals from different sessions on the same day and **2)** using larger datasets from different days. **3)** The proposed biometric system must manage the problem of reducing the number of channels for real-time use, as well as for portability and comfort. However, it must be able to train a model for recognizing the subjects with just a few instances, as in fingerprint and face-recognition systems. In this context, another important problem that must be tackled, which is also important for most BCI applications, is related to data augmentation. Collecting a few EEG instances and then creating artificial instances with information from the collected signal will increase the feasibility of the biometric system. Thus, this proposal will be more competitive with current biometric systems.

Data augmentation methods will be proposed in an attempt to solve this

problem and will also help in the transfer learning problem related to epileptic seizure classification. 4) In the case of epilepsy, the machine-learning models must be able to recognize the seizures of new subjects in the database, without adding any seizure data, but by first testing whether it is improved by adding instances from the new subject to be analyzed, as well as adding new artificial instances for increasing the performance of the models.

The dataset used in the second approach of **case study 2** consists of EEG signals from a single session (see Section 5.5), which limits the experimental configurations and does not allow evaluation of whether one can create models for each subject from a certain session and be able to recognize the subjects or reject them using data from another session.

Future steps will be focused on tackling this problem and analyzing a possible way to use new correctly-classified instances to decrease session-to-session variability, data augmentation techniques, and comparing current progress in transfer learning, using machine-/deep-learning methods for this problem [282, 309].

The use of deep-learning techniques for real-time applications in EEG is still a challenge, due to the normally high computational cost. However, an interesting future study is related to the use of auto-encoders for one-class classification and will compare their performance to that of LOF and OCSVM [317].

The use of ever-larger datasets (i.e., a larger number of subjects) is still necessary using EEG data from different sessions and of different lengths, as well as considering fewer instances for training for both studying epileptic-seizure classification and creating a biometric system. Additionally, whether solving the problems related to EMD (best spline, end effects, mode mixing, etc.) or using different EMD-based algorithms, such as multivariate EMD (MEMD) [318] or Adaptive EMD (AEMD) [319], etc., can improve the results presented in both study cases will be evaluated.

As mentioned in Section 3.5, various approaches for channel selection in motor imagery classification have been proposed, but there has been no evaluation between all these techniques to identify a set of EEG channels [172, 174, 176, 179, 188, 196, 198, 199]. Therefore, future efforts will also focus on testing the various approaches for the classification of motor imagery and the

selection of channels to compare them with the methods proposed in this thesis.

The energy and fractal features extracted from the sub-bands obtained after applying DWT or EMD were shown to be useful and robust across experimental setups and for both study cases. However, as mentioned in the discussion of Chapter 5, future work will include selection of the best subset of features by including it during the optimization process (which could be by using a big bag-of-features). This would make it possible to verify whether this set is still the best for these and new EEG-based applications and whether there are new features capable of extracting useful patterns from EEG signals.

Future efforts will also be focused on feature selection by using NSGA-III or recent proposals in multi-objective optimization, such as multi-objective evolutionary algorithms based on decomposition (MOEA/D) [320]. These could be used to select the best levels of decomposition from DWT or the best IMFs from EMD by selecting the best subsets of features while reducing the number of required EEG channels, which could be for epileptic-seizure classification and prediction, improving the biometric system, or for a different task associated with EEG signal analysis.

Towards finding a unique set of channels for EEG signal processing, it will be necessary to test whether it is possible to force NSGA-based (especially NSGA-III) or MOEA/D-based algorithms to select a single array of EEG channels by running different folds in parallel while using the same chromosome for selecting the channels and the necessary parameters for one-class or multi-class classification.

Future studies will focus on all these relevant aspects, involving the optimization of multiple parameters related to feature extraction and machine-learning methods by using discrete values for representing the chromosomes, as carried out in the second approach of biometric systems presented in Section 5.5, and not only as a binary sequence.

References

- [1] Elena Ratti, Shani Waninger, Chris Berka, Giulio Ruffini, and Ajay Verma. Comparison of medical and consumer wireless EEG systems for use in clinical trials. *Frontiers in human neuroscience*, 11:398, 2017.
- [2] Herbert Jasper. Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalogr Clin Neurophysiol*, 10:370–375, 1958.
- [3] Robert Oostenveld and Peter Praamstra. The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical neurophysiology*, 112(4):713–719, 2001.
- [4] American Electroencephalographic Society. Guideline thirteen: Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, 11(1):111–3, 1994.
- [5] Marc R Nuwer, Giancarlo Comi, Ronald Emerson, Anders Fuglsang-Frederiksen, Jean-Michel Guérit, Hermann Hinrichs, Akio Ikeda, Fransisco Jose C Luccas, and Peter Rappelsburger. IFCN standards for digital recording of clinical EEG. *Electroencephalography and clinical Neurophysiology*, 106(3):259–261, 1998.
- [6] Jeffrey M Rogers, Stuart J Johnstone, Anna Aminov, James Donnelly, and Peter H Wilson. Test-retest reliability of a single-channel, wireless EEG system. *International Journal of Psychophysiology*, 106:87–96, 2016.
- [7] Silvia Erika Kober and Christa Neuper. Sex differences in human EEG theta oscillations during spatial navigation in virtual reality. *International Journal of Psychophysiology*, 79(3):347–355, 2011.
- [8] Yuji Wada, Yuko Takizawa, Jiang Zheng-Yan, and Nariyoshi Yamaguchi. Gender differences in quantitative EEG at rest and during photic stimulation in normal young adults. *Clinical Electroencephalography*, 25(2):81–85, 1994.
- [9] Nsreen Alahmadi, Sergey A Evdokimov, Yury Juri Kropotov, Andreas M Müller, and Lutz Jäncke. Different resting state EEG features in children from Switzerland and Saudi Arabia. *Frontiers in human neuroscience*, 10:559, 2016.
- [10] Jeannette McGlone. Sex differences in human brain asymmetry: A critical survey. *Behavioral and brain sciences*, 3(2):215–227, 1980.
- [11] Rytis Maskeliunas, Robertas Damasevicius, Ignas Martisius, and Mindaugas Vasiljevas. Consumer-grade EEG devices: are they usable for control tasks? *PeerJ*, 4:e1746, 2016.
- [12] Richard Caton. Electrical currents of the brain. *The Journal of Nervous and Mental Disease*,

- 2(4):610, 1875.
- [13] Lindsay F Haas. Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(1):9–9, 2003.
- [14] Anton Coenen and Oksana Zayachkivska. Adolf Beck: A pioneer in electroencephalography in between Richard Caton and Hans Berger. *Advances in cognitive psychology*, 9(4):216, 2013.
- [15] Anton Coenen, Edward Fine, and Oksana Zayachkivska. Adolf Beck: a forgotten pioneer in electroencephalography. *Journal of the History of the Neurosciences*, 23(3):276–286, 2014.
- [16] Hans Berger. Über das elektroenkephalogramm des menschen. *Archiv für psychiatrie und nervenkrankheiten*, 87(1):527–570, 1929.
- [17] Christoph M Michel and Micah M Murray. Towards the utilization of EEG as a brain imaging tool. *Neuroimage*, 61(2):371–385, 2012.
- [18] Jeffrey W Britton, Lauren C Frey, Jennifer L Hopp, Pearce Korb, Mohamad Z Koubeissi, William E Lievens, Elia M Pestana-Knight, and EK Louis St. *Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants*. American Epilepsy Society, Chicago, 2016.
- [19] Fabian Pedregosa-Izquierdo. *Feature extraction and supervised learning on fMRI: from practice to theory*. PhD thesis, Université Pierre et Marie Curie, 2015.
- [20] Arthur W Toga. *Brain mapping: An encyclopedic reference*. Academic Press, 2015.
- [21] John William Carey Medithe and Usha Rani Nelakuditi. Study of normal and abnormal EEG. In *2016 3rd International conference on advanced computing and communication systems (ICACCS)*, volume 1, pages 1–4. IEEE, 2016.
- [22] Maria Emilia Cosenza Andraus and Soniza Vieira Alves-Leon. Non-epileptiform EEG abnormalities: an overview. *Arquivos de Neuro-Psiquiatria*, 69(5):829–835, 2011.
- [23] Claudio Babiloni, Robert J Barry, Erol Başar, Katarzyna J Blinowska, Andrzej Cichocki, Wilhelmus HIM Drinkenburg, Wolfgang Klimesch, Robert T Knight, Fernando Lopes da Silva, Paul Nunez, et al. International Federation of Clinical Neurophysiology (IFCN)–EEG research workgroup: Recommendations on frequency and topographic analysis of resting state EEG rhythms. Part 1: Applications in clinical research studies. *Clinical Neurophysiology*, 131(1):285–307, 2020.
- [24] Catherine Tallon-Baudry. Oscillatory synchrony and human visual cognition. *Journal of Physiology-Paris*, 97(2-3):355–363, 2003.
- [25] Lawrence M Ward. Synchronous neural oscillations and cognitive processes. *Trends in cognitive sciences*, 7(12):553–559, 2003.
- [26] Derk-Jan Dijk, Daniel P Brunner, Domien GM Beersma, and Alexander A Borbély. Electroencephalogram power density and slow wave sleep as a function of prior waking and circadian phase. *Sleep*, 13(5):430–440, 1990.
- [27] Jean Reither, Michel Beaudry, and Charles P Leduc. Temporal intermittent rhythmic delta activity (TIRDA) in the diagnosis of complex partial epilepsy: sensitivity, specificity and predictive value. *Canadian journal of neurological sciences*, 16(4):398–401, 1989.
- [28] Chetan S Nayak and Arayamparambil C Anilkumar. Eeg normal waveforms. *StatPearls [Internet]*, 2020.

-
- [29] José Luis Cantero and Mercedes Atienza. Alpha burst activity during human REM sleep: descriptive study and functional hypotheses. *Clinical neurophysiology*, 111(5):909–915, 2000.
- [30] Jose L Cantero, Mercedes Atienza, and Rosa M Salas. Human alpha oscillations in wakefulness, drowsiness period, and REM sleep: different electroencephalographic phenomena within the alpha band. *Neurophysiologie Clinique/Clinical Neurophysiology*, 32(1):54–71, 2002.
- [31] Paul Gerrard and Robert Malcolm. Mechanisms of modafinil: a review of current research. *Neuropsychiatric disease and treatment*, 3(3):349, 2007.
- [32] Robert B Aird and Y Gastaut. Occipital and posterior electroencephalographic rhythms. *Electroencephalography and clinical neurophysiology*, 11(4):637–656, 1959.
- [33] Martica Hall, Julian F Thayer, Anne Germain, Douglas Moul, Raymond Vasko, Matthew Puhl, Jean Miewald, and Daniel J Buysse. Psychological stress is associated with heightened physiological arousal during NREM sleep in primary insomnia. *Behavioral sleep medicine*, 5(3):178–193, 2007.
- [34] Gert Pfurtscheller and FH Lopes Da Silva. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology*, 110(11):1842–1857, 1999.
- [35] Greg Worrell and Jean Gotman. High-frequency oscillations and other electrophysiological biomarkers of epilepsy: clinical studies. *Biomarkers in medicine*, 5(5):557–566, 2011.
- [36] Nicole Ille, Patrick Berg, and Michael Scherg. Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of clinical neurophysiology*, 19(2):113–124, 2002.
- [37] Peter Anderer, Stephen Roberts, Alois Schlögl, Georg Gruber, Gerhard Klösch, Werner Herrmann, Peter Rappelsberger, Oliver Filz, Manel J Barbanoj, Georg Dorffner, et al. Artifact processing in computerized analysis of sleep EEG—a review. *Neuropsychobiology*, 40(3):150–157, 1999.
- [38] Jose Antonio Urigüen and Begoña Garcia-Zapirain. EEG artifact removal—state-of-the-art and guidelines. *Journal of neural engineering*, 12(3):031001, 2015.
- [39] William O Tatum, Barbara A Dworetzky, and Donald L Schomer. Artifact and recording concepts in EEG. *Journal of clinical neurophysiology*, 28(3):252–263, 2011.
- [40] Mehrdad Fatourechi, Ali Bashashati, Rabab K Ward, and Gary E Birch. EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical neurophysiology*, 118(3):480–494, 2007.
- [41] Franklin F Offner. The EEG as potential mapping: the value of the average monopolar reference. *Electroencephalography and clinical neurophysiology*, 2(2):213, 1950.
- [42] Pablo F Diez, Vicente Mut, Eric Laciari, and Enrique Avila. A comparison of monopolar and bipolar EEG recordings for SSVEP detection. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 5803–5806. IEEE, 2010.
- [43] Marc Saab. Basic concepts of surface electroencephalography and signal processing as applied to the practice of biofeedback. *Biofeedback*, 36(4):128, 2008.
- [44] Christoph M Michel and Denis Brunet. EEG source imaging: a practical review of the analysis steps. *Frontiers in neurology*, 10:325, 2019.
- [45] Uros Topalovic, Zahra M Aghajan, Diane Villaroman, Sonja Hiller, Leonardo Christov-Moore,

- Tyler J Wishard, Matthias Stangl, Nicholas R Hasulak, Cory Inman, Tony A Fields, et al. Wireless Programmable Recording and Stimulation of Deep Brain Activity in Freely Moving Humans. *bioRxiv*, 2020.
- [46] GE Chatrian, E Lettich, and PL Nelson. Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *American Journal of EEG technology*, 25(2):83–92, 1985.
- [47] Catherine J Chu. High density EEG—What do we have to lose? *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, 126(3):433, 2015.
- [48] I Pisarenco, M Caporro, C Prosperetti, and M Manconi. High-density electroencephalography as an innovative tool to explore sleep physiology and sleep related disorders. *International Journal of Psychophysiology*, 92(1):8–15, 2014.
- [49] Amanda K Robinson, Praveen Venkatesh, Matthew J Boring, Michael J Tarr, Pulkit Grover, and Marlene Behrmann. Very high density EEG elucidates spatiotemporal aspects of early visual processing. *Scientific reports*, 7(1):1–11, 2017.
- [50] Anders Bach Justesen, Mette Thrane Foged, Martin Fabricius, Christian Skaarup, Nizar Hamrouni, Terje Martens, Olaf B Paulson, Lars H Pinborg, and Sándor Beniczky. Diagnostic yield of high-density versus low-density EEG: The effect of spatial sampling, timing and duration of recording. *Clinical Neurophysiology*, 130(11):2060–2064, 2019.
- [51] Andres Soler, Pablo A Muñoz-Gutiérrez, Maximiliano Bueno-López, Eduardo Giraldo, and Marta Molinas. Low-Density EEG for Neural Activity Reconstruction Using Multivariate Empirical Mode Decomposition. *Frontiers in Neuroscience*, 14, 2020.
- [52] Phattarapong Sawangjai, Supanida Hompoonsup, Pitshaporn Leelaarporn, Supavit Kongwudhikunakorn, and Theerawit Wilaiprasitporn. Consumer grade eeg measuring sensors as research tools: A review. *IEEE Sensors Journal*, 20(8):3996–4024, 2019.
- [53] John LaRocco, Minh Dong Le, and Dong-Guk Paeng. A systemic review of available low-cost EEG headsets used for drowsiness detection. *Frontiers in neuroinformatics*, 14, 2020.
- [54] Nikolas Williams, Genevieve M McArthur, and Nicholas A Badcock. 10 years of epoc: A scoping review of emotiv’s portable eeg device. *BioRxiv*, 2020.
- [55] Jérémy Frey. Comparison of an open-hardware electroencephalography amplifier with medical grade device in brain-computer interface applications. *arXiv preprint arXiv:1606.02438*, 2016.
- [56] Marta Molinas, Audrey Van der Meer, Nils Kristian Skjærvold, and Lars Lundheim. David versus Goliath: single-channel EEG unravels its power through adaptive signal analysis-FlexEEG. *Research project*, 2018.
- [57] Luis Alfredo Moctezuma, Andres Felipe Soler Guevara, Erwin Habibzadeh Tonekabony Shad, Alejandro Antonio Torres-Garcia, and Marta Molinas. David versus Goliath: Low-density EEG unravels its power through adaptive signal analysis – FlexEEG. In *4th HBP Student Conference On Interdisciplinary Brain Research*, 2020.
- [58] Marta Molinas, Trond Ytterdal, Audrey Van der Meer, and Luis Romundstad. FlexEEG: EEG scanning for highly portable, real-time functional brain mapping. *Research project*, 2018.
- [59] Lloyd M Nirenberg, John Hanley, and Edwin B Stear. A new approach to prosthetic control:

- EEG motor signal tracking with an adaptively designed phase-locked loop. *IEEE Transactions on Biomedical Engineering*, BME-18(6):389–398, 1971.
- [60] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain–computer interfaces for communication and control. *Clinical neurophysiology*, 113(6):767–791, 2002.
- [61] Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering*, 4(2):R1, 2007.
- [62] Jonathan R Wolpaw and Dennis J McFarland. Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans. *Proceedings of the national academy of sciences*, 101(51):17849–17854, 2004.
- [63] Jonathan R Wolpaw. Brain–computer interfaces as new brain output pathways. *The Journal of physiology*, 579(3):613–619, 2007.
- [64] Jose M Carmena, Mikhail A Lebedev, Roy E Crist, Joseph E O’Doherty, David M Santucci, Dragan F Dimitrov, Parag G Patil, Craig S Henriquez, and Miguel AL Nicolelis. Learning to control a brain–machine interface for reaching and grasping by primates. *PLoS biol*, 1(2):e42, 2003.
- [65] Dawn M Taylor, Stephen I Helms Tillery, and Andrew B Schwartz. Direct cortical control of 3D neuroprosthetic devices. *Science*, 296(5574):1829–1832, 2002.
- [66] Mijail D Serruya, Nicholas G Hatsopoulos, Liam Paninski, Matthew R Fellows, and John P Donoghue. Instant neural control of a movement signal. *Nature*, 416(6877):141–142, 2002.
- [67] B Wodlinger, JE Downey, EC Tyler-Kabara, AB Schwartz, ML Boninger, and JL Collinger. Ten-dimensional anthropomorphic arm control in a human brain– machine interface: difficulties, solutions, and limitations. *Journal of neural engineering*, 12(1):016011, 2014.
- [68] Aya Rezeika, Mihaly Benda, Piotr Stawicki, Felix Gemblar, Abdul Saboor, and Ivan Volosyak. Brain–computer interface spellers: A review. *Brain sciences*, 8(4):57, 2018.
- [69] Reza Abiri, Soheil Borhani, Eric W Sellers, Yang Jiang, and Xiaopeng Zhao. A comprehensive review of EEG-based brain–computer interface paradigms. *Journal of neural engineering*, 16(1):011001, 2019.
- [70] Monica Fabiani, Gabriele Gratton, Demetrios Karis, Emanuel Donchin, et al. Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. *Advances in psychophysiology*, 2(S 1):78, 1987.
- [71] John Polich. Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, 118(10):2128–2148, 2007.
- [72] Pietro Cipresso, Laura Carelli, Federica Solca, Daniela Meazzi, Paolo Meriggi, Barbara Poletti, Dorothée Lulé, Albert C Ludolph, Vincenzo Silani, and Giuseppe Riva. The use of P300-based BCIs in amyotrophic lateral sclerosis: from augmentative and alternative communication to cognitive assessment. *Brain and behavior*, 2(4):479–498, 2012.
- [73] Lawrence Ashley Farwell and Emanuel Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, 70(6):510–523, 1988.

- [74] Theresa M Vaughan, Jonathan R Wolpaw, and Emanuel Donchin. EEG-based communication: prospects and problems. *IEEE transactions on rehabilitation engineering*, 4(4):425–430, 1996.
- [75] Reza Fazel-Rezai, Brendan Z Allison, Christoph Guger, Eric W Sellers, Sonja C Kleih, and Andrea Kübler. P300 brain computer interface: current challenges and emerging trends. *Frontiers in neuroengineering*, 5:14, 2012.
- [76] Lynn M McCane, Eric W Sellers, Dennis J McFarland, Joseph N Mak, C Steve Carmack, Debra Zeitlin, Jonathan R Wolpaw, and Theresa M Vaughan. Brain-computer interface (BCI) evaluation in people with amyotrophic lateral sclerosis. *Amyotrophic lateral sclerosis and frontotemporal degeneration*, 15(3-4):207–215, 2014.
- [77] Jinhu Xiong, Liangsuo Ma, Binqun Wang, Shalini Narayana, Eugene P Duff, Gary F Egan, and Peter T Fox. Long-term motor training induced changes in regional cerebral blood flow in both task and resting states. *Neuroimage*, 45(1):75–82, 2009.
- [78] EUGENE V Golanov, SEIJI Yamamoto, and DONALD J Reis. Spontaneous waves of cerebral blood flow associated with a pattern of electrocortical activity. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 266(1):R204–R214, 1994.
- [79] Dante Mantini, Mauro G Perrucci, Cosimo Del Gratta, Gian L Romani, and Maurizio Corbetta. Electrophysiological signatures of resting state networks in the human brain. *Proceedings of the National Academy of Sciences*, 104(32):13170–13175, 2007.
- [80] CJ Stam, T Montez, BF Jones, SARB Rombouts, Y Van Der Made, YAL Pijnenburg, and Ph Scheltens. Disturbed fluctuations of resting state EEG synchronization in Alzheimer’s disease. *Clinical neurophysiology*, 116(3):708–715, 2005.
- [81] Peter Putman. Resting state EEG delta–beta coherence in relation to anxiety, behavioral inhibition, and selective attentional processing of threatening stimuli. *International journal of psychophysiology*, 80(1):63–68, 2011.
- [82] Jun Wang, Jamie Barstein, Lauren E Ethridge, Matthew W Mosconi, Yukari Takarae, and John A Sweeney. Resting state EEG abnormalities in autism spectrum disorders. *Journal of neurodevelopmental disorders*, 5(1):24, 2013.
- [83] Lin Gao, Wei Cheng, Jinhua Zhang, and Jue Wang. EEG classification for motor imagery and resting state in BCI applications using multi-class Adaboost extreme learning machine. *Review of Scientific Instruments*, 87(8):085110, 2016.
- [84] Rui Zhang, Dezhong Yao, Pedro A Valdés-Sosa, Fali Li, Peiyang Li, Tao Zhang, Teng Ma, Yongjie Li, and Peng Xu. Efficient resting-state EEG network facilitates motor imagery performance. *Journal of neural engineering*, 12(6):066024, 2015.
- [85] Yang Di, Xingwei An, Feng He, Shuang Liu, Yufeng Ke, and Dong Ming. Robustness Analysis of Identification Using Resting-State EEG Signals. *IEEE Access*, 7:42113–42122, 2019.
- [86] Luis Alfredo Moctezuma and Marta Molinas. Sex differences observed in a study of EEG of linguistic activity and resting-state: Exploring optimal EEG channel configurations. In *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, pages 1–6. IEEE, 2019.
- [87] Luis Alfredo Moctezuma and Marta Molinas. Towards a minimal EEG channel array for a biometric system using resting-state and a genetic algorithm for channel selection. *Scientific Reports*, 10(1):1–14, 2020.

-
- [88] Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.
- [89] Dr Lehmann, H Ozaki, and I Pal. EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. *Electroencephalography and clinical neurophysiology*, 67(3):271–288, 1987.
- [90] Arjun Khanna, Alvaro Pascual-Leone, Christoph M Michel, and Faranak Farzan. Microstates in resting-state EEG: current status and future directions. *Neuroscience & Biobehavioral Reviews*, 49:105–113, 2015.
- [91] Michael D Greicius, Ben Krasnow, Allan L Reiss, and Vinod Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100(1):253–258, 2003.
- [92] Thomas Koenig, Leslie Prichep, Dietrich Lehmann, Pedro Valdes Sosa, Elisabeth Braeker, Horst Kleinlogel, Robert Isenhardt, and E Roy John. Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. *Neuroimage*, 16(1):41–48, 2002.
- [93] Dietrich Lehmann, Roberto D Pascual-Marqui, and Christoph Michel. EEG microstates. *Scholarpedia*, 4(3):7632, 2009.
- [94] Anna Custo, Dimitri Van De Ville, William M Wells, Miralena I Tomescu, Denis Brunet, and Christoph M Michel. Electroencephalographic resting-state networks: source localization of microstates. *Brain connectivity*, 7(10):671–682, 2017.
- [95] Christoph M Michel and Thomas Koenig. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *Neuroimage*, 180:577–593, 2018.
- [96] Saam Iranmanesh and Esther Rodriguez-Villegas. A 950 nW analog-based data reduction chip for wearable EEG systems in epilepsy. *IEEE Journal of Solid-State Circuits*, 52(9):2362–2373, 2017.
- [97] M Rajya Lakshmi, TV Prasad, and Dr V Chandra Prakash. Survey on EEG signal processing methods. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(1), 2014.
- [98] Mamunur Rashid, Norizam Sulaiman, Anwar PP Abdul Majeed, Rabiun Muazu Musa, Ahmad Fakhri Ab Nasir, Bifta Sama Bari, and Sabira Khatun. Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review. *Frontiers in Neuroinformatics*, 2020.
- [99] Jesus Minguillon, M Angel Lopez-Gordo, and Francisco Pelayo. Trends in EEG-BCI for daily-life: Requirements for artifact removal. *Biomedical Signal Processing and Control*, 31:407–418, 2017.
- [100] Stefan Debener, Cornelia Kranczioch, and Maarten De Vos. Electroencephalography: Current Trends and Future Directions. In *Neuroeconomics*, pages 359–373. Springer, 2016.
- [101] Mamunur Rashid, Norizam Sulaiman, Mahfuzah Mustafa, Sabira Khatun, Bifta Sama Bari, and Md Jahid Hasan. Recent Trends and Open Challenges in EEG Based Brain-Computer Interface Systems. In *InECCE2019*, pages 367–378. Springer, 2020.
- [102] David Looney, Preben Kidmose, Cheolsoo Park, Michael Ungstrup, Mike Lind Rank, Karin Rosenkranz, and Danilo P Mandic. The in-the-ear recording concept: User-centered and

- wearable brain monitoring. *IEEE pulse*, 3(6):32–42, 2012.
- [103] Martin G Bleichner and Stefan Debener. Concealed, unobtrusive ear-centered EEG acquisition: cEEGrids for transparent EEG. *Frontiers in human neuroscience*, 11:163, 2017.
- [104] Alexander J Casson, Shelagh Smith, John S Duncan, and Esther Rodriguez-Villegas. Wearable EEG: what is it, why is it needed and what does it entail? In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5867–5870. IEEE, 2008.
- [105] Alexander J Casson, David C Yates, Shelagh JM Smith, John S Duncan, and Esther Rodriguez-Villegas. Wearable electroencephalography. *IEEE engineering in medicine and biology magazine*, 29(3):44–56, 2010.
- [106] Michal Teplan et al. Fundamentals of EEG measurement. *Measurement science review*, 2(2):1–11, 2002.
- [107] Rodney J Croft and Robert J Barry. Removal of ocular artifact from the EEG: a review. *Neurophysiologie Clinique/Clinical Neurophysiology*, 30(1):5–19, 2000.
- [108] Chi Qin Lai, Haidi Ibrahim, Mohd Zaid Abdullah, Jafri Malin Abdullah, Shahrel Azmin Suandi, and Azlinda Azman. Artifacts and noise removal for electroencephalogram (EEG): A literature review. In *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pages 326–332. IEEE, 2018.
- [109] Xiao Jiang, Gui-Bin Bian, and Zean Tian. Removal of artifacts from EEG signals: a review. *Sensors*, 19(5):987, 2019.
- [110] Jun Lu, Dennis J McFarland, and Jonathan R Wolpaw. Adaptive Laplacian filtering for sensorimotor rhythm-based brain–computer interfaces. *Journal of neural engineering*, 10(1):016002, 2012.
- [111] Kai Keng Ang, Juanhong Yu, and Cuntai Guan. Extracting effective features from high density nirs-based BCI for assessing numerical cognition. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2233–2236. IEEE, 2012.
- [112] Syahrull Hi Fi Syam, Heba Lakany, RB Ahmad, and Bernard A Conway. Comparing common average referencing to laplacian referencing in detecting imagination and intention of movement for brain computer interface. In *MATEC Web of Conferences*, volume 140, 2017.
- [113] Yash Paul. Various epileptic seizure detection techniques using biomedical signals: a review. *Brain informatics*, 5(2):6, 2018.
- [114] Yizhang Jiang, Dongrui Wu, Zhaohong Deng, Pengjiang Qian, Jun Wang, Guanjin Wang, Fu-Lai Chung, Kup-Sze Choi, and Shitong Wang. Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(12):2270–2284, 2017.
- [115] Sanjeev Kumar Dhull, Krishna Kant Singh, et al. A Review on Automatic Epilepsy Detection from EEG Signals. In *Advances in Communication and Computational Technology*, pages 1441–1454. Springer, 2021.
- [116] 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.
- [117] Norden E Huang. *Hilbert-Huang transform and its applications*, volume 16. World Scientific, 2014.
- [118] Norden E Huang, Man-Li C Wu, Steven R Long, Samuel SP Shen, Wendong Qu, Per Gloersen, and Kuang L Fan. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 459(2037):2317–2345, 2003.
- [119] ZHAO Jin-Ping and Huang Da-ji. Mirror extending and circular spline function for empirical mode decomposition method. *Journal of Zhejiang University-Science A*, 2(3):247–252, 2001.
- [120] Liu Zhengkun and Zhang Ze. The improved algorithm of the EMD endpoint effect based on the mirror continuation. In *2016 Eighth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pages 792–795. IEEE, 2016.
- [121] LV Chenhuan, ZHAO Jun, WU Chao, GUO Tiantai, and CHEN Hongjiang. Optimization of the end effect of Hilbert-Huang transform (HHT). *Chinese Journal of Mechanical Engineering*, 30(3):732–745, 2017.
- [122] Jian Wang, Wenyuan Liu, and Shuai Zhang. An approach to eliminating end effects of EMD through mirror extension coupled with support vector machine method. *Personal and Ubiquitous Computing*, 23(3-4):443–452, 2019.
- [123] Yunchao Gao, Guangtao Ge, Zhengyan Sheng, and Enfang Sang. Analysis and solution to the mode mixing phenomenon in EMD. In *2008 Congress on Image and Signal Processing*, volume 5, pages 223–227. IEEE, 2008.
- [124] Zhaohua Wu and Norden E Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, 1(01):1–41, 2009.
- [125] J. Jebaraj and R. Arumugam. Ensemble empirical mode decomposition-based optimised power line interference removal algorithm for electrocardiogram signal. *IET Signal Processing*, 10(6):583–591, 2016.
- [126] Gabriel Rilling, Patrick Flandrin, Paulo Goncalves, et al. On empirical mode decomposition and its algorithms. In *IEEE-EURASIP workshop on nonlinear signal and image processing*, volume 3, pages 8–11. NSIP-03, Grado (I), 2003.
- [127] Douglas Baptista de Souza, Jocelyn Chanussot, and Anne-Catherine Favre. On selecting relevant intrinsic mode functions in empirical mode decomposition: An energy-based approach. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 325–329. IEEE, 2014.
- [128] Daoud Boutana, Messaoud Benidir, and Braham Barkat. On the selection of intrinsic mode function in EMD method: application on heart sound signal. In *2010 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2010)*, pages 1–5. IEEE, 2010.
- [129] Albert Ayenu-Prah and Nii Attoh-Okine. A criterion for selecting relevant intrinsic mode functions in empirical mode decomposition. *Advances in Adaptive Data Analysis*, 2(01):1–24, 2010.

- [130] Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 1989.
- [131] HM Teager and SM Teager. Evidence for nonlinear sound production mechanisms in the vocal tract. In *Speech production and speech modelling*, pages 241–261. Springer, 1990.
- [132] Firas Jabloun and A Enis Cetin. The Teager energy based feature parameters for robust speech recognition in car noise. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)*, volume 1, pages 273–276. IEEE, 1999.
- [133] Emmanuel Didiot, Irina Illina, Dominique Fohr, and Odile Mella. A wavelet-based parameterization for speech/music discrimination. *Computer Speech & Language*, 24(2):341–357, 2010.
- [134] Truong Quang Dang Khoa, Vo Quang Ha, and Vo Van Toi. Higuchi fractal properties of onset epilepsy electroencephalogram. *Computational and mathematical methods in medicine*, 2012, 2012.
- [135] Luis Alfredo Moctezuma and Marta Molinas. Classification of low-density EEG epileptic seizures by energy and fractal features based on EMD. *Journal of Biomedical Research*, 2019.
- [136] Benoit B Mandelbrot. Self-affine fractals and fractal dimension. *Physica scripta*, 32(4):257, 1985.
- [137] Włodzimierz Klonowski. Fractal Analysis of Electroencephalographic Time Series (EEG Signals). In *The Fractal Geometry of the Brain*, pages 413–429. Springer, 2016.
- [138] Luis Alfredo Moctezuma and Marta Molinas. Multi-objective optimization for eeG channel selection and accurate intruder detection in an eeG-based subject identification system. *Scientific Reports*, 10(1):1–12, 2020.
- [139] Agostino Accardo, M Affinito, M Carrozzì, and F Bouquet. Use of the fractal dimension for the analysis of electroencephalographic time series. *Biological cybernetics*, 77(5):339–350, 1997.
- [140] Werner Lutzenberger, Hubert Preissl, and Friedemann Pulvermüller. Fractal dimension of electroencephalographic time series and underlying brain processes. *Biological Cybernetics*, 73(5):477–482, 1995.
- [141] Karolina Lebiecka, Urszula Zuchowicz, Agata Wozniak-Kwasniewska, David Szekely, Elzbieta Olejarczyk, and Olivier David. Complexity analysis of EEG data in persons with depression subjected to transcranial magnetic stimulation. *Frontiers in physiology*, 9:1385, 2018.
- [142] Tomoyuki Higuchi. Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, 31(2):277–283, 1988.
- [143] Carlos Gómez, Ángela Mediavilla, Roberto Hornero, Daniel Abásolo, and Alberto Fernández. Use of the Higuchi’s fractal dimension for the analysis of MEG recordings from Alzheimer’s disease patients. *Medical engineering & physics*, 31(3):306–313, 2009.
- [144] Elisabeth Ruiz-Padial and Antonio J Ibáñez-Molina. Fractal dimension of EEG signals and heart dynamics in discrete emotional states. *Biological psychology*, 137:42–48, 2018.
- [145] Sladana Spasic, Aleksandar Kalauzi, G Grbic, Ljiljana Martac, and Milka Culic. Fractal

- analysis of rat brain activity after injury. *Medical and Biological Engineering and Computing*, 43(3):345–348, 2005.
- [146] Arthur Petrosian. Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns. In *Proceedings Eighth IEEE Symposium on Computer-Based Medical Systems*, pages 212–217. IEEE, 1995.
- [147] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.
- [148] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, 15(3):031005, 2018.
- [149] Meysam Golmohammadi, Amir Hossein Harati Nejad Torbati, Silvia Lopez de Diego, Iyad Obeid, and Joseph Picone. Automatic analysis of EEGs using big data and hybrid deep learning architectures. *Frontiers in human neuroscience*, 13:76, 2019.
- [150] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, 16(5):051001, 2019.
- [151] Gen Li, Chang Ha Lee, Jason J Jung, Young Chul Youn, and David Camacho. Deep learning for EEG data analytics: A survey. *Concurrency and Computation: Practice and Experience*, 32(18):e5199, 2020.
- [152] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, 3(3):1–13, 2007.
- [153] Faraz Akram, Seung Moo Han, and Tae-Seong Kim. An efficient word typing P300-BCI system using a modified T9 interface and random forest classifier. *Computers in biology and medicine*, 56:30–36, 2015.
- [154] David Steyrl, Reinhold Scherer, Josef Faller, and Gernot R Müller-Putz. Random forests in non-invasive sensorimotor rhythm brain-computer interfaces: a practical and convenient non-linear classifier. *Biomedical Engineering/Biomedizinische Technik*, 61(1):77–86, 2016.
- [155] Chongsheng Zhang, Changchang Liu, Xiangliang Zhang, and George Almpanidis. An up-to-date comparison of state-of-the-art classification algorithms. *Expert Systems with Applications*, 82:128–150, 2017.
- [156] Stuart J Russell and Peter Norvig. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.
- [157] Thorsten Joachims. Making large-scale svm learning practical. Technical Report 1998,28, Universität Dortmund, <http://hdl.handle.net/10419/77178>, 1998.
- [158] Abdiansah Abdiansah and Retantyo Wardoyo. Time complexity analysis of support vector machines (SVM) in LibSVM. *International journal computer and application*, 2015.
- [159] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.
- [160] Keinosuke Fukunaga and Patrenahalli M. Narendra. A branch and bound algorithm for computing k-nearest neighbors. *IEEE transactions on computers*, 100(7):750–753, 1975.

- [161] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.
- [162] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [163] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomForest. *R news*, 2(3):18–22, 2002.
- [164] Mia Stern, Joseph Beck, and Beverly Park Woolf. Naive Bayes classifiers for user modeling. *Center for Knowledge Communication, Computer Science Department, University of Massachusetts*, 1999.
- [165] David Martinus Johannes Tax. *One-class classification: Concept learning in the absence of counter-examples*. PhD thesis, Delft University of Technology, 2002.
- [166] Iwan Syarif, Adam Prugel-Bennett, and Gary Wills. SVM parameter optimization using grid search and genetic algorithm to improve classification performance. *Telkomnika*, 14(4):1502, 2016.
- [167] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. LOF: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104, 2000.
- [168] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.
- [169] Max Kuhn, Kjell Johnson, et al. *Applied predictive modeling*, volume 26. Springer, 2013.
- [170] Claude Sammut and Geoffrey I Webb. *Encyclopedia of machine learning*. Springer Science & Business Media, 2011.
- [171] 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, 2015.
- [172] Muhammad Zeeshan Baig, Nauman Aslam, and Hubert PH Shum. Filtering techniques for channel selection in motor imagery EEG applications: a survey. *Artificial intelligence review*, 53(2):1207–1232, 2020.
- [173] Luis Alfredo Moctezuma and Marta Molinas. Subject identification from low-density EEG-recordings of resting-states: A study of feature extraction and classification. In *Future of Information and Communication Conference*, pages 830–846. Springer, 2019.
- [174] Yanru Bai, Zhiguo Zhang, and Dong Ming. Feature selection and channel optimization for biometric identification based on visual evoked potentials. In *2014 19th International Conference on Digital Signal Processing*, pages 772–776. IEEE, 2014.
- [175] Ying Wang, Xi Long, Hans van Dijk, Ronald Aarts, and Johan Arends. Adaptive EEG channel selection for nonconvulsive seizure analysis. In *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, pages 1–5. IEEE, 2018.
- [176] Tao Yang, Kai Keng Ang, Kok Soon Phua, Juanhong Yu, Valerie Toh, Wai Hoe Ng, and Rosa Q So. Eeg channel selection based on correlation coefficient for motor imagery classification: A study on healthy subjects and als patient. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1996–1999. IEEE, 2018.
- [177] Mustafa Turan Arslan, Server Göksel Eraldemir, and Esen Yildirim. Channel selection from

- EEG signals and application of support vector machine on EEG data. In *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, pages 1–4. IEEE, 2017.
- [178] Huijuan Yang, Cuntai Guan, Chuan Chu Wang, and Kai Keng Ang. Maximum dependency and minimum redundancy-based channel selection for motor imagery of walking EEG signal detection. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1187–1191. IEEE, 2013.
- [179] Huijuan Yang, Cuntai Guan, Kai Keng Ang, Kok Soon Phua, and Chuanchu Wang. Selection of effective EEG channels in brain computer interfaces based on inconsistencies of classifiers. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 672–675. IEEE, 2014.
- [180] Karim Ansari-Asl, Guillaume Chanel, and Thierry Pun. A channel selection method for EEG classification in emotion assessment based on synchronization likelihood. In *2007 15th European Signal Processing Conference*, pages 1241–1245. IEEE, 2007.
- [181] Yongkoo Park and Wonzoong Chung. Optimal Channel Selection Using Correlation Coefficient for CSP Based EEG Classification. *IEEE Access*, 8:111514–111521, 2020.
- [182] Zhong-Min Wang, Shu-Yuan Hu, and Hui Song. Channel selection method for eeg emotion recognition using normalized mutual information. *IEEE Access*, 7:143303–143311, 2019.
- [183] Michael Schröder, Thomas Navin Lal, Thilo Hinterberger, Martin Bogdan, N Jeremy Hill, Niels Birbaumer, Wolfgang Rosenstiel, and Bernhard Schölkopf. Robust EEG channel selection across subjects for brain-computer interfaces. *EURASIP Journal on Advances in Signal Processing*, 2005(19):174746, 2005.
- [184] Fatma Ibrahim, Saly Abd-Elateif El-Gindy, Sami M El-Dolil, Adel S El-Fishawy, El-Sayed M El-Rabaie, Moawaed I Dessouky, Ibrahim M Eldokany, Turkey N Alotaiby, Saleh A Alshebeili, and Fathi E Abd El-Samie. A statistical framework for EEG channel selection and seizure prediction on mobile. *International Journal of Speech Technology*, 22(1):191–203, 2019.
- [185] Jonas Duun-Henriksen, Troels Wesenberg Kjaer, Rasmus Elsborg Madsen, Line Sofie Remvig, Carsten Eckhart Thomsen, and Helge Bjarup Dissing Sorensen. Channel selection for automatic seizure detection. *Clinical Neurophysiology*, 123(1):84–92, 2012.
- [186] Jianhai Zhang, Ming Chen, Shaokai Zhao, Sanqing Hu, Zhiguo Shi, and Yu Cao. ReliefF-based EEG sensor selection methods for emotion recognition. *Sensors*, 16(10):1558, 2016.
- [187] M Murugappan and Sazali Yaacob. Asymmetric ratio and FCM based salient channel selection for human emotion detection using EEG. *WSEAS Transactions on Signal Processing*, 2008.
- [188] Yi-Hung Liu, Shiuan Huang, and Yi-De Huang. Motor imagery EEG classification for patients with amyotrophic lateral sclerosis using fractal dimension and Fisher’s criterion-based channel selection. *Sensors*, 17(7):1557, 2017.
- [189] Ahmed Al-Ani and Mostefa Mesbah. EEG rhythm/channel selection for fuzzy rule-based alertness state characterization. *Neural Computing and Applications*, 30(7):2257–2267, 2018.
- [190] Annushree Bablani, Damodar Reddy Edla, Diwakar Tripathi, Shubham Dodia, and Sridhar Chintala. A synergistic concealed information test with novel approach for EEG channel selection and SVM parameter optimization. *IEEE Transactions on Information Forensics and Security*, 14(11):3057–3068, 2019.

- [191] Jianhua Yang, Harsimrat Singh, Evor L Hines, Friederike Schlaghecken, Daciana D Iliescu, Mark S Leeson, and Nigel G Stocks. Channel selection and classification of electroencephalogram signals: an artificial neural network and genetic algorithm-based approach. *Artificial intelligence in medicine*, 55(2):117–126, 2012.
- [192] Mahnaz Arvaneh, Cuntai Guan, Kai Keng Ang, and Chai Quek. Optimizing the channel selection and classification accuracy in EEG-based BCI. *IEEE Transactions on Biomedical Engineering*, 58(6):1865–1873, 2011.
- [193] Ahmed Al-Ani and Akram Al-Sukker. Effect of feature and channel selection on EEG classification. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2171–2174. IEEE, 2006.
- [194] Beatriz A Garro, Rocio Salazar-Varas, and Roberto A Vazquez. EEG Channel Selection using Fractal Dimension and Artificial Bee Colony Algorithm. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 499–504. IEEE, 2018.
- [195] Vikram Shenoy Handiru and Vinod A Prasad. Optimized bi-objective eeg channel selection and cross-subject generalization with brain–computer interfaces. *IEEE Transactions on Human-Machine Systems*, 46(6):777–786, 2016.
- [196] Hao Sun, Jing Jin, Wanzeng Kong, Cili Zuo, Shurui Li, and Xingyu Wang. Novel channel selection method based on position priori weighted permutation entropy and binary gravity search algorithm. *Cognitive Neurodynamics*, pages 1–16, 2020.
- [197] Alejandro A Torres-García, Carlos A Reyes-García, Luis Villaseñor-Pineda, and Gregorio García-Aguilar. Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification. *Expert Systems with Applications*, 59:1–12, 2016.
- [198] Lin He, Youpan Hu, Yuanqing Li, and Daoli Li. Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG. *Neurocomputing*, 121:423–433, 2013.
- [199] Chea-Yau Kee, Sivalinga Govinda Ponnambalam, and Chu-Kiong Loo. Multi-objective genetic algorithm as channel selection method for P300 and motor imagery data set. *Neurocomputing*, 161:120–131, 2015.
- [200] Luis Alfredo Moctezuma and Marta Molinas. EEG Channel-selection method for epileptic-seizure classification based on multi-objective optimization. *Frontiers in Neuroscience*, 14:593, 2020.
- [201] Douglas Rodrigues, Gabriel FA Silva, João P Papa, Aparecido N Marana, and Xin-She Yang. EEG-based person identification through binary flower pollination algorithm. *Expert Systems with Applications*, 62:81–90, 2016.
- [202] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2009.
- [203] Patrenahalli M. Narendra and Keinosuke Fukunaga. A branch and bound algorithm for feature subset selection. *IEEE Transactions on computers*, pages 917–922, 1977.
- [204] Iman Foroutan and Jack Sklansky. Feature selection for automatic classification of non-gaussian data. *IEEE Transactions on Systems, Man, and Cybernetics*, 17(2):187–198, 1987.

-
- [205] Jihoon Yang and Vasant Honavar. Feature subset selection using a genetic algorithm. In *Feature extraction, construction and selection*, pages 117–136. Springer, 1998.
- [206] Luis Alfredo Moctezuma and Marta Molinas. Event-related potential from eeg for a two-step identity authentication system. In *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, volume 1, pages 392–399. IEEE, 2019.
- [207] Kalyanmoy Deb. *Multi-objective optimization using evolutionary algorithms*, volume 16. John Wiley & Sons, 2001.
- [208] Tinkle Chugh, Karthik Sindhya, Jussi Hakanen, and Kaisa Miettinen. A survey on handling computationally expensive multiobjective optimization problems with evolutionary algorithms. *Soft Computing*, 23(9):3137–3166, 2019.
- [209] Oliver Kramer. *Genetic algorithm essentials*, volume 679. Springer, 2017.
- [210] Nidamarthi Srinivas and Kalyanmoy Deb. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, 2(3):221–248, 1994.
- [211] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.
- [212] Kalyanmoy Deb and Himanshu Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601, 2013.
- [213] Himanshu Jain and Kalyanmoy Deb. An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: handling constraints and extending to an adaptive approach. *IEEE Transactions on Evolutionary Computation*, 18(4):602–622, 2013.
- [214] Indraneel Das and John E Dennis. Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM journal on optimization*, 8(3):631–657, 1998.
- [215] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.
- [216] António Dourado, M Le Van Quyen, B Schelter, G Favaro, A Schulze-Bonhage, S Sales, and V Navarro. EPILEPSIAE-EVOLVING PLATFORM FOR IMPROVING LIVING EXPECTATION OF PATIENTS SUFFERING FROM ICTAL EVENTS: E595. *Epilepsia*, 50:210–211, 2009.
- [217] Iyad Obeid and Joseph Picone. The temple university hospital EEG data corpus. *Frontiers in neuroscience*, 10:196, 2016.
- [218] Ali Hossam Shoeb. *Application of machine learning to epileptic seizure onset detection and treatment*. PhD thesis, Massachusetts Institute of Technology, 2009.
- [219] Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw. BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043, 2004.
- [220] Perrin Margaux, Maby Emmanuel, Daligault Sébastien, Bertrand Olivier, and Mattout Jérémie.

- Objective and subjective evaluation of online error correction during P300-based spelling. *Advances in Human-Computer Interaction*, 2012:4, 2012.
- [221] Luis Alfredo Moctezuma. Distinción de estados de actividad e inactividad lingüística para interfaces cerebro computadora. Master's thesis, Benemérita Universidad Autónoma de Puebla, 2017.
- [222] Luis Alfredo Moctezuma, Alejandro A Torres-García, Luis Villaseñor-Pineda, and Maya Carrillo. Subjects identification using EEG-recorded imagined speech. *Expert Systems with Applications*, 118:201–208, 2019.
- [223] Luis Alfredo Moctezuma and Marta Molinas. EEG-based Subjects Identification based on Biometrics of Imagined Speech using EMD. In *International Conference on Brain Informatics*, pages 458–467. Springer, 2018.
- [224] Petre Lameski, Eftim Zdravevski, Riste Mingov, and Andrea Kulakov. SVM parameter tuning with grid search and its impact on reduction of model over-fitting. In *Rough sets, fuzzy sets, data mining, and granular computing*, pages 464–474. Springer, 2015.
- [225] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.
- [226] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [227] Julian Blank and Kalyanmoy Deb. pymoo: Multi-objective Optimization in Python. *IEEE Access*, 8:89497–89509, 2020.
- [228] Matthew Rocklin. Dask: Parallel computation with blocked algorithms and task scheduling. In *Proceedings of the 14th python in science conference*, pages 130–136. Citeseer, 2015.
- [229] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, António H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.
- [230] Charles R. Harris, K. Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585:357–362, 2020.
- [231] Gregory R Lee, Ralf Gommers, Filip Waselewski, Kai Wohlfahrt, and Aaron O'Leary. PyWavelets: A Python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, 2019.

-
- [232] Jaidev Deshpande. pyhht Documentation. <https://pyhht.readthedocs.io/en/latest/>, 2018. Accessed: 2021-01-01.
- [233] Magnus Sjölander, Magnus Jahre, Gunnar Tufte, and Nico Reissmann. EPIC: An energy-efficient, high-performance GPGPU computing research infrastructure. *arXiv preprint arXiv:1912.05848*, 2019.
- [234] Florian Mormann, Ralph G Andrzejak, Christian E Elger, and Klaus Lehnertz. Seizure prediction: the long and winding road. *Brain*, 130(2):314–333, 2006.
- [235] Rajendra Kale. Bringing epilepsy out of the shadows: Wide treatment gap needs to be reduced, 1997.
- [236] Jr J Engel. A practical guide for routine EEG studies in epilepsy. *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, 1(2):109–142, 1984.
- [237] Hojjat Adeli and Samanwoy Ghosh-Dastidar. *Automated EEG-based diagnosis of neurological disorders: Inventing the future of neurology*. CRC press, 2010.
- [238] Orrin Devinsky. Diagnosis and treatment of temporal lobe epilepsy. *Rev Neurol Dis*, 1(1):2–9, 2004.
- [239] Jerome Engel Jr. Mesial temporal lobe epilepsy: what have we learned? *The neuroscientist*, 7(4):340–352, 2001.
- [240] Vairavan Srinivasan, Chikkannan Eswaran, and N. Sriraam. Artificial neural network based epileptic detection using time-domain and frequency-domain features. *Journal of Medical Systems*, 29(6):647–660, 2005.
- [241] Yatindra Kumar, ML Dewal, and RS Anand. Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine. *Neurocomputing*, 133:271–279, 2014.
- [242] Yusuf Uzzaman Khan, Nidal Rafiuddin, and Omar Farooq. Automated seizure detection in scalp EEG using multiple wavelet scales. In *2012 IEEE International Conference on Signal Processing, Computing and Control*, pages 1–5. IEEE, 2012.
- [243] Morteza Zabihi, Serkan Kiranyaz, Ali Bahrami Rad, Aggelos K Katsaggelos, Moncef Gabbouj, and Turker Ince. Analysis of high-dimensional phase space via Poincaré section for patient-specific seizure detection. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(3):386–398, 2015.
- [244] Muhammad Sohaib J Solaija, Sajid Saleem, Khawar Khurshid, Syed Ali Hassan, and Awais Mehmood Kamboh. Dynamic mode decomposition based epileptic seizure detection from scalp EEG. *IEEE Access*, 6:38683–38692, 2018.
- [245] Abhijit Bhattacharyya and Ram Bilas Pachori. A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform. *IEEE Transactions on Biomedical Engineering*, 64(9):2003–2015, 2017.
- [246] Yinda Zhang, Shuhan Yang, Yang Liu, Yexian Zhang, Bingfeng Han, and Fengfeng Zhou. Integration of 24 feature types to accurately detect and predict seizures using scalp EEG Signals. *Sensors*, 18(5):1372, 2018.
- [247] U Rajendra Acharya, Filippo Molinari, S Vinitha Sree, Subhagata Chattopadhyay, Kwan-Hoong Ng, and Jasjit S Suri. Automated diagnosis of epileptic EEG using entropies. *Biomedical*

- Signal Processing and Control*, 7(4):401–408, 2012.
- [248] Rajeev Sharma and Ram Bilas Pachori. Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. *Expert Systems with Applications*, 42(3):1106–1117, 2015.
- [249] Vipin Gupta and Ram Bilas Pachori. Epileptic seizure identification using entropy of FBSE based EEG rhythms. *Biomedical Signal Processing and Control*, 53:101569, 2019.
- [250] Vipin Gupta, Abhijit Bhattacharyya, and Ram Bilas Pachori. Automated identification of epileptic seizures from EEG signals using FBSE-EWT method. In *Biomedical Signal Processing*, pages 157–179. Springer, 2020.
- [251] José Antonio de la O Serna, Mario R Arrieta Paternina, Alejandro Zamora-Méndez, Rajesh Kumar Tripathy, and Ram Bilas Pachori. EEG-Rhythm Specific Taylor-Fourier filter bank Implemented with O-splines for the Detection of Epilepsy using EEG Signals. *IEEE Sensors Journal*, 2020.
- [252] Rahul Sharma, Ram Bilas Pachori, and Pradip Sircar. Seizures classification based on higher order statistics and deep neural network. *Biomedical Signal Processing and Control*, 59:101921, 2020.
- [253] Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E Elger. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6):061907, 2001.
- [254] P Fiedler, P Pedrosa, S Griebel, C Fonseca, F Vaz, E Supriyanto, F Zanow, and J Haueisen. Novel multipin electrode cap system for dry electroencephalography. *Brain topography*, 28(5):647–656, 2015.
- [255] Selenia di Fronso, Patrique Fiedler, Gabriella Tamburro, Jens Haueisen, Maurizio Bertollo, and Silvia Comani. Dry EEG in sport sciences: a fast and reliable tool to assess individual alpha peak frequency changes induced by physical effort. *Frontiers in Neuroscience*, 13:982, 2019.
- [256] Nidal Rafiuddin, Yusuf Uzzaman Khan, and Omar Farooq. Feature extraction and classification of EEG for automatic seizure detection. In *2011 International Conference on Multimedia, Signal Processing and Communication Technologies*, pages 184–187. IEEE, 2011.
- [257] Vairavan Srinivasan, Chikkannan Eswaran, and Natarajan Sriraam. Approximate entropy-based epileptic EEG detection using artificial neural networks. *IEEE Transactions on information Technology in Biomedicine*, 11(3):288–295, 2007.
- [258] Abdulhamit Subasi and M Ismail Gursoy. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert systems with applications*, 37(12):8659–8666, 2010.
- [259] CHRYSOTOMOS P Panayiotopoulos and MICHALIS Koutroumanidis. The significance of the syndromic diagnosis of the epilepsies. *National Society for Epilepsy*, 2005.
- [260] Yong Won Cho and Keun Tae Kim. The Latest Classification of Epilepsy and Clinical Significance of Electroencephalography. *Journal of Neurointensive Care*, 2(1):1–3, 2019.
- [261] Ena Bingham and Victor Patterson. A telemedicine-enabled nurse-led epilepsy service is acceptable and sustainable. *Journal of Telemedicine and Telecare*, 13(3_suppl):19–21, 2007.

- [262] Phil Smith. Telephone review for people with epilepsy. *Practical neurology*, 16(6):475–477, 2016.
- [263] Najib Kissani, Yilédoma Thierry Modeste Lengané, Victor Patterson, Boulenouar Mesraoua, Eliashiv Dawn, Cigdem Ozkara, Graeme Shears, Harmiena Riphagen, Ali A Asadi-Pooya, Alicia Bogacz, et al. Telemedicine in epilepsy: How can we improve care, teaching, and awareness? *Epilepsy & Behavior*, page 106854, 2020.
- [264] Carmen Terranova, Vincenzo Rizzo, Alberto Cacciola, Gaetana Chillemi, Alessandro Calamuneri, Demetrio Milardi, and Angelo Quartarone. Is there a future for non-invasive brain stimulation as a therapeutic tool? *Frontiers in neurology*, 9:1146, 2019.
- [265] Siu Kwan Lam, Antoine Pitrou, and Stanley Seibert. Numba: A llvm-based python jit compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, pages 1–6, 2015.
- [266] Emmanuel K Kalunga, Sylvain Chevallier, Quentin Barthélemy, Karim Djouani, Eric Monacelli, and Yskandar Hamam. Online SSVEP-based BCI using Riemannian geometry. *Neurocomputing*, 191:55–68, 2016.
- [267] Robin Tibor Schirrmester, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human brain mapping*, 38(11):5391–5420, 2017.
- [268] Anil K Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1):4–20, 2004.
- [269] Anil K Jain, Arun Ross, and Umut Uludag. Biometric template security: Challenges and solutions. In *2005 13th European signal processing conference*, pages 1–4. IEEE, 2005.
- [270] Umut Uludag and Anil K Jain. Attacks on biometric systems: a case study in fingerprints. In *Security, steganography, and watermarking of multimedia contents VI*, volume 5306, pages 622–633. International Society for Optics and Photonics, 2004.
- [271] Seyed Abolfazl Valizadeh, Franziskus Liem, Susan Mérillat, Jürgen Hänggi, and Lutz Jäncke. Identification of individual subjects on the basis of their brain anatomical features. *Scientific reports*, 8(1):1–9, 2018.
- [272] Katharine Brigham and BVK Vijaya Kumar. Subject identification from electroencephalogram (EEG) signals during imagined speech. In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, 2010.
- [273] Gonzalo Safont, Addisson Salazar, Antonio Soriano, and Luis Vergara. Combination of multiple detectors for EEG based biometric identification/authentication. In *2012 IEEE International Carnahan Conference on Security Technology (ICCST)*, pages 230–236. IEEE, 2012.
- [274] Matteo Fraschini, Arjan Hillebrand, Matteo Demuru, Luca Didaci, and Gian Luca Marcialis. An EEG-based biometric system using eigenvector centrality in resting state brain networks. *IEEE Signal Processing Letters*, 22(6):666–670, 2014.
- [275] Jae-Hwan Kang, Young Chang Jo, and Sung-Phil Kim. Electroencephalographic feature evaluation for improving personal authentication performance. *Neurocomputing*, 287:93–101, 2018.

- [276] Alejandro Riera, Aureli Soria-Frisch, Marco Caparrini, Carles Grau, and Giulio Ruffini. Unobtrusive biometric system based on electroencephalogram analysis. *EURASIP Journal on Advances in Signal Processing*, 2008:18, 2008.
- [277] Bin Hu, Quanying Liu, Qinglin Zhao, Yanbing Qi, and Hong Peng. A real-time electroencephalogram (EEG) based individual identification interface for mobile security in ubiquitous environment. In *2011 IEEE Asia-Pacific Services Computing Conference*, pages 436–441. IEEE, 2011.
- [278] Qiong Gui, Maria V. Ruiz-Blondet, Sarah Laszlo, and Zhanpeng Jin. A survey on brain biometrics. *ACM Comput. Surv.*, 51(6):112:1–112:38, February 2019.
- [279] JX Chen, ZJ Mao, WX Yao, and YF Huang. EEG-based biometric identification with convolutional neural network. *Multimedia Tools and Applications*, pages 1–21, 2019.
- [280] Yingnan Sun, Frank P-W Lo, and Benny Lo. EEG-based user identification system using 1D-convolutional long short-term memory neural networks. *Expert Systems with Applications*, 125:259–267, 2019.
- [281] Theerawit Wilaiprasitporn, Apiwat Dittthapron, Karis Matchaparn, Tanaboon Tongbuasirilai, Nannapas Banluesombatkul, and Ekapol Chuangsuwanich. Affective EEG-based person identification using the deep learning approach. *IEEE Transactions on Cognitive and Developmental Systems*, 2019.
- [282] Ozan Özdenizci, Ye Wang, Toshiaki Koike-Akino, and Deniz Erdoğmuş. Adversarial deep learning in EEG biometrics. *IEEE Signal Processing Letters*, 26(5):710–714, 2019.
- [283] Philip Davis, Charles D Creusere, and Jim Kroger. Subject identification based on EEG responses to video stimuli. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 1523–1527. IEEE, 2015.
- [284] Thiago Schons, Gladston JP Moreira, Pedro HL Silva, Vitor N Coelho, and Eduardo JS Luz. Convolutional network for EEG-based biometric. In *Iberoamerican Congress on Pattern Recognition*, pages 601–608. Springer, 2017.
- [285] Xiang Zhang, Lina Yao, Salil S Kanhere, Yunhao Liu, Tao Gu, and Kaixuan Chen. MindID: Person identification from brain waves through attention-based recurrent neural network. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–23, 2018.
- [286] Longbin Jin, Jaeyoung Chang, and Eunyi Kim. EEG-Based User Identification Using Channel-Wise Features. In *Asian Conference on Pattern Recognition*, pages 750–762. Springer, 2019.
- [287] Daria La Rocca, Patrizio Campisi, Balazs Vegso, Peter Cserti, György Kozmann, Fabio Babiloni, and F De Vico Fallani. Human brain distinctiveness based on EEG spectral coherence connectivity. *IEEE transactions on Biomedical Engineering*, 61(9):2406–2412, 2014.
- [288] Alessandra Crobe, Matteo Demuru, Luca Didaci, Gian Luca Marcialis, and Matteo Fraschini. Minimum spanning tree and k-core decomposition as measure of subject-specific EEG traits. *Biomedical Physics & Engineering Express*, 2(1):017001, 2016.
- [289] Marco Garau, Matteo Fraschini, Luca Didaci, and Gian Luca Marcialis. Experimental results on multi-modal fusion of EEG-based personal verification algorithms. In *2016 International Conference on Biometrics (ICB)*, pages 1–6. IEEE, 2016.

- [290] Kavitha P Thomas and A Prasad Vinod. Biometric identification of persons using sample entropy features of EEG during rest state. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 003487–003492. IEEE, 2016.
- [291] Kavitha P Thomas and A Prasad Vinod. Utilizing individual alpha frequency and delta band power in EEG based biometric recognition. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 004787–004791. IEEE, 2016.
- [292] Silvio Barra, Andrea Casanova, Matteo Frascini, and Michele Nappi. Fusion of physiological measures for multimodal biometric systems. *Multimedia Tools and Applications*, 76(4):4835–4847, 2017.
- [293] Su Yang, Farzin Deravi, and Sanaul Hoque. Task sensitivity in EEG biometric recognition. *Pattern Analysis and Applications*, 21(1):105–117, 2018.
- [294] Patrizio Campisi and Daria La Rocca. Brain waves for automatic biometric-based user recognition. *IEEE transactions on information forensics and security*, 9(5):782–800, 2014.
- [295] Mohammed Abo-Zahhad, Sabah Mohammed Ahmed, and Sherif Nagib Abbas. State-of-the-art methods and future perspectives for personal recognition based on electroencephalogram signals. *IET Biometrics*, 4(3):179–190, 2015.
- [296] Amir Jalaly Bidgoly, Hamed Jalaly Bidgoly, and Zeynab Arezoumand. A survey on methods and challenges in EEG based authentication. *Computers & Security*, page 101788, 2020.
- [297] Salahiddin Altahat, Michael Wagner, and Elisa Martinez Marroquin. Robust electroencephalogram channel set for person authentication. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 997–1001. IEEE, 2015.
- [298] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deep: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.
- [299] Zijing Mao, Wan Xiang Yao, and Yufei Huang. EEG-based biometric identification with deep learning. In *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 609–612. IEEE, 2017.
- [300] Alejandro Gonzalez, Isao Nambu, Haruhide Hokari, and Yasuhiro Wada. EEG channel selection using particle swarm optimization for the classification of auditory event-related potentials. *The Scientific World Journal*, 2014, 2014.
- [301] Nobuaki Mizuguchi, Hiroki Nakata, Takuji Hayashi, Masanori Sakamoto, Tetsuro Muraoka, Yusuke Uchida, and Kazuyuki Kanosue. Brain activity during motor imagery of an action with an object: a functional magnetic resonance imaging study. *Neuroscience research*, 76(3):150–155, 2013.
- [302] Kai J Miller, Gerwin Schalk, Eberhard E Fetz, Marcel den Nijs, Jeffrey G Ojemann, and Rajesh PN Rao. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proceedings of the National Academy of Sciences*, 107(9):4430–4435, 2010.
- [303] Wolfgang Taube, Michael Mouthon, Christian Leukel, Henri-Marcel Hoogewoud, Jean-Marie Annoni, and Martin Keller. Brain activity during observation and motor imagery of different balance tasks: an fMRI study. *cortex*, 64:102–114, 2015.

- [304] Su Yang and Farzin Deravi. On the usability of electroencephalographic signals for biometric recognition: A survey. *IEEE Transactions on Human-Machine Systems*, 47(6):958–969, 2017.
- [305] Erwin HT Shad, Marta Molinas, and Trond Ytterdal. Impedance and Noise of Passive and Active Dry EEG Electrodes: A Review. *IEEE Sensors Journal*, 2020.
- [306] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.
- [307] Mahnaz Arvaneh, Cuntai Guan, Kai Keng Ang, and Chai Quek. EEG data space adaptation to reduce intersession nonstationarity in brain-computer interface. *Neural computation*, 25(8):2146–2171, 2013.
- [308] Hohyun Cho, Minkyu Ahn, Kiwoong Kim, and Sung Chan Jun. Increasing session-to-session transfer in a brain-computer interface with on-site background noise acquisition. *Journal of neural engineering*, 12(6):066009, 2015.
- [309] Feng Li, Yi Xia, Fei Wang, Dengyong Zhang, Xiaoyu Li, and Fan He. Transfer Learning Algorithm of P300-EEG Signal Based on XDAWN Spatial Filter and Riemannian Geometry Classifier. *Applied Sciences*, 10(5):1804, 2020.
- [310] Sara Hegdahl Åsly. Supervised learning for classification of EEG signals evoked by visual exposure to RGB colors. Master’s thesis, NTNU, 2019.
- [311] Shobiha Premkumar. Subject Identification using EEG Signals and Supervised Learning. Master’s thesis, NTNU, 2020.
- [312] Julie Haga. Biometric system using EEG signals from resting-state and one-class classifiers. Master’s thesis, NTNU, 2020.
- [313] Sara H Åsly, Luis Alfredo Moctezuma, Marta Molinas, and Monika Gilde. Towards EEG-based signals classification of RGB color-based stimuli. In *GBCIC*, 2019.
- [314] Alejandro A Torres-García, Luis Alfredo Moctezuma, Sara Asly, and Marta Molinas. Discriminating between color exposure and idle state using EEG signals for BCI application. In *International Conference on e-Health and Bioengineering (EHB)*, 2019.
- [315] Alejandro A. Torres-García., Luis Alfredo Moctezuma., and Marta Molinas. Assessing the Impact of Idle State Type on the Identification of RGB Color Exposure for BCI. In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 4: BIOSIGNALS.*, pages 187–194. INSTICC, SciTePress, 2020.
- [316] Andres Felipe Soler Guevara, Luis Alfredo Moctezuma, Eduardo Giraldo, and Marta Molinas. Low-density EEG source reconstruction with channel selection enabled by evolutionary optimization. *arXiv preprint*, 2019.
- [317] Pierre Baldi. Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML workshop on unsupervised and transfer learning*, pages 37–49, 2012.
- [318] Naveed Rehman and Danilo P Mandic. Multivariate empirical mode decomposition. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2117):1291–1302, 2010.
- [319] Mruthun R Thirumalaisamy and Phillip J Ansell. Fast and adaptive empirical mode decomposition for multidimensional, multivariate signals. *IEEE Signal Processing Letters*, 25(10):1550–1554, 2018.

- [320] Qingfu Zhang and Hui Li. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on evolutionary computation*, 11(6):712–731, 2007.

ISBN 978-82-471-9693-9 (printed ver.)
ISBN 978-82-471-9970-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)



NTNU

Norwegian University of
Science and Technology