Machine Learning for Analysis of Brain Signals

Machine Learning for Analysis of Brain Signals

BY

FATEMEH ARMAN FARD, M.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

© Copyright by Fatemeh Arman Fard, October 15, 2020

All Rights Reserved

Master of Applied Science (2020) McMaster University

Electrical and Computer Engineering Hamilton, Ontario, Canada

TITLE: Machine Learning For Analysis Of Brain Signal

AUTHOR: Fatemeh Arman Fard,

B.Sc. (Software Engineering)

McMaster University, Hamilton, Canada

SUPERVISOR: Dr. James Reilly

Dr. John Connolly

NUMBER OF PAGES: xi, 66

Abstract

Event-Related Potential (ERP) measures derived from the electroencephalogram (EEG) have been widely used in outcome prediction of brain disorders. Recently, the ERPs that are transient (EEG) responses to auditory, visual, or tactile stimuli, have been introduced as useful predictors of a positive coma outcome (i.e, emergence from coma).

In this study, machine learning techniques were applied for detecting the Mismatch Negativity (MMN) component, which is a transient EEG response to auditory stimuli and its existence has a high correlation with coma awakening, through analyzing ERPs signals recorded from healthy control brain signals. To this end, two different dimensionality reduction methods, Localized Feature Selection (LFS) and minimum-redundancy maximum-relevance (mRMR) were employed, where a localized classifier and the support vector machine (SVM) with radial basis function (RBF) kernel are used as classifiers. We trained both LFS and mRMR algorithms using signals of healthy brains and evaluated their performance for MMN detection on both healthy subjects and coma patients. The evaluation on healthy subjects, using leave-one-subject-out cross-validation technique, shows the detection accuracy performance of 86.6% (using LFS) and 86.5% (using mRMR).

In addition to analyzing brain signals for MMN detection, we also implemented a machine learning algorithm for discriminating healthy subjects from those who have experienced TBI. The EEG signals used in the TBI study were recorded using an ERP paradigm.

However, we treated the recorded signals as resting state signals. To this end, we used the mRMR feature selection method and fed the selected features into the SVM classifier that outputs the estimated class labels. This method gives us a poor performance compared to the methods that directly used ERP components (without considering them as resting signals.). We conclude that our hypothesis of treating ERP data as resting data is not valid for TBI detection.

Acknowledgments

First of all, I would like to appreciate my supervisor, Dr. James P.Reilly, who has diligently guided me in the last two years. The door to his office was always open whenever I faced any challenge in my research.

Second, I would like to thank Dr. John Connolly, for much needed guidance in the field of TBI and coma outcome prediction.

I also would like to express my gratitude to my friends and colleagues at McMaster University who helped me go through my Master's study.

In the end, I would like to dedicate this thesis to my wonderful parents, who have sacrificed and invested 27 years of their lives for me to grow, learn and eventually teach. I would like to thank my brother and sisters for their consistent encouragement in the last two years. Thank you for always believing in me and pushing me to try harder.

Notation and abbreviation

BDI II	Beck Depression Inventory II
DAI	Diffuse Axonal Injury
DoC	Disordes of Consciousness
DRL	Driven Right-Leg
EEG	Electroencephalography
EOG	Electrooculogram
ERP	Event Related Potential
EP	Evoked Potential
GCS	Glasgow Coma Scale
HI-REB	Hamilton Integrated Research Ethics Board
ImPACT	Immediate Post-Concussion Assessment and Cognitive Test
ICA	Independent Component Analysis
ISI	Inter-Stimulus Interval
KNN	K Nearest Neighbor
LOO	Leave One-subject Out
LDA	Linear Discriminant Analysis
LFS	Locolized Feature Selection
ML	Machine Learning

mRMR	minimum Redundancy Maximum Relevant
MMN	Mismatch Negativity
PLV	Phase-Locking Value
PCSS	Post-Concussion Symptom Scale
PSD	Power Spectral Density
PCA	Principal Component Analysis
RBF	Radial Basis Function
rCFL	retired Canadian Football League
SF-36	Short Form Health Survey
SON	Subject's Own Name
SVM	Support Vector Machine
TBI	Traumatic Brain Injury
YLD	Years Lived with Disability

Table of Contents

List of Figures

Chapter 1

1 Introduction

Coma is a state of prolonged unconsciousness that has a variety of etiologies, e.g. traumatic brain injury, stroke, brain tumor, drug or alcohol intoxication (Young, Ropper, and Bolton 1997). Clinicians will often use neuroimaging techniques such as MRI or CT to fully evaluate the extent of the patient's injuries. Prognostication in these cases requires the ongoing review of these tests but is highly subjective and dependent on the individual clinician performing the assessment. In more recent years, there has been a strong push to move away from these highly subjective tests and towards more objective measures (Armanfard et al. 2016).

Coma state usually lasts for a few weeks up to one month, and transitions into either unresponsive wakefulness syndrome (also known as vegetative state) or minimally conscious state, and is generally the result of bihemispheric lesions of the cortex or white matter, focal lesions of the paramedian tegmentum, or bilateral thalamic damage (Connolly et al. 2019).

The limited availability of intensive-care treatment, the requirements of planning individual patient management, and the need for counseling family members with realistic expectations, make result prediction valuable for patients, their family and attendant medical staff. Online assessment of comatose patients is very important because it provides us the potential to detect short rises in the level of consciousness, thus improving both outcome prediction and the rehabilitation process (Kane, Nicholas M and Butler, Stuart R and Simpson 2000; N Armanfard, Reilly, and Komeili 2018; Armanfard et al. 2019).

Traumatic brain injury (TBI) is a nondegenerative, noncongenital insult to the brain from an external mechanical impact (Freire et al. 2011). The mechanical impact possibly leads to permanent or temporary impairment of cognitive, physical, and psychosocial functions, with an associated diminished or altered state of consciousness (Syed et al. 2007). Moderate to severe traumatic brain injury may result in prolonged or permanent changes in a person's state of consciousness, awareness, or responsiveness (Purbhoo 2018). the current methodology for TBI detection is based on verbal questions, requires considerable training and expertise to administer, and is also highly subjective (Prince and Bruhns 2017).

Brain injury has become a significant issue at the local, provincial, national, and international levels (Perel et al. 2008). Among all types of injuries in the world, injuries to the brain are among the most probable causes of death or permanent disability (Dennis et al. 2017). Therefore, early detection of TBI can help prevent serious impairment while simultaneously improving the efficacy of the health care system.

The proposed Machine Learning (ML) process requires a training set, which consists of EEG data from healthy subjects as well as data from TBI patients. Each data record is labeled as either healthy or TBI. The objective of our ML process is to discriminate between these two classes. If this proves possible, then we can implicitly detect the presence of TBI. The training data is provided by Prof. John Connolly, the Master's co-supervisor of the applicant.

The potential impact of the TBI project is the development of an objective method to detect the existence of TBI using EEG signals. As previously mentioned, early detection of TBI is vital to saving human lives and in mitigating permanent impairment to the brain. Such objective measure offers an inexpensive brain scanning tool that can significantly decrease the economic burden on the healthcare system and to individual families.

1.1 Motivation

Mental disorders have been considered as the highest burden among global health problems, contributing about 32.4% years lived with disability (YLDs) and a cost of 2.5 trillion US dollars including both the direct and indirect costs(Trautmann, Rehm, and Wittchen 2016; Vigo, Thornicroft, and Atun 2016; Whiteford et al. 2013) which is expected to double by 2030(Cao and Reilly 2019).

We wish to diagnose or identify various disorders of consciousness (DoC) by comparing a patient's brain responses to those of healthy controls. To do so we elicit event-related potentials (ERPs) from both the patient and controls and compare their responses using machine learning techniques. Any abnormal difference in the patient's response is an indication of a potential brain deficit or disorder, which may have arisen e.g., from a brain injury or some congenital condition.

1.2 Thesis Organization

The following is an overview of the contents of each chapter in this thesis. This study has four main parts: first, an introduction to the required background and related works. Second, in Chapter 2 the methodology and the related datasets for both TBI and coma research will be discussed. Third, Chapter 3 presents the experimental results and gives a brief discussion for each research topic. Fourth, in Chapter 4 we conclude our findings for this study and present suggestions for future work.

1.3 Electroencephalography and the Event Related Potential

Electroencephalography (EEG) allows reading brain signals which can be measured non-invasively from the scalp, with a higher temporal resolution on the order of milliseconds rather than seconds by applying smart signal processing techniques (Anwar, Batool, and Majid 2019). EEG measures the electrical activity of large postsynaptic potentials in the brain with electrodes placed on the scalp (Light et al. 2010).

Event-related potentials (ERPs) are EEG brain responses evoked to specific types of stimuli. Since these signals are usually hard to detect in continuous EEG recordings, it is common to average windows of preprocessed EEG data (trials or segments) across individual occurrences of a type of stimulus in question (Sculthorpe-Petley et al. 2015). Some of the early potentials are elicited or emitted due to direct sensory, cognitive or motor input (exogenous) and are mostly referred to as evoked potentials (EPs) (Sculthorpe-Petley et al. 2015).

One of the most frequent paradigms used in the literature investigates "pre-attentive" processing - that is, neural processing at a lower level of conscious awareness in the individual, yet returns selective processing of a stimulus by its deviance from a settled sequence of stimulation. This paradigm is known as the *oddball paradigm* and basically includes two types of auditory stimuli: standard tones and deviant tones, where repetitive standard tones are interspersed with slightly deviant stimuli. This useful paradigm elicits two different long-latency ERP components: the N1(N100) and the Mismatch Negativity (MMN) (Armanfard et al. 2019; Duncan et al. 2009; Armanfard et al. 2016). Examples of ERPs and EPs used in the present thesis are discussed below.

1.3.1 The N1 and the MMN (from Armanfard, 2019)

"The oddball paradigm elicits two different long latency ERP components: the N1 and the Mismatch Negativity (MMN). The presence of N1 and MMN (elicited at respectively about 100 and 150 millisecond post-stimulus) provides evidence of basic brain function at a level reflecting cortical function. The N1 is an obligatory sensory response evoked by each tone (i.e. both standard and deviant) and highlights the encoding of acoustic input in the auditory cortex. The MMN is an automatic response to auditory stimuli that deviate from the ongoing context of identical auditory stimuli. It reflects automatic sensory memory processes (R Näätänen et al. 2007; R Näätänen, Gaillard, and Mäntysalo 1978). Although the MMN is often referred to as a "pre-attentive" response, the evidence from sleep and anesthesia research indicates that a state of consciousness is required for the response to occur."(Armanfard et al. 2019)

The Mismatch Negativity is an ERP component that appears when a passive odd-ball auditory stimulus sequence is applied (Armanfard et al. 2019, 2016). This component is elicited from a deviant stimulus from an established pattern. It is a negative deflection, peaking around 150–250 ms after the onset of the deviant stimulus (Beres 2017; Todd et al. 2008). The MMN has been shown to arise due to several types of deviant stimuli (Todd et al. 2008). Even though the temporal aspect of the MMN is similar to that of the N100, studies have shown that the MMN is dissociated from the N100 in its function. The MMN has been argued to be a manifestation of a part of the underlying mechanism for auditory awareness (Risto Näätänen, Jacobsen, and Winkler 2005; Risto Näätänen 2001).

Generally, using 5–10 active electrodes to record MMN is sufficient, which should consist of at least Fz, Cz, C3, C4, and mastoid locations. The preferred reference is the nose (Duncan et al. 2009). The MMN is normally extracted from a difference waveform achieved by subtracting the averaged standard ERP waveform from the averaged ERP response to the deviant stimulus (Duncan et al. 2009).

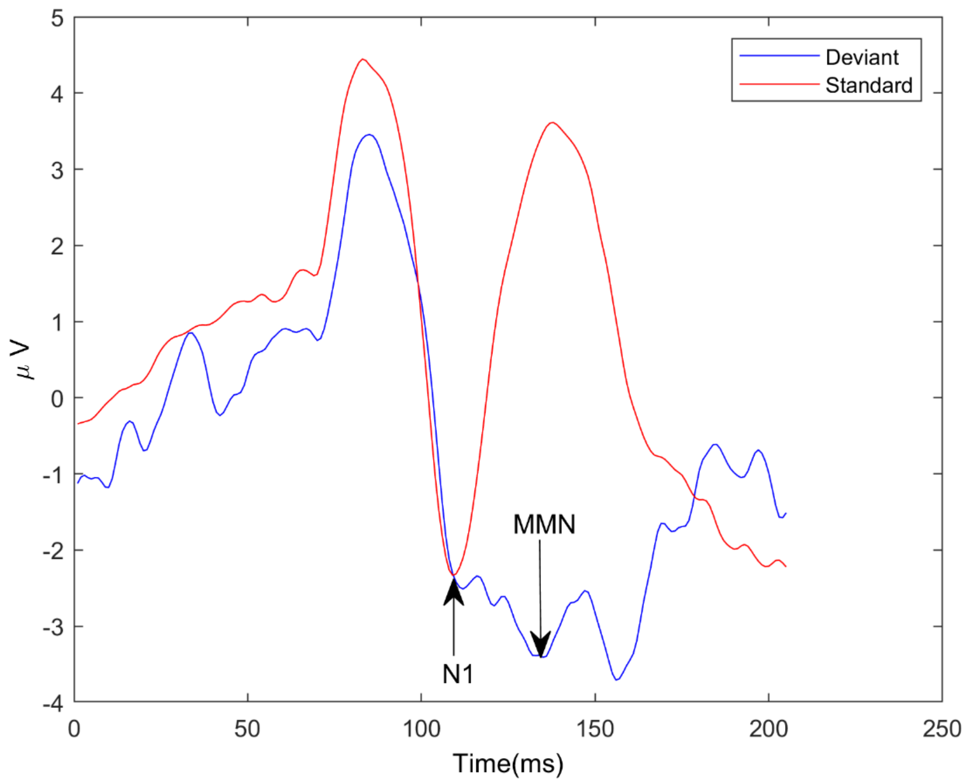

Figure 1-1: Average of de-artifacted epochs corresponding to standard (red) and deviant (blue) stimuli of one of the healthy subjects at channel Fz.

In Figure 1-1, two averaged signals for a typical training subject (at channel Fz) are shown. The standard signal has only the N1 component and the deviant signal has both the N1 and MMN components Figure 1-1: Average of de-artifacted epochs corresponding to standard (red) and deviant (blue) stimuli of one of the healthy subjects at channel Fz.

If a patient's response to a deviant stimulus appears similar to the response from a standard stimulus, that indicates potential for a possible brain disorder. It is therefore a major objective of this study to use machine learning methods to discriminate between standard and deviant responses in healthy subjects and in DoC subjects, so that anomalous responses may be identified. More specifically, work by ((Fischer et al. 1999; Morlet and Fischer 2014) has shown that if the MMN appears in comatose patients, then it is highly likely the patient will emerge (i.e. the presence of the MMN has a high positive predictive value). But only about 30% of patients who had regained consciousness showed MMN (i.e. a low sensitivity). This is an indication that the MMN may be present but difficult to detect in many cases. We, therefore intend to investigate whether machine learning methods may improve the detectability of the MMN and therefore lead to an improved test for coma emergence. This is equivalent to being able to distinguish the difference between standard responses and deviant responses in coma patients.

1.4 Machine Learning Background

Machine Learning (ML) is a field of technology that allows machines to extract knowledge from data and self-improve as new data becomes available. The ML field has been one of the most ubiquitous methods encountered in different aspects of our modern life.

The machine learning paradigm is a promising new technique for analysis of the EEG for applications in neuroscience. The brain is far too complicated an organism to enable modeling by classical means, a process that would typically involve the use of mathematical and physical models of brain structure and function to predict brain behavior in some way (Cao and Reilly 2019).

To give an example of how ML works, imagine your machine's goal is to use a given dataset of e.g. images (called the training set) to learn to distinguish the class of each certain image. The class label should be already defined, e.g. vehicle or text. Since the machine has no a priori knowledge about the classes, it creates a large pool of measurements (candidate features) from each image sample. Candidate features could be any type of parameter that ideally changes value between the classes. Most of the time, the number of candidate features is far too large. To reduce the number of features to a more manageable level, the features are fed into a feature selection algorithm. The selected features are the candidate features that are the most effective in class discrimination. Afterward, a designed classifier uses the selected features to separate the two classes of the training set, as far as possible, into separate regions in a Cartesian co-ordinate space, whose axes are the selected features. This is referred to as the feature space. Each region corresponds to a class. In this step, the final machine learning model has been identified. However, we might wish to apply a cross-validation procedure to calculate the accuracy of the model to avoid misclassification as much as possible (Cao and Reilly 2019).

There are roughly three types of Machine Learning algorithms. Supervised, Unsupervised, and Semi-Supervised Learning. In supervised learning, a labeled dataset would be provided to train the model in the so-called training phase. It requires a training set of labeled documents and returns a function that maps data samples to the predefined class labels

(Arzucan ozgür 2002). In unsupervised learning methods, the machine is given a set of unlabeled data and is tasked with finding various forms of information to distinguish the data into meaningful groups (D'Urso and De Giovanni 2018). Lastly, in semi-supervised learning, the machine trains in the presence of both labeled and unlabeled data. The goal of semi-supervised learning is to find an algorithm that takes advantage of a combination of labeled and unlabeled data (Zhu, Xiaojin and Goldberg 2009).

The odd-ball ERP paradigm (as described in the following sub-section) is applied to each subject as described in Sect. 1.3 Since it is known beforehand whether the respective trial is standard or deviant, our class labels are available, and therefore in the present study, we consider only supervised learning methods.

1.4.1 Feature Selection

Candidate feature counts in typical applications in neuroscience, psychiatry, and generally in a lot of medical applications, have a tendency to be large; however, there are limited available training samples. Generally, feature selection methods aim to recognize those features whose level of statistical dependency with the class label is high. Therefore, the value of the selected features changes substantially with the class. A well-designed feature selection algorithm must analyze every potential combination of every existing N candidate features for relevance. Since feature selection is a well studied topic, there are quite large number of feature selection algorithms in the literature. Armanfard, Reilly, and Komeili (2018) provides several feature selection methods. The minimum redundancy maximum relevance (mRMR) method is a feature selection method that has been demonstrated to work effectively for applications in brain research. In the mRMR method, mutual information is used as a measure of statistical dependency. This method is an iterative greedy approach. In this approach, a single feature which has the maximum mutual information with the class labels (relevance) is chosen in the first iteration. Then in subsequent iterations, the selected features must have a combination of minimum mutual information (redundancy) with the features selected in the previous iterations (redundancy) and maximum relevance with respect to the class labels (Cao and Reilly 2019).

Other feature selection approaches such as Principal Component Analysis (PCA) (Al-Kandari and Jolliffe 2005), Linear Discriminant Analysis (LDA) (Duda, Hart, and Stork 2001) and Independent Component Analysis (ICA) (Hyvärinen and Oja 2000), facilitate dimensionally reduction by merging original features to acquire a new set of features.

Resulting features from this approach usually lose their physical interpretation in terms of the original features. Dimensionality reduction, with no transformation, by choosing a subset of the original features is the basis for the feature selection approaches. Therefore, feature selection approaches keep the physical interpretability property in terms of the selected features. For this research study, we use feature selection methods (Armanfard, 2017).

Batch methods and online algorithms are two categories of feature selection algorithms. In batch methods, the process of feature selection task is conducted offline where features of training instances are available. However, for online feature selection algorithms, the full feature space is considered unknown in advance. The applications of online methods are where the training samples or features arrive in a sequential manner (Wang et al. 2014; Wu et al. 2013; Yu et al. 2014). For this study, the batch algorithms are used (Armanfard, 2017).

1.4.2 Classification

Classification is a very mature topic and subsequently, there are many types of classification methods. Following feature selection, the samples from each class in the training set divide (i.e. cluster) as well as possible into two separate regions in the feature space (Cao and Reilly 2019). Typical feature selection methods choose an optimal global feature subset that is applied over all regions of the sample space (Armanfard, 2017). The classifier determines the most likely cluster that a test point belongs to. Note that points that fall into an overlap region between clusters may not classify correctly (Cao and Reilly 2019).

There are various types of classifier. The support vector machine (SVM)(Hastie, Trevor and Tibshirani, Robert and Friedman 2009; Haykin 2009) is one of the most well-established classification methods that has shown reasonable performance in neuropsychiatric applications. Rather than SVM, we can also mention K Nearest Neighbor (KNN), the Linear Discriminant Analyzer (LDA) and, decision trees. These methods are all described in detail in (Hastie, Trevor and Tibshirani, Robert and Friedman 2009; Rumelhart, Hinton, and Williams 1986). The first such approach which has proven useful in brain studies is the mRMR feature selection scheme in conjunction with an SVM classifier (Cao and Reilly 2019).

The LFS method selects a feature subset such that, within a localized region, within-class and between-class distances are respectively minimized and maximized. This allows the feature set to optimally adapt to local variations of the sample space. The process of computing a specific feature subset for each region is independent of those of other regions and hence can be performed in parallel; it is also an appropriate approach for the case where the data are distributed on a non-linear and/or a disjoint manifold. The method selects only relevant features, so the LFS method is not overly sensitive to the overfitting problem (Armanfard, 2017). The LFS method is suitable to the "data poor" case where the number of candidate features far exceeds the number of training samples, and is also resistant to the overfitting problem. The LFS method has proven to be successful in predicting emergence in coma patients (Armanfard et al., 2016).

1.4.3 Validation

A very important component of machine learning model is validation. The usual form of validation is *cross-validation*, where the available training set is split into two parts, where one part is larger than the other. The larger is referred to as a *validation set*, and the other the *test set*. The machine learning model only uses the *validation set* to train (Cao and Reilly 2019) and uses the smaller test set for the testing step. In the following, we introduce two important cross-validation methods; K-fold and Leave One Out (LOO).

K-Fold: in this method, the entire training set divides into K contiguous folds. For a trustful evaluation, we shuffle the dataset before dividing. Then, for training, the model uses one fold for the testing set and the remaining folds feed into the model as the training set. This process iterates K times, where each fold is held out exactly once. Therefore in the end, we have K accuracy results, which are averaged over the folds and reported as the final result (Cao and Reilly 2019).

Leave One Out (LOO): This method basically acts like K-Fold but with a slight difference. For a dataset with N instances, we leave out one instance as a test set, and the remaining data are used as the training set, and repeat over all instances so that each instance is left out once. In the end, we have N results which we average and report the final result (Armanfard et al., 2016; Linden, 2019).

Chapter 2

2 Methodology

2.1 Introduction

In this chapter, First, the basics of machine learning and various machine learning methods are explained. In this research different ML algorithms are applied to the datasets. In this study, as explained in Sect.2.2 below, two datasets have been used so that a broader range of applications of EEG/ERP signals may be explored. The coma dataset includes 26 healthy controls and 2 coma patients. The traumatic brain injury (TBI) dataset includes fourty-three subjects, twenty of which were retired football players who had experienced concussion.

2.2 Datasets

2.2.1 Coma data

The study recruited two coma patients. Patient 1: age = 29, Gender: male, Glasgow Coma Scale (GCS): 5, Etiology: traumatic. EEG recording was conducted 13 days post-injury. Patient 2: age = 21, Gender: male, GCS: 4, Etiology: motor vehicle accident - diffuse axonal injury (DAI) and hypoxic ischemia. EEG recording was conducted for 27 days post-injury.

Twenty-six control subjects who had no history of concussion or any other type of neurological disorder and were recruited through the local newspaper, personal contacts, and McMaster University.

In this study, the N1 and MMN components were elicited using an alteration of a classic auditory oddball paradigm, as described in part in (Fischer, Dailler, and Morlet 2008). Stimuli consisted of deviant tones (14%), the subject's own name (SON), which was spoken by a native female English speaker with a neutral voice (3%), the novel sound of dog bark recorded digitally (3%), and standard tones (80%). These stimuli were randomly presented; however, each deviant was preceded by at least two standard tones. It used a duration deviant, which is known to be one of the stronger types of "deviant" features, both for evoking the MMN but also for producing one of the more stable MMN waveforms over time (Escera et al. 2000).

Both healthy and coma subjects were examined with the passive oddball paradigm, comprised of deviant and standard tones of 800 Hz with a duration of 30 ms and 75 ms, respectively (Armanfard et al. 2019). In this process, 2000 stimuli were presented in total -- 1600 standard tones, 280 deviants, 60 SON, and 60 novels (dog bark). The stimuli were pseudorandomized so that no two deviant/novels were presented consecutively; there were at least two standards in between each deviant or novel. The total duration of the EEG recording was approximately 25 minutes.

With regard to healthy/controls participants, the EEG was recorded from a 64-channel BioSemi ActiceTwo system and a 0.01– 100 Hz bandpass that was digitally sampled at

512 Hz. Five Ag/AgCl external electrodes configured in a 10-20 montage were placed on the subject's nose, left and right mastoids, and above and over the outer canthus of the left eye. For coma patients (depending on the patient), the EEG has recorded bedside in the ICU from either a 32 or 8 channel BioSemi ActiceTwo system as described earlier.

Electrodes were placed on the scalp according to the standard 10/20 positioning using a 32-electrode cap. Vertical and horizontal electrooculogram signals were monitored by electrodes placed above and over the outer canthus of the left eye. References were recorded bilaterally from the mastoids and at the nose for offline rereferencing. In the case of a skull fracture or any obstruction to the placement of a regular cap, a customized cap was used with a reduced number of electrodes. Similarly, data from healthy controls were recorded using a 64-channel EEG cap.

Initially, we collected data at the Hamilton General Hospital from comatose patients with a new protocol as described in detail in Connolly et al. (2019). However, due to the COVID-19 pandemic, both patient and healthy control recording was stopped. In the following, we present a brief summary of the protocol which was originally intended for this study.

Data is to be collected from 50 coma patients. EEG/ERP data will be recorded for 24 consecutive hours at a maximum of five times points covering 30 days from the date of recruitment to track participants' progression. The study employs paradigms designed to elicit brainstem potentials, middle-latency responses, mismatch negativity, P300, N100, and N400. In the case of patient emergence, data are recorded on that occasion to form an additional basis for comparison. A relevant healthy controls data set will be collected from the testing of 20 participants, each extends over a 15-hour recording period to formulate a baseline. (Connolly et al. 2019)

2.2.2 TBI data

The study, approved by the Hamilton Integrated Research Ethics Board (HI-REB), Hamilton, Ontario, Canada, recruited twenty retired Canadian Football League (CFL) athletes (rCFL) with histories of concussions (mean age = 57.6, range = 45–66 years) and twenty-three healthy age-matched control subjects (mean age = 53.7, range 45–61). Control subjects had no history of concussion or any other type of neurological disorder and were recruited through the local newspaper, personal contacts, and McMaster University. All participants (who were native English speakers and self-reported as having no hearing issues) provided informed consent, in accordance with the ethical standards of the Declaration of Helsinki, prior to participation in the experiment. Participants were assessed using the Immediate Post-Concussion Assessment and Cognitive Test (ImPACT), Beck Depression Inventory II (BDI II), Short Form Health Survey (SF-36), and the Post-Concussion Symptom Scale (PCSS) (Ruiter et al. 2019).

Two different protocols were used to examine two distinct cognitive processes. The first protocol, adapted from Todd et al. (2008), was a P300 auditory oddball task that consisted of one Standard tone (ST, 1000 Hz, 80 dB SPL [sound pressure level], 50 ms duration) and three deviant tones differing from ST in Frequency (FT, 1200 Hz, 80 dB SPL, 50 ms), Intensity (IT, 1000 Hz, 90 dB SPL, 50 ms), and Duration (DT, 1000 Hz, 80 dB SPL, 100 ms). The protocol employed an inter-stimulus interval (ISI) of 1000 ms. Each deviant tone was presented 36 times representing 6% of the stimulus set while the ST was presented 492 times representing 82% of the stimulus set. Participants were asked to left-click to every ST and right-click to all deviant tones to be sure they were responding to stimuli; this procedure was counterbalanced within-subjects halfway through the protocol. The response requirement in this task was designed to engage active attentional processes and invoke the P3b (Ruiter et al. 2019).

The second protocol, developed by Todd et al. (2008), was an extended version of the same auditory oddball task used in the first protocol, but with different procedures to enable the examination of pre-attentive processes as manifested by the MMN. A total of 2400 tones, with a 500 ms ISI, were used in this experiment with each deviant tone being presented 144 times representing 6% of the stimulus set, while the ST was presented 1968 times representing 82% of the stimulus set. Instead of attending to the stimuli, participants were informed that the tones were of no relevance to the study and instructed that they need to only watch a nature movie. The film was an edited version of a nature program with the auditory track removed and only visually neutral scenes shown (Ruiter et al. 2019).

Finally, protocols 1 and 2 were presented in that order but were separated by an additional experiment requiring participants to judge the grammaticality of spoken sentences and make a ‘‘correct/ incorrect” manual response to each sentence. This task created a distraction break of 10–15 minutes between the two oddball tasks (Ruiter et al. 2019).

EEG was recorded from 64 Ag/AgCl electrodes (International 10–20 system) using a BioSemi ActiveTwo system and a 0.01– 100 Hz bandpass (with a 60 Hz notch filter employed) that was digitally sampled at 512 Hz. Five Ag/AgCl external electrodes were placed on the subject's nose, left and right mastoids, and above and over the outer canthus of the left eye. The EOG (electrooculogram) was recorded (using the same bandpass and sampling rate) from the external electrodes placed above and over the outer canthus of the left eye. EEG acquisition was referenced to the driven right-leg (DRL) and re-referenced offline to the average of the mastoids.

2.3 Methods

2.3.1 Coma project

The goal of this project was to apply a supervised machine learning algorithm to detect Mismatch Negativity on comatose patients. In this section, we propose a supervised machine learning-based algorithm for automatic and continuous assessment of a subject. We had twenty-six subjects from which two of them were comatose patients and the remaining subjects are healthy controls. More details about subjects and how ERP stimuli(standard and deviant) are generated are explained in section 2.2.1.

A large quantity of EEG data, recorded under an auditory oddball paradigm, was available. A previous study (Armanfard et al. 2019) used this data to detect Mismatch Negativity components evoked from an oddball paradigm using the Local Feature Selection (LFS) method. Specifically, in this present project, we wish to extend the previous study with different methods to compare to the LFS algorithm.

The supervised Machine Learning approach has two phases: phase 1, which is the training phase, and phase 2, which is the test phase. The machine learning process consists of these stages: 1) pre-processing 2) feature extraction and 3) feature selection, 4) classification, and 5) validation.

A. Learning phase

The ML algorithm has two classes. The first class D corresponds to the presence of N1 + MMN components of deviant tones and the second class S correspond to the presence of N1 component of standard tones. The required training points for both classes D and S are recorded from healthy brain responses to both deviant tones (providing N1 plus MMN) and standard tones (providing N1) respectively (Armanfard 2017).

In the training phase, we only use healthy control dataset. The pre-processing step is to eliminate artifacts and filter the signals. Data preprocessing was conducted using the BrainVision Analyzer 2 platform. To eliminate as much noise (eye blinks and muscle artifacts) as possible, we filtered the raw EEG signals by a band-pass FIR filter from 2Hz to 30Hz with a filter order of 40 (Morlet and Fischer; N Armanfard, Komeili, Reilly, and Pino 2016; Armanfard, Komeili, Reilly, Mah, et al. 2016). Each subject has three corresponding files. The *.dat* file contains the EEG data itself, *.vhdr* is the history file that contains the preprocessing analyses, and the *.vmrk* contains the event marker information. The most important markers (i.e., stimuli) for this study are the S11 (standard) and S16 (duration deviant). Also there are the S1 (subject’s Own Name) and S6 (dog bark) markers.

The filtered data feeds into the EEGlab package to extract relevant components corresponding to the Standard and Deviant responses. EEGlab is an open-source MATLAB toolbox for processing different formats of data (Delorme and Makeig 2004; Brunner, Delorme, and Makeig 2013). In this step, we first load each control data segment in *vhdr* format, then specify the S16 and S11 stimulus intervals, and then extracting an analysis time window extending from 0 to 300 ms after each stimulus onset, from the entire ERP interval which extends from -100 to 1000 msec. (This is done because the data outside the smaller window is not relevant for our purposes). Since the test data only has 32 channels, we select the same channels from the training data to match the two datasets.

Furthermore, to provide reliable and stable training samples, we average the deartifacted epochs corresponding to each class, for each healthy training subject. So in the end, we have 2*26 32-channel training samples available. (There are 26 healthy subjects, each providing a standard and deviant averaged response on each of 32 channels.) Figure 2-1 demonstrates a comparison of both Standard and Deviant averaged epochs for a healthy control participant at channel Fz.

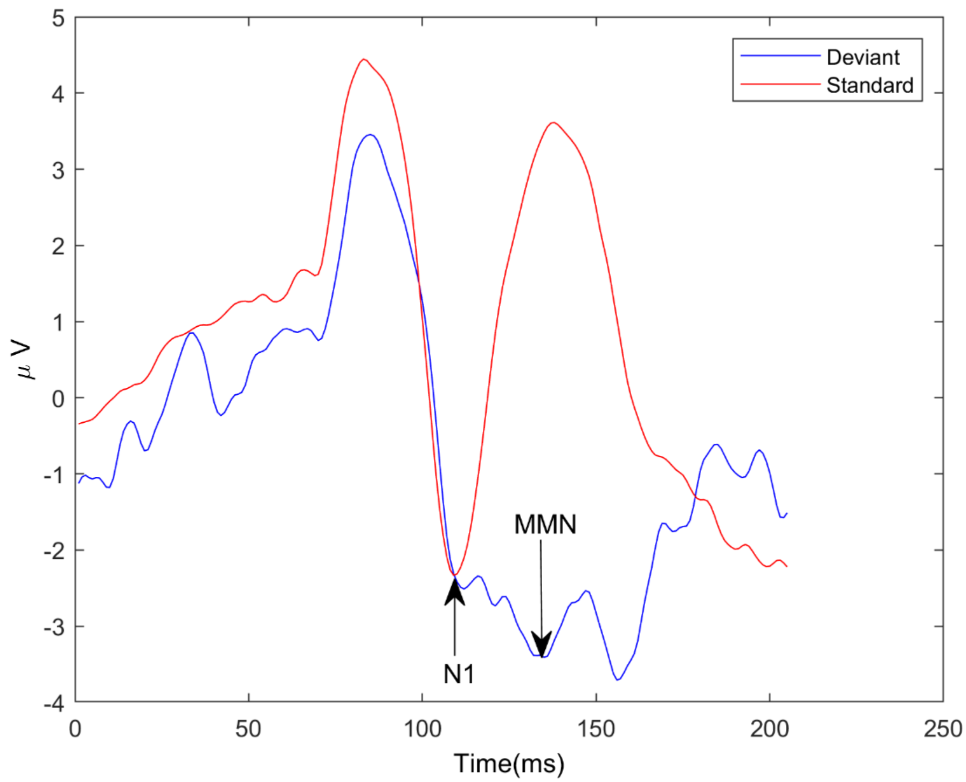

Figure 2-1: Average of de-artifacted epochs corresponding to standard (red) and deviant (blue) stimuli of one of the healthy subjects at channel Fz.

As expected, it is shown in Figure 2-1, in class D we have epochs that contain MMN and N1 components and Class S contains only the N1 component.

In the feature extraction step, a large number *(M)* of candidate features are created that represent each training point obtained from the pre-processing section. The candidate feature set in this study contains a variety of statistical quantities at each channel. The features fed into the machine learning algorithm are wavelet coefficients, kurtosis, variance, maximum, minimum, skewness, and power in eight different frequency bands: Alpha-band (8Hz to 13Hz), Beta1-band (13Hz to 20 Hz), Delta-band (1Hz to 4Hz), Lower-band (1Hz to 8Hz), Total-band (1Hz to 30 Hz), Beta-band (13Hz to 30 Hz), Beta2-band (20Hz to 30Hz) and Theta-band (4Hz to 8Hz). The wavelet 'rbio6.8' (available in MATLAB) at level 3 is used for the wavelet decomposition. Wavelets have proven effective for previous EEG (Armanfard et al. 2019). The set of candidate features were chosen based on the fact that they have been effective in previous studies (Khodayari-Rostamabad et al. 2013; Ravan et al. 2012; Orme-Johnson and Haynes 1981; Torsvall and Akerstedt 1987).

Subsequently, we concatenate all of the extracted features *M= 268* from each channel together for each subject. Then we make a matrix for each subject data from its extracted features. So after concatenating the features, for each subject there is *M=32*268* features. In the end, for each subject, there are two feature sets which are extracted from deviant and standard epochs respectively. The labels assigned to the deviant and standard stimuli are 0 and 1 respectively.

Afterward, there is feature selection step. Since *M* is large and the number of training samples small, we need to reduce the number of candidate features for efficient classification. By using feature selection methods, only the discriminative features will be selected and the irrelative ones will be ignored. In this research, both mRMR and LFS are used for feature selection.

mRMR: The greedy, iterative feature selection method mRMR (minimum Redundancy, Maximum Relevance)(Hanchuan Peng, Fuhui Long, and Ding 2005) has been used. This algorithm selects the most relevant features according to a maximal statistical dependency criterion based on mutual information. The mRMR method ranks each feature to maximize its relevance with the target class *y* and simultaneously minimizes redundancy with features selected in previous iterations. The better a feature is deemed to be, the higher the rank it is assigned.

Finally, the reduced set of features obtained from the previous step is then fed into a classifier that outputs the estimated class of an input data sample. For the purpose of this project, we use a support vector machine (SVM) classifier (Haykin 2009; Hastie, Trevor and Tibshirani, Robert and Friedman 2009). SVM is a discriminative classifier formally defined by a separating hyper-plane. In other words, given labeled training data (*supervised learning*), the algorithm outputs an optimal hyper-plane which optimally separates the two classes in the feature space. In two-dimensional space, this hyper-plane is a line dividing the feature space into two parts where each class lies on either side.

Kernelized SVM is used for classes that cannot be divided linearly. It maps the data to a higher dimension using a kernel function, where the classes can be separated linearly. In this project, a linear kernel is used.

Fitcsvm and the *predict* are predefined MATLAB commands and both are used in this program. For cross-validation, one subject's deviant and standard epochs are kept separate for testing, and the remaining ones are used to train the model (Leave One out). The average error rate is returned as the final average error for a given number of features (K best features). In this study, different values for K ranging from 1 to 20 are examined. This whole process (feature selection, classification, and cross-validation) was averaged over the 26 available subjects to provide an aggregated measure of performance.

LFS: Almost all of the previous feature selection methods select a global common feature subset for all regions of the sample space. These methods may not be appropriate for complex classification problems (such as classification of biological signals) (Armanfard et al. 2019). The LFS method selects a feature subset such that, within a localized region, within-class and between-class distances are respectively minimized and maximized. This allows the feature set to optimally adapt to local variations of the sample space. The process of computing a specific feature subset for each region is independent of those of other regions and hence can be performed in parallel. It is also an appropriate approach for the case where the data are distributed on a non-linear and/or a disjoint manifold. The method selects only relevant features, so the LFS method is not overly sensitive to the overfitting problem (Armanfard 2017). The LFS method is suitable to the "data poor" case where the number of candidate features far exceeds the number of training samples, and is also resistant to the overfitting problem.

In this step, all the created features are fed to LFS algorithm. Again the Leave One Out (LOO) cross-validation process was applied here. Default values for the parameters associated with the LFS algorithm are used, except for the parameter α, which is suggested to be set at the value 19 (Armanfard 2017). Finally, overall performance is obtained by averaging over the 26 subjects.

The training process for the LFS algorithm identifies a distinct set of features for each available training sample. The features are selected so that locally, in the region surrounding the training sample under consideration, other training samples of the same class cluster as closely as possible to the considered training point, and other training samples of the opposite class as removed as far as possible from the considered point. The training process involves identifying the unique set of features associated with each training sample.

B. Testing phase

In this phase, only coma data was used as testing data. EEG data from a test subject is preprocessed in the same way as described in the training phase. Preprocessed data is fed into EEGlab to extract relevant components corresponding only to deviant stimuli since MMN only appears on deviant tones. This process is the same as that explained in the previous section; however, in the training phase, we extracted standard epochs as well. Then, we average over deviant trials. With previous methods, we average over all available data of the same stimulus type, where the recording interval could extend over a few days. In this study, we instead create short 2-minute windows with a minute overlap for each patient's EEG data. In this manner, we can monitor the patient's responses over every window, thus providing a finer time scale of patient behavior, relative to the averages taken over days. The use of a shorter window also allows us to track changes in the patient's response over the recording interval, that may reflect a waxing and waning in the level of the patient's consciousness.

For each test sample, we extract the same features that were used in the training phase. For patient 1, we have $M=32*268 = 8576$ candidate features in each of 310 windows. For patient 2, we have $M=8576$ features over 210 windows.

We use two methods for feature selection -- these are the LFS and the mRMR methods. The testing procedure for LFS proceeds as follows. We associate a hypersphere of a specified radius, centered on the training sample, in the coordinate space specified by the features identified in the training phase for the respective training sample. The hypersphere adopts the class of the corresponding training sample. We then test how many hyperspheres of each class contain the test sample, and assign the class of the test sample by a majority vote.

A useful property of the LFS method is that we can establish a *similarity measure* (between 0 and 1) of a test point to each class. This is achieved by taking the ratio of the number of hyperspheres containing the test point for a particular class, to the total number of training samples of that class.

For the mRMR method, we identify a global set of features using the entire training set. Using the selected features, we can specify an SVM classifier, and then apply a cross-validation technique to assess accuracy.

2.3.2 TBI project

This study was done under MacData fellowship with cooperation from Fatemeh Yasdanpanah, an undergrad student in the Electrical and Management program at McMaster who participated in this project as an undergraduate thesis project. The goal of this project is to apply supervised machine learning techniques to detect prior concussion in individuals, among a pool of test subjects. We had forty-three subjects, 20 of which were retired football players who had experienced concussion, and the remaining subjects are healthy controls. A large quantity of EEG data, recorded under an auditory odd-ball paradigm, was available. A previous study (Ruiter et al. 2019) used this data to indicate that MMN and P300 components evoked from an oddball paradigm are significantly altered in retired Canadian Football League athletes, even though they had been retired for up to a few decades. Specifically in this project, we wish to test whether a supervised machine learning algorithm can discriminate brain injury in the retired athletes, under the hypothesis that this EEG data (which was recorded as ERPs, i.e., under an odd-ball paradigm, synchronized with a stimulus train) can be treated as resting EEG data, as if the stimulus were not present. In other words, the synchrony of the recorded ERP data with the stimulus presentation was ignored when processing the data for this study. If this hypothesis approves, the diagnose of TBI would be faster and easier.

The supervised Machine Learning approach has two phases: phase 1) the training phase and phase 2) the test phase. In the training phase, which is offline, we build a Machine Learning model using training data. In the testing phase which is online, we test a new unseen object to detect if the athlete suffers from TBI. The proposed traumatic brain injury scheme in the training phase consists of five successive stages: 1) pre-processing, 2) feature extraction, 3) feature selection, 4) classification, and 5) cross-validation.

The pre-processing step is to characterize EEG signals corresponding to the traumatic brain injury state. First, we use recorded EEG signals from healthy and TBI participants. Then, to eliminate as much noise as possible, we bandpass filter the raw EEG signals from 2Hz to 30Hz. These signals are then divided into two groups: 1) training set- a subset to train a model and 2) test set- a subset to test the trained model. The training set is large enough to yield a statistically meaningful representation of the underlying data. The test set is representative of the data set as a whole. In other words, we attempt to select a test set with similar characteristics to the training set.

In the feature extraction step, a large number of candidate features are extracted. The features fed into the machine learning algorithm are Power Spectral Density (PSD), Coherence, Fractal dimension, and Phase-Locking Value (PLV). Not every feature in the candidate feature set is equally relevant to the TBI state. Features may not be independent of each other. Having irrelevant or dependent features may degrade the accuracy and efficiency of the ML prediction and lead to overfitting. Therefore, to improve the discrimination performance between the TBI vs. healthy states, and to avoid the overfitting issue, the candidate feature set extracted in the feature extraction step is reduced to a set of

$\widehat{M}$ most relevant features ($\widehat{M} \ll M$) using a feature selection process that selects only those features which are most statistically indicative of the TBI vs. healthy classes.

The greedy, iterative feature selection method mRMR (minimum Redundancy, Maximum Relevance)(Hanchuan Peng, Fuhui Long, and Ding 2005) has been used. This algorithm selects the most relevant features according to a maximal statistical dependency criterion based on mutual information. The mRMR method ranks each feature to maximize its relevance with the target class y and simultaneously minimizes redundancy with features selected in previous iterations. The better a feature is deemed to be, the higher the rank it is assigned.

Finally, the reduced set of features obtained from the previous step is then fed into a classifier that outputs the estimated alertness class. For the purpose of this project, we use a support vector machine (SVM) classifier (Haykin 2009; Hastie, Trevor and Tibshirani, Robert and Friedman 2009). SVM is a discriminative classifier formally defined by a separating hyper-plane. In other words, given labeled training data (*supervised learning*), the algorithm outputs an optimal hyper-plane which optimally separates the two classes in the feature space. In two-dimensional space, this hyper-plane is a line dividing the feature space into two parts where each class lies on either side.

Kernelized SVM is used for classes that cannot be divided linearly. It maps the data to a higher dimension using a kernel function, where the classes can be separated linearly. In this project, a linear kernel is used.

Svmtrain and *svmclassify* are predefined MATLAB commands and both are used in this program. The code is divided into five scripts. The following diagram shows the relationship between the scripts, in sequential order of execution.

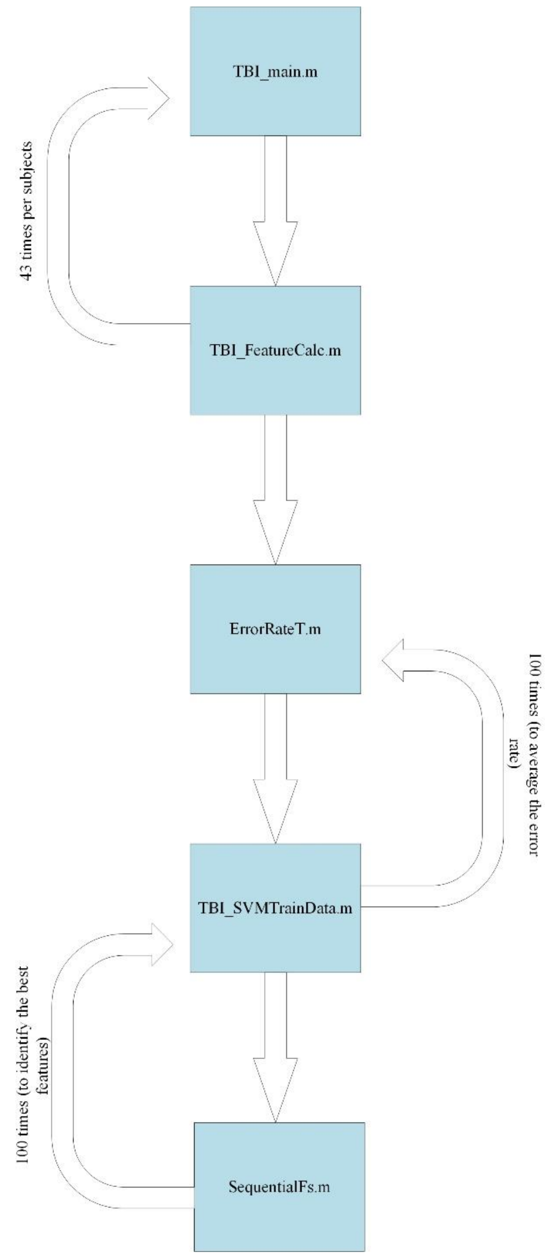

Figure 2-2: Users of this program only need to run the script “TBI_main”. “TBI_main” first runs “TBI_FeatureCalc” and then runs “ErrorRateT” which calls “TBI_SVMTrainData” and “sequentialFS”.

The script "TBI_FeatureCalc.m" takes the raw data files one subject at a time, and calculates features (PSD, Coherence, Fractal, and PLV). The calculated features are appended into one file, ready to get fed into the SVM program.

In the script "TBI_SVMTrainData.m", 25% of subjects are kept separate for testing, and the remaining 75% is used to train the model. Note that selection of subjects is random since the subject index matrix gets shuffled before every run. TBI_SVMTrainData calls the script "sequentialFS.m". This script contains the same logic for training and classification. Through the call of sequentialFs.m from within the TBI_SVMTrain, every run of the program selects features, trains, and classifies 100 times, and selects the best K-many features which result in an error rate less than sixty-four percent. Computation of the error rate is done in the script "ErrorRateT.m". Note that the development of the training model in each loop of the sequentialFS is done only using the previously selected 75% of the data. This means that the initially separated 25% of the data in TBI_SVMTrain is isolated and not involved in the training. Once the 100th loop is done, the K-many most frequently occurred features are returned to "TBI_SVMTrainData.m", where a final train and classification is done against the initially separated 25% subjects, using only those K-many features. The file "TBI_SVMTrainData.m" is called in a loop 100 times, from the script "ErrorRateT.m". The average error rate for every 100 runs is returned as the final average error for a given K. The script "ErrorRateT.mat" contains an additional loop, in case if the user wants to run the entire program for multiple values of K. For example, the user might want to run the program for K=5, 15, 25 that is to get the average error rate when the corresponding features are selected by the mRMR feature selection algorithm.

In summary, the 100 runs in the script "sequentialFS" are done so that the most frequently-occurring features while resulting in an error rate of less than sixty-four percent, are identified. Then, in "TBI_SVMTrainData" those "best" features are used to train the initially separated 75% subjects and test it against the 25%. In order to get an average of error, this whole process has been run 100 times from the script "ErrorRateT.mat". The flow diagram below provides a visual demonstration of the logic explained above:

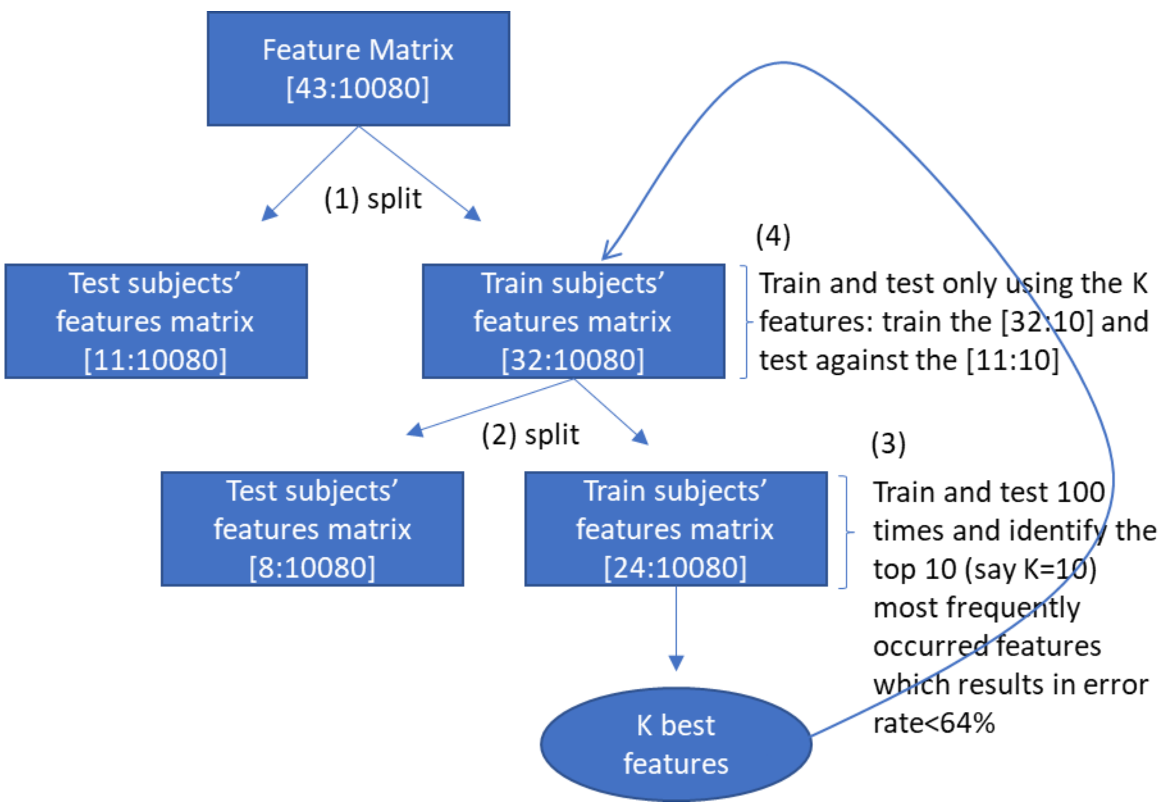

Figure 2-3: Process of choosing the best K-many features.

Four different features were calculated; PSD, Coherence, PLV, and Fractal dimension. In all cases, the number of features was significantly larger than the number of subjects. Therefore, to reduce the number of features, only 35 electrodes were chosen to be used, rather than all 64 electrodes. For Example, when PSD, Coherence, and fractal are calculated over all the electrodes, it results in a total of 35424 features per subject. This number reduces to 10745 per subject when 29 of the 64 electrodes are ignored. When features are Coherence and PSD only, and over the 35 electrodes, the total number of features is 10080 per subject. When PLV is calculated and used instead of the Coherence, and over the 35 electrodes, it results in 1155 features per subject.

Chapter 3

3 Results and Discussion

3.1 Coma project: Using conventional ERP analysis

In the training phase, we trained the two different methods for feature selection, which are mRMR and LFS, where SVM is used as a classifier in conjunction with mRMR. LFS has an associated localized classifier which allows incorporating multiple feature subsets when performing classification. We trained the method on normal subjects using Leave One Out (LOO) cross-validation technique. We also applied the ML models trained on normal subjects to two coma patients to examine the model's performance on prediction coma outcome.

3.1.1 Performance of mRMR

We trained the mRMR feature selection method using LOO cross-validation. mRMR selects the most relevant features. Classification is performed in the feature sub-space defined by mRMR where we used the support vector machine (SVM) with RBF kernel as our classifier. The parameters of the SVM (with RBF) are set to their default value in MATLAB.

The accuracy scheme corresponding to this experiment is shown in Figure 3-1. This figure shows the high performance of our ML model for classification of deviant vs. standard samples. After multiple runs and closely watching the results, it was found that the combination of mRMR and SVM has the highest accuracy. Using only the best single feature, which is extracted from mRMR, an accuracy of 86.5% was obtained.

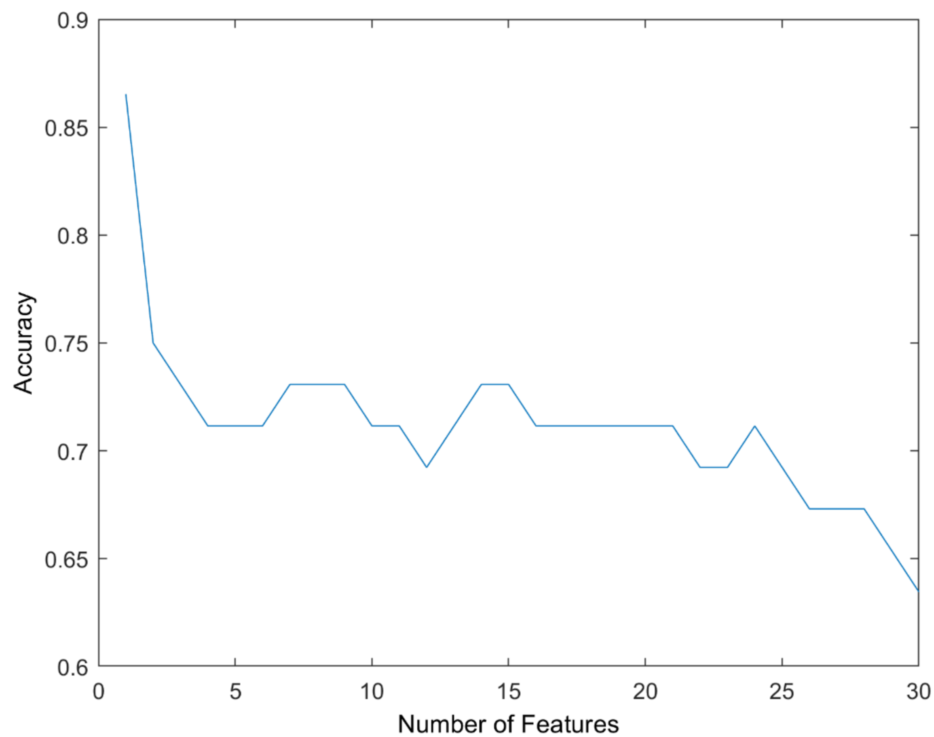

Figure 3-1: Accuracy of the ML model using different number of features that picked as best features by mRMR.

In testing phase, SVM predicted both patients as Class D. In other words, the method claims that both test data belongs to the Deviant class. The mRMR feature selection method selected the 1238th feature (wavelet decomposition) as the most discriminative feature among all the feature set. Using other less important features decreases the accuracy and harms the performance of the model.

3.1.2 Performance of LFS

On the other hand, using the LFS method combined with LOO cross-validation shows a noteworthy accuracy of 86.6% in the training phase. In another words, the trained model can predict the class of unknown labeled test data with a *probability* of 86.5%.

Furthermore, in the testing phase, we only use the deviant component of the coma patients' data as test input.

In the LFS algorithm, we only discuss the similarity value of each coma data interval to the deviant class. Since healthy controls were used for training, the similarity measure, in this case, may be interpreted as a pseudo-probability that the coma patient's deviant response is the same as that of a control deviant response. We expect if some of the similarity of intervals are high, the brain of the comatose patient will behave as a healthy brain in the future; i.e., there is indication that the patient will emerge.

In Figure 3-2 and Figure 3-3, we plot the similarity measures of the deviant responses of our two coma patients, versus the index of the respective 2-minute recording interval; i.e., the horizontal axis may be interpreted as time. We can see that the similarity measures are quite high in some intervals, giving a positive indication of emergence. We note these two patients did in fact emerge.

We also note that the similarity measure for both patients waxes and wanes over time. This is an indication that the patient's level of consciousness varies with time. This is consistent with clinical observations of coma patients being partially aware for short periods as they progress towards emergence.

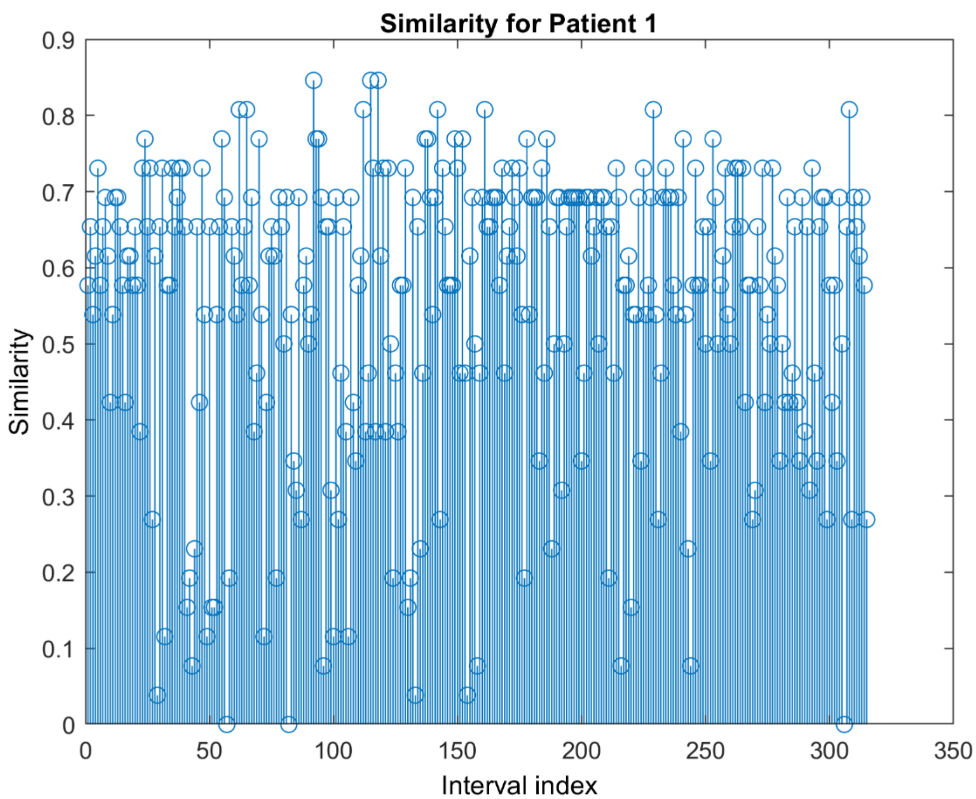

Figure 3-2: Similarity for patient 1. This patient showed a very high similarity in most of its intervals.

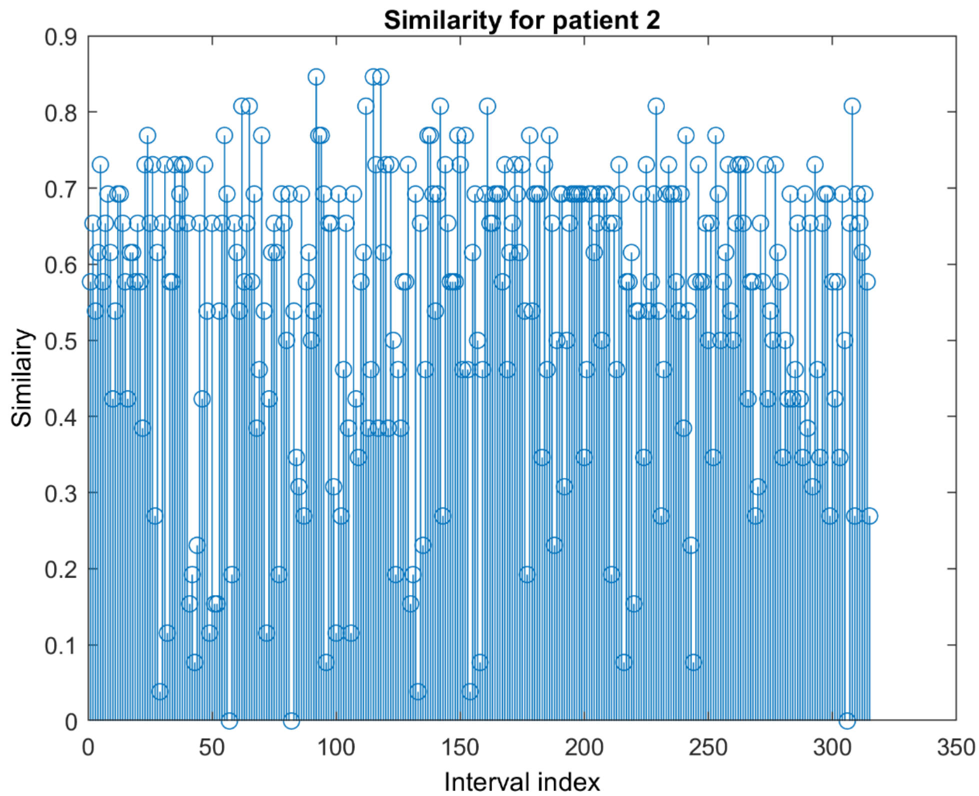

Figure 3-3: Similarity for patient 2. This patient showed a very high similarity in most of its intervals

The SVM classifier also demonstrates a high level of performance using the mRMR method. We note the LFS method can give us a “soft” output in terms of a similarity value of a patient to a healthy control, while the SVM/mRMR method only gives a categorical 0 – 1 output value.

3.2 TBI project: Treating ERP data as resting EEG

Different combinations of the features mentioned in the previous chapter in section 2.3.2, were used for feature selection. After multiple runs and closely observing the results, it was found that the combination of Coherence and Power Spectral Density (PSD) or a combination of Phase Locking Value (PLV) and PSD gives the lowest error rates. Coherence and PLV values result in very similar error rates. Fractal features were rarely selected by the feature selection algorithm and therefore got eliminated.

	ntff=64, cut=16, no fractal, selected electrodes only, Coherence and PSD
k	Accuracy (%)
5	60.6
15	**65.8**
25	63.4
35	64.6
45	63.8
55	63.8
65	64.8
75	61
85	59.4
95	62.4
Number of features prior to feature selection	43 X 10080

Figure 3-4: explored over different values of K (number of best features), from 5 to 95 using only Coherence and PSD.

	ntff=64, cut=16, no fractal, selected electrodes only, PLV and PSD
k	Accuracy (%)
5	61.3
15	59.9
25	62.1
35	62.2
45	63
55	**65.3**
65	60.6
75	63.5
85	63.3
95	64.1
Number of features prior to feature selection	43 X 1155

Figure 3-5: explored over different values of K (number of best features), from 5 to 95 using only PLV and PSD.

With regard to the low accuracy of the proposed method of 65.8%, and since the accuracy of discrimination between controls and athletes in the original study (i.e., which exploits the ERP structure in the data) is very high, it appears that our hypothesis of treating ERP data as resting data is not valid in this case. It appears we have lost a significant amount of important information by ignoring the ERP structure and the synchronization with the stimulus, making it difficult for the model to distinguish between the classes.

Chapter 4

4 Conclusions

4.1 Research summary

The main objective of this research was to apply different machine learning methods on brain signals for two EEG/ERP datasets. In the first (coma) dataset, our objective was to discriminate between standard and deviant responses in comatose patients. This is equivalent to testing whether or not they will emerge. With the second (TBI) dataset, our objective was to discriminate healthy subjects from those who have experienced TBI, while treating the data recorded from an ERP paradigm as resting state data.

With regard to the coma part of the study, we presented a machine learning approach for automatic and continuous assessment of ERPs for identifying the presence of the MMN component, which has a good correlation with coma awakening. Experimental results on normal and comatose subjects demonstrate the effectiveness of the proposed method. We had twenty-six subjects, two of which were comatose patients and the remaining subjects were healthy controls. In the training phase, we trained the two different feature selection methods, mRMR, and LFS, where SVM is used as the classifier with mRMR.

We trained the mRMR feature selection method using LOO cross-validation and achieved an accuracy of 86.5%. We used SVM with an RBF kernel as our classifier. In the testing phase, our model predicted both coma patient's class.

LFS has an associated localized classifier which allows incorporating multiple feature subsets when performing classification. We trained the method on healthy subjects evaluated using a Leave One Out (LOO) cross-validation technique and achieved an accuracy of 86.6%. We also applied the ML models trained on healthy subjects to two coma patients to examine the model's performance on prediction coma outcome. Both coma patients predicted emergence, correctly.

LFS and mRMR methods both represented high performance, but LFS's prediction is more reliable since it gives us a similarity measure of a test sample to each of the classes. Finding the similarity gives us a heads up about each patient's brain signal state compared to a healthy brain.

For the Traumatic Brain Injury study, the main goal was to find an automatic method, using supervised Machine Learning analysis of the electroencephalogram (EEG), to detect TBI in patients. In this part, ERPs are treated like *resting* EEG. This objective is hard to achieve using traditional methods, such as through responses to questions. For example, patients might be afraid, to tell the truth or they might be not sure about their condition and would not be able to express their situation accurately. Consequently, the diagnosis would be made on an incorrect basis. Since the proposed study is based on EEG analysis, the diagnosis is much more trustworthy, in comparison to previous methods which mostly rely on the patient's verbal responses. The data was collected from retired football players and healthy individuals. Each data record is labeled as either healthy or TBI. The objective of our ML process is to discriminate between these two classes.

With regard to the low accuracy of the proposed method of 65.8%, and since the accuracy of discrimination between controls and athletes in the original study (i.e., which exploits the ERP structure in the data) is very high, it appears that our hypothesis of treating ERP data as resting data is not valid in this case. It appears we have lost a significant amount of important information in the data by ignoring ERPs components, making it difficult for the model to distinguish between the classes.

Moreover, in the coma study, collecting new EEG data from comatose patients has been an important and challenging part to extend this study, but due to the pandemic of Covid-19, we had to stop all the process and use the previous dataset.

4.2 Future work

In this thesis, we proposed two different methods for feature selection and classification in the present context. One suggestion for future work is to extend our results to include a wider variety of machine learning methods. One example is the use of adaptive connectivity measures that can track responses throughout an ERP interval.

The dataset used for this research was small, so as future work we suggest extending the database to include significantly more DoC and comatose patients.

For the LFS method, we evaluated the similarity of the EEG data empirically. For future studies when more data becomes available, determining a suitable threshold, above which the patient is deemed to emerge, is a necessary step for the application of the proposed method in the clinic.

Bibliography

Al-Kandari, Noriah M, and Ian T Jolliffe. 2005. "Variable Selection and Interpretation in Correlation Principal Components." *Environmetrics* 16 (6): 659–72. https://doi.org/10.1002/env.728.

Anwar, S, T Batool, and M Majid. 2019. "Event Related Potential (ERP) Based Lie Detection Using a Wearable EEG Headset." In *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, 543–47. https://doi.org/10.1109/IBCAST.2019.8667131.

Armanfard, M Komeili, J P Reilly, and J F Connolly. 2019. "A Machine Learning Framework for Automatic and Continuous MMN Detection With Preliminary Results for Coma Outcome Prediction." *IEEE Journal of Biomedical and Health Informatics* 23 (4): 1794–1804. https://doi.org/10.1109/JBHI.2018.2877738.

Armanfard, Komeili, James P. Reilly, Richard Mah, and John F. Connolly. 2016. "Automatic and Continuous Assessment of ERPs for Mismatch Negativity Detection." In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2016-Octob:969–72. https://doi.org/10.1109/EMBC.2016.7590863.

Armanfard, N, M Komeili, J P Reilly, and L Pino. 2016. "Vigilance Lapse Identification Using Sparse EEG Electrode Arrays." In *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 1–4. https://doi.org/10.1109/CCECE.2016.7726846.

Armanfard, N, J P Reilly, and M Komeili. 2018. "Logistic Localized Modeling of the Sample Space for Feature Selection and Classification." *IEEE Transactions on Neural Networks and Learning Systems* 29 (5): 1396–1413. https://doi.org/10.1109/TNNLS.2017.2676101.

Armanfard, Narges. 2017. "Localized Feature Selection for Classicatione." McMAster University. https://macsphere.mcmaster.ca/bitstream/11375/20944/2/Armanfard_Narges_20171_PhD.pdf.

Arzucan ozgür, Arzucan ozgür Arzucan. 2002. "SUPERVISED AND UNSUPERVISED MACHINE LEARNING TECHNIQUES FOR TEXT DOCUMENT CATEGORIZATION."

Beres, Anna M. 2017. "Time Is of the Essence: A Review of Electroencephalography (EEG) and Event-Related Brain Potentials (ERPs) in Language Research." *Applied*

Psychophysiology and Biofeedback 42 (4): 247–55. https://doi.org/10.1007/s10484-017-9371-3.

Brunner, Clemens, Arnaud Delorme, and Scott Makeig. 2013. "EEGLAB-AN OPEN SOURCE MATLAB TOOLBOX FOR ELECTROPHYSIOLOGICAL RESEARCH." *Degruyter.Com*. https://doi.org/10.1515/bmt-2013-4182.

Cao, Bo, and Jim Reilly. 2019. "Major Challenges and Limitations of Big Data Analytics." In *Personalized Psychiatry: Big Data Analytics in Mental Health*, 15–36. Springer International Publishing. https://doi.org/10.1007/978-3-030-03553-2_2.

Connolly, John F., James P. Reilly, Alison Fox-Robichaud, Patrick Britz, Stefanie Blain-Moraes, Ranil Sonnadara, Cindy Hamielec, Adianes Herrera-Díaz, and Rober Boshra. 2019. "Development of a Point of Care System for Automated Coma Prognosis: A Prospective Cohort Study Protocol." *BMJ Open* 9 (7). https://doi.org/10.1136/bmjopen-2019-029621.

D'Urso, Pierpaolo, and Livia De Giovanni. 2018. "Unsupervised Learning," November, 1–23. https://doi.org/10.1002/047134608x.w8379.

Delorme, Arnaud, and Scott Makeig. 2004. "EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis." *Journal of Neuroscience Methods*. Vol. 134. http://www.sccn.ucsd.edu/eeglab/.

Dennis, E L, F Rashid, N Jahanshad, T Babikian, R Mink, C Babbitt, J Johnson, C C Giza, R F Asarnow, and P M Thompson. 2017. "A Network Approach to Examining Injury Severity in Pediatric TBI." In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 105–8. https://doi.org/10.1109/ISBI.2017.7950479.

Duda, Richard, Peter Hart, and David Stork. 2001. "Pattern Classification." In *Wiley Interscience*. Vol. xx.

Duncan, Connie C, Robert J Barry, John F Connolly, Catherine Fischer, Patricia T Michie, Risto Näätänen, John Polich, Ivar Reinvang, and Cyma Van Petten. 2009. "Event-Related Potentials in Clinical Research: Guidelines for Eliciting, Recording, and Quantifying Mismatch Negativity, P300, and N400." *Clinical Neurophysiology* 120 (11): 1883–1908. https://doi.org/https://doi.org/10.1016/j.clinph.2009.07.045.

Escera, Carles, Elena Yago, M.Dolores Polo, and Carles Grau. 2000. "The Individual Replicability of Mismatch Negativity at Short and Long Inter-Stimulus Intervals." *Clinical Neurophysiology* 111 (3): 546–51. https://doi.org/https://doi.org/10.1016/S1388-2457(99)00274-6.

Fischer, C, Frédéric Dailler, and Dominique Morlet. 2008. "Novelty P3 Elicited by the

Subject's Own Name in Comatose Patients." *Clinical Neurophysiology* 119 (10): 2224–30. https://doi.org/https://doi.org/10.1016/j.clinph.2008.03.035.

Fischer, C, D Morlet, P Bouchet, J Luaute, C Jourdan, and F Salord. 1999. "Mismatch Negativity and Late Auditory Evoked Potentials in Comatose Patients." *Clinical Neurophysiology* 110 (9): 1601–10. https://doi.org/https://doi.org/10.1016/S1388-2457(99)00131-5.

Freire, Fabio Rios, Fernanda Coelho, Juliana Rhein Lacerda, Marcio Fernando da Silva, Vanessa Tome Gonçalves, Sergio Machado, Bruna Velasques, et al. 2011. "Cognitive Rehabilitation Following Traumatic Brain Injury." *Dementia & Neuropsychologia* 5 (1): 17–25. https://doi.org/10.1590/S1980-57642011DN05010004.

Hanchuan Peng, Fuhui Long, and C Ding. 2005. "Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (8): 1226–38. https://doi.org/10.1109/TPAMI.2005.159.

Hastie, Trevor and Tibshirani, Robert and Friedman, Jerome. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science \& Business Media.

Haykin, S S. 2009. *Neural Networks and Learning Machines*. Neural Networks and Learning Machines. Prentice Hall. https://books.google.ca/books?id=K7P36lKzI_QC.

Hyvärinen, A, and Erkki Oja. 2000. "Oja, E.: Independent Component Analysis: Algorithms and Applications. Neural Networks 13(4-5), 411-430." *Neural Networks : The Official Journal of the International Neural Network Society* 13: 411–30. https://doi.org/10.1016/S0893-6080(00)00026-5.

Kane, Nicholas M and Butler, Stuart R and Simpson, Tom. 2000. "Coma Outcome Prediction Using Event-Related Potentials: P3 and Mismatch Negativity." *Audiology and Neurotology* 5: 186--191.

Khodayari-Rostamabad, Ahmad, James P Reilly, Gary M Hasey, Hubert de Bruin, and Duncan J MacCrimmon. 2013. "A Machine Learning Approach Using EEG Data to Predict Response to SSRI Treatment for Major Depressive Disorder." *Clinical Neurophysiology* 124 (10): 1975–85. https://doi.org/https://doi.org/10.1016/j.clinph.2013.04.010.

Light, Gregory A, Lisa E Williams, Falk Minow, Joyce Sprock, Anthony Rissling, Richard Sharp, Neal R Swerdlow, and David L Braff. 2010. "Electroencephalography (EEG) and Event-Related Potentials (ERPs) with Human Participants." *Current Protocols in Neuroscience* 52 (1): 6.25.1-6.25.24.

https://doi.org/10.1002/0471142301.ns0625s52.

Linden, Ariel. 2019. "LOOCLASS: Stata Module for Generating Classification Statistics of Leave-One-Out Cross-Validation for Binary Outcomes."

Morlet, Dominique, and Catherine Fischer. 2014. "MMN and Novelty P3 in Coma and Other Altered States of Consciousness: A Review." *Brain Topography* 27 (4): 467–79. https://doi.org/10.1007/s10548-013-0335-5.

Näätänen, R, A W K Gaillard, and S Mäntysalo. 1978. "Early Selective-Attention Effect on Evoked Potential Reinterpreted." *Acta Psychologica* 42 (4): 313–29. https://doi.org/https://doi.org/10.1016/0001-6918(78)90006-9.

Näätänen, R, P Paavilainen, T Rinne, and K Alho. 2007. "The Mismatch Negativity (MMN) in Basic Research of Central Auditory Processing: A Review." *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology* 118 (12): 2544–90. https://doi.org/10.1016/j.clinph.2007.04.026.

Näätänen, Risto. 2001. "The Perception of Speech Sounds by the Human Brain as Reflected by the Mismatch Negativity (MMN) and Its Magnetic Equivalent (MMNm)." *Psychophysiology* 38 (1): 1–21. https://doi.org/10.1111/1469-8986.3810001.

Näätänen, Risto, Thomas Jacobsen, and István Winkler. 2005. "Memory-Based or Afferent Processes in Mismatch Negativity (MMN): A Review of the Evidence." *Psychophysiology* 42 (1): 25–32. https://doi.org/10.1111/j.1469-8986.2005.00256.x.

Orme-Johnson, David W, and Christopher T Haynes. 1981. "EEG Phase Coherence, Pure Consciousness, Creativity, and Tm—Sidhi Experiences." *International Journal of Neuroscience* 13 (4): 211–17. https://doi.org/10.3109/00207458108985804.

Perel, Pablo, Miguel Arango, Tim Clayton, Phil Edwards, Edward Komolafe, Stuart Poccock, Ian Roberts, Haleema Shakur, Ewout Steyerberg, and Surakrant Yutthakasemsunt. 2008. "Predicting Outcome after Traumatic Brain Injury: Practical Prognostic Models Based on Large Cohort of International Patients." *BMJ (Clinical Research Ed.)* 336 (7641): 425–29. https://doi.org/10.1136/bmj.39461.643438.25.

Prince, Carolyn, and Maya E Bruhns. 2017. "Evaluation and Treatment of Mild Traumatic Brain Injury: The Role of Neuropsychology." *Brain Sciences* 7 (8). https://doi.org/10.3390/brainsci7080105.

Purbhoo, K K. 2018. "Severe Traumatic Brain Injury." *Southern African Journal of Anaesthesia and Analgesia* 24: S13–20.

Ravan, M, D MacCrimmon, G Hasey, J P Reilly, and A Khodayari-Rostamabad. 2012. "A Machine Learning Approach Using P300 Responses to Investigate Effect of

Clozapine Therapy.” In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 5911–14. https://doi.org/10.1109/EMBC.2012.6347339.

Ruiter, Kyle I, Rober Boshra, Mitchell Doughty, Michael Noseworthy, and John F Connolly. 2019. “Disruption of Function: Neurophysiological Markers of Cognitive Deficits in Retired Football Players.” *Clinical Neurophysiology* 130 (1): 111–21. https://doi.org/https://doi.org/10.1016/j.clinph.2018.10.013.

Rumelhart, D, Geoffrey E Hinton, and R J Williams. 1986. “Learning Internal Representations by Error Propagation.” In .

Sculthorpe-Petley, Lauren, Careesa Liu, Sujoy Ghosh Hajra, Hossein Parvar, Jason Satel, Thomas P Trappenberg, Rober Boshra, and Ryan C N D’Arcy. 2015. “A Rapid Event-Related Potential (ERP) Method for Point-of-Care Evaluation of Brain Function: Development of the Halifax Consciousness Scanner.” *Journal of Neuroscience Methods* 245: 64–72. https://doi.org/https://doi.org/10.1016/j.jneumeth.2015.02.008.

Syed, A T, N A Lone, M Afzal Wani, and A S Bhat. 2007. “Clinical Management of Patients with Minor Head Injuries.” *International Journal of Health Sciences* 1 (1): 131–40.

Todd, Juanita, Patricia T Michie, Ulrich Schall, Frini Karayanidis, Hirooki Yabe, and Risto Näätänen. 2008. “Deviant Matters: Duration, Frequency, and Intensity Deviants Reveal Different Patterns of Mismatch Negativity Reduction in Early and Late Schizophrenia.” *Biological Psychiatry* 63 (1): 58–64. https://doi.org/https://doi.org/10.1016/j.biopsych.2007.02.016.

Torsvall, Lars, and Torbjörn Akerstedt. 1987. “Sleepiness on the Job: Continuously Measured EEG Changes in Train Driversfile:///C:/Users/Armanfaf/OneDrive - McMaster University/COMA/Papers/Citations/Citation-325903978.Bib.” *Electroencephalography and Clinical Neurophysiology* 66 (6): 502–11. https://doi.org/https://doi.org/10.1016/0013-4694(87)90096-4.

Trautmann, Sebastian, Jürgen Rehm, and Hans-Ulrich Wittchen. 2016. “The Economic Costs of Mental Disorders.” *EMBO Reports* 17 (9): 1245–49. https://doi.org/10.15252/embr.201642951.

Vigo, Daniel, Graham Thornicroft, and Rifat Atun. 2016. “Estimating the True Global Burden of Mental Illness.” *The Lancet Psychiatry* 3 (2): 171–78. https://doi.org/10.1016/S2215-0366(15)00505-2.

Wang, J, P Zhao, S C H Hoi, and R Jin. 2014. “Online Feature Selection and Its Applications.” *IEEE Transactions on Knowledge and Data Engineering* 26 (3): 698–710. https://doi.org/10.1109/TKDE.2013.32.

Whiteford, Harvey A., Louisa Degenhardt, Jürgen Rehm, Amanda J. Baxter, Alize J. Ferrari, Holly E. Erskine, Fiona J. Charlson, et al. 2013. "Global Burden of Disease Attributable to Mental and Substance Use Disorders: Findings from the Global Burden of Disease Study 2010." *The Lancet* 382 (9904): 1575–86. https://doi.org/10.1016/S0140-6736(13)61611-6.

Wu, X, K Yu, W Ding, H Wang, and X Zhu. 2013. "Online Feature Selection with Streaming Features." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (5): 1178–92. https://doi.org/10.1109/TPAMI.2012.197.

Young, G, A Ropper, and C Bolton. 1997. "Coma and Impaired Consciousness: A Clinical Perspective.".

Yu, K, X Wu, W Ding, and J Pei. 2014. "Towards Scalable and Accurate Online Feature Selection for Big Data." In *2014 IEEE International Conference on Data Mining*, 660–69. https://doi.org/10.1109/ICDM.2014.63.

Zhu, Xiaojin and Goldberg, Andrew B. 2009. "Introduction to Semi-Supervised Learning (Synthesis Lectures on Artificial Intelligence and Machine Learning." *Morgan and Claypool Publishers* 14 (June): 1–116. https://doi.org/10.2200/S00196ED1V01Y200906AIM006.