Università degli Studi di Padova

Department of Information Engineering

Master Thesis in Bioengineering

Feature extraction and classification of Electrooculography signals from Locked-In patients

Supervisor
Alessandra Bertoldo
Università di Padova

Co-supervisor
Niels Birbaumer
Ujwal Chaudhary

Master Candidate
Giovanni Zanella

Date
09/12/2019

Dedicated to family and friends who helped and supported me

Abstract

The ability to communicate without using speech or hand gestures poses a great improvement in the quality of life of patients that suffer from movement impairment. Human-machine interaction tools are being studied and developed in order to optimize the usage of biological signals that survive the individual's disease. Among different approaches, Electrooculography signals are an alternative for those who still have a residual eye control capability. An interface was designed to record, process and classify EOG activity. Four ALS patients were recorded while controlling the interface for yes/no answering. This work proposes the use of different features extraction techniques for classification of EOG signal. The challenge stems from the need to extract spatial and temporal patterns from noisy multidimensional time series obtained from patients with different clinical conditions. Time and frequency domain feature extraction methods are proposed by the use of Discrete Cosine Transform, Autoencoder and Complex Wavelet Transform. The different algorithms allowed to reach an average classification accuracy up to 75 %, 92 %, 94 % and 98 % in the four different patients.

Preface

This dissertation is the result of a study conducted at the Institute of Medical Psychology and Behavioral Neurobiology of Tübingen, where I had the privilege to join the team led by Niels Birbaumer and Ujwal Chaudhary. With this experience I had the opportunity to actively contribute to the study and development of EEG, EOG, fNIRS based human-machine interfaces for patients affected by ALS.

Contents

Listing of figures

Listing of tables

1 Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive disease of the lower and often upper motor neurons that usually leads to complete paralysis within 2–5 years [1]. Each year, approximately 5600 people in the United States are diagnosed with ALS at a mean age of 55 years. About 30,000 people are living with ALS in the United States alone (ALS Association, 2013). As the disease progresses, individuals may use their remaining muscle control to operate a variety of assistive communication devices for interacting with the world [2]. However, as their disease progresses these devices may become ineffective. Human-Machine Interface (HMI) technology can allow people with severe motor disabilities (i.e., ALS) to use alternative tools, rather than voice or muscles, to communicate and to control their environments [3][4].

The term 'locked-in syndrome' (LIS) was first introduced by Plum and Posner[5] [6]. The term denotes a neurological condition consisting of tetraplegia and paralysis of all cranial nerves except voluntary eye movements[7]. Consciousness is fully preserved and can be demonstrated by voluntary blinking. The patients are able to communicate complex ideas for instance by blinking Morse code [5].

Patients with a neurodegenerative disease like ALS, lose the voluntary control of their muscles and transit to LIS as time passes. When the patients lose control of their eye movement, which is usually the last active muscle [8], the patient is in the completely locked-in state (CLIS) [9]. The patients in CLIS are then left without any means of communication[7].

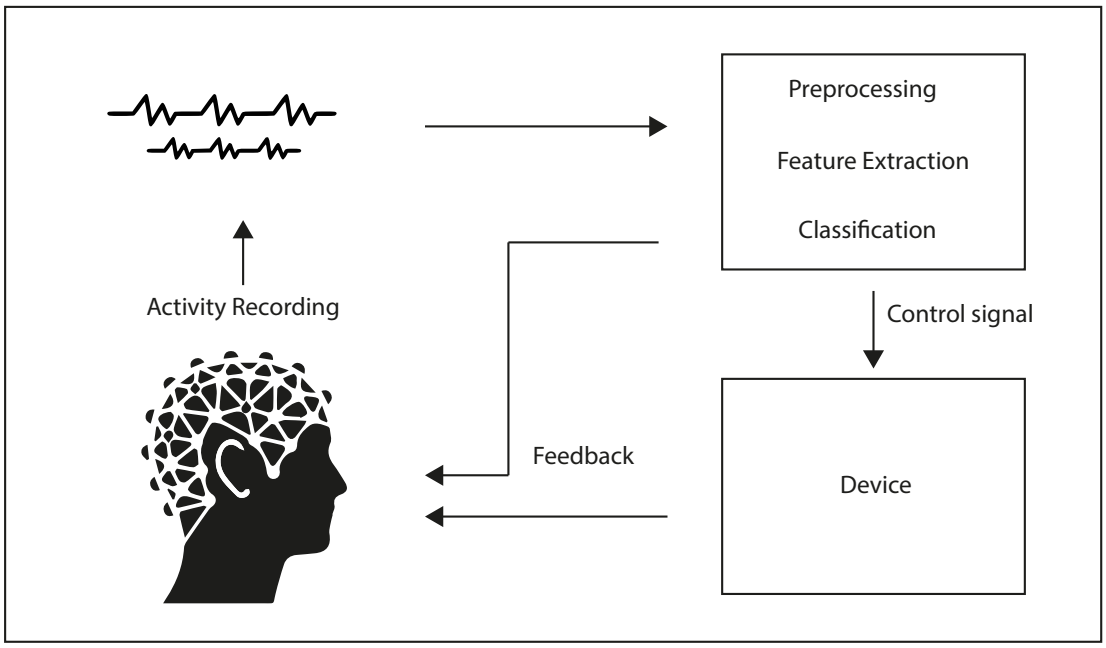

Figure 1.1: High-level structure of a Human-Machine Interface

Many interfaces have been designed and successfully used in the last years using different strategies. For instance, the use of invasive BCIs involve surgical implantation of electrodes or multi-electrode grids. They measure activity patterns of neurons, which encode behaviourally relevant information. Different types of brain activity are measured with invasive BCIs such as local field potentials, multi-unit activity, and calcium channel permeability [10]. Even if having many advantages compared to other non-invasive techniques, they carry all the drawbacks and costs of brain surgery. Other non-invasive BCI techniques have been developed. For instance, in [4], functional near-infrared spectroscopy (fNIRS) was used to

measure and classify cortical oxygenation and deoxygenation after the presentation of a question, resulting in satisfactory classification accuracy for the patients' answers. Another valid strategy is the use of EEG signals; using different paradigms as motor imagery [11] or inner speech [12] is possible to restore the communications with the patients, however both fNIRS and EEG set up require a time consuming set up and return a much more complex signal compared to eye-movement based techniques. The most used communication systems for ALS patients are in fact eye tracker based. The problem with this communication technology is that it primarily relies on the gaze fixation ability of the patients. In the transition from LIS to CLIS, this capability is lost by the patients, forcing the caregivers to abandon the communication. In this study, an electrooculography (EOG) based interface will be presented as it can overcome some of the previously described shortcomings. As will be discussed in the following chapters, an ALS patient in transition from LIS to CLIS is still able to perform voluntary eye movements and EOG signal can be exploited to detect it and to control an interface. A deeper analysis of this technology will be proposed in the next chapters.

EOG based HMIs have been object of study in the last decades for different applications. In [13] for instance, a computer input device was developed. The device acquires EOG signals generated by eye movements and blinks and produces output for emulating the PC mouse. From the acquired signals is possible to estimate the direction and amplitude of eye movement and detect blinks using a microcontroller. In [14] and [15]the EOG signal is used to control an electric wheelchair. In the latter application, the wheelchair dynamic is controlled by constantly estimating the eyes rotation, modeling the process as a transition of states. In [16] and [17] using two different strategies, the EOG signal is used as the mean to control a speller. The promising results from Tonin et al. confirmed how it is possible to use EOG based control algorithm to restore communication with ALS patients.

All the previously mentioned possible applications of EOG based HMI (as many others in literature) have a similar pattern in the signal processing pipeline. From the preprocessed signal a set of features are extracted with the goal of representing in a compact set the useful information produced by the eye movement. This process is the core strategy of many classification pipelines, in the next chapters will be analyzed different methods and their results.

A work with an interesting perspective about feature extraction optimization has been done in [18]. Features are extracted and selected through a scheme called 'Genetic Programming'. It consists in generating and combining ensembles of features that can maximize the so called 'fitness function'. In this iterative heuristic procedure, every combination of features competes with the others and just the fittest are selected. The algorithm is stopped when a threshold value is achieved by the fitness function or when the maximum number of iteration is reached. This method represents an attempt of optimality criteria in the designing of the feature set.

Another endeavor towards the removing of subjectivity in the feature extraction process is the usage of neural network algorithms for classification. In [19] for example, the EOG signal is directly used to feed the input layer, while the output layer is the class membership. In this way, the feature extraction step is embedded in the network and the features are represented by the neurons. In this structure, the subjectivity of the results due to the ability of the experimenter to describe the signal by the feature set is heavily reduced. The back propagation step allows replacing empirical research with a gradient descent, optimizing classification accuracy.

According to this philosophy, in this study will be proposed three data driven approaches for the feature extraction of EOG signals from LIS patients. In the process of signals classification, is important to have an a-priori knowledge of the explored phenomena, but hand

selecting the features that should discriminate the class membership might lead to loss of useful information. The proposed methods exploit advanced signal processing techniques whose goal is retaining the maximum amount of information from the signal space to the feature set. The effort made in this study has been to combine the two approaches.

In chapter 2 will be presented the methods and the instrumentation used retrieving the data from the patients. In chapter 3 will be proposed an overview of the EOG signal in the specific case of the LIS patients. In chapter 4 three different methods for feature extraction will be proposed. In chapter 5 will be presented the results and a comparison of the different techniques. And finally in chapter 6 will be proposed an analysis of the study highlighting the potentialities of the discussed techniques and the future works.

2

Methods & instrumentation

In this study four lateral sclerosis (ALS) patients are analyzed. The data used from these subjects is retrieved in a period of transition from a locked-in-state (LIS) (with the ability to control some muscle of the body) to completely-locked-in-state (CLIS) (without the possibility to control any muscle in the body). With the progression of the disease the patients gradually lose the possibility to control eye-tracker devices and are left without any mean of communication. An EOG based interface was developed to allow patients to answer questions controlling the interface through eye movements. The majority of eye-tracker based communication systems, strongly depend on the gaze fixation ability of the patient and rely on a sharp control of eye movements. The aggravation of ALS leads to a progressive decreasing of quality of vision and an advancing paralysis of eye and lid muscles[20]. As discussed in [21] and [4] an auditory based interface can be a solution to tackle the aforementioned problems.

For data collection, EEG+EOG channels were recorded with a 16 channel EEG amplifier (V-Amp DC, Brain Products, Germany) with Ag/AgCl active electrodes. A total of 7 EEG electrodes (with locations C1, Cz, C2, and FC3, F3, F4, FC4 and F4, depending on experimental conditions) and 4 EOG (electrooculogram channel placed above the eyebrow of the right eye (EOGU) , electrooculogram channel placed below the right eye (EOGD), electrooculogram channel placed on the side of the right eye (EOGR) and electrooculogram channel placed on the side of the left eye (EOGL) electrodes. All the channels were referenced to an electrode on the right mastoid and grounded to the electrode placed at FPz. For the montage, electrode impedance is kept below 10 kΩ. Frequency sampling was 500 Hz.

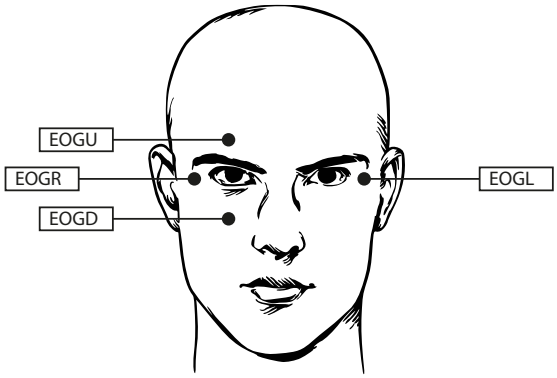

Figure 2.1: EOG Electrode Placement

Before the recording of the data, more than 200 questions were provided from the patients' families. The contents of the questions cover trivial information that is known to patients and families. For instance: 'Is Berlin the capital of Germany?', or 'Were you born in 1969?' For every known question with a 'yes' answer a semantically identical question which requests a 'no' answer was constructed and vice versa, in order to maintain a balanced class

representation. Each recording session started with a 'Training block ' of 20 questions presentation, 10 of which had positive answers and the other 10 negative answers. The yes-no questions were presented randomly. The training blocks were repeated until a satisfactory model was built in the software classification (usually 75% or more of prediction accuracy). To clarify the data retrieval structure and the terminology in Fig 2.2 there is a summarizing scheme of the visits procedure.

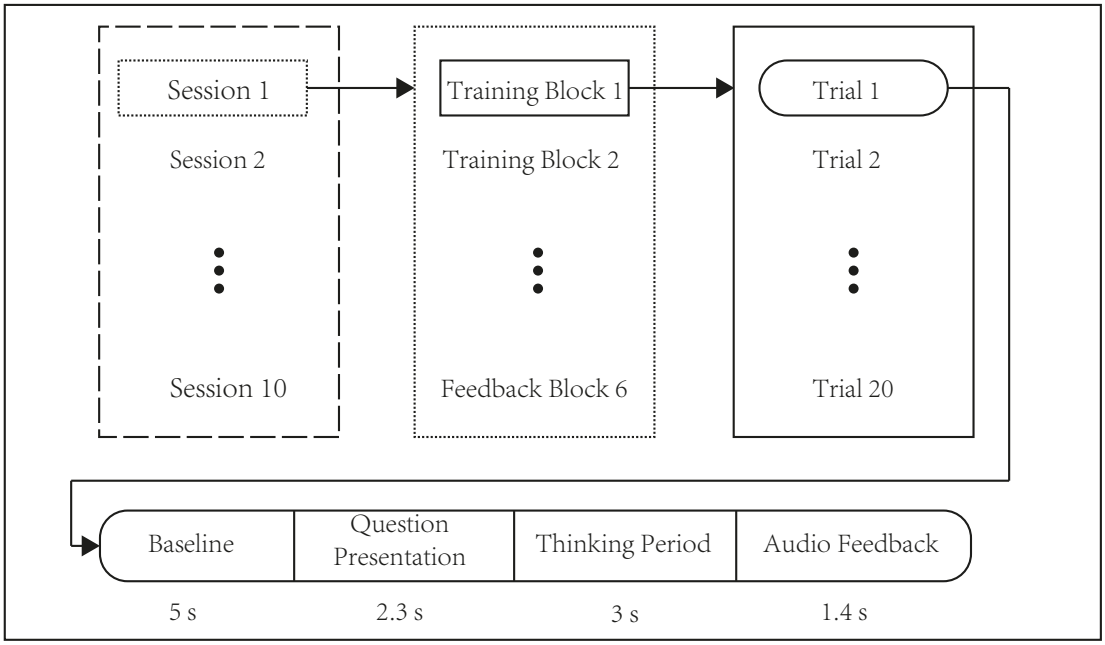

Figure 2.2: Paradigm & structure of the retrieved data

After the achievement of an above chance classification 'Feedback Blocks' can be performed. Each block is composed of 20 'Trials', each of them consisting in the following paradigm. The first phase of resting state is followed by a question presentation, once terminated the patient is instructed to move the eyes for answering yes and to do nothing for answering no (thinking period), and the last seconds are dedicated to providing a feedback to the user to let him know the finish of the trial. The difference between the two blocks types relies on

the interface response to the patient. During training blocks, the auditory feedback given to the user is simply 'Danke', the German word for 'Thank you', while in the feedback block it would be 'Deine Antwort wurde als Ja/Nein erkannt' ('Your answer was recognized as yes/no'). The signal processing for the classification will be discussed in the next sections, but in order to understand the data retrieval process is important to state that the classification algorithms used in this study have to be trained in a supervised fashion framework. This technique allows achieving better classification accuracy than the unsupervised approach but as a drawback needs training sets in order to build a model for the classification. For this reason, performing training blocks is necessary before moving to feedback blocks, in this way is possible to minimize boredom and frustration in the user trying to control the interface. During the feedback block, the patients can actually learn how to control the interface finding the most suitable technique for eye movement as a compromise between good classification accuracy and personal effort. Simple tasks as eye movement can result in fatigue and sometimes exhaustion for ALS patients, and this is important while evaluating the classification performance as it indicates how the classification outcomes are not fully determined by the experiment efficacy, but they also have a certain grade of uncertainty rooted in the stage of the disease and patient's motivation.

2.2 The patients

Four patients are analyzed in this study, but even sharing the locked-in-state there are important differences that are worth to be mentioned, as they are reflected in EOG signals. Here follows a brief summary of each patient most salient clinical traits.

2.2.1 Patient 11

Male, 33 years old, was diagnosed with not-bulbar ALS in August 2015. He lost speech and capability to walk by the end of 2015. He has been fed through a percutaneous endoscopic gastrostomy tube and artificially ventilated since July 2016 and is in home care. He started using an eye-tracking-based communication system in August 2016 for communication. From August 2017 onwards, he could not use the eye-tracker for communication because of his inability to fixate his gaze, after that the family developed its own spelling system to communicate with the patient by observing the eye movements, which did not require the patient to fixate his gaze. Any visible eye movement was identified as a "yes" response, no movements as "no". Detection of eye-movements by relatives became increasingly difficult and errors made communication attempts virtually impossible up to the point where communication had to be abandoned. The follow up time of Patient 11 (P11) spans from March 2018 to March 2019

2.2.2 Patient 13

Male, 59 years old in June of 2018, was diagnosed with bulbar ALS at Jan of 2011. He lost speech and capability to walk by the end of 2011. He has been fed through a percutaneous endoscopic gastrostomy tube since July 2011, artificially ventilated since December 2011, and is in home care. He started using an eye-tracking based communication system for communication in February 2012 and used it until April 2017. From May 2017 onwards when he could not use the eye-tracking system for communication the family and caretakers communicated reliably with him based on his eye movements (looking straight means "no", looking to right means "yes"). The fidelity of the answer of the patient depends on the experience of the caretaker to decipher the patient's eye movement. The follow up time of Patient 13 (P13)

spans from June 2018 to May 2019

2.2.3 Patient 15

Female, 63 years old, was diagnosed with lower motor neuron predominant-ALS in February 2017. She has been fed through a percutaneous endoscopic gastrostomy tube since May 2017, artificially ventilated since May 2017, and is in a shared caring community. She started using an eye-tracking based communication system in April 2018. Since November 2018 the family and caretakers communicated with her relying on her eye movements (looking to the right means ”no”, looking to the left means ”yes”) because of her inability to calibrate the eye-tracker. The patient was visited for the first time in February 2019. The visits of Patient 15 (P15) were performed in the month of February 2019

2.2.4 Patient 16

Male, 56 years old in Feb 2019, was diagnosed with lower motor neuron ALS at the end of 2012. He has been fed through a percutaneous endoscopic gastrostomy tube since June 2015, artificially ventilated since June 2015, and is in home care. He started using an eye-tracking based assistive communication device in 2017. Since June 2018 he has not been able to reliably communicate using the eye-tracker. The family and caretakers communicate with him since June 2018 based on his eye movements (looking straight means ”no”, looking to the right means ”yes”). The follow up time of Patient 16 (P16) spans from February to May 2019

3

Signals and classification

Electrooculography (EOG) is a method for sensing eye movement and is based on recording the standing corneal-retinal potential arising from hyperpolarizations and depolarizations existing between the cornea and the retina; this is commonly known as an electrooculogram. This potential can be considered as a steady electrical dipole with a negative pole at the fundus and a positive pole at the cornea [22]. The standing potential in the eye can thus be estimated by measuring the voltage induced across a system of electrodes placed around the eyes as the eye gaze changes, thus obtaining the EOG. The EOG value varies from 50 to 3500 μV with a frequency range of about dc-100 Hz. Its behavior is practically linear for gaze angles of $\pm 30^{o}$ and changes approximately 20 microvolts for each degree of eye movement. The variability of the electrooculogram reading depends on many factors that are difficult to determine: perturbations caused by other biopotentials such as EEG (electroencephalogram), EMG (electromyogram); moreover has to be considered the variability introduced by the acquisition system, plus the one due to positioning of the electrodes, skin-electrode contacts, lighting conditions, head movements, blinking, etc [23].

Another factor that is important to consider, is that the ability to control eye movement varies from patient to patient and changes with the progress of ALS. In Fig 3.1 and 3.2 is possible to have an insight into the strategy adopted by each different patient. Both figures represent the differential channel (EOGL - EOGR). All the experimenters were instructed to move the eyes in any direction to answer 'yes' and not to move them to say 'no'. However, they were also encouraged to find themselves the most suitable strategy during the feedback sessions. Is interesting to notice how the differences of patients' conditions are reflected in the EOG signals.

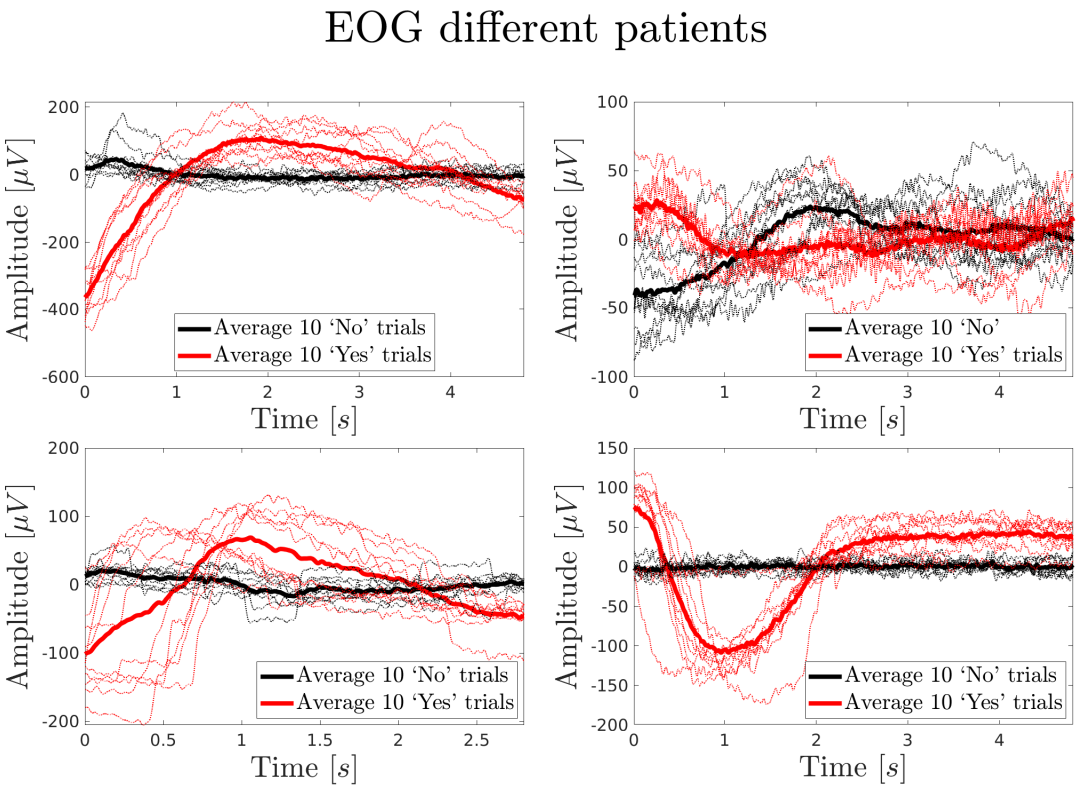

Figure 3.1: P11 (top-left), P13 (top-right), P15 (bottom-left), P16 (bottom right)

From these figures is possible to observe how the patients' response can anticipate the thinking period trigger during data acquisition. With this windowing, an important part of

the signal information is lost as it is directly discarded. In order to try to maximize classification accuracy and for the sake of integrity in the signal might be useful considering also part of the signal acquired during the question presentation. During data retrieval, the training and feedback blocks were performed using the hard windowing just described. Even if this strategy might result fallacious, in this study is kept for two reasons. The first is for consistency of the data acquisition process: during the feedback blocks the patients learn how to control that particular interface, giving the feedback to the patients re-basing the signal processing on a different time window might lead to different results. The second reason regards the first steps of this study, that started with a qualitative comparison of proposed methods with the online results in classification accuracy. In this phase has been important working on the same reference signals. The objective of the analysis proposed in the next chapters is not focused on the maximization of results, but is a comparison between different feature extraction methods. In this sense, while the different techniques are extracting features from the same chunk of data, the comparison is not corrupted. In the next sections, all the figure and results will be obtained using this standard.

3.0.1 Patient 16

Patient 16 dataset consists of two visits, with a total of 20 between training and feedback blocks. As will be discussed in the following sections, the performance in terms of classification is the best between the group and it reflects in a clearer EOG signal. Even having lost the gaze fixation ability, the eye movement control is sufficient for a correct classification during the follow up period.

3.0.2 Patient 15

Patient 15 dataset consists of two visits, with a total of 20 between training and feedback blocks. The performance in term of classification is comparable with the ones of P16. Even having lost the gaze fixation ability, the eye movement control is sufficient for a correct classification during the follow up period.

3.0.3 Patient 13

Patient 13 dataset consists of four visits, with a total of 40 between training and feedback blocks. One peculiarity of P13 is that in the very early session he decided to move the eyes on one side for 'yes' and to the other for 'no'.

3.0.4 Patient 11

Patient 11 dataset consists of ten visits from March 2018 to March 2019 with a total of 79 between training and feedback blocks. His case probably represents the most interesting study because of the longest follow up where is possible to understand how the ALS course is affecting the control on eye movement. Fig 3.2

An overview of the dataset structure can be found in Table A.1

The following sections will be dedicated to the data processing: from raw signal to classification. In this chapter, in particular, will be taken into consideration one specific feature extraction strategy as reference for a deeper analysis in the following chapters.

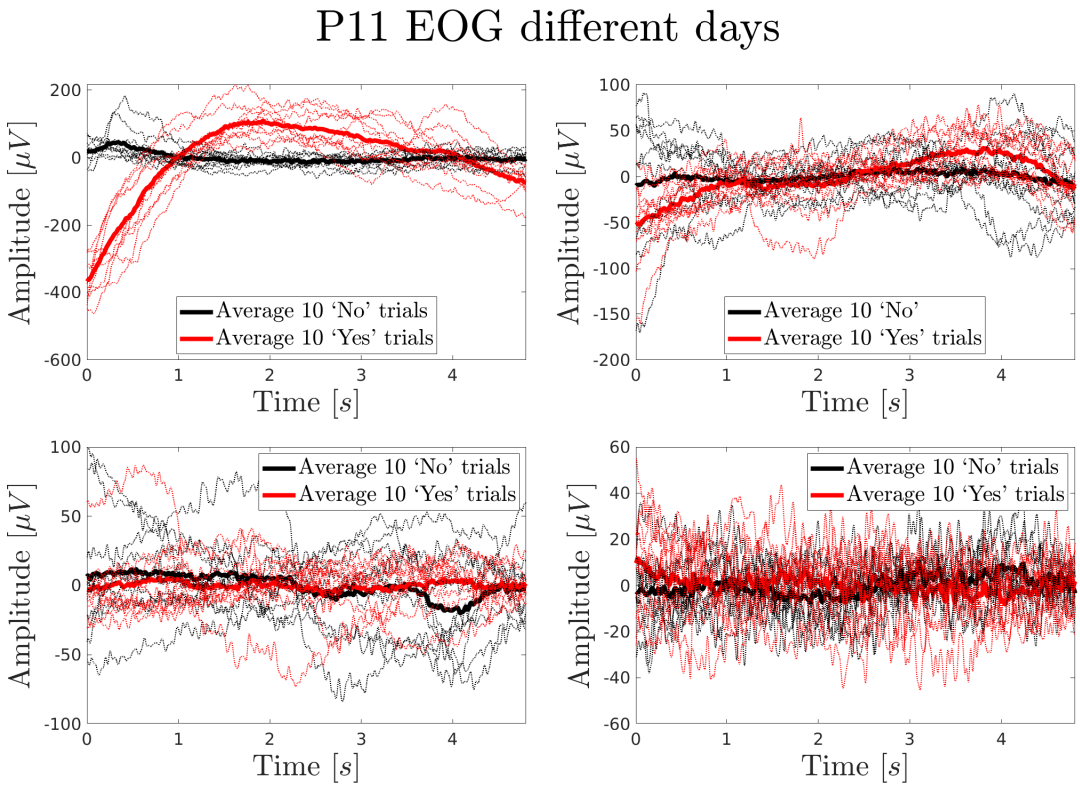

Figure 3.2: March 2018 (top-left), May 2018 (top-right), September 2018 (bottom-left), March 2019 (bottom-right)

3.1 Preprocessing & feature extraction

The first proposed feature extraction method relies on the idea that is possible to estimate the EOG general trend during a thinking period by extrapolating from each channel the maximum and minimum values and their respective time location. It is a simple yet effective time domain analysis that provides a compact and intuitive feature coefficients set that can discriminate eye movements period from resting states[17]. From each trial, the EOG signal is first divided into the different subsections keeping the structure shown in Fig 2.2. The following step is to apply a zero phase-shift digital filter consisting of an Infinite Impulse Response passband Butterworth. Similarly to what is done in [24], the selected frequency

band is $0.5 - 30$ Hz. Once cleared the signal from high frequency noise, four feature functions are applied channel-wise to each thinking period. This strategy is hereafter referred to as MinMax.

- *Maximum value:* Returns the maximum value of the amplitude of the signal in the time domain.
- *Minimum value:* Returns the minimum value of the amplitude of the signal in the time domain.
- *Maximum Location:* Returns the time location relative to the maximum amplitude recorded value.
- *Minimum Location:* Returns the time location relative to the minimum amplitude recorded value.

This feature extraction method has several advantages: the computational burden is very lightweight, it is easy to be implemented, and returns a set of values of immediate comprehension. As presented in the results section, it shows great reliability in healthy patients and doesn't need any tuning of complex parameters. The main drawback is that trying to represent each channel with just four salient coefficients there is an uncontrolled loss of information from the time domain signal to the feature set. In the results section, will be explained how this is not critical for patients with high responsiveness, but might be crucial in the late stages of ALS.

3.2 Feature selection

Generalizing the step of feature extraction is important to state that some methods could return a high dimension feature set. In this case, before classification, some feature extraction methods can be taken into consideration in order to improve the classifier performance [25]. Using techniques of dimensionality reduction can be used for the following reasons:

– *Reducing computational burden:* With the perspective of using a supervised classifier, reducing the number of features can lead to faster training and predictions.

– *Reducing overfitting:* High dimension feature sets usually reflect in complex models built in the training step. This downside might reduce the generalization properties of the classifier, with high training accuracy and poor prediction performance.

– *Discarding highly correlated features:* Some features can be highly dependent one another. In this case they should be removed before classification, because useless and possibly biasing the results.

– *Discarding noise:* Some features can be both very sensitive to signal noise, or irrelevant regarding the investigated phenomena. In both cases, they should be removed as adding unnecessary complexity to the classifier model leading to worse classification performance.

In this study, the proposed feature selection method follows what is called 'minimal redundancy maximal relevance criterion' (mRMR) [26]. In terms of mutual information, the purpose of feature selection is to find a feature set S with m features x_i, which jointly have the largest dependency on the target class c. This scheme, called Max-Dependency, has the following form:

$$maxD(S,c), \quad D = I(\{x_i, i = 1, ...m\}; c) \tag{3.1}$$

The Max-Relevance criterion is to search features satisfying the following equation, which approximates $D(S,c)$ in 3.1 with the mean value of all mutual information values between individual feature x_i and class c:

$$maxD(S,c), \quad D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) \tag{3.2}$$

It is likely that features selected according to Max-Relevance could have rich redundancy. When two features highly depend on each other, the respective class-discriminative power

would not change much if one of them were removed. Therefore, the following minimal redundancy (Min-Redundancy) condition can be added to select mutually exclusive features

$$minR(S), \quad R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \tag{3.3}$$

The criterion combining the above two constraints is called "minimal-redundancy-maximal-relevance" (mRMR). In [27] are proposed two different strategies to define it :

$$\Phi_{MID} = D - R \tag{3.4}$$

$$\Phi_{MIQ} = D/R \tag{3.5}$$

These different schemes are respectively called 'Mutual Information Difference' and 'Mutual Information Quotient'. For a deep characterization and differences between the two strategies is possible to find an exhaustive work in [27]. In this study, the strategy adopted is the one in equation 3.5.

3.3 Classification

After feature selection, the following critical step is the choice of the classification strategy. Depending on the preprocessing, the feature extraction, and feature selection methods, different classifiers could be more suitable than others. In this study, three different classifiers are taken into consideration in order to evaluate the performance of the different feature extraction technique proposed.

3.3.1 Support Vector Machine

The features of 'Support Vector Machine' (SVM) make it one of the most used tools for binary classification of EEG/EOG signal and suitable for this study [28]. The SVM classifies data by finding the *best* hyperplane that separates data points of one class from those of the other class [29]. Some binary classification problems do not have a simple hyperplane as a useful separating criterion. For those problems, there is a variant of the mathematical approach that is based on a non-linear transformation kernel and retains nearly all the simplicity of an SVM separating hyperplane. It is an easy yet effective tool that can be used to discriminate the state of an EOG signal from a set of extracted features [30].

3.3.2 Decision Tree

The second classifier taken into consideration is based on a 'Decision Tree' strategy. The goal of a Decision Tree is to create a model that predicts the value of a target variable based on several input variables [31]. A tree is built by splitting the source set, constituting the root node of the tree, into subsets which constitute the successor children. The splitting is based on a set of splitting rules based on classification features. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node has all the same values of the target variable, or when splitting no longer adds value to the predictions. This process of top-down induction of decision trees is one of the most common and widely used for its simplicity and almost 'out-of-the-box' usage [32]. Individual decision trees tend to overfit, and one possible enhancement here considered is the Bootstrap-aggregated (bagged) strategy. Bootstrap-aggregated decision trees combine the results of many decision trees, which reduces the effects of overfitting and improves generalization. Bagged Trees grow the decision trees in the ensemble using bootstrap

samples of the data [33].

3.3.3 K-Nearest-Neighbors

The third classifier taken into consideration is the 'K-Nearest-Neighbors' algorithm (K-NN). In K-NN classification, the input consists of the k closest training examples in the feature space (the training set), the output is a class membership. An element is classified by a plurality vote of its neighbors, with the element being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small) [34].

4

Automatic feature extraction

As discussed in the previous sections, one of the core challenges in the EEG/EOG signal classification is the feature extraction. In this chapter will be proposed three techniques of feature extraction exploiting data compression algorithms. A substantial part of existing projects involving EOG classification focus on the design of functions that try extrapolate some features that could help to discriminate the class of the signal [35] [36]. This strategy has many benefits: each feature function (depending on its simplicity) can have a straightforward meaning, informative on the signal nature, moreover feature functions usually have a low computational burden and don't need any training dataset. However using this approach, from the signal to the extracted feature, there is an uncontrolled loss of information, and the effectiveness of the process depends by an a-priori knowledge on the nature of the signal by the feature functions designer. The dimensionality reduction from the original signal space to the feature space can be considered as an attempt to encode the signal characteristics, to feed a classifier with a compressed version of the signal. In this sense, the next sections will be focused on exploring the application of well-known data compression algorithm for an au-

tomatic feature extraction. The first technique will be based on the 'Discrete Cosine Transform ' (DCT), the second will be based on an Autoencoder (AE) framework and the third on 'Discrete Wavelet Transform' (DWT).

4.1 Automatic feature extraction for eye movement in ALS patients

The transition from LIS to CLIS brings to the loss of control of the eye movements reflecting in a notable degradation of the signal. In Fig 3.2 is possible to detect in patient 11 an eye activity at first sight (in an early stage of the transition); for later stages of the disease is possible to note the fading of a clear response. The eye movement detection for ALS patients is thus a multifaceted problem depending not only on the algorithms' performance but also on the patients' responsiveness and the stage of ALS. Patient 16 and patient 15, in fact, proved to have the best eye control among all the patients, with a clear signal response and good prediction accuracy. Patient 13 after the first training sessions decided his own strategy to control the response, and started to move the eyes right for yes and left for no. Being able to achieve discrete prediction accuracy with this strategy the patient was encouraged to stick to his technique. P11, compared to the patients, is in the most advanced stage of ALS, reflecting a poorer response in the EOG signals for yes/no questions. Moreover, P11 is the one with the longest follow-up spanning from March 2018 to March 2019. This period is characterized by an aggravation of P11 conditions, reflecting in a notable reduction of the signal amplitude response from visit to visit. Refer to chapter 2 and 3 for a deeper analysis of the patients' conditions.

All these peculiarities suggest that eye movement recognition in ALS patients is a non-trivial problem, in this sense there is the need for an adaptive feature extraction technique that can keep up with the patients' changes, a data driven approach that can follow the muta-

ble nature of the signal. The proposed feature extraction methods aim to preserve the EOG signal information in the bands of interest.

4.2 Feature extraction through DCT

The first approach that will be explored, is an application of discrete cosine transform (DCT) on EOG signals. DCT takes correlated input data and concentrates its energy in just first few transform coefficients. Successfully used for the feature extraction in BCI applications [37], this method allows data size reduction without losing the low frequency information [38]. DCT is a transformation method for converting a time series signal into basic frequency components. Low frequency components are concentrated in the first coefficients and high frequency in the last ones. The one-dimensional DCT of a signal x of length N, and with δ_{kh} the Kronecker delta, the transform is defined by:

$$Z(k) = \sqrt{\frac{2}{N}} \sum_{n=1}^{N} \frac{x(n)}{\sqrt{1+\delta_{k1}}} \cos\left(\frac{\pi}{2N}(n-1)(k-1)\right) \qquad k = 1, 2, ..., N \qquad (4.1)$$

The input is a set of N data values and the output is a set of N DCT transform coefficients Z. DCT exhibits good energy compaction for highly correlated signals. If the input data consists of correlated quantities, then most of the N transform coefficients produced by the DCT are zeros or small numbers [37]. Applying this feature extraction to EOG signals allows compressing useful data to the first few coefficients. Therefore, only these coefficients can be used for classification using machine learning algorithms. This kind of data compression dramatically reduce input vector size and can capture features useful for classification [39] [37].

Using this approach leads to several advantages:

- Solves the problem of data filtering for high frequency noise
- Allows to select the bandwidth of interests
- Reduces dataset size
- Captures significant features for classification

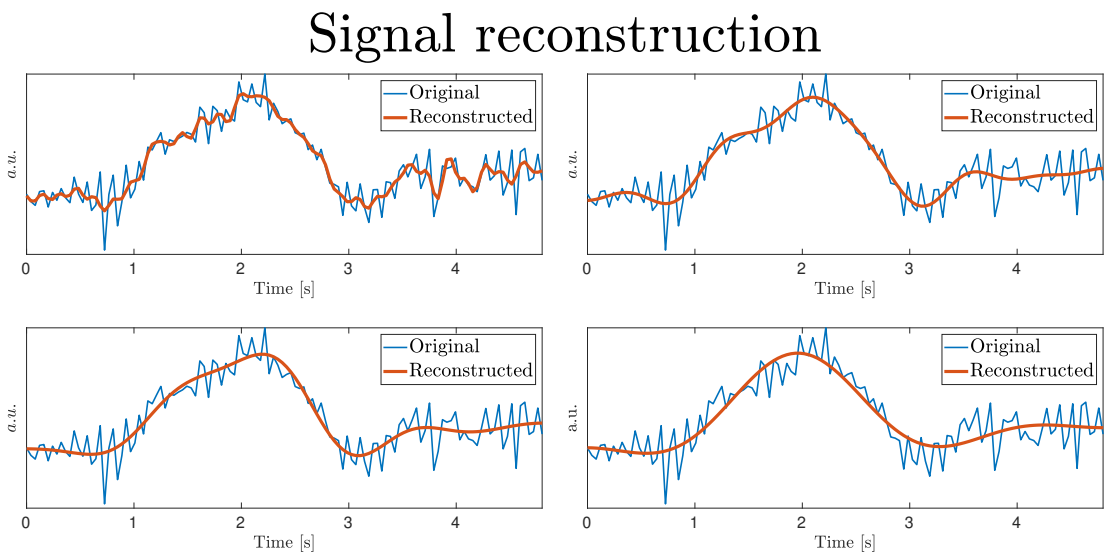

Figure 4.1: Example of signal reconstruction through inverse DCT from different number of coefficients with compression rate equal to 2 (top-left) 7 (top right) 11 (bottom left) 20 (bottom right)

In order to capture an eye movement dynamic that hardly exceeds 12 Hz [40], a 500 Hz sampling rate is heavily oversampling the bandwidth of interest. It is thus possible to decrease the input signal size by a down-sampling step as a first straight-forward technique for dimensionality reduction. The values used in this study for results and figures can be found in appendix A.

In Fig 4.1 is notable how is possible to reconstruct the original signal from just a small set of the first DCT coefficients and the progressive degradation of the inverse DCT results, as function of the increasing compression rate. Is clear how the lowest frequencies can maintain

the information relative to the patient's eye movement, capturing the slowest oscillations. The reconstruction is compared with the original signal filtered in the (0.5-30 Hz) bandwidth with zero-phase 'Back-Forward' Butterworth filter.

A crucial decision is the choice of the number of coefficients. Three metrics are proposed here to capture the fidelity of the reconstruction. The Pearson correlation coefficient, the Percentage-Root-Square Error (PRD) and the Spearsman ρ. Values are calculated as an average performance of signal reconstruction between all the available trials. Only significant values are considered for Pearson correlation and Spearsman ρ with a 5% threshold. In Fig 4.2 is possible to notice how trying to reconstruct the signal with smaller subset of DCT coefficient the error is increasing and correlations decreasing as cause of the removal of the higher frequencies.

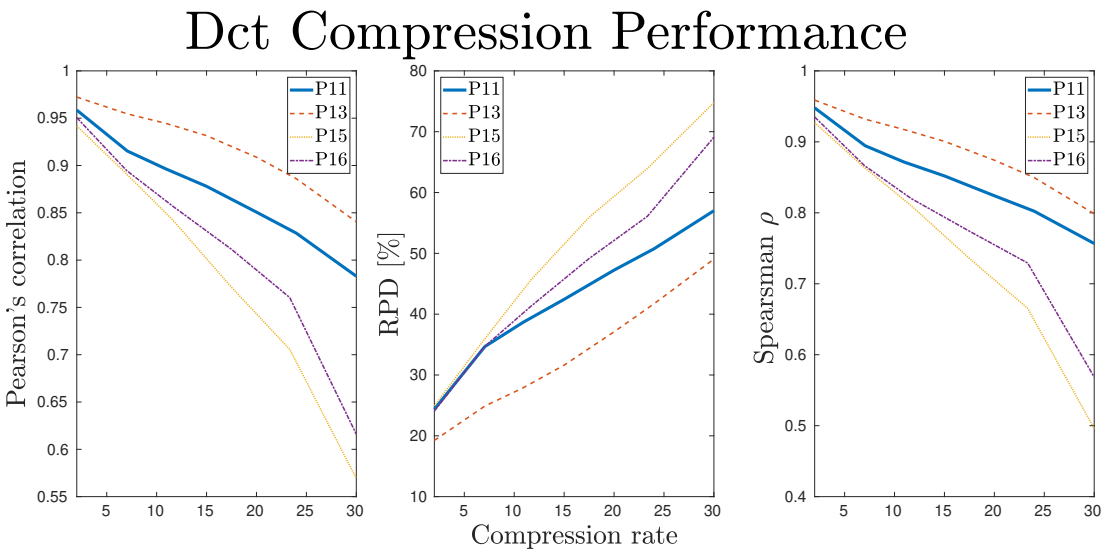

Figure 4.2: DCT reconstruction performance for the four different patients

4.3 Feature extraction through Autoencoder

In this section, will be proposed the use of an Autoencoder-based technique for dimensionality reduction. An Autoencoder (AE) is a simple neural network that consists of three main layers: input layer X, hidden layer Z and output layer $\hat{X}$ [41]. The goal of the Autoencoder is to train the network with the target values set to be equal to the inputs $X = \hat{X}$ through a reduced number of dimensions. In this way, the AE is forced to find a signal representation in the hidden space Z, much smaller than the original signal. AEs are currently in wide use in deep learning applications. They have been utilized in BCI for feature extraction and classification in [42] [43]. Exploiting the high correlation existing in the EOG channels, AE framework can reach extreme compression rate preserving the original signal information.

The input matrix shape depends on visits and patients. In P11 for example, only three channels are consistently used throughout all the visits (EOGD, EOGL, EOGR), and four are used for P13, P15 and 16 (EOGU, EOGD, EOGL, EOGR). The thinking period duration changes from patient to patient depending on the ability of the participant to control the interface (better results, shorter thinking periods). In any case, for every patient, the input signal can be arranged in a matrix with size $N_C \times N_S \times N_T$. Being N_C the number of channels, N_S the number of samples in the thinking period and N_T the number of trials. Each trial has been filtered channel/wise with a Butterworth filter between 0.5-30 Hz. The filter has been applied with a 'Backward-Forward' technique in order to obtain a zero phase distortion. In order to capture a dynamic that hardly exceeds 12 Hz [40], a down-sampling step is performed as a first straight-forward technique for dimensionality reduction, from a 500 Hz sampling rate with a factor of 20. The input is subsequently normalized in the interval [-1,1], in order to exploit the central region of the activation function selected. The

resulting preprocessed matrix size will then be indicated as $N_C \times \hat{N}_S \times N_T$.

Autoencoders convergence is a high demanding computational task, requiring time and quantity of data that are not compatible with an online application. For these reasons, the learning framework is divided into two steps. The first one entirely off-line and unsupervised to enable the AE to find the weights' values for the representation of a generic set of EOG signal, and a second step online often referred to as 'fine-tuning', supervised, that can be performed online. For a mathematical description of the unsupervised learning the following definitions will be used:

Definition 1. $D_x = N_C * \hat{N}_S$ dimension of the input space.

Definition 2. $D_z = N_z$ be the dimensions of the hidden layer.

Definition 3. $D_{\hat{x}} = N_C * \hat{N}_S$ the dimensions of the output space.

Definition 4. $x \in \mathbb{R}^{D_x \times 1}$ is an element of the input matrix.

Definition 5. $z \in \mathbb{R}^{D_z \times 1}$ represents the hidden space.

Definition 6. $W^{(1)} \in \mathbb{R}^{D_z \times D_x}$ is the weights matrix of the first layer.

Definition 7. $b^{(1)} \in \mathbb{R}^{D_z \times 1}$ is the bias vector of the first layer.

Definition 8. $h^{(1)} : \mathbb{R}^{D_z} \to \mathbb{R}^{D_z}$ indicates the punctual application of the activation function.

Exploiting the previous definitions the 'encoding' layer from the signal space to the feature space can be described as:

$$z = h^{(1)}(W^{(1)}x + b^{(1)}) \tag{4.2}$$

Similarly, the decoding part from the hidden layer to the reconstructed signal can be defined with an analogous structure remembering that the output layer $\hat{x}$ has the same shape of the input layer x:

$$\hat{x} = h^{(2)}(W^{(2)}z + b^{(2)}) \tag{4.3}$$

In order to exploit the correlation inter- and intra-channels of the EOG data a 'Dense' layer has been chosen for both the encoding and decoding side. Moreover, considering that the information that must be preserved lies in the lower frequencies, where most of the signal's power is concentrated, a 'Mean-Squared-Error' (MSE) strategy can be used as core metric for the cost function represented by the error:

$$E = MSE + \lambda * \Omega_w + \beta * \Omega_s$$

With MSE the mean-squared-error calculated between the input and the reconstructed signal, Ω_w is the L_2 regularization, and Ω_s sparsity regularization. The coefficients λ and β are hyper-parameters left to the user choice and reported in appendix A. The first term is thus the mean square error defined as:

$$MSE = \frac{1}{N_x}\sum_{n=1}^{N_x}(x_n - \hat{x_n})^2$$

Ω_s is often referred to as Sparsity regularizer. This term attempts to enforce a constraint on the sparsity of the output from the hidden layer. Sparsity can be encouraged by adding a regularization term that takes a large value when the average activation value, $\hat{\rho}_i$, of a neuron i and its desired value, ρ, are not close in value [44]. One such sparsity regularization term can be the Kullback-Leibler divergence.

$$\Omega_s = \sum_{i=i}^{D_z} KL(\rho||\hat{\rho}_i) = \sum_{i=i}^{D_z} \rho \log\left(\frac{\rho}{\hat{\rho}_i}\right) + (1-\rho)\log\left(\frac{1-\rho}{1-\hat{\rho}_i}\right) \tag{4.4}$$

Kullback-Leibler divergence is a function for measuring how different two distributions are. In this case, it takes the value zero when ρ and $\hat{\rho}_i$ are equal to each other, and becomes larger as they diverge from each other. Minimizing the cost function forces this term to be small, hence ρ and $\hat{\rho}_i$ to be close to each other.

When training a sparse autoencoder, it is possible to make the sparsity regulariser small by increasing the values of the weights w and decreasing the values of z [44]. Adding a regularization term on the weights to the cost function prevents it from happening. This term is called the L2 regularization term and is defined by:

$$\Omega_w = \frac{1}{2}\sum_{l}^{L}\sum_{j}^{N_T}\sum_{i}^{\hat{N_S}}(w_{ij}^{(l)})^2 \tag{4.5}$$

where L is the number of hidden layers, N_T is the number of observations, and $\hat{N_S}$ is the number of variables in the training data.

The selected training algorithm is known as Scaled Conjugate Gradient Descent, more information can be found in the original paper [45].

In fig 4.5 is possible to see the AE unsupervised simple structure. The unsupervised training has been performed 'patient-wise'.

A critical parameter in the AE framework is the number of hidden neurons N_z. In Fig 4.4 is possible to note the performance of the Autoencoder evaluating the similarity of the reconstructed signal from the latent space from the original signal. Three metrics are proposed to capture the fidelity of the reconstruction. The Pearson correlation coefficient, the Percentage-Root-Square Error (PRD) and the Spearsman ρ.

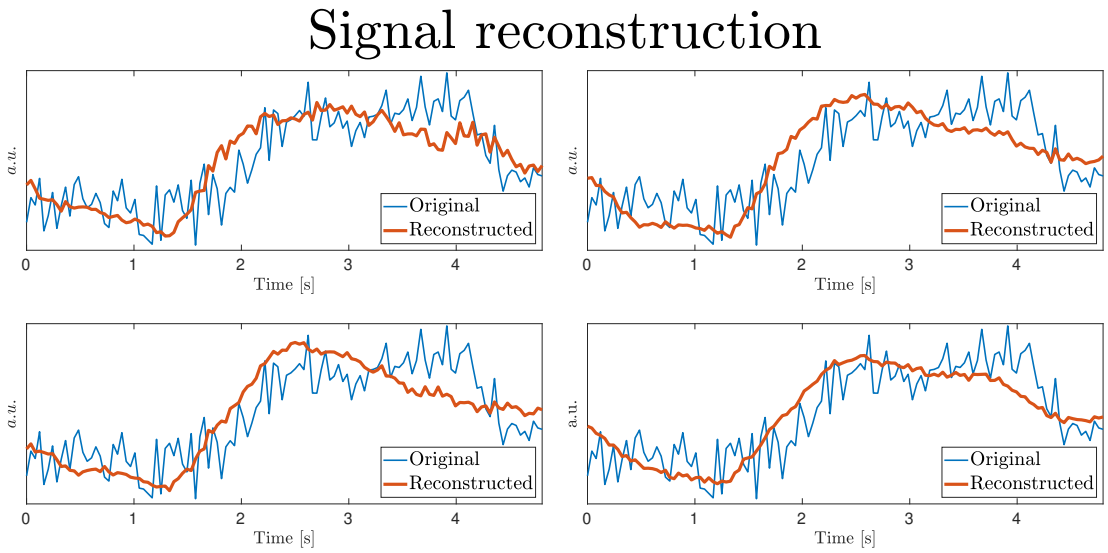

Figure 4.3: Example of signal reconstruction through Autoencoder from different hidden size N_Z with compression rate equal to 5 (top-left) 15 (top right) 25 (bottom left) 50 (bottom right)

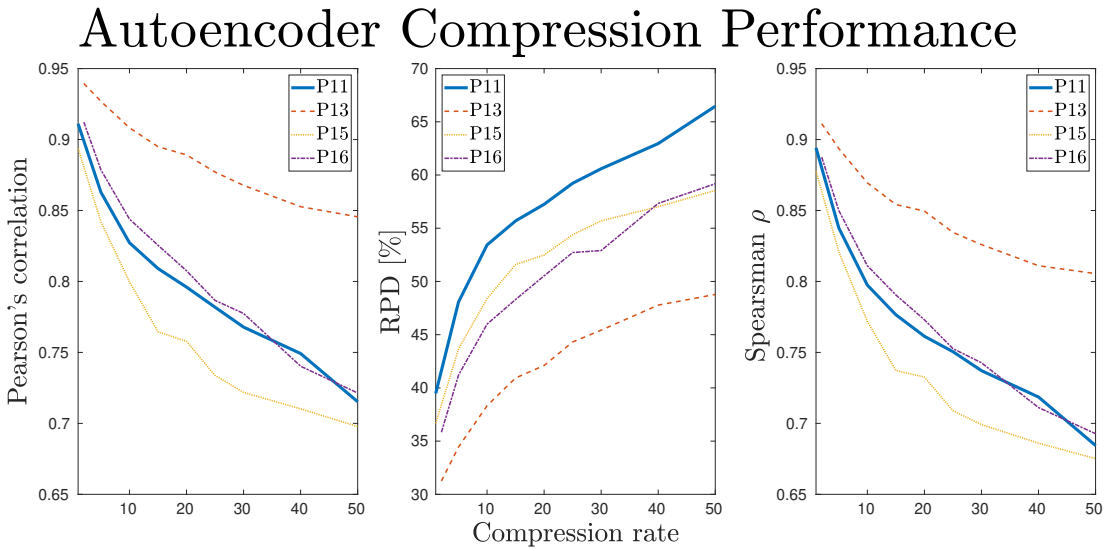

Figure 4.4: Average performance of signal reconstruction as function of compression rate, for the four different patients

Despite the other two techniques the Autoencoder framework exploits the inter-channel EOG correlation allowing a much more efficient encoded version of the signal. In Fig 4.3 is

possible to see how the reconstruction of EOG time series from the different hidden space follows the original signals. Is clear how the choice of the cost function during the training phase allows the hidden layer to capture the slowest component of the EOG, with a progressive degradation of the reconstruction caused by more aggressive compression rates.

As discussed in chapter 2, in order to build a model for one of the proposed classifiers, a supervised training step is necessary. In this phase is possible to exploit the labeled training set to back propagate the class membership in order to fine tune the network weights for a discrimination of the signal class (yes/no). The pretrained encoding side is thus linked to a softmax layer for a supervised fashioned second training. During the fine tuning step the input layer consists of different trial of the training set, and the output layer is set as the 'one-hot-encoding' representation of the relative class membership. This strategy, often used as classification step, lacks a proper amount of consistent data. As discussed in the previous sections the nature of the signal is heavily changing between patients, and sometimes also in the same patients and different visits. For this reason, this supervised step should be performed 'day-wise' and wouldn't allow satisfying convergence for classification. However, it is still performed as it can adaptively augment the discrimination power of the feature space.

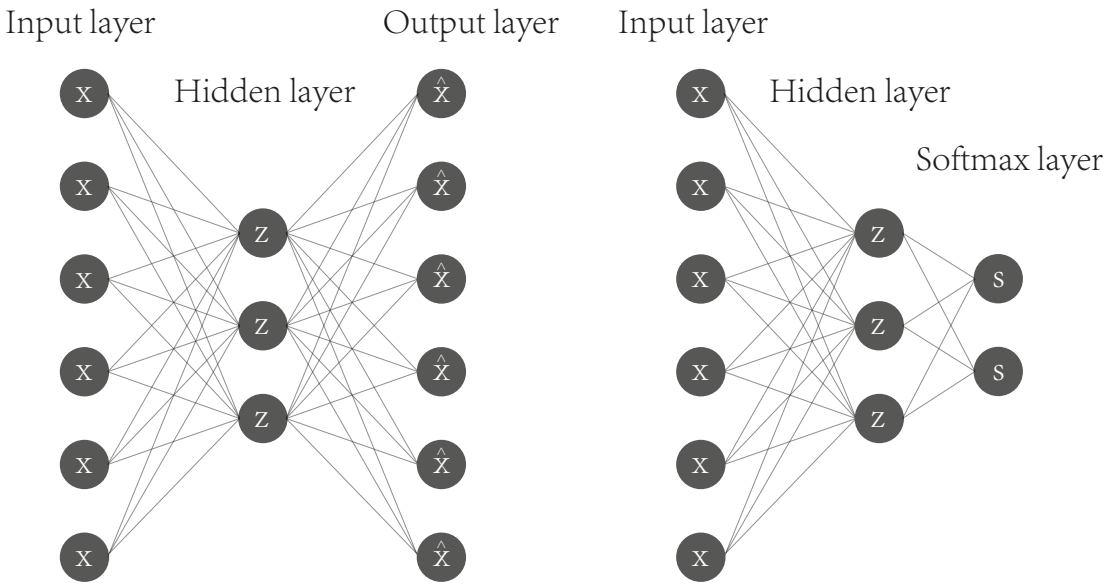

Figure 4.5: Structure of the Autoencoder framework in the unsupervised configuration (left), and fine tuning configuration (right)

4.4 Feature extraction through Dwt

Wavelet analysis is a common tool for analyzing variations of power within a time series. By decomposing a time series into time–frequency space, is possible to determine both the dominant modes of variability and how those modes vary in time. Wavelet Transform can be represented as a linear transformation $Y = WX$, where X, Y are input and output of the transformation and W is orthogonal mother wavelet transformation matrix [46]. Mother wavelet is defined as:

$$\Psi_{u,s}(t) = \frac{1}{\sqrt{s}}\Psi\left(\frac{t-u}{s}\right) \tag{4.6}$$

Wavelets are oscillating functions of time that must satisfy several conditions: A wavelet Ψ has zero time average and unit energy corresponds to orthonormality property of wavelets.

The amplitudes of a wavelet have large fluctuations within a designated time period and extremely small values outside of that time while being band-limited in terms of their frequency content. The continuous wavelet transform CWT of a signal $f(t)$ can be calculated using equation 4.7.

$$F(u,s) = \int x(t) \frac{1}{\sqrt{s}} \Psi \left(\frac{t-u}{s} \right) dt \tag{4.7}$$

By varying the values for s and u results in an infinite number of combinations and can be used to decompose the signal $x(t)$. Here u and s are the translation and dilation respectively. A much more computationally efficient approach is the use of the discrete wavelet transform (DWT), which was developed in 1989 [47]. Knowing only the values of the DWT coefficients, the waveform can be perfectly reconstructed. All of the extra coefficients of the CWT create redundancy in calculation as they are highly correlated with the ones of the DWT. In implementation, the DWT can even outperform CWT because waveforms are already digitally sampled and have finite duration so the number of coefficients is limited. DWT or CWT can be seen as a number on the time scale plane representing the correlation between the signal vector and the wavelet function at a given time-scale point. The DWT can produce as many wavelet coefficients as there are samples in the original signal. In Fig 4.6 is shown a structure of a DWT.

The original signal is convolved with a low and high pass filter whose impulse response is determined by the wavelet chosen. The output of each filter produces the same number of samples as the original signal, so both outputs are downsampled by 2 resulting in the approximation and detail coefficients each with half the number of points as the original signal. The coefficients represent a correlation between the signal of interest and wavelet chosen at different scales and during translation. Because all of the coefficients are preserved, the origi-

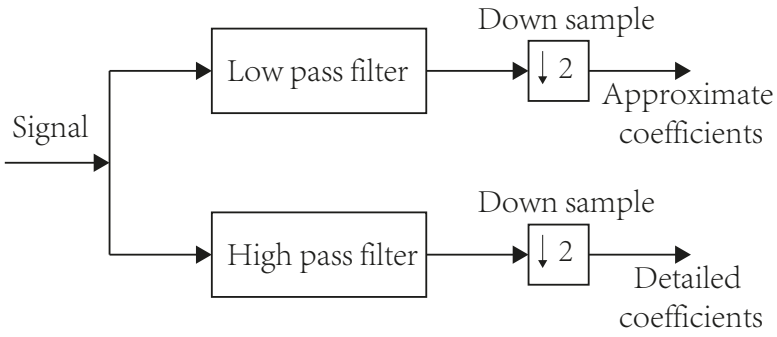

Figure 4.6: Discrete Wavelet Transform interpretation as filter-banks

nal signal or any level of decomposition can be reconstructed using a filter scheme similar to decomposition shown in Fig 4.6. The process can be reversed; doing so the coefficients are upsampled(interpolated), filtered, and summed.

Dual tree complex wavelet transform (DTCWT) is an enhancement to the discrete wavelet transform (DWT) which has additional properties useful for the problem considered. Published in 1999, Kingsbury's work [48] first highlighted the potentiality and features of complex wavelets for signal processing that proved to have desirable properties compared to real wavelet transform.

The framework of DTCWT is shown in Fig 4.7 and is structured as follows. DTCWT employs two real DWTs where the first DWT gives the real part of the transform while the second DWT gives the imaginary part. The analysis filter bank structure used to implement DTCWT uses two different sets of filters which satisfy perfect reconstruction conditions. When the dual-tree CWT is applied to a real signal, the output of the upper and lower filter banks give the real and imaginary parts of the complex coefficients respectively.

The filter design is a non trivial task in a DTCWT framework. In this work, a strategy known as 'Q-shift solution' has been adopted. More information can be found in[49].

The proposed method of feature extraction consists in first decomposing the EOG signal

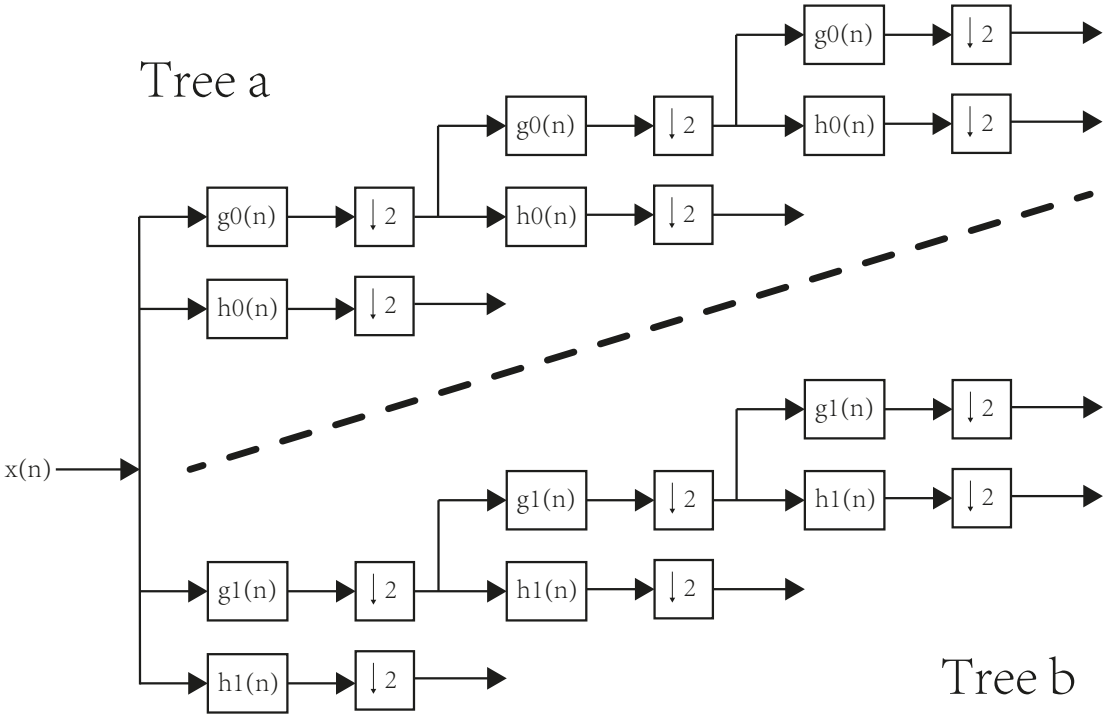

Figure 4.7: Structure of a Double Tree Complex Wavelet Transform as an ensemble of two real DWT

of each trial with dual tree complex wavelet transform into different levels and then extracting suitable features from the lower bands. Since forward transform of DTCWT gives two branches containing real and imaginary coefficients, it gives a rich way of analyzing EOG signals in both time and frequency domain. DTCWT proved to overcome several DWT shortcomings [50] [48], but the one particularly interesting property of the DTCWT is that compared to the DWT it is nearly shift invariant. The delay of response of the patients depends on many uncontrollable factors. The double tree framework reflects lower changes in its transform coefficients for a time shifted signal compared to the real DWT, and this feature allows the classification to achieve results that are less dependent on the patient's delay in the response.

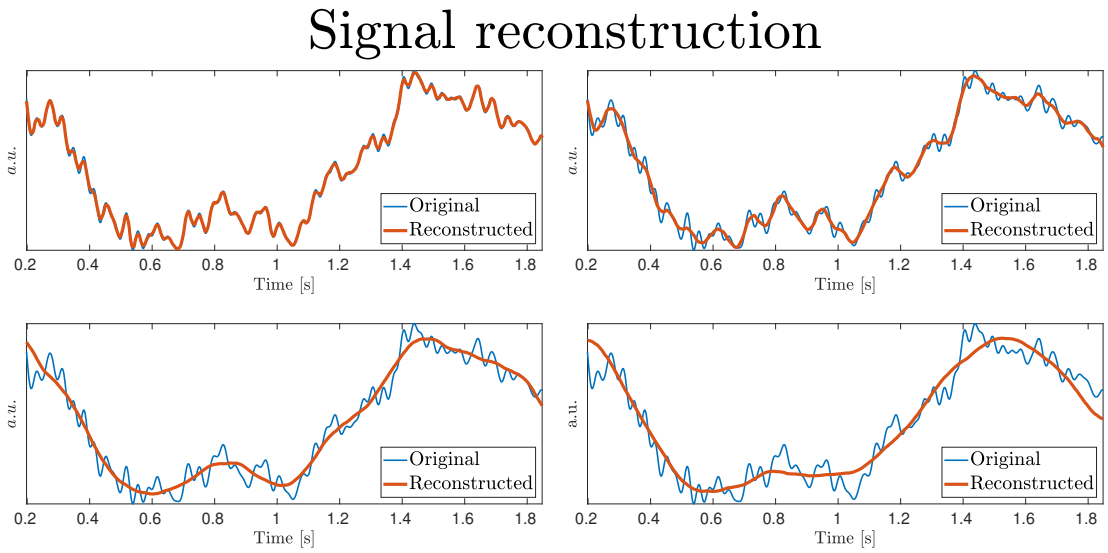

Figure 4.8: Example of signal reconstruction through inverse Dwt from different number of coefficients with compression rate equal to 3 (top-left) 5 (top right) 22 (bottom left) 44 (bottom right)

In Fig 4.8 is possible to notice how the signal reconstruction is performed from different levels of the DTCWT and how the lowest bands can follow the EOG dynamic (initial and final artifact omitted). Is important to notice that each next level of the transformation thanks to the down-sampling step allows the reconstruction from half of the previous level samples. Fig 4.9 shows the trend of three different metrics to evaluate the signal reconstruction depending on the compression rate. While trying to compare the compression rates between the current and the previous techniques is important to take into consideration the different preprocessing. While DCT and AE were preceded by a heavy downsampling, this is not the case for DTCWT feature extraction. The compression rates in the previous section indicate the reduction from the downsampled signal and the feature set. In the current framework (as notable in Fig 4.7) the downsampling step is embedded in the filter bank cascade. For this reason, the possible comparison between the compression rate should consider a scale factor equal to the downsampling rate.

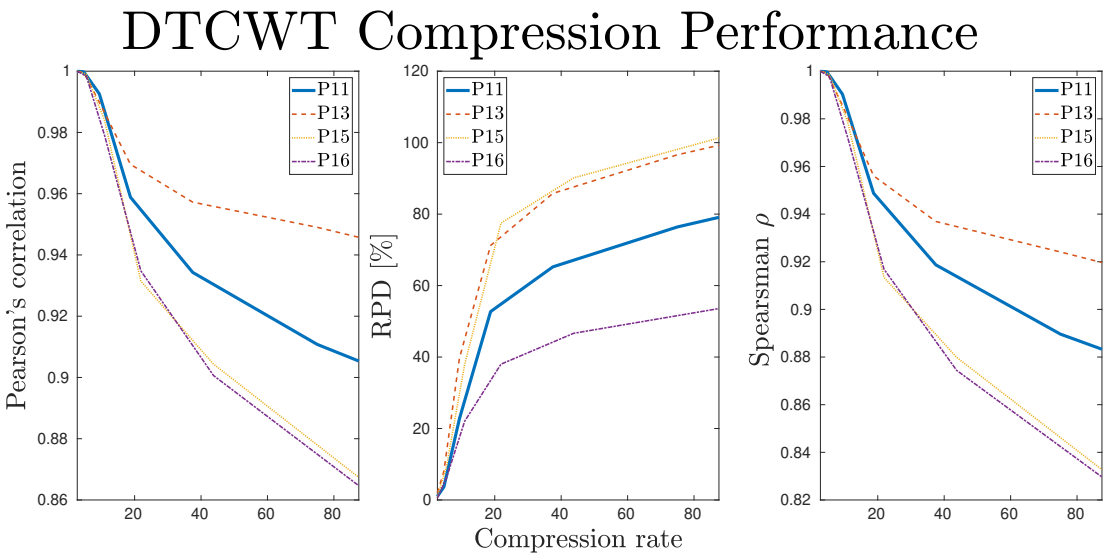

Figure 4.9: Average performance of signal reconstruction as function of compression rate, for the four different patients

5 Results

The offline results obtained in this section try to replicate the structure of the retrieval of the data, in order to analyze a possible implementation of the aforementioned feature extraction technique. Following the scheme shown in Fig 2.2, for every different day a train set and a test set are created. Considering the whole trials of one day the first 70% are considered part of the training set and the remaining 30% as part of the test set.

A critical choice that had to be made is regarding the classifier hyperparameters. Every previous step in the processing pipeline (preprocessing, feature extraction, feature selection) would lead to long and costly effort about the parameters tuning. A possible solution for achieving the best test accuracy would be an exhaustive grid search of every parameter in the pipeline, including classification. This solution was not considered for two reasons. The first reason is that the computational burden of such an optimization would lead to training and testing one different classifier for each one of the previous configurations; this combinatorial explosion makes this strategy nearly impossible. The second reason why this option is not considered is that the main goal of this study is to investigate the potentiality of the proposed

feature extraction techniques. For these reasons, the classifier parameters are kept fixed for every patient and feature extraction method. Doing so, in this case, the classifiers are used as a fixed reference and exploited to evaluate the different feature extraction methods efficacy. For these reasons, in this study is made the debatable assumption that the results are depending just on the feature extraction effectiveness.

Is important to be stated that the following results are biased in favor of the MinMax strategy. This feature extraction method can be set in three different configurations: using only minimum and maximum amplitude of the signal, using just their relative time locations, or using all the four coefficients at the same time. In each trial of the following results just the best performing one is selected. In this way, the results reported regarding MinMax, loose their meaning of *one-method* effectiveness, but on the other hand the proposed feature extraction techniques are compared with the best performing solution within the three.

In this section is shown the classification accuracy of the three proposed methods with the classifier previously described: SVM, K-NN, Decision Tree. The bar plots represent the performance of test classification accuracy on a subset of single days. Depending by how many blocks were performed that day each date might represent a different number of trials having different statistical weights when comparing the results.

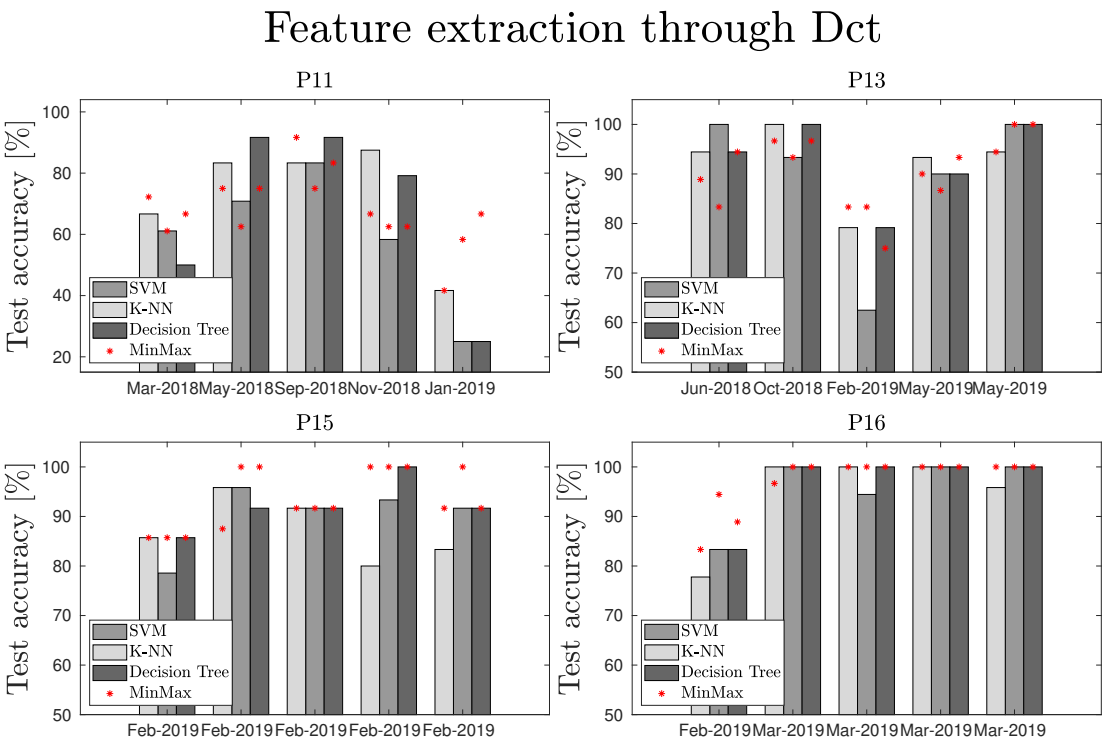

Figure 5.1: Classification Performance with DCT feature extraction with four different patients

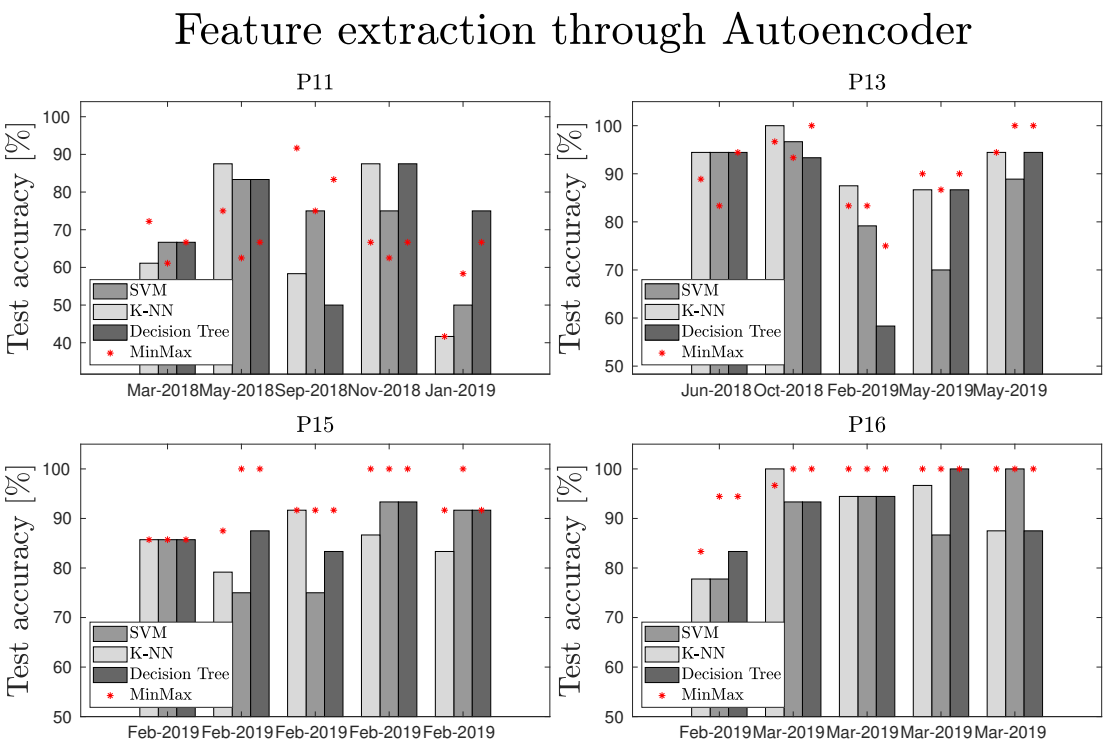

Figure 5.2: Classification Performance with Autoencoder feature extraction with four different patients

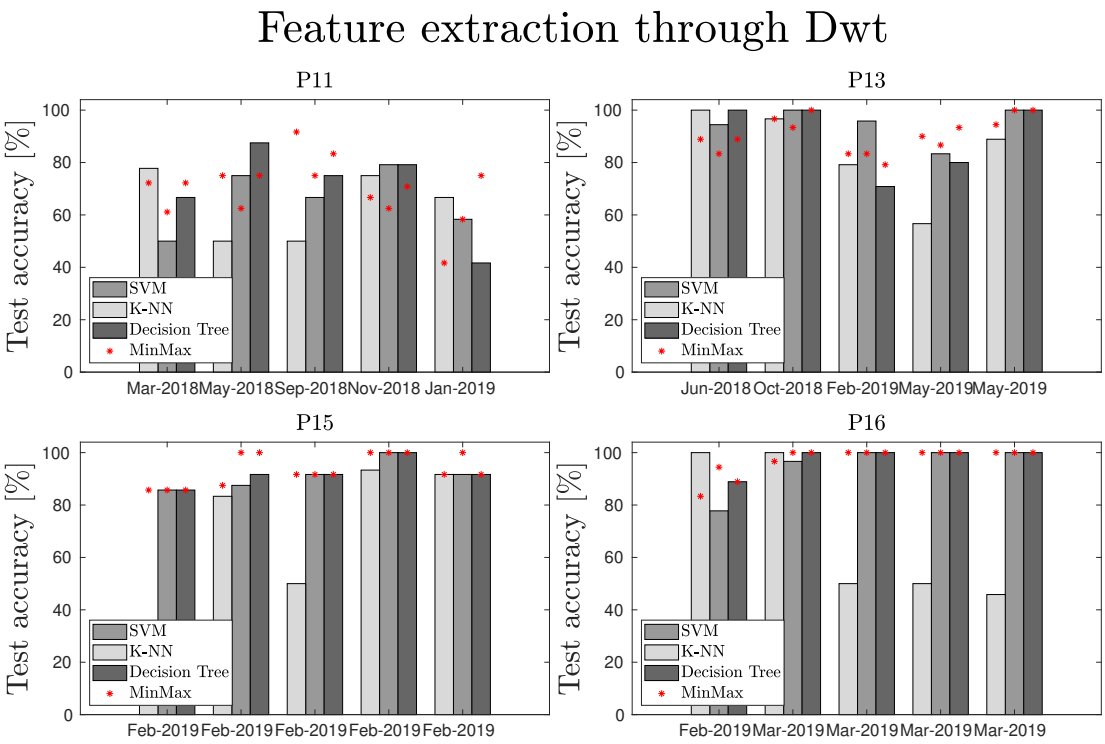

Figure 5.3: Classification Performance with DTCWT feature extraction with four different patients

5.1 Overall Results

The results reported in this section are calculated over all the trials available. Every table shows the percentage of success of each pair feature extraction - classifier, for the four different patients. A further division of the patient visits was performed for P11 in order to analyze any time related performance of the classification. The number of visits available for P15 and P16 were not enough to justify the splitting and the ones from P13 didn't show any significant changes. Specifically, the 'Early Stage' and 'Advanced Stage' notation refers to the first and last five days available.

Table 5.1: P11 All days

Feature Extraction	Support Vector Machine	K-Nearest Neighbors	Decision Tree
MinMax	68.38	66.03	70.94
Dct	73.29	64.53	73.93
Autoencoder	75.21	70.30	74.79
Autoenc+Dct	73.93	64.96	73.29
DTCWT	62.61	63.68	66.67

Table 5.2: P11 Early Stage

Feature Extraction	Support Vector Machine	K-Nearest Neighbors	Decision Tree
MinMax	72.62	67.86	72.62
Dct	72.02	68.45	67.86
Autoencoder	76.19	73.21	68.45
Autoenc+Dct	74.40	70.83	70.24
DTCWT	55.36	66.67	70.24

Table 5.3: P11 Advanced Stage

Feature Extraction	Support Vector Machine	K-Nearest Neighbors	Decision Tree
MinMax	64.44	62.22	62.22
Dct	72.22	58.89	73.33
Autoencoder	76.67	68.89	71.11
Autoenc+Dct	77.78	61.11	73.33
DTCWT	58.89	58.89	54.44

Table 5.4: P13 All Days

Feature Extraction	Support Vector Machine	K-Nearest Neighbors	Decision Tree
MinMax	87.92	85.42	90
Dct	91.67	85.83	92.08
Autoencoder	89.17	84.58	85.83
Autoenc+Dct	91.25	84.17	89.58
DTCWT	63.75	88.33	87.92

Table 5.5: P15 All Days

Feature Extraction	Support Vector Machine	K-Nearest Neighbors	Decision Tree
MinMax	90.91	96.10	94.81
Dct	88.31	90.91	90.91
Autoencoder	84.42	83.12	80.52
Autoenc+Dct	89.61	93.51	92.21
DTCWT	70.13	90.91	92.21

Table 5.6: P16 All Days

Feature Extraction	Support Vector Machine	K-Nearest Neighbors	Decision Tree
MinMax	96.67	99.17	98.33
Dct	95.83	96.67	96.67
Autoencoder	95.83	95	90
Autoenc+Dct	96.67	95.83	98.33
DTCWT	69.17	95.83	98.33

6
Conclusion

The first consideration coming from the previous chapter results, is that the three proposed solutions (plus the MinMax strategy) are a valid alternative for the classification of eye movements of ALS patients in transition from LIS to CLIS. The three methods proved (with one exception) to have comparable results one another, showing classification accuracy consistent with the patients' conditions. The only coupling that seems to be widely sub-optimal is the DTCWT feature extraction with SVM classifier. As shown in Tab 5.6 the coupling of DTCWT with K-NN and Decision tree on the same dataset are satisfactory, with an almost 25% difference in term of classification performance for P15 and P16. It suggests that the particular shape of the feature set extracted with the DTCWT framework are hardly divisible by the SVM hyperplane. One could speculate that if the information is retained in the wavelet coefficients (as shown with K-NN and Decision Tree) it should exist a non linear transformation that could make the feature set separable allowing SVM to reach performance up to expectations. Such an optimization was not taken into consideration because the argument would go beyond the purpose of this study as discussed in Chapter 5. But can be stated that

this configuration proves to be strongly sub-optimal.

Other considerations can be done regarding the Autoencoder framework, the first one is about the offline/online nature of the results. P11 'Early Stage', P13 'Early Stage', P15 and P16 can't be considered comparable to online performance as their data are used for the unsupervised training of the AE. One could argue that this procedure goes against the train/test splitting philosophy prejudicing the transparency of the results. That would not be completely correct as the splitting regards the classification step and its integrity is here preserved. During the prediction of a trial class (in the test-set), what is guaranteed is the fact that is the first time for the classifier to 'see' that trial, but is not for the Autoencoder. Even if the classifier performance can be directly compared to the online results, the Autoencoder has thus to be considered an offline application in the aforementioned subset. The choice to use the first part of the dataset in the case of patients 11 is debatable. In the first instance, this strategy was chosen to maintain a chronological consecutiveness between the training and the prediction. In this manner is possible to evaluate a 'real world' application of the technique. As discussed in Chapter 3 the signal's nature might vary during time. The right selection of the time window used as training set might be crucial. Large training set might weaken representation ability of the AE (as it would be trained to represent different signals at the same time), on the other hand, a small dataset would lead to a lack of generalization from the AE and consecutive overfitting.

Some of the previous choices in the AE framework are forced by the lack of data. The scarcity and singularity of the patients make the signal retrieval costly and difficult. A promising strategy that could tackle the problem would be the one known as 'transfer learning'. In some measure, the new patients (with no dataset available) can use pre-trained AE with other patients data.

Table 6.1: Transfer table

Transfer	Support Vector Machine	K-Nearest Neighbors	Decision Tree
P13 -> P11	68.80	61.75	69.44
P11 -> P13	93.65	86.51	87.30
P16 -> P15	83.12	80.52	79.22
P15 -> P16	93.33	93.33	90.83

Table 6.1 represents the results of the classification through AE feature extraction swapping the patients' training set. The notation 'P13 -> P11' for instance indicates the prediction accuracy on P11 test set using an AE trained with P13 train-set. These results prove that this technique can be applied also to new patients and refined with the usage and the increasing dataset. Further cross validation results weren't obtained because of the different paradigm between the patients. The thinking period of P11 and P13 was set to 5 s while for P15 and P16 it was 3 s. For this reason the strict structure of the AE wouldn't allow a straightforward comparison.

Another interesting feature regarding the AE framework comes from the comparison with the feature functions strategy. Observing tables 5.2 and 5.3, is possible to notice a difference of about -8% -5% -10% from 'Early' to 'Late' stage of P11 with MinMax strategy, while the same comparison with AE is about 0% -4% $+3\%$. From these results seems that AE strategy performance holds *better* with the progression of ALS in P11. In order to asses the significance of the difference between the two strategies a Wilcoxon test has been performed. In Fig 6.1 is shown the difference between the performance of two different groups: the MinMax and AE feature extraction with SVM classification.

The 10 days grouping (with one day sliding window) has been chosen as a compromise between having high sample population for significant results and time resolution: a few days grouping would lead to non significant results (because of the lack of data), and a many

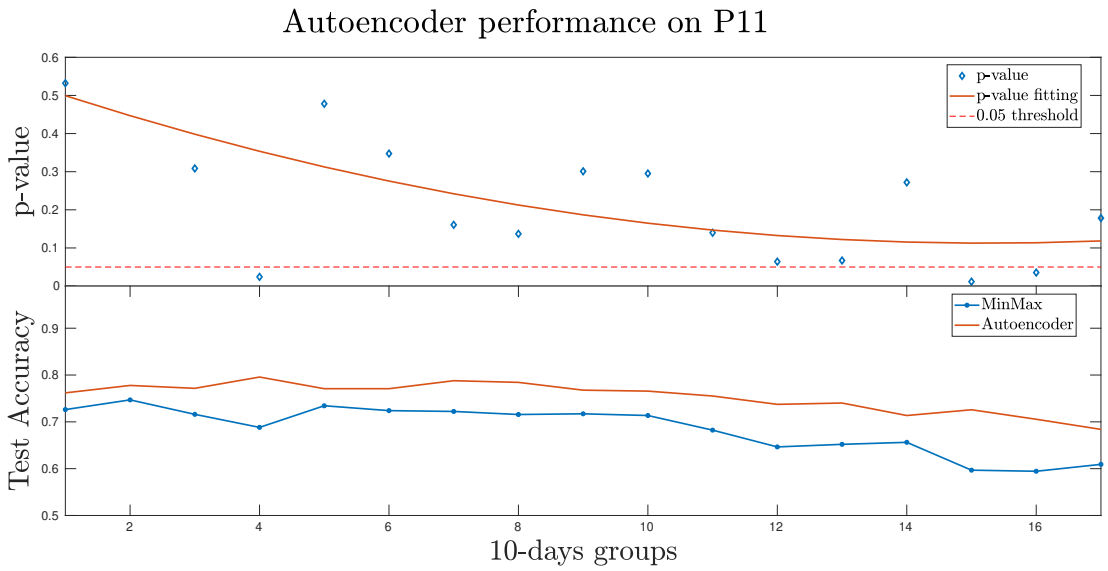

Figure 6.1: Test accuracy and relative p-value in 10-days groups along time

days grouping would lose the time resolution, losing the time analysis purpose of the statistic. From the Fig 6.1 is possible to see that in this statistical set up the AE classification accuracy is grater than MinMax in every group and is even more interesting to notice that the significance of the difference is increasing along time as shown from the decreasing of the p-value.

6.0.1 Future works

The passage from LIS to CLIS is a very unpredictable transition where the decline of the disease can follow various trends. The decay of the muscle control ability is a subjective path different for each person. In this study has been analyzed four different patients that showed different signal characteristic, with a particularly interesting evolution in the case of P11. The goal of the interface that has been created is to allow communication in the shortest period possible. The duration of patients' answering is one of the critical parameters in the paradigm that has not been discussed in this study. An interesting possible future work

would be developing a method and an optimality criteria for the selection of the right interval.

Another interesting topic related to the advanced stage of ALS regards the patients' collaboration. If it is simple to asses the subjects' response looking at their eye movement in the early stage, is more difficult if not impossible, getting closer to CLIS; confusion, distraction and frustration for instance could be factors that could weaken the hypothesis of collaboration. A possible future study could be on the aforementioned hypothesis of patients' responsiveness. Having more information on the patients' effort would be interesting for many reasons. First of all it would be a potential psychophysical index about the subject, but apart from that it could allow selecting the 'good' trials performed by the patient to build a more robust classifier model.

In conclusion, there are still many possibilities for improvement of the proposed system and it is of fundamental importance to continue this research mainly for two reasons. On the one hand, finding a stable and performing system also means better understanding the ALS progression, giving fundamental information about the development, diagnosis, and treatment of this syndrome. On the other hand, a system capable of restoring communication between a conscious person but unable to relate to the environment around him could greatly improve the psychological well-being of the patient himself but in the same way of his family and caregivers.

A Appendix

A.1 Dataset specifics

Table A.1: Dataset information

Patient	Visits period	Visits #	Days #	Blocks #	Trials #	Thinking	Channels #
P11	Mar 2018-Mar 2019	10	26	79	1580	5s	3
P13	Jun 2018-May 2019	4	12	40	800	5s	4
P15	Feb 2019-Feb 2019	2	5	20	400	3s	4
P16	Feb 2019-May 2019	2	5	20	400	3s	4

A.2 Dct parameters

Table A.2: Parameters for Dct feature extraction

Method	Downsampling rate	Selected coefficients	Total features #	Features selected #
Dct	20	25	$25 \times N_c$	20

A.3 AE parameters

Table A.3: Parameters for AE feature extraction

Method	Downsampling rate	Hidden size	Activation	λ	β	ρ	Features selected #
AE	20	25	Sigmoid	0.001	1	0.05	20

A.4 DTCWT parameters

Table A.4: Parameters for DTCWT feature extraction

Method	Downsampling rate	Tree Level	Wavelet	Total features #	Features selected #
DTCWT	None	6	Kingsbury-Q-shift	$N_s/2^6 \times N_c$	20

References

[1] L. M. McCane, S. M. Heckman, D. J. McFarland, G. Townsend, J. N. Mak, E. W. Sellers, D. Zeitlin, L. M. Tenteromano, J. R. Wolpaw, and T. M. Vaughan, "P300-based brain-computer interface (bci) event-related potentials (erps): People with amyotrophic lateral sclerosis (als) vs. age-matched controls," *Clinical Neurophysiology*, vol. 126, no. 11, pp. 2124 – 2131, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S138824571500067X

[2] M. Doyle and B. Phillips, "Trends in augmentative and alternative communication use by individuals with amyotrophic lateral sclerosis," *Augmentative and Alternative Communication*, vol. 17, no. 3, pp. 167–178, 2001. [Online]. Available: https://doi.org/10.1080/aac.17.3.167.178

[3] J. Wolpaw and E. Wolpaw, "Brain-computer interfaces: Something new under the sun," *Brain-Computer Interfaces: Principles and Practice*, 01 2012.

[4] U. Chaudhary, B. Xia, S. Silvoni, L. G. Cohen, and N. Birbaumer, "Brain–computer interface–based communication in the completely locked-in state," *PLOS Biology*, vol. 15, no. 1, pp. 1–25, 01 2017. [Online]. Available: https://doi.org/10.1371/journal.pbio.1002593

[5] D. R. Gress, "Plum and posner's diagnosis of stupor and coma, 4th edition," *Neurology*, vol. 72, no. 3, pp. 295–295, 2009. [Online]. Available: https://n.neurology.org/content/72/3/295

[6] G. Bauer, F. Gerstenbrand, and E. Rumpl, "Varieties of the locked-in syndrome," *Journal of Neurology*, vol. 221, no. 2, pp. 77–91, Aug 1979. [Online]. Available: https://doi.org/10.1007/BF00313105

[7] M. K. Ardali, A. Rana, M. Purmohammad, N. Birbaumer, and U. Chaudhary, "Semantic and bci-performance in completely paralyzed patients: Possibility of language attrition in completely locked in syndrome," *Brain and Language*, vol.

194, pp. 93 – 97, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0093934X18301792

[8] A. R. Murguialday, J. Hill, M. Bensch, S. Martens, S. Halder, F. Nijboer, B. Schoelkopf, N. Birbaumer, and A. Gharabaghi, "Transition from the locked in to the completely locked-in state: A physiological analysis," *Clinical Neurophysiology*, vol. 122, no. 5, pp. 925 – 933, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1388245710006619

[9] E. Smith and M. Delargy, "Locked-in syndrome," *BMJ*, vol. 330, no. 7488, pp. 406–409, 2005. [Online]. Available: https://www.bmj.com/content/330/7488/406

[10] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain-computer interfaces for communication and rehabilitation," *Nature Reviews Neurology*, vol. 12, pp. 513 EP –, Aug 2016, review Article. [Online]. Available: https://doi.org/10.1038/nrneurol.2016.113

[11] G. Townsend, B. Graimann, and G. Pfurtscheller, "Continuous eeg classification during motor imagery-simulation of an asynchronous bci," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 12, no. 2, pp. 258–265, June 2004.

[12] M. Matsumoto and J. Hori, "Classification of silent speech using support vector machine and relevance vector machine," *Applied Soft Computing*, vol. 20, pp. 95 – 102, 2014, hybrid intelligent methods for health technologies. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494613003578

[13] S. H. Kwon and H. C. Kim, "Eog-based glasses-type wireless mouse for the disabled," in *Proceedings of the First Joint BMES/EMBS Conference. 1999 IEEE Engineering in Medicine and Biology 21st Annual Conference and the 1999 Annual Fall Meeting of the Biomedical Engineering Society (Cat. N*, vol. 1, Oct 1999, pp. 592 vol.1–.

[14] D. R. Kher and S. Shah, "Design of electrooculogram based wheelchair control," 05 2016.

[15] R. Barea, L. Boquete, M. Mazo, and E. Lopez, "System for assisted mobility using eye movements based on electrooculography," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 4, pp. 209–218, Dec 2002.

[16] S. He and Y. Li, "A single-channel eog-based speller," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1978–1987, Nov 2017.

[17] A. Tonin, A. Jaramillo-Gonzalez, A. Rana, M. Khalili Ardali, N. Birbaumer, and U. Chaudhary, "Brain computer interface based communication system for patients in transition from locked-in to complete locked-in state," 2019.

[18] R. Medeiros, A. C. S. Souza, and G. F. Rodrigues, "Mouse control interface using electrooculogram and genetic programming," in *XXVI Brazilian Congress on Biomedical Engineering*, R. Costa-Felix, J. C. Machado, and A. V. Alvarenga, Eds. Singapore: Springer Singapore, 2019, pp. 335–339.

[19] A. Güven and S. Kara, "Classification of electro-oculogram signals using artificial neural network," *Expert Systems with Applications*, vol. 31, no. 1, pp. 199 – 205, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417405002058

[20] L. Jacobs, D. Bozian, R. R. Heffner, and S. A. Barron, "An eye movement disorder in amyotrophic lateral sclerosis," *Neurology*, vol. 31, no. 10, pp. 1282–1282, 1981. [Online]. Available: https://n.neurology.org/content/31/10/1282

[21] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767 – 791, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1388245702000573

[22] A. Bulling, J. A. Ward, H. Gellersen, and G. Troster, "Eye movement analysis for activity recognition using electrooculography," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 741–753, April 2011.

[23] R. Barea, L. Boquete, M. Mazo, and E. Lopez, "System for assisted mobility using eye movements based on electrooculography," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 4, pp. 209–218, Dec 2002.

[24] S. Yathunanthan, L. U. R. Chandrasena, A. Umakanthan, V. Vasuki, and S. R. Munasinghe, "Controlling a wheelchair by use of eog signal," in *2008 4th International Conference on Information and Automation for Sustainability*, Dec 2008, pp. 283–288.

[25] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 1, pp. 131 – 156, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1088467X97000085

[26] Hanchuan Peng, Fuhui Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug 2005.

[27] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," in *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003*, Aug 2003, pp. 523–528.

[28] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[29] N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines and other kernel-based learning methods," 2000.

[30] R.-E. Fan, P.-H. Chen, and C.-J. Lin, "Working set selection using second order information for training support vector machines," *J. Mach. Learn. Res.*, vol. 6, pp. 1889–1918, Dec. 2005. [Online]. Available: http://dl.acm.org/citation.cfm?id=1046920.1194907

[31] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324

[32] O. T. Yıldız and E. Alpaydın, "Model selection in omnivariate decision trees," in *Machine Learning: ECML 2005*, J. Gama, R. Camacho, P. B. Brazdil, A. M. Jorge, and L. Torgo, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 473–484.

[33] W.-Y. Loh, "Regression trees with unbiased variable selection and interaction detection," 2002.

[34] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-15, no. 4, pp. 580–585, July 1985.

[35] M. Vidal, A. Bulling, and H. Gellersen, "Analysing eog signal features for the discrimination of eye movements with wearable devices," in *Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-based Interaction*, ser. PETMEI '11. New York, NY, USA: ACM, 2011, pp. 15–20. [Online]. Available: http://doi.acm.org/10.1145/2029956.2029962

[36] M. M. Rahman, M. I. H. Bhuiyan, and A. R. Hassan, "Sleep stage classification using single-channel eog," *Computers in Biology and Medicine*, vol. 102, pp. 211 – 220, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010482518302427

[37] D. Birvinskas, V. Jusas, I. Martisius, and R. Damasevicius, "Eeg dataset reduction and feature extraction using discrete cosine transform," in *2012 Sixth UKSim/AMSS European Symposium on Computer Modeling and Simulation*, Nov 2012, pp. 199–204.

[38] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. C-23, no. 1, pp. 90–93, Jan 1974.

[39] R. V. Wankar, P. Shah, and R. Sutar, "Feature extraction and selection methods for motor imagery eeg signals: A review," in *2017 International Conference on Intelligent Computing and Control (I2C2)*, June 2017, pp. 1–9.

[40] E. Dong, C. Li, and C. Chen, "An eog signals recognition method based on improved threshold dual tree complex wavelet transform," in *2016 IEEE International Conference on Mechatronics and Automation*, Aug 2016, pp. 954–959.

[41] J. Gehring, Y. Miao, F. Metze, and A. Waibel, "Extracting deep bottleneck features using stacked auto-encoders," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 3377–3381.

[42] S. Narejo, E. Pasero, and F. Kulsoom, "Eeg based eye state classification using deep belief network and stacked autoencoder," *International Journal of Electrical and Computer Engineering*, vol. 6, p. 3131, 12 2016.

[43] M. Dai, D. Zheng, R. Na, S. Wang, and S. Zhang, "Eeg classification of motor imagery using a novel deep learning framework," *Sensors*, vol. 19, no. 3, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/3/551

[44] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision Research*, vol. 37, no. 23, pp. 3311 – 3325, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0042698997001697

[45] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, no. 4, pp. 525 – 533, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608005800565

[46] M. S. Reddy, B. Narasimha, E. Suresh, and K. S. Rao, "Analysis of eog signals using wavelet transform for detecting eye blinks," in *2010 International Conference on Wireless Communications Signal Processing (WCSP)*, Oct 2010, pp. 1–4.

[47] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989. [Online]. Available: http://dx.doi.org/10.1109/34.192463

[48] B. W. Silverman, J. C. Vassilicos, and N. Kingsbury, "Image processing with complex wavelets," *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 357, no. 1760, pp. 2543–2560, 1999. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rsta.1999.0447

[49] N. Kingsbury, "Design of q-shift complex wavelets for image processing using frequency domain energy minimization," in *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, vol. 1, Sep. 2003, pp. I–1013.

[50] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, Nov 2005.

Acknowledgments

This study wouldn't have been possible without the amazing research performed by Niels Birbaumer and Ujwal Chaudry in the last years. So the first thanksgiving goes to them and their work. A special thank goes to all my colleagues from the laboratory of Institute of Medical Psychology and Behavioral Neurobiology of Tübingen, that helped me with my work and made my days enjoyable.

And finally I wanted to thank my friends and family for supporting me along these years of university.