UCLouvain

École polytechnique de Louvain

Is it possible to identify anxiety, depression and/or alcoholism from EEG (resting EEG and event related potentials) and structural MRI data using machine learning algorithms?

Author: **Jean GILLAIN**
Supervisor: **Benoit MACQ**
Readers: **Gaëtan RENSONNET, Phillipe DE TIMARY, Nicolas DELINTE, Jean-Charles DELVENNE**
Academic year 2020–2021
Master [120] in Mathematical Engineering

Abstract

Identifying robust and non-invasive biomarkers of psychiatric diseases in the brain is still an open challenge. These unknowns prevent the effective treatment of patients suffering from alcoholism, depression and anxiety for example. In this work, we investigate whether it is possible using machine learning algorithms (SVM, random forest, ..) to predict from EEG and structural MRI data whether subjects are alcoholic, depressive or anxious. To do this, we used two datasets collected at St-Luc University Hospital. The first one includes 210 alcoholic, depressive, anxious and control subjects for whom structural MRI data have been collected. The second dataset includes 43 alcoholics with anxiety and depression comorbidities at the end of their withdrawal, for which EEG data (resting state and event related potentials) were collected. The results based on our two datasets showed that, (i) the structural MRI data allowed to differentiate alcoholics from controls, (ii) the most different variables between alcoholics and controls corresponded to the cortical thickness of the brain structures, and (iii) it was possible to identify slight differences between the ERP components of depressive and non depressive alcoholic patients. These results indicate that EEG and structural MRI data could be used to help diagnose psychiatric patients and that their potential should be exploited to devise new treatments and to study more complicated problems such as relapse in alcoholic patients.

Acknowledgements

I would like to thank Benoit Macq for giving me the opportunity to work on this subject during my master's thesis.

I would also like to show my gratitude to Gaetan and Nicolas who took the time to answer my questions on both medical and machine learning issues.

I would also like to thank the people at the hospital who made this project possible and helped me to understand the medical data. More specifically I am addressing Philippe de Timary, Géraldine Petit, Sophie Leclercq and Laurence Dricot.

Contents

Abstract **ii**

Acknowledgements **iii**

Contents **iv**

List of Figures **vii**

List of Tables **xi**

Introduction **1**

I State of the art **2**

1 Cerebral examinations **3**
1.1 Electroencephalogram . . . 3
1.1.1 Structure of the neuron . . . 3
1.1.2 Brain structure . . . 4
1.1.3 System of Electrodes Placement . . . 5
1.1.4 EEG frequency ranges/ Frequency Bands . . . 6
1.1.5 Event related potentials (ERP) . . . 7
1.2 Magnetic resonance imaging (MRI) . . . 10
1.2.1 Magnetic properties . . . 10
1.2.2 Radiofrequency pulses(RF) . . . 11
1.2.3 Spin-echo sequence . . . 12
1.2.4 Structural MRI features . . . 13

2 Psychiatric diseases **15**
2.1 Alcohol Use Disorder . . . 15
2.1.1 General Overview . . . 15
2.1.2 MRI and AUD [39] . . . 16
2.2 Depression . . . 16
2.2.1 General overview . . . 16
2.2.2 Resting EEG and Depression . . . 17
2.2.3 ERP and depression . . . 18
2.2.4 Structural MRI and depression [68] . . . 18
2.3 Anxiety . . . 19
2.3.1 General overview . . . 19
2.3.2 Resting EEG and anxiety . . . 19
2.3.3 Structural MRI and anxiety [83] . . . 20

II Prediction of alcoholism, depression and anxiety from EEG and MRI data: A case study **22**

3 Methods **23**
3.1 Data . . . 23
3.1.1 Group segmentation and problem statement . . . 24
3.1.2 MRI-based dataset . . . 25
3.1.3 EEG-based dataset . . . 26
3.2 Preprocessing . . . 27
3.2.1 Preprocessing of Resting EEG signals . . . 27
3.2.2 Preprocessing of Event-related-potiential . . . 29
3.2.3 Preprocessing of MRI data . . . 30
3.3 Features extraction . . . 30
3.3.1 Resting state EEG features . . . 30
3.3.2 ERP features . . . 31
3.3.3 MRI features . . . 32
3.3.4 Summary of extracted features . . . 33
3.4 Machine learning methods . . . 33
3.4.1 Support Vector Machine (SVM) . . . 33
3.4.2 Logistic regression . . . 35
3.4.3 Linear Discriminant Analysis . . . 35
3.4.4 Random Forest . . . 36
3.4.5 Comparison of methods . . . 37
3.4.6 Standardisation and correlation . . . 37
3.4.7 Feature selection methods . . . 38
3.4.8 Training and test phase . . . 40
3.4.9 Parameters tuning . . . 41
3.4.10 Performance metrics . . . 43
3.5 Strategies . . . 44
3.5.1 All-sMRI strategy . . . 44
3.5.2 Separate-sMRI strategy . . . 45
3.5.3 Resting EEG and ERP strategy . . . 46
3.5.4 Multimodal strategy using sMRI data . . . 47

4 Results **49**
4.1 All-sMRI strategy . . . 49
4.1.1 Problem 1 : Detection of alcoholism . . . 49
4.1.2 Problem 2: Detection of anxiety . . . 50
4.1.3 Problem 3: Detection of depression . . . 50
4.1.4 Problem 4: Detection of anxiety for alcoholic patients . . . 51
4.1.5 Problem 5: Detection of depression for alcoholic patients . . . 51
4.2 Separate-sMRI strategy . . . 52
4.2.1 Problem 1 : Detection of alcoholism . . . 52
4.2.2 Problem 2 : Detection of anxiety . . . 53
4.2.3 Problem 3 : Detection of depression . . . 53
4.2.4 Problem 4 : Detection of anxiety for alcoholic patients . . . 54
4.2.5 Problem 5 : Detection of depression for alcoholic patients . . . 54
4.3 Resting EEG and ERP strategy . . . 55
4.3.1 Problem n°4 : Detection of anxiety for alcoholic patients . . . 55
4.3.2 Problem n°5 : Detection of depression for alcoholic patients . . . 55
4.4 Multimodal strategy using sMRI data . . . 56

5 Discussion **57**
5.1 Limitations . . . 57
5.2 Classification of alcoholism . . . 58
5.2.1 All-sMRI strategy . . . 58

5.2.2 Separate-sMRI strategy . . . 59
5.2.3 Multimodal strategy using sMRI data . . . 60
5.3 Classification of anxiety state . . . 61
5.4 Classification of depression . . . 62
5.5 Classification of anxiety for alcoholic patients . . . 63
5.6 Classification of depression for alcoholic patients . . . 64
5.6.1 MRI based dataset . . . 64
5.6.2 ERP-based dataset . . . 65
5.6.3 Resting EEG dataset . . . 65
5.7 Choices . . . 65
5.7.1 Choice of the classifier . . . 65
5.7.2 Choice of the performance metrics . . . 66
5.8 Interest of using EEG and MRI features for the prediction of diseases . . . 67
5.9 Future perspectives . . . 67

Conclusion **70**

Bibliography **71**

Appendix **82**

A Atlases **82**
A.1 Destrieux Atlas . . . 82
A.2 Desikan Atlas . . . 85

B Feature selections **87**
B.1 Wrapper methods . . . 87
B.2 Feature importance for random forest [192] . . . 88

C Machine learning results : extension **90**
C.1 All-sMRI strategy . . . 90
C.1.1 Problem n° 2 : Detection of anxiety . . . 90
C.1.2 Problem n°4 : Detection of anxiety for AUD patients . . . 90
C.2 Separate-sMRI strategy . . . 91
C.2.1 Problem n°2 : Detection of anxiety . . . 91
C.2.2 Problem n°4 : Detection of anxiety for AUD patients . . . 91
C.3 EEG and ERP strategy . . . 92
C.3.1 Problem n°4 : Detection of anxiety for AUD patients . . . 92

D Most important features **93**
D.1 Problem n°1 : Detection of alcoholism . . . 93
D.2 Problem n°2 : Detection of Anxiety . . . 96
D.3 Problem n°3 : Detection of depression . . . 99
D.4 Problem n°4 : Detection of anxiety for AUD patients . . . 102
D.5 Problem n°5 : Detection of depression for AUD patients . . . 105

E Correlation between Resting state EEG and microbiota **109**
E.1 Data . . . 109
E.2 Methods and results . . . 109

List of Figures

1.1 General illustration of the neuron structure, from [2]. . . . 4
1.2 General overview of how the EEG signal is recorded and its origin, from [5]. . . . 4
1.3 Illustration of the position of the electrodes on the scalp of the 10-20 System, from [10]. . . . 5
1.4 Illustration of the delta, theta, alpha, beta and gamma bands of an electroencephalogram wave from [11]. The recorded EEG signal (above) is the mixture of different sub-signals (below), each oscillating in a specific frequency. . . . 6
1.5 Example of exogenous (until almost 100ms after the onset of a stimulus) and endogenous (peaking later) ERP waveforms from [16]. . . . 7
1.6 Example of CNV recorded wave at the Cz electrode highlighting early and late CNV, adapted from [27]. . . . 9
1.7 Formation of parallel and anti parallel spins due to magnetic field B_0 from [29]. . . . 10
1.8 Illustration of the system after 90° RF(left) and 180° RF(right) after being initially in resting state , see figure 1.7 from [29]. . . . 11
1.9 Spin echo sequence from [31, 32]. . . . 12
1.10 Illustration of T1 weighted image recording sequence (left) and obtained result (right), from [31, 32]. 13
1.11 Illustration of T1 weighted image recording sequence (left) and obtained result (right), from [31, 32]. 13
1.12 Mean curvature and Gaussian curvature illustrated from [33]. Above figures show mean curvature in 2 and 3 dimensions. The below figures illustrate mean and Gaussian curvature obtained on a brain. Positive and negative curvature are drawn in red and green. . . . 14

2.1 Brain structural MRI imaging. First picture shows most impacted regions by alcohol, the second one shows significant differences between AUD and controls [39]. . . . 16
2.2 Cortical thickness of Orbitofrontal, cingulate cortex and insula, and temporal lobe are the most impacted regions for MDD, from [43]. . . . 18
2.3 GAD mainly affects volumes of the amygdal, anterior cingulate cortex, postcentral gyrus, dorsolateral and ventrolateral prefrontal cortex. [95]. . . . 20

3.1 Pipeline of the project separated in 5 steps: data acquisition, preprocessing, features extraction, machine learning and strategies. . . . 23
3.2 Type of data at disposal for the moment at St-Luc hospital. Data is mainly available for sMRI and EEG in T1 and T2. . . . 24
3.3 Data distribution of the MRI-dataset showing the correlation between AUD and anxious subjects (left) correlation between AUD and depressed subjects (right). . . . 26
3.4 Power spectral density of the recorded EEG signals at the 23 electrodes(each corresponds to a different color) before (left) and after (right) applying a Notch filter where the peak around 50Hz is then deleted. . . . 28
3.5 Type of artefacts presented on eeglab. Linear trend, discontinuity and transient high frequency are often due to environmental artefacts or electrodes malfunction, while low frequency and high frequency artefacts are respectively due to eye blinking and muscular movements. . . . 29
3.6 Report from `ANT Neuro` for the P50 wave, from [118]. . . . 32
3.7 Example of data separation with a Support vector machine by maximising the margin, illustration from [124]. . . . 34
3.8 Example of majority vote for random forest algorithm from [130]. For each tree of the forest, a new matrix of size $N \times \sqrt{m}$ given as input, each returning a class ($\pm$ 1). The output of the random forest is given by the majority vote of these outputs. . . . 36

3.9 Feature selection methods: sub-figure a represents filter methods, sub-figure b represents embedded methods and sub-figure c wrapper methods. Projection methods are not the figure but can be compared to filter methods, instead of importance score, projection in a new feature subspace is done. . . . 39
3.10 Grid search cross-validation from [142]. Above figure shows data separation in training and test. The bottom figure shows 5-split cross-validation where the blue part is for each split the validation set and the green part the training set. . . . 41
3.11 Example of confusion matrix for binary classification problem (where A = ill, B = healthy) from [146]. . . . 43
3.12 Scheme of the all-sMRI strategy. All sMRI data is given as input features of the classifiers, data is then split in a training (80%) and a test set(20%). Filter or embedded methods or PCA are then used in order to reduce the number of dimensions (the number of dimensions is chosen by cross-validation as the hyperparameters of the model). The classifier models are then fitted with the training set and hyperparameters are found by cross-validation. Performances of the classifiers are then evaluated on the test set (repeated 20 times). . . . 45
3.13 Scheme of the separate-sMRI strategy. There is one classifier per MRI data, the same steps as for the all-sMRI strategy are then performed. . . . 46
3.14 Scheme of the EEG strategy. The EEG dataset is splitted in two (resting EEG and ERP dataset), each giving rise to a input feature matrix. Same steps as for the all-SMRI strategy are then performed, except that training and test set are split by LOO. . . . 47
3.15 Mulitmodal strategy using sMRI data : combination of structural MRI data. . . . 48

4.1 Accuracy and Cohen's κ using model all-sMRI strategy in order to predict alcoholism. . . . 49
4.2 Recall and Precision using model all-sMRI strategy in order to predict alcoholism. . . . 50
4.3 Accuracy and Cohen's κ using all-sMRI strategy in order to predict depression. . . . 50
4.4 Recall and Precision using model all-sMRI strategy in order to predict depression. . . . 51
4.5 Accuracy and Cohen's κ using all-sMRI strategy in order to predict depression for AUD patients. 51
4.6 Recall and Precision using all-sMRI strategy in order to predict depression for AUD patients. . . 52
4.7 Accuracy and Cohen's κ using separate-sMRI strategy in order to predict alcoholism. . . . 52
4.8 Recall and Precision using separate-sMRI strategy in order to predict alcoholism. . . . 53
4.9 Accuracy and Cohen's κ using separate-sMRI strategy in order to predict depression. . . . 53
4.10 Recall and Precision using separate-sMRI strategy in order to predict depression. . . . 54
4.11 Accuracy and Cohen's κ using separate-sMRI strategy in order to predict depression of AUD patients. . . . 54
4.12 Recall and Precision using separate-sMRI strategy in order to predict depression of AUD patients. 55
4.13 Accuracy and Cohen's κ obtained with resting EEG dataset in order to predict depression for AUD patients. . . . 55
4.14 Accuracy and Cohen's κ obtained with ERP strategy in order to predict depression for AUD patients. . . . 56

5.1 MRI data for alcoholic patients and controls projected in the most important variables plane for the filter method(p-value) showing strong data separation. For a better understanding of the abbreviations, see Appendices A.2 and A.1. The suffix ct_ means it is a cortical thickness variable. 58
5.2 15 Most important features highlighted by random forest for detection of AUD patients with MRI features highlighting cortical thickness features as a strong biomarker for the detection of AUD. . 59
5.3 Multi modal neuroimaging method using NAC, RWR, CTH, CSH GMD data to classify people as alcoholic or not from [158]. . . . 60
5.4 Probability estimation to be alcoholic according the four different types of MRI features. Probabilities obtained with SVM using other features than mean cortical thickness are not accurate enough improve the accuracy of the model obtained with mean cortical thickness. . . . 61
5.5 Statistically-significant differences do not necessarily imply high classification accuracy, and conversely, from [166]. . . . 62
5.6 Best area under curve is given in top left corner of the graph, while worst area under curve is a straight line from (0,0) to (1,1). . . . 67
5.7 Three last steps of the potential future project : Predicting relapse of alcoholic patients. . . . 68

A.1 Destrieux atlas [111] . . . 82
A.2 Desikan atlas [112] . . . 86

C.1 Accuracy and Cohen's κ using all-sMRI strategy in order to predict anxiety. . . . 90
C.2 Accuracy and Cohen's κ using all-sMRI strategy in order to predict anxiety for AUD patients. . 91
C.3 Accuracy and Cohen's κ using separate-sMRI strategy in order to predict anxiety. . . . 91
C.4 Accuracy and Cohen's κ using separate-sMRI strategy in order to predict anxiety for AUD patients. 92
C.5 Accuracy and Cohen's κ obtained with ERP data in order to predict anxiety for AUD patients. . 92
C.6 Accuracy and Cohen's κ obtained with resting EEG data in order to predict anxiety for AUD patients. . . . 92

D.1 5 most statically significant Gaussian Curvature features comparing AUD and controls. . . . 94
D.2 5 most statically significant cortical thickness standard deviation features comparing AUD and controls. . . . 94
D.3 5 most statically significant mean cortical thickness features comparing AUD and controls. . . . 95
D.4 5 most statically significant volume features comparing AUD and controls. . . . 95
D.5 Group of AUD and controls in the plan of most important features (p-value obtained with Welch method). . . . 96
D.6 Most important features obtained with random forest algorithm based on Gini impurity for AUD and controls. . . . 96
D.7 5 most significant Gaussian Curvature features comparing anxious and non-anxious. . . . 97
D.8 5 most significant cortical thickness standard deviation features comparing anxious and controls. 97
D.9 5 most significant volume features comparing anxious and non-anxious. . . . 98
D.10 Group of anxious and non-anxious in the plan of most important features (p-value obtained with Welch method). . . . 98
D.11 Most important features obtained with random forest algorithm based on Gini impurity for Anxious and non-anxious. . . . 99
D.12 5 most statistically significant Gaussian Curvature features comparing depressed and non-depressed. . . . 99
D.13 5 most significant statistically cortical thickness standard deviation features comparing depressed and non-depressed. . . . 100
D.14 5 most statistically significant volume features comparing depressed and non-depressed. . . . 100
D.15 5 most statistically significant mean cortical thickness features comparing depressed and non-depressed. . . . 101
D.16 Group of depressed and non- depressed in the plan of most important features (p-value obtained with Welch method). . . . 101
D.17 Most important features obtained with random forest algorithm based on Gini impurity for the detection of depression from MRI features. . . . 102
D.18 5 most statistically significant Gaussian Curvature features comparing anxious AUD and non-anxious AUD. . . . 102
D.19 5 most statistically significant Standard Deviation features comparing anxious AUD and non-anxious AUD. . . . 103
D.20 5 most statistically significant volume features comparing Anxious AUD and non-anxious AUD. 103
D.21 5 most statistically significant mean cortical thickness features comparing anxious AUD and non-anxious AUD. . . . 104
D.22 Group of anxious AUD and non-anxious AUD in the plan of most important features (p-value obtained with Welch method). . . . 104
D.23 Most important features obtained with random forest algorithm based on Gini impurity for anxious AUD and non-anxious AUD. . . . 105
D.24 5 most statistically significant Gaussian Curvature features comparing depressed AUD and non-depressed AUD. . . . 105
D.25 5 most statistically significant cortical thickness Standard Deviation features comparing depressed AUD and non-depressed AUD. . . . 106
D.26 5 most statistically significant volume features comparing depressed AUD and non-depressed AUD. . . . 106

D.27 5 most statistically significant mean cortical thickness features comparing depressed AUD and non-depressed AUD. 107
D.28 Group of depressed AUD and non-depressed AUD in the plan of most important features (p-value obtained with Welch method). 107
D.29 Most important features obtained with random forest algorithm based on Gini impurity for depressed AUD and non-depressed AUD. 108

E.1 The left and central figures highlight *Delta* PSD for patients with a low and high permeability. The last figure shows where those changes are significant thanks to p-values for EC. 110
E.2 The left and central figures highlight *Delta* PSD for patients with a low and high permeability. The last figure shows where those changes are significant thanks to p-values for EO. 110
E.3 Regression model linking permeability and mean PSDs. 111

List of Tables

1.1 Relaxation times and water importances of brain tissues at 1.5 Tesla, modified from [29]. 12

3.1 Description of the experimentation in order to trigger the P300 component. 27
3.2 Table regrouping the features from different clinical exams. 33
3.3 Advantages and drawbacks of machine learning methods used for this thesis. 37
3.4 Hyperparameters of the different methods. Words in bold are choices that were not optimised during grid-search. 42

4.1 Results of Multimodal structural MRI strategy in order to detect alcoholism and depression. . . 56

E.1 Summary of clinical population for permeability experiment. 109

Glossary

AUDIT Alcohol Use Disorders Identification Test. 15

BDI Beck Depression Inventory. 17

CF Cerebrospinal Fluid. 59

EC Eyes Closed. 30

EEG Electroencephalogram. 1

EO Eyes Open. 30

ERP Event Related Potentials. 3

FFT Fast Fourrier Transform. 31

FN False Negative. 43

FP False Positive. 43

GAD General Anxiety Disorder. 19

GM Grey Matter. 4

LDA Linear discriminant analysis . 35

LOO Leave-One-Out. 41

MDD Major Depression. 16

MEG Magnetoencephalograph. 17

NAC Nucleus Accumbens Connectivity. 59

OCD Obsessive-Compulsive Disorder. 19

PSD Power Spectral Density. 31

PTSD Post Traumatic Stress Disorder. 19

rEEG Resting EEG. 3

ROC Receiver Operator Curve. 66

RWR Reward Response. 59

sMRI Structural Magnetic Resonance Imaging. 1

STAI State-Trait Anxiety Inventory. 19

SVM Support Vector Machine. 34

TE Echo Time. 12

TN True Negative. 43

TP True Positive. 43

TR Repetition Time. 12

WLLN Weak Low of Large Numbers. 43

WM White Matter. 4

Introduction

The links between psychiatric pathologies and their pathophysiological mechanisms are still unclear. Indeed, the study of the brain was still limited a few years ago, which has caused research to lag behind. Moreover, the difficulties in producing explanations between mental, biological and physical states could be due to intrinsic differences between biology and psychology. Given the current limit of scientific knowledge of psychiatric pathologies, some different pathologies are believed to be the same because they have a number of similar symptoms, but their neurological causes are not always the same. Thus, they are often treated in the same way, which results in their treatment not always being effective. Nowadays, clinical diagnoses in Belgium are mainly made by a psychiatrist after a discussion with the patient, using the Diagnostic and Statistical Manual of Mental Disorders (DSM V), and/or via the results of questionnaires answered by the patient.

With the advent of new technologies, mainly structural Magnetic Resonance Imaging (sMRI) and Electroencephalogram (EEG), other types of clinical diagnosis are now available to improve the diagnosis of psychiatric patients. EEG is a method which records the neuronal activity of the brain using electrodes placed on the scalp. Structural MRI is a technique which produces images of the anatomy of the brain. These methods are currently not widely used in psychiatric wards in Belgium and their potential is not fully exploited. The aim of this thesis is therefore to identify, using machine learning algorithms, whether MRI and EEG data can be used to identify alcoholism and its comorbid diseases such as anxiety and depression. The research question is therefore: "Is it possible to identify anxiety, depression and/or alcoholism from EEG (resting EEG and event related potentials) and structural MRI data? ".

To achieve this objective, we focus on one EEG and one structural MRI dataset, each collected at the end of the alchohol withdrawal period, i.e., approximately twenty days after the start of the patient's care by the psychiatrist. These data were collected from the archives of clinical examinations performed in recent years in psychiatrist department of St-Luc Hospital.

The MRI dataset is composed of 210 subjects, including 141 alcoholics and 69 controls. Among these subjects, 69 are also in a depressive state and 44 in an anxiety state. For the latter, we will apply machine learning methods that will allow, based on the MRI data, to classify the subjects as alcoholic or not, in a depressive state or not, in an anxious state or not with the whole dataset or only within the subset of alcoholic patients.

The EEG dataset contains 43 alcoholic subjects from the withdrawal ward of St-Luc Hospital. This one was also collected after the withdrawal period and contains features extracted from evoked potentials and EEG at rest. For the EEG dataset, we will apply machine learning methods to classify the alcoholic patients in an anxious and/or depressive state.

The report is divided into two subparts: the literature review and the proof of concept. First, a theorical background relates the characteristics of the EEG and MRI signals, the diseases studied, and the machine learning methods used. The second section is a proof of concept and focuses first on the types of data that were used and the characteristics of the signals that were extracted. Then, a section is dedicated to the machine learning algorithms and the strategies implemented. It concludes with a section on the results we have obtained from the machine learning algorithms we implemented and a discussion of these results.

Part I

State of the art

Chapter 1

Cerebral examinations

This chapter focuses on the two types of medical brain data that have been used as input to our machine learning algorithms. The first part of this chapter focuses on electroencephalogram. It tries to answer the following questions: What is an electroencephalogram ? What does it record? What is the interpretation of the characteristics of the EEG signal? The second part of this chapter will focus on MRI and will explain the physics behind the recording of this signal. It will also present the different characteristics that are studied for this signal.

1.1 Electroencephalogram

Electroencaphalography is a non invasive method introduced by Adolf Beck in 1890 [1], giving a global overview of the functional work of the brain. Electroencephalogram (EEG) is a record of the oscillations of brain electrical potential recorded via some electrodes on the scalp. Brain cells communicate information to other neurons through other cells via electric or chemical signals. The transmission of information through neurons leads to the creation of an electromagnetic field which can be recorded by electrodes on the scalp. In the literature, we often consider two kinds of EEG, resting state EEG (rEEG) and Event-Related Potentials (ERP). As the name suggests it, resting EEGs provide information about the neural activity of a subject at rest, whereas ERPs are concerned with the neural activity of a subject after a certain task or stimulus.

1.1.1 Structure of the neuron

Neurons are the cells that form the basis of the EEG signal. A neuron is a cell of the nervous system specialized in the communication and treatment of information. The total number of neurons in the brain is more than 100 billion and are therefore capable of creating an incredibly complex network, sometimes with more than 100,000 synapses per neuron.

The neuron is composed of the following parts [2]:

- A cell body, also called the nucleus, which contains chromosomes, responsible for the production of proteins.
- Numerous dendritic branches from which information comes from.
- An axon, through which the information is diffused. Synapses enable to put into contact axons and dendrites of different cells which enable the passage of information.

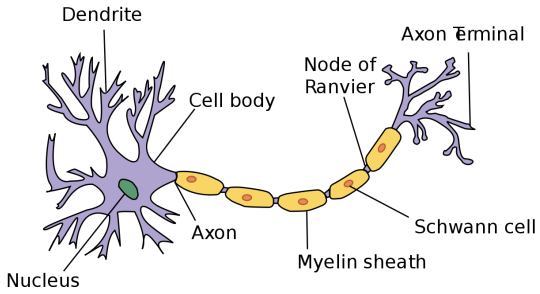

Figure 1.1: General illustration of the neuron structure, from [2].

Neuroscientsts claim that brain behaviors are related to the result of interaction of neurons and assemblies of neurons [3]. Neurons are excitable cells, which means that a stimulus (voltage) can lead to the formation of a bioelectric signal in the cell, which can be transmitted to other neurons or to other tissues to activate them (muscles, secretory glands, etc.) [4]. Those stimuli can be measured with an electroencephalogram. The scalp potentials which are measured by electrodes on the scalp are generated by millisecond-scale modulations of synaptic current sources at the cortical surfaces [5, 6], see figure 1.2. More precisely, the movement of ions in the membrane of neurons creates an excitatory postsynaptic potential. An excitatory postsynaptic potential is characterized by a temporary reduction in the membrane potential. This can be excitatory inducing extracellular potentials and creating a current. The electrical potential recorded with the EEG is created via those extracellular currents. Those currents are measured at far distance from its source and can thus be approximated by a dipole moment per volume [5]. This approximation is valid in the case of EEG, for a mathematical proof see [7].

1.1.2 Brain structure

As the EEG signals record the brain activity, a short reminder of the brain structure is needed. The brain is composed of white and grey matter. Grey matter (GM) is composed of different cell bodies and few axons. White matter (WM) is mainly composed of axons and is responsible for communication between the different neuronal cells. The brain is divided in three primary regions, they are called the brainstem, cerebellum and cerebrum (see figure 1.2).

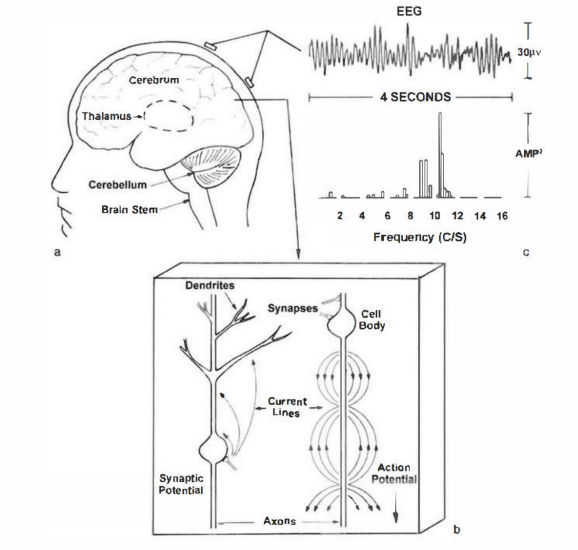

Figure 1.2: General overview of how the EEG signal is recorded and its origin, from [5].

The brain is also separated in two cerebral hemispheres, the left and the right one and are connected via the corpus callosum, containing axonal fibers, enabling interaction between the two hemispheres. Hemisphere

interaction occur thanks to more than 100 millions callosal axons connecting the two hemispheres and that is the reason why the brain structure is often compared to a neural network or even to an electrical circuit. Circuit elements could be compared to unique neurons or even to groups of neurons of different size. The axons could then be imagined as wires connecting the circuit elements. In this simple picture, the circuit elements would be controlled via electrical and chemical inputs form the brainstem [5].

1.1.3 System of Electrodes Placement

As the recording of the EEG signals depends on the locations of the electrodes, a general system placement for electrodes is often used [8]. To record EEG signals, at least three electrodes are needed, two measurement and one grounding electrodes. For most system placements, the grounding electrode is placed on the neck or on the ears. It provides a reference voltage which enables to prevent amplifier drift and to reject environmental artefacts. The measuring electrodes are placed on the rest of the scalp and, as the name suggests, they record the electrical activity of the scalp.

The most widely used electrode placement system is the 10-20 System (see figure 1.3), used by *Cognitrace*, the company having sold its electrodes to the St-Luc hospital. It is an international method which was introduced by the American Electroencephalographic Society [9]. This method ensures that

- Inter-electrode spacing is equal.
- Tests are reproducible.
- The complete brain region is covered.

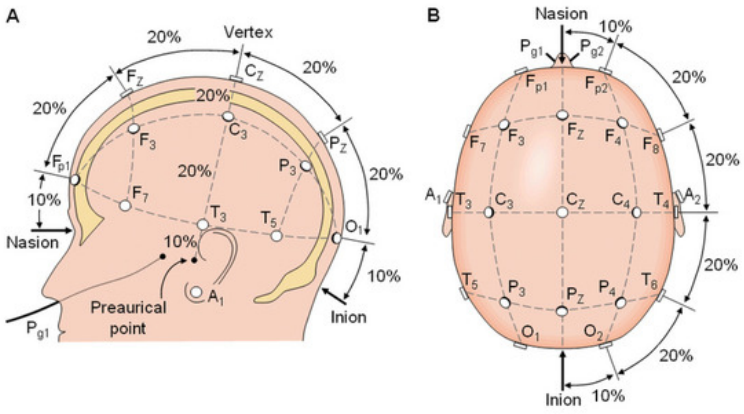

Figure 1.3: Illustration of the position of the electrodes on the scalp of the 10-20 System, from [10].

The code/name of each electrode, composed of a sequence of characters followed by a number, gives information about its position on the scalp [10]:

- A : Eear Lobe
- C : Central Lobe
- Fp : Frontral Polar
- O : Occipital Lobe
- P : Parietal Lobe
- Pg : Nasopharyngeal
- T : temoral Lobe

1.1.4 EEG frequency ranges/ Frequency Bands

We often consider the resting EEG signal as the mixture of different signals with varying frequency ranges, considered to reflect certain cognitive states. The EEG signal is thus often decomposed in different subbands : *Delta* band (1-4Hz), *Theta* band (4-8Hz), *Alpha* band (8-12 Hz), *Beta* band (13-30 Hz) and *Gamma* band (>30Hz) [11]. For a comparison between the different brain waves, see figure 1.4.

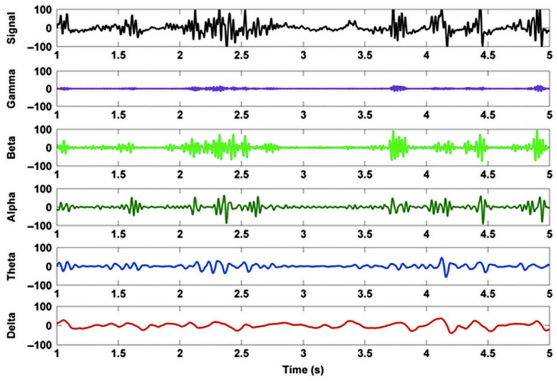

Figure 1.4: Illustration of the delta, theta, alpha, beta and gamma bands of an electroencephalogram wave from [11]. The recorded EEG signal (above) is the mixture of different sub-signals (below), each oscillating in a specific frequency.

Delta Waves

The delta waves have a frequency in the range 1-4Hz. They are high-amplitude waves associated to deep sleep, coma or hypnosis. The delta waves are also linked to different brain functions other than deep sleep, e.g., high frontal delta waves in awake subjects are associated with cortical plasticity [11]. Reductions of delta amplitudes are often observed for schizophrenia and alcoholic patients during sleep.

Theta Waves

The theta waves are in the frequency range 4-8Hz. They are more often observed when people are in a drowsy state and they are more common for children than for adults. The hippocampus contains large layers which are major contributors of the theta waves formation. Increase in the theta activity for awake subjects is often related to brain abnormalities or diseases [12].

Alpha Waves

The alpha waves have their frequencies in the range 8-12Hz. Activity in alpha waves is related to different stages of sleep. Those waves are having higher amplitudes when people are relaxed [12]. The boundary between theta and alpha waves is not clear, they are both associated to memory actions [13].

Beta Waves

The beta waves are in the range 13-30 Hz (on average). They have a lower amplitude than the waves described before. The activity of beta waves has been observed during different activities such as: activeness, anxious thinking or concentration. Engel and Fries proved that beta activity was linked with cognitive processing and the motor system [14]. The beta waves over the motor system are correlated with muscle contractions. When the activity of beta waves is more important, it is often due to the use of sedatives or drugs.

Gamma Waves

Gamma waves are the waves with a frequency higher than 30 Hz. They often get small amplitude and high contamination due to noise, due e.g. to muscle artifacts or eye blinking. Still some studies proved that high activities in temporal locations are associated with memory processes [12].

1.1.5 Event related potentials (ERP)

The event related potential is a part of the recorded EEG signal, starting after a stimulus or a certain event. The ERP components can be quantified by averaging the activity in the EEG according to a specific event. The result is a waveform associated with the processing of that specific event. The characteristics of the ERP components are [15, 16]

- Its position : the brain-response will not be the same in the central or occipital lobes.
- Its polarity : positive or negative; P300 and N100 peaking respectively positively and negatively.
- Its latency : P50 and P300 peaking on average 50 and 300 ms after a stimulus.

We observe two types of event related potentials, displayed on figure 1.5, after a certain stimulus [15]:

- Sensory or exogenous waves :
  - They peak roughly the first 100 ms after a stimulus.
  - They depend on physical features of the sensory stimulus.
  - They are independent of the level of concentration and cognition processes.
- Cognitive or endogenous waves :
  - They peak later than the exogenous waves.
  - They do not depend on physical features of the sensory stimulus.
  - They depend of the level of concentration and of the resources required for stimulus processing.
  - They are influenced by cognition processes.

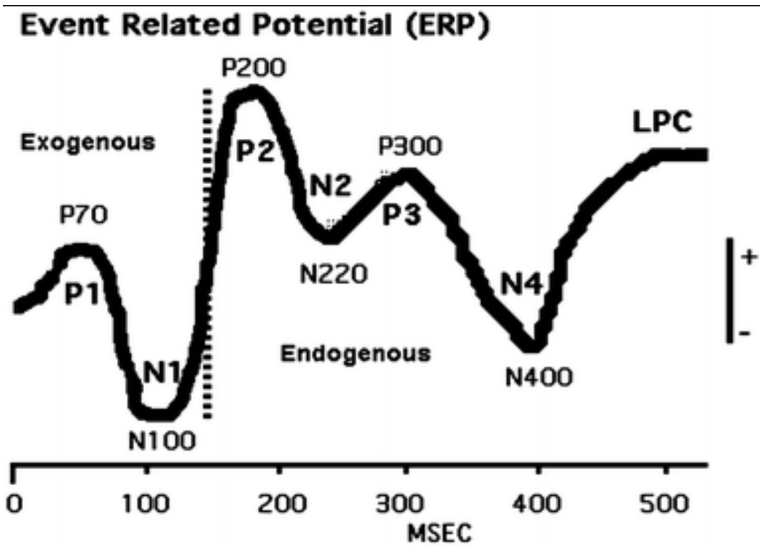

Figure 1.5: Example of exogenous (until almost 100ms after the onset of a stimulus) and endogenous (peaking later) ERP waveforms from [16].

.

In the end of this section, we mainly focus on the ERP components that have been used in our case-study.

P50

The P50 can be observed after a pair clicked paradigm [17]: two auditive clicks (S1,S2) are presented after each other to the patients with a certain interstimulus interval (ISI). Each click enhances a stimulation in the brain and creates what we call the P50 S1 and P50 S2 component. The P50 is the ERP component which peaks first, on average 50 ms after an auditory click and it is the one with the smallest amplitude of all [18]. The P50 is one of the first step of the signal processing related to the attentional filtering, occuring in the brain after the stimulus. The P50 is sometimes also called P1, as on figure 1.5.

P200

The P200 waveform component is observed between 150 and 280 ms and peaks around 200 ms after a stimulus. The P200 component could return pre-attentive alerting mechanisms that help improving stimulus perception. For our case study, the P200 wave was examined after an oddball auditory stimulus, see next section for a better explanation.

P300

EEG low-frequency components, especially delta bands, are the primary contributor to the P300 peak of event-related potentials. The P300 wave would take its source in the limbic system and the hippocampus [19]. Different methods have been used to highlight the P300 wave among which the oddball paradigm. The oddball paradigm was discovered by Nancy Squires *et al* [20]. Different stimulus are presented to the subject, with a less frequent stimulus referred to as the oddball. This infrequent stimulus triggers the P300 response.

The P300 wave peaks roughly between 280 and 500 after the oddball. The latency is often associated to the speed of the stimulus classification such that shorter latencies are associated to people with higher mental performances. The amplitude of the P300 wave is associated with and short memory [21]. "Reduced P300 amplitude is an indicator of the broad neurobiological vulnerability that underlies disorders within the externalizing spectrum, such as alcohol dependence, drug dependence, nicotine dependence, conduct disorder and adult antisocial behavior" [22]. Studies proved that the P300 latencies and amplitudes decrease with age [23].

Contigent Negative Variation (CNV)

The CNV is an ERP component introduced by Walter *et al.* in 1964 [24]. The CNV can be generated by a standard reaction time paradigm (S1-S2-motor response) [15]. A warning signal (S1) is broadcasted to the subject, a sound or an image on a screen for example. This signal is followed by an imperative signal (S2) to which the patient must react, this signal is potentially different from S1. The CNV component is the part of the EEG signal recorded between S1 and S2. Two different waves are observed in the CNV, the early CNV taking place 1000 up to 2000 ms after the warning stimulus and the late CNV which takes place 2000 to 3000 ms after the warning stimulus. The early CNV would be generated by the warning stimulus [25], while the late CNV would be linked to the motor preparation of the second stimulus [26].

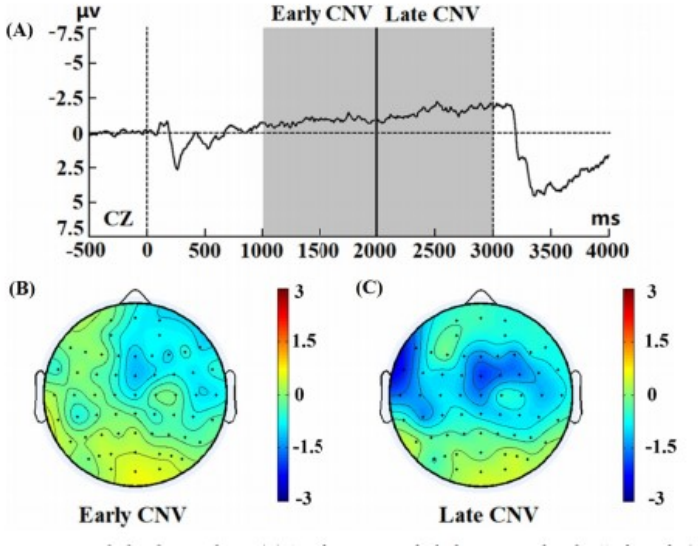

Figure 1.6: Example of CNV recorded wave at the Cz electrode highlighting early and late CNV, adapted from [27].

1.2 Magnetic resonance imaging (MRI)

Magnetic resonance imaging (MRI) is a medical non-invasively imaging technique that provides two and three-dimensional views of the interior of the body and enables to offer great contrast of soft tissues (gray and white matter). Compared to computed tomography (CT), MRI often gives less accurate images but it is safer for human since it does not produce any radiation. The first MRI on humans was done in 1977 and is since used on average $\frac{52}{1000}$ inhabitants/year [28].

1.2.1 Magnetic properties

In quantum mechanics, all atomic nuclei with an odd atomic number (^{1}H, ^{13}C, ^{31}P) have a spin, also commonly called angular momentum, the fundamental principle on which MRI images are obtained. Due to the large number of hydrogen atoms bound to water molecules in the human body, ^{1}H-MRI is the most commonly used technique.

A spinning charge and therefore a spinning ^{1}H particle creates an electric field and a magnetic field (μ) aligned with its spin axis. When no magnetic field is applied to a set of rotating charges, the resulting sum of these magnetic fields is equal to the vector null.

However, when a B_0 external magnetic field is applied to this same set rotating charges (protons for ^{1}H), these charges align themselves parallel (spin $\frac{1}{2}$, low energy state) or anti-parallel (spin $-\frac{1}{2}$, high energy state) to the magnetic field [29]. These magnetic moments induce a torque proportional to μ and to the B_0 field so that the rotation axis of the protons is not perfectly aligned with B_0, see figure 1.7.

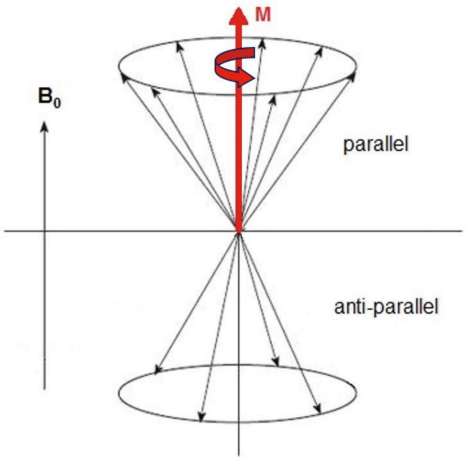

Figure 1.7: Formation of parallel and anti parallel spins due to magnetic field B_0 from [29].

The Larmor frequency describes the precession frequency of protons subjected to a magnetic field B_0. The Larmor frequency is given by the following equations where γ represents the gyromagnetic ratio, equal to $2\pi \times 10^8 \ \frac{\text{rad}}{sT}$.

$$w_0 = \gamma B_0 \tag{1.1}$$

or

$$v_0 = \frac{\gamma}{2\pi} B_0 \tag{1.2}$$

Due to the external magnetic field B_0, the magnetic fields induced by each proton sum up together and create a magnetic moment, called macroscopic magnetisation.

$$\vec{M} = \sum_i \vec{\mu_i} \tag{1.3}$$

The macroscopic magnetisation induces an energy difference between protons with a parallel and anti-parallel spin.

$$\Delta E = h\nu_0 = \frac{h\gamma}{2\pi} B_0 \tag{1.4}$$

The state occupancy between the two states is given with Boltzmann equation.

$$\frac{n_{\frac{1}{2}}}{n_{-\frac{1}{2}}} = e^{\frac{\Delta E}{kT}} = e^{\frac{\Delta h\gamma B_0}{2}} \tag{1.5}$$

Where h, k, T are respectively Planck's constant, Boltzmann's constant, temperature in Kelvins.

1.2.2 Radiofrequency pulses(RF)

Transition between the two states of energy can be done by applying RF pulses with a frequency equal to the frequency of protons such that resonance takes place. There are two main kinds of impulse :

- 90° RF : The number of protons in the low energy state is initially higher. This pulse ensures that there are as many protons in the high and low energy states and that they are all in phase and aligned such that a non-zero resultant in the xy plane is created and the M_z component is equal to zero, see figure 1.8.
- 180° RF : the pulse flips the total net magnetization from $\vec{M}$ to $-\vec{M}$, protons in the lower energy state transit to higher energy state. This is seen on figure 1.8, where $\vec{M}$ was initially in the same orientation as $\vec{B_0}$ and is then flipped.

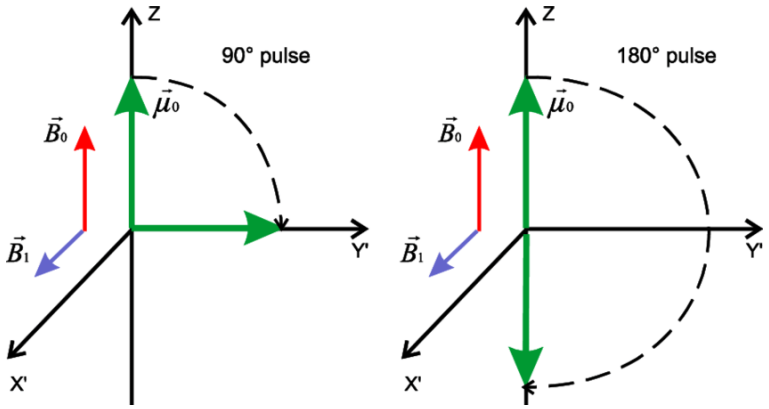

Figure 1.8: Illustration of the system after 90° RF(left) and 180° RF(right) after being initially in resting state , see figure 1.7 from [29].

With a 90° RF, protons return from a high energy state to their initial state. The time needed to pass from one state to the other is called relaxation. We mainly encounter two types of relaxation in the literature, the T1 and T2 relaxation.

T1 relaxation

The longitudinal or T1 relaxation is characterized by an exchange of energy between the spins and the surrounding lattice (spin-lattice relaxation). The recovery of the longitudinal magnetization follows an exponential curve whose recovery rate is characterized by the tissue-specific time constant T1. After the time T1, the longitudinal magnetization returns to 63 % of its final value. The closer the collision frequency between the molecules is to the Larmor frequency, the smaller the T1 is, and the reverse is also true.

T2 relaxation

The transverse relaxation also called spin-spin relaxation is due to the phase shift of the spins. Interactions between the induced magnetic fields of the spins can occur when the spins move together. These magnetic fields then influence the precession velocity of the spins which causes a cumulative phase loss that results in a decrease of the transverse magnetization. The transverse magnetization loss is characterized by an exponential equation characterized by the time constant T2. After a time T2, the transverse magnetization lost 63 % of its initial value. The T2 time varies according to the molecular structure of the tissue and the state of the material, an example is shown in table 1.1. Those changes in relaxation times can be used with MRI sequences in order to differentiate brain tissues.

Brain tissue	T1(ms)	T2(ms)	$\%H_2O$
Gray Matter	900	100	80
White Matter	780	90	72
Fat	260	80	0

Table 1.1: Relaxation times and water importances of brain tissues at 1.5 Tesla, modified from [29].

Bloch equations

The T1 and T2 relaxations are deduced from the Bloch equations when $\beta_z = B_0$ [30].

$$\frac{dM_x(t)}{dt} = \gamma(\vec{M(t)} \times \vec{B(t)})_x - \frac{M_x(t)}{T_2} \tag{1.6}$$

$$\frac{dM_y(t)}{dt} = \gamma((\vec{M(t)} \times \vec{B(t)})_y - \frac{M_y(t)}{T_2} \tag{1.7}$$

$$\frac{dM_z(t)}{dt} = \gamma((\vec{M(t)} \times \vec{B(t)})_z - \frac{M_z(t) - M_0}{T_1} \tag{1.8}$$

$$\tag{1.9}$$

1.2.3 Spin-echo sequence

The spin-echo sequence is one of the first MRI sequences that was introduced. The simplest form of spin-echo sequence consists of a -90° pulse followed by a 180° pulse and an echo. A spin echo sequence is characterised by its echo time (TE) which corresponds to the time between the 90° impulse and the echo. The sequence is also characterised by the repetition time (TR) which corresponds to the time between the two 90° RF impulses. A scheme of the sequence is illustrated on figure 1.9.

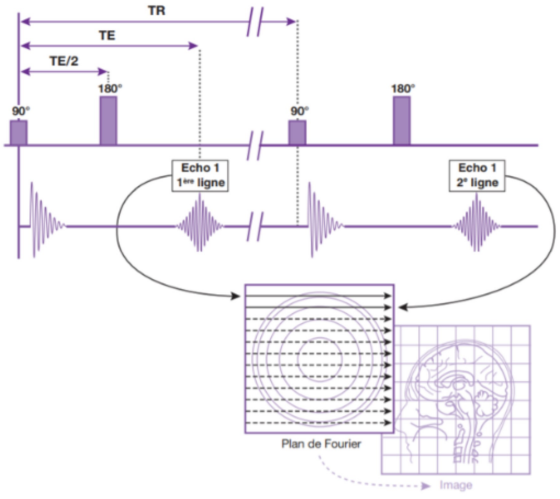

Figure 1.9: Spin echo sequence from [31, 32].

T1 weighted images

The T1-weighted image highlights the differences in T1 relaxation times of tissues. It allows to differentiate anatomical structures on the basis of T1 relaxation times of brain tissues, see table 1.1. In order to have T1 weighted images, both TE and TR should be short.

The choice of a short TR allows that not all protons find their alignment with the main magnetic field, which would display an image with a uniform intensity. The choice of a TR shorter than the recovery time of the tissues also allows to differentiate the different types of tissues. Figure 1.10 shows how to obtain this type of image and what it looks like. "The white matter is represented in light grey, the grey matter in a darker shade of grey and the fluids in black" [32].

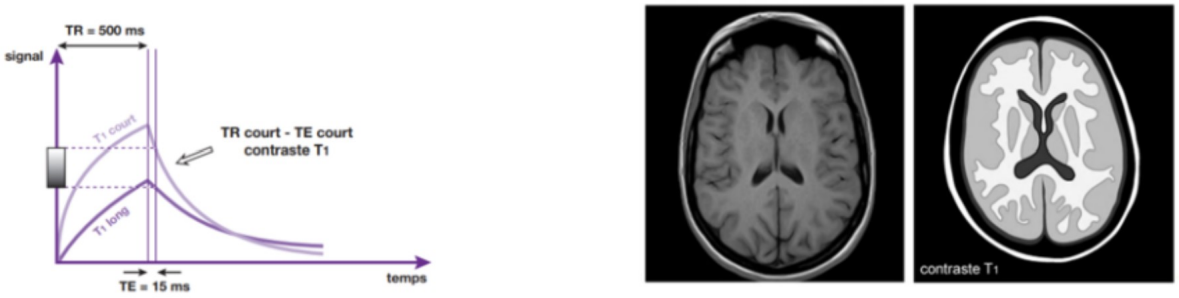

Figure 1.10: Illustration of T1 weighted image recording sequence (left) and obtained result (right), from [31, 32].

T2 weighted images

The T2-weighted image highlights the differences in T2 relaxation times of tissues. This is done by applying a spin-echo sequence with a strong T2 contrast with a long TE and TR. Figure 1.11 shows how to obtain this type of image and what it looks like."The white matter is in dark grey,the grey matter is in light grey and the cerebrospinal fluid is in white" [32].

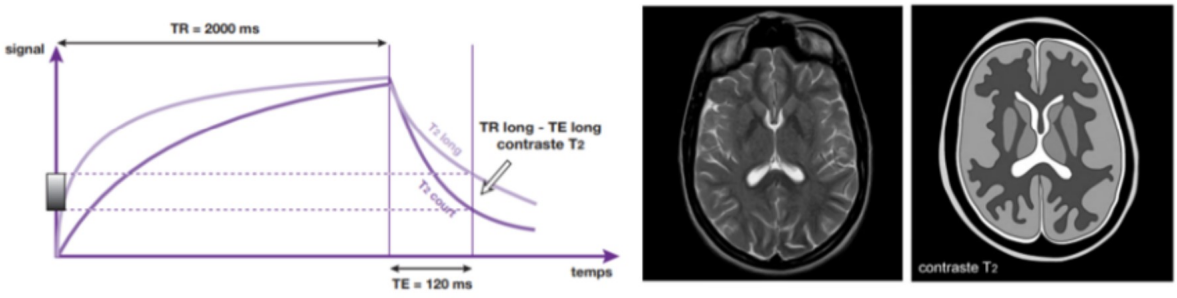

Figure 1.11: Illustration of T1 weighted image recording sequence (left) and obtained result (right), from [31, 32].

1.2.4 Structural MRI features

In the medical field, different features of the brain structures are extracted from MRI images such as the volume, area, mean and standard deviation of the cortical thickness of the brain structures (grey matter). The curvature of brain structures is also studied. To understand this concept of curvature better, let's first explain what the principal curvatures of a surface at a given point are. The principal curvatures c_1 and c_2

are the eigenvalues of the shape operator S at that point, given by the negative derivated of the unit Normal vector field of the surface at that point.

$$cu = S(u)u$$

They measure how the surface curves by different amounts in the main directions at each point of the surface. The mean curvature at a certain point is given by the average of $c_1$and c_2 or by the trace of S at that point. In two dimensions, the mean curvature at a certain point of a line is given by the inverse of the radius length of the osculating circle [33], see Figure 1.12. Gaussian curvature is obtained via the product of c_1 and c_2, the determinant of S a that point. The mean curvature depicts the extrinsic shape, while the Gaussian curvature is an intrisic feature, it quantifies expansion or distortion of the surface. The principal curvature c, at a point on a line is defined as the inverse of the radius of the inscribed circle at that point, see the upper left of figure 1.12. Same happens in 2 dimension but this time with R_1 and R_2, see upper right of figure 1.12.

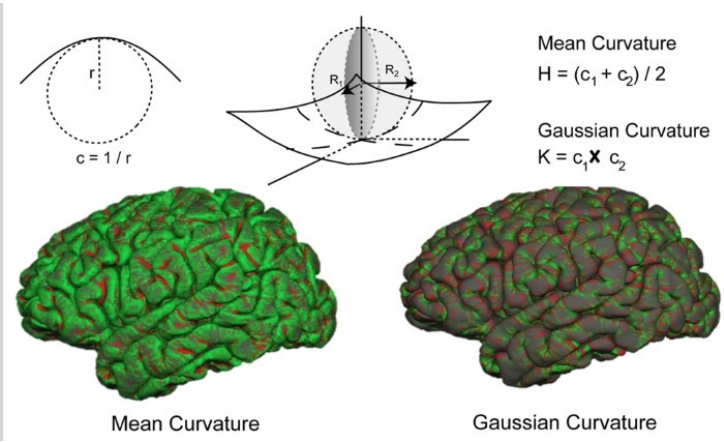

Figure 1.12: Mean curvature and Gaussian curvature illustrated from [33]. Above figures show mean curvature in 2 and 3 dimensions. The below figures illustrate mean and Gaussian curvature obtained on a brain. Positive and negative curvature are drawn in red and green.

Chapter 2

Psychiatric diseases

This section covers the three psychiatric diseases that were studied for our case study, alcoholism, anxiety and depression. Each section begins with a general overview of the disease : its symptoms, the questionnaires to identify it and its links with other psychiatric diseases. Each section also looks at EEG and MRI findings related to these psychiatric diseases.

2.1 Alcohol Use Disorder

This section focuses on the first disease we studied, alcoholism. This section is divided into two subsections, the first focusing on the on the psychiatric side and the second on the link between brain structures and alcoholism.

2.1.1 General Overview

Alcohol use disorder (AUD) is a psychiatric disease which is characterized by harmful drinking patterns leading to negative emotional, physical and social ramifications [34]. AUD is an illness spread overall the world, each year 3 million deaths are due to alcohol on average. In the age group 20–39 years approximately 13.5 % of the total deaths are alcohol-attributable. Excessive alcohol use can damage all organ systems, but it particularly affects the brain [35]. The major causes of alcoholism are not yet known. However, certain factors can increase the risk of alcoholism [36]:

- High weekly consumption: more than 15 drinks for a man, more than 12 drinks for a woman.
- Binge drinking: high consumption once a week (more than 5 drinks in one day).
- Genetics, family history of alcoholism.
- Psychiatric illnesses: depression, anxiety, etc.

There is currently no cure for alcoholism. However, professional treatment in a rehabilitation centre can help people with alcohol use disorders to recover and stay sober. Detoxification is an important first step towards recovery from an AUD, but it often does not lead to long-term abstinence. In fact, the risk of relapse after a detoxification is around 80% within a year [37].

The Alcohol Use Disorders Identification Test(AUDIT) was developed in 1982 by the World Health organisation. Since it was introduced, it has been the most used psychological test to detect alcoholism. The test is composed of 10 multiple choice questions about the frequency of alcohol consumption, drinking patterns and alcohol-related problems. The test can help identify excessive drinking as the cause of the presenting illness [38]. Each question is ranked from 0 to 4, the total score of each subject is then quantified in order to assign a certain level of alcohol use disorder. The most used cutoff for the moment is the following

- Audit score < 8 : Control
- Audit score ≥ 8 : AUD

2.1.2 MRI and AUD [39]

Alcoholism affects a fairly large number of structures in the brain and for the most part, a reduction in volume, area and/or density of white matter or gray matter is observed in these structures. In particular, several studies have found a reduction in gray matter in the cerebral cortex [39, 40]. Alcoholic patients who smoke experience even more extensive gray matter changes, this was for example reported in the prefontal cortex and the anterior cingulate cortex [41].

Regarding cortical atrophy, the most affected areas by alcoholism are the frontal lobes and the posterior cingulate, [42, 43] which are related to memory and attention [44] and also the cerebellum, related to balance and other motor functions, see figure 2.1. Aging does also alter the cortical atrophy, one studied showed that cortical thickness was even smaller for old and young alcoholic with the same consumption of alcohol during their life [45].

The subcortical structures,which are located below the cerebral cortex and are responsible for memory and emotions are also affected by alcoholism. White matter reductions of volumes in the corpus callosum, pons, thalamus, hippocampus, ganglia, putamen have been reported [39].

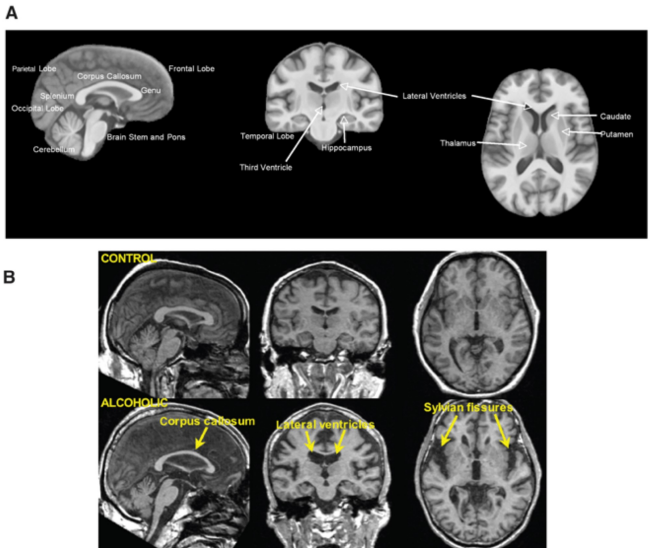

Figure 2.1: Brain structural MRI imaging. First picture shows most impacted regions by alcohol, the second one shows significant differences between AUD and controls [39].

For a more detailed explanation of brain structural imaging and alcoholism, see [32].

2.2 Depression

This section gives a brief overview of the depression disease. It is divided into four parts: the first part deals with psychiatric side of the disease, the second and third with the findings linking EEG, ERP and depression. The last part focuses on links between brain structures and depression.

2.2.1 General overview

Depression is certainly the current major health and economic concern. Depression is a mental illness which affects the way you feel, think and act. Depression is characterized by a high feeling of sadness and desperation [46]. We identify different types of depression such as Major Depression (MDD) and bipolar depression which are not treated the same way. We will mainly focus on major depression in the following part. The lifetime prevalence of major depression is at least 10% with the risk in women twice that in men [47]. Depression is a disease which may be influenced by different factors such as

- Biochemistry : Chemical or biological differences in the brain [48].
- Genetics: Heritabilty is about 40 to 50% in twins studies [49].
- Personality: People who have less self-esteem or people who are more stressed are more likely to be depressed [50].
- Environmental factors: Violence or events can lead to depression [50].
- Other mental diseases : alcoholism [36], anxiety [51].

The correlation between alcohol or other substances consumption and mental illness such as depression is a subject that creates some conflicting evidences. For example, Hanyes *et al.* told in 2005 that high AUD may not be correlated with depression [52]. Plant claimed that the correlation is one sided that after heavy drinking problems, people get often more and more depressive [53]. In addition, alcohol-related problems have been shown to make recovery from depression more difficult [54]. At the inverse, Babor *et al.* in 1997 suggested that the correlation between both was bidirectional, alcohol affects mental wellness and mental wellness affects alcohol consumption [55].

Efficient treatments exist for depression. Contrary to AUD, depression is a disease for which 80 to 90% of subjects response well to treatment and almost all patients finally recover from their illness. Depression can be treated by medication, psychotherapy and by electroconvulsive therapy.

The Beck Depression Inventory (BDI) was first published in 1961 by Aaron T. Beck. The BDI-II was modified in 1996 to be more consistent with the new criteria of depression and is currently one of the most well known tools used to asses depression [56]. The inventory contains 21 self-reported multiple questions. Each answer is ranked from 0 to 3, the total score of each subject is then quantified in order to assign a certain level of depression. The most used cutoffs are the following [56]:

- 1-9 : No or small depression
- 10 -18 : Mild mood disturbance
- 19 -29 : Moderate depression
- $>$ 30 : Severe depression

2.2.2 Resting EEG and Depression

Relations between resting EEG and depression have been widely studied in the literature. The main analytical paradigm studied is the power spectral decomposition of EEG bands. Studies showed that there was a reduction in the delta activity and an increase in the alpha and beta activity for depressed subjects [57, 58]. New studies focusing on the gamma band suggest that

- There are differences between major depressed subjects and controls. A study observed that there was a reduction in the gamma power spectral density for MDD in the anterior cingulate cortex [59].
- Medications for depression, such as ketamine could alter the gamma band [60].
- There are differences between unipolar and bipolar depression. Some studied using magnetoencephalograph (MEG) showed that the gamma power spectral density was superior for MDD in the whole brain after an auditive stimulus in comparison with subjects suffering from bipolar depression [61].

Some studies tried to classify MDD thanks to resting state EEG data. A report studying correlation between resting state EEG and depression for almost 100 subjects achieved 70% sensitivity and 76% specificity using power spectral density (PSD) of the bands of the EEG signals. They first used Linear Discriminant Analysis (LDA) to change the initial feature space (reduction of dimension) and then the Genetic algorithm was used to find the most important features. Decision trees were then used to classify depressed from non-depressed subjects [62]. Another research using resting state EEG achieves 97 % of accuracy in order to detect control subjects and MDD. The different features extracted from the EEG signals were the power spectral density and EEG alpha interhemispheric. The most important characteristics were selected by assigning a weight to each characteristic that was more significant if it could separate each class on its own. Different classifier models were used such as Logistic Regression, Support Vector machine or Naïve Bayesian [63].

2.2.3 ERP and depression

A research studying the P300 wave elicited after an oddball paradigm for 17 melancholic, drug-free depressed patients, 22 controls and 13 patients after recovery from MDD, showed that there was no significant difference for the latency between depressed, controls and for subjects after recovery. At the inverse, there was a significant difference in the amplitude of the P300 wave. They claimed that the P300 amplitude for MDD subjects was smaller compared to controls and recovered patients, was negatively correlated with the severity of depression and kept increasing after recovery [64].

Another study using an auditory oddball paradigm to assess significant differences between P200 and P300 waves between MDD and control subjects found that the latency of P300 waves of MDD was significantly higher as the amplitude of the P200 wave [65].

Features of P300 waves such as amplitudes and latencies of the Fz, Cz and Pz electrodes have also been used in order to classify MDD and controls. By using t-test to take most significant features into account and by using logistic regression, the highest classification accuracy achieved was 85.1% [66]. A last study using P300 features with an auditory oddball task using more than 100 subjects obtained 73% of accuracy using logistic regression in order to classify MDD and and post traumatic disorder patients and healthy controls [67].

2.2.4 Structural MRI and depression [68]

Researchers who studied depression and its correlation with brain structure noted an overall decrease in the volume of gyrus material in the frontal lobe and particularly in the orbitofrontal cortex [68, 69, 70]. Other studies also agreed that there was a reduction in the volume of the temporal lobe, the hippocampus, the gyrus and the left posterior cingulate for depressed subjects [71, 72]. An Enigma mass study including 1728 MDD and 7956 control patients was performed to analyze differences in subcortical volumes and cortical thickness. Depressed subjects did not show significant differences with control adolescents in subcortical volumes. Adult patients showed a reduction of the cortical thickness of the orbitofrontal cortex, the anterior and posterior cingulate cortex, the insula, and the temporal lobe. It was also observed for these patients that their brain appeared on average more elevated and that this was even amplified when they were taking antidepressants [43]. A better overview of the cortical thickness changes (in yellow) for MDD is shown on figure 2.2.

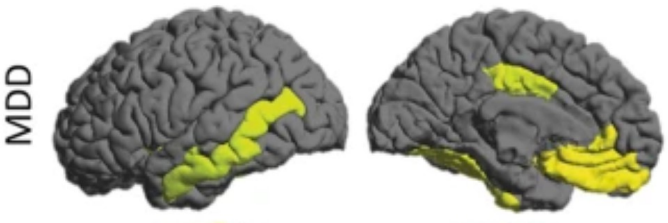

Figure 2.2: Cortical thickness of Orbitofrontal, cingulate cortex and insula, and temporal lobe are the most impacted regions for MDD, from [43].

Reductions in subcortical volumes such as the putamen and basal ganglia are also observed, these are problems also observed for alcoholic patients [73].

2.3 Anxiety

The last psychiatric illness we looked at is anxiety. In this section, we first focus on the different types of anxiety that exist, the risks that can induce anxiety and the questionnaires that can assess anxiety. We then look at the studies linking resting EEG, structural MRI and anxiety.

2.3.1 General overview

Anxiety disorder is one of the most psychiatric illnesses spread in the world. It affects more than 40 millions adults in the US each year. Despite being one of the most widely spread mental illnesses in the world, it is also certainly one which is the most misunderstood as only 27.6% of people suffering from anxiety gets a treatment for it [74]. Anxiety can be decomposed in five different classes [75]:

- Generalized Anxiety Disorder (GAD) is characterized by exaggerated worry.
- Obsessive-Compulsive Disorder (OCD) is a disease which is characterized by compulsive behavior which can be observed which different tests such as the OCDS obsessions.
- Panic disorder is characterized by repeated panic attacks and fear moments.
- Post Traumatic Stress Disorder (PTSD) is an an anxiety illness which can occur after some traumatic events such as personal assaults, moment of excess fear, etc..
- Social anxiety disorder, also called social phobia, is characterised by strong anxiety and excessive self-consciousness.

Several factors can increase the risk to develop anxiety :

- Personality [76]: people who are more stressed are more likely to be depressed.
- Psychiatric illnesses: alcoholism [36], depression [51]. Withdrawal can cause or worsen anxiety.
- Stress [75]: induced by an illness or an event.

Anxiety is mainly treated in two different ways. The first way, psychotherapy, is to work on these anxiety disorders. For this, cognitive behavioural therapy is often used. If therapy is not effective, or as a supplement, medication is also used to treat anxiety.

The State-Trait Anxiety Inventory (STAI) was introduced by psychologist Charles Spielberger in 1983 [77]. The STAI is a multiple-choice questionnaire with each item ranging from 1 to 4, with the total score ranging from 20 to 80. A cutoff score of 40 is generally used to define likely clinical levels of anxiety [78]. The STAI can measure two different types of anxiety, state anxiety and trait anxiety. The STAI-state assesses the patient's current state of anxiety, using items that measure subjective feelings of apprehension, tension, nervousness and worry. The STAI-trait instead assesses patients' self-confidence, calmness and safety.

2.3.2 Resting EEG and anxiety

There are few studies which focused on the correlation between the power spectral density of EEG bands and generalised anxiety disorder. One of those researches found that the power spectral density of the gamma bands in the posterior sites was superior for anxious subjects compared to controls and suggested that it could have a correlation with negative affect of those people [79].

Another study working on patients with and without panic attacks and more precisely with agoraphobic disorders found that agoraphobic anxious subjects had a reduction in the power spectral density (PSD) of the alpha and beta waves in the right hemisphere. Plus, they also emphasized that anxious subjects had a significant increase of the PSD of the theta wave in the temporal areas of the right hemisphere [80]. A study counting 52 patients suffering from panic disorder and 104 controls investigates the correlation with overall PSD of subjects and anxiety. Their major finding was that the overall power spectral density of panic disorder subjects was lower compared to controls and this mostly when eyes are closed [81].

A study using SVM with different kernels such as Gaussian kernel and polynomial kernel obtained an average of 83% of accuracy by using as features the PSD of the EEG bands. To increase the number of subjects, they cut the EEG signal of each subject in different signal and consider each segment as a different subject such that they obtain more than 5000 patients [82] (maybe overfitting as soon as each patient is similar to itself). Given the noise of the EEG signal, by slicing the signal into 2-second bands, the spectral density of each band of each electrode may be completely different from one slice of the signal to another. For clinical purpose, it is then difficult to choose a slice of the signal that completely characterizes a patient.

2.3.3 Structural MRI and anxiety [83]

The main studies using MRI structural to study PTSD have mainly found differences in the hippocampus, which is thought to be the origin of posttraumatic symptoms [83, 84, 85]. However, this is not the case for all studies and this could be due to the fact that these changes are very fine and therefore very complicated to detect [86]. A more advanced study comparing hippocampal volume reduction for PTSD can be found in [87]. Other studies also noted reductions of the corpus callopsum [88] volume, which may involves cognitive impairments [68].

The main studies concerning OCDS noted a reduction in the orbitofrontal cortex of the left and right hemisphere [83, 89, 90]. Another study showed that people with OCDS had increased white matter volume compared to controls and that the volume of the right and left thalamus was greater [91], but most of the studies did not agree on regions other than the orbitofrontal cortex [83].

Concerning panic disorder, a reduction of the volume in the left temporal lobe, the hippocampus, the gyrus and the putamen was detected [83, 92, 93, 94]. It was also suggested that the duration of the disease was positively correlated with the volume of the left hippocampus [92].

Concerning GAD, studies were carried out indicating differences in structures related to emotions and social behaviors such as the amygdala, the prefrontal cortex and the temporal areas [95, 96]. A summary of the most impacted brain regions by GAD are reported on figure 2.3.

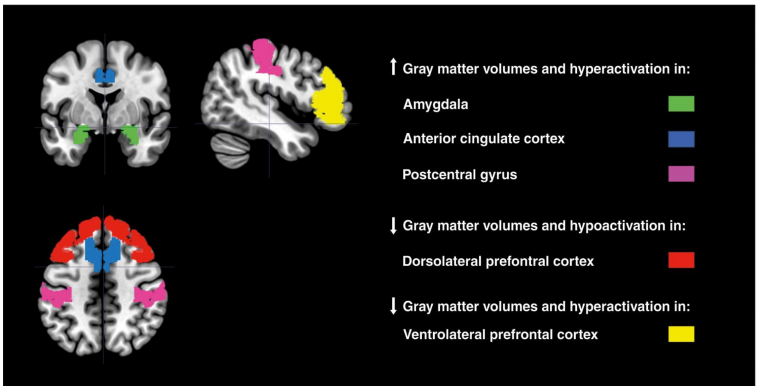

Figure 2.3: GAD mainly affects volumes of the amygdal, anterior cingulate cortex, postcentral gyrus, dorsolateral and ventrolateral prefrontal cortex. [95].

It is interesting to see that reductions in the thalamus, hippocampus, putamen and ganglia are observed for alcoholic and anxious patients which show that those two diseases are also correlated in medical symptoms. Moreover, it is important to note that different types of anxiety do not affect the same brain regions even if they are classified the same way by questionnaires.

A study using Markov boundary selection algorithm in order to find most important features with 561 subjects [97] and using a SVM algorithm in order to classify post traumatic stressed soldiers obtained an area under curve in the range [85-91]% using symptoms scores as the criteria features.[1] This shows that even with symptoms scores and questionnaires scores, the identification of post traumatic (anxious) patients is not straightforward.

[1] For more information of Markov boundary selection, see [98].

Part II

Prediction of alcoholism, depression and anxiety from EEG and MRI data: A case study

Chapter 3

Methods

This chapter outlines the different steps we have taken during this work. Figure 3.1 summarizes the different steps required to predict alcoholism, depression and/or anxiety from the EEG and MRI data. First data should be recorded, the EEG and MRI signal should then be preprocessed. From these preprocessed signals, features can be extracted. From these features, machine learning algorithms could be applied to predict if someone is alcoholic or not, depressed or not, anxious or not (only binary problems). This is explained in further details in the machine learning section. The strategies section is strongly linked to the machine learning section as it explains how the machine learning algorithms have been applied on our data.

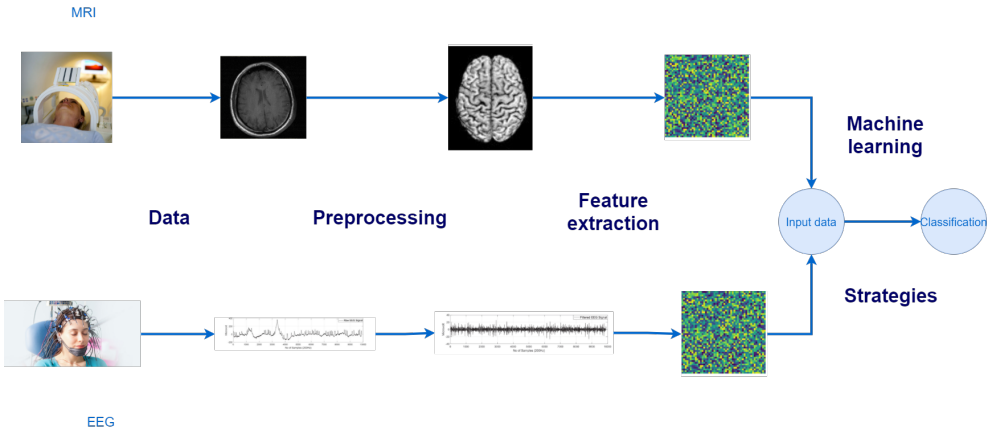

Figure 3.1: Pipeline of the project separated in 5 steps: data acquisition, preprocessing, features extraction, machine learning and strategies.

3.1 Data

To understand what kind of data were used to perform our experiment, it is first necessary to understand how the collection of MRI and EEG data is organized during the withdrawal of alcoholics in Belgium and more specifically at the St-Luc hospital in Brussels. This collection is carried out in three distinct periods.

First, the T1 period takes place during the first days of the alcoholics' internment. During this period, the alcoholic patients have consultations with a hospital psychiatrist, answer questionnaires about their state of depression, anxiety and craving. Most of these patients also receive EEG examinations, as this is the only clinical brain examination reimbursed by the National Institute for Health and Disability Insurance (INAMI). Some patients also undergo MRI examinations, which are diverse (see figure 3.2) : structural Magnetic Resonance Imaging (sMRI), Diffusion Magnetic Resonance Imaging (dMRI) and functional Magnetic Resonance Imaging (fMRI).

Then, when patients have completed their withdrawal, they can leave the hospital and go home. This is the T2 period. At the end of the 20 days of withdrawal, a new clinical examination is often performed,

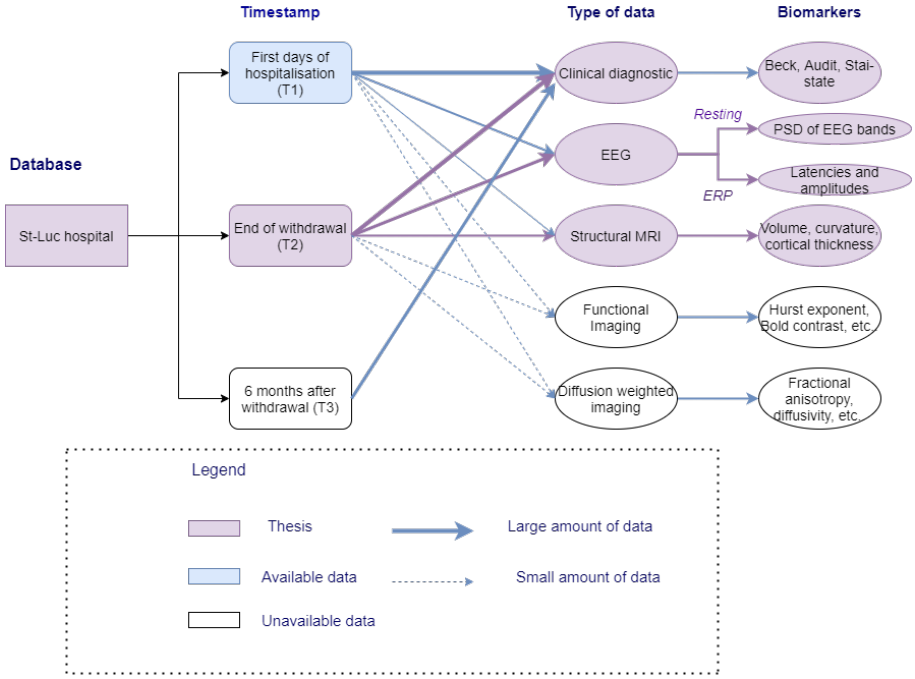

Figure 3.2: Type of data at disposal for the moment at St-Luc hospital. Data is mainly available for sMRI and EEG in T1 and T2.

adapted to each individual case. Patients with alcohol use disorders are asked to complete a number of questionnaires. At this point of treatment, alcohol-dependent patients may still receive for the most part clinical EEG examinations and for some MRI examinations at a certain frequency, see figure 3.2.

Finally, the T3 period occurs six months after their hospital stay. At this period, electrophysiology or MRI medical examinations may be carried out. To date, very few examinations of this kind have been performed.

For this work, only data collected in T2 will be used. Two different datasets were used to carry out our experiments and analyze whether from EEG or MRI data we could predict psychiatric diseases :

- **MRI-based dataset** : This dataset includes 210 subjects, 141 alcoholics at the end of their withdrawal (14 to 18 days) and 69 control subjects for who structural MRI data were obtained between the years 2015 and 2018 at St-Luc hospital.
- **EEG-based dataset** : This dataset includes 42 alcoholic patients at the end of their withdrawal (14 to 18 days). Resting state EEG data and event related potentials were recorded: P50, P200, P300 and CNV. Data was obtained between the years 2016 and 2020 at St-Luc hospital.

3.1.1 Group segmentation and problem statement

For our study, we define four possible conditions or states

- healthy or control subjects.
- Subjects with alcohol use disorder based on their alcohol consumption, more than 500g of alcohol per week on average.
- Subjects in anxiety state if their STAI score ≥ 50.
- Subjects in depressive state if their Beck score ≥ 8.

Note that combinations of these states are possible. E.g., a patient suffering from AUD can be in a depressed state or not or can be both in a depressed and an anxious state simultaneously.

From there we define five machine learning prediction tasks

1. AUD vs controls (MRI dataset)
2. Anxiety state vs non-anxious (MRI dataset)
3. Depressive state vs non-depressed (MRI dataset) [1]
4. AUD in an anxiety state vs AUD not in anxiety state (MRI and EEG datasets)
5. AUD in depressive state vs AUD not in depressive state (MRI and EEG datasets)

Machine learning algorithms with structural MRI features were performed to resolve problems 1 up to 5. As EEG signals that were recorded concerned only AUD patients, classifications with EEG features tried to answer problems 4 and 5.

3.1.2 MRI-based dataset

A total of 210 subjects were recruited to conduct this study. 141 alcoholic patients were recruited from the full-time alcohol detoxification unit of the Cliniques Universitaires St-Luc, Brussels, Belgium. The patients had to abstain for alcohol for at least 14 to 18 days (T2). The criteria for inclusion were the following [99] :

- Meet the Diagnostic and Statistical Manual of Mental Disorders criteria for alcoholism, see [100].
- Be able to speak French.
- Consume more than 500g of alcohol per week before being hospitalized.
- Providing written consent for participation before engaging in the experimental procedures.

69 control subjects were also recruited to realise the experiment. The criteria for inclusion were the following [99] :

- Be able to speak French.
- A reasonable drinking consummation, i.e., less than 14 standard drinks a week for a woman and 21 standard drinks a week for a man.

For each subject, the following parameters were collected: age (in years), gender, BECK and STAI state score.

For the structural MRI data, 3D T1 weighted images were recorded on a 3 T Achieva Philips scanner(Philips Healthcare) with a 32-channel phased array head coil. Data was acquired through a spin-echo sequence with an inversion prepulse (Turbo Field Echo) acquired in the sagittal plane (it is an anatomical plane which divides the body into right and left parts). The spin echo sequence was performed using the following parameters [99]:

- TR = 9.1 ms, TE = 4.6 ms
- flip angle = 8°
- 150 slices; slice thickness = 1 mm
- in-plane resolution = 0.81 × 0.95 mm^2 (reconstructed in 0.75 × 0.75 mm^2)
- FOV 220 × 197 mm^2, acquisition matrix = 296 × 247 (reconstruction 320× 320)
- SENSE factor = 1.5 (parallel imaging)

[1] One could say that it is more interesting to compare depressed patients to control patients but this would not be interesting because it would almost be like comparing alcoholic patients to controls given our dataset. It is the same for anxiety, see figure 3.3.

The following figures give a better view of how the data is distributed.

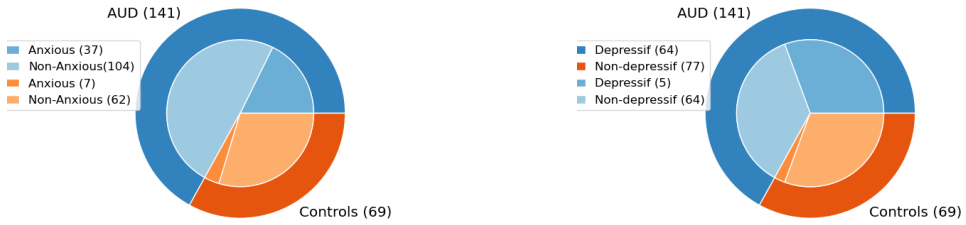

Figure 3.3: Data distribution of the MRI-dataset showing the correlation between AUD and anxious subjects (left) correlation between AUD and depressed subjects (right).

As shown on figure 3.3, most people who are in an anxious state are also alcoholic, most depressed patients are also alcoholic but at a lower level. Almost half of the alcoholic patients are in a depressive state, which shows to what extent these two disorders are correlated. It is not shown in the above figure, but almost all patients who are in anxious state are also in a depressive state, 39 of the 44 anxious subjects.

3.1.3 EEG-based dataset

The electroencephalogram data was acquired from 42 subjects suffering from AUD from the full-time alcohol detoxification unit of the Cliniques Universitaires St-Luc, Brussels, Belgium. Participants had abstained from alcohol consumption for 14 to 18 days before being enrolled in the study. The electroencephalogram data was acquired with the program EEVOKE 3.1.5 a the hospital between 2016 and 2020. The total number of patients who were interned during this period for alcohol problem is higher, but EEG signals were not collected for all patients. Most of them got EEG examinations at the end of their withdrawal, but all clinical diagnostics were not available. The EEG tests collected are varied: resting state EEG and ERP components such as P50, P200, P300 and CNV. The characteristics of the amplificator used to record the EEG signals are the following :

- Type : Refa 8 Clinical.
- 23 electrodes (see figure 1.3 for their positions):
  - 19 basic electrodes.
  - Reference electrodes : A1 and A2.
  - Fpz and Oz.
- Sampling frequency : 256 Hz

To conduct this study, we had initially EEG data for 42 patients. For each patient, data was collected on a 2-day time period. Some EEG signals were not available for each patient. Some of them got only resting state EEG or event related potentials. For huge dataset, when some features are unavailable for certain subjects, those features are often replaced by the mean values of the available data or by other techniques. This was not possible in our case as the dataset was really small. By replacing some missing values by the mean values of the population, it could strongly bias our data. The best way to use the largest amount of data was to thus to split our initial dataset in two :

- A resting EEG-based dataset using features of the resting EEG signal.
- An ERP-based dataset using CNV, P50, P200 and P300 features.

Resting EEG-based dataset

The resting EEG-based dataset contains 36 AUD patients, 12 in a depressive state and 7 in anxious state. On the first day of the recording, the patients got 10 min of resting EEG with different periods of eyes open and closed. The Eyes Open protocol lasted 4 minutes as the Eyes Closed protocol. They also got 3 min of recording during hyperventilation and visual flashs in order to detect epilepsy.

ERP-based dataset

The ERP-based dataset contains 38 AUD patients, 6 in an anxious state and 16 in an depressive state. On the second day of the recording, the patients got 3 different types of ERP examinations.

First, electroencephalograms were recorded while an oddball stimulus was presented to the patients in order to extract the P200 and P300 waves. Two types of tones were presented, with varying times and probabilities and the patient was asked to respond directly after hearing the rare tone. A richer description of the parameters of the experience is presented at table 3.1.

Tone	Form	Frequency	Duration	Rise,fall time	Probability
Frequent	sine wave	1000Hz	60 ms	10 ms	0.8
Rare	sine wave	2000Hz	60 ms	10 ms	0.2

Table 3.1: Description of the experimentation in order to trigger the P300 component.

CNV was ellicited with a S1-S2 paradigm. The CNV paradigm consists first of an auditory beep (S1), which warns the subject. The warning stimulus is followed by an auditory beep, the imperative signal(S2). The subject has to respond as quickly as possible to this beep by pressing a button. The auditory beep lasted 60 ms and had a frequency of 1000Hz. The duration of the S2 flash was 1000ms.

The P50 wave was extracted via two paired-click stimulation with an interstimulus interval (ISI) of 300ms between the S1 and S2 clicks. Time between two experiences was about 3200 ms. S1 and S2 clicks had the same properties : they lasted about 5.6 ms, the frequency was 1500 Hz and the rise/fall time was 1 ms. The total recording time was 5:20 minutes for about 180 stimuli in total.

3.2 Preprocessing

The first step realised after data recording is preprocessing, see figure 3.1. This step consists in transforming raw data to a smoother signal from which we can extract interesting features. This section is separated in three parts : the preproccesing of rEEG, ERP and MRI signals.

3.2.1 Preprocessing of Resting EEG signals

The preprocessing of EEG signals is the simple procedure of transforming raw EEG data to a more suitable format in order to be able to extract biomedical features from the signal. The EEG signals contain a lot of noise and should thus be filtered. Some bad channels are sometimes also not well recorded and should be removed. The preprocessing of resting state EEG signals is similar to the one applied for event related potentials.

Filtering

A finite impulse response high pass filter at 0.5 Hz on `eeglab` [101] was first applied to the raw rEEG to remove low frequency noise. A low pass filter was also applied to remove high frequency noise and to remove frequencies greater than the Nyquist limit.

The electrical circuits surrounding the electrodes create some noise around 50 Hz. This is illustrated on figure 3.4, which displays the power spectral density of resting EEG signals. A peak around 50 Hz is presented and should be discarded [102]. This peak was discarded by applying a Notch filter to the signals between 49 and 51 Hz [102]. The effect of the filter can be seen on next figure.

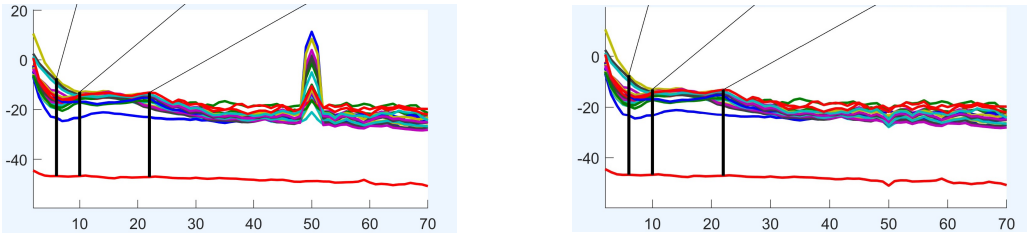

Figure 3.4: Power spectral density of the recorded EEG signals at the 23 electrodes(each corresponds to a different color) before (left) and after (right) applying a Notch filter where the peak around 50Hz is then deleted.

Removing bad channels

Bad channels are channels which do not provide accurate information. It is really important to detect and to delete them in order to be able to interpret extracted features further. Bad channels are due to different problems :

- Some channels are malfunctioned initially.
- The electrodes were not properly placed or were not in contact with the scalp.
- Electrodes can get saturated when electrodes are wet.

Bad channels can be discarded by inspection and this was realised for this study. Some numerical techniques are also used on `eeglab` in order to reject bad channels. Bad channels are removed if the signal of one electrode flats more than 5 second or if the correlation signal recorded by one electrode with is nearby channels is lower than 0.8 in absolute values. Bad channels are also removed if the standard deviation of the signal is higher than 4 μV [101]. After detection, bad channels were interpolated by spherical splines. The method which is proposed on `eeglab` and `mne` [103] consists of 3 different steps [104] :

1. The channel locations are projected on a sphere, based on the positions proposed on figure 1.3.
2. Compute a matrix correlation between the *good* and *bad* electrodes.
3. Finally, the *bad* channels are interpolated via strongly correlated *good* channels.

Artefact correction

Artefacts are elements of the signals that are recorded by the electroencephalograms, but do not actually originate from the brain, see figure 3.5 for an illustration. In the literature, two types of artefacts are often reported [5] :

- Environmental artefacts are due to outside-world interference, electrical noise or electrodes losing contact. Those effects can be minimized by adjusting the environment, f.e. shielding the room and properly secure the electrodes.
- Biological artefacts find their sources from the body. Those artefacts are mainly due to movements and eye blinking.

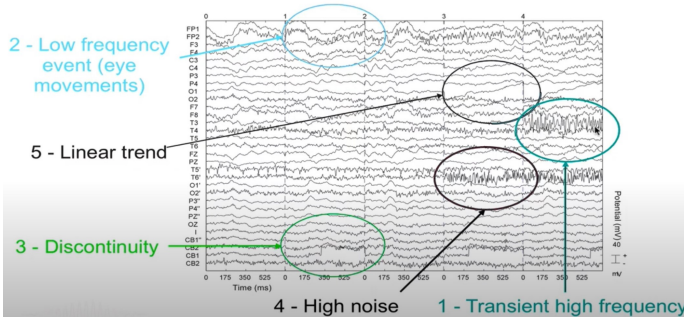

Figure 3.5: Type of artefacts presented on eeglab. Linear trend, discontinuity and transient high frequency are often due to environmental artefacts or electrodes malfunction, while low frequency and high frequency artefacts are respectively due to eye blinking and muscular movements.

Independent component analysis (ICA) was used to remove such artefacts [105, 106]. ICA is a signal processing method which aims at separating independent sources which are mixed in the different electrodes. By this way, ICA can separate the artefacts as they are supposed to be independent [105, 106].

3.2.2 Preprocessing of Event-related-potiential

The preprocessing of the P300 and the P200 component was realised via `erplab` [107], an extension of `eeglab`. The filtering of the EEG signals in order to extract ERP components, is similar to that applied for the resting EEG signals. A high pass filter was also applied at 0.5 Hz. No Notch filter was applied as a low pass filter was applied at 30 Hz.

Extraction of events

For event related potentials, we only focused on the cognitive reactions of patients due to a certain stimulus. To do so, we must extract data epochs time-locked to events of interest, for example for the oddball paradigm, frequent and rare tones. The epoch limits were from 100 ms before the event to 800 ms after the time-locking event. A mean baseline value was removed for each data epoch to counter low-frequency drifts [107]. Event related potentials are then created by averaging epochs, ignoring those with artefacts.

Artefacts detection

The best method on `erplab` to detect artefacts within a data epoch is called moving window peak-to-peak threshold and enables to detect *bad* epochs. The method finds the maximum peak-to-peak voltage, the difference between the maximum and minimum value of the signal, within a moving window that slides across the epoch. This algorithm takes different parameters as input such as [107] :

- The threshold voltage, initially set to 100 μV : if the peak to peak difference of the signal recorded by an electrode is higher than this threshold, then the epoch is suppressed.
- The moving window : peak to peak is not calculated on the whole data epoch but on on small periods of chosen length, called moving window.
- Window step : the algorithm first finds the peak of the first window and then shifts the window by a certain amount, the window step.

This method is really efficient to find artefacts for which the voltage changes really fast such as eye blinking. At the inverse is not really reliable in order to highlight high noise or linear trends that are presented on figure 3.5.

3.2.3 Preprocessing of MRI data

The preprocessing of the MRI images was performed by L. Dricot and will be briefly explained in this section. The preprocessing contains 5 major steps on `freesurfer` [108]. First, an affine registration is performed between the 3D MRI image and the MNI305 space [109, 110], which is assumed to be insensitive to brain diseases. This step consists in bringing these two images into alignment. Second, a first segmentation of the brain structures is performed. The next step consists in correcting the magnetic field intensity B_0. Then, a non-linear alignment with the MNI305 atlas is performed. Finally, the atlas is labeled using different atlases.

The reference atlases that were used, were initially made from images labeled by hand by brain experts indicating for each voxel of the MRI images to which brain structure it belonged. Each image was then projected in the same space to obtain a voxel by voxel correspondence for all subjects. The labeling of each point in the space to a certain brain structure is done by finding the segmentation that maximizes the probability that this voxel belongs to a certain structure knowing the probabilities that each voxel belongs to each structure in the reference atlases.

For our case study, three atlases were used in order to create the structures of interest :

- Destrieux atlas (see appendix A.1) which uses 74 brain areas [111] .
- Desikan atlas (see appendix A.2) which uses 34 brain areas [112].
- Brodmann atlas [113], using only 12 areas (BA1, BA2, BA2, BA3a, BA3b, BA4a, BA4a, BA4p, BA6, BA44, BA45, V1, V2, MT).

3.3 Features extraction

When the EEG and MRI signals are preprocessed, we can then extract features from the signals. This section is separated in three parts explaining how features were extracted from the resting EEG, the ERP and from the MRI signals. It lasts with a summary of each extracted feature for each signal.

3.3.1 Resting state EEG features

For the resting state EEG, the power spectral density for each electrode and each band (*delta*, *theta*, *alpha*, *beta*, *gamma*) has been calculated for eyes open (EO) and eyes closed (EC). The power spectrum gives information about the magnitude of the signal at each frequency. It can thus be used in order to assess activity of the signal at certain frequencies, but it gives no information about the phase of the signal. The

basic idea to calculate the power spectral density (PSD) of a temporal signal x of length N, is to calculate its Fast Fourrier Transform (FFT) and then its magnitude :

$$PSD(f) = \frac{1}{N}\sum_{n=0}^{N-1} |x_n e^{\frac{-j2\pi fn}{N}}|^2 \tag{3.1}$$

However the EEG signal is a non-stationary signal. Its spectral content varies constantly due to the neural activity taking place under the scalp, so there is very little chance that a long EEG signal can be decomposed as a sum of sinusoids. The method presented at equation 3.1 works well for stationary signals, but is biased for non-stationary signals. Averaging the power spectral density obtained with 3.1 over short segments over the original long EEG allows to reduce the bias, this is known as Welch's method [114][2]. This comes at the cost, however, of a lower frequency resolution f_{res} as the frequency resolution is defined by:

$$f_{res} = \frac{f_s}{N} = \frac{f_s}{f_s T} = \frac{1}{T} \tag{3.2}$$

where f_s and T are the sampling frequency and the temporal length of the signal. The most common approach is to take a window length which contains at least two full cycles of the lowest frequency of interest, as in our case it was 1 Hz we choose a window size of $\frac{2}{1} = 2$ seconds [115]. PSDs of the different bands have been calculated during EO and EC periods . Significant differences were found in PSD of the alpha band as it was suggested in the literature [116]. Indeed, Welch's t-test was performed to test the hypothesis that the means of the PSD of the EEG signals were the same during EC and EO. The sample t-test was calculated with the library `scipy` [117]. It highlighted that the Welch PSDs of the alpha band for EC and EO in the P3, Cz, C4, T6, O1, O2, P4, F8, T4, F7 and F4 (see figure 1.3 for a better scheme of the placement of the electrodes) were statistically different ($\alpha < 0.05$).

3.3.2 ERP features

The P300 and P200 components were observed after an oddball paradigm. The P200 component after an oddball paradigm is typically observed 150 to 250 ms after the auditory stimulus. The P300 component is observed 300 to 600 ms after the auditory stimulus. The absolute latency, fraction peak latency and amplitude of those components were obtained via `erplab` at the Fz, Cz and Pz electrodes. Amplitudes and latencies were measured including the local peak option. This option prevents rising edge of a component near to the edge of the measurement window from being selected as the peak. The latency is the time interval between the exposition of the stimulus and the peak of the EEG signal. It is expressed in milliseconds and it represents the time needed for the brain to generate the ERP component. The amplitude represents the maximal peak (negative or positive) that was recorded. It is expressed in μV and it gives information about the number of neurons that were stimulated and its activation synchrony during the component generation. The fractional peak latency is useful to register the onset of a component. Because of the variability in latency across trials, the start time of an average waveform will be significantly earlier than the average start time of individual trials, and the 50% maximum amplitude point may be more representative of the average start time of an individual trial [107].

The P50 features, the peak amplitude and the peak latency, were extracted at the electrodes T3, Cz and T4 after the first click stimulus (S1) and after the second click stimulus (S2). Amplitudes at the Fz, Cz and Pz channels of the M1 (500 - 700 ms post S1), M2(200 - 0 ms pre S2), PINV (500-700 ms post S2) were extracted from the CNV component. `ANT Neuro` [118][3], the company which installed and sold the electroencephalograms to St-Luc, also sold a program which extracts features of the ERP waves and stores them in pdf and/or doc files. In order to read those files in python, the packages `pypdf2` [119] and `docx2txt` [120] were used. An overview of those reports is shown at figure 3.6.

[2]Welch's method was applied on `mne` for our case study.

[3]The preprocessing that was performed via `ANT Neuro` in order to extract the P50 and CNV components was similar than the one done on `eeglab`. For a better overview see [118].

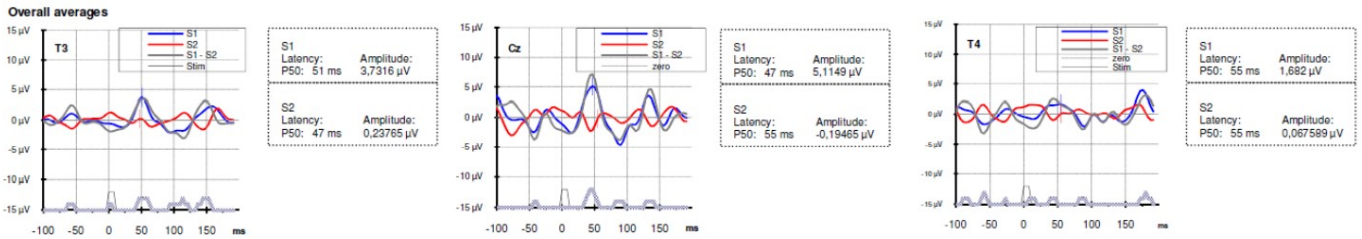

Figure 3.6: Report from `ANT Neuro` for the P50 wave, from [118].

3.3.3 MRI features

When the preprocessing was over, `freesurfer` rendered the following information from the MRI images for each atlas:

- Volume of brain structures
- Area of brain structures
- Mean and standard deviation of cortical thickness of brain structures
- Curvature of brain structures

Age has been shown to influence the brain structure [121, 122]. it was advised to correct this effect before using the MRI data for machine learning algorithms. Laurence Dricot proposed to calculate the linear model linking the different MRI metrics and age. Then, for each metric of each patient, calculate the prediction via the linear model linking each metric and each patient. Finally, calculate the residual = prediction - initial metric and keep this residual value as MRI metric. In this way, the MRI metrics are no longer correlated with age, but a lot of information about the MRI data is lost. This transformation was initially perform but the results obtained with the machine learning algorithms were not conclusive such that this method was finally abandoned.

3.3.4 Summary of extracted features

The following table summarises the features that were extracted for each clinical examination:

Medical Examination	Features	Position	Software (preprocess)	Software (features calculation)
Resting State EEG	Welch Power Spectral Density for EC and EO	23 electrodes	`eeglab` (`Matlab`)	`mne` (`Python`)
Paired clicked paradigm	Amplitude and latency of P50 wave after S1 and S2	T3, Cz, T4	`ANT Neuro`	`ANT Neuro`
Oddball Paradigm	Amplitude, fraction peak latency and latency of P200 and P300 wave after frequent and rare event	Fz, Cz, Pz	`erplab` and `erplab` (`Matlab`)	`erplab`
Standard reaction time paradigm	Amplitudes of M1, M2, PINV	Fz, Cz, Pz	`ANT Neuro`	`ANT Neuro`
Structural MRI	Volume, Area, cortical thickness mean and gaussian curvature, standard deviation of brain structures	Brain structures relative to Desikan, Destrieux and Brodman atlas	`freesurfer`	`freesurfer`

Table 3.2: Table regrouping the features from different clinical exams.

3.4 Machine learning methods

How can alcoholism, anxiety and/or depression influence the EEG and structural MRI data collected from subjects?

In this section, we try to answer this question using machine learning methods. We try to create a predictor function from EEG or sMRI data that would predict the subject's class introduced in section 3.1.1 :

$$f : x \to y \in \{\pm 1\}$$

Where x represents the structural MRI or EEG data and y the class (disease) of the subjects. y will be equal to 1 if a subject is ill (depressive, alcoholic, anxious) and to -1 if he is not. For notational purposes, we will define by $X \in \mathcal{R}^{N \times m}$ the feature matrix, where N and m are respectively the total number of samples and features. We will denote by x^i the i^{th} observation and by X_k the k-th column of X, which represents a single feature.

In this part, we start by defining in a global way the different machine learning methods we have used : support vector machine, logistic regression, linear discriminant analysis and random forest. Next, we focus on the transformations that need to be performed on our data for our machine learning algorithms to work properly : standardisation, correlation analysis and feature selection. We conclude by explaining how our algorithms were tested and how the hyperparameters of our algorithms were determined.

3.4.1 Support Vector Machine (SVM)

The support vector machine was developed in 1995 by Vladimir Vapnik and is currently one of the most well known classification algorithms [123].

The idea behind the SVM algorithm is to find a m-dimensional hyperplane (the set of points x which satisfy $w^T x + w_0 = 0$), separating the features of the two classes. Features on one side of the hyperplane would belong to a certain class, while features on the other side would belong the other one (blue and red points on figure 3.7). To do so, the SVM algorithm tries to find w, which maximises the width of the margin (dashed lines on figure 3.7), equal to $\frac{2}{||w||_2}$.

The optimisation algorithm related to the SVM algorithm can be formulated as (if we define $y_i = 1$ if element i belongs to class 1 and $y_i = -1$ if element i belongs to class -1) as :

$$\max_{w,w_0} \quad \frac{2}{||w||_2} \tag{3.3}$$

$$s.t \quad y^i(w^T x^i + w_0) \geq 1 \quad \forall i\{1, 2, .., N\} \tag{3.4}$$

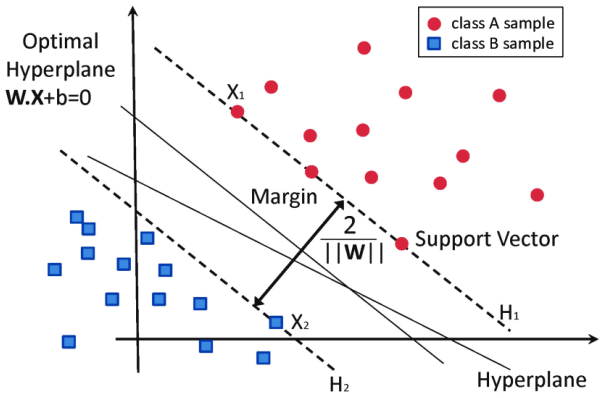

Figure 3.7: Example of data separation with a Support vector machine by maximising the margin, illustration from [124].

However, the optimisation problem defined previously is not always feasible and the problem is then called non-linearly separable. To tackle this problem, the concept of slack variables is introduced, denoted by ϵ^i attached to each point x^i. When x^i is misclassified, ϵ^i measures the distance between x^i and the margin relative to its class, otherwise ϵ^i is equal to 0. SVM also deals with unbalanced dataset, a pretty common problem. This problem happens when a dataset is composed of more instances of one class than the other one. In that case, machine learning can easily fail to work well on a dataset independent of the training set by overfitting and classifying most elements in the class the most present in the training set. If we consider a dataset of size N, composed of two classes each containing $\mathcal{N}$ and $\mathcal{P}$ elements, we can tackle these two problems by adapting the SVM optimisation problem [125]:

$$\min_{w,w_0} ||w||^2 + C(\beta^- \sum_{i \in \mathcal{N}} \epsilon^i + \beta^+ \sum_{i \in \mathcal{P}} \epsilon^i) \tag{3.5}$$

$$s.t \quad y_i(w^T x^i + w_0) \geq 1 - \epsilon^i \quad \forall i \in \{1, 2, .., N\} \quad \epsilon_i \geq 0 \tag{3.6}$$

$$\tag{3.7}$$

where the β^+ and β^- are adapted such that misclassification of elements of the class containing the least elements is more costly. The value of C should also be chosen before optimisation and corresponds to the general misclassification cost. When w^* is found, a new element x, is assigned to class 1 if $w^{*T} x + w_0^* > 0$ and to class -1 otherwise.

3.4.2 Logistic regression

The assumptions of logistic regression are the following [126]:

- Observations should be independent of each other, observations should not come from repeated clinical exam for example.
- Little or non multicollinearlity among the independent variables.
- Linearity of independent variables and log odds.

The last assumption can be rewritten as :

$$\mathcal{P}(y^i = \pm 1|x^i, w) = \frac{e^{w^T x^i}}{1 + e^{y^i w^T x^i}} = h_w(x^i) \tag{3.8}$$

Equation 3.8 assumes that the probability to belong to a class is linearly correlated with x and weights w, the parameters to be found for this method. The weights w^* which maximise the likehood of observing (X,y) following equation 3.8, can be found by maximising the log likehood function :

$$l(w) = \frac{1}{N}\sum_{i=1}^{N}(-y^i w^T x^i + \ln(1 + e^{w^T x^i})) \quad \text{s.t} \quad ||w||_2 \leq C \tag{3.9}$$

Where a norm constraint is added on w in order to have a bounded w and to avoid overfitting. Finding w^* for equation 3.9 can be done using gradient descent as the function is convex in w. The problem can easily be adapted for an unbalanced dataset. If we consider a dataset of size N, composed of two classes each containing $\mathcal{N}$ and $\mathcal{P}$ elements, we can rewrite the empirical loss function as [126]:

$$l(w) = \frac{1}{N}(\beta^+ \sum_{i\in\mathcal{P}}(-y^i w^T x^i + \ln(1 + e^{w^T x^i})) + \beta^- \sum_{i\in\mathcal{N}}(-y^i w^T x^i + \ln(1 + e^{w^T x^i}))) \quad \text{s.t} \quad ||w||_2 \leq C \tag{3.10}$$

where the β^+ and β^- are adapted such that misclassification of elements of the class containing the least elements is more costly. When w^* is found, a new element x is assigned to class 1 if $P(y = 1|x, w^*) > \frac{1}{2}$, else to class -1.

3.4.3 Linear Discriminant Analysis

Linear discriminant analysis (LDA) was developed by R.A. Fisher in 1936. For binary classification, the algorithm supposes that the features belonging to each class follow a multivariate normal distribution with means, $\mu_{i=\pm 1}$, with the same covariance matrix Σ [127]. Under those assumptions and knowing the prior probabilities to belong to a class $\pi_{i=\pm 1}$, the probability that an element x belongs to a class can be estimated using Bayes rule. The class that gets the highest probability is the output class and a prediction is made : element x is assigned to the class j if [4]

$$j = \underset{i}{\text{argmax}} \quad x^T\Sigma^{-1}\mu_i - \frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i + \log\pi_i \tag{3.11}$$

Equation 3.11 is linear in x such that boundaries between classes in the feature space can only be linear. If we assume that the covariance matrix is class specific, boundaries can be quadratic and we assign x to class j if :

$$j = \underset{i}{\text{argmax}} -\frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(x - \mu_i)^T\Sigma_i^{-1}(x - \mu_i) + \log\pi_i \tag{3.12}$$

[4]The principle of this method is to find a new feature space(linear) which maximises the separation between features.

Assuming that the covariance is the same for each class is as strong assumption and often not realistic. To the contrary, assuming that the covariance matrix is different for each class is more realistic but the number of parameters to estimate for the model then increases quadraticly with the number of features. The estimation is thus often poor when the number of samples is low. A tradeoff between both methods is to use shrinkage, individual covariance matrix shrink toward a common pooled covariance [128]:

$$\Sigma_k(\alpha) = (1-\alpha)\Sigma_k + \alpha S \tag{3.13}$$

Where Σ_k, S and $\Sigma_k(\alpha)$ are respectively the estimated covariance of class k (poor when the number of features is low, S the diagonal matrix which contains the variance elements and the shrinkage covariance matrix.

3.4.4 Random Forest

Random forest is an ensemble method, introduced by Breimann in 2001 [129]. His idea was that simple classifiers (decision trees) combined together could outperform more sophisticated models.

Decision trees[5] are very convenient to use as they handle both categorical and continuous input variables. They are also not sensitive to outliers. Moreover, trees contain integrated methods for feature importance and can be used for classification and regression. Despite all these advantages, decision trees can easily overfit and that is the reason why they are not often used alone in practice.

To tackle the overfitting problem of decision trees, the random forest algorithm randomly generates an ensemble of different decision trees. More precisely, random forest uses bagging, short for Bootstrap aggregating. In its construction algorithm, a training sub-sample of size N with replacement of the original data is taken, to serve as input of each decision tree, containing approximately 63% of the dataset. Instead of using all the features of X as input features of the decision trees, $\sqrt{m}$ features are randomly chosen for each tree to create a new feature matrix. This ensures that each tree in the forest sees a different subset of the training samples. The tree growing algorithm is then randomised in order to decrease the correlation between the trees. The predicted value of a random forest classifier is given by the majority votes of the trees in its forest (the one which is the most predicted by the decision trees), in the case of classification, or the average of each terminal leaf for regression, see figure 3.8.

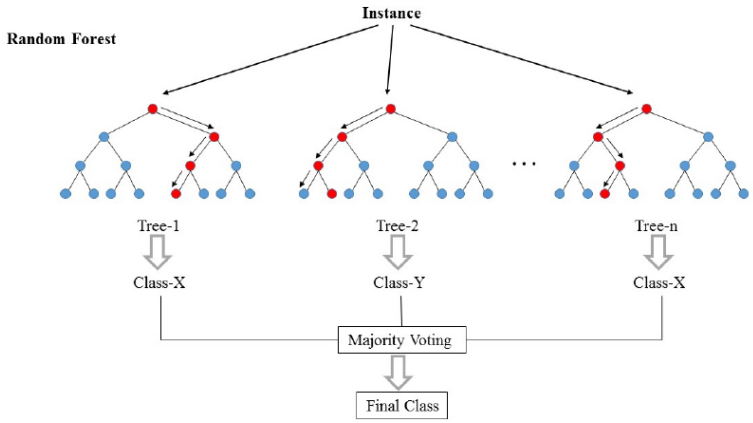

Figure 3.8: Example of majority vote for random forest algorithm from [130]. For each tree of the forest, a new matrix of size $N \times \sqrt{m}$ given as input, each returning a class ($\pm$ 1). The output of the random forest is given by the majority vote of these outputs.

[5] See appendix B.2 for a more in-depth approach to tree formation.

3.4.5 Comparison of methods

Table 3.3 shows the major differences between the machine learning algorithms presented previously:

Algorithms	SVM	Logistic regression	LDA	Random Forest
Type of data	Continuous	Continuous	Continuous	Continuous and categorical
Feature importance	weights assigned to each feature (w) in equation 3.5	weights assigned to each feature (w) in equation 3.9	Impossible	Ginni Importance or Permutation Importance (see appendix B.2)
Non linear boundaries	Possible with kernel trick	Possible with kernel trick	Possible but limited to quadratic boundary with QDA	Impossible
Tricks for unbalanced Dataset	Cost associated to misclassification	Cost associated to misclassification	None (similar to prior probabilities)	Cost associated to misclassification
Prone to overfitting	A little	YES	NO	NO
Handle outliers well	YES	NO	NO	YES

Table 3.3: Advantages and drawbacks of machine learning methods used for this thesis.

3.4.6 Standardisation and correlation

The first two transformations performed on our dataset were standardisation and the removal of linearly correlated variables.

Standardisation

Standardisation consists of rescaling variables such that each new standardized variable has its mean equal to 0 and its standard deviation equal to 1. Standardisation (scaling more precisely), is needed in our case for different purposes.

First, as we explained it before, the empircal loss for the logistic regression model is convex in w (see 3.9), so that gradient descent is used in order to find optimal w. Gradient descent is an iterative method which looks for mininum/maximum of convex/concave functions. At each iteration t, a new w^t is calculated depending on previous w^{t-1} [131] :

$$w^t = w^{t-1} + \alpha \nabla l(w^{t-1}) \tag{3.14}$$

By looking at equation 3.9, we see that $\nabla l(w)$ depends on the input features such that the step size of the gradient descent depends on the scaling of the features. If the features are not scaled properly, it results in different step sizes for each weight element. This can lead to too large step sizes and prevent the convergence of this method. Scaling enables lower step sizes and can thus tackle this problem.

The other reason why scaling was needed for logistic regression and SVM was the norm constraint on w. If features are not scaled, features which are smaller in magnitude will have higher weights than other features which have higher magnitudes and which have potentially as much importance as them for classification. However, due to the norm constraint which avoids overfitting, those weights will be forced to be lower such that all variables are not on the same level.

Plus the performances of random forest and linear discriminant analysis are not affected by the scaling of the variables such that rescaling was applied before applying each algorithm. Although for the variable importance methods of random forest (see the section on feature selection), this can cause some problems if the variables are not scaled cleanly [132].

Correlation Analysis

A correlation matrix was calculated with all features in order to delete redundant variables, the ones which are strongly correlated with each other. This was performed using the Pearson coefficient [133]. The Pearson coefficient between two random variable X and Y can be estimated with :

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.15}$$

where $\bar{x}$, $\bar{y}$ are respectively the sample mean of X and Y and n is the total number of samples X and Y. $r_{x,y}$ ranges between -1 and 1, the closer $|r_{x,y}|$ is to 1, the more the linear correlation between X an Y. If $|r_{x,y}| = 1$, it means that there is a perfect linear correlation between X and Y.

If the Pearson coefficient between two features was above 95% one of both variables was deleted as the one that was kept could derive the other one. Deletion of correlated features does not necessarily improves the performances of the algorithm, but the simpler the better.

3.4.7 Feature selection methods

Since our dataset was small and the number of features was high :

- 210 subjects for problem 1, 2, 3 with 1800 structural MRI features
- 141 subjects for problem 4 and 5 with 1800 structural MRI features
- 37 subjects for problem 4 and 5 with 55 ERP features
- 36 subjects for problem 4 and 5 with 100 resting EEG features.

This has resulted in a feature matrix that is much wider than it is long. This is a phenomenon known as the curse of dimensionality. This can lead to problems such as overfitting [134]. Overfitting happens when the machine learning algorithm using the training samples fits too closely to the datapoints. In such a case, the model learns the details too much of the training samples, different from the test set. When faced to the test set, the model does not work as well as in the training phase as noise was used as learning concept for the predictor [135]. Overfitting occurs very often when the number of input features is large compared to the number of samples. Thus, several methods have been implemented to reduce the number of dimensions while keeping the variables contributing most to the classification. This is referred to as feature selection.

Feature selection is a process by which the features contributing the most to classify output features of the machine learning algorithms are selected. The feature selection methods that are proposed in the next section are heuristics since the problem of finding best features as input of machine learning models is NP hard [136]. Feature selection methods thus do not necessarily identify the variables that contribute the most to the output variables [137]. For this thesis, we present four types of feature selection methods, see figure 3.9 for a global overview:

- Filter methods
- Projection methods
- Wrapper methods
- Embedded methods

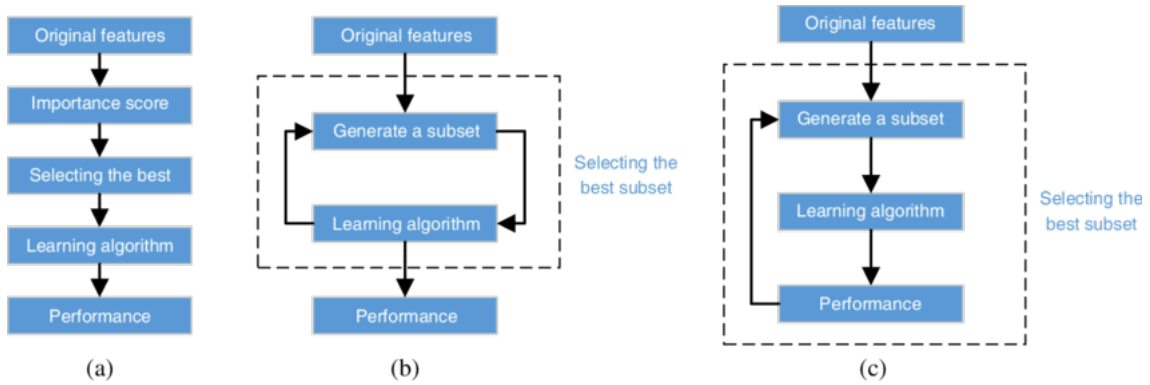

Figure 3.9: Feature selection methods: sub-figure a represents filter methods, sub-figure b represents embedded methods and sub-figure c wrapper methods. Projection methods are not the figure but can be compared to filter methods, instead of importance score, projection in a new feature subspace is done.

Filter methods

Filter methods use statistics of the data in order to find the best features. Anova's t-test was applied in order to assess any statistical differences between the EEG and MRI features between the different groups. The Anova's t-test is a statistical test that is used to find out whether the distribution of a variable is statistically different from one population to another. It assumes that the distributions of the two groups have equal variances, follow a normal distribution and are independent.

From those assumptions, we can calculate the t-statistic of feature X^i to check whether y affects X^i or not :

$$t(X^i) = \frac{|\bar{X}_1^i - \bar{X}_{-1}^i|}{\sqrt{\frac{(\sigma_1^i)^2}{N_1} + \frac{(\sigma_{-1}^i)^2}{N_{-1}}}}$$

Where $\bar{X}_j^i, \sigma_j^i$ are the sample mean, standard deviation of the ith feature for class j. N_j corresponds to the number of samples for class j.

After calculating the values of the t-statistic for each feature, we sort these values in descending order in order to select the most important features.
Applying this method before feeding SVM, logistic regression, LDA and random forest algorithms are respectively called **filter SVM, filter log regression, filter LDA and filter random forest**.

Projection methods

Dimensional reduction can also be performed using PCA [138]. PCA reduces the number of dimensions in the dataset while keeping as much variation in the dataset as possible. To do this, the original variables are transformed into a new set of variables, the principal components of X (our dataset), which are uncorrelated with each other and such that the first principal components contain the largest variance of the dataset. The first k principal component directions of the standardised matrix X, $V_{m\times k}$, can be found by solving the following optimisation problem :

$$V_{m\times k} = \text{argmax}_{AA^T=I_k} \text{trace}(A^T \Sigma A) \tag{3.16}$$

where Σ is the covariance matrix of the dataset given by $\frac{1}{N}\sum_{i=1}^{N}(x^i - \bar{x})(x^i - \bar{x})^T$ with $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x^i$.
The easiest way to apply PCA for dimensional reduction is to compute $H = AA^T$. From H, we can then project each sample of the dataset X_i into the space of the first k principal components. It is interesting to note that the principal components with high variance are not necessarily useful for separating two populations. This is because the direction of the highest variance may not separate the sample means well after projection [139]. Other projection methods such as Isomap, Locally Linear Embedding and ICA[6] were

[6]The results will not be displayed as the obtained results were not better than with PCA.

also tested. Applying PCA before logistic regression, SVM and LDA are respectively called **PCA logistic, PCA SVM and PCA LDA**.

Wrapper methods

The two methods that were proposed before based on the statics of the data. Wrapper methods are a little bit different, they are dependant of the learning algorithm, see see figure 3.9 for a better overview. Filter and projection methods can be executed without prior knowledge of the machine learning algorithm. In contrast, wrapper methods select the best features according to the performances of the algorithm. The features that are selected by wrapper methods depend on the algorithm used and may not be suitable for another algorithm. They are more easily subject to overfitting since they involve training with multiples sets of feature combinations [140]. A better explanation of those methods is given in appendix B.1. Those methods were not retained for this thesis.

Embedded methods

Embedded methods also depend on the machine learning algorithm (see figure 3.9) used and cannot be realized for each of the machine learning algorithms listed above.

Embedded methods using logistic regression are called Lasso regressions. In order to understand this method, let's reintroduce the empirical loss function [7] of the logistic regression.

$$l(w) = \frac{1}{N}\sum_{i=1}^{N}(-y^i w^T x^i + \ln(1 + e^{w^T x^i})) \tag{3.17}$$

The goal of the logistic regression algorithm is to find w^* that minimizes the loss function 3.17. For standard logistic regression, a L_2 norm constraint is added on w to avoid overfitting, it is simply written as $||w||_2 \leq C$, where C is a hyperparameter of the model. When the L_1 norm is used for this constraint : $||w||_1 \leq C$, we talk about Lasso regression. In that case, the coefficients related to minor contributions are put to 0 and w becomes sparse and it can thus also be seen as feature selection method. This method can not work as a feature selection method if the features are not scaled properly. Lasso regression is computational less expensive and less prone to overfitting compared to wrapper methods [141]. This method will be called **Lasso regression** or simply **embedded log regression**.

Random forest gets different embedded methods. For this project, we mainly focused on the Gini importance of features derived of the Gini impurity. For a better explanation on random forest embedded methods and Gini impurity, take a look at annex B.2. This method will be called **embedded random forest** until the end of this thesis.

3.4.8 Training and test phase

To evaluate the performances of the different methods, a two-stage approach was adopted. The first one is called the training phase. The second one is called the test phase. The training phase is used to find the best parameters for the machine learning algorithms and to train the models. There are two types of parameters:

- Model Parameters : these are the parameters specific to the model, they depend on the data and the hyperparameters given to the model (f.e. w in equations 3.7 and 3.17).
- Hyperparameters : those parameters that are taken as input to the models, they should be tuned in order to obtain the best performances of the machine learning algorithms, see table 3.4 for a list of them (f.e. C in equations 3.7 and 3.17).

[7] It corresponds to the maximum likehood of the probability density defined at equation 3.8.

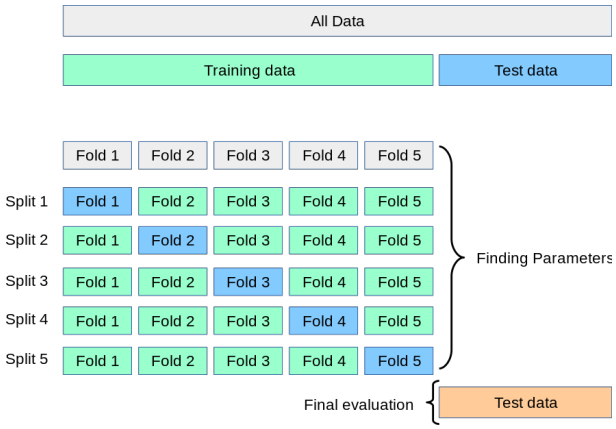

Figure 3.10: Grid search cross-validation from [142]. Above figure shows data separation in training and test. The bottom figure shows 5-split cross-validation where the blue part is for each split the validation set and the green part the training set.

Before the training and test phase, the whole dataset is splitted in a training and a test set, respectively containing 80% and 20% of the data. During the training phase, the training set is splitted in 5 different sets. $\frac{4}{5}$ of the training data is used to train our model and $\frac{1}{5}$ of the training set, the validation set, providing an unbiased evaluation model fitted on the training set, is used to evaluate the model. This is repeated 5 times, see figure 3.10. The value of the best performing hyperparameters, averaged over the five validation tests, is retained for the test phase. We call this 5-fold-cross grid search validation, see figure 3.10 for a better understanding.

During the test phase, the model obtained during the training phase is used (the one with the best hyperparameters and fitted with the training data set) to predict the results of the test set. The predicted outcomes are then compared to the test outputs to evaluate the performances of our models. It is really important to separate the training set from the test set, such that no overfitting can take place. Moreover, it is very often recommended when the dataset is small to repeat this training and test phase several times as the performances of the classifiers depend on the choice of the training and test set. In our case, the procedure was repeated 20 times with a random shuffling of the MRI data before each repetition.

When the dataset is really small, another strategy is often used to assess the performances of machine learning algorithms, called Leave-One-Out (LOO). LOO is similar to what we have seen previously, the training phase remains the same, the splitting is just different. Instead of initially splitting the data in a training set containing 80% of the data and a test set containing 20% of the data and repeating this 20 times, we iterate over each element of the dataset. At each iteration, the training set contains the whole dataset except the iterated element and the test set contains the iterated element. This method is computationally more expensive and it is the reason why it was not applicated for the MRI dataset, but it ensures that all elements are taken into account in the test set, that there is no bias due to the random shuffling and that the size of the training set is higher (more learning). When the dataset size is larger, the performances obtained by the two test methods presented above do not vary greatly [143].

3.4.9 Parameters tuning

When our machine learning models are chosen and our dataset is created, the elements that will influence the most the performances of our machine learning algorithms will be the hyperparameters. The next paragraphs explain what values the hyperparameters could take for the grid search selection, presented in table 3.4.

Method	**Parameters**	**Value(s)**
All methods except method Lasso regression	Number of features to keep (number of principal components for PCA)	Ranged from 1 to 60
All methods except LDA methods	Cost associated to misclassification error for example β^+ and β^- in equations 3.5 and 3.10.	Ranged from 0.2 up to 0.8
SVM methods	C in equation 3.5	Ranged from 0.0001 to 0.01
Logistic regression with filter and projection methods	C in equation3.10	Ranged from 0.01 to 2
Lasso regression	C in equation 3.10 with L_1 norm constraint on w	Ranged from 0.0001 to 0.01
LDA methods	Shrinkage, α in equation 3.13	Ranged from 0.1 to 1
All random forest method	Number of features per decision tree	$\mathbf{log_2}$(**features**)
All random forest method	Number of trees	**400**
Embedded random forest	Feature importance method	**Ginni importance** (see appendix B.2), permutation importance

Table 3.4: Hyperparameters of the different methods. Words in bold are choices that were not optimised during grid-search.

Some values of hyperparameters of the machine learning algorithms always obtained better performances by cross-validation than others. Therefore, to test the performances of our machine learning algorithms, these values were no longer optimised to save time. These values are shown in bold in the table 3.4.

The higher the **number of input features**, the more the machine learning algorithm can learn. However, the more input features, the more chance of overfitting during the training phase. That is why, the number of features to keep for feature selection method was limited, see table 3.4. If the number of the samples was really higher, there would be no meaning to select the most important features as the machine learning algorithm would select them by itself during its training phase.

Cost associated to misclassification error is needed in order to handle unbalanced dataset. Initially, misclassification cost were put to be proportional to the number of elements of each class in the training set, so that the class with the least number of elements was favored, but after a few tries it was seen that this cost could be adapted and could lead to better performances. Another way to counter this problem is to use subsampling. Subsampling can be performed using down-sampling or up-sampling [144]. Down-sampling consist in taking a training set which does not contain 80% of the data, but to take a training set with a lower size but which contains as many elements of each class in the training set [144]. Elements of the most predominant class are removed from the training set until the number of elements of both classes in the training set is the equal. The problem with this method is that the training set is now having a lower size such that the machine learning methods learns less. Up-sampling consists in taking a training set containing 80% of the data but elements from the minority class with random replacement are added (so that there may be duplicates) until the number of elements from both classes are equal in the training set [144]. By adding new data, we add more importance to some elements in the minority class, but we also add bias as we have duplicates.

The higher the value of **C for SVM [125] and logistic regression [126]**, the higher the probability of overfitting. Indeed, when the number of samples is low and the number of features is high for logistic regression and SVM [125], the weights w are chosen during the training phase in order to have the best performances and will then often fit too close to the data. By adding, a term of regularisation, the values of the weight parameters w are limited and thus prevent overfitting. The regularisation term is regulated by C. A high value of C can promote overfitting, while a low value of C can lead to underfitting, which makes the choice of C tedious.

The **shrinkage** parameter α [128] for LDA is as C for SVM and logistic regression a term of regularisation. The value of α is always between 0 and 1. When the value of α is equal to 0, this corresponds to zero shrinkage and we are in the case of QDA, which does not work well when the number of samples is low and the number of features is high. This will tend to overfitting. Conversely, a value of α of 1, implies a complete shrinkage which implies that the diagonal matrix of variances will be used as an estimate of the covariance matrix of class k and will tend to underfitting.

The number of trees for the random forest methods were not tuned and kept to 400. For binary classification problems, each decision tree in the forest returns the value 1 or 0. The decision is then made by majority vote. This can be thought of as taking the average of these outcomes and applying the step function. The application of the weak law of large numbers (WLLN)[8] in this case implies that, for each sample, the set will tend towards a particular average prediction value for that sample as the number of trees tends to infinity [145]. By using different numbers of trees to generate our forest, we realised that from 400 trees the results tended to this value.

3.4.10 Performance metrics

There are many ways to assess the performances of a binary classification problem. To understand them, let's first introduce the concept of confusion matrix, central in the assessment of performance. To do so, let's consider a simple example where the classification problem consists in assessing that people are ill or healthy. The true positive (TP) are the samples which are predicted as ill and are ill, the false positive (FP) are the samples which are predicted as ill and which are healthy. The false negative (FN) are the samples predicted as healthy and which are ill. The true negative (TN) are the ones which are predicted as healthy and which are healthy [146].

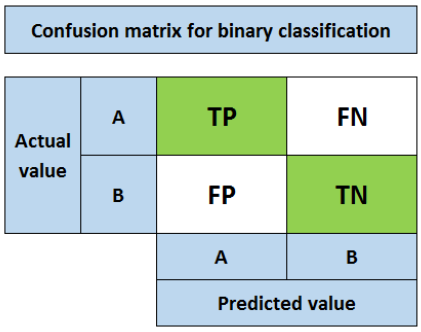

Figure 3.11: Example of confusion matrix for binary classification problem (where A = ill, B = healthy) from [146].

For our algorithms, the first metric that was evaluated was the accuracy. The accuracy is the ratio of number of correct predictions to the total number of input samples and is given by the following equation [147] :

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Sample}} \tag{3.18}$$

The problem with accuracy is that when the population is not balanced, which means that there is more samples of one class than the other one, the accuracy can be really high even if the machine learning algorithm always predicts the same class. For example imagine a test sample with 100 individuals, 80 AUD patients and 20 controls. The train accuracy can easily get 80% of accuracy by simply predicting each individual as AUD patient. That is why, accuracy can not be used alone as performance metric for unbalanced dataset.

[8]The weak law of large numbers is applicable as the decision of all trees is identically distributed and that since each tree returns the value 0 or 1, that each experiment has a variance of finite size.

Let's take a look at another performance metric widely used in the literature. The precision of a class is the number of correct positive results divided by the number of positive results that were predicted by the classifier [147].

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{3.19}$$

The recall or sensitivity of a class measures the proportion of positive that are correctly labeled [147].

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{3.20}$$

Recall and precision are relevant metrics to analyse which class is well classified and which class is not.

The last metric that was used to assess performances of the machine learning algorithms was the Cohen's Kappa κ, introduced by Cohen in 1960 [148]. The Cohen's κ enables interrater and intrater reliability. It ranges between -1 and 1, where 0 means that the predicted outputs were obtained by chance and 1 means that the predicted and the actual outputs are completely similar. Negative Cohen's κ is really rare. For example, if we go back to our previous example, classifying all the items in the AUD class will give a Cohen's κ value of 0 and an accuracy of 80% [149]. Cohen's κ is given by the following formula:

$$\kappa = \frac{p_a - p_c}{1 - p_c} \tag{3.21}$$

where κ, p_a, p_c are respectively inthe Cohen's κ, the percentage of agreement between y predicted and the true y and, the chance agreement.

An interpretation of the value of κ is listed below [149]:

- <0 : no agreement
- 0-20 % : slight
- 20-40% : fair
- 40-60% : moderate
- 60-80% : substantial
- 80-100% : almost perfect agreement

However there is no consensus on this interpretation of the value of Cohen's κ in the literature [147].

3.5 Strategies

This section covers how the EEG and MRI data have been used in order to address the predictive tasks introduced in the section 3.1.1. Four different strategies have been introduced to try to solve machine learning prediction tasks. The first and second strategies use only the MRI data, while the third strategy uses the EEG data. The last strategy tries to use the MRI and EEG data together, but as the number of samples which had EEG and MRI data was too low, a multi-modal MRI strategy was used.

3.5.1 All-sMRI strategy

The all-sMRI strategy uses the structural MRI data. All structural MRI data is initially taken into account to make predictions. MRI images are first preprocessed and different MRI features (see figure 3.12) are extracted from each region which are described by the Desikan, Destrieux and Brodman atlases. The input matrix is then subdivided in a training (80%) and a test set (20%), cross-validation on the training set is used to find the best hyperparameters and the most important features (see table 3.4). Finally, performance metrics are evaluated on the test set. The purpose of this strategy is to see to which extent structural MRI data can be useful to detect psychiatric diseases. For this strategy, the machine learning algorithms listed below are used :

- SVM, logistic regression, LDA and random forest with filter methods.
- SVM, logistic regression, LDA with PCA.
- Ridge Regression and random forest with Gini importance.

A better scheme of this method is shown on figure 3.12.

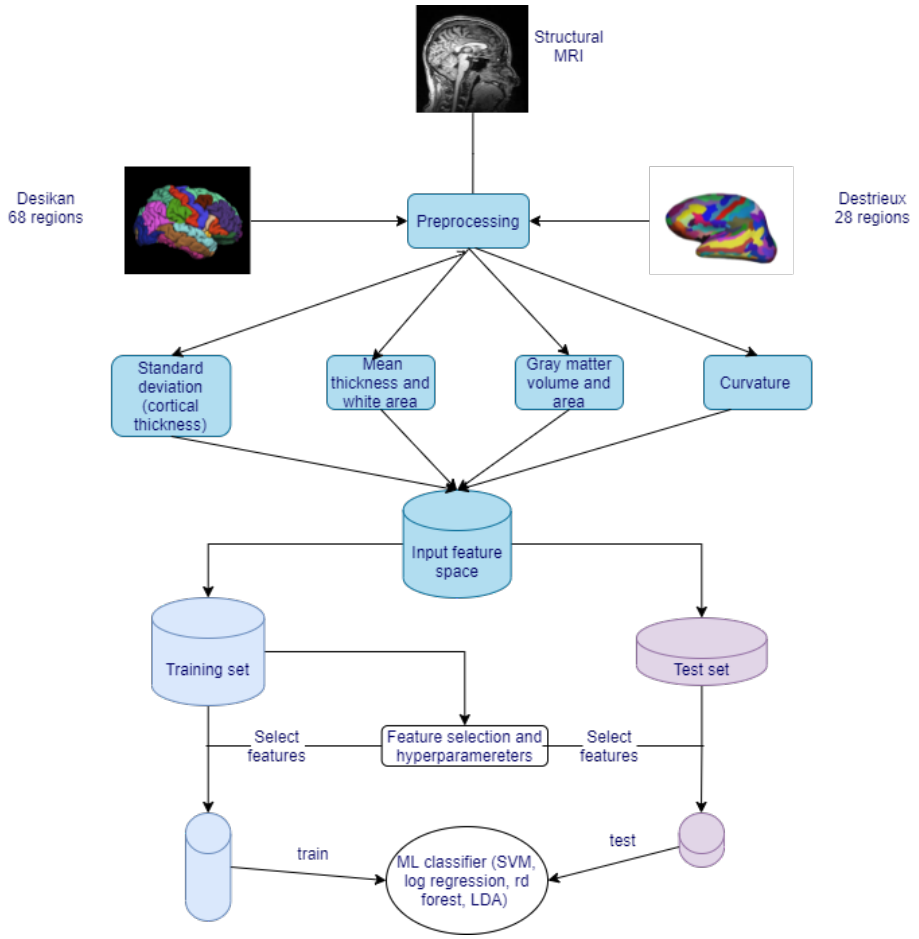

Figure 3.12: Scheme of the all-sMRI strategy. All sMRI data is given as input features of the classifiers, data is then split in a training (80%) and a test set(20%). Filter or embedded methods or PCA are then used in order to reduce the number of dimensions (the number of dimensions is chosen by cross-validation as the hyperparameters of the model). The classifier models are then fitted with the training set and hyperparameters are found by cross-validation. Performances of the classifiers are then evaluated on the test set (repeated 20 times).

3.5.2 Separate-sMRI strategy

The separate-sMRI strategy is similar to the previous one except that using all structural MRI as input of the machine learning algorithms, the MRI features, standard deviation of cortical thickness, mean thickness, gray matter volume and curvature, are studied separately. Instead of using them together to create a huge dataset containing a high number of features and a low number of samples, each MRI feature gives rise to an input dataset, see figure 3.13. By this way, the problem of curse of dimensionality is easier to address since the number of features of each dataset is lower. It also allows us to see which variables are important for our different classification tasks. Plus, this strategy enables to look at the agreement between the predictions performed with different information. For this strategy, not all methods presented in section are used, only filter random forest and filter SVM have been implemented as those were performing well with the previous strategy and are not too complex. After separation of the MRI data, the data is then standardised and the

most correlated features are removed. Filtering methods are then used to reduce the number of dimensions, see figure 3.13. Performances of the classifiers are then evaluated on the test set (repeated 20 times).

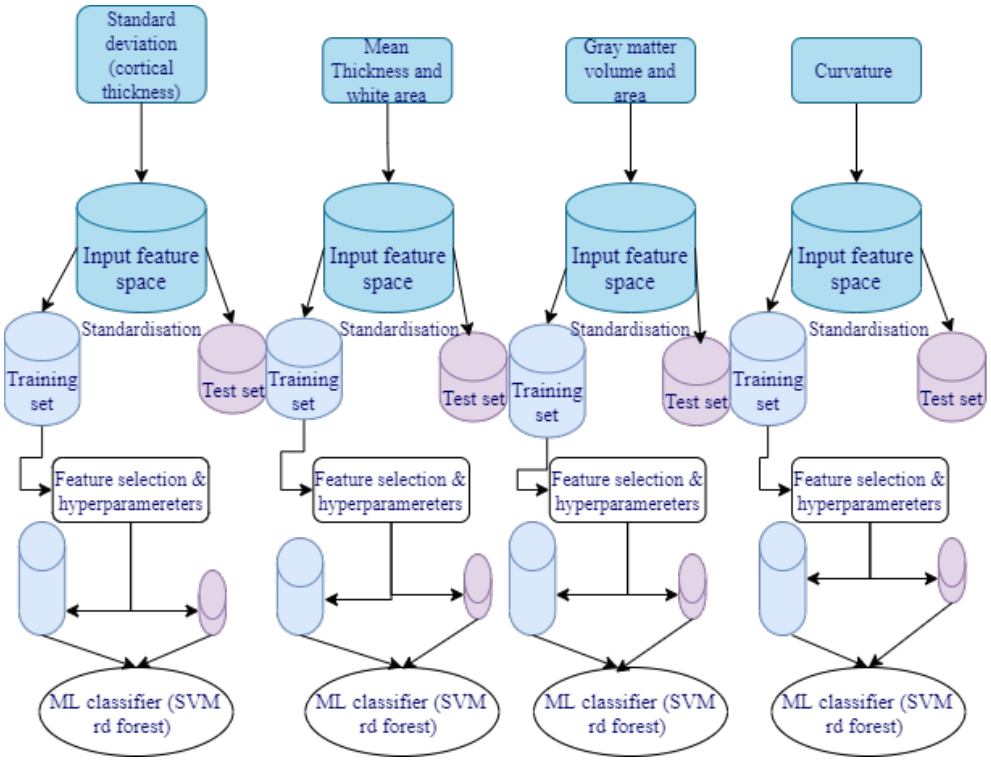

Figure 3.13: Scheme of the separate-sMRI strategy. There is one classifier per MRI data, the same steps as for the all-sMRI strategy are then performed.

3.5.3 Resting EEG and ERP strategy

This strategy uses the same algorithms than the all-sMRI strategy, see 3.5.1, to handle the classifications tasks with the ERP and resting EEG features. The EEG signals are first preprocessed and EEG features are extracted from each of them, see table 3.2 for the complete list of them. The resting EEG and ERP features are not put together in the same dataset to resolve classification tasks 4 and 5 as some patients did not get some ERP or resting EEG exams. Plus, some resting EEG exams were really noisy such that it was not possible to extract consistent resting EEG features from them. That is why, the resting EEG and the ERP features were fed to machine learning algorithms separately, see figure 3.14. The data is then standardised and the most correlated features are removed. The data are further separated into a training set and a test set using the Leave-One-Out method. Filtering, embedded or PCA methods are then used to reduce the number of dimensions. The classifier models are then fitted with the training set and hyperparameters are found by cross-validation. Performances of the classifiers are then evaluated on the test set (for each element of the dataset). A complete overview of this model can be seen on image 3.14.

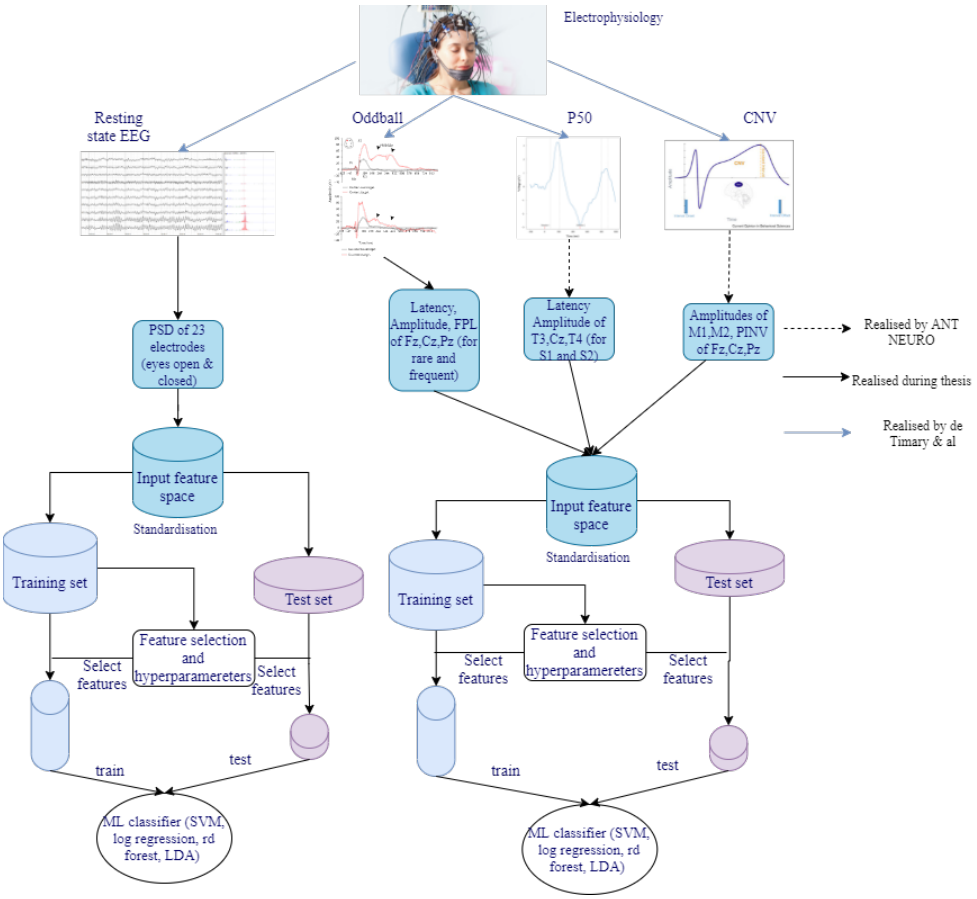

Figure 3.14: Scheme of the EEG strategy. The EEG dataset is splitted in two (resting EEG and ERP dataset), each giving rise to a input feature matrix. Same steps as for the all-SMRI strategy are then performed, except that training and test set are split by LOO.

3.5.4 Multimodal strategy using sMRI data

Initially, the purpose of this thesis was to combine EEG and MRI features together in order to detect depression state and anxiety state for alcoholic patients. However, this was finally not an option as the number of individuals having EEG and MRI features was really low, 23 patients[9]. The idea was to take the outputs of SVM filters obtained via EEG and MRI features separately and used them as input of a perceptron. A perceptron is a simple single-layer neural network. The perceptron takes its input values and multiplied them by a weight, specif to each input feature, learned during the training phase. Those values are then summed and applied to an activation functions, which mapped the weighted sum to the values 0 or 1. The perceptrons do not perform well when the input features are categorical, which means that the perceptron would not perform well and improve the results obtained with the MRI and EEG features separately if we put the predictive class of each model as input to the perceptron. To address this problem, instead of putting the output classes({-1: healthy, 1: ill} for example) as input for the perceptron, distances to hyperplanes of SVM could be put as input of the perceptron. A negative distance would mean that the individual belongs to class -1 and a positive one would mean that it belongs to class 1. Moreover, further away points are from the hyperplane, more the points are considered to be well classified such that perceptron. From the distances of the points to the SVM hyperlane, Platt devised a model that would estimate the probability of an element belonging to the class 1 or the class -1 [150]. The parameters of this model are estimated by cross-validation. The model then takes as input a distance from a point to the SVM plane and as output a number p between 0 and 1. This number indicates the probability estimated by the SVM model that this element is sick ($y = 1$). In principle, if $p > 0.5$, it means that the element is estimated to be

[9]Moreover, data was strongly unbalanced for the 23 patients only five elements were depressive and 4 anxious which would result in classifying all elements as control.

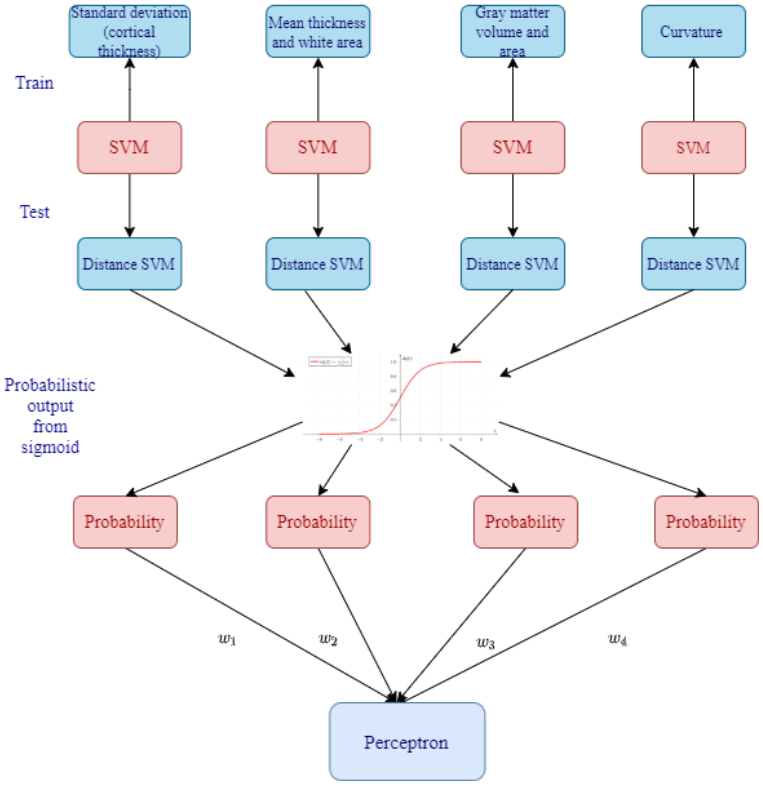

Figure 3.15: Mulitmodal strategy using sMRI data : combination of structural MRI data.

sick by the model. The closer p is to 1, the greater the probability that the subject is ill and vice versa. To facilitate the learning of the perceptron and to have a better understanding of the inputs to the perceptron, these probabilities p were given as input to the perceptron, this is the probabilities case on figure 3.15.

This strategy is different for the all-sMRI data as it combines all structural MRI data. Indeed, as it was explained earlier, as the number of features is too high and the number of samples too low, feature selection is needed. For this selection, for example, the most important features for alcoholism were usually information on cortical thickness and other types of features were often not taken into account. The purpose of the fourth model was to see if each MRI information could be combined.

80% of the initial dataset was used to train the SVM (train on figure 3.15). 20% of the dataset was left as test set. By cross-validation, the probability density of belonging to a class was estimated, the probabilities were then estimated for the elements of the test set. This was then repeated 5 times to get a new set, which is the concatenation of the 5 previous test sets. From this new dataset, 80% of the set was used as training for the perceptron and 20% to evaluate the performances of the classifier. This step was repeated 20 times and mean values of the performances were retained. The predictions of the training samples to belong to a class to train the SVM could not be used to train the perceptron as they were biased due to overfitting.

Chapter 4

Results

This section is separated in four parts, one for each strategy we proposed in the previous section[1]. For each strategy, we then present the results we obtained for the classification tasks we defined in section 3.1.1. The Cohen's κ and accuracy are displayed as the recall and precision of our methods.

For this chapter, it is important to understand what a baseline is. Since we are dealing with unbalanced datasets, it is easy to get an accuracy above 50%, but this does not mean that learning is done, see section 3.4.10. The baseline corresponds to the proportion of elements of the predominant case in our dataset. The baseline is therefore more or less[2] equal to the accuracy of the classifier that would classify all the elements in the most represented class in the training set. To see if learning is taking place from our data, we will mainly compare the accuracy of our methods with the baselines.

4.1 All-sMRI strategy

In this section, we present the results obtained in order to detect depression, alcoholism and anxiety from all structural MRI data, obtained with the strategy defined in section 3.5.1.

4.1.1 Problem 1 : Detection of alcoholism

The following figure shows the performances we obtained in order to detect alcoholism using all-sMRI strategy.

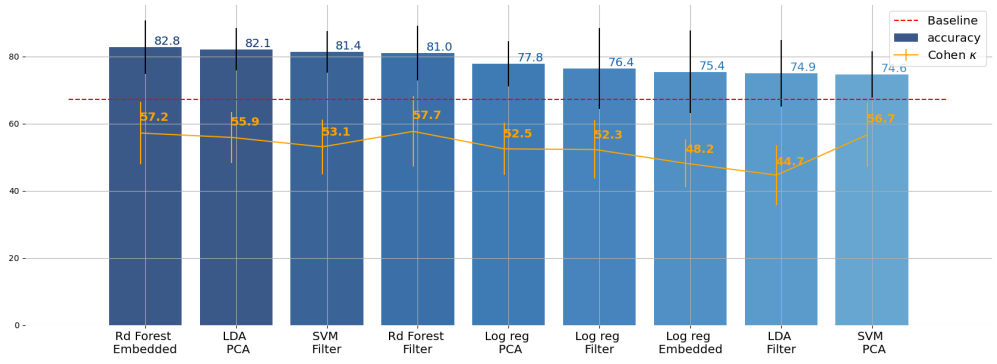

Figure 4.1: Accuracy and Cohen's κ using model all-sMRI strategy in order to predict alcoholism.

[1]The methods that were used for each strategy were implemented using the package *sklearn*.

[2]Some bias is possible due to the random shuffling of the data before creating the training and test set. For the Leave-One-out test method, both are equal, see section 3.4.8 for a reminder of the test methods.

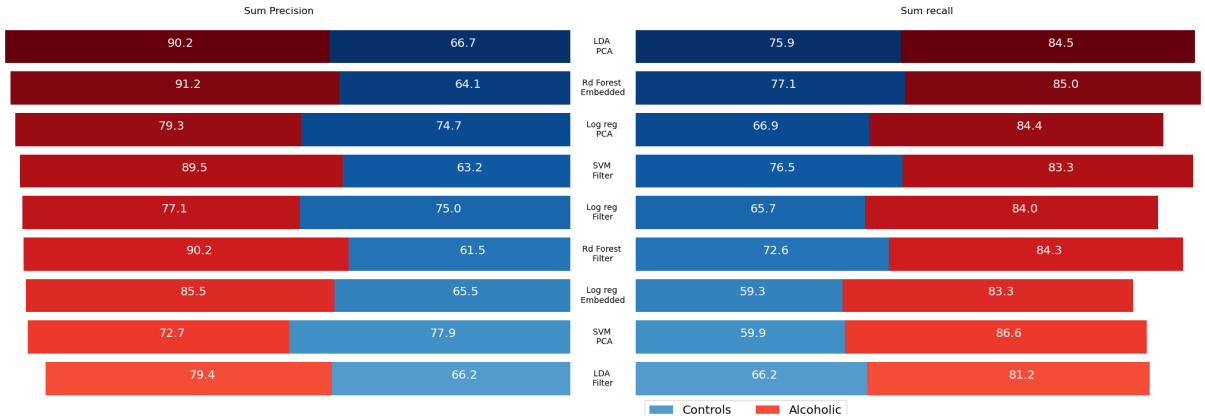

Figure 4.2: Recall and Precision using model all-sMRI strategy in order to predict alcoholism.

All methods are performing better than the baseline, showing evidence of the effect of alcohol consumption on the brain. Moreover, we can see that the Cohen's κ of all methods is higher than 40%, meaning that the agreement between the predicted outputs and test outputs is moderate. With the second image, we can see that the recall for AUD patients is higher than for controls, certainly due to class imbalance, it is prefered to classify an element as AUD.

4.1.2 Problem 2: Detection of anxiety

The performances of the machine learning algorithms implemented for this problem are not better than the baseline, equal to 80%. For this problem, the recall and precision are not very informative. The dataset is strongly unbalanced and no learning was done such that the precision for the controls was almost equal to 100%, while the precision, as the recall, for the anxious patients are next to 0 for almost all methods, see appendix C.1.1 for a better overview of the results.

4.1.3 Problem 3: Detection of depression

The following figure shows the performances of the different methods we obtained in order to detect depression with all structural MRI information.

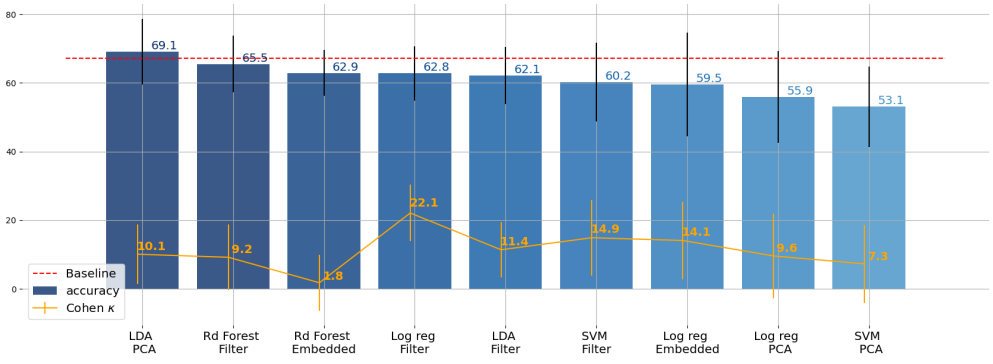

Figure 4.3: Accuracy and Cohen's κ using all-sMRI strategy in order to predict depression.

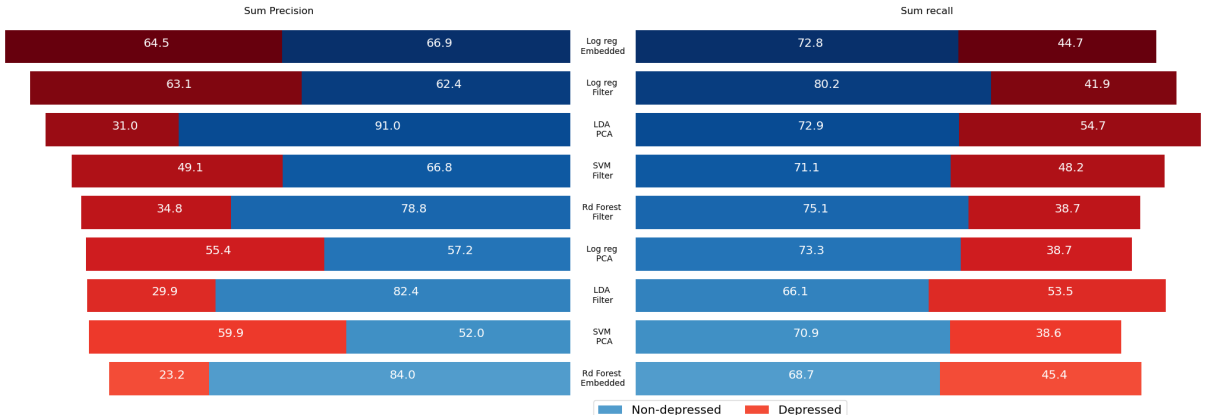

Figure 4.4: Recall and Precision using model all-sMRI strategy in order to predict depression.

The method obtaining the best results was LDA with PCA with an accuracy of 69.1%. Still the accuracy is just above the baseline (67%) and given the value of the Cohen's κ, this method is not reliable. No any method is performing significantly better than the baseline and the best strategy consists in classifying almost all elements as controls, value of Cohen's κ near 0. The recall is also higher for non-depressed subjects due to the missclassification weights.

4.1.4 Problem 4: Detection of anxiety for alcoholic patients

The performances of the machine learning algorithms implemented for this detection of anxiety for alcoholic patients are not better than the baseline, equal to 74%. For this problem, the recall and precision are not very informative. The dataset is strongly unbalanced and no learning is done such that the precision for the controls was almost equal to 100%, while the precision, as the recall, for the anxious patients are next to 0 for almost all methods, see appendix C.1.2 for a better overview of the results.

For this problem, we get the same problem as for problem 2, due to the class imbalance. the best method is simply to assign each patient as anxious alcoholic and this renders results not better than the baseline.

4.1.5 Problem 5: Detection of depression for alcoholic patients

The following figures give an overview the performances of the different methods implemented in order to detect alcoholism with structural MRI information.

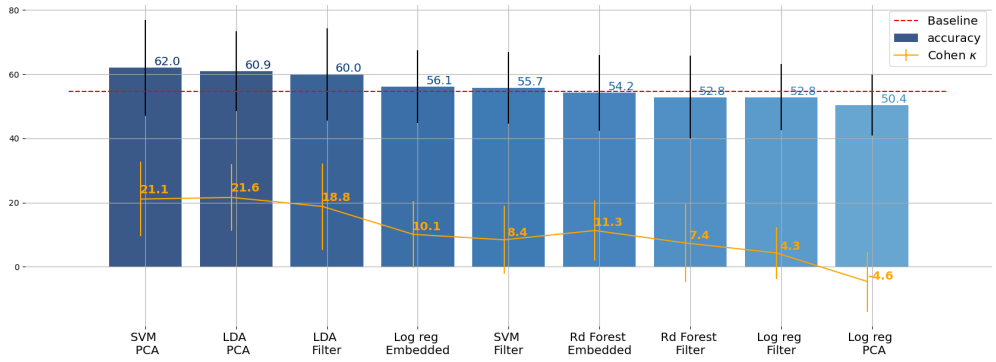

Figure 4.5: Accuracy and Cohen's κ using all-sMRI strategy in order to predict depression for AUD patients.

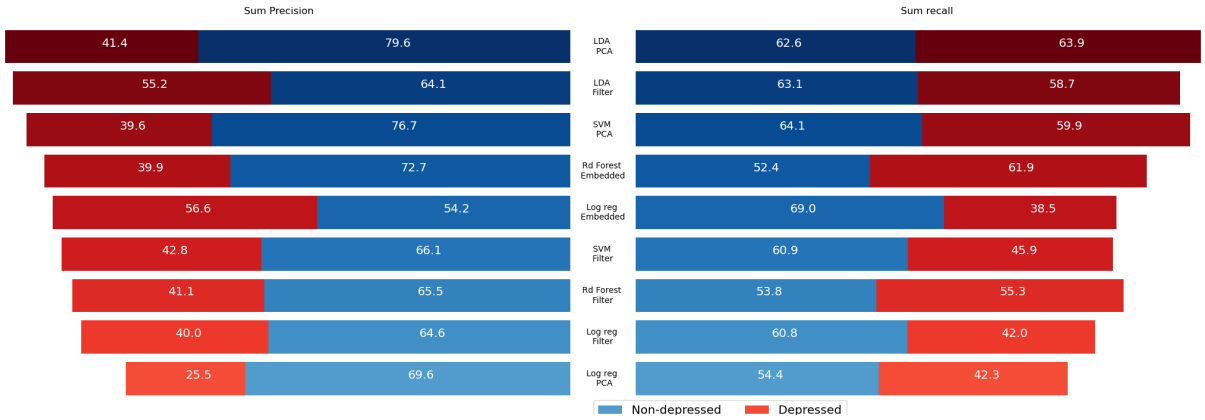

Figure 4.6: Recall and Precision using all-sMRI strategy in order to predict depression for AUD patients.

There is an amelioration compared to the baseline, 55% by using machine learning algorithms in order to asses depression for alcoholic patients. It looks like that the only way to classify patients as depressive or not is by transforming the initial feature space to a new one, in our case by using PCA. For most of the methods, the recall is almost the same for AUD and depressed AUD.

4.2 Separate-sMRI strategy

In this section, we present the results obtained in order to detect depression, alcoholism and anxiety from MRI data separately using random forest and SVM.

4.2.1 Problem 1 : Detection of alcoholism

The two next figures show the results we obtained with the separated sMRI strategy respectively using random forest and SVM to classify patients as alcoholic or not :

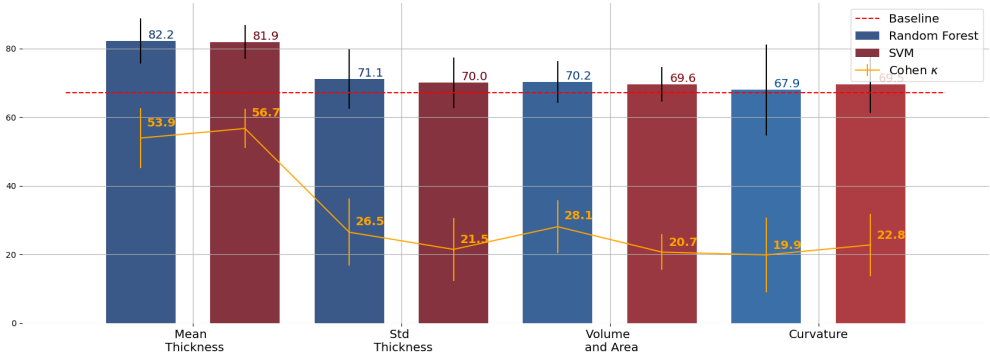

Figure 4.7: Accuracy and Cohen's κ using separate-sMRI strategy in order to predict alcoholism.

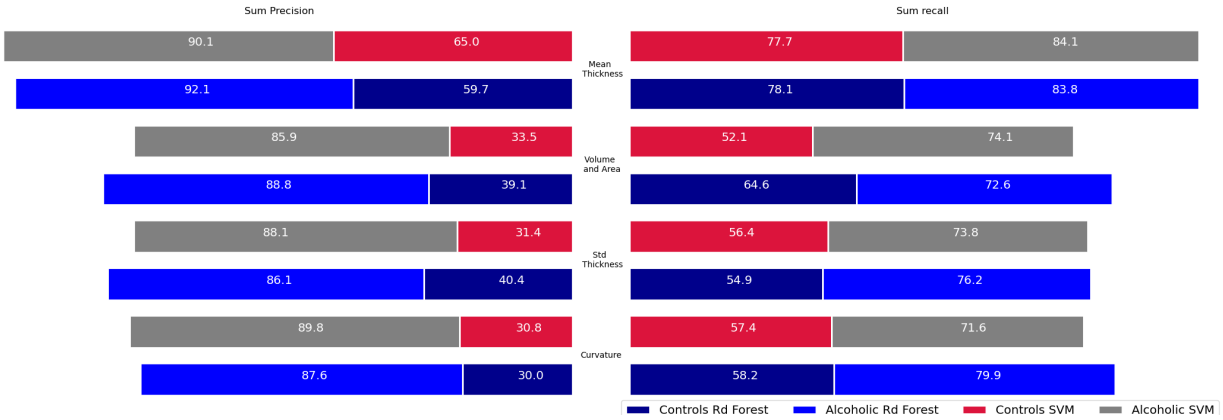

Figure 4.8: Recall and Precision using separate-sMRI strategy in order to predict alcoholism.

The machine learning models obtained with mean cortical thickness are the ones which are working the best, while the other models have an accuracy just above the baseline. From the other metrics than cortical thickness, we see that most of the elements are well classified for AUDs but much less for controls which is due to the large class imbalance and the low learning.

4.2.2 Problem 2 : Detection of anxiety

No method is working better than the baseline for this problem with the separate-sMRI strategy. The most accurate method consists in classifying all items as controls, see appendix C.2.1 for a better overview of the results.

4.2.3 Problem 3 : Detection of depression

The two next figures show the results of the separated sMRI strategy we obtained, using respectively random forest and SVM to classify patients as depressive or not :

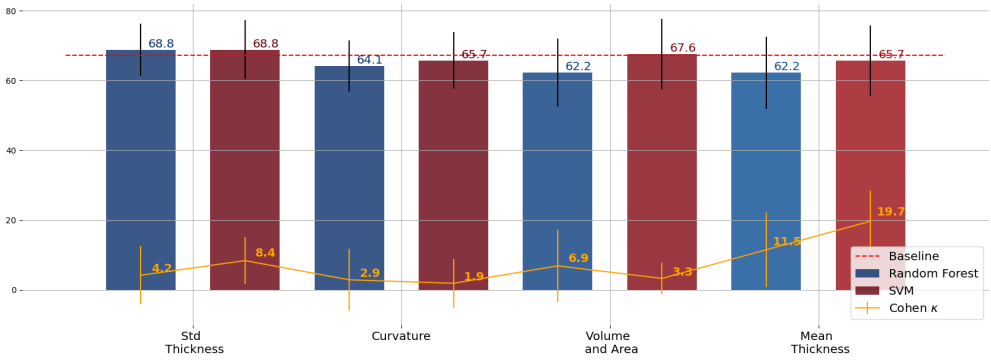

Figure 4.9: Accuracy and Cohen's κ using separate-sMRI strategy in order to predict depression.

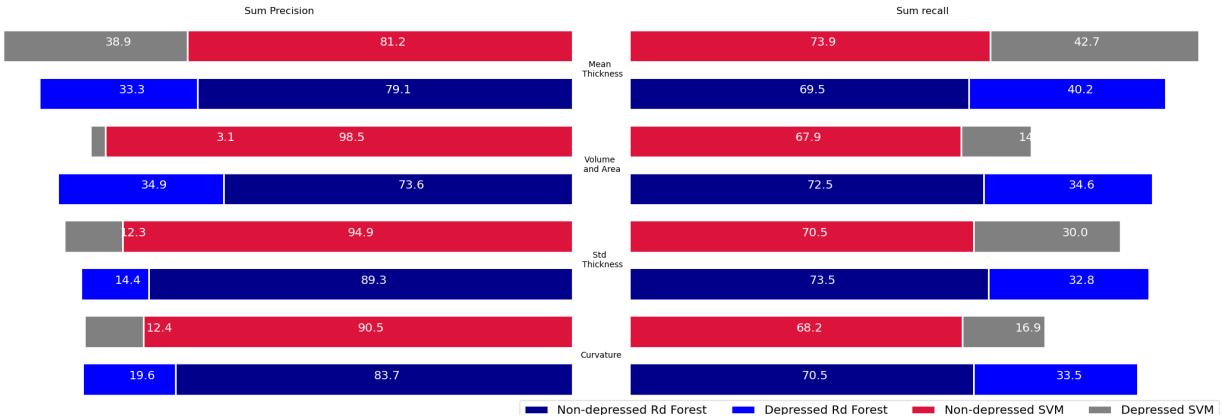

Figure 4.10: Recall and Precision using separate-sMRI strategy in order to predict depression.

There is no amelioration compared to the baseline with none of the methods. Moreover, the Cohen's κ is always close to 0, meaning there is no learning.

4.2.4 Problem 4 : Detection of anxiety for alcoholic patients

For this problem and with this strategy, the most accurate method is to classify all items as controls. The accuracy is just below the baseline for all methods, for a better overview of the results see appendix C.2.2.

4.2.5 Problem 5 : Detection of depression for alcoholic patients

The two following figures give a summary of the performances we obtained with the separate-sMRI strategy respectively using random forest and SVM to classify alcoholic patients as depressive or not:

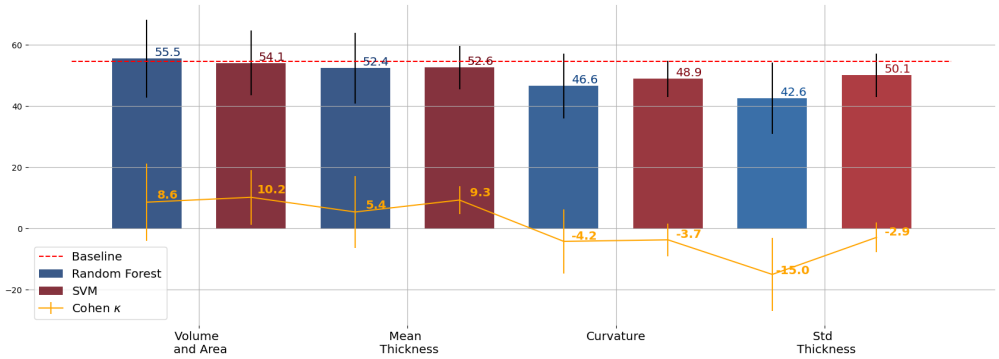

Figure 4.11: Accuracy and Cohen's κ using separate-sMRI strategy in order to predict depression of AUD patients.

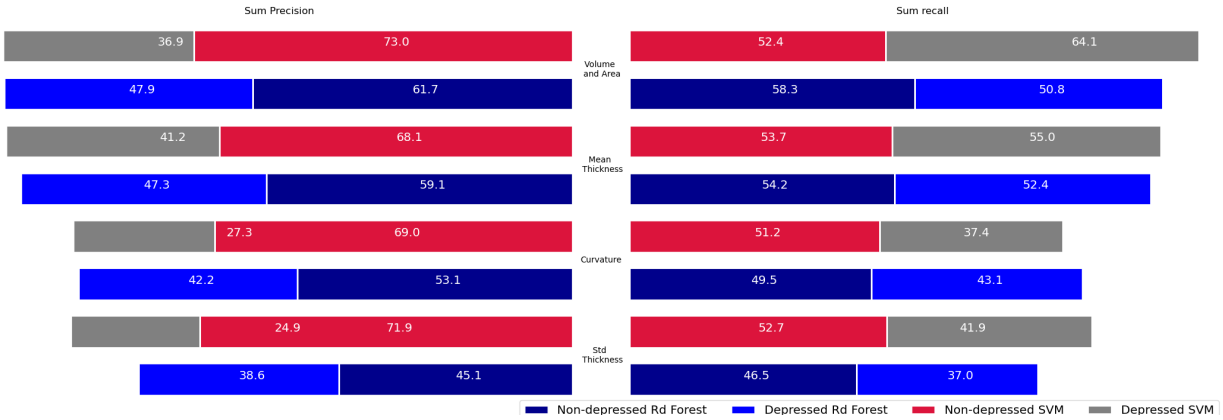

Figure 4.12: Recall and Precision using separate-sMRI strategy in order to predict depression of AUD patients.

There is no learning when using structural MRI separately and when no projection methods are used.

4.3 Resting EEG and ERP strategy

In this section, we present the results obtained in order to detect depression and anxiety from EEG data.

4.3.1 Problem n°4 : Detection of anxiety for alcoholic patients

The performances of the machine learning algorithms with the EEG and ERP datasets are not better than the baseline to predict anxiety, for a better overview of the results, see appendix C.3.1.

4.3.2 Problem n°5 : Detection of depression for alcoholic patients

The following figures give an overview of the performances of the machine learning implemented on the ERP and resting EEG datasets :

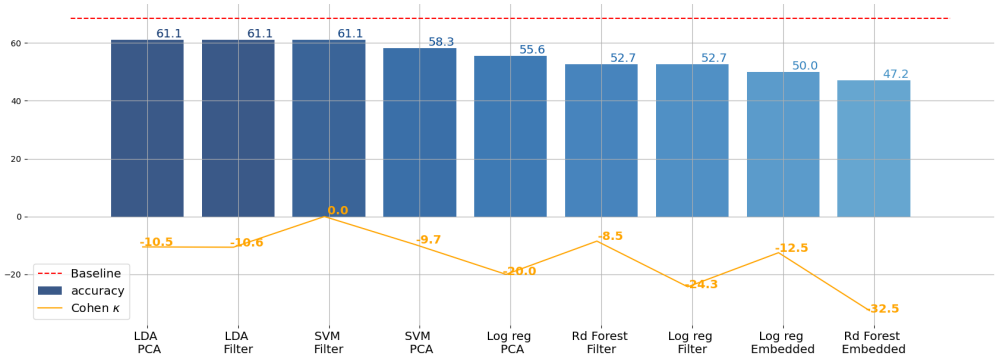

Figure 4.13: Accuracy and Cohen's κ obtained with resting EEG dataset in order to predict depression for AUD patients.

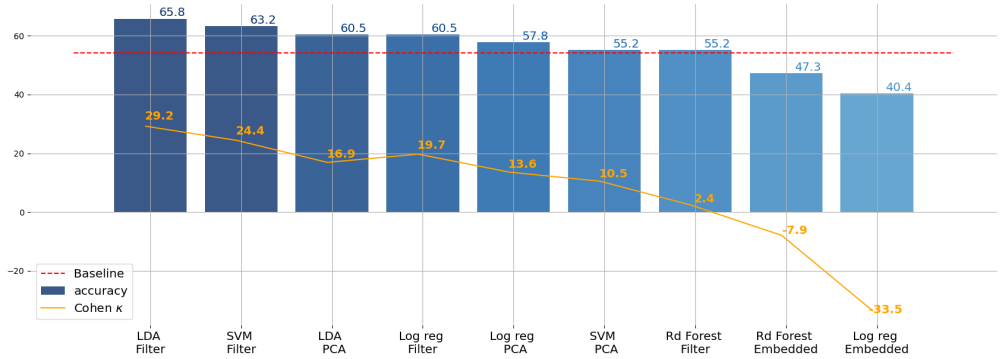

Figure 4.14: Accuracy and Cohen's κ obtained with ERP strategy in order to predict depression for AUD patients.

With the resting state EEG data, there is no learning, the accuracy is always under the baseline. However, with the ERP data, some methods have an accuracy higher than the baseline. In the next chapter, we will analyse if those results can be considered as significant or not.

4.4 Multimodal strategy using sMRI data

There was no real interest to use this strategy in order to predict anxiety as each SVM model classified patients as non-anxious due to the class imbalance. Still, we use this model in order to predict alcoholism and depression. The results of this strategy are shown on the next table :

Metric	Alcoholism	Depression for AUD	Depression
Accuracy	80.3 ±9.4	56.7 ±15.7	60.3 ±15.5
Precision illness	88.1 ±15.1	37.1 ±30.5	67.8 ±30.5
Recall controls	73.4 ±20.3	60.7 25.3	63.8 ±20.5
Recall illness	85.1 ±12.8	23.6 ±20.1	12.1 ±17.2
Cohen's κ	52.4 ±15.4	4.4 ±14.7	0.0 ±10.0

Table 4.1: Results of Multimodal structural MRI strategy in order to detect alcoholism and depression.

The results obtained with this strategy are worse than the best results obtained with the all-sMRI strategy. For the detection of alcoholism, some learning is made. This is not the case for depression.

Chapter 5

Discussion

This section is separated into 8 different parts, starting with a brief summary of the limitations of our work. It then looks at the results obtained for our 5 classification problems. A section is also dedicated to the choices we made for this project and what could have been considered. It continues with a section on why the use of EEG and/or MRI data is interesting for predicting psychiatric diseases, especially alcoholism. The chapter ends with the perspectives to be considered after having carried out this work.

5.1 Limitations

The limitations of the MRI-based dataset are first its size and also its unbalance. Compared to other state of the art datasets, 210 samples is relatively high for medical based problems. Still this a small number for machine learning algorithms, especially when the number of features is high. Feature selection methods ensure performances of the algorithms. However, in absolute terms, by dropping some features, we drop information that could be interesting for the algorithms. By increasing the amount of data, the machine learning algorithms will be able to learn the features that are important and will not need feature selection methods anymore, this could have be done with neural networks for example if the number of data was higher [151, 152, 153]. Furthermore, given class imbalance, even by using higher costs when the algorithm is wrong for the minority class, class imbalance is a problem since the algorithms will tend to overfit and make the most predominant class the most easily forecast to increase the accuracy of the algorithms.

The same limitations arise for the EEG-based dataset to an even larger extent. Indeed the dataset has a real small size, 37 patients. It will be difficult to be able to draw any conclusions from a so small dataset as learning is not really possible when the correlation between the features and the output is not straightforward.

Another limiting factor for classification of psychiatric disorders is the label noise of the psychiatric diagnoses [154] and also the heterogeneity that exists between psychiatric diseases, perhaps a little less so in terms of alcoholism, but which is quite high between depression and anxiety [155, 156].

One problem that arose for the MRI dataset was variables such as age and gender that were varying in our population. It has been shown that those features influence the brain structures in previous studies [121, 122]. As explained in section 3.3.3, one strategy to remove these correlations, for MRI features, is to compute a linear model relating age and MRI features and to take the residuals of the linear model as MRI features. This transformation of the MRI data was abandoned because it did not allow the machine learning algorithms to better classify the data. Moreover, to make sure that age did not influence the MRI features too much and that no transformation on them should be done, the median age of the dataset was calculated to create two subgroups, the first one containing patients older than the median and the second one containing patients that were younger than the median. After that, a SVM was implemented in order to classify subjects as younger or older than the median based on the MRI data. The method obtained an accuracy on the test set around 70%. If there was one variable which was highly linearly correlated with age, the accuracy could easily be higher than 90%, which is not the case. An accuracy around 70% (with a 50% baseline) stills means that age indeed influences the MRI data and could thus influences the performances of our classifiers, but slightly.

5.2 Classification of alcoholism

Classification of alcoholism could only be performed with the MRI-based dataset as the EEG-based dataset contained only AUD patients. As a reminder, the MRI-based dataset contains 210 patients, 141 AUD and 69 controls.

5.2.1 All-sMRI strategy

As it is shown on figures 4.1 and 4.2, the classification of alcoholic patients using structural MRI features is accurate. Of course, class imbalance plays a role as the precision and recall are higher for the AUD patients than for controls, but still the recall for controls is still acceptable. The algorithm performing the best is the embedded random forest, which is the most sophisticated algorithm. Its accuracy is equal to 82.8% with a baseline equal to 66% and its Cohen's κ is equal to 57.2%, meaning almost substantial agreement between y predicted and y test [149]. As the number of control patients was lower, weights were put such that misclassified control patients were more penalised. Thus, classification as control is preferred and this can be seen on figures 4.7 as the precision for controls is low and still the recall is high. These weights improve the results but are not necessary to classify patients as alcoholic or control. Methods using LDA have an accuracy 15% higher than the baseline without using class weights.

The good performances of our classifiers are mainly due to a number of structural MRI variables, called important variables. Filter methods highlight features which are statically the most different, the ones having the lowest p-value. An illustration of how the data behaves in the most important filter features space is shown on figure 5.1. It is impressive to see that even with two dimensions, namely the cortical thickness of the precuneus of the left hemisphere and the subparietal sulcus of the left hemisphere, clusters separating alcoholic patients from controls are created. Those two features are related to each other as the subparietal sulcus separates the precuneus from the cingulate gyrus, see appendixes A.1 and A.2. The separation is even increased when the feature dimension is increased on figure 5.1. Thinning, within precuneus, agrees broadly with the literature which has shown that the precuneus may be a cortical site for neuroplastic changes related to drugs (alcohol and cigarettes) [157] .

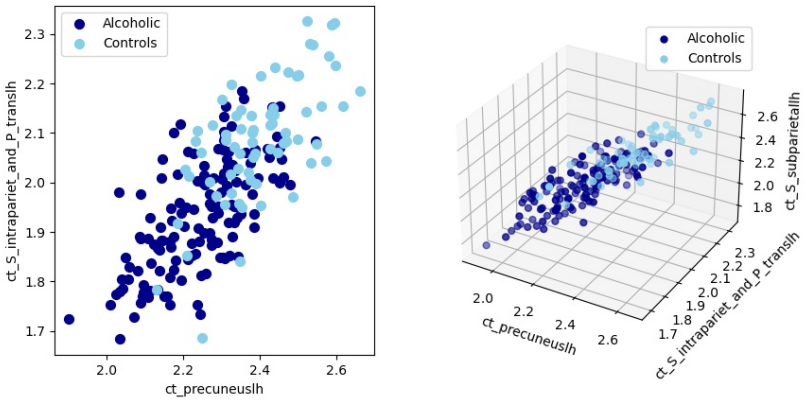

Figure 5.1: MRI data for alcoholic patients and controls projected in the most important variables plane for the filter method(p-value) showing strong data separation. For a better understanding of the abbreviations, see Appendices A.2 and A.1. The suffix ct_ means it is a cortical thickness variable.

Random forest also has feature importance methods to highlight the most important features. The 15 most important features that were highlighted from it are shown on figure 5.2. It is first interesting to see that the most important features given with Gini importance are similar to the ones reported with filter methods, not necessarily in the same order. By looking more closely at the features, we note that most important variables are cortical thickness features. This is very much in line with what we have seen in the literature which indicated a higher atrophy for alcoholic patients in the different parts of the brain [39, 42].

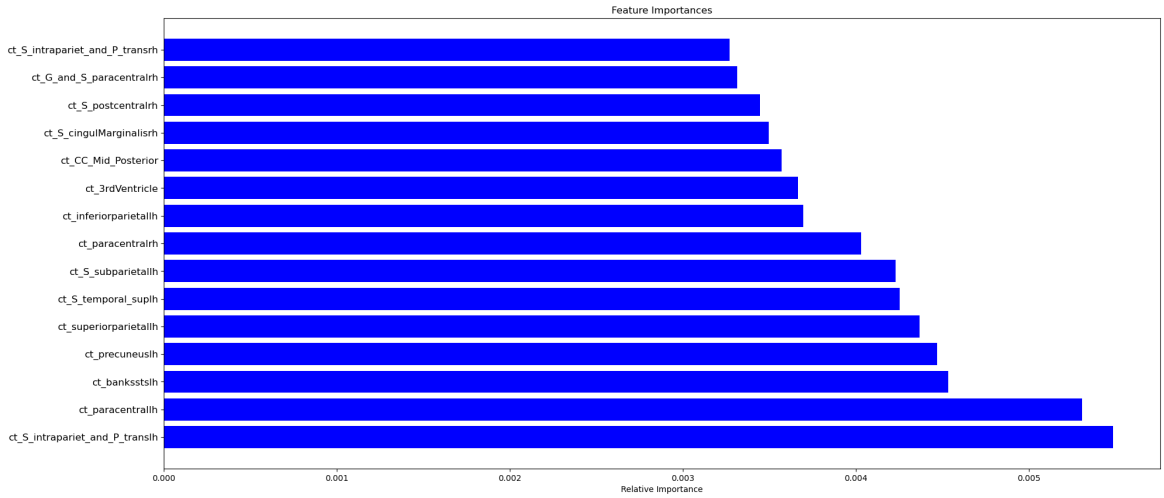

Figure 5.2: 15 Most important features highlighted by random forest for detection of AUD patients with MRI features highlighting cortical thickness features as a strong biomarker for the detection of AUD.

5.2.2 Separate-sMRI strategy

The importance of cortical thickness features for classification of AUD patients is also reported on figure 4.7 where the machine learning methods using the mean cortical thickness and white matter density of brain structures are highlighted to be the ones performing the best for the classification of patients as alcoholic or not (accuracy = 82.2% and Cohen's κ = 56.7%). The performances of the machine learning algorithms with the other features results obtained better than the baseline with a Cohen's κ higher than 20%, which means that there is potentially some learning curve with the other features.

It would be interesting to compare our results to what is currently done in the literature. A study used 5 MRI features: grey-matter density (GMDD), cerebrospinal fluid (CF) cortical thickness (CT), a basic reward response (RWR) signal from task-based functional MRI and nucleus accumbens connectivity (NAC) from resting-state functional MRI to classify subjects as alcoholic or control [158]. The structural MRI features were also extracted using `freesurfer` from T1 images with the Glasser anatomical atlas [159], which contains 358 regions. Their dataset contains 119 AUD patients and 97 controls, which is a more balanced set than ours. They used balanced accuracy to look at the performances of machine learning algorithms. It is given by the mean recall of each class. In our case, the balanced accuracy of our best method using all MRI features equals 81.0 %. With a multimodal approach, they obtained a maximum balanced accuracy of 79.3%. To do so, they separated their data into 5, (see figure 5.3), each data type gave rise to a SVM with radial basis. They then used the outputs of the SVMs to feed a perceptron.

In their study, the feature which gave the best performances was the grey-matter density (volume for our study) with a balanced accuracy of 76.3%, which is better compared to us with this modality, balanced accuracy of 68.6%. However, using cortical thickness they did not obtain a balanced accuracy better than 65%, which is really low compared to our performances using cortical thickness, 81% of balanced accuracy. There are three reasons that could explain why our studies presented different results. The first reason could be that they suffer from overfitting as they do not appear to have used feature selection methods. In addition, they did not use exactly the same variables as we did, as they did not use the same atlases. Finally, our datasets are both small and it is possible that the populations we used are not representative and therefore different. Overall, mean cortical thickness still appears to be a more important biomarker of alcoholism than grey matter density. In fact, a new study using structural MRI data to classify patients as alcoholic or control also identified cortical thickness and area of structures as the most important variables. This study included more than 2000 subjects (Enigma) and obtained with these important variables 77% area under the receiver operating characteristic curve [160].

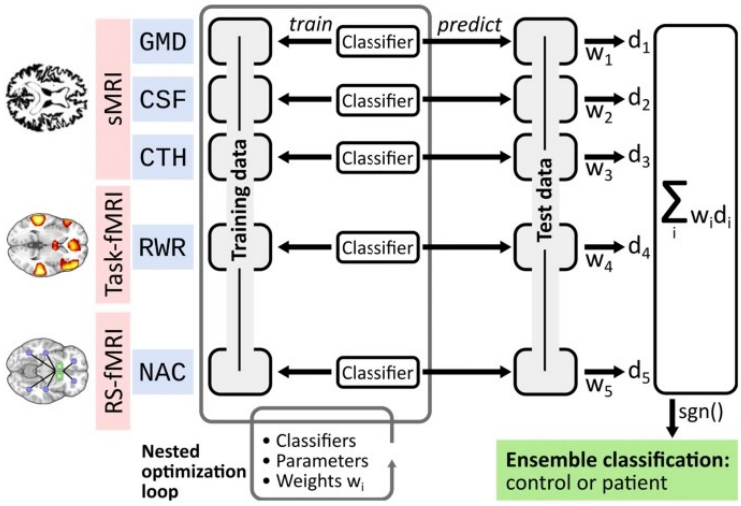

Figure 5.3: Multi modal neuroimaging method using NAC, RWR, CTH, CSH GMD data to classify people as alcoholic or not from [158].

5.2.3 Multimodal strategy using sMRI data

The strategy shown on figure 5.3 is similar to our multimodal strategy using structural MRI data. As a reminder, we used the probability of belonging to a class, deduced from the distances of the samples to the SVM hyperplane, trained with four structural MRI modalities, as input of a perceptron. The results obtained, 80.5% of accuracy and a Cohen's κ of 52.4%, are similar (less accurate) than the ones obtained with the all-sMRI strategy or the separate-sMRI strategy. SVM with features other than mean cortical thickness data perform slightly better than the baseline and therefore do not estimate a probability density close to reality. Figure 5.4 shows the probability densities estimated from the SVM models fitted with curvature data, standard deviation of the cortical thickness and volume and area data compared to the mean cortical thickness. Probability higher than $\frac{1}{2}$ means that the classifier will assign this element to the alcoholic class. In this figure, we can see that when features are classified in the same way by both types of data, on the bottom left and top right of each image, very few elements are misclassified. Conversely, when the results obtained by the different types of data are different, in the top left and bottom right of the figure, the predictions from the mean cortical thickness data are less wrong. When the results obtained with the other types of data are correct (and thus the one with cortical thickness incorrect), the estimated probability that these data are correct are close to $\frac{1}{2}$ and therefore do not have enough weight to be able to influence the choice of the perceptron. Moreover, the perceptron's weight for probability obtained from the mean cortical thickness data is 4 times higher than the others. The perceptron therefore relies mainly on this input to perform its classification. The outputs of the other classifiers are also used to perform the prediction but to a lesser extent. Given the lower performances of this model, this could indicate what we said earlier, that there is very little learning from data other than mean cortical thickness.

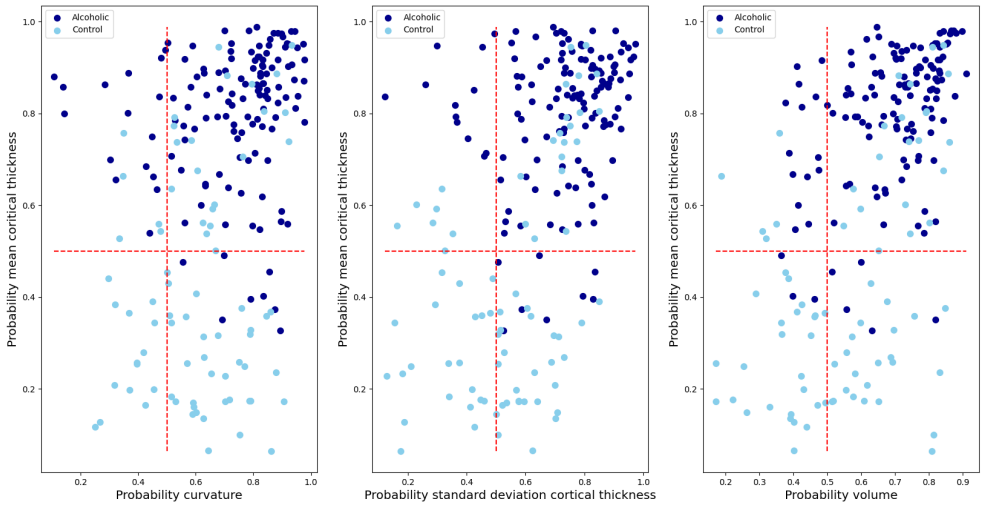

Figure 5.4: Probability estimation to be alcoholic according the four different types of MRI features. Probabilities obtained with SVM using other features than mean cortical thickness are not accurate enough improve the accuracy of the model obtained with mean cortical thickness.

5.3 Classification of anxiety state

Classification of anxiety state could only be performed with the MRI-based dataset as the EEG-based dataset contained only AUD patients. As a reminder, the MRI-based dataset contains 210 patients, 44 subjects in anxious mood and 166 non-anxious.

The classification of anxiety for all patients with structural MRI features was not a success. Indeed it is possible to see on figures C.1 and C.3 that there are not any machine learning that had an accuracy better than the baseline. It means that the best way for the machine learning to have the best accuracy was to classify all elements as control. This can be seen through the Cohen's κ almost equal to 0 and can maybe be explained by three different reasons.

The first one is the class imbalance problem as there is almost 80% of the total population that is control, even if the algorithm does find some differences between both groups, it is safer for it to classify all elements in the same class. "Even by using methods which counter imbalancement, the learning process of most classification algorithms is often biased toward the majority class examples, so that minority ones are not well modeled into the final system "[161].

The second reason why the classifier does not perform well can be explained by the fact that detecting anxiety with structural MRI information is not an easy task. For example, a study involving 19 patients with GAD and 24 controls used GM and WM areas and volumes as input to a SVM to predict patient anxiety [162]. They obtained a classification accuracy of 68% which is higher than the baseline 55%, but not statistically significant. Plus, the way the machine learning algorithm was evaluated is not really clear. Indeed to evaluate their performances, they used Leave-One-Out and it is not clear whether they used this method separately to find the hyperparameters of the model and to test their performances[1].

The final reasons that may influence the results of our algorithms are related to the way we have considered patients as anxious. Indeed, as the STAI-state questionnaires were reported before to make a MRI scanner which is a stress full moment, it could increase the STAI-state score of patients [163]. Moreover anxiety state

[1]If the test set is dependant of the validation set, their performances will be better as their algorithm will learn on the hyperparameters of the model.

is a comorbid state with the begin of alcohol withdrawal and may disappear later [164], such that people who scored above 50 at the STAI-state questionnaire are maybe not anxious but just in an anxiety state due to their withdrawal. The use of questionnaires could also distort the results because patients answer these questionnaires on their own, whereas the best way to classify a person as anxious is through a psychiatrist. Furthermore, as noted in section 2.3.3, the questionnaires classify all anxious patients in the same class, even though different types of anxiety do not affect the brain in the same way. In addition, the STAI-state questionnaire may not be the best questionnaire to indicate the anxiety level of patients. A study claimed that the most correlated questionnaire with anxiety was the STAI-trait questionnaire and not the STAI-state [165]. Finally, our "control" group for this problem also contained patients with alcohol use disorder. It has been proven in the state of the art that these two diseases affect certain parts of the brain in the same way and are also closely related.

Those are limitations, but differences are still observed between the MRI features of anxious and non-anxious patients, see appendix D.2. Thus, one might ask how it is possible that our algorithms do not work well when we notice sMRI statistically different variables between anxious and non-anxious patients?

To answer the previous question, let's have a look at a simple example using one feature and a simple classifier which uses thresholding (this can be seen as a decision tree) and containing two populations each containing 100 samples. Let's look at three different cases. In the first case, a simple t-test is performed to distinguish the means of the two groups and a p-value of 0.001 is obtained which is statistically significant (<0.05). However, weak performances of the classifier are observed, see figure 5.5.A . This because the values for that feature overlap each other. This means that even if a feature is significantly different from one group to another, it does not imply high classification results. However, the opposite is not true either and can be seen on figure 1.B. In this case, the thresholding cannot be done with only one threshold, although with only one threshold the results will be better than for example A. Finally, it is still possible to have a feature that is significantly different and that allows to have quite high classification performances (see picture C) [166].

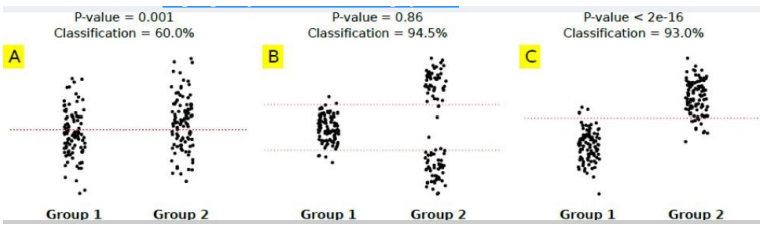

Figure 5.5: Statistically-significant differences do not necessarily imply high classification accuracy, and conversely, from [166].

5.4 Classification of depression

Classification of depression could only be performed with the MRI-based dataset as the EEG-based dataset contained only AUD patients. As a reminder, the MRI-based dataset contains 210 patients, 69 subjects in depressive mood and 141 non-depressed.

As shown on figures 4.3 and 4.9, the classification of depression for the whole population did not work well. The best method consists in using LDA and using Principal Component Analysis in order to reduce the number of dimensions. This method has an accuracy which is near the baseline (68%). We have a balanced accuracy of 63% which could indicate learning from the MRI data, but it is not the case as the Cohen's κ is really low. This method is not reliable and could thus not be taken into account for psychiatric diagnostics. The reasons why the performances of our classifiers for this problem are bad can potentially be explained by the choice of the input data and our dataset.

Depression is an illness which mainly affects the functionality of the brain. Studies conducted on major depression implementing machine learning algorithms to classify MDD patients vs controls showed better

performances with fMRI data than sMRI data. For example, a study conducted to classify MDD from controls obtained 74% of accuracy with a population of 30 controls and 30 MDD. The input features were based on a combination of fMRI biomarkers, which corresponded to the neural responses of individuals to neutral faces, important rewards and safety [167]. Indices derived from rs-fMRI such as the Hurst exponent seem to be predictive biomarkers. Two studies used those biomarkers to predict depression. The first study used 40 subjects, 20 controls and 20 MDD and obtained 90 % of accuracy with a SVM [168]. The second study did not only try to classify people as depressed or not, but also tried to classify patients as remitted or not. To do this, they had access to the Hurst exponent of fMRI signals of 9 current MDD, 19 remitted MDD and 19 controls. By using a SVM with a radial basis function, they obtained 81% of accuracy between remitted MDD and controls, 85 % of accuracy between current MDD and controls and 78% of accuracy between remitted MDD and current MDD [169]. This shows how powerful the Hurst exponent is in order to detect major depression even with small populations.

Some studies focused on classifying major depressed individuals compared to controls by using structural MRI biomarkers and their performances were not as good as with fMRI data. For example, a study involving 33 teenagers, 18 who get depressed within the 5 years and 15 who did not, used cortical thickness of brain regions of these teenagers to predict whether they will suffer from depression within 5 years and got 69% of accuracy. Those performances are certainly seen on the rise as they obtain those scores by cross-validation [170]. Other studies, using volume of brain structures as input to detect major depression obtained accuracy which ranged between 64 and 85% [171, 172, 173, 174]. The same purpose was conducted with another study, they used PCA and locally linear embending methods in order to extract the most important features of gray and white matter volumes with a SVM and obtained respectively 85.7 and 77% of accuracy [175]. Finally, a last study using cortical thickness and curvature with a SVM obtained respectively 69% and 48% of accuracy [172].

Those results are worse than the one obtained with classifiers using fMRI biomarkers. However, most of the studies using structural MRI data in order to classify patients as MDD obtained better results than the baseline. Other reasons linked to our dataset could explain those poor performances.

First of all, it is important to remember that depression and alcoholism are strongly correlated [52, 53, 55]. Most of the subjects considered as depressed are also alcoholic in our study. The Beck scores of the subjects considered as depressed for our study are really low compared to MDD, but higher than 8, synonym of a low depressive state, which is characteristic of alcoholic patients at the end of their withdrawal [176]. This depressive state is often temporary and tends to fade gradually [176]. Thus, the "depression" (the depressive state) that we describe in this thesis may not have severe effects on the brain and therefore may not be visible on MRI images.

In addition, studies that looked at depression compared depressed patients and healthy subjects. Our "control" group for this problem also contained patients with alcohol dependence, potentially influencing the results. The purpose of this problem was therefore different from what we see in the literature and the results are hardly comparable. Our aim was to see whether alcohol-related depression had sufficiently large effects on the brain to be visible on MRI, compared to alcoholic patients and healthy subjects.

Finally, the Beck scores indicate depression but they are not 100% reliable. Indeed, those questionnaires are personal and do not always represent reality. The people who are considered as depressed in our survey are not clinically depressed, at least they are not considered as depressed by a psychiatrist, which is currently the safest way to diagnostic people as depressed. The "sick" and "control" subjects for this study are therefore perhaps poorly considered.

5.5 Classification of anxiety for alcoholic patients

For this problem, we could use the MRI-based dataset, containing 141 AUD patients, 26% of them being in an anxiety state; the resting EEG-based dataset containing 36 patients, 16% being in an anxious mood and the ERP-based dataset containing 38 patients, 18.5% of them being in an anxious mood.

Figures C.2, C.4, C.5 and C.6 show evidence that the prediction of anxiety for alcoholic patients is not something that has been full-filled with the MRI and or EEG features.

The resting EEG dataset was strongly unbalanced, 6 anxious and 30 controls. The results obtained were not better than the baseline, but it is difficult to draw any conclusions from those results given the size of the dataset. Other studies have found alpha symmetry to be an effective marker for rEEG-based classification of anxiety in AUD patients [177, 178]. Alpha symmetry features is therefore a promising avenue for future research on our data.

The classification of anxiety state with structural MRI data for alcoholic patients did not show interesting results either and the reason explaining that are similar to what is explained in section 5.3 :

- **Temporary anxiety** : Anxiety, like depression, is often associated with the onset of withdrawal and symptoms tend to decline gradually [164, 179].
- **Type of questionnaire**: The STAI-status questionnaire may not be the best questionnaire for assessing anxiety, the STAI-trait is more accurate [165].
- The **use of questionnaires** may bias the results as the questionnaires are done personally by the patients and not by a psychiatrist.
- The **strong class unbalance**.

5.6 Classification of depression for alcoholic patients

For this problem, we could use the sMRI-based dataset, containing 141 AUD patients, 44% of them being in a depressive mood, the resting EEG-based dataset containing 36 patients, 33% being in a depressive mood and the ERP-based dataset containing 38 patients, 42% of them being in a depressive mood. As a reminder, for this problem, patients considered as depressed alcoholic are those who answered the BECK questionnaire and obtained a score higher than 8 in T2. The scores of the patients considered as depressed are close to 8, associated with the end of withdrawal for a large number of alcoholic patients, more prone to relapse [180, 181, 182]. This depressive state is often temporary and tends to fade gradually and may therefore affect the brain only slightly. In addition, Beck scores are self-completed questionnaires and are therefore not always 100% reliable. These two factors are certainly the main reasons why our classifiers do not achieve significant performances. This section is separated in three subsections, each explaining the results obtained with the 3 different types of data.

5.6.1 MRI based dataset

As shown on figure 4.5, a PCA SVM model with sMRI features obtained 62% accuracy, an improvement of 6% compared the baseline 56%. However, the 10% confidence interval of the model's test-set accuracy contained that baseline performance, which we attribute to our limited sample size. Moreover, by using features separately and without any projection methods, see figure 5.11, the machine learning methods do not have an accuracy higher than the baseline. It is thus difficult to indicate that it is possible to use structural MRI features in order to detect depression for AUD patients.

However, there are variables that are statistically significantly different between depressed AUDs and AUDs for the whole population. The most significant variables that are retrieved are mainly changes of volume. We can see that there is a reduction in the volume of the cingulate gyrus for AUD depressed patients compared to AUD patients, see figure D.26. This was also suggested in different studies [43, 71, 72] which analysed structural MRI differences between controls and major depressed patients. The posterior and the isthmus of the cingulate cortex are the most affected areas. The posterior cingulate cortex cortex is used to evaluate emotions and to concentrate [183, 184], while the isthmus is involved in memory and pain processing [185]. This may potentially indicate a correlation between major depression and depression related to the end of withdrawal.

5.6.2 ERP-based dataset

As evidence on figure 4.14, event-related potential features fed to machine learning algorithms obtained a mean accuracy higher than 60%, thus improving on the baseline 58%. The best method obtained 66%, 8% above the baseline. While encouraging, it is worth noting that an 8% improvement only represents 3 subjects, given the size of our dataset. However it does not look like at it as both logistic and SVM methods were able to highlight those results. To analyze these results, consider a simple classifier that would assign the AUD control label to patients with probability $p = \frac{21}{38} \simeq 0.55$ and the AUD label in a depressed state with probability $q = 1 - p = \frac{17}{38} \simeq 0.45$. These numbers are not trivial and correspond to the proportion of each class in the ERP dataset. The probabilities that an element is well classified is given by the probability to pick a depressed AUD subject and to classify it as depressed AUD + the probability to pick an non-depressed AUD patient and to classify it as non-deprresed AUD : $\mathcal{P}(\text{well classified}) = p^2 + q^2$. The probability of misclassification is thus given by : $\mathcal{P}(\text{misclassified}) = pq + qp$. the sum is equal to 1 since $q^2 + p^2 + 2pq = (q + p)^2 = 1$. The probability of correctly classifying exactly k patients with this classifier using a Leave-One-Out method for testing with the ERP dataset is given by the mass density of a binomial distribution.

$$\mathcal{P}(X = k) = \binom{N}{k} \cdot (1 - 2pq)^k (2pq)^{N-k} \tag{5.1}$$

.

Where N, the number of samples equals 38. In our case, we are interested in the probability that this classifier has a number of correct answers greater than 25 so that its accuracy is greater than $\frac{25}{38} \simeq 65.7\%$. This is given by

$$\mathcal{P}(X \geq 25) = \sum_{i=25}^{38} \mathcal{P}(X = i) = 1 - \sum_{i=1}^{24} \mathcal{P}(X = i) \simeq 5\% \tag{5.2}$$

Equation 5.2 shows that our results could not be considered as statistically significant, but at least encouraging. The non reliability of this model was already highligted with the low value of Cohen's κ, 29.5 %. The fact that the use of ERP (mostly P300) features in order to detect depression is encouraging has already been noted in the literature [64, 65, 66]. However, in those cases, the studies compared patients with MDD and healthy controls and not AUD patients with minor depression and AUD patients.

5.6.3 Resting EEG dataset

Finally, by using resting EEG features, we can not draw any conclusion in order to assess that AUD patients are in a depressive state or not. The non-correlation between PSD of patients and Beck and STAI-State score was already highlighted in a previous study conducted on resting state EEG of AUD patients at St-Luc [186].

5.7 Choices

This section focuses on the machine learning choices that were made for this study and explains what other algorithms could have been used, what other metrics could have been analyzed.

5.7.1 Choice of the classifier

The choice of the different classifiers used are based on our literature research and on the size of the dataset, as we would have used deep learning if the number of samples was higher. The SVM algorithm is certainly the most widely used algorithm in order to assess brain diseases, often with kernel tricks [166][2]. This is also the case for LDA, but at a lower level [166]. The choice of the other different methods are based

[2]Gaussian kernel was tried for the classification problem of alcoholism with the MRI dataset, as the results (80.5% of mean accuracy for classification) were not better than the ones obtained with linear kernel (initial SVM) and as the number of presented results were already high, we decide not to show them. The fact that simple models (linear SVM) perform better than non-linear models is completely in line with what we see in the literature when the number of data is low [187]. But, they get really outperformed when the number of samples increase (>500), and if other studies on the subject using more data should use SVMs or kernels or neural networks as they allow more learning but are more prone to overfitting [187] (those improvements also depend on the dataset and correlation between y and X are sometimes linear such that non linear kernel do

on their properties. The main advantage of the random forest algorithm is its ability to evaluate the most important features, which is an interesting tool in the medical field, as it could help doctors to examine certain features when diagnosing patients. The main advantage of logistic regression is its output meaning. Indeed, the output of the logistic regression is a probability level that indicates the level of class membership, which could also help patients when diagnosticating as doctors thus have a meaning of the output and not simply a label.

Other more methods in the medical field are often used such as K-nearest neighbor. K-nearest is one of the simplest classification method, the only hyperparameter to tune is the number of neighbors. The method assumes that the predictor function is locally smooth such that the decision function of each point can be approximated by its k nearest neighbors. In practice, this method is the simplest form of ensemble classifier as it simply assign a label by the majority vote of its k nearest neighbor. This method has not been considered since it tends to highly suffer from the curse of dimensionality. Naive Bayes could also have been considered, as it also often used for machine learning algorithms using medical data [166]. It is often associated to logistic regression in the literature, as both are based on a probability distribution. The Naive Bayes classifier uses Bayes rule and maximises the maximum likehood a posteriori, while the logisitic regression model maximises the maximum likehood in order to find w.

5.7.2 Choice of the performance metrics

In machine learning studies related to disease detection, the most used performance metric is certainly the accuracy [166]. However, even in this study, this performance metric showed its limitation when the dataset is unbalanced. Some studies thus put interest in balanced accuracy to assess the performances of their classifier. It has also shown it limitation as it does not indicate which class is well labeled and the one which is not. That is why, we prefered to highlight accuracy, precision, recall of each class and the Cohen's κ.

Still, other performance metrics could have been used in order to condensate the results of our classifiers. First, let's introduce the F-measure, introduced at the Fourth Message Understanding Conference [188]. The F_1 score is given by the harmonic mean between the recall and the precision:

$$F_1 = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} \tag{5.3}$$

The F-score can also be adapted in order to put more weight on the recall or on the precision, we then talk about F_β. The F-score enables to give the global overview of the precision and the recall of one class and has the advantage of being more concise than the recall and precision calculation. However, it does not allow to know exactly how accurate the precision and recall of a class are and this is why when it is possible to indicate several results, it is preferable to calculate these two metrics separately.

In the medical field, another metric is often used, called the Receiver Operator Curve (ROC). The ROC curve shows the performances of a classifier by showing the TP rate in function of the FP rate while varying a hyperparameter. From this curve, it is possible to extract the area under curve metric, it ranges between 0.5 and 1, where the value 0.5 is a random classifier and the value 1 means a perfect classifier. The area under curve metric indicates the discriminant capacity of the classifier. It indicates the probability that a patient would be well classified compared to a control. A quick interpretation of the area under curve is shown on figure 5.6.

not always outperform simple models when the number of samples increase). SVM with a Gaussian kernel looks like to be a promising avenue for future research.

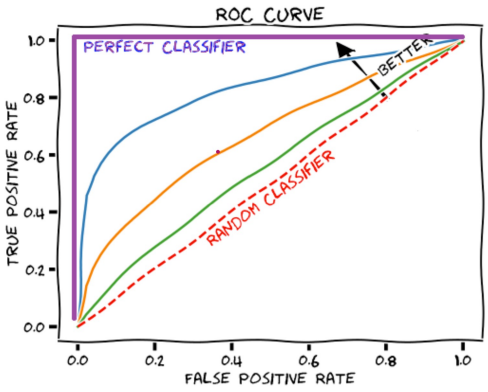

Figure 5.6: Best area under curve is given in top left corner of the graph, while worst area under curve is a straight line from (0,0) to (1,1).

5.8 Interest of using EEG and MRI features for the prediction of diseases

This section tries to explain what could be the interest of using the machine learning algorithms presented previously on EEG or MRI data to solve the classification tasks presented in section 3.1.1, this section is the result of a discussion with Sophie Leclerq and exchanges with psychiatrists such as Dr de Timary. In fact, there is not much interest in using machine learning algorithms on EEG and/or MRI data (that are difficult to extract) in order to predict a level of depression, anxiety or alcoholism which are not so difficult to collect with questionnaires.

The interest of using these EEG and/or MRI data to indicate diseases such as alcoholism, anxiety and or depression is nevertheless important. Let's focus on alcoholism for which the relapse rate is high (about 80% after 1 year) and for which drug treatments do not work well [37]. It is necessary to find new biomarkers to predict relapse. By proving with this work, that there are biomarkers that can predict alcoholism (these biomarkers may be the effects of alcohol and may not be the reason why people drink, these can be identified with feature importance methods discussed in section 3.5.2), it can be envisaged that some of these biomarkers can identify relapse of patients and that they could be modified by an intervention. For example, using microbiota, see Appendix E for the link between resting EEG and microbiota, one could imagine that certain bacteria living in the gut could be "predictors" such as potentially intestinal permeability, potentially intestinal permeability [189]. If this is the case, there is a way to vary their levels via nutritional approaches [190]

- A probiotic approach: giving the live bacteria orally (in the form of food supplements for example).
- A prebiotic approach: giving dietary fibers that would allow the growth of these "good" bacteria.

However, since measuring intestinal permeability is a tedious task (24-hour urine collection, ingestion of radioactive products), using an EEG or MRI markers rather than measuring intestinal permeability would be interesting (in terms of saving time and money, and safety for the patient and the clinician).

5.9 Future perspectives

This thesis was initially a proof of concept to see if it was possible to use EEG and/or MRI data to improve the management of alcohol-dependent patients at St-Luc Hospital. The ultimate goal of Prof. de Timary *et al.* is to use the data they collect at T1 and T2 to predict the physical state of alcoholic patients at T3. The major objective would be to predict which patients will relapse, knowing that today more than 80% relapse within a year. To this end, all type of information could be collected:

- Personal: age, sex, financial situation, marital status etc..
- Alcohol: Type of consumption, average quantity per week, medication, number of withdrawals, etc..

- Questionnaires in T1 and T2: BECK, AUDIT, STAI scores.
- Medical examinations: EEG, ERP, MRI, microbiota[3] in T1 and T2.

Since very few studies have succeeded in predicting relapse in alcohol-dependent patients, we cannot rely on any one type of data. Collecting all the types of data listed above will require a lot of effort and will create a very large database. For this study, we already had more than 1800 different variables counting only the MRI data. Before we can hope to predict patient relapse from machine learning algorithms, it is necessary to identify a limited number of markers that are potentially predictive of relapse.

It is therefore necessary to set up a systematic collection of data from alcoholic patients returning to hospital for withdrawal. One of the major difficulties encountered in this work was to collect all available patient data. A large proportion of the patients admitted to the psychiatric ward had EEG examinations, but it was very difficult to find their clinical diagnosis.

To carry out this project, we could imagine a common server for the whole psychiatric department. This server would contain a very simple database, a list of folders. Each folder would be specific to a patient and would contain four sub-folders, containing the files related to the name of the folder:

- Questionnaires and medical examinations in T1
- Questionnaires and medical examinations in T2
- Questionnaires and medical examinations in T3
- Personal and alcohol-related information

After setting up this common server, it would be necessary to collect this information for a large number of patients, at least 300 or 400 as the proportion of patients who do not relapse is really low. This work would be long and tedious and it might be interesting to use only the EEG data as a medical examinations since these are the only reimbursable medical examinations, see step 2 on figure 5.7.

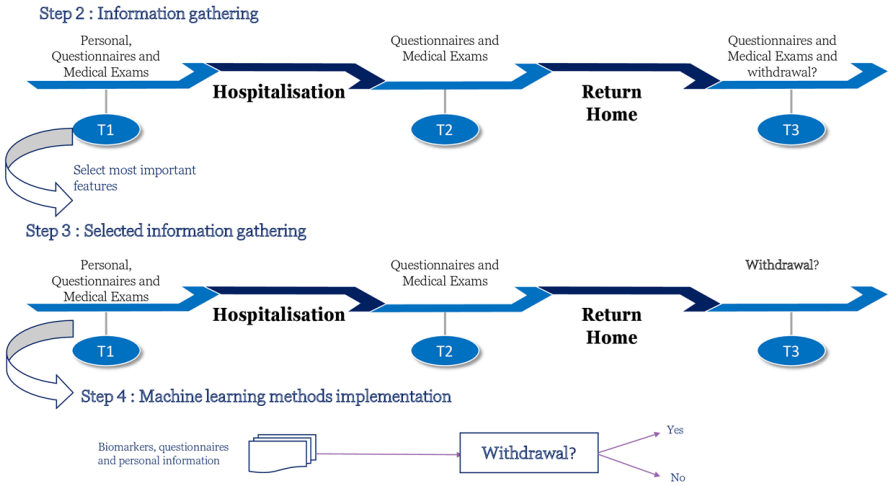

Figure 5.7: Three last steps of the potential future project : Predicting relapse of alcoholic patients.

Once this information has been collected and knowing the relapse output, it will be possible to identify some markers potentially predictive of relapse. The collection of biomarkers would make it possible to identify the differences between those who relapsed and those who did not relapse and how these also evolved during withdrawal.

[3]This type of information may be removed due to the difficulty of sampling.

Then, data should continue to be collected but only for those markers that were identified important in the previous step, this step corresponds to step 3 in the figure. This would save time and money.

Finally, when a sufficient number of data has been collected, about 1000 patients[4], a machine learning algorithm could be used to link all the predictive markers together and potentially predict patient relapse, see step 4 on figure 5.7.

[4]There is no real rule for how many samples are needed, but it is known that the more data, the greater the learning potential. Furthermore, given the large amount of data that could be collected and the class imbalance, 1000 patients seems to be necessary to be able to draw conclusions.

Conclusion

In this work, we were interested in the links between EEG and structural MRI data and certain psychiatric diseases. More specifically, we were interested in the use of machine learning methods using EEG and/or structural MRI data to classify alcoholic and control subjects, depressive and control subjects, anxious and control subjects, alcohol-depressive and alcoholic subjects, alcohol-anxious and alcoholic subjects.

To do so, we first conducted a literature search on EEG and MRI signals, specifically on biomarkers that could be extracted to study these diseases. We the developed two different datasets containing MRI and EEG features studied in the literature research. The first one contains structural MRI data for 210 subjects. The second one contains electrophysiology data for 43 alcoholic patients. We then researched the classification methods that would be most suitable for these data and applied them to our data.

The results of the implemented machine learning methods obtained from the structural MRI data showed that the brain structures were significantly different for alcoholic and control patients and that classification regarding alcoholism from structural MRI data was possible. More precisely, this work highlighted that cortical thickness data were the most interesting variables for the prediction of alcoholism. In addition, classification via sMRI data for depression in alcoholic patients showed encouraging but not yet statistically-significant results due to the small number of data available.

The classification according to anxiety and depression did not allow to draw interesting conclusions from sMRI and EEG data in resting state. However, classification of depression of alcoholic patients from ERP data yielded encouraging results. Given the number of subjects, these results do not allow us to conclude a significant difference. The classification of anxiety of alcoholic patients based on ERP data did not allow to conclude significant differences between anxious and alcoholic patients.

The first limitation of this work was the number of subjects for our two datasets, which if new studies on the same topic were to be initiated, would have to be increased. On the one hand, this would allow for more learning and better results and on the other hand it would allow for more certainty in our results. One study indicated that for this type of study to have a confidence interval of 0.1, the number of test items should be greater than 140, i.e. a set of at least 500 subjects with our method fo testing [191].

In addition, it would be interesting to change the way patients were classified as depressed or anxious. If diagnoses could be made by psychiatrists and not by answers to questionnaires, this could improve the quality of our data. This would certainly lead to better performances for the machine learning algorithms and also allow more certain conclusions to be drawn.

However, even with these limitations, the results we have obtained with our two datasets are already very encouraging. They indicate that EEG and structural MRI data could be used to help diagnose psychiatric patients and their potential should be exploited to design new treatments. The EEG and MRI data could be used to study more complex problems such as relapse in alcohol-dependent patients, see section 5.9.

Bibliography

[1] Coenen, A., Fine, E., Zayachkivska, O. (2014). Adolf Beck: A forgotten pioneer in electroencephalography. *Journal of the History of the Neurosciences, 23*(3), 276-286.

[2] Structure of a neuron, `https://owlcation.com/stem/Structure-of-a-Neuron`, [Accessed 10-04-2021].

[3] Lippold, O. C. J. (1977). Mass action in the nervous system—Examination of the neurophysiological basis of adaptive behaviour through the EEG: By WJ Freeman, xx+ 489 pages, 191 illustrations, 18 tables, Academic Press, New York, San Francisco, London, 1975, US. *Journal of the Neurological Sciences, 32*(1), 146.

[4] Grabowski, S. R., Tortora, G. J. (2000). *Principles of anatomy and physiology*. New York, NY: Wiley.

[5] Nunez, P. L., Srinivasan, R. (2006). *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA.

[6] Lopes da Silva, F. H., Storm van Leeuwen, W. (1978). The cortical alpha rhythm in dog: the depth and surface profile of phase. *Architectonics of the cerebral cortex*, 319-333.

[7] Nunez, M. D., Nunez, P. L., Srinivasan, R., Ombao, H., Linquist, M., Thompson, W., Aston, J. (2016). Electroencephalography (EEG): neurophysics, experimental methods, and signal processing. *Handbook of neuroimaging data analysis*, 175-197.

[8] Sanei, S., Chambers, J. A. (2013). *EEG signal processing*. John Wiley Sons.

[9] Acharya, J. N., Hani, A. J., Cheek, J., Thirumala, P., Tsuchida, T. N. (2016). American clinical neurophysiology society guideline 2: guidelines for standard electrode position nomenclature. *The Neurodiagnostic Journal, 56(4)*, 245-252.

[10] Malmivuo, J., Plonsey, R. (1996). Bioelectromagnetism. *Medical and Biological Engineering and Computing, 34*, 9-12.

[11] Malik, A. S., Amin, H. U. (2017). *Designing EEG experiments for studying the brain: Design code and example datasets*. Academic Press.

[12] Roohi-Azizi, M., Azimi, L., Heysieattalab, S., Aamidfar, M. (2017). Changes of the brain's bioelectrical activity in cognition, consciousness, and some mental disorders. *Medical journal of the Islamic Republic of Iran, 31*, 53.

[13] Baars, B., Gage, N. M. (2013). *Fundamentals of cognitive neuroscience: a beginner's guide*. Academic Press.

[14] Engel, A. K., Fries, P. (2010). *Beta-band oscillations—signalling the status quo?. Current opinion in neurobiology, 20*(2), 156-165.

[15] Sur, S., Sinha, V. K. (2009). Event-related potential: An overview. *Industrial psychiatry journal, 18*(1), 70.

[16] Molfese, P. J. (2012). Imaging studies of reading disabilities in children. Reading, *Writing, Mathematics and the Developing Brain: Listening to Many Voices*, 49-64.

[17] Dalecki, A., Croft, R. J., Johnstone, S. J. (2011). An evaluation of P50 paired-click methodologies. *Psychophysiology, 48*(12), 1692-1700.

[18] Kuntz, A., Missonnier, P., Merlo, M., Gothuey, I. (2018). Clinical utility of evoked potentials in the field of addiction medicine. *Revue medicale suisse, 14*(610), 1189-1192.

[19] Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of clinical neurophysiology, 9*(4), 456-479.

[20] Squires, N. K., Squires, K. C., Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and clinical neurophysiology, 38*(4), 387-401.

[21] Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology, 118*(10), 2128-2148.

[22] Patrick, C. J., Bernat, E. M., Malone, S. M., Iacono, W. G., Krueger, R. F., McGue, M. (2006). P300 amplitude as an indicator of externalizing in adolescent males. *Psychophysiology, 43*(1), 84-92.

[23] Fjell, A. M., Walhovd, K. B. (2004). Life-span changes in P3a. *Psychophysiology, 41*(4), 575-583.

[24] Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., Winter, A. L. (1964). Contingent negative variation: an electric sign of sensori-motor association and expectancy in the human brain. *Nature, 203*(4943), 380-384.

[25] Weerts, T. C., Lang, P. J. (1973). The effects of eye fixation and stimulus and response location on the contingent negative variation (CNV). *Biological Psychology, 1*(1), 1-19.

[26] Gaillard, A. W. K. (1978). Slow brain potentials preceding task performance.

[27] Guo, Z., Tan, X., Pan, Y., Liu, X., Zhao, G., Wang, L., Peng, Z. (2019). Contingent negative variation during a modified cueing task in simulated driving. *PLoS one*, 14(11), e0224966.

[28] Indicators, O. E. C. D. (2015). Health at a Glance 2011. *OECD Indicators, OECD Publishing*, Paris DOI: *https://doi org/101787/health_glance-2015-en Accessed February, 15*, 2016.

[29] Adrian ION-MĂRGINEANU et al, Machine learning for classifying abnormal brain tissue progression based on multi-parametric Magnetic Resonance data, KUL, October 2017.

[30] Torrey, H. C. (1956). Bloch equations with diffusion terms. *Physical review, 104*(3), 563.

[31] Kastler, B., Vetter, D. (2018). *Comprendre l'IRM: manuel d'auto-apprentissage*. Elsevier Health Sciences.

[32] Delinte, N., Macq, B., Dricot, L. " Quantifying the microstructural changes in the brain of patients with alcohol use disorder after abstinence via diffusion MRI."

[33] Medic, N., Kochunov, P., Ziauddeen, H., Ersche, K. D., Nathan, P. J., Ronan, L., Fletcher, P. C. (2019). BMI-related cortical morphometry changes are associated with altered white matter structure. *International Journal of Obesity, 43*(3), 523-532.

[34] Tawa, E. A., Hall, S. D., Lohoff, F. W. (2016). Overview of the genetics of alcohol use disorder. *Alcohol and alcoholism, 51*(5), 507-514.

[35] World Hearth Organisation, `https://www.who.int/news-room/fact-sheets/detail/alcohol`, [Accessed 30-04-2021].

[36] Kendler, K. S., Walters, E. E., Neale, M. C., Kessler, R. C., Heath, A. C., Eaves, L. J. (1995). The structure of the genetic and environmental risk factors for six major psychiatric disorders in women: Phobia, generalized anxiety disorder, panic disorder, bulimia, major depression, and alcoholism. *Archives of general psychiatry, 52*(5), 374-383.

[37] Kirshenbaum, A. P., Olsen, D. M., Bickel, W. K. (2009). A quantitative review of the ubiquitous relapse curve. *Journal of substance abuse treatment, 36*(1), 8-17.

[38] Bohn, M. J., Babor, T. F., Kranzler, H. R. (1995). The Alcohol Use Disorders Identification Test (AUDIT): validation of a screening instrument for use in medical settings. *Journal of studies on alcohol, 56*(4), 423-432.

[39] Rosenbloom, M. J., Pfefferbaum, A. (2008). Magnetic resonance imaging of the living brain: evidence for brain degeneration among alcoholics and recovery with abstinence. *Alcohol Research Health.*

[40] Cardenas, V. A., Studholme, C., Meyerhoff, D. J., Song, E., Weiner, M. W. (2005). Chronic active heavy drinking and family history of problem drinking modulate regional brain tissue volumes. *Psychiatry Research: Neuroimaging, 138*(2), 115-130.

[41] Brody, A. L., Mandelkern, M. A., Jarvik, M. E., Lee, G. S., Smith, E. C., Huang, J. C., ... London, E. D. (2004). Differences between smokers and nonsmokers in regional gray matter volumes and densities. Biological psychiatry, 55(1), 77-84.

[42] Pfefferbaum, A., Sullivan, E. V., Mathalon, D. H., Lim, K. O. (1997). Frontal lobe volume loss observed with magnetic resonance imaging in older chronic alcoholics. *Alcoholism: Clinical and Experimental Research, 21*(3), 521-529.

[43] Thompson, P. M., Jahanshad, N., Ching, C. R., Salminen, L. E., Thomopoulos, S. I., Bright, J., ... Sønderby, I. E. (2020). ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Translational psychiatry*, 10(1), 1-28.

[44] Chayer, C., Freedman, M. (2001). Frontal lobe functions. *Current neurology and neuroscience reports, 1*(6), 547-552.

[45] Wang, G. J., Volkow, N. D., Roque, C. T., Cestaro, V. L., Hitzemann, R. J., Cantos, E. L., ... Dhawan, A. P. (1993). Functional importance of ventricular enlargement and cortical atrophy in healthy subjects and alcoholics as assessed with PET, MR imaging, and neuropsychologic testing. *Radiology, 186*(1), 59-65.

[46] What is depression, Pscychiatry.org, `https://www.psychiatry.org/patients-families/depression/what-is-depression`, [Consulted on 29-03-2021].

[47] Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., Eaves, L. J. (1993). The lifetime history of major depression in women: reliability of diagnosis and heritability. *Archives of general psychiatry, 50*(11), 863-870.

[48] Mendels, J., Stern, S., Frazer, A. (1976). Biochemistry of depression. *Diseases of the nervous system.*

[49] Moldin, S. O., Reich, T., Rice, J. P. (1991). Current perspectives on the genetics of unipolar depression. *Behavior Genetics, 21*(3), 211-242.

[50] Peterson, C., Seligman, M. E. (1984). Causal explanations as a risk factor for depression: theory and evidence. *Psychological review, 91*(3), 347.

[51] Strine, T. W., Mokdad, A. H., Balluz, L. S., Gonzalez, O., Crider, R., Berry, J. T., Kroenke, K. (2008). Depression and anxiety in the United States: findings from the 2006 behavioral risk factor surveillance system. *Psychiatric services, 59*, 1383-1390.

[52] Haynes, J. C., Farrell, M., Singleton, N., Meltzer, H., Araya, R., Lewis, G., Wiles, N. J. (2005). Alcohol consumption as a risk factor for anxiety and depression: results from the longitudinal follow-up of the National Psychiatric Morbidity Survey. *The British Journal of Psychiatry, 187*(6), 544-551.

[53] McBride, A., Petersen, T. (Eds.). (2002). *Working with substance misusers: A guide to theory and practice.* Psychology Press.

[54] Sullivan, L. E., Fiellin, D. A., O'Connor, P. G. (2005). The prevalence and impact of alcohol problems in major depression: a systematic review. *The American journal of medicine, 118*(4), 330-341.

[55] Finn, P. R., Sharkansky, E. J., Viken, R., West, T. L., Sandy, J., Bufferd, G. M. (1997). Heterogeneity in the families of sons of alcoholics: The impact of familial vulnerability type on offspring characteristics. *Journal of Abnormal Psychology, 106*(1), 26.

[56] Beck, A. T., Steer, R. A., Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical psychology review, 8*(1), 77-100.

[57] Pollock, V. E., Schneider, L. S. (1990). Quantitative, waking EEG research on depression. *Biological Psychiatry, 27*(7), 757-780.

[58] Knott, V., Mahoney, C., Kennedy, S., Evans, K. (2001). EEG power, frequency, asymmetry and coherence in male depression. *Psychiatry Research: Neuroimaging, 106*(2), 123-140.

[59] Pizzagalli, D. A., Jahn, A. L., O'Shea, J. P. (2005). Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. *Biological psychiatry, 57*(4), 319-327.

[60] Hakami, T., Jones, N. C., Tolmacheva, E. A., Gaudias, J., Chaumont, J., Salzberg, M., ... Pinault, D. (2009). NMDA receptor hypofunction leads to generalized and persistent aberrant oscillations independent of hyperlocomotion and the state of consciousness. *PloS one, 4*(8), e6755.

[61] Isomura, S., Onitsuka, T., Tsuchimoto, R., Nakamura, I., Hirano, S., Oda, Y., ... Kanba, S. (2016). Differentiation between major depressive disorder and bipolar disorder by auditory steady-state responses. *Journal of affective disorders, 190*, 800-806.

[62] Mohammadi, M., Al-Azab, F., Raahemi, B., Richards, G., Jaworska, N., Smith, D., ... Knott, V. (2015). Data mining EEG signals in depression for their diagnostic value. *BMC medical informatics and decision making, 15*(1), 1-14.

[63] Mumtaz, W., Xia, L., Ali, S. S. A., Yasin, M. A. M., Hussain, M., Malik, A. S. (2017). Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). *Biomedical Signal Processing and Control, 31*, 108-115.

[64] Gangadhar, B. N., Ancy, J., Janakiranaiah, N., Umapathy, C. (1993). P300 amplitude in non-bipolar, melancholic depression. *Journal of affective disorders*, 28(1), 57-60.

[65] Vandoolaeghe, E., van Hunsel, F., Nuyten, D., Maes, M. (1998). Auditory event related potentials in major depression: prolonged P300 latency and increased P200 amplitude. *Journal of affective disorders, 48*(2-3), 105-113. 57-60.

[66] Mumtaz, W., Malik, A. S., Ali, S. S. A., Yasin, M. A. M. (2015, October). P300 intensities and latencies for major depressive disorder detection. In 2015 *IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (pp. 542-545). IEEE.

[67] Wang, C., Rapp, P., Darmon, D., Trongnetrpunya, A., Costanzo, M. E., Nathan, D. E., ... Keyser, D. (2018). Utility of P300 ERP in monitoring post-trauma mental health: a longitudinal study in military personnel returning from combat deployment. *Journal of psychiatric research, 101*, 5-13.

[68] Zhuo, C., Li, G., Lin, X., Jiang, D., Xu, Y., Tian, H., ... Song, X. (2019). The rise and fall of MRI studies in major depressive disorder. *Translational psychiatry, 9*(1), 1-14. 448-471.

[69] Botteron, K. N., Raichle, M. E., Drevets, W. C., Heath, A. C., Todd, R. D. (2002). Volumetric reduction in left subgenual prefrontal cortex in early onset depression. *Biological psychiatry, 51*(4), 342-344.

[70] Bremner, J. D., Vythilingam, M., Vermetten, E., Nazeer, A., Adil, J., Khan, S., ... Charney, D. S. (2002). Reduced volume of orbitofrontal cortex in major depression. *Biological psychiatry, 51*(4), 273-279.

[71] Videbech, P., Ravnkilde, B. (2004). Hippocampal volume and depression: a meta-analysis of MRI studies. *American Journal of Psychiatry, 161*(11), 1957-1966.

[72] Vakili, K., Pillay, S. S., Lafer, B., Fava, M., Renshaw, P. F., Bonello-Cintron, C. M., Yurgelun-Todd, D. A. (2000). Hippocampal volume in primary unipolar major depression: a magnetic resonance imaging study. *Biological psychiatry, 47*(12), 1087-1090.

[73] Jaworska, N., Yang, X. R., Knott, V., MacQueen, G. (2015). A review of fMRI studies during visual emotive processing in major depressive disorder. *The World Journal of Biological Psychiatry, 16*(7), 448-471.

[74] Alonso, J., Liu, Z., Evans-Lacko, S., Sadikova, E., Sampson, N., Chatterji, S., ... WHO World Mental Health Survey Collaborators. (2018). Treatment gap for anxiety disorders is global: Results of the World Mental Health Surveys in 21 countries. *Depression and anxiety, 35*(3), 195-208.

[75] Anxiety Disorders, *National Institute of Mental Health*, `https://www.nimh.nih.gov/health/topics/anxiety-disorders`, [Accessed 15-04-2021].

[76] Matsudaira, T., Kitamura, T. (2006). Personality traits as risk factors of depression and anxiety among Japanese students. *Journal of clinical psychology, 62*(1), 97-109.

[77] Spielberger, C. D. (1983). Manual for the State-Trait Anxiety Inventory; Palo Alto, CA, Ed.

[78] Emons, W. H., Habibović, M., Pedersen, S. S. (2019). Prevalence of anxiety in patients with an implantable cardioverter defibrillator: measurement equivalence of the HADS-A and the STAI-S. *Quality of Life Research, 28*(11), 3107-3116.

[79] Oathes, D. J., Ray, W. J., Yamasaki, A. S., Borkovec, T. D., Castonguay, L. G., Newman, M. G., Nitschke, J. (2008). Worry, generalized anxiety disorder, and emotion: Evidence from the EEG gamma band. *Biological psychology, 79*(2), 165-170.

[80] Gordeev, S. A. (2008). Clinical-psychophysiological studies of patients with panic attacks with and without agoraphobic disorders. *Neuroscience and behavioral physiology, 38*(6), 633-637.

[81] Wise, V., McFarlane, A. C., Clark, C. R., Battersby, M. (2011). An integrative assessment of brain and body function 'at rest'in panic disorder: a combined quantitative EEG/autonomic function study. *International Journal of Psychophysiology, 79*(2), 155-165.

[82] Chen, C., Yu, X., Belkacem, A. N., Lu, L., Li, P., Zhang, Z., ... Ming, D. (2021). EEG-Based Anxious States Classification Using Affective BCI-Based Closed Neurofeedback System. *Journal of Medical and Biological Engineering*, 1-10.

[83] Ferrari, M. C. F., Busatto, G. F., McGuire, P. K., Crippa, J. A. S. (2008). Structural magnetic ressonance imaging in anxiety disorders: an update of research findings. *Brazilian Journal of Psychiatry, 30*(3), 251-264.

[84] Bremner, J. D., Randall, P., Scott, T. M., Bronen, R. A., Seibyl, J. P., Southwick, S. M., ... Innis, R. B. (1995). MRI-based measurement of hippocampal volume in patients with combat-related posttraumatic stress disorder. *The American journal of psychiatry, 152*(7), 973.

[85] Bremner, J. D., Randall, P., Vermetten, E., Staib, L., Bronen, R. A., Mazure, C., ... Charney, D. S. (1997). Magnetic resonance imaging-based measurement of hippocampal volume in posttraumatic stress disorder related to childhood physical and sexual abuse—a preliminary report. *Biological psychiatry, 41*(1), 23-32..

[86] Li, L., Chen, S., Liu, J., Zhang, J., He, Z., Lin, X. (2006). Magnetic resonance imaging and magnetic resonance spectroscopy study of deficits in hippocampal structure in fire victims with recent-onset posttraumatic stress disorder. *The Canadian Journal of Psychiatry, 51*(7), 431-437.

[87] Smith, M. E. (2005). Bilateral hippocampal volume reduction in adults with post-traumatic stress disorder: A meta-analysis of structural MRI studies. *Hippocampus, 15*(6), 798-807.

[88] Villarreal, G., Hamilton, D. A., Graham, D. P., Driscoll, I., Qualls, C., Petropoulos, H., Brooks, W. M. (2004). Reduced area of the corpus callosum in posttraumatic stress disorder. *Psychiatry Research: Neuroimaging, 131*(3), 227-235.

[89] Choi, J. S., Kang, D. H., Kim, J. J., Ha, T. H., Lee, J. M., Youn, T., ... Kwon, J. S. (2004). Left anterior subregion of orbitofrontal cortex volume reduction and impaired organizational strategies in obsessive-compulsive disorder. *Journal of Psychiatric Research, 38*(2), 193-199.

[90] Kang, D. H., Kim, J. J., Choi, J. S., Kim, Y. I., Kim, C. W., Youn, T., ... Kwon, J. S. (2004). Volumetric investigation of the frontal-subcortical circuitry in patients with obsessive-compulsive disorder. *The Journal of neuropsychiatry and clinical neurosciences, 16*(3), 342-349.

[91] Atmaca, M., Yildirim, H., Ozdemir, H., Tezcan, E., Poyraz, A. K. (2007). Volumetric MRI study of key brain regions implicated in obsessive–compulsive disorder. *Progress in neuro-psychopharmacology and Biological Psychiatry, 31*(1), 46-52.

[92] Uchida, R. R., Del-Ben, C. M., Santos, A. C., Araujo, D., Crippa, J. A., Guimaraes, F. S., Graeff, F. G. (2003). Decreased left temporal lobe volume of panic patients measured by magnetic resonance imaging. *Brazilian Journal of Medical and Biological Research, 36*(7), 925-929.

[93] Massana, G., Serra-Grabulosa, J. M., Salgado-Pineda, P., Gastó, C., Junqué, C., Massana, J., Mercader, J. M. (2003). Parahippocampal gray matter density in panic disorder: a voxel-based morphometric study. *American Journal of Psychiatry, 160*(3), 566-568.

[94] Yoo, H. K., Kim, M. J., Kim, S. J., Sung, Y. H., Sim, M. E., Lee, Y. S., ... Lyoo, I. K. (2005). Putaminal gray matter volume decrease in panic disorder: an optimized voxel-based morphometry study. *European Journal of Neuroscience, 22*(8), 2089-2094.

[95] Madonna, D., Delvecchio, G., Soares, J. C., Brambilla, P. (2019). Structural and functional neuroimaging studies in generalized anxiety disorder: a systematic review. *Brazilian Journal of Psychiatry, 41*(4), 336-362.

[96] Etkin, A., Prater, K. E., Schatzberg, A. F., Menon, V., Greicius, M. D. (2009). Disrupted amygdalar subregion functional connectivity and evidence of a compensatory network in generalized anxiety disorder. *Archives of general psychiatry, 66*(12), 1361-1372.

[97] Karstoft, K. I., Statnikov, A., Andersen, S. B., Madsen, T., Galatzer-Levy, I. R. (2015). Early identification of posttraumatic stress following military deployment: Application of machine learning methods to a prospective study of Danish soldiers. *Journal of affective disorders, 184*, 170-175.

[98] Aliferis, C. F., Statnikov, A., Tsamardinos, I., Mani, S., Koutsoukos, X. D. (2010). Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: algorithms and empirical evaluation.*Journal of Machine Learning Research, 11*(1).

[99] Rolland, B., Dricot, L., Creupelandt, C., Maurage, P., De Timary, P. (2020). Respective influence of current alcohol consumption and duration of heavy drinking on brain morphological alterations in alcohol use disorder. *Addiction biology, 25*(2), e12751.

[100] Edition, F. (2013). Diagnostic and statistical manual of mental disorders. Am Psychiatric Assoc, 21.

[101] Delorme A Makeig S (2004) EEGLAB: an open-source toolbox for analysis of single-trial EEG dynamics, *Journal of Neuroscience Methods 134*,19-21.

[102] Qian, X., Xu, Y. P., Li, X. (2005). A CMOS continuous-time low-pass notch filter for EEG systems. *Analog Integrated Circuits and Signal Processing, 44*(3), 231-238.

[103] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience, 7*(267):1–13, 2013.

[104] Electroencephalography and Clinical Neurophysiology Volume 95, Issue 3, September 1995, Pages 178-188,

[105] Arnauddelorme, ICA for Dummies, `http://arnauddelorme.com/ica_for_dummies/`, [Accessed 22/03/2021].

[106] Hyvarinen, A., Oja, E. (2000). Neural Networks Research Centre. Helsinki University of Technology,*"Independent Component Analysis: Algorithms and Applications" Neural Networks, 13*, 411-430.

[107] Lopez-Calderon, J., Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience, 8*, 213.

[108] Fischl, B. (2012). FreeSurfer. *Neuroimage, 62*(2), 774-781.

[109] Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., Peters, T. M. (1993, October). 3D statistical neuroanatomical models from 305 MRI volumes. In *1993 IEEE conference record nuclear science symposium and medical imaging conference (pp. 1813-1817)*. IEEE.

[110] Collins, D. L., Neelin, P., Peters, T. M., Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of computer assisted tomography, 18*(2), 192-205.

[111] Destrieux, C., Fischl, B., Dale, A., Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature.

[112] Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage, 31*(3), 968-980.

[113] Maldjian, J. A., Laurienti, P. J., Kraft, R. A., Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage, 19*(3), 1233-1239.

[114] Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics, 15*(2), 70-73.

[115] Heinzel, G., Rüdiger, A., Schilling, R. (2002). Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new at-top windows.

[116] Barry, R. J., Clarke, A. R., Johnstone, S. J., Magee, C. A., Rushby, J. A. (2007). EEG differences between eyes-closed and eyes-open resting conditions. Clinical neurophysiology, 118(12), 2765-2773.

[117] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods, 17*(3), 261-272.

[118] ANT Neuro, `https://www.ant-neuro.com/`, [Accessed 28-03-2021].

[119] Github, PyPDF, `https://github.com/mstamy2/PyPDF2`, [Accessed 30-04-2021].

[120] Github, Doc2txt, `https://github.com/ankushshah89/python-docx2txt`, [Accessed 30-04-2021].

[121] Gu, T., Fu, C., Shen, Z., Guo, H., Zou, M., Chen, M., ... Song, X. (2019). Age-related whole-brain structural changes in relation to cardiovascular risks across the adult age spectrum. *Frontiers in aging neuroscience, 11*, 85.

[122] Brickman, A. M., Habeck, C., Zarahn, E., Flynn, J., Stern, Y. (2007). Structural MRI covariance patterns associated with normal aging and neuropsychological functioning. *Neurobiology of aging, 28*(2), 284-295.

[123] Vapnik, V., Guyon, I., Hastie, T. (1995). Support vector machines. *Mach. Learn, 20*(3), 273-297.

[124] García-Gonzalo, E., Fernández-Muñiz, Z., García Nieto, P. J., Bernardo Sánchez, A., Menéndez Fernández, M. (2016). Hard-rock stability analysis for span design in entry-type excavations with learning classifiers.*Materials, 9*(7), 531.

[125] Drish, J. (2001). Obtaining calibrated probability estimates from support vector machines. *Technique Report, Department of Computer Science and Engineering, University of California, San Diego*, CA.

[126] Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., Klein, M. (2002). *Logistic regression*. New York: Springer-Verlag.

[127] Xanthopoulos, P., Pardalos, P. M., Trafalis, T. B. (2013). Linear discriminant analysis. In *Robust data mining* (pp. 27-33). Springer, New York, NY.

[128] Peck, R., Van Ness, J. (1982). The use of shrinkage estimators in linear discriminant analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* (5), 530-537.

[129] Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

[130] Linkedin,Random Forest Algorithm, An Interactive Discussion,https://www.linkedin.com/pulse/random-forest-algorithm-interactive-discussion-niraj-kumar/, June 2016. [Consulted on 02-05-2021]

[131] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

[132] Strobl, C., Boulesteix, A. L., Zeileis, A., Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics, 8*(1), 1-21.

[133] Benesty, J., Chen, J., Huang, Y., Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing* (pp. 1-4). Springer, Berlin, Heidelberg.

[134] Bengio, Y., Delalleau, O., Le Roux, N. (2005). The curse of dimensionality for local kernel machines. *Techn. Rep, 1258*, 12.

[135] Webb, G. I. (2010). Overfitting.

[136] Roffo, G. (2016). Feature selection library (MATLAB toolbox). *arXiv preprint arXiv*:1607.01327.

[137] Hall, M. A., Smith, L. A. (1999, May). Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper. In FLAIRS conference (Vol. 1999, pp. 235-239).

[138] Shi, S., Nathoo, F. (2018, December). Feature Learning and Classification in Neuroimaging: Predicting Cognitive Impairment from Magnetic Resonance Imaging. In 2018 4th International Conference on *Big Data and Information Analytics* (BigDIA) (pp. 1-5). IEEE.

[139] Jolliffe, I. T. (2003). Principal component analysis. *Technometrics, 45*(3), 276.

[140] Wan, C. (2019). Feature selection paradigms. In *Hierarchical Feature Selection for Knowledge Discovery* (pp. 17-23). Springer, Cham.

[141] Khaire, U. M., Dhanalakshmi, R. (2019). Stability of feature selection algorithm: A review. *Journal of King Saud University-Computer and Information Sciences*.

[142] Sklearn, Grid search cross validation, `https://scikit-learn.org/stable/modules/cross_validation.html`, [Consulted 02-05-2021].

[143] Vehtari, A., Gelman, A., Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing, 27*(5), 1413-1432.

[144] Provost, F. (2000, July). Machine learning from imbalanced data sets 101. In *Proceedings of the AAAI'2000 workshop on imbalanced data sets* (Vol. 68, No. 2000, pp. 1-3). AAAI Press.

[145] Probst, P., Boulesteix, A. L. (2017). To Tune or Not to Tune the Number of Trees in Random Forest. *J. Mach. Learn. Res., 18*(1), 6673-6690.

[146] Towardsdatascience, Confusion Matrix-Explained, `https://towardsdatascience.com/confusion-matrix-explained-34e4be19b3ec`, [Consulted 15-04-2021].

[147] Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

[148] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement, 20*(1), 37-46.

[149] McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica, 22*(3), 276-282.

[150] Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers, 10*(3), 61-74.

[151] Setiono, R., Liu, H. (1997). Neural-network feature selector. *IEEE transactions on neural networks, 8*(3), 654-662.

[152] Verikas, A., Bacauskiene, M. (2002). Feature selection with neural networks. *Pattern recognition letters, 23*(11), 1323-1335.

[153] Cai, J., Luo, J., Wang, S., Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing, 300*, 70-79.

[154] Santelmann, H., Franklin, J., Bußhoff, J., Baethge, C. (2015). Test–retest reliability of schizoaffective disorder compared with schizophrenia, bipolar disorder, and unipolar depression-a systematic review and meta-analysis. *Bipolar disorders, 17*(7), 753-768.

[155] Wardenaar, K. J., de Jonge, P. (2013). Diagnostic heterogeneity in psychiatry: towards an empirical solution. *BMC medicine, 11*(1), 1-3.

[156] Jablensky, A. (2016). Psychiatric classifications: validity and utility. *World Psychiatry, 15*(1), 26-31.

[157] Courtney, K. E., Ghahremani, D. G., London, E. D., Ray, L. A. (2014). The association between cue-reactivity in the precuneus and level of dependence on nicotine and alcohol. *Drug and alcohol dependence, 141*, 21-26.

[158] Guggenmos, M., Schmack, K., Veer, I. M., Lett, T., Sekutowicz, M., Sebold, M., ... Sterzer, P. (2020). A multimodal neuroimaging classifier for alcohol dependence. *Scientific reports, 10*(1), 1-12.

[159] Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature, 536*(7615), 171-178.

[160] Hahn, S., Mackey, S., Cousijn, J., Foxe, J. J., Heinz, A., Hester, R., ... Garavan, H. (2020). Predicting alcohol dependence from multi-site brain structural measures. *Human Brain Mapping*. .

[161] Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., Herrera, F. (2018). *Learning from imbalanced data sets* (Vol. 11). Berlin: Springer.

[162] Hilbert, K., Lueken, U., Muehlhan, M., Beesdo-Baum, K. (2017). Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: A multimodal machine learning study. *Brain and behavior, 7*(3), e00633.

[163] van Minde, D., Klaming, L., Weda, H. (2014). Pinpointing moments of high anxiety during an MRI examination. *International journal of behavioral medicine, 21*(3), 487-495.

[164] Saitz, R. (1998). Introduction to alcohol withdrawal. *Alcohol health and research world, 22*(1), 5.

[165] Kabacoff, R. I., Segal, D. L., Hersen, M., Van Hasselt, V. B. (1997). Psychometric properties and diagnostic utility of the Beck Anxiety Inventory and the State-Trait Anxiety Inventory with older adult psychiatric outpatients. *Journal of anxiety disorders, 11*(1), 33-47.

[166] Arbabshirani, M. R., Plis, S., Sui, J., Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage, 145*, 137-165.

[167] Hahn, T., Marquand, A. F., Ehlis, A. C., Dresler, T., Kittel-Schneider, S., Jarczok, T. A., ... Fallgatter, A. J. (2011). Integrating neurobiological markers of depression. *Archives of general psychiatry, 68*(4), 361-368.

[168] Wei, M., Qin, J., Yan, R., Li, H., Yao, Z., Lu, Q. (2013). Identifying major depressive disorder using Hurst exponent of resting-state brain networks. Psychiatry Research: *Neuroimaging, 214*(3), 306-312.

[169] Jing, B., Long, Z., Liu, H., Yan, H., Dong, J., Mo, X., ... Li, H. (2017). Identifying current and remitted major depressive disorder with the Hurst exponent: a comparative study on two automated anatomical labeling atlases. *Oncotarget, 8*(52), 90452.

[170] Foland-Ross, L. C., Sacchet, M. D., Prasad, G., Gilbert, B., Thompson, P. M., Gotlib, I. H. (2015). Cortical thickness predicts the first onset of major depression in adolescence. *International Journal of Developmental Neuroscience, 46*, 125-131.

[171] Sacchet, M. D., Livermore, E. E., Iglesias, J. E., Glover, G. H., Gotlib, I. H. (2015). Subcortical volumes differentiate major depressive disorder, bipolar disorder, and remitted major depressive disorder. *Journal of psychiatric research, 68*, 91-98.

[172] Qiu, L., Huang, X., Zhang, J., Wang, Y., Kuang, W., Li, J., ... Gong, Q. (2014). Characterization of major depressive disorder using a multiparametric classification approach based on high resolution structural images. *Journal of psychiatry neuroscience: JPN, 39*(2), 78.

[173] Sankar, A., Zhang, T., Gaonkar, B., Doshi, J., Erus, G., Costafreda, S. G., ... Fu, C. H. (2016). Diagnostic potential of structural neuroimaging for depression from a multi-ethnic community sample. *BJPsych open, 2*(4), 247-254.

[174] Johnston, B. A., Steele, J. D., Tolomeo, S., Christmas, D., Matthews, K. (2015). Structural MRI-based predictions in patients with treatment-refractory depression (TRD). *PloS one, 10*(7).

[175] Liu, F., Guo, W., Yu, D., Gao, Q., Gao, K., Xue, Z., ... Chen, H. (2012). Classification of different therapeutic responses of major depressive disorder with multivariate pattern analysis method based on structural MR scans. *PloS one, 7*(7).

[176] Schuckit, M. A., Irwin, M. R. (1995). Alcoholism and affective disorder: clinical course of depressive symptoms. *Am J Psychiatry, 152*(1), 45-52.

[177] Enoch, M. A., Rohrbaugh, J. W., Davis, E. Z., Harris, C. R., Ellingson, R. J., Andreason, P., ... Goldman, D. (1995). Relationship of genetically transmitted alpha EEG traits to anxiety disorders and alcoholism. *American journal of medical genetics, 60*(5), 400-408.

[178] Mathersul, D., Williams, L. M., Hopkinson, P. J., Kemp, A. H. (2008). Investigating models of affect: relationships among EEG alpha asymmetry, depression, and anxiety. *Emotion, 8*(4), 560.

[179] Roelofs, S. M., Dikkenberg, G. M. (1987). Hyperventilation and anxiety: alcohol withdrawal symptoms decreasing with prolonged abstinence. *Alcohol, 4*(3), 215-220.

[180] Drummond, S. P., Gillin, J. C., Smith, T. L., DeModena, A. (1998). The sleep of abstinent pure primary alcoholic patients: natural course and relationship to relapse. *Alcoholism: Clinical and Experimental Research, 22*(8), 1796-1802. *Translational psychiatry, 6*(4), e788-e788.

[181] Nubukpo, P., Girard, M., Sengelen, J. M., Bonnefond, S., Varnoux, A., Marin, B., Malauzat, D. (2016). A prospective hospital study of alcohol use disorders, comorbid psychiatric conditions and withdrawal prognosis. *Annals of general psychiatry, 15*(1), 1-11.

[182] Huertas, D., Bautista, S., Sanjoaquin, A., Chamorro, L., Gilaberte, I. (1999). Subsyndromal depressive semiology in severe alcoholism. *Actas espanolas de psiquiatria*, 27(4), 223-227.

[183] Maddock, R. J., Garrett, A. S., Buonocore, M. H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human brain mapping, 18*(1), 30-41.

[184] Leech, R., Sharp, D. J. (2014). The role of the posterior cingulate cortex in cognition and disease. *Brain, 137*(1), 12-32.

[185] Nielsen, F. Å., Balslev, D., Hansen, L. K. (2005). Mining the posterior cingulate: segregation between memory and pain components. *Neuroimage, 27*(3), 520-532.

[186] Mohring V., Delzenne N., Leclercq S., Impact du microbiote intestinal sur l'appétence à l'alcool, UCL, 2020.

[187] Schulz, M. A., Yeo, B. T., Vogelstein, J. T., Mourao-Miranada, J., Kather, J. N., Kording, K., ... Bzdok, D. (2020). Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. *Nature communications, 11*(1), 1-15.

[188] Sundheim, B. M. (1992). *Overview of the fourth message understanding evaluation and conference.* NAVAL COMMAND CONTROL AND OCEAN SURVEILLANCE CENTER RDT AND E DIV SAN DIEGO CA.

[189] Leclercq, S., Matamoros, S., Cani, P. D., Neyrinck, A. M., Jamar, F., Stärkel, P., ... Delzenne, N. M. (2014). Intestinal permeability, gut-bacterial dysbiosis, and behavioral markers of alcohol-dependence severity. *Proceedings of the National Academy of Sciences, 111*(42), E4485-E4493.

[190] Leclercq, S., Stärkel, P., Delzenne, N. M., de Timary, P. (2019). The gut microbiota: A new target in the management of alcohol dependence?. *Alcohol, 74*, 105-111.

[191] Beleites, C., Neugebauer, U., Bocklitz, T., Krafft, C., Popp, J. (2013). Sample size planning for classification models. *Analytica chimica acta, 760*, 25-33.

[192] Weissteiner, J. (2018). *Variable importance measures in regression and classification methods.* (Master Thesis). Vienna University of Economics and Business, Vienna.

[193] Croux, C., Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical methods applications, 19*(4), 497-515.

Appendix A

Atlases

This section covers the different regions that were used for the brain atlases we used for this thesis.

A.1 Destrieux Atlas

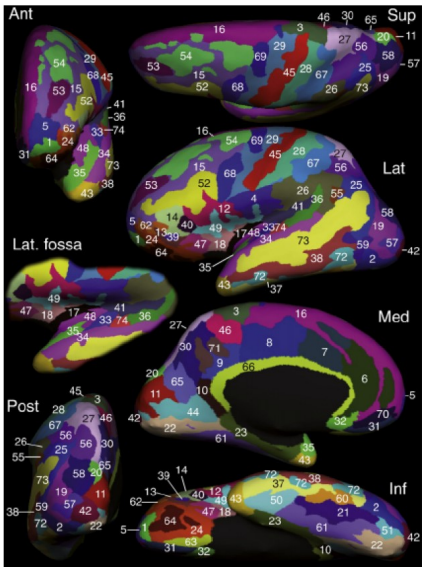

Figure A.1: Destrieux atlas [111]

The labels of the Desikan atlas are labelled below [112]:

1. G_and_S_frontomargin Fronto-marginal gyrus (of Wernicke) and sulcus
2. G_and_S_occipital_inf Inferior occipital gyrus (O3) and sulcus
3. G_and_S_paracentral Paracentral lobule and sulcus
4. G_and_S_subcentral Subcentral gyrus (central operculum) and sulci
5. G_and_S_transv_frontopol Transverse frontopolar gyri and sulci
6. G_and_S_cingul-Ant Anterior part of the cingulate gyrus and sulcus (ACC)
7. G_and_S_cingul-Mid-Ant Middle-anterior part of the cingulate gyrus and sulcus (aMCC)
8. G_and_S_cingul-Mid-Post Middle-posterior part of the cingulate gyrus and sulcus (pMCC)
9. G_cingul-Post-dorsal Posterior-dorsal part of the cingulate gyrus (dPCC)

10. G_cingul-Post-ventral Posterior-ventral part of the cingulate gyrus (vPCC, isthmus of the cingulate gyrus)
11. G_cuneus Cuneus
12. G_front_inf-Opercular Opercular part of the inferior frontal gyrus
13. G_front_inf-Orbital Orbital part of the inferior frontal gyrus
14. G_front_inf-Triangul Triangular part of the inferior frontal gyrus
15. G_front_middle Middle frontal gyrus (F2)
16. G_front_sup Superior frontal gyrus (F1)
17. G_Ins_lg_and_S_cent_ins Long insular gyrus and central sulcus of the insula
18. G_insular_short Short insular gyri
19. G_occipital_middle Middle occipital gyrus (O2, lateral occipital gyrus)
20. G_occipital_sup Superior occipital gyrus (O1)
21. G_oc-temp_lat-fusifor Lateral occipito-temporal gyrus (fusiform gyrus, O4-T4)
22. G_oc-temp_med-Lingual Lingual gyrus, ligual part of the medial occipito-temporal gyrus, (O5)
23. G_oc-temp_med-Parahip Parahippocampal gyrus, parahippocampal part of the medial occipito-temporal gyrus, (T5)
24. G_orbital Orbital gyri
25. G_pariet_inf-Angular Angular gyrus
26. G_pariet_inf-Supramar Supramarginal gyrus
27. G_parietal_sup Superior parietal lobule (lateral part of P1)
28. G_postcentral Postcentral gyrus
29. G_precentral Precentral gyrus
30. G_precuneus Precuneus (medial part of P1)
31. G_rectus Straight gyrus, Gyrus rectus
32. G_subcallosal Subcallosal area, subcallosal gyrus
33. G_temp_sup-G_T_transv Anterior transverse temporal gyrus (of Heschl)
34. G_temp_sup-Lateral Lateral aspect of the superior temporal gyrus
35. G_temp_sup-Plan_polar Planum polare of the superior temporal gyrus
36. G_temp_sup-Plan_tempo Planum temporale or temporal plane of the superior temporal gyrus
37. G_temporal_inf Inferior temporal gyrus (T3)
38. G_temporal_middle Middle temporal gyrus (T2)
39. Lat_Fis-ant-Horizont Horizontal ramus of the anterior segment of the lateral sulcus (or fissure
40. Lat_Fis-ant-Vertical Vertical ramus of the anterior segment of the lateral sulcus (or fissure)
41. Lat_Fis-post Posterior ramus (or segment) of the lateral sulcus (or fissure)
42. Pole_occipital Occipital pole

43. Pole_temporal Temporal pole
44. S_calcarine Calcarine sulcus
45. S_central Central sulcus (Rolando's fissure)
46. S_cingul-Marginalis Marginal branch (or part) of the cingulate sulcus
47. S_circular_insula_ant Anterior segment of the circular sulcus of the insula
48. S_circular_insula_inf Inferior segment of the circular sulcus of the insula
49. S_circular_insula_sup Superior segment of the circular sulcus of the insula
50. S_collat_transv_ant Anterior transverse collateral sulcus
51. S_collat_transv_post Posterior transverse collateral sulcus
52. S_front_inf Inferior frontal sulcus
53. S_front_middle Middle frontal sulcus
54. S_front_sup Superior frontal sulcus
55. S_interm_prim-Jensen Sulcus intermedius primus (of Jensen)
56. S_intrapariet_and_P_trans Intraparietal sulcus (interparietal sulcus) and transverse parietal sulci
57. S_oc_middle_and_Lunatus Middle occipital sulcus and lunatus sulcus
58. S_oc_sup_and_transversal Superior occipital sulcus and transverse occipital sulcus
59. S_occipital_ant Anterior occipital sulcus and preoccipital notch (temporo-occipital incisure)
60. S_oc-temp_lat Lateral occipito-temporal sulcus
61. S_oc-temp_med_and_Lingual Medial occipito-temporal sulcus (collateral sulcus) and lingual sulcus
62. S_orbital_lateral Lateral orbital sulcus
63. S_orbital_med-olfact Medial orbital sulcus (olfactory sulcus) _orbital-H_Shaped Orbital sulci (H-shaped sulci)
64. S_parieto_occipital Parieto-occipital sulcus (or fissure)
65. S_pericallosal Pericallosal sulcus (S of corpus callosum)
66. S_postcentral Postcentral sulcus
67. S_precentral-inf-part Inferior part of the precentral sulcus
68. S_precentral-sup-part Superior part of the precentral
69. S_suborbital Suborbital sulcus (sulcus rostrales, supraorbital sulcus)
70. S_subparietal Subparietal sulcus
71. S_temporal_inf Inferior temporal sulcus
72. S_temporal_sup Superior temporal sulcus (parallel sulcus
73. S_temporal_transverse Transverse temporal sulcus

A.2 Desikan Atlas

The different structures retrieved in the Desikan atlas are given below[112]:

- Banks superior temporal sulcus
- Caudal anterior-cingulate cortex
- Caudal middle frontal gyrus
- Cuneus cortex
- Entorhinal cortex
- Fusiform gyrus
- Inferior parietal cortex
- Inferior temporal gyrus
- Isthmus– cingulate cortex
- Lateral occipital cortex
- Lateral orbital frontal cortex
- Lingual gyrus
- Medial orbital frontal cortex
- Middle temporal gyrus
- Parahippocampal gyrus
- Paracentral lobule
- Pars opercularis
- Pars orbitalis
- Pars triangularis
- Pericalcarine cortex
- Postcentral gyrus
- Posterior-cingulate cortex
- Precentral gyrus
- Precuneus cortex
- Rostral anterior cingulate cortex
- Rostral middle frontal gyrus
- Superior frontal gyrus
- Superior parietal cortex
- Superior temporal gyrus
- Supramarginal gyrus
- Temporal pole
- Transverse temporal cortex

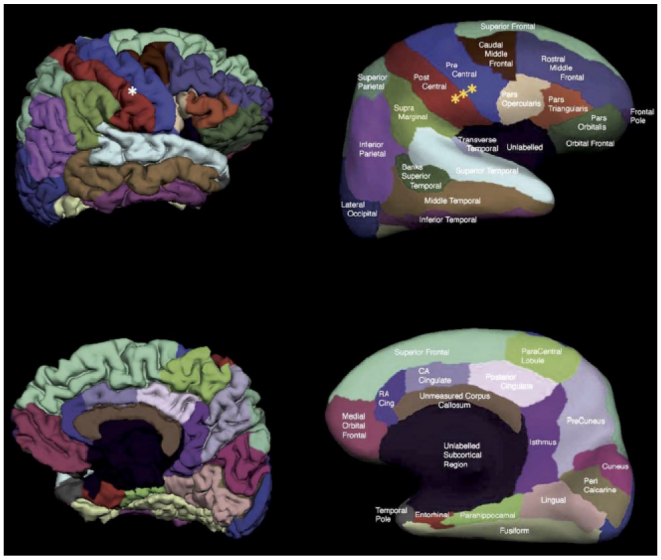

Figure A.2: Desikan atlas [112]

Appendix B

Feature selections

This appendix will try to fill in the gaps in the feature selection sub-section. We will look in more detail at embedded methods that have not been used during this thesis, but that could at first sight be a good idea in the future, although the computation time limits its use when the input feature space is large. We are also interested in the embedded methods available for random forest.

B.1 Wrapper methods

Wrapper methods follow a greedy search approach. The idea is to evaluate a large number of different variable combinations as input variables to the machine learning algorithm and to choose the set that performs best on the validation set. When the number of features is high, these algorithms require a very high computation time and are not very efficient. There exists 3 main wrapper methods [140]:

- Forward sequential selection
- Backward Sequential selection
- Hill Climbing

To illustrate the different methods, let's take a simple example : predict whether a person is an alcoholic based on the following characteristics: age, gender, weekly alcohol consumption and BECK score. The only constraint of this problem is that we can simply use a random forest algorithm and two input variables. Let's imagine that for this purpose we have access to a population of 100 individuals, 50 controls and 50 AUD patients. Let's imagine that we use 20 individuals to test our performances and 80 for train, we then divide our train set in a training set and a validation set see figure .

In order to select our two most important features, we start by using the sequential feature selection algorithm. The goal of this one is to start from an empty set of important features to train our machine learning algorithm on the training set. The algorithm is iterative and at each iteration we will add a feature in our set of important features based on the performance of our algorithm with this feature in the set of important variables. For the example we presented above, the algorithm will work as follows

- On the first iteration: it will try to take each variable (age, gender, alcohol consumption per week and BECK score) one by one and simply predict from it if the person is alcoholic or not. Let's imagine that the accuracy on the validation set using as input variable separately the age, the sex, the weekly alcohol consumption and the BECK score is respectively 50%, 52%, 92%, 70%. As the performance are the best with the variable alcohol consumption per week, this variable is kept in the set of important variables.
- At the second iteration: We look at which variable will help the most the classification scores on the validation set if we add it to the variable alcohol consumption per week. We thus now try to predict alcoholism thanks two variables on the validation set. The accuracy of our machine learning algorithm with the 3 different sets (age and alcohol consumption, sex and alcohol consumption, alcohol and alcohol consumption) is compared. The set obtaining the best accuracy is kept.

The backward elimination is similar to what we have just seen except that instead of starting from an empty set of important variables, we start from a set with all the variables and at each iteration instead of adding a variable in the set of important variables, we remove the one that when removed from the set deters the least the results or improves them the most [140].

Hill climbing is a bit different, the beginning is similar to froward selection, the first iteration is exactly the same. Then a second variable is added and a backward elimination iteration, one of the two is removed if the performance is not better, is applied to see if the two variables go "well" together. The algorithm iterates until it finds a set of features of size k [140].

B.2 Feature importance for random forest [192]

In order to understand embedded random forest methods, it is necessary to understand how the random forest algorithm works. As explained previously, the random forest algorithm generates a set of decision trees in a random way (Bootstrap aggregation) so that each input of each decision tree of the forest is different from the others and as independent as possible. This is also due to the fact that only $\sqrt{m}$ features are taken as input for each decision tree.

As the training data is sampled with replacement, we can use the remaining data, on average 37% of the data to calculate the out-of-bag prediction error of each tree as the empirical risk on the remaining, unseen training data. The final out-of-Bag is given by the average over all these values.

The generation of a decision tree consists in choosing nodes and decisions. To make it simple, let us consider a decision tree which for each split, takes into account only one feature[1]. At each node of each decision tree, the data is split in two different subsets R_1 and R_2 according to a splitting point s^*, such that

$$R_1(j,s) := \{(X,Y)|X_j \leq s\} \in N_{node} \tag{B.1}$$

and

$$R_2(j,s) := \{(X,Y)|X_j \geq s\} \in N_{node}. \tag{B.2}$$

For a binary classification tree, we define the proportion level of class l (equal to 0 or 1) as

$$p_{kl} := \frac{1}{|R_k(j,s)|} \sum_{y_i \in R_k(j,s)} 1_{y_i=l} \tag{B.3}$$

To find the splitting point at node k, we use an algorithm minimising the weighted Gini impurity. The Gini impurity is a measurement of the likelihood of the incorrect classification of a new instance of a random variable. The Gini impurity of a class is defined at node k by

$$GI_k(j,s) := \sum_{l=0}^{1} p_{kl}(1-p_{kl}) \tag{B.4}$$

The Gini impurity of R_k will be the largest possible if and only if there are as many elements belonging to one class as to the other in R_k, which means that during the training phase, the split obtaining R_k was useless because it does not allow to separate the data at all. We hope to find during our training phase, the smallest possible Gini impurities and thus very discriminating splits. j and s are then found at node k by solving the minimisation problem :

$$\min_{j,s} \sum_{k=1}^{2} GI_k(j,s) \frac{|R_k(j,s)|}{\sum_{l=1}^{2} |R_l(j,s)|} \tag{B.5}$$

[1]This problem can easily be adapted to ours where each split is done by taking into account a set of features of size log(features)

The Gini importance measure is born naturally from the construction of each tree. To study the importance of variable X_j, the improvement in the splitting criterion for splits made by this variable X_j could be accumulated across all trees. This is easily implemented in the existing code, the only thing to be adjusted is to store the ΔG_j during training and then take the k elements for which the accumulated Gini impurities is the smallest. This method works well in practice and is very convenient as it doesn't require much calculation, but results are not always 100% accurate as it looks at the *in-bag* samples, inverse of *out-bag* samples, described above and thus sometimes introduces some bias.

A radically different mechanism can be to try to remedy these flaws. *Permutation measures* employ the *out-bag* samples which remain after the selection of the bagged samples. The prediction accuracy is first evaluated on this unmodified validation data. The values for the variable X_j are then shuffled within the *out-bag* dataset. For instance, the cortical thickness of the temporal lobe and the precuneus cortex are replaced by each other. The impact of this modification on the prediction accuracy then describes the importance of the variable for the classification model. It has not been implemented because for such a large number of input variables, this kind of method does not work very well because if we want to be able to swap all the variables, the calculation time would be too long.

Appendix C

Machine learning results : extension

This appendix covers the obtained results which were not conclusive for our study.

C.1 All-sMRI strategy

C.1.1 Problem n° 2 : Detection of anxiety

This figure shows the Cohen's Kappa and accuracy obtained with the different machine learning methods with the all-sMRI strategy to detect anxiety.

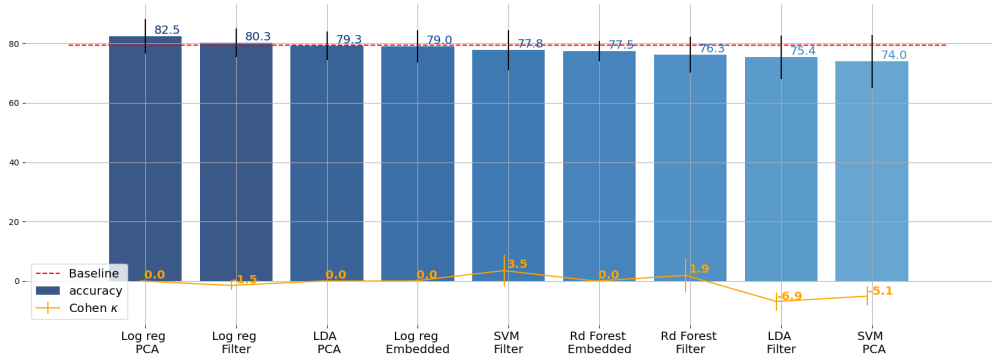

Figure C.1: Accuracy and Cohen's κ using all-sMRI strategy in order to predict anxiety.

C.1.2 Problem n°4 : Detection of anxiety for AUD patients

This figure shows the Cohen's Kappa and accuracy obtained with the different machine learning methods with the all-sMRI strategy to detect anxiety for AUD patients.

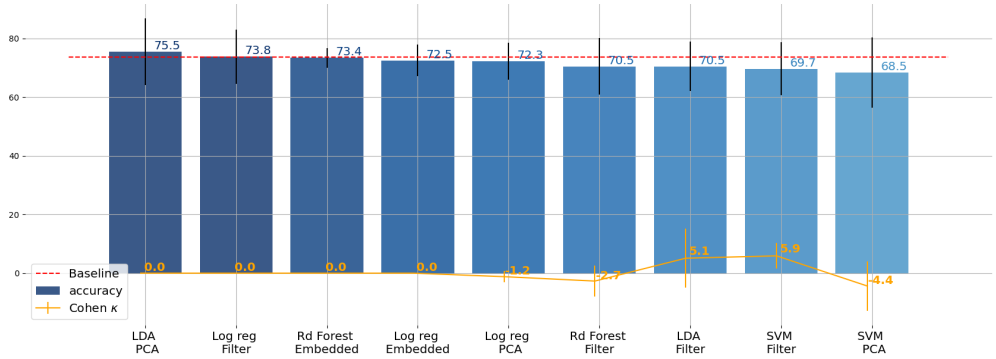

Figure C.2: Accuracy and Cohen's κ using all-sMRI strategy in order to predict anxiety for AUD patients.

C.2 Separate-sMRI strategy

C.2.1 Problem n°2 : Detection of anxiety

This figure shows the Cohen's Kappa and accuracy obtained with SVM and random forest methods with the separated-sMRI strategy to detect anxiety.

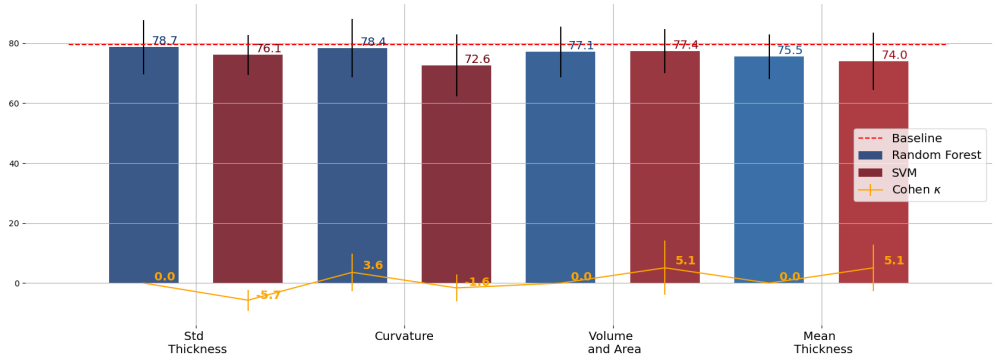

Figure C.3: Accuracy and Cohen's κ using separate-sMRI strategy in order to predict anxiety.

C.2.2 Problem n°4 : Detection of anxiety for AUD patients

This figure shows the Cohen's Kappa and accuracy obtained with the different machine learning methods with the separated-sMRI strategy to detect anxiety for AUD patients.

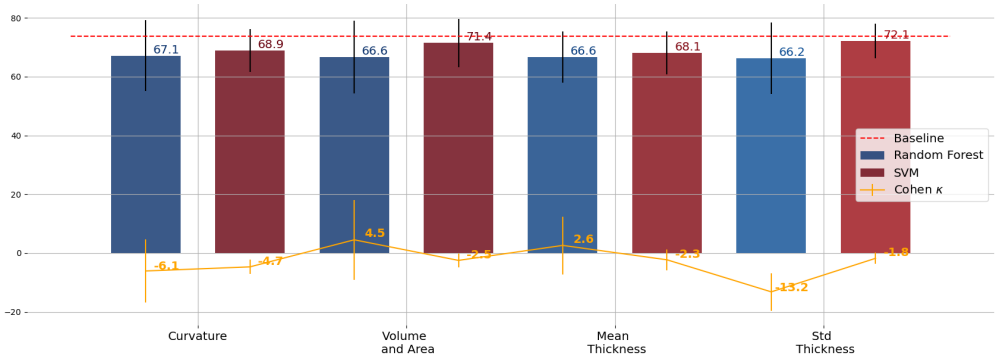

Figure C.4: Accuracy and Cohen's κ using separate-sMRI strategy in order to predict anxiety for AUD patients.

C.3 EEG and ERP strategy

C.3.1 Problem n°4 : Detection of anxiety for AUD patients

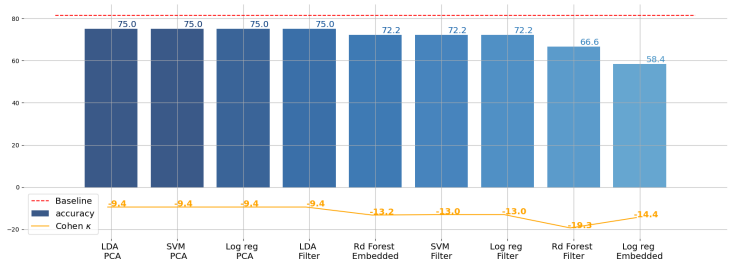

Figure C.5: Accuracy and Cohen's κ obtained with ERP data in order to predict anxiety for AUD patients.

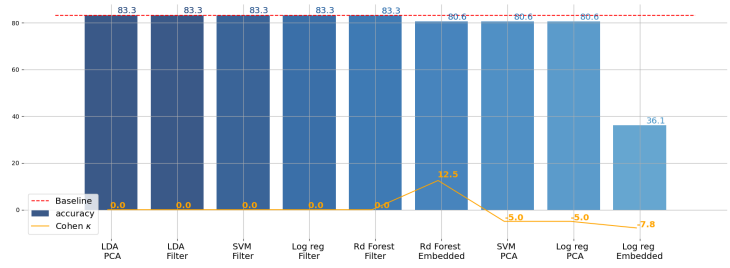

Figure C.6: Accuracy and Cohen's κ obtained with resting EEG data in order to predict anxiety for AUD patients.

Appendix D

Most important features

This appendix summarises the most important variables that have been highlighted from our feature importance methods for the MRI structural dataset. Specifically, we will first look at the variables that are statistically most different between groups (alcoholics vs. controls, anxious vs. non-anxious, etc). For each classification problem studied in this thesis, we will highlight the 5 most statically different variables concerning mean curvature, volume, mean cortical thickness and standard deviation of cortical thickness of brain structures. For the latter, we will indicate the significance level of the variables (α value). The data will then be plotted in the space of the most important variables to see if a separation is impossible. Finally, we also look at the most important variables highlighted with the Gini importance for all MRI structural variables this time.

Some abbreviations have been used in this appendix for the different features:

- lh : left hemisphere
- rg : right hemisphere
- cm_ : mean curvature
- cg_ : gaussian curvature
- v_ : volume
- s_ : surface
- wm_ : white matter area
- std_ : standard deviation of cortical thickness
- ct_ : mean cortical thickness.

D.1 Problem n°1 : Detection of alcoholism

The following figures show the most statically different features to differentiate between alcoholic and control subjects.

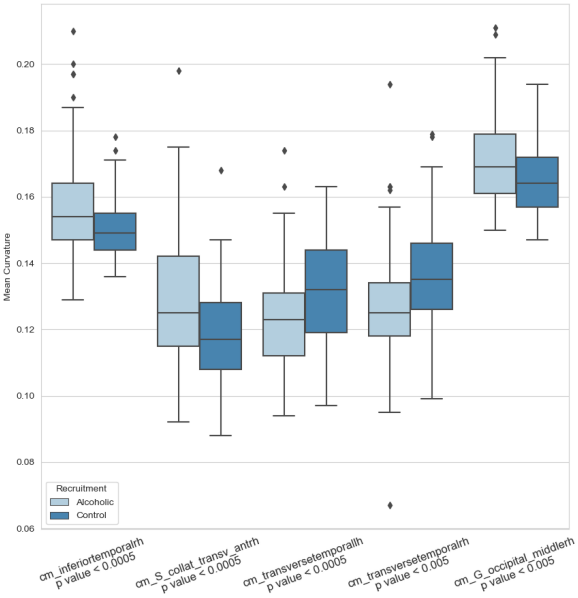

Figure D.1: 5 most statically significant Gaussian Curvature features comparing AUD and controls.

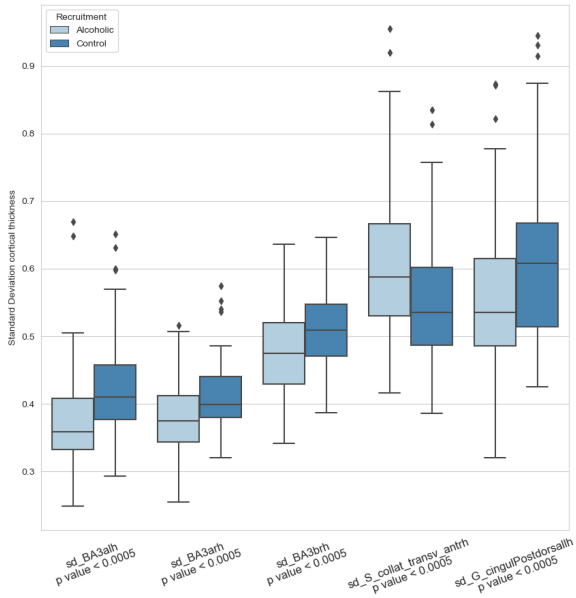

Figure D.2: 5 most statically significant cortical thickness standard deviation features comparing AUD and controls.

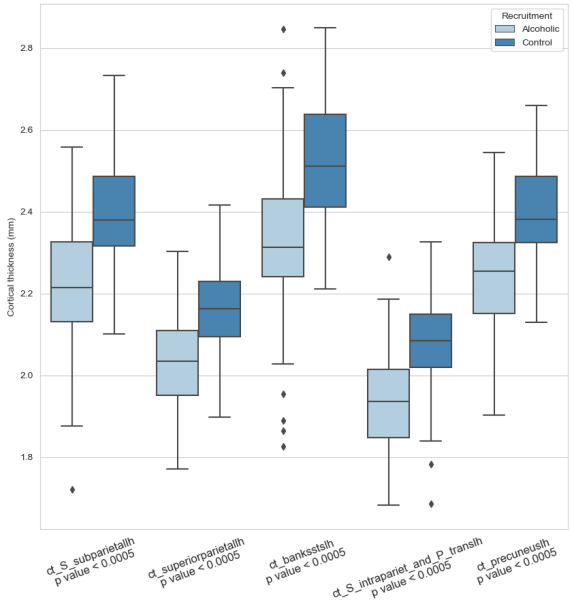

Figure D.3: 5 most statically significant mean cortical thickness features comparing AUD and controls.

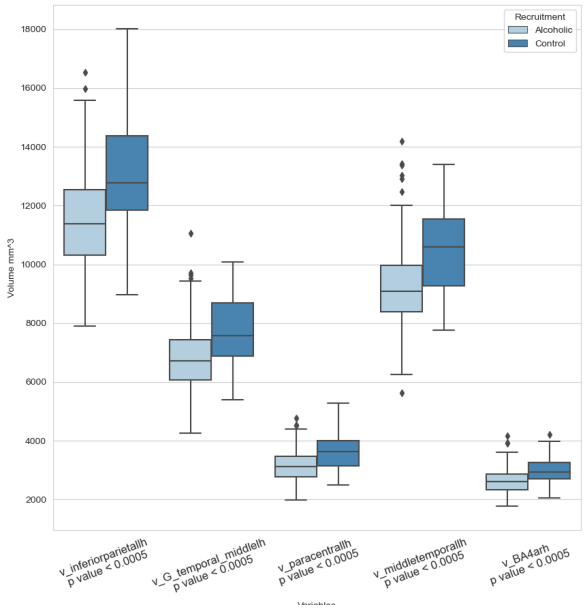

Figure D.4: 5 most statically significant volume features comparing AUD and controls.

As you can see, we have variables that are statistically different for all modalities, but the variables that are most different between groups are the mean cortical thickness variables.

Finally, the following graph shows how the data behaves in the most static feature plane. we can see that in this feature space (mean cortical thickness) the data are very separable.

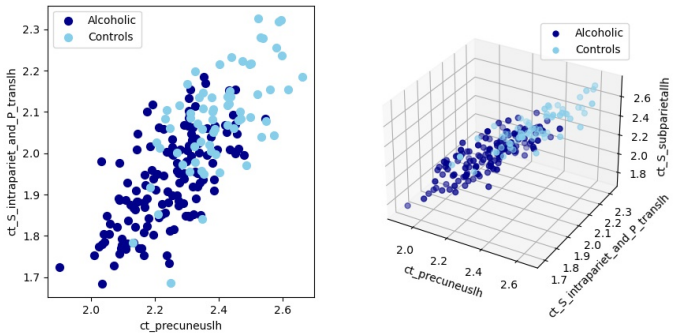

Figure D.5: Group of AUD and controls in the plan of most important features (p-value obtained with Welch method).

The following figure shows most important features rendered by Gini impurity:

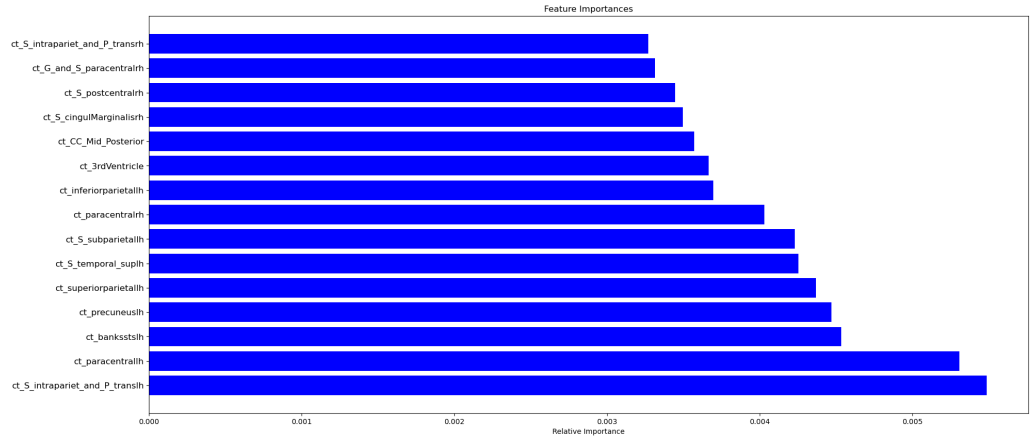

Figure D.6: Most important features obtained with random forest algorithm based on Gini impurity for AUD and controls.

The most important features are mostly mean cortical thickness variables, which is very much in line with what we saw with the filter (statistic) methods.

D.2 Problem n°2 : Detection of Anxiety

The following figures show the most statically different features to differentiate between anxious and non anxious subjects.

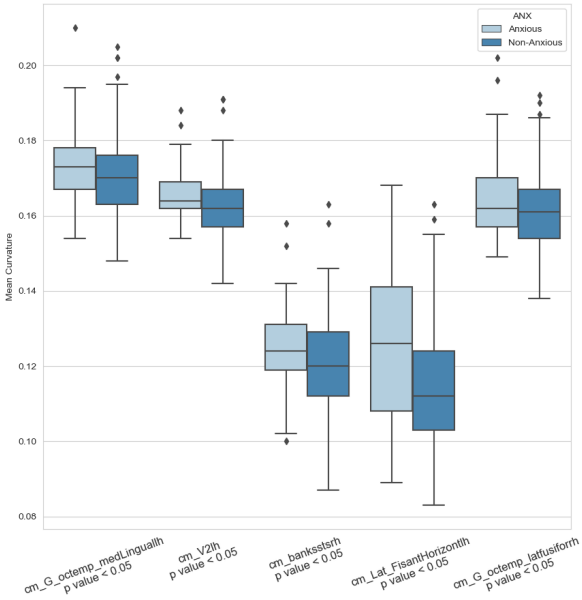

Figure D.7: 5 most significant Gaussian Curvature features comparing anxious and non-anxious.

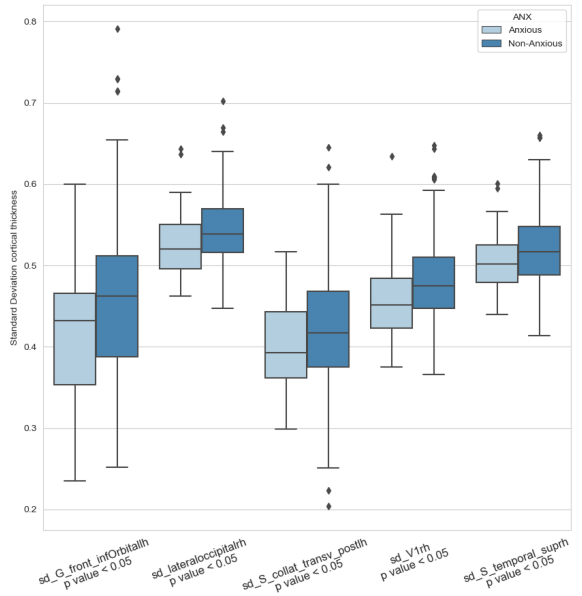

Figure D.8: 5 most significant cortical thickness standard deviation features comparing anxious and controls.

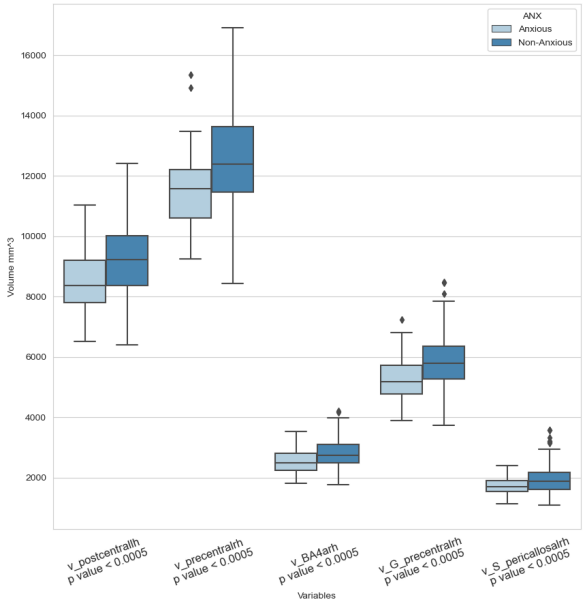

Figure D.9: 5 most significant volume features comparing anxious and non-anxious.

It can be seen that the variables that are most significantly different are the volume variables. However, we can see from the boxplot that the variables are not very separable. In fact, in the design of the most significantly different variables, one cannot observe two clusters.

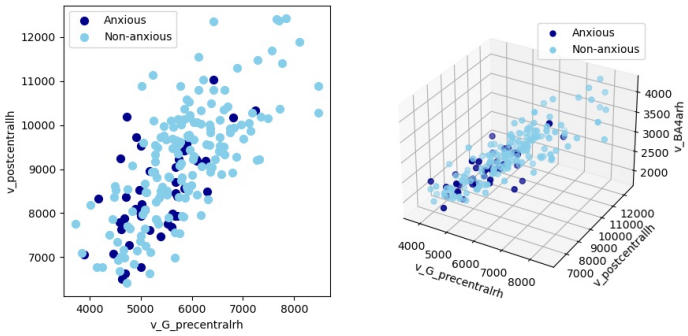

Figure D.10: Group of anxious and non-anxious in the plan of most important features (p-value obtained with Welch method).

The following figure shows most important features rendered by the random forest algorithm, which are not necessarily volume variables. There is no real type of data that allows differentiation between the two groups.

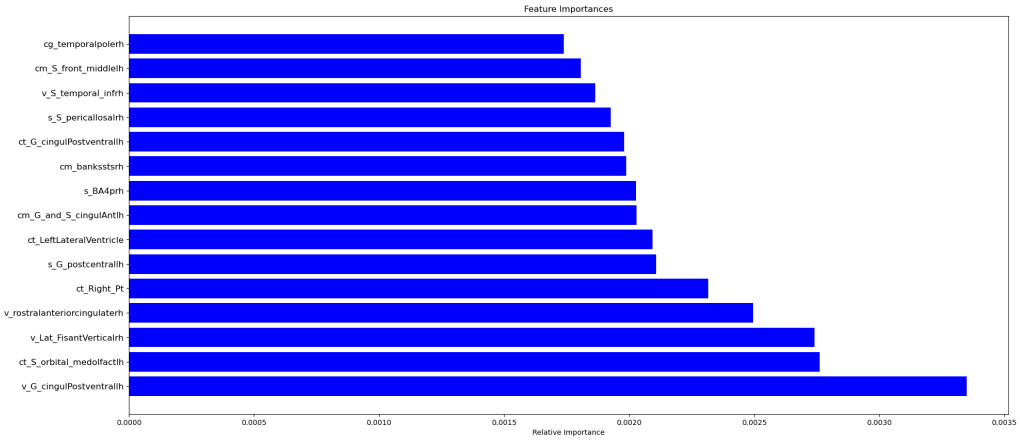

Figure D.11: Most important features obtained with random forest algorithm based on Gini impurity for Anxious and non-anxious.

D.3 Problem n°3 : Detection of depression

The first four graphs show the most statistically significant different features between the depressed and non-depressed subjects.

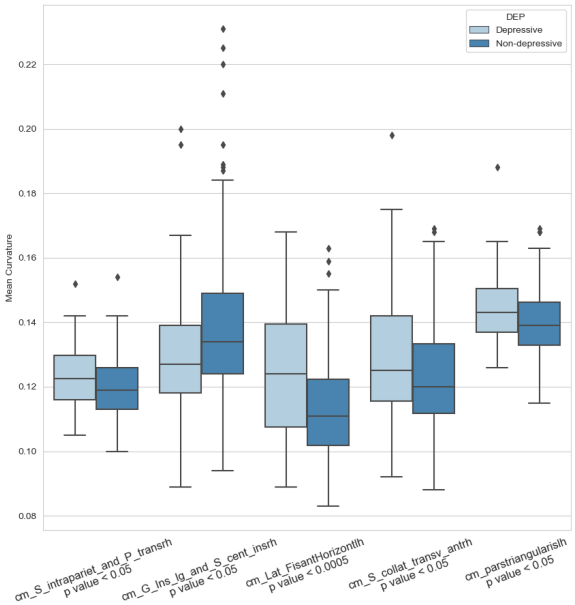

Figure D.12: 5 most statistically significant Gaussian Curvature features comparing depressed and non-depressed.

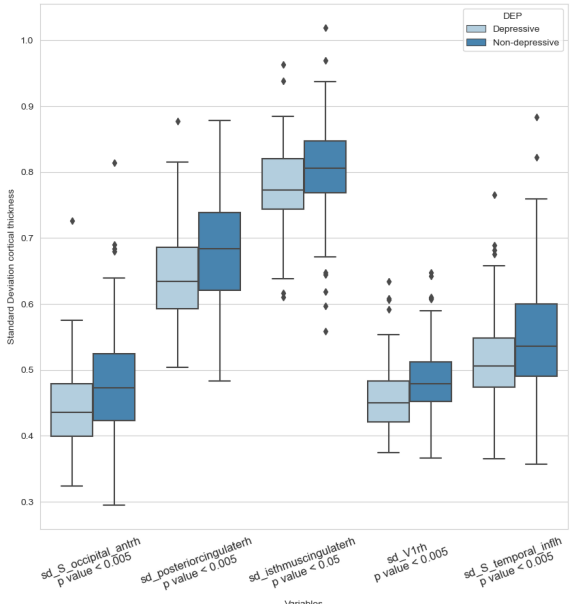

Figure D.13: 5 most significant statistically cortical thickness standard deviation features comparing depressed and non-depressed.

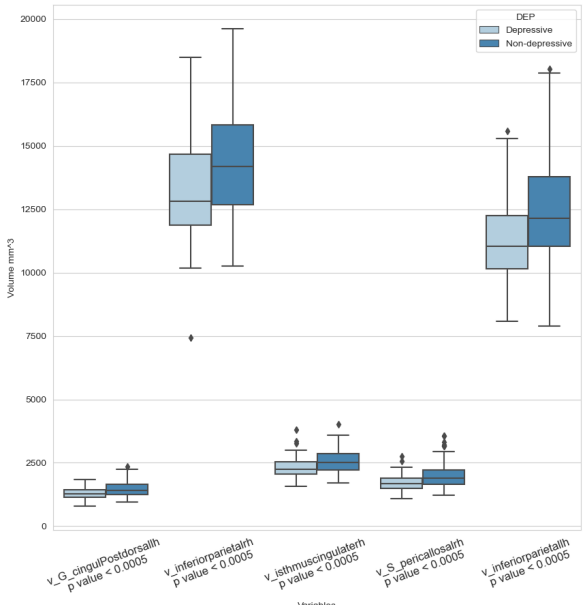

Figure D.14: 5 most statistically significant volume features comparing depressed and non-depressed.

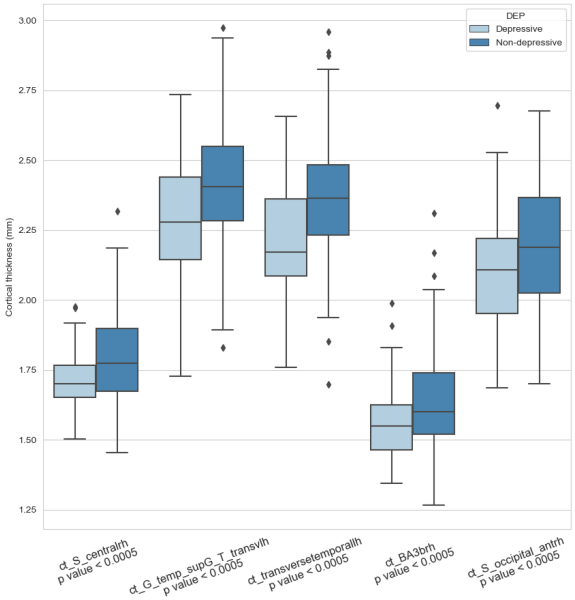

Figure D.15: 5 most statistically significant mean cortical thickness features comparing depressed and non-depressed.

It can be seen from the previous boxplots that the most significant variables for this problem are the cortical volume and thickness variables, this may be related to the fact that most of the patients considered as depressed are also alcoholics in our dataset. This can also be seen in the plot of the most significantly different variables, namely the cortical thickness of the transverse temporal cortex, the volume of the pericallosal sulcus and the volume of the inferior parietal cortex. This space of variables does not allow the creation of real clusters.

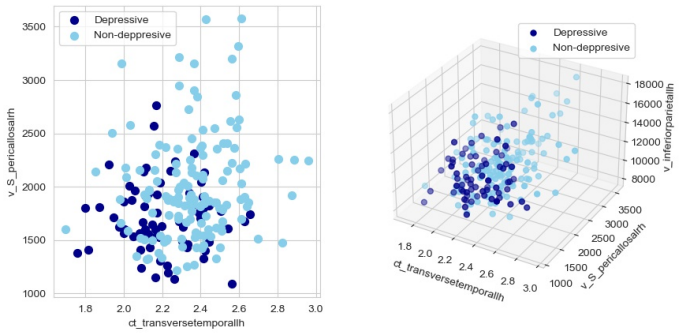

Figure D.16: Group of depressed and non- depressed in the plan of most important features (p-value obtained with Welch method).

The following figure shows most important features rendered by the random forest algorithm:

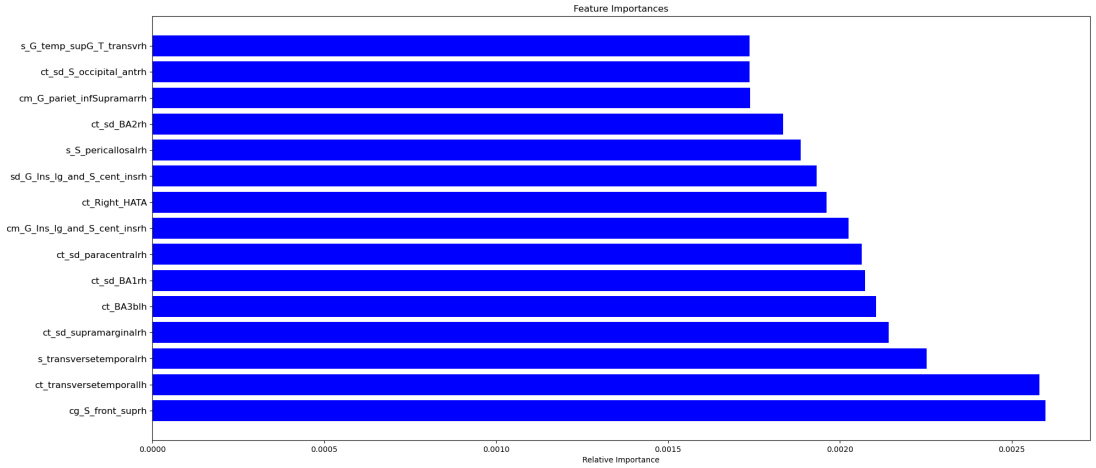

Figure D.17: Most important features obtained with random forest algorithm based on Gini impurity for the detection of depression from MRI features.

This method does not work very well and does not highlight one type of data.

D.4 Problem n°4 : Detection of anxiety for AUD patients

The following figures show the most statically different features to differentiate between anxious alcoholic and alcoholic subjects.

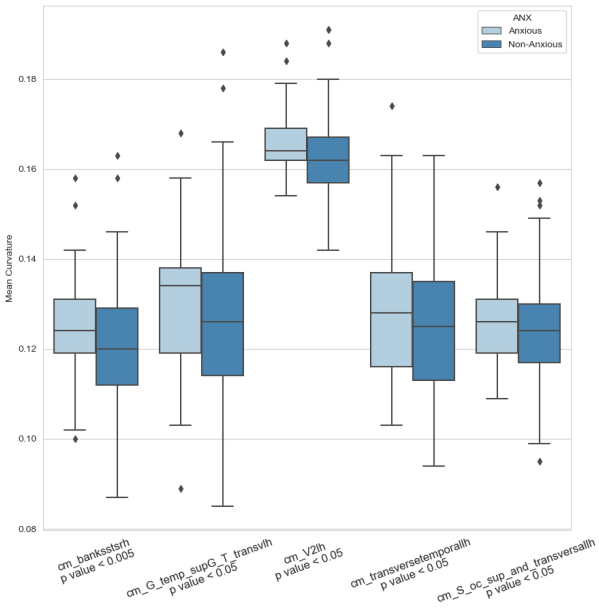

Figure D.18: 5 most statistically significant Gaussian Curvature features comparing anxious AUD and non-anxious AUD.

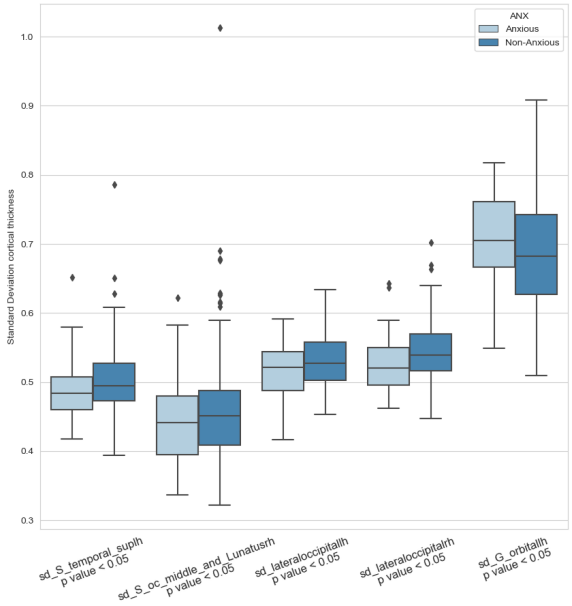

Figure D.19: 5 most statistically significant Standard Deviation features comparing anxious AUD and non-anxious AUD.

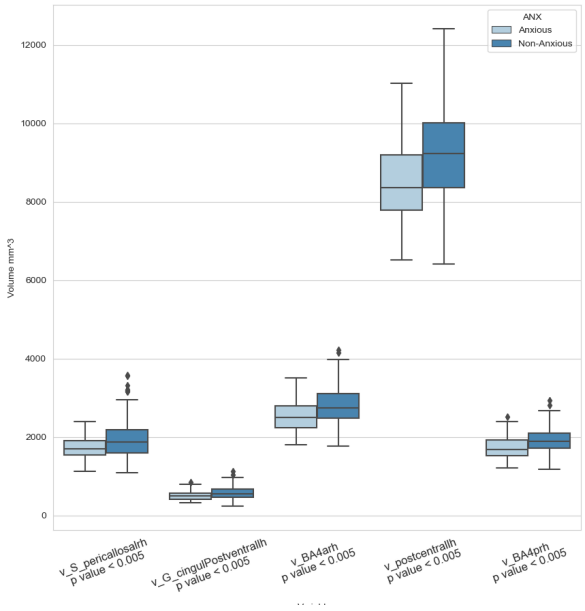

Figure D.20: 5 most statistically significant volume features comparing Anxious AUD and non-anxious AUD.

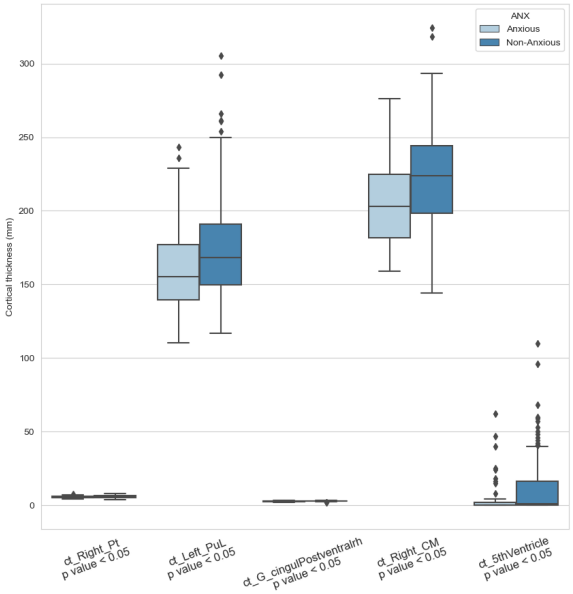

Figure D.21: 5 most statistically significant mean cortical thickness features comparing anxious AUD and non-anxious AUD.

There are no real variables which are significatively different from one group to the other except eh volume ones. the following graph shows how the data behaves in the features plan, which does not create any cluster.

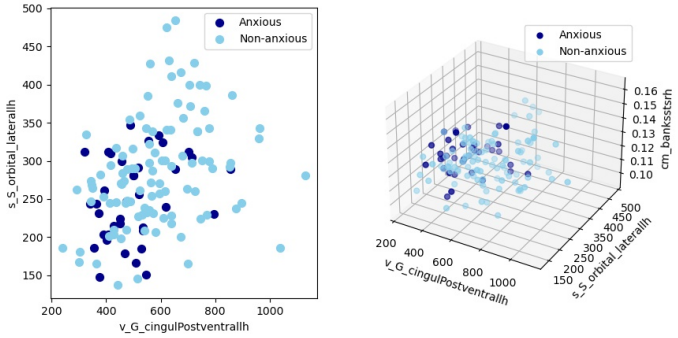

Figure D.22: Group of anxious AUD and non-anxious AUD in the plan of most important features (p-value obtained with Welch method).

The following figure shows most important features rendered by Gini impurity:

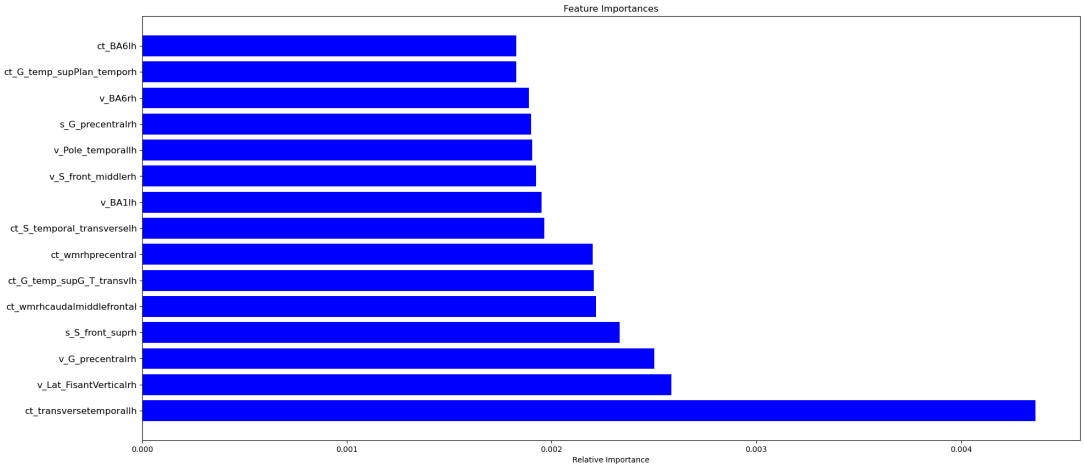

Figure D.23: Most important features obtained with random forest algorithm based on Gini impurity for anxious AUD and non-anxious AUD.

D.5 Problem n°5 : Detection of depression for AUD patients

The following figures show the most statically different features to differentiate between depressed alcoholic and alcoholic subjects.

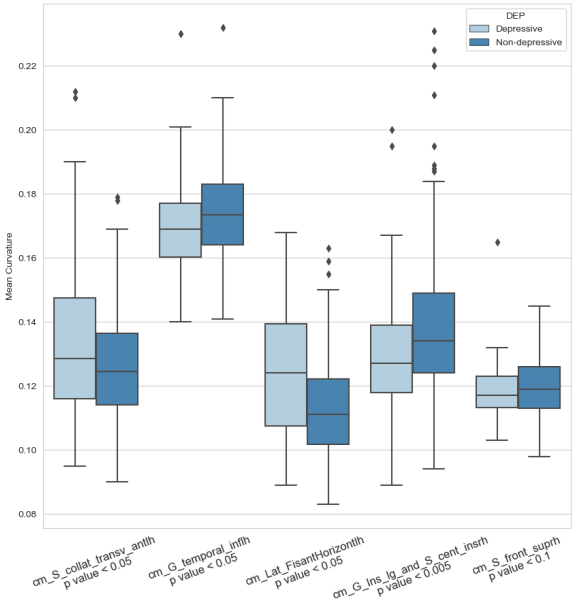

Figure D.24: 5 most statistically significant Gaussian Curvature features comparing depressed AUD and non-depressed AUD.

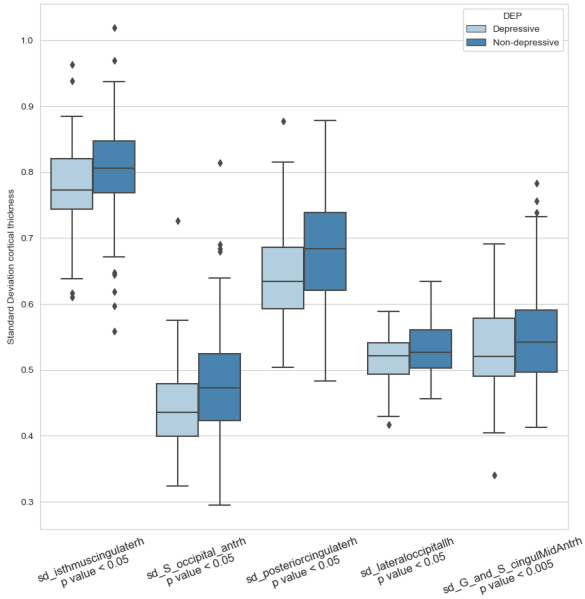

Figure D.25: 5 most statistically significant cortical thickness Standard Deviation features comparing depressed AUD and non-depressed AUD.

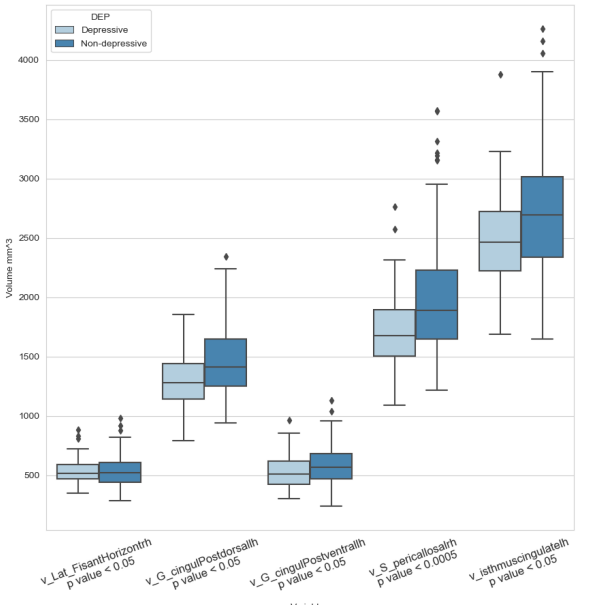

Figure D.26: 5 most statistically significant volume features comparing depressed AUD and non-depressed AUD.

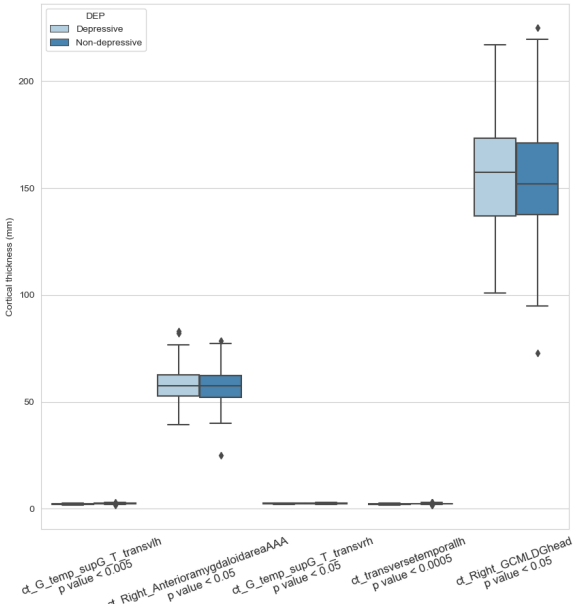

Figure D.27: 5 most statistically significant mean cortical thickness features comparing depressed AUD and non-depressed AUD.

The most impacted features are the mean cortical thickness and the volumes. the following graph shows how the data behaves in the features plan, which creates a semblance of a cluster. However, using these variables to perform our machine learning algorithms directly would be considered overfitting and cheating. The most important variables should be found from the training set and not from the whole set.

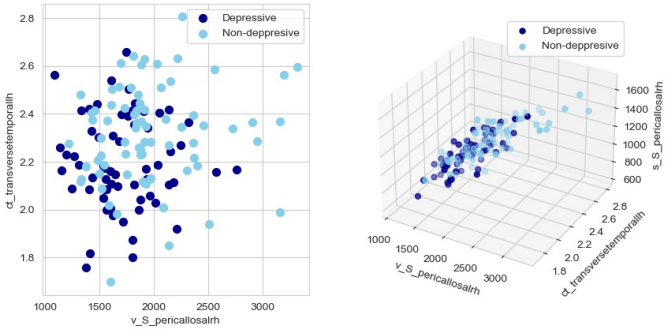

Figure D.28: Group of depressed AUD and non-depressed AUD in the plan of most important features (p-value obtained with Welch method).

The following figure shows most important features rendered by the random forest algorithm:

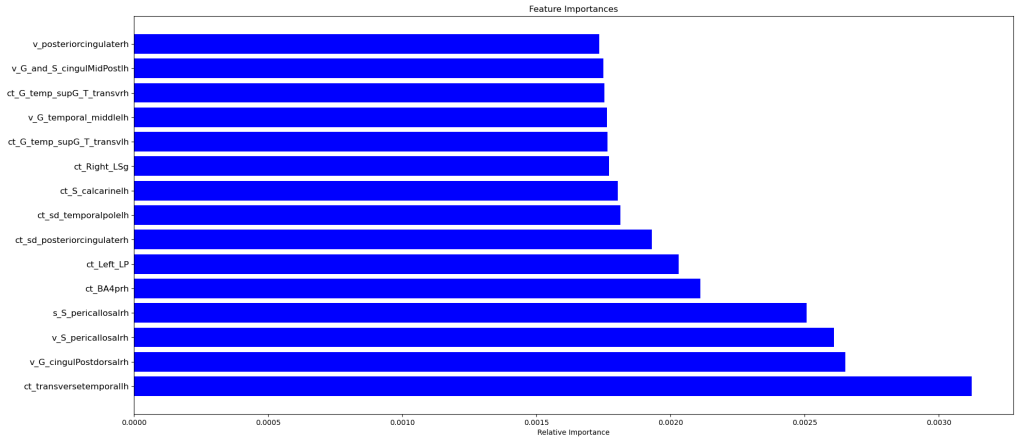

Figure D.29: Most important features obtained with random forest algorithm based on Gini impurity for depressed AUD and non-depressed AUD.

Appendix E

Correlation between Resting state EEG and microbiota

The vast majority of this master thesis has focused on the relationship between EEG, MRI and psychiatric illnesses. In this appendix we will focus on the link between EEG data and the gut microbiota of alcoholic patients. It has been argued that the microbiota may be an incriminating factor in alcoholism [189, 190]. New approaches are considering the use of probiotic or prebiotic methods to treat some forms of alcoholism. However, extracting the elements of gut permeability is very slow. If we could predict the intestinal permeability of alcoholic patients on the basis of EEG examinations (easy to perform and inexpensive), we could have an index of the intestinal micriobiota of alcoholic patients and potentially adapt their treatment, while saving time and money.

E.1 Data

The initial aim of this work was toto predict from the EEG data the intestinal permeability of patients with alcohol dependence. For this purpose, we analysed a dataset of 19 AUD patients for whom gut microbiota data and resting state EEG signals had been collected at St-Luc Hospital. The amount of data was too small to perform prediction methods based on machine learning algorithms. This is why limited ourselves to a statistical analysis of these data. The intestinal permeability of patients was measured by ingestion of 51Cr-EDTA, a radioactive molecule that is very weakly resorbed by a normal a normal intestine, thus a low permeability intestine. A part of this molecule is absorbed by the intestines and then excreted by urinary tract. The intestinal permeability was measured by urine and corresponded to the % 51Cr-EDTA compared to the ingested dose normalized for one gram of creatinine. If this measurement was below 2, the permeability was considered a normal, synonymous with healthy patients. Above, the intestinal permeability was considered as high and would be associated with a higher appetite (craving) for alcohol [189] .The summary of the data is shown in next table.

Metric	High permeability	Low permeability	P-value or Chi-squared P value
% 51Cr-EDTA / g creat	1.26 ± 0.3	3.23 ± 1.19	0.0004
Sexe	3 W - 6M	3W - 7M	0.876
Age	50.4 ± 11.5	46.6 ± 9.0	0.425
Alcohol consumption (g/day)	186.1 ± 137.7	177.7 ± 93.5	0.877

Table E.1: Summary of clinical population for permeability experiment.

E.2 Methods and results

The preprocessing of the EEG data was similar to the one we performed for the resting state data in the methods section. The signals were filtered on *eeglab*. The noisy segments were deleted on *eeglab*. ICA was applied via *mne* and the Welch power spectral densities of the different bands of the EEG electrodes were

calculated via *mne*. The power spectral densities were calculated in the *Delta* (1-4HZ), *Theta* (4-8Hz), *Alpha* (8-12Hz), *Beta* (13-30 Hz) and *Gamma* (>30Hz) bands of the EEG electrodes. The PSDs were calculated for eyes open and eyes closed separetely.

Groups differences

To perform our statistical analyses, we used the python package *scipy* to perform a Welch t-test in order to compare the mean values of the PSDs of our two subgroups. We observed higher PSDs for patients with a higher permeability. The main differences between the high permeability and the low permeability population were found in the *Delta* band for EC. The most affected electrodes, the one with p-values smaller than 0.05 were the A1, A2, C3, O2, Oz, T4, T5 and T6, see figure E.1.

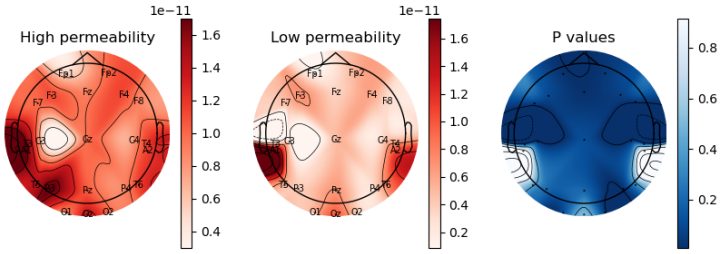

Figure E.1: The left and central figures highlight *Delta* PSD for patients with a low and high permeability. The last figure shows where those changes are significant thanks to p-values for EC.

For the EO, the only significant difference between the high permeability and the low permeability population was found in the T4 *Delta* band, see figure E.2.

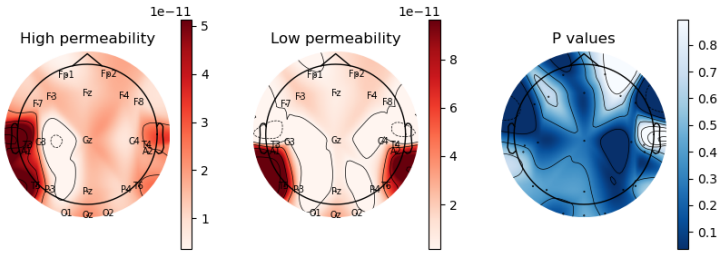

Figure E.2: The left and central figures highlight *Delta* PSD for patients with a low and high permeability. The last figure shows where those changes are significant thanks to p-values for EO.

Simple regression models

There were no other significant differences between the two groups for the other bands. Pearson [133] and Spearman [193] correlations did not show strong correlations between the PSDs of the EEG bands and permeability. Nevertheless, by some transformation of the feature space (logarithmic transformation), it was possible to see some correlations between the measurements. On figure E.3, we can see on the left figure, the correlation between the logarithm permeability and the logarithm of the PSD of the A2 delta wave for EC (p-value = 0.0009). On the right figure, the correlation between the logarithm permeability and the logarithm of PSD of T6 delta wave for EC (p-value=0.04) is shown.

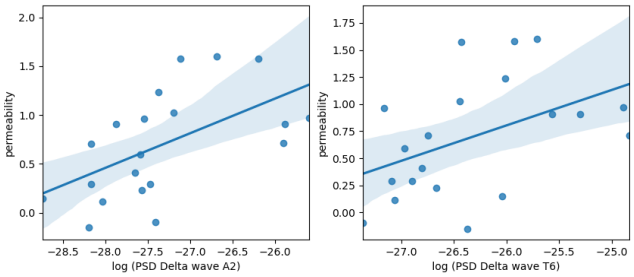

Figure E.3: Regression model linking permeability and mean PSDs.

These results encourage us to say that there are links between the EEG data and the intestinal permeability of alcoholic patients. Specifically, gut permeability would be positively correlated with the power spectral density of the EEG bands at rest during eye closure. However, due to the small number of data, these results cannot be considered significant. It would be interesting to carry out this study with a larger sample size to see if there are still statistical differences between the two groups. If this the case, it would then be plausible to use machine learning algorithms to predict the patients' intestinal permeability from the PSDs of the EEG signals.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain
Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/epl**