NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

MARIA MARGARIDA MENDES ANTUNES

Bachelor of Science in Biomedical Engineering

# TOWARDS AMYOTROPHIC LATERAL SCLEROSIS INTERPRETABLE DIAGNOSIS USING SURFACE ELECTROMYOGRAPHY

MASTER IN BIOMEDICAL ENGINEERING

NOVA University Lisbon
November, 2021

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

# TOWARDS AMYOTROPHIC LATERAL SCLEROSIS INTERPRETABLE DIAGNOSIS USING SURFACE ELECTROMYOGRAPHY

MARIA MARGARIDA MENDES ANTUNES

Bachelor of Science in Biomedical Engineering

**Adviser:** Prof. Dr. Hugo Filipe Silveira Gamboa
*Associate Professor, NOVA University of Lisbon*

**Examination Committee:**

**Chair:** Dr. Carla Maria Quintão Pereira
*Assistant Professor, FCT-NOVA*

**Rapporteur:** Dr. Cláudia Regina Pereira Quaresma
*Assistant Professor, FCT-NOVA*

**Adviser:** Dr. Hugo Filipe Silveira Gamboa
*Associate Professor, FCT-NOVA*

MASTER IN BIOMEDICAL ENGINEERING

NOVA University Lisbon
November, 2021

**Towards Amyotrophic Lateral Sclerosis interpretable diagnosis using surface electromyography**

Copyright © Maria Margarida Mendes Antunes, NOVA School of Science and Technology, NOVA University Lisbon.
The NOVA School of Science and Technology and the NOVA University Lisbon have the right, perpetual and without geographical boundaries, to file and publish this dissertation through printed copies reproduced on paper or on digital form, or by any other means known or that may be invented, and to disseminate through scientific repositories and admit its copying and distribution for non-commercial, educational or research purposes, as long as credit is given to the author and editor.

This document was created using the (pdf/Xe/Lua)LaTeX processor, based on the NOVAthesis template, developed at the Dep. Informática of FCT-NOVA by João M. Lourenço. [1]

# Acknowledgements

The accomplishment of this important stage of my life would not have been possible without the support of several people who deserve my most sincere gratitude.

First, I thank my academic supervisor, Professor Hugo Gamboa, for his encouragement and expertise and giving me the opportunity to be a part of such a meaningful and exciting project. I also thank him for the stimulating questions on Artificial Intelligence, which made me realize that this was the field I ought to pursue. I thank Professor Carla Quintão for the last five years of guidance and her collaboration by providing the data for this thesis. I would also like to express my gratitude to all the members of *Fraunhofer AICOS* for welcoming me and teaching me so much. You gave me the opportunity to grow as a professional, and I am very proud to say that I belong to this team. A special thanks to Marília Barandas for always being ready to help me with my doubts and discuss my work. Finally, I will never be able to repay everything that Duarte Folgado has done for me. I thank him for his patience, availability, support, and friendship. You set the example of what a Data Scientist should be.

To Marta, Marta, Rita, Joana, Nisa, Beatriz, Bruno, André, Daniel, Pedro, Diogo, Luís. There are no words to describe our connection. In particular, I have to thank you, Bruno Ribeiro, for always being by my side, answering my endless questions, and making me laugh during stressful times.

*Queria agradecer à minha família por estarem sempre presentes e por acreditarem em mim. A ti mana pelo exemplo de resiliência e coragem. A ti pai pelo exemplo de dedicação e trabalho. A ti mãe pelo exemplo de compaixão e grandeza de espírito. Espero um dia conseguir ser um bocadinho daquilo que vocês são.*

And last but certainly not least, to all the people diagnosed with Amyotrophic Lateral Sclerosis. There is hope.

# Abstract

Amyotrophic Lateral Sclerosis (ALS) is a fast-progressing disease with no cure. It is diagnosed through the assessment of clinical exams, such as needle electromyography, which measures the muscles' electrical activity by inserting a needle into the muscle tissue. Nevertheless, surface electromyography (SEMG) is emerging as a more practical and less painful alternative. Even though these exams provide relevant information regarding the electric structures conducted in the muscles, ALS symptoms are similar to those of other neurological disorders, preventing a faster detection of the disease.

This dissertation focuses on implementing and analyzing innovative SEMG features related to the morphology of the functional structures present in the signal. To assess the efficiency of these features, a framework is proposed, aiming to distinguish healthy from pathological signals through the use of a classification algorithm. The classification task was performed using SEMG signals acquired from the upper limb muscles of healthy and ALS subjects.

The results show the utility of employing the proposed set of features for ALS diagnosis, with an $F_1$ Score higher than 80% in most experimental conditions. The novel features improved the model's overall performance when combined with other state-of-art SEMG features and also demonstrated efficiency when used individually. These outcomes are of significant importance in supporting the use of SEMG as a complementary diagnosis exam. The proposed features demonstrate promising contributions for better and faster detection of ALS and increased classification interpretability.

**Keywords:** Amyotrophic Lateral Sclerosis, Surface Electromyography, Time Series, Signal Processing, Feature Selection, Machine Learning

# Resumo

A Esclerose Lateral Amiotrófica (ELA) é uma doença incurável de progressão rápida. O seu diagnóstico é feito através da avaliação de exames clínicos como a eletromiografia de profundidade, que mede a atividade elétrica muscular com agulhas inseridas no músculo. No entanto, a eletromiografia de superfície (SEMG) surge como uma alternativa mais prática e menos dolorosa. Embora ambos os exames forneçam informações relevantes sobre as estruturas elétricas conduzidas nos músculos, os sintomas da ELA são semelhantes aos de outras doenças neurológicas, impedindo uma identificação mais precoce da doença.

Esta dissertação foca-se na implementação e análise de atributos inovadores de SEMG relacionados com a morfologia das estruturas funcionais presentes no sinal. Para avaliar a eficiência destes atributos, é proposto um *framework*, com o objetivo de distinguir sinais saudáveis e sinais patológicos através de um algoritmo de classificação. A tarefa de classificação foi realizada utilizando sinais de SEMG adquiridos dos músculos dos membros superiores de indivíduos saudáveis e com ELA.

Os resultados demonstram a utilidade do conjunto de atributos proposto para o diagnóstico de ELA, com uma métrica de classificação $F_1$ superior a 80% na maioria das condições experimentais. Os novos atributos melhoraram o desempenho geral do modelo quando combinados com outros atributos de SEMG do estado da arte, e também se comprovaram eficientes quando aplicados individualmente. Estes resultados são de grande importância na justificação da aplicabilidade da SEMG como um exame complementar de diagnóstico da ELA. Os atributos apresentados demonstram ser promissores para um melhor e mais rápido diagnóstico, e facilitam a explicação dos resultados da classificação devido à sua interpretabilidade.

**Palavras-chave:** Esclerose Lateral Amiotrófica, Eletromiografia de Superfície, Séries Temporais, Processamento de Sinal, Seleção de Atributos, Aprendizagem Automática

# Contents

**List of Figures** **xv**

**List of Tables** **xvii**

**Abbreviations** **xix**

**1 Introduction** **1**
1.1 Context . . . 1
1.2 Motivation . . . 2
1.3 Literature Review . . . 2
1.4 Objectives . . . 5
1.5 Thesis Overview . . . 6

**2 Theoretical Background** **7**
2.1 The Motor Unit and the Electrical Conduction of Nervous Impulses . . . 7
2.2 Amyotrophic Lateral Sclerosis (ALS) . . . 11
2.2.1 Pathology . . . 11
2.2.2 Causes and Risk Factors . . . 11
2.2.3 Symptoms . . . 11
2.2.4 Pathophysiology . . . 12
2.2.5 Diagnosis . . . 12
2.2.6 Prognosis and Treatment . . . 13
2.2.7 Disease Progression Quantification . . . 13
2.3 Electromyography (EMG) . . . 14
2.3.1 Introduction . . . 14
2.3.2 Electromyography in Amyotrophic Lateral Sclerosis . . . 16
2.4 Machine Learning . . . 17
2.4.1 Introduction . . . 17
2.4.2 Feature Extraction and Selection . . . 19
2.4.3 Classification . . . 21

2.4.4 Validation and Performance Metrics . . . . . 22

**3 Novel Morphological Features for Surface Electromyography Characterization 27**
3.1 State-of-the-art SEMG features . . . . . 27
3.1.1 Time-Domain Features . . . . . 27
3.1.2 Spectral-Domain Features . . . . . 31
3.1.3 Statistical-Domain Features . . . . . 33
3.1.4 Fractal-Domain Features . . . . . 35
3.2 Novel Features . . . . . 36

**4 A Framework for Amyotrophic Lateral Sclerosis Diagnosis based on SEMG 43**
4.1 Preprocessing . . . . . 44
4.2 Feature Extraction . . . . . 45
4.3 Feature Selection . . . . . 45
4.4 Supervised Learning . . . . . 46
4.4.1 Classifier's Parameters Optimization . . . . . 47
4.4.2 Sample Rejection . . . . . 48
4.4.3 Voting System or Diagnosis Criteria . . . . . 48

**5 Performance Evaluation 51**
5.1 Dataset . . . . . 51
5.2 Classification . . . . . 53
5.3 Feature Selection . . . . . 57
5.4 Feature Importance . . . . . 59

**6 Conclusions 67**
6.1 Overall Conclusions . . . . . 67
6.2 Future Work . . . . . 68

**Bibliography 71**

**Appendices**

**A Computational Tools 87**
A.1 Python Packages . . . . . 87

**B Additional Results 89**

# List of Figures


2.1 Schematic representation of the areas of the spinal cord. . . . . 8
2.2 Schematic representation of the LMNs and the UMNs. . . . . 9
2.3 Schematic representation of multiple MUs. . . . . 10
2.4 Examples of different MUAP waveforms originated from different MUs. . . 10
2.5 Amplified SEMG and NEMG signals from the biceps brachii of the same subject. . . . . 15
2.6 Abnormalities in muscle activity. . . . . 16
2.7 Interference NEMG patterns. . . . . 17
2.8 Normal versus Neuropathic MUAP. . . . . 17
2.9 Conventional pipeline of a Machine Learning task. . . . . 19
2.10 AdaBoost scheme. . . . . 23
2.11 Examples of model fits. . . . . 24
2.12 Confusion Matrix for a binary classification task (two classes). . . . . 26

3.1 Quantitative characteristics of MUAP waveforms. . . . . 37

4.1 Schematic representing the framework's pipeline. . . . . 43
4.2 Signal processing for onset detection. . . . . 45
4.3 Schematic representation of SFS. . . . . 46
4.4 Voting system schematic, showing the contribution of each onset for the final subject's class. . . . . 49

5.1 Experimental setup showing the position of both limbs, and the Bioplux device with the four SEMG sensors and ground electrode. . . . . 52
5.2 Confusion Matrix for the subjects' classification task, when using all features. 55
5.3 Most commonly selected features, when selecting the 30 best from all features extracted from the three channels. . . . . 58
5.4 Most commonly selected features, when selecting the 30 best from all features extracted from channel 3 only. . . . . 60

5.5 Most commonly selected features, when selecting the best feature subset from all features extracted from channel 3 only. . . . . 60
5.6 Average impurity decrease of the most important features, when using all features. . . . . 61
5.7 Average accuracy decrease of the most important features, when using all features. . . . . 62
5.8 Boxplots for the peaks features. . . . . 63
5.9 Boxplots for the peak-to-peak features. . . . . 64
5.10 Boxplots for the rise time and MUAP duration features. . . . . 65
5.11 Boxplots for the MUAP integrated area, phases and turns. . . . . 66

# List of Tables


3.1 Time-domain features extracted from SEMG signal. . . . . 29
3.2 Spectral-domain features extracted from SEMG signal. . . . . 32
3.3 Statistical-domain features extracted from SEMG signal. . . . . 33
3.4 Fractal-domain features extracted from SEMG signal. . . . . 36
3.5 Description of peak-related features . . . . . 38
3.6 Description of MUAP morphology features . . . . . 38

5.1 Onset classification results obtained by using each feature group separately. 54
5.2 Subject classification results obtained by using each feature group separately. 54
5.3 Onset classification results obtained by using all feature groups and all features except the proposed ones. . . . . 54
5.4 Subject classification results obtained by using all feature groups and all features except the proposed ones. . . . . 54
5.5 Onset classification results obtained by using all feature groups before and after sample rejection. . . . . 56
5.6 Onset classification results obtained by using all features from each channel and from all channels simultaneously. . . . . 56
5.7 Onset classification results obtained by using all feature groups and all features except the proposed ones extracted from channel 3 only. . . . . 57
5.8 Onset classification results obtained by using each feature group separately extracted from channel 3 only. . . . . 57
5.9 Onset classification results obtained by using all feature groups and 30 features selected with SFS. . . . . 58
5.10 Onset classification results obtained by using all feature groups but varying parameters for the number of features and the number of channels. . . . . 59

B.1 Onset classification results obtained by each individual feature group combined with the proposed group. . . . . 90
B.2 Onset classification results obtained by duos of feature groups, excluding the proposed group. . . . . 90

B.3 Onset classification results obtained by duos of feature groups in combination with the proposed group. . . . . 90
B.4 Onset classification results obtained by trios of feature groups, excluding the proposed group. . . . . 91
B.5 Onset classification results obtained by trios of feature groups in combination with the proposed group. . . . . 91
B.6 Subject classification results obtained by each individual feature group combined with the proposed group. . . . . 92
B.7 Subject classification results obtained by duos of feature groups, excluding the proposed group. . . . . 92
B.8 Subject classification results obtained by duos of feature groups in combination with the proposed group. . . . . 92
B.9 Subject classification results obtained by trios of feature groups, excluding the proposed group. . . . . 93
B.10 Subject classification results obtained by trios of feature groups in combination with the proposed group. . . . . 93
B.11 Subject classification results obtained by using all feature groups before and after sample rejection. . . . . 94
B.12 Subject classification results obtained by using all features from each channel and from all channels simultaneously. . . . . 94
B.13 Subject classification results obtained by using all feature groups and all features except the proposed ones extracted from channel 3 only. . . . . 94
B.14 Subject classification results obtained by using each feature group separately extracted from channel 3 only. . . . . 94
B.15 Subject classification results obtained by using all feature groups and 30 features selected with SFS. . . . . 94
B.16 Subject classification results obtained by using all feature groups but varying parameters for the number of features and the number of channels. . . . . 95

# Abbreviations

| | |
|---|---|
| **AI** | Artificial Intelligence |
| **ALS** | Amyotrophic Lateral Sclerosis |
| **ALSFRS-R** | Revised ALS Functional Rating Scale |
| **AUC** | Area Under the ROC Curve |
| **CNS** | Central Nervous System |
| **CWT** | Continuous Wavelet Transform |
| **DFA** | Detrended Fluctuation Analysis |
| **ECDF** | Empirical Cumulative Distribution Function |
| **ECG** | Electrocardiogram |
| **EMG** | Electromyography |
| **fALS** | Familial Amyotrophic Lateral Sclerosis |
| **FD** | Fractal Dimension |
| **FFT** | Fast Fourier Transform |
| **FP** | Fasciculation Potential |
| **FVC** | Forced Vital Capacity |
| **HC** | Healthy Control |
| **HDSEMG** | High-density Surface Electromyography |
| **HFD** | Higuchi's Fractal Dimension |
| **LMN** | Lower Motor Neuron |
| **LZ** | Lempel-Ziv |
| **MFL** | Maximum Fractal Length |

| | |
|---|---|
| **ML** | Machine Learning |
| **MND** | Motor Neurone Disease |
| **ms** | Milliseconds |
| **MSE** | Multiscale entropy |
| **MU** | Motor Unit |
| **MUAP** | Motor Unit Action Potential |
| **mV** | Millivolts |
| **NEMG** | Needle EMG |
| **PSD** | Power Spectral Density |
| **SAMME** | Stagewise Additive Modeling using a Multi-class Exponential Loss Function |
| **SBS** | Sequential Backward Selection |
| **SEMG** | Surface EMG |
| **SFS** | Sequential Feature Selection |
| **TKEO** | Teager–Kaiser Energy Operator |
| **UMN** | Upper Motor Neuron |

# 1 Introduction

## 1.1 Context

Amyotrophic Lateral Sclerosis (ALS) is a chronic neurodegenerative disease with no cure that leads to muscle atrophy due to progressive loss of motor neurons [2]–[5]. According to a review on epidemiological studies proposed by [6], there has been an increasing number of patients diagnosed with ALS over the last years, with an approximate incidence between 2.1 to 3.8 per 100 000 person-years in Europe. Symptoms such as atrophy, weakness, and fasciculations usually begin in one limb and tend to spread to the others. However, the presentation of the disease can be very distinctive for each patient [7]. Gradually, the condition deteriorates up to a severely disabled state [5], leading to an average survival from the onset of symptoms of approximately three years, with respiratory failure as the major cause of death [5], [7].

Since it has not yet been found one established biomarker for the disease diagnosis, this is primarily achieved through clinical rather than medical examination, which means that the process of identifying the disease is based mostly on symptoms and signs and less in laboratory reports and medical exams. However, rarely does this diagnosis routine meets established criteria for a definite diagnosis of ALS. The authors of [8] stress the significance of clinical neurophysiological examination, which can uncover the involvement of body regions not initially regarded as affected by the disease. One of the most frequently performed exams as part of the medical examination is Electromyography (EMG), more specifically concentric Needle EMG (NEMG) [7]. This exam accurately measures the electric potentials generated by muscular cells.

Artificial Intelligence (AI) is a growing field that has numerous open opportunities in the most different domains. Its growth was mostly powered by the increasing amount of data collection and computational power [9]. With large quantities of acquired data as input, AI systems are able to "learn" and therefore make predictions when faced with new data. One of the many areas AI has been applied is the medical field. These systems are extremely helpful not only for disease detection and screening but also for prognosis prediction, primary care delivery, creation of drugs, management of health records, auxiliary

devices development, etc. Particularly in ALS, these methods may be useful to improve the way the disease diagnosis is performed and enhance ALS clinical studies design [10].

## 1.2 Motivation

The initial symptoms of ALS often mimic those of other neuromuscular conditions, with the clinical diagnosis taking longer to demonstrate than the medical one. This frequently leads to a delayed diagnosis, with the authors of [6] reporting that the mean diagnostic delay could go up to 24 months, which is an extremely high value for a disease with such low survival.

It is of strong relevance that more resources are allocated to ALS research, not only to speed up diagnosis but also to provide an efficient prognosis to patients and healthcare providers. This can improve the expectancy of life of these patients by anticipating the therapeutic techniques that should be applied, as well as guarantee their earlier adaption to support devices.

NEMG is used as the forefront medical examination tool in the electrodiagnosis of ALS. This exam allows for the recognition of abnormalities in the structures composing the signal related to pathological states of the muscle. On the other hand, ALS is a fast-progressing disease that demands constant re-evaluation and frequent medical examination to understand the advancement of the neuron's degeneration. NEMG is not a practical exam to include in routine medical appointments, as it is painful, and it can not be performed in an outpatient regime, since it requires professional expertise. That is one of the primary reasons for the recent growing interest in Surface EMG (SEMG) [11], a method that also measures electrical activity of skeletal muscles, but instead of using invasive electrodes, it relies on surface ones. Recently, a review on SEMG use in ALS supported its practical flexibility, due to its non-invasive nature, and potential to help determine an effective biomarker of the disease [12].

From these findings arises the need to understand the role and utility of SEMG in predicting the ALS diagnosis through AI models. This is of particular interest if relations can be established between the SEMG characteristics and the actual pathophysiology phenomena of the disease. The technique has been around for a few years. Still, the medical community has not been receptive to its use due to the often difficulty interpreting the SEMG signal correctly. However, the advancement of hardware technology and software of signal processing tools has allowed the improvement of the technique, leading to a regained interest in its use.

## 1.3 Literature Review

The EMG signal consists of multiple action potentials originated in the Motor Units (MUs) of the muscles, named Motor Unit Action Potentials (MUAPs). MUAPs are unique in shape and size for each MU. The discrimination of single MUAPs is possible and

undoubtedly relevant for anatomical and physiological analysis. Nevertheless, it often implies the use of complex techniques, which can limit its practical application in a clinical framework, and still presents numerous restrictions, such as the difficulty in separating superimposed MUAPs [13] and the SEMG decomposition strong dependency on the experimental conditions [14].

The interpretation of quantitative data from a SEMG signal can be very informative, namely through the use of time, frequency, and time-frequency features. A thorough review of such features used in various types of contexts was made in [15]. Several frequency-domain methods have been applied successfully for different purposes using SEMG, such as implementing Fast Fourier Transform (FFT) derived feature sets [16], and autoregressive coefficients as features [17]. Others have used the time-frequency distribution to characterize the variations of the spectral components of the SEMG signal [18]. Time-domain features can sometimes be preferred due to the simpler data processing methods they require [19]. Recent studies also explore the possibility of using wavelet-based approaches, which can help to remove unwanted noise and improve feature extraction [20], [21]. Additionally, fractal properties have demonstrated effectiveness when used in recognition of upper limb movements. Fractal Dimension (FD) is considered a truthful indicator of the complexity of the signal, and Detrended Fluctuation Analysis (DFA) is a well-established algorithm commonly applied in EMG analysis. More recently, Maximum Fractal Length (MFL) has been proposed as an accurate measure of the level of muscle activity [22].

Several works have been proposed in the context of ALS diagnosis using similar features and automated learning architectures to those presented above. The experimental data collection setup on these experiments is often categorized into NEMG, High-density Surface Electromyography (HDSEMG), and SEMG.

In the context of NEMG, these advances may also be relevant, especially since extracting spectral, temporal, statistical, and amplitude measures from NEMG has already proven valuable in discriminating the disease [23]–[25]. NEMG has also been used by [26] to evaluate the natural history and earliest changes in motor unit physiology of ALS patients.

The authors of [27] proposed an automated machine learning pipeline with features extracted from NEMG using *tsfresh* [28]. They used a dataset composed of 65 subjects (20 with inclusion body myositis, 20 with ALS, and 25 Healthy Controls (HCs)). Two classification strategies were designed: muscle-level, meaning the prediction is accomplished for each muscle of all subjects individually; patient-level, which relied on voting ensembles on the muscles from the same patient. They classified each subject as either diseased or healthy and achieved an Area Under the ROC Curve (AUC) score of 81.7% and 81.5% for muscle-level and patient-level, respectively. FFT coefficients ranked highest according to impurity-based importance scores.

The interest of time-frequency analysis of the NEMG signal was demonstrated in [29], where neuropathy activity was differentiated through the visualization of Continuous

Wavelet Transform (CWT). In parallel, an accuracy of 95% was obtained when using time and frequency features, such as entropy and standard deviation, for ALS discrimination.

Nevertheless, some of these techniques rely on the characteristics of the interference patterns of HDSEMG rather than on the single MU discrimination. HDSEMG is a SEMG technique that uses arrays of individual electrodes designed to record simultaneously. Some studies have addressed the potential advantages of inspecting the Fasciculation Potentials (FPs), namely by examining their amplitude and frequency [30], and by improving their detection [31]. The authors of [11] acquired data at different levels of voluntary muscle contraction from 10 ALS subjects and 11 HCs. When using a combination of statistical features and the clustering index of the HDSEMG, they were able to discriminate between the two classes with 90% sensitivity.

In the context of conventional SEMG, the analysis of parameters like coherence and phase-locking factor has shown potential in indicating motor neuron integrity in patients with ALS [32]. Classification tasks with multiple features have also shown promising results, as [33] used a set of statistical, temporal, complexity, and fractal features from SEMG recordings in the upper limbs. The authors used a dataset composed of 33 subjects (13 with ALS and 20 HCs) and tested several Machine Learning (ML) classifiers, being the decision tree, random forest, and AdaBoost, the ones that achieved the greatest performance. The authors reached an average accuracy of 77% by combining differences between features extracted from the hand and forearm recordings.

Multiscale entropy is a measure that provided great results when used as a feature for the SEMG signal in [34], aiming to differentiate healthy and pathological signals that included myopathic and neuropathic subjects. A dataset with a total of 27 subjects was used, and SEMG signals at different contraction levels were acquired. Through the use of C-support vector classification, the accuracy for the binary classification task reached 81.5%.

For further works using NEMG or SEMG in the context of ALS diagnosis, you might refer to [35], which recently conducted a systematic literature review on machine learning techniques and biomedical signals in the context of ALS. The number of works using NEMG has been higher compared to SEMG, which motivates the need for more contributions in the context of SEMG.

The quantitative measurements calculated using feature extraction from the studies identified above do not place a strong emphasis on the MUAPs morphology. Quite often, prior studies use a set of statistical, temporal, and spectral features to evaluate the EMG into a more high-level setting. Therefore, there is a gap in the literature regarding using morphological features for MUAPs, which can later be used for learning algorithms. These features would be representative of the morphology of the MUAPs based on their surface SEMG representations. Furthermore, they could more objectively depict the changes in the EMG caused by ALS: reinnervation potentials, which result in higher amplitude MUAPs; loss of MUs, which results in an increased firing rate of the active MUs; and evidence of FPs, marked by abrupt spikes.

One of the limitations of using such features would be the process of detecting each MUAP of the SEMG exam, since surface detected action potentials from different motor units are identical [36], there is a higher amount of superposition of action potentials from different MUs, and an augmented noise component [37] when compared to a NEMG signal. Detecting the MUAPs of a SEMG signal is the process of identifying and isolating the surface representations of MUAPs. Since the measured muscle has a limited number of MUs, each originating a unique MUAP, the signal can be decomposed into the different firing MUAPs, which involves first the detection of MUAPs, followed by their categorization into one of the originating MUs.

The SEMG decomposition into MUAPs is per si a challenging topic in signal processing. However, MUAP decomposition is out of the scope of this work. Although it might increase the robustness of the extracted morphology and peak-related features, we adopted a simple approach for MUAP detection. Additionally, we can extract relevant statistical information from the morphology of the surface portrayals of the MUAPs contained in the signal. Therefore, we developed an optimized detection of local maxima in the SEMG signal with the hypothesis that includes MUAPs, although we do not cluster them according to their originating MU.

## 1.4 Objectives

In this dissertation, we propose a set of morphological features for SEMG analysis. Our contributions are focused on the introduction of a signal processing pipeline to calculate the morphological features on EMG and a validation study for automated ALS diagnosis using a cohort of control and ALS patients. The use of morphological features to complement the feature sets which are typically utilized would present some advantages. For example, those features might capture complementary discriminate behaviors on the waveform as the disease progresses. Additionally, since those features are inspired by clinical interpretation, they can lead to more interpretable predictions from the classifiers.

Consequently, two research questions were identified:

- Which quantitative measurements are currently used to characterize MUAPs during NEMG and support the diagnosis criteria of ALS? Can these quantitative measurements be extracted from SEMG?
- Can the SEMG quantitative measurements be used as features for machine learning algorithms to predict the diagnosis of ALS?

The following objectives were outlined:

- Develop algorithms to calculate morphological features from MUAPs on SEMG data;

- Propose a signal processing and machine learning pipeline for automated ALS diagnosis based on SEMG;
- Validate the proposed approach using SEMG from on a cohort of control and ALS patients;
- Study the relevance of the proposed features through feature importance and feature selection techniques.

## 1.5 Thesis Overview

This document contains six chapters and two appendices. The current chapter presents the context and motivation behind the developed research, in addition to a review of the state-of-the-art and the principal objectives of the dissertation. The next chapter details the theoretical concepts that support the comprehension of this project, including a description of the ALS condition, the EMG exam, and the ML concepts. Chapters 3 and 4 specify the methodologies applied throughout the thesis. A summary of the proposed features is found in Chapter 3. Chapter 4 explains the applied approach specific to this diagnosis task in ALS using muscle signals, focusing on signal processing, feature extraction and selection, and classification algorithms. Lastly, the results and their discussion are addressed in Chapter 5, which also contains a description of the analyzed dataset. The last chapter summarizes the contributions and limitations of this work, along with some recommendations for future work. Appendix A lists the employed computational tools. Appendix B contains the complete set of results originated from multiple experiences with the classification algorithm.

# 2

# Theoretical Background

This chapter encompasses the theoretical background information regarding the themes that the dissertation will approach. First, a characterization of the nervous system and its conduction mechanisms is introduced, followed by a description of ALS. Thirdly, the medical exam of EMG will be explained, and lastly, the ML basic concepts will be discussed.

## 2.1 The Motor Unit and the Electrical Conduction of Nervous Impulses

Neurons are excitable nerve cells located in the brain or spinal cord, the two constituents of the Central Nervous System (CNS). The spinal cord has an inner core of gray matter surrounded by a layer of white matter. When seen on cross-section, the gray matter is an H-shaped pillar with anterior and posterior gray columns or horns. The white matter is divided into anterior, lateral, and posterior white columns [2].

Multiple pairs of spinal nerves are attached to the spinal cord by the anterior and posterior roots. The dendrites and axons of the neurons are often referred to as nerve fibers. Efferent fibers are nerve fibers that conduct nervous impulses from the CNS to the peripheral nervous system, muscles, and glands and can be found in the anterior root. The efferent fibers that connect to skeletal muscles are called motor fibers. Their cells of origin are located in the anterior gray horn of the spinal cord. Afferent fibers are responsible for conducting nervous impulses to the CNS and lie in the posterior root. The cell bodies of these nerve fibers are situated on the posterior root ganglion. These areas can be visualized in Figure 2.1 [2].

Neurons can be classified into different categories according to their form or function. When they are classified functionally, two types of neurons can be recognized: sensory neurons, responsible for carrying information from the receptor organs and muscles to the CNS; motor neurons, located in the anterior horn of the spinal cord and capable of carrying impulses from the CNS to the muscles and glands [2], [38]. Additionally, interneurons exist throughout the grey matter of the spinal cord, which are smaller neurons

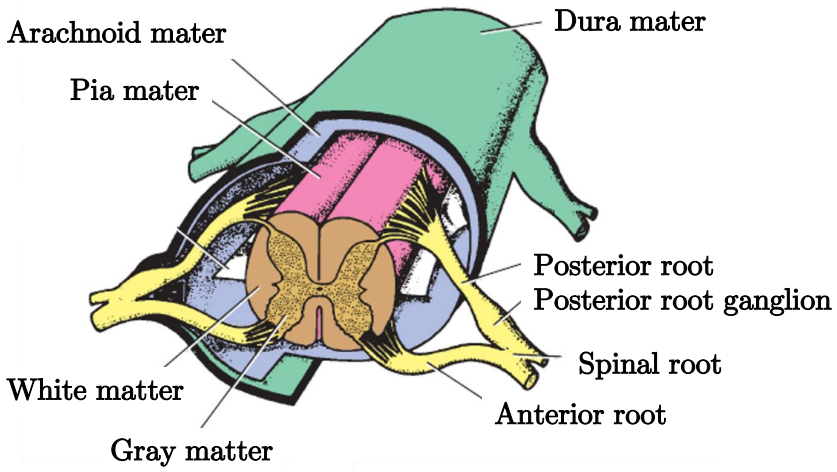



Figure 2.1: Schematic representation of the regions of the spinal cord. Adapted from [2].

responsible for connecting different types of neurons [38].

The motor neurons are also divided into two types, both in the autonomic nervous system, as in the somatic nervous system. In the latter, we find Lower Motor Neurons (LMNs) and Upper Motor Neurons (UMNs). LMNs are contained in the gray horns of the spinal cord, and they innervate skeletal muscle through the anterior roots of the spinal nerves. On the other hand, UMNs, whose cell bodies are in the motor area of the cerebral cortex, are composed of supraspinal neurons and respective descending tracts. These tracts are nerve fibers from various supraspinal nerve centers that can be grouped into nerve bundles when they descend through the white matter in the spinal cord, as represented in Figure 2.2, being responsible for conducting impulses to the spinal cord from the brain. In other words, the UMNs come from the cerebral cortex and descend to relay in the motor nuclei of the cranial nerves and anterior horn cells of the spinal cord. Additionally, the LMNs arise from the cranial nerve nuclei and anterior horn cells to innervate skeletal muscles [38].

That are multiple descending tracts, and each is related to a different function. The corticospinal tracts are of particular importance in ALS since these pathways are concerned with voluntary and skilled movements of the distal parts of the limbs [2].

The LMN is seen as "the final common pathway" to the muscles [2]. Most of these neurons are large, multipolar, and supply the extrafusal skeletal muscle fibers responsible for muscle contraction. The remaining are equally multipolar, smaller, and connected to the intrafusal muscle fibers, which have a sensory function [39]. The first ones are called alpha motor neurons and the second ones are gamma motor neurons.

The long tubular extensions of these neurons are called axons, which branch out to multiple muscle fibers, forming neuromuscular junctions where their terminals connect with the muscle fiber. Each muscle fiber has only one neuromuscular junction, although each neuron innervates more than one fiber. The muscle fibers supplied by the same

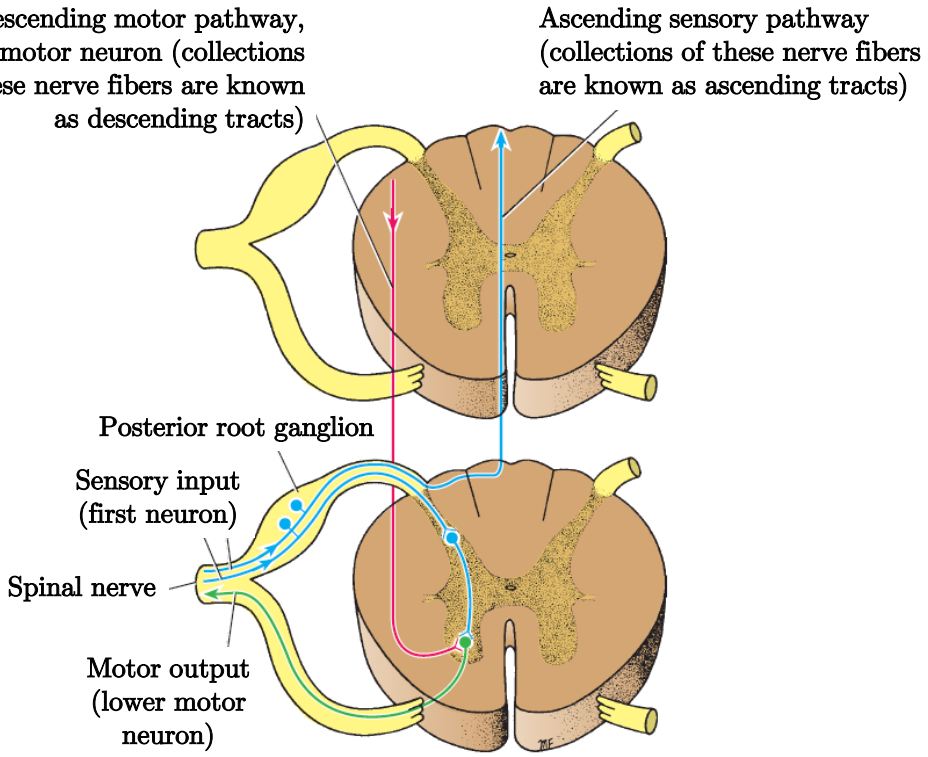



Figure 2.2: Schematic representation of the LMNs and the UMNs. Adapted from [2].

neuron are spread throughout the muscle, assuring that if one neuron degenerates, others will be able to compensate and make the entire muscle contract anyway [40].

The MU is an anatomical and functional element of the neuromuscular system [5], since it can be described as a single alpha motor neuron and all of the muscle fibers that it supplies [2], [38], [40], as illustrated by Figure 2.3. The size of the MU depends on the muscles it innervates. If a muscle is required for precise movements, it tends to have smaller MUs, with fewer muscle fibers. The opposite happens with big, weight-bearing muscles [38], [40].

The nervous impulses are conducted through the afferent nerve fibers to the spinal cord, where they make synaptic connections with motor neurons [2]. The depolarization of the motor neuron generates an action potential that is transmitted through the axon, or motor fiber, to as many muscle fibers as that particular motor neuron is connected to. From there, the action potential propagates in all directions from the neuromuscular junctions, resulting in the contraction of all the muscle fibers of that specific MU [40].

The contraction of a muscle is achieved with the activation of an increasing number of MUs of that muscle with a simultaneous reduction of activity of antagonizing muscles' MUs [2]. When a muscle initiates its contraction, smaller MUs are recruited since these present lower thresholds of excitability. With increasing intensity, firstly MUs fire at a higher rate and then progressively larger MUs are recruited, which is known as a normal

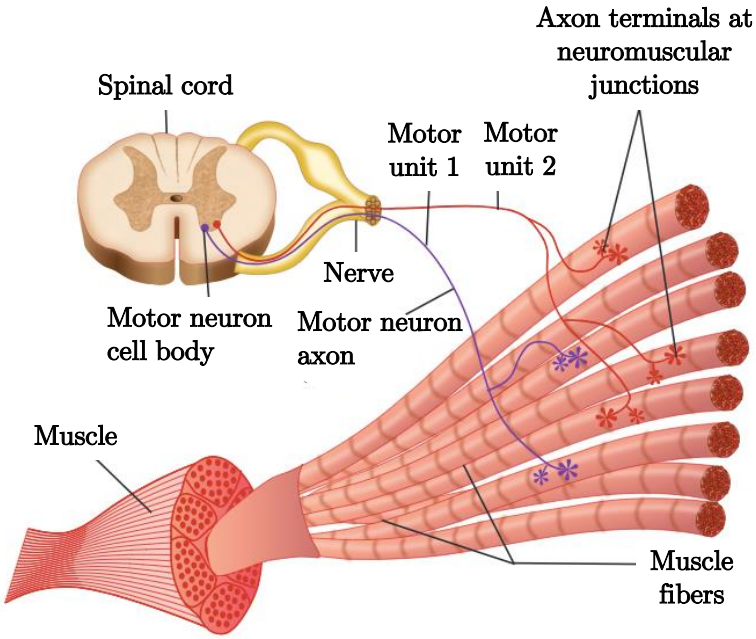



Figure 2.3: Schematic representation of multiple MUs. Adapted from [40].

recruitment pattern [2], [5].

The summation of all the muscle fiber action potentials of one MU results in a MUAP [5], which is distinguishable in shape and size for each MU [41], [42], as can be seen in Figure 2.4.

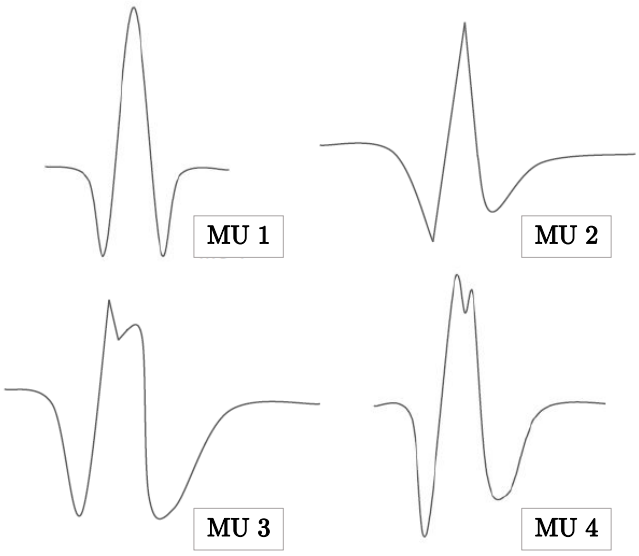



Figure 2.4: Examples of different MUAP waveforms originated from different MUs.

## 2.2 Amyotrophic Lateral Sclerosis (ALS)

### 2.2.1 Pathology

ALS is a neurodegenerative acquired disorder that leads to muscle atrophy and paralysis due to dysfunction of the somatic muscles of the body. This disease is characterized by: 1) a progressive loss of UMNs; 2) a progressive loss of LMNs [4], [7], [43]; 3) hypertrophy of glial cells in the areas of degeneration of neurons, particularly in the motor cortex and spinal cord [3].

Since the motor neuron undergoes apoptosis, its axon degenerates. This provokes the destruction of the neuromuscular junction, leading to denervation and subsequent atrophy of the muscle fibers supplied by that axon or motor fiber.

ALS is the most common form of a broader spectrum of disorders, called Motor Neurone Diseases (MNDs), all of them characterized by the progressive degeneration of motor neurons [3], [43].

The molecular pathway that leads to the destruction of these neurons is unknown; however, it is likely similar to that of other neurodegenerative illnesses [4], [44].

### 2.2.2 Causes and Risk Factors

It has not yet been found a cause for ALS [2]. Most cases are considered "sporadic," with no underlying genetic or external cause, and 5 to 10% hereditary, related to Familial Amyotrophic Lateral Sclerosis (fALS), an autosomal dominant disease caused by mutations in specific genes [4]. However, sporadic ALS and fALS are indistinguishable both phenotypically and pathologically [3], [43].

The disorder can manifest itself at any period during adulthood; however, it is more common to be diagnosed between 54 and 69 years, with a peak age of ALS onset between 51 and 66 years, as shown in a recent epidemiological analysis of ALS [6].

Even though the etiology of the disease remains unknown, a few risk factors have been identified. These include male, older age, family history of ALS, exposure to insecticides, and smoking [6], [43].

### 2.2.3 Symptoms

Symptoms such as atrophy, weakness, and fasciculations usually begin in one limb and tend to spread to the others, leading to the primary form of the disease or spinal onset [3], [5]. These signs can manifest either distally or proximally and asymmetrically in the upper and lower limbs. Bulbar onset tends to result in speaking and swallowing difficulties [6] and limitations in the movements of the tongue and face, leading to dysarthria (speech dysfunction) [4], [43]. Approximately two-thirds of patients present the spinal form of the disease [4], [6], however the presentation of the disease can be very distinctive for each patient [7]. Most patients with spinal onset go on to develop bulbar symptoms

and eventually respiratory symptoms within 1–2 years. About 5% of patients present respiratory issues without significant limb or bulbar signs [4], [43].

The loss of the UMNs results in specific signs on neurological examination, such as: Babinski's sign, which is the extension of the great toe on plantar stimulation; pathological hyperreflexia that leads to loss of dexterity; muscle weakness, which is proven through Hoffman's sign (flexion and adduction of the thumb after flicking the fingernail); stiffness detected with the jaw jerk reflex; and clonus (involuntary muscular contractions and relaxations) [4], [43].

The disappearance of LMNs leads to muscle atrophy and weakness, FPs (involuntary muscle twitching), and cramps detected on muscle biopsy [4], [7], [43].

Approximately half of the patients show a component of frontotemporal dementia characterized by symptoms such as personality changes, abnormal eating and hygiene habits, and language dysfunction.

Other less frequent symptoms may be reported, such as ocular motility disturbance in the late stages of the disease [43].

### 2.2.4 Pathophysiology

The molecular pathway that leads to motor neurons degeneration is unknown; however, this pathway is likely similar to that of other neurodegenerative illnesses. Usually, multiple pathogenic cellular mechanisms are intertwined, and these may be genetic factors, excitotoxicity (neuronal injury due to excessive glutamate), oxidative stress, mitochondrial dysfunction, impaired axonal transport, neurofilament aggregation, protein aggregation, inflammatory dysfunction and contribution of non-neuronal cells, deficits in neurotrophic factors and dysfunction of signaling pathways [4], [44]. In addition, the cell death process in ALS is thought to be apoptosis.

### 2.2.5 Diagnosis

ALS is diagnosed through clinical history and medical exams capable of detecting ALS signs, including EMG, muscle biopsy, and neurological examination since a valid diagnostic biomarker has not been established so far. However, a recent study indicated cortical hyperexcitability as a possible diagnostic biomarker for ALS [44].

However, these results alone are not enough for a concrete diagnosis. According to the Revised El Escorial criteria [45], diagnosis is only achieved after: 1) evidence of both LMN and UMN degeneration by clinical, electrophysiological or neuropathologic examination; 2) progressive appearance of symptoms within a region or to other regions; 3) absence of electrophysiological or pathological, and neuroimaging evidence of other disease processes. *Definite ALS* diagnosis is achieved when a patient manifests UMN and LMN signs in three regions of the neuraxis (cervical, thoracic, lumbosacral, and bulbar). *Probable ALS* is diagnosed when two of these regions are implicated. The diagnosis is *Probable ALS - Laboratory supported* when there are clinical signs of UMN and LMN dysfunction

in only one region, or only UMN signs in one region and LMN signs in two regions by EMG criteria, complemented with neuroimaging and clinical laboratory studies to exclude other causes. When patients present signs of UMN and LMN degeneration in one region or UMN signs in at least two regions or LMN signs rostral to UMN signs, the diagnosis is *Possible ALS*. The variability of possible diagnoses demonstrates the difficulty in achieving a final certain diagnosis for the disease [5], [45].

Diagnostic tests are performed to exclude other possibilities in the differential diagnosis, like blood tests, genetic tests when family history suggests a familial disorder, neuroimaging tests are usually performed to exclude other conditions that might superficially mimic ALS, pulmonary function tests, etc. [4]

The various clinical phenotypes of MNDs can be grossly classified depending on the area of degeneration of motor neurons, the degeneration intensity level, and the pattern of onset [3]. Typical ALS involves simultaneous UMN and LMN signs and progressive spread of symptoms over time and space [3], [4]. Other syndromes included in MNDs are Primary Lateral Sclerosis (PLS), Progressive Muscular Atrophy (PMA), Bulbar onset ALS, or Progressive Bulbar Palsy (PBP), and Pseudobulbar Palsy [3], [4], [7].

Unfortunately, initial misdiagnoses are common, which further delay the correct diagnosis of ALS, mainly because the initial symptoms are not specific to the disease [4].

### 2.2.6 Prognosis and Treatment

Respiratory complications, such as respiratory muscle failure and aspiration pneumonia, are the usual cause of death in ALS. Mean survival from onset of symptoms is approximately two to five years; however, 10% of patients present a slow form of the disease and can survive beyond ten years [4], [6], [43].

Treatment for ALS is palliative, multidisciplinary, and based on symptom management, intended to prolong survival and optimize healthcare delivery [4], [43].

The only medicine that has shown the potential to prolong survival by 2 to 3 months for ALS itself is riluzole [7], [43], [46]. Nevertheless, other pharmacologic interventions may be applied, targeting specific symptoms.

Other therapies applied include nutritional management, non-invasive ventilation, and invasive mechanical ventilation (tracheostomy) [43], [46].

Additionally, assistive equipment is of great importance for patient rehabilitation. Some devices that may be useful are cough-assist devices, percutaneous gastrostomy (PEG) tube, augmentative speech systems, and cervical and foot orthosis [46].

### 2.2.7 Disease Progression Quantification

The ALS functional rating scales are instruments for evaluating the functional status of patients with ALS. The scales track the functional change in a patient over time. The Revised ALS Functional Rating Scale (ALSFRS-R) [47] is the most widely used score in both clinical and research settings. It is achieved through a questionnaire that measures

physical function in carrying out activities of daily living (ADL) and global function of patients with ALS. Forced Vital Capacity (FVC) measurement complements the ALSFRS-R in both prediction of survival and representation of disease progression [48]. The scale group contains four domains encompassing gross motor tasks, fine motor tasks, bulbar functions, and respiratory functions [47], [49], [50].

The deterioration of ALSFRS-R over time is a marker of disease progression and can be used to predict survival. Despite the current knowledge of ALS, there is still a need to improve quantitative measures of disease progression, as they are vital for the shared medical decision-making process and for designing clinical trials.

Using a functional rating scale instead of other primary outcome measures is advantageous because it evaluates results that cannot be directly measured [51] and it is easy to administrate.

## 2.3 Electromyography (EMG)

### 2.3.1 Introduction

EMG is the recording process of the electrical activity produced by skeletal muscles. It employs electrodes on the skin's surface, also known as SEMG, or inserted into the muscle, called NEMG. A standard EMG machine is equipped with amplifiers, a control panel, and a digital computer for data sampling, storing, and processing. The EMG unit hardware might optionally offer stimulation devices to trigger responses from the nerves or muscles being studied. The amplifier is responsible for intensifying the difference between the inputs of two electrodes. One of them is the reference electrode, which is positioned in a part of the body where minimal muscle activation will happen. This differential amplification eliminates background noise, and Electrocardiogram (ECG) potential interference since these potentials are recorded by both electrodes equally, and therefore the difference between signals is zero [52].

Acquisition protocols may involve different degrees of contraction. The resulting signal is usually contaminated with noise from AC power line interference or movement artifacts, making signal processing indispensable to improve spectral resolution. Conventional filters are able to reduce noise without compromising the usable EMG signal [53].

Nowadays, the use of NEMG is standard practice for electrodiagnosis. NEMG is an invasive technique that uses a needle recording electrode inserted directly into the muscle. NEMG provides accurate potential measurements with a higher selectivity and signal-to-noise ratio, i.e., amplified signal voltage with attenuated noise, and it depicts individual MUAPs. Figure 2.5(B) depicts a normal recruitment pattern, measured with NEMG, with increasing force as displayed in Figure 2.5(A), leading to the firing of increasingly larger MUs, as explained in Section 2.1. The signal is usually expressed in millivolts. However, this invasive technique has the disadvantage of being particularly painful for

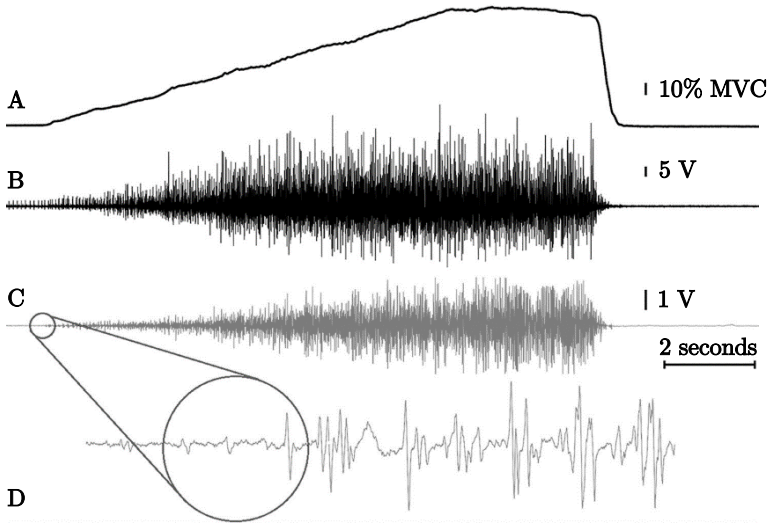



Figure 2.5: Amplified SEMG and NEMG signals from the biceps brachii of the same subject. A) Force according to the percentage of maximal voluntary contraction (MVC). B) NEMG signal. C) SEMG signal. D) Zoomed SEMG signal, where individual MUAPs are identifiable at low levels of force (approximately 0–15% MVC). Adapted from [55].

patients, which hinders its adoption from tracking disease progression in routine medical appointments, and it is a technically demanding exam. Furthermore, it can only provide a limited amount of information since NEMG electrodes imply a small detection site, so recordings are obtained from a relatively small number of motor units. Additionally, these recordings often present variability due to small movements of the electrodes with higher muscular force [41].

SEMG is a non-invasive technique that uses surface electrodes attached to the skin. This signal has the advantage of recording multiple MUs activity patterns, which are the summation of the MUAPs generated by these MUs [18]. As shown in Figure 2.5(C), the sum of smaller and slower firing MUAPs results in a smaller amplitude interference pattern in SEMG, that increases with higher contraction levels. Despite their versatility, surface electrodes usually record with a lower signal-to-noise ratio compared to NEMG. The tissues underlying the electrodes act as a low-pass filter, causing similar shapes in potentials from different motor units [36]. The signal is usually expressed in microvolts. Since surface electrodes are farther away from the muscle fibers, the recorded MUAPs are lower in amplitude and with higher probability superposition of multiple MUAPs. For this reason, SEMG does not allow for the identification of individual MUAPs, like NEMG does. These drawbacks have resulted in the long medical community's discrediting of SEMG potential in clinical practice [54]. Nevertheless, changes in the shape of surface representations of MUAPs of the interference pattern are helpful to identify possible pathological changes in MU activity patterns [55]. Exceptionally, when the level of muscle contraction is low, individual MUAPs are visible, as displayed in Figure 2.5(D).

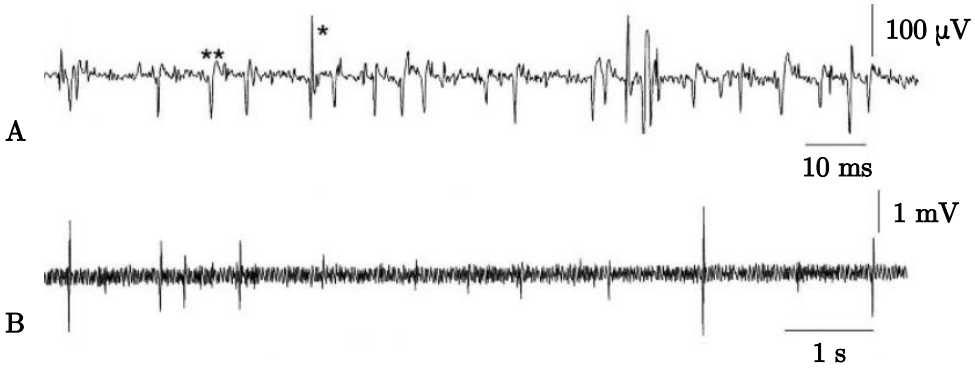



Figure 2.6: Abnormalities in muscle activity. A) Fibrillation potentials (*) and positive sharp waves (**). B) Fasciculations. Adapted from [56].

### 2.3.2 Electromyography in Amyotrophic Lateral Sclerosis

The ALS diagnostic guidelines defined in 2008 during the Awaji consensus meeting [8] support the utility of NEMG as an attempt to improve the diagnostic accuracy of the Revised El Escorial criteria [45]. As explained in Section 2.2.1, muscle fiber denervation occurs because of the degeneration of motor neurons. NEMG is able to provide electrophysiological evidence of such acute or ongoing denervation, namely FPs, fibrillation potentials, and/or positive sharp waves. Figure 2.6 shows these irregularities, which are forms of abnormal electrical activity while the muscle is at rest. Fibrillation potentials are spontaneous discharges originated from hypersensitive acutely denervated muscle fibers. Positive sharp waves usually are found in association with fibrillation potentials. Both arise from the needle insertion into the muscle fiber during NEMG, persisting for approximately 3 seconds after the needle movement has stopped. Involuntary discharges of part or the whole motor unit result in FPs. These present a variable waveform morphology, and have a irregular firing rate, much slower than that of voluntary MUAPs [56].

Another manifestation of denervation is reduced MU recruitment, defined by a higher firing rate of a lower number of MUs, which can be verified in Figure 2.7. When a patient contracts the analyzed muscle with increasing force, a regular recruitment pattern would result in a higher MU firing rate and then recruitment of more MUs, as mentioned in Section 2.1. As ALS patients have fewer MUs, the remaining ones fire more rapidly to be able to produce enough force for the contraction, which is known as a reduced recruitment pattern. Muscle reinnervation usually follows, and it leads to pathological MUs, with more muscle fibers than usual, resulting in abnormal characteristics that can be detected through NEMG, such as polyphasic, high amplitude, and long duration MUAPs [5], [8], as shown in Figure 2.8.

The latest technological advances in sensing hardware, computing capacity, and signal processing for SEMG have contributed to an increasing better acceptance of this technique in clinical practice. According to a recent systematic review by [12], SEMG offers significant practical and analytical flexibility compared to NEMG and there is a need

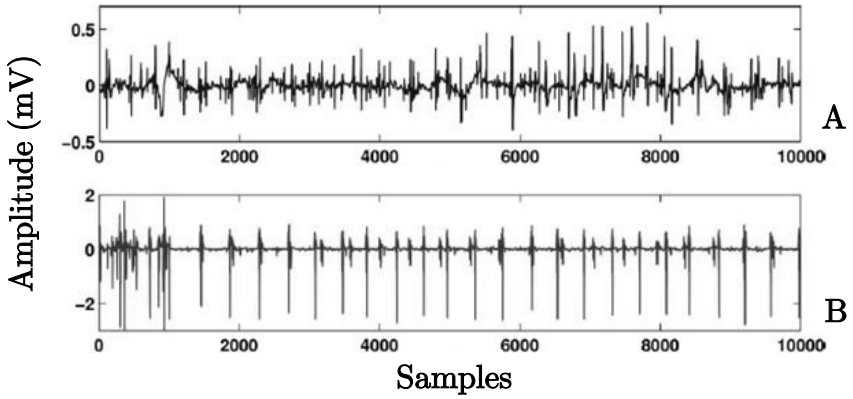



Figure 2.7: Interference NEMG patterns. A) Normal subject's signal. B) ALS subject's signal. Adapted from [57].

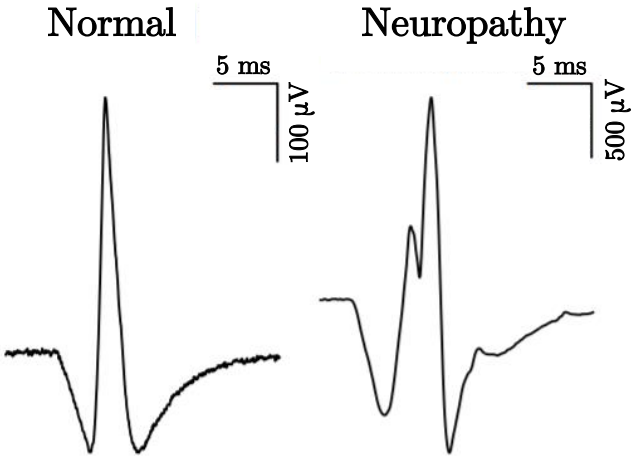



Figure 2.8: Normal versus Neuropathic MUAP. Adapted from [58].

for multi-disciplinary research collaboration on the topic. The technological advances have leverage SEMG techniques with significant advantages in the diagnosis, prognosis, monitoring and pathoetiological resolution of ALS.

## 2.4 Machine Learning

### 2.4.1 Introduction

One of the sub-fields of AI is ML, a type of computer program based on mathematical models that can learn autonomously and automatically from previous experience [59]. Experience in this context is understood as the past information that was presented to the algorithm [60]. While the performing operation needs to be specified in a traditional algorithm, ML algorithms make predictions without being explicitly programmed to perform a specific task [61]. These algorithms are given example samples associated with

the respective outcomes and find in which way the samples are mapped to the outcomes, therefore having the ability to extract information from input data by themselves and, consequently, generate accurate predictions when unseen data is presented as input [62]. Usually, the quality of an algorithm is measured through its performance, and time and space complexity. With ML algorithms, sample complexity must also be accounted for so that we know the sample size required for the algorithm to learn [60].

A table can store a dataset, and it contains all the features, $X$, and the targets to predict, $y$. Each row in the dataset is an instance or entry $i$ composed of the feature's values, $x^i$, and the respective target, $y_i$. The feature $j$ has a value for each instance $i$, $x^i_j$, and the vector with that features' values for all instances is a column in the dataset, $x_j$. The trained model is $\hat{f}$, which is a hypothesis of the real function $f$, and therefore a prediction for the instance $i$ is described by [62], [63]:

$$\hat{y_i} = \hat{f}(x^i) \tag{2.1}$$

When the targets are discrete, they can be named classes or labels.

These algorithms can be developed in two different types of scenarios. *Supervised learning* learns a rule to map inputs to outputs, and therefore the training data is a set of labeled examples. The predictor learns using a known training dataset, consequently being able to infer future outcomes. The most common supervised ML algorithms include Decision Tree, Naive Bayes, and Support Vector Machine. *Unsupervised learning* finds patterns or structures on its own, using unlabeled data. The most famous unsupervised ML algorithms include K-Means Clustering and Principal Component Analysis [64]. A combination of the *supervised* and *unsupervised* scenarios may happen, resulting in *Semi-supervised learning* [60], [61], [65], [66].

The main ML tasks, i.e., types of inferences being made, include: **classification**, where class labels are assigned to each item; **regression**, where continuous values are predicted for each item; **ranking**, where items are ordered according to some condition; **clustering**, used to group instances of data into clusters with similar characteristics; and **dimensionality reduction**, used to reduce the dimension of the items while preserving specific properties [60].

One of the most critical characteristics these systems have to integrate is the ability to generalize, which is affected by sample size and complexity. If a more complex system is chosen and the sample size is relatively small, poor generalization is obtained, also known as *overfitting*. Over-fitted models tend to memorize the training data, including unavoidable noise, instead of learning the hidden patterns behind it [67]. The opposite can also happen, leading to an algorithm with decreased accuracy [60].

The traditional ML algorithm development process is represented in Figure 2.9. Data preprocessing may include tasks such as data cleaning or dealing with missing data points. After data preprocessing, data are split in order to obtain train and test datasets. The classifier then receives the extracted features from the training dataset and is consequently evaluated with the test dataset. This process is redone, with the classifier's parameters

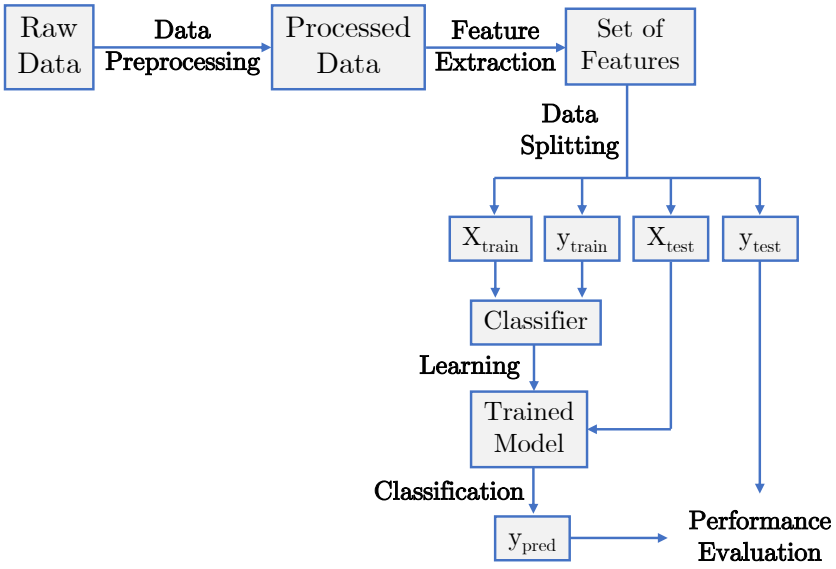



Figure 2.9: Conventional pipeline of a Machine Learning task.

being repeatedly tuned until a satisfying model is obtained, with adequate predictive accuracy. Thus this iterative process is model "learning," where the model is the learned program that links inputs to predictions. ML uses the training set to establish a regressor or classifier and assesses its performance through the test set [63].

Even if we understand the underlying mathematical principles of most AI models, some of them are applied in a black-box manner, i.e., no information is given about what in the input data made them actually reach a particular predictive outcome [68]. Deep Neural Networks are examples of black box models [9], as they may have impressive results but are always limited by their inefficiency in understandably explaining its reasoning. By not revealing their internal mechanisms, these systems prevent the user from verifying the conclusions obtained by the model [63].

### 2.4.2 Feature Extraction and Selection

Features can be extracted from multiple data types. The one used in this work is time series, which can be defined as a chronologically ordered sequence of data points taken at equally spaced time intervals [65].

Feature extraction is the process of gathering the most informative set of measurable properties of a signal. A feature vector is the most common form of data input for ML models, as these can be quantitative (discrete or continuous) or qualitative (binary or categorical) measurements [62], [66]. Feature construction is the first step of this operation, and it requires the conversion of raw data into a usable set of variables or attributes. For example, time series features often include outcomes resulting from the time-domain, frequency-domain, and time-frequency processing [65]. Of course, it is essential to be cautious not to lose any relevant information regarding the signal, but withholding an

exceedingly large number of features implies a dimensionality problem.

Feature selection can help solve this problem by finding the most compact and pertinent subset of features that contains maximum knowledge regarding the discrimination of classes or labels of the input data. This subset can still efficiently describe the data while reducing the number of redundant variables and features with no correlation to the classes, both acting as noise for the predictor. This process can help reduce data storage and computation time, improve the model's performance, and better understand input data [62].

Any feature selection strategy should account for three aspects: 1) the **evaluation criterion**, that measures the relevance level of each feature with regard to the classes; 2) the **feature subset generation**, which is the process of choosing the most effective features; 3) the **assessment method** of the outcome. Supervised feature selection can be achieved using filters, wrappers, or embedded methods briefly explained in the following paragraphs [62], [69].

The main parameter that distinguishes filters from the remaining techniques is the evaluation criterion. Filters generally use a relevance index based on correlation or statistical analysis, therefore relying on the characteristics of the input data without depending on the used learning algorithm [70], [71]. This index orders the features by relevance and filters out the variables whose ranking is below a previously defined threshold [69]. As a result of the simplicity of the ranking calculation, filter methods tend to require fewer computational costs, and their independence from the predictor allows for better generalization and limited *overfitting* to the data [70]. However, it does not account for the fact that features are not independent of each other. Therefore features can be individually irrelevant but more informative when applied with others; or relevant on their own but redundant when combined with the rest [62]. Calculating the rejection threshold and consequently choosing the number of selected features can also be an issue [69]. Some examples of filter methods include Pearson correlation coefficient and Relief.

Whereas filters do not involve any learning by the model, wrappers use a learning model trained with a candidate feature subset to evaluate the performance of the given subset [62], [72]. Wrappers are based on search strategy when they select the group of features that maximizes the model's classification performance [69], [72]. According to [62], the search strategy can be optimal, stochastic or sequential. The first one is an exhaustive search evaluating all possible subsets, therefore having a high computational cost. Stochastic methods iteratively update the candidate subsets through a certain mechanism according to a given optimization function. Lastly, sequential selection is an iterative method that can be applied in two manners. Sequential Feature Selection (SFS) is when initially there is an empty set and the first added feature is the one that maximizes the value of the optimization function. The process is repeated until the performance of the predictor can no longer be improved by the addition of any feature or until the desired feature subset size is reached. The same process performed backwards is Sequential Backward Selection (SBS), in fact, it starts with all available features and removes the feature

that leads to the best results [62], [69], [72]. Wrapper methods usually perform better than filters since the selection process is optimized for a specific predictor [73], [74].

Hybrid methods apply a wrapper approach to a subset of features generated from the filter's ranking, which is particularly indispensable when the number of features is very high [70], [71].

We have seen that filters do not incorporate model learning. Wrappers require the re-computation of a model for each subset of features but do not assimilate information about the classification's specific structure. On the other side, embedded methods do not segregate the learning from the feature selection part by incorporating the search for an optimal group of features as part of the model's training process [62], [75].

### 2.4.3 Classification

The classification process happens after the algorithm is trained with the extracted features, enabling the model to classify new data into one of the possible classes. Before doing so, one must choose the appropriate algorithm depending on the used type of data and the research goals. Many approaches have been developed in a supervised ML context based on logic, perceptrons, instances, or bayesian network [66]. This section will focus solely on supervised ML approaches used during our work, including Random Forest Classifier and AdaBoost with Random Forest base. A brief explanation for each classifier follows:

- **Decision Tree** describes possible decisions to be made, the options for those decisions, and the outcomes of different combinations of decisions and options. Each node in a tree represents a decision point related to a feature, and each branch is a possible value for that feature. The root node should be associated with the feature that best discriminates training data. From there, multiple splits or partitions are created at each node according to a range of criteria, like minimizing classification error or gain in impurity [76], [77]. Subsequently, instances are sorted based on their features values until they reach the terminal node that attributes them to a class [65], [66]. Its highly interpretable graphics, easy comprehension, and straightforward implementation make it one of the most frequently used algorithms [76], [78].

- **Random Forest** creates multiple Decision Trees by training each of them with a different subset of the training data. It uses bootstrapping resampling by repeatedly drawing a smaller set of samples from the training data. The set is restored in the original training data after the creation of each tree. It is considered an ensemble method since multiple models, in this case, individual Decision Trees, are combined into one single Random Forest during the learning procedure. Tree construction is slightly different when compared to what happens in a Decision Tree since the Decision Tree chooses the best split considering all features, and Random Forest

selects the best partition among a random subset of features. In the end, the several trees are averaged together, which is extremely advantageous since Decision Trees are more prone to overfitting. This way, it is possible to obtain more accurate predictions [65], [76]. Furthermore, Random Forest takes into account parameters related to the trees themselves, like the number of random features used for each tree, and the tree depth, which is the longest path between the root node and the terminal node; and it also considers the number of samples for the bootstrapping process and the number of trees in the forest [77], [79].

- **AdaBoost** is also an ensemble method but based on the boosting technique, which is the concept of creating a strong classifier by combining many relatively weak classifiers [80], as it can be verified in Figure 2.10. This algorithm originates from an iterative operation that is initially trained with the unweighted training set. Then AdaBoost builds a classifier, for example, a Decision Tree, that predicts class labels. These results are weighted in the following manner: if a training instance is wrongly classified, the weight of that point is boosted or increased; if is correctly predicted, its weight is decreased. A second classifier is subsequently built, using the same samples but with the new weights, and the procedure is repeated with sample weights being adjusted in each iteration. The aim of this is to oblige the weak classifier to focus on the more complex cases that were missed by the preceding weak classifier. The final step is to compute the classifier as a weighted majority vote of the weak classifiers since a score is assigned to each of them [81], [82].

A ML pipeline may integrate optimization techniques in an effort to enhance the classifier's performance. Grid Search Cross-Validation is an example of such, as it can be applied to determine the best parameters for the estimator before it performs any classification tasks. This is achieved through an exhaustive search over the list of possible parameter values of that particular classifier. The performance of every combination of parameters is evaluated using a performance score, and the parameters that maximize that score are selected [83]. In this case, the performance metrics are the results of Cross-Validation, which will be explained further in Section 2.4.4. Parameter adjustment can be time-consuming, especially when the learning algorithm has numerous parameters [84], [85].

### 2.4.4 Validation and Performance Metrics

A model should be trained with a different set of samples than the ones used for its testing. Otherwise, it would obtain a perfect score because it would have seen all testing samples during the train. The model needs to be able to generalize well to new data, so it is important to implement techniques that prevent *overfitting*. At the same time, the results we obtain can vary a lot depending on the sample division that is made, and simpler models can fall into the tendency of *underfitting* by not capturing well the linkage

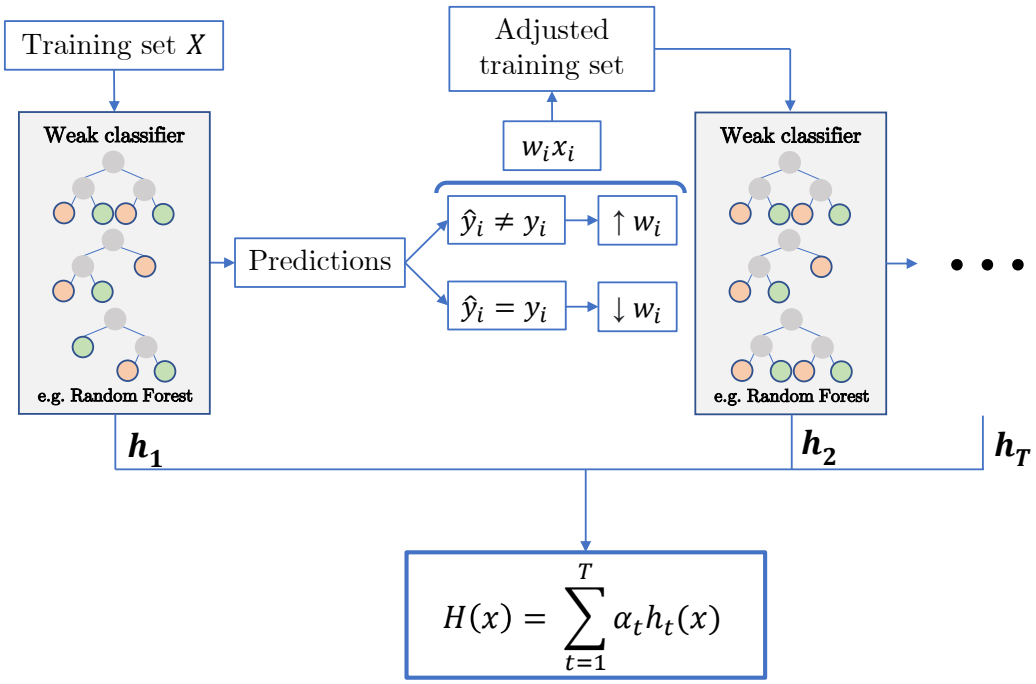



Figure 2.10: AdaBoost scheme. $x_i$ is a sample of the feature dataset, $y_i$ is the real class of the sample, $\hat{y}_i$ is the predicted class of the sample, $w$ represents the weights assigned to each sample, $h$ represents the trained weak classifier, and $\alpha$ represents the weight assigned to each weak classifier.

between variables (Figure 2.11). That is why standard validation techniques include data resampling like Cross-Validation, whose variations depend on how the division of training and test sets is accomplished. A brief explanation for a few of these variations follows:

- **K-Fold Cross-Validation** divides the dataset into $k$ disjoint subsets or folds of approximately equal size. The model is then trained with $k$ - 1 subsets, which together originate the training set and is tested with the remaining subset. This process is repeated $k$ times until all subsets have been used as test sets. If the dataset is imbalanced, Cross-Validation can include *stratified* random sampling, guaranteeing that the class proportions in the individual subsets reflect the actual proportions present in the dataset [87], [88].

- **Leave-One-Out Cross-Validation** is an exceptional situation of K-Fold Cross-Validation when $k$ equals the total number of samples in the dataset. In this case, each sample is an individual subset, so what happens is that the model trains with all available samples except one and is tested with the remaining sample [88]. Typically, this type of validation is only used when the dataset is small.

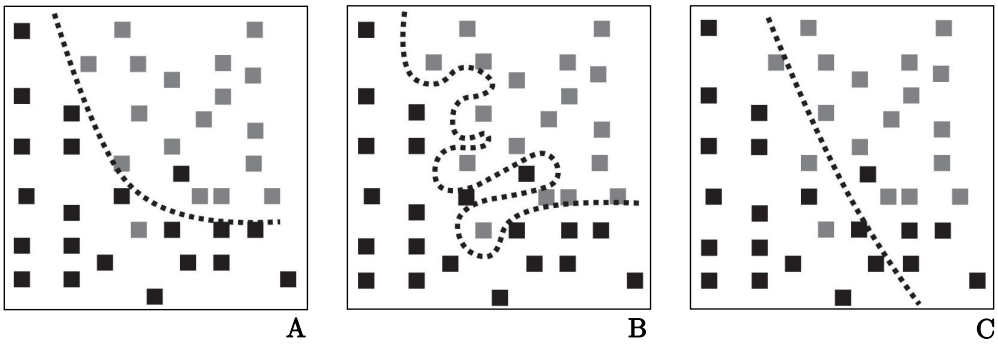



Figure 2.11: Examples of model fits. The squares represent data points, with each color representing a different class. The model is presented as the data separation curve. A) Adequate model fit. B) Overfitted model. C) Underfitted model. Source: [86].

- **Shuffle-Group(s)-Out Cross-Validation** splits the dataset according to an external group, which contains information regarding specific stratifications of the samples. For example, when a study involves multiple acquisitions by the same subjects, it can be significant to allocate all samples of one subject either on the training set or on the test set [85], [89].

- For **Leave-One-Group-Out Cross-Validation**, each subset separated by the iterator is composed of all samples associated with one group. Thus, for the given example, the model trains using all instances of all subjects except one and tests on the samples of the remaining subject.

When feature selection is part of the ML pipeline, the selected features must be determined using only the training set and not the entire dataset so that the produced performance estimate is not biased. Another application for Cross-Validation is model parameter optimization, by performing Cross-Validation for different values of the tuning parameter, and choosing the one that minimizes the cross-validated error.

Finally, the performance of the final model is estimated using the average performance of the models on each subset. The model's overall performance can be visually interpreted using a confusion matrix, which indicates the true labels in the rows against the predicted ones in the columns, as represented in Figure 2.12. This square matrix depicts the raw results about the classifiers' predictions as it reports the number of True Positives ($T_P$), True Negatives ($T_N$), False Positives ($F_P$) and False Negatives ($F_N$) [90]. These values can quantify different metrics, the most common being:

- **Accuracy** is the fraction of correct predictions to the total of test samples.

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{2.2}$$

- **Precision** assesses the overall positive predictions, comparing the true positive cases to all the predicted positives, both correctly and not.

$$Precision = \frac{T_P}{T_P + F_P} \tag{2.3}$$

- **Recall** relates the positive predictions to the total of true positive cases.

$$Recall = \frac{T_P}{T_P + F_N} \tag{2.4}$$

- $\mathbf{F_1}$ **score** is the harmonic mean of the model's precision and recall.

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2.5}$$

The $F_1$, precision and recall scores measure how effective the model is when classifying the positive class. When we intend to take into account the performance of the model when classifying both classes in a binary classification problem, we can apply a macro-average. This average computes the metric independently for each class and then calculates their arithmetic mean. Macro-averaging ensures that all classes get equal weight when contributing to the metric scores. This might not be an appropriate calculation when dealing with class imbalance.

The choice of the proper evaluation metric depends on the provided dataset and the goals of the classification task. A poorly selected metric may lead to fallacious conclusions. This can have even more significant repercussions in a medical scenario since we have to consider that it is probably better for a model to incorrectly identify a disease as being present when it is not (False Positive) than the opposite (False Negative) [86]. For this purpose, recall would be an adequate measure since it informs about the chance the model has of correctly identifying the disease. Another example is regarding accuracy, which is not an appropriate measure when dealing with class imbalance since it is possible to achieve a high score even if the model performs poorly in the less probable class [91]. For the purpose of analyzing imbalanced datasets, $F_1$ score may be a more appropriate metric since it encompasses both precision and recall.

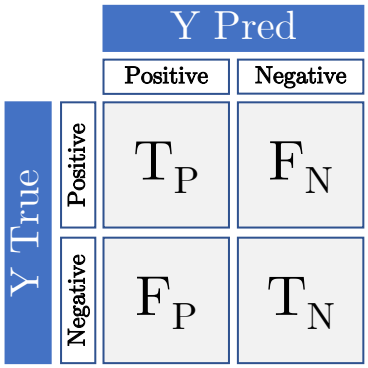



Figure 2.12: Confusion Matrix for a binary classification task (two classes). The counts from predicted values for each class are on the columns, and the counts from actual values on the rows.

# 3

# Novel Morphological Features for Surface Electromyography Characterization

The right features can make a huge difference in a model's predictive power. As stated by Professor Pedro Domingos, "At the end of the day, some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used" [92]. In the present work, a set of state-of-the-art EMG features was combined with a proposed novel set of features designed to take into consideration the quantitative measurements taken by clinicians to diagnose neuropathies.

This chapter encompasses the description of all considered features, along with a more thorough explanation of the implemented methods to extract the proposed features.

## 3.1 State-of-the-art SEMG features

An extensive group of features has been previously used in EMG studies for different purposes, including muscle fatigue analysis, movement classification, and disease classification. In this work, these features were explored to understand their discriminating value in SEMG for ALS diagnosis and how the novel features could improve that. Among the state-of-the-art features, some have reported high computational efficiency and excellent classification performance for diagnosis tasks.

### 3.1.1 Time-Domain Features

The Table 3.1 presents the time-domain features that were extracted from the SEMG signal. Most of these features are objective signal parameters that several research projects have explored [93]–[103]. Most time-domain features are computed using the raw EMG time series, so their implementation is typically straightforward and does not require any significant transformation. Compared to other features in the frequency domain, this group of features has the potential to provide better classification performance in low

noise environments and computational simplicity. However, the raw recording usually contains much interference that can be a disadvantage for these calculations, especially for features related to energy properties [94], [95].

Features that integrate full-wave rectification help retain the energy content of the signal, such as the absolute value of the summation of square root and the integrated EMG [93]. The latter is commonly used for onset detection since it closely relates to the signal's sequence firing point [94]. The absolute value of the summation of the exponential root provides an approximate measure of the power of the SEMG, giving a better perception of the amplitude of the signal [93]. An estimate of the total amount of activity in one time window can be obtained with the mean value of the square root.

The V-Order and log detector implicitly estimate muscle contraction force, and Willison amplitude gives an approximate estimate of the number of active MUs, which is also related to the muscle contraction levels [101].

Waveform length is a measure of the waveform complexity when the muscle is active [103]. The average amplitude change and difference absolute standard deviation values are the waveform length's average and standard deviation, respectively [94]. Thus, these features contribute with information regarding the frequency, duration, and waveform amplitude. An enhanced version of this feature has been introduced by [99], which attributes higher weights to values in the center of the signal, where more valuable information can be found.

Zero crossing rate, myopulse percentage rate, Willison amplitude, and slope sign change provide frequency content about the signal defined in the time-domain since these features quantify or average how many times a particular parameter exceeds a predefined threshold. The computation associated with a threshold value intends to attenuate the effect of background noise. Cardinality also uses a threshold after data sorting, and the consecutive values are only considered distinct if their difference is above that threshold [94].

Auto-regressive coefficients are obtained through a prediction model that expresses the signal as a combination of past samples and a white noise term. The coefficients were computed until the fourth order as suggested in [94], resulting in four variables so four features. This model gives information regarding the frequency distribution of the signal.

By applying the logarithmic transform to the summation of the Teager–Kaiser Energy Operator (TKEO) in all signal windows, the TKEO information was included in the feature set, which is an energy operator that measures instantaneous frequencies of the signal and estimates the time-varying amplitude envelope [100].

The mean absolute value slope is the difference between the average value of adjacent segments covering the signal. The number of segments considered in this study was five, leading to five features [103].

A total of 48 time-domain features was considered.

Table 3.1: Time-domain features extracted from SEMG signal.

| Feature | Description |
|---|---|
| **Absolute energy** | Sum over the squared values of the signal |
| **Absolute value of the summation of exponential root** | An approximate measure of the power of the signal |
| **Absolute value of the summation of square root** | Absolute value of the sum of the square root of all values of the signal |
| **Area under the curve** | Magnitude of the product of the intensity of the signal and time |
| **Autocorrelation** | Correlation of the signal with a delayed version of itself |
| **Autoregressive coefficients** | Coefficients in the regression equation, that models the signal a linear autoregressive time series |
| **Average amplitude change** | Average of the waveform length |
| **Cardinality** | Total number of distinct values |
| **Centroid** | Arithmetic average position of all values along the time axis |
| **Difference absolute mean value** | Average of the absolute difference between consecutive values of the signal |
| **Difference absolute standard deviation value** | Standard deviation absolute value of the difference between consecutive values of the signal |
| **Difference variance value** | Variance value of the difference between consecutive values of the signal |
| **Enhanced wavelength** | Waveform length with an additional parameter weighting each value |
| **Entropy** | Entropy of the signal using the Shannon Entropy |
| **Integrated EMG** | Summation of the absolute values of the signal |
| **Log detector** | A logarithmic estimate of muscle contraction force defined from a mathematical model of the signal generation |
| **Log difference absolute standard deviation** | Logarithmic transformation of the difference absolute standard deviation value |
| **Log difference absolute mean value** | Logarithmic transformation of the difference absolute mean value |
| **Log Teager-Kaiser energy operator** | Logarithmic transformation of the TKEO |

Table 3.1: Time-domain features extracted from SEMG signal (continued).

| Feature | Description |
|---|---|
| **Mean absolute difference** | Average absolute difference of two random and independent values from the same probability distribution |
| **Mean difference** | Average difference between consecutive values of the signal |
| **Mean absolute value slope** | Differences between the mean absolute values of consecutive segments |
| **Mean value of square root** | A measure that gives an estimate of the total amount of activity in the signal, given by the average of the square root of all the values of the signal |
| **Median absolute difference** | Median of the absolute differences between consecutive values of the signal |
| **Median difference** | Median of the differences between consecutive values of the signal |
| **Myopulse percentage rate** | Average of the myopulse output, which is considered one when the absolute value of the signal exceeds a threshold and zero otherwise |
| **Negative turning points** | Number of minimum points, where the slope changes sign from negative to positive |
| **New zero crossing** | Total number of times the signal crosses a predefined amplitude level |
| **Neighbourhood peaks** | Number of peaks from a limited neighbourhood of the signal |
| **Peak-to-peak distance** | Distance between maximum and minimum peaks |
| **Positive turning points** | Number of maximum points, where the slope changes sign from positive to negative |
| **Signal distance** | Total distance traveled by the signal |
| **Slope** | Slope by fitting a linear equation to the signal |
| **Slope sign change** | Total number of times the slope of the signal changes sign associated with a threshold function |
| **Sum absolute difference** | Summation of the absolute differences between consecutive values of the signal |
| **Temporal moment** | The absolute value of the higher-order temporal moments, which are characteristic of the time-dependent response curve |
| **Total energy** | Area under the squared magnitude of the signal |

Table 3.1: Time-domain features extracted from SEMG signal (continued).

| Feature | Description |
|---|---|
| **V-Order** | An estimate of muscle contraction force defined from a mathematical model of the signal generation |
| **Waveform length** | A measure of the fluctuations of the signal described as the cumulative length of the signal's waveform |
| **Willison amplitude** | Total number of times the amplitude difference between two consecutive segments of the signal exceeds a threshold |
| **Zero crossing rate** | Total number of times the signal crosses the zero amplitude level and the difference between consecutive points exceeds a threshold |

### 3.1.2 Spectral-Domain Features

Frequency or spectral-domain features are commonly used to study the fatigue of the muscle and MU recruitment. The complete list of used spectral features can be found in Table 3.2.

The Power Spectral Density (PSD) refers to the spectral energy distribution as a function of frequency, giving an idea about the frequencies that compose the signal. It is obtained through the FFT of the autocorrelation function of the signal [104]. PSD is an essential tool for signal analysis in the frequency-domain, and a few explicit properties from this spectrum can be extracted, such as its maximum, bandwidth, centroid, roll-off, and roll-on points. Others are obtained indirectly, such as the amount of decrease, entropy, kurtosis, skewness, slope, spread, and variation [105].

Cepstrum coefficients inform about the rate changes and periodic structures in different frequency spectrum bands of the signal. Even though they are more commonly used for speech-related tasks, several studies have shown their relevance in the EMG domain [101], [106]. The first 12 coefficients were extracted, since these contain most of the information.

A number of features can be obtained after applying the CWT, which performs the convolution of the signal with a set of wavelets at different scales, in this case nine scales, allowing for variable time-frequency resolution. Wavelets are mathematical functions that describe a wave-like oscillation [29]. Each of these parameters resulted in nine features.

This resulted in a total of 79 spectral-domain features.

Table 3.2: Spectral-domain features extracted from SEMG signal.

| Feature | Description |
|---|---|
| **Fundamental frequency** | Representation of the content of the signal spectrum |
| **0.6-2.5 Hz range energy ratio** | Ratio between the energy in the frequency band 0.6-2.5 Hz and the whole energy band |
| **Linear prediction cepstral coefficients** | Cepstral coefficients derived from the linear prediction |
| **Mel-frequency cepstral coefficients** | Cepstral coefficients based on a cosine transform of the logarithmic power spectrum expressed on a mel-frequency scale |
| **Maximum power spectrum** | Maximum value of the PSD |
| **Maximum frequency** | Maximum frequency of the signal |
| **Median frequency** | Median frequency of the signal |
| **Power bandwidth** | PSD bandwidth, which is the width of the frequency band that contains 95% of its power |
| **Spectral centroid** | Location of the barycenter of the PSD |
| **Spectral decrease** | Amount of decrease of the PSD amplitude with an emphasis on the slopes of the lower frequencies |
| **Spectral distance** | Distance of the signal's cumulative sum of the FFT elements to the respective linear regression |
| **Spectral entropy** | Entropy measure of the PSD of the signal |
| **Spectral kurtosis** | Flatness of a distribution around its mean |
| **Spectral positive turning points** | Number of positive turning points of the FFT magnitude signal |
| **Spectral roll-off** | Frequency value which contains 95% of the signal magnitude |
| **Spectral roll-on** | Frequency value which contains 5% of the signal magnitude |
| **Spectral skewness** | Measure of asymmetry of a distribution around its mean value |
| **Spectral slope** | Amount of decrease of the PSD |
| **Spectral spread** | Standard deviation around the spectral centroid |

Table 3.2: Spectral-domain features extracted from SEMG signal (continued).

| Feature | Description |
|---|---|
| **Spectral variation** | Amount of variation of the PSD, which comes from the normalized cross-correlation between two consecutive amplitude spectra |
| **Wavelet absolute mean** | CWT absolute mean value of the wavelet scales |
| **Wavelet energy** | CWT energy of the wavelet scales |
| **Wavelet entropy** | CWT entropy of each wavelet scale |
| **Wavelet standard deviation** | CWT standard deviation of each wavelet scale |
| **Wavelet variance** | CWT variance of each wavelet scale |

### 3.1.3 Statistical-Domain Features

The statistical features extracted from the signal can be found in Table 3.3. One of the most commonly used features is the mean absolute value, which is used as a measure of the signal's amplitude and, therefore, an indicator of muscle contraction levels. A few variations of this result have been introduced to improve the robustness of the feature [99] by assigning higher weights to more central values. In particular, modified mean absolute value type 2 uses a continuous function to improve the smoothness of the weights.

Features such as average energy, root mean square, and variance are power indexes of the signal. Root mean square is a widely used feature for multiple purposes, and in the EMG context, is an indicator of muscle fatigue [107].

Kurtosis analysis has shown to be correlated with muscle contraction force [108].

The interquartile range, standard deviation, variance, coefficient of variation, and its logarithmic transform are estimates of the degree of dispersion of the signal, so these features gain relevance due to the importance of EMG variability in pathology recognition [18].

The histogram values inform about the frequency with which the signal reaches different amplitudes, considering ten equal-width bins, consequently leading to ten features [105].

Ten values of the empirical cumulative distribution function along the time axis were analyzed, a function which provides information about the empirical measure of a sample. The percentile values and percentile counts of the function were also obtained, adding four more features to this subset [105].

The total of statistical-domain features comes down to 44.

Table 3.3: Statistical-domain features extracted from SEMG signal.

| Feature | Description |
|---|---|
| **Average energy** | Mean of the squared values of the signal |

Table 3.3: Statistical-domain features extracted from SEMG signal (continued).

| Feature | Description |
|---|---|
| **Values of the empirical cumulative distribution function (ECDF)** | Values of ECDF, which is an estimate of the cumulative distribution function that generated the points in the sample |
| **ECDF percentile** | Percentile values of the ECDF |
| **Coefficient of variation** | Ratio of standard deviation to mean of the signal |
| **ECDF percentile count** | Cumulative sum of samples lower than the percentile |
| **ECDF slope** | Slope of the ECDF |
| **Enhanced mean absolute value** | Mean absolute value with an additional discrete parameter weighting each value |
| **Histogram** | Values of the histogram, which represents a frequency distribution of the signal |
| **Interquartile range** | Difference between the upper and lower percentiles of the signal |
| **Kurtosis** | Measure of how the tails of the signal's distribution differ from the tails of a normal distribution |
| **Log coefficient of variation** | Logarithmic transformation of the coefficient of variation |
| **Maximum** | Maximum value of the signal |
| **Mean** | Mean value of the signal |
| **Mean absolute deviation** | Average distance between each value of the signal and its mean |
| **Mean absolute value** | Average of the absolute values of the signal |
| **Median** | Median value of the signal |
| **Median absolute deviation** | Median of the absolute deviations from the signal's median |
| **Minimum** | Minimum value of the signal |
| **Modified mean absolute value type 1** | Mean absolute value with an additional discrete parameter weighting each value |
| **Modified mean absolute value type 2** | Mean absolute value with an additional continuous parameter weighting each value |
| **Root mean square** | Square root of the arithmetic mean of the squares of the signal's values |
| **Skewness** | Measure of asymmetry that deviates from the normal distribution |

Table 3.3: Statistical-domain features extracted from SEMG signal (continued).

| Feature | Description |
|---|---|
| **Standard deviation** | Measure of the dispersion of the signal relative to its mean |
| **Variance** | Square of the standard deviation |

### 3.1.4 Fractal-Domain Features

Lastly, the fractal-domain features used in this work are listed in Table 3.4. The FD correlates self-similar signal patterns or figures and the used scale for their measurement. It can be obtained from the logarithmic relationship between the changes in the length of the curve and the measurement scale. Thus it evaluates the complexity of a figure by understanding how the used scale affects the detail in the pattern. Biosignals' patterns, which are small-scale structures, present self-similarity with larger-scale structures, and therefore this relationship can be studied with fractal analysis. The FD is useful in EMG tasks since it evaluates non-linear dynamics, and the EMG signals are not linear nor stationary. Additionally, it has been established that FD is related to the muscle size and complexity of the muscle properties and not to the contraction strength [22], [109], [110]. Higuchi's Fractal Dimension (HFD) is one of the most popular fractal time series algorithms, and it has been demonstrated that it leads to a better estimation of the FD for physiological signals. Thus, in our work, HFD was used to compute the FD.

MFL can be useful when measuring low-level muscle activation since FD is only an interesting measure for high contraction levels. Muscle force is obtained through the recruitment of MUAPs, each MUAP causing a singularity in the signal. The height and density of singularities significantly influence the length of the signal, and thus MFL is a solid indicator of muscle contraction force [110], [111].

The self-similarity of a physiological time series can be estimated using DFA, a fractal time series algorithm. It provides a scaling exponent related to spectral techniques, but it does not rely on the selection of wavelets, thus achieving lower computational complexity [112].

Lempel-Ziv (LZ) complexity implementation begins with converting the signal to a binary sequence by comparing each value with a threshold of 0, followed by the identification of the normalized number of distinct patterns contained in the sequence. This measure has been applied successfully for measuring the deterministic complexity of EMG signals. In particular, it reflects the duration and firing rate of MUAPs after the computation of the number of different patterns present in the signal within a sliding window [113], [114]. The LZ complexity was also computed for the absolute signal, using a threshold of 0.4, resulting in two features related to this measure.

The signal's entropy is affected by the density of the SEMG interference pattern and

its amplitude distribution, both relevant in pathology discrimination. When using Multiscale entropy (MSE), the entropy analysis is performed over multiple time scales. Abnormal physiology is usually associated with more regularity, and therefore this measure is beneficial due to its capacity to provide better insight into the non-linear dynamic properties of the signal [34], [115]. Two features were obtained from this measure, its average and standard deviation.

A more detailed description of how these methods were applied can be found in [116]. A total of seven fractal-domain features was explored in our work.

Table 3.4: Fractal-domain features extracted from SEMG signal.

| Feature | Description |
|---|---|
| **Detrended Fluctuation Analysis** | Modified root mean square analysis of the integrated signal |
| **Fractal Dimension** | Measure of non-linear properties that reflects how the detail in the fractal changes with the used scale |
| **Lempel-Ziv Complexity** | Measure of finite sequences randomness |
| **Maximum Fractal Length** | Length obtained from the y-axis interception of the logarithmic plot of length versus scale |
| **Multiscale entropy** | Sum of sample entropy values computed from multiple downsampled versions of the signal |

## 3.2 Novel Features

The clinical interpretation of an EMG exam relies on the analysis of MUAPs waveforms. Quantitative measurements are conducted by the clinician, such as MUAPs amplitude, duration, number of phases, firing rate, among others [117], as found in Figure 3.1. As identified in Section 1.3, these measurements are seldom used as features for learning algorithms.

For this reason, we developed features focused on extracting information regarding the morphology of the signal. These features were designed on the assumption that they can be more interpretable, as they measure morphological characteristics that represent physiological processes that are taken into account in clinical interpretation. Two groups of features are proposed: **peak-related features** and **MUAP morphology features**. A brief description for each feature of the peak-related and MUAP morphology groups can be found in Tables 3.5 and 3.6, respectively. The first group quantifies statistical characteristics related to all the detected maxima of the signal. The second one provides MUAP morphology measures, which can be visually interpreted in Figure 3.1.

Peaks difference and all MUAP morphology features generate two variables, the average of the mentioned measure and the respective standard deviation. The extraction of the standard deviation alongside the average was due to the standard deviation being

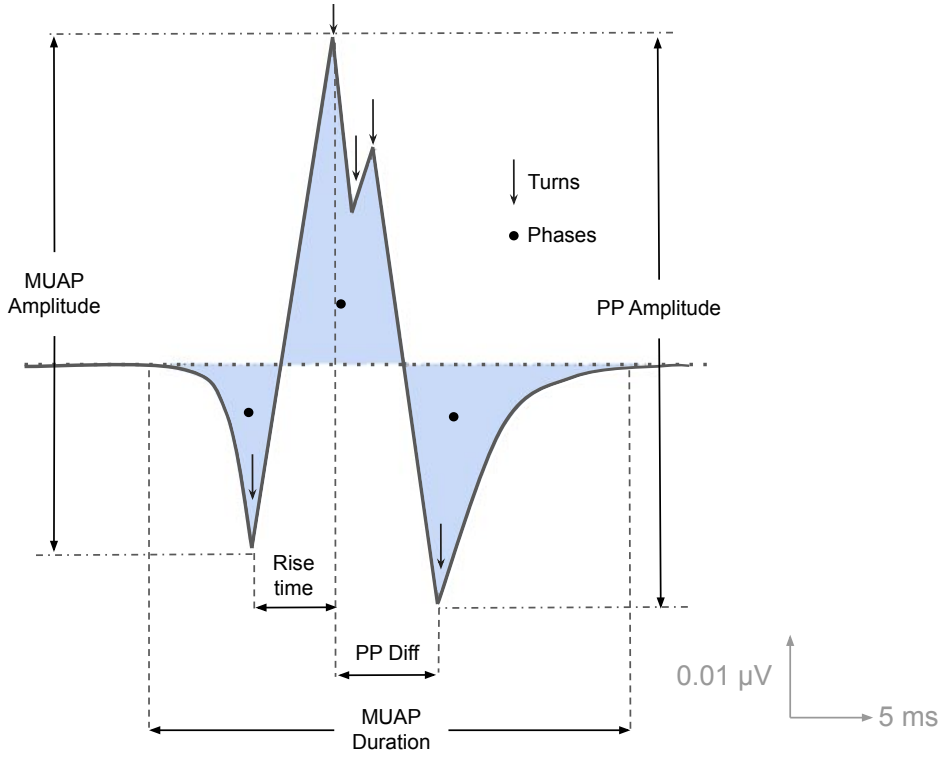



Figure 3.1: Quantitative characteristics of MUAP waveforms. "PP" corresponds to peak-to-peak.

related to the variation within each measure, which could present differences between pathological and healthy signals. The remaining features originate one variable, the direct measure. This results in a total of 18 new features.

The applied methodology to obtain the novel features is based on detecting the signal's peaks, and decomposing the SEMG signal to identify the surface representations of individual MUAPs. The MUs activated during muscle contraction generate MUAPs and their summation yields the SEMG signal. The SEMG decomposition consists of segmenting and identifying the constituent MUAPs and, therefore, the superimpositions of SEMG signals can be avoided to some extent. Over the last years, there have been several proposed approaches for EMG decomposition [118], [119]. These approaches are complex and still have some shortcomings, particularly for SEMG, which is more challenging due to its signal complexity. We opted to conduct the SEMG decomposition using a simpler method. Although our proposed method might come short compared to more sophisticated alternative methods, we argue that it can still be applied for a preliminary assessment between control and ALS groups using our proposed set of features.

Our naive approach for SEMG decomposition consists of (1) identifying all the significant local maxima during muscle activation periods; and (2) post-process the detected peaks to isolate the MUAP window around them.

For the first task, Algorithm 1 was implemented. As shown in this algorithm, a local maxima is considered significant if its amplitude is higher than 98-th percentile of the

Table 3.5: Description of peak-related features

| Feature | Description |
|---|---|
| **Number of peaks** | Total number of detected peaks |
| **Peaks difference** | Time interval between consecutive peaks |
| **Peaks rate** | Number of peaks per second |

Table 3.6: Description of MUAP morphology features

| Feature | Description |
|---|---|
| **Peak-to-peak amplitude** | Amplitude from the maximum negative peak to the maximum positive peak of the MUAP |
| **Peak-to-peak difference** | Time interval between the maximum negative peak to the maximum positive peak of the MUAP |
| **MUAP duration** | Time interval during which the MUAP occurs |
| **MUAP integrated area** | Absolute area of the MUAP |
| **MUAP rise time** | Time interval between maximum negative peak and the following minimum positive peak within the duration of the MUAP |
| **MUAP phases** | Number of baseline crossings within the duration where amplitude exceeds the mean of the signal |
| **MUAP turns** | Number of positive and negative peaks where the differences from the preceding and following turn exceed 25 $\mu V$ |

onset moment and the distance between consecutive maxima has a minimum value of 20 Milliseconds (ms), to prevent the detection of peaks related to noise. Using the obtained peaks in this phase, the peak-related features were extracted for each contraction window of the signal. The features number of peaks and peaks rate were directly quantified. The peaks difference was determined for each pair of consecutive peaks and then averaged, resulting in two features which were the average and standard deviation of this measure.

**Algorithm 1** Peaks detection

**Input:**

window $w$ ▷ Signal corresponding to a contraction window.

min_height ▷ Minimum peak height.

min_dist ▷ Minimum distance between consecutive peaks.

**Output:**

peaks ▷ Detected peaks' indexes.

1: $min_height \leftarrow percentile(w, 98)$
2: $min_dist \leftarrow 20$
3: $peaks \leftarrow find_peaks(w, min_height, min_dist)$ ▷ Compute the peaks' indexes.

The second part involved delimiting the MUAP windows around the peaks detected by Algorithm 1. This is represented by Algorithm 2. The MUAP is considered the window of 20 ms centered in each detected peak, since the average MUAP duration goes from 10

**Algorithm** 2 MUAP detection

**Input:**
window $w$ ▷ Signal corresponding to a contraction window.
peaks ▷ Detected peaks' indexes.
win_len ▷ Length of the MUAP window.
**Output:**
MUAPS ▷ All MUAP values.
**Require:**
win_len $\in [10, 30]$ ms
1: **for** peak in peaks **do**
2: MUAP $\leftarrow w[peak - \frac{win_len}{2} : peak + \frac{win_len}{2}]$ ▷ Save the signal's values within the window.
3: MUAPS $\leftarrow append(MUAP)$ ▷ Add the obtained MUAP to a list.
4: **end for**

ms to approximately 30 ms [120].

Due to the variety of the designed features, the MUAP was consequently delimited between its maximum negative and positive peaks, in order to be able to obtain peak-to-peak measures within the MUAP waveform, as shown in Algorithm 3. This algorithm incorporated the detection of negative local minima within the MUAP window, whose absolute height had to be at least the mean of all the negative values in the window. The detected positive and negative peaks of a part of the samples were visually inspected, and most of them were correctly identified. However, as expected, some surface MUAPs may be closer together than what happens in NEMG, or even superimposed, which resulted in the detection of a few abnormal peaks that did not appear representative of a MUAP. Four features were obtained from this algorithm, related to the peak-to-peak amplitude and time difference. These peak-to-peak measures were determined for each MUAP within a contraction window, and then their mean and respective standard deviation were calculated.

Lastly, the waveform of the MUAP was isolated through the detection of stable zones around the main peak, which allowed the calculation of the remaining MUAP morphology features, as shown by Algorithm 4. This stage is of particular importance since within the duration of the MUAP, we find intrinsic physiologic information about the MU, namely regarding the number of muscle fibers in it and the temporal dispersion of their firing [121]. First we inspect the order in which the peaks detected by Algorithm 3 appear in the MUAP window, and then we subtract one ms to the one that comes first and add one ms to the one that comes second, referring to them as first and second points. Then the values to the left and right of the first and second points, respectively, are assessed and compared to a threshold. This threshold is defined as $\frac{1}{15}$ of the peak-to-peak amplitude of the MUAP, previously determined by the same Algorithm 3. The first left and right values greater than the threshold are the considered starting and ending points, respectively. This process allows the identification of the waveform region that approximates to what

**Algorithm 3** MUAP peak-to-peak limitation

**Input:**
MUAPS ▷ All MUAP values.
peaks ▷ Detected peaks' indexes.
**Output:**
pos_peaks ▷ Positive peaks' indexes.
neg_peaks ▷ Negative peaks' indexes.
1: **for** MUAP in MUAPS and peak in peaks **do**
2: $thresh_inv \leftarrow mean(-muap[-muap > 0])$ ▷ Calculate the threshold for the detection of negative peaks.
3: $inv_peaks \leftarrow find_peaks(-muap, thresh_inv)$ ▷ Compute the negative peaks' indexes.
4: $pos_peak \leftarrow peak$ ▷ Compute the maximum positive peak index.
5: $neg_peak \leftarrow where(min(muap[inv_peaks]))$ ▷ Compute the minimum negative peak index.
6: $pos_peaks \leftarrow append(pos_peak)$ ▷ Add the obtained positive peak to a list.
7: $neg_peaks \leftarrow append(neg_peak)$ ▷ Add the obtained negative peak to a list.
8: **end for**

the MUAP waveform is expected to be like. This algorithm allowed the extraction of the remaining 10 features, which are related to measures within the waveform of the MUAP, namely the MUAP duration, integrated area, rise time, phases and turns. In a similar way to what happened with the peak-to-peak measures, first the MUAP measures were determined for each MUAP within a contraction window, and then the values were averaged. The average and respective standard deviation are the considered features.

We acknowledge that this method has limitations. It does not ensure that all detected peaks are actually representative of a MUAP, and it does not take into account that MUAP duration presents a high degree of variability [121]. Additionally, some MUAPs may not be noticed through this method and superimposed MUAPs can not be distinguished. Nevertheless, MUAP detection is rarely flawless due to the fact that SEMG signals have a lower precision, and for that reason, MUAPs are not as distinctive as in NEMG. Also, it must be noted that this work investigates the interpretation of the surface representations of MUAPs, not their intramuscular representation. Given these circumstances, our approach allows a significant amount of surface MUAPs to be considered that are representative of the overall pattern present in the surface signal.

**Algorithm 4** MUAP waveform limitation

**Input:**

MUAPS ▷ All MUAP values.
peak_to_peak_amp ▷ Peak-to-peak amplitude for each MUAP.
amp_frac ▷ Fraction of the peak-to-peak amplitude to consider.
pos_peaks ▷ Positive peaks' indexes.
neg_peaks ▷ Negative peaks' indexes.

**Output:**

start_points ▷ Starting points of the MUAP waveforms.
end_points ▷ Ending points of the MUAP waveforms.

$amp_frac \leftarrow \frac{1}{15}$
**for** i in len(MUAPS) **do**
  $MUAP \leftarrow MUAPS[i]$
  $pos_peak \leftarrow pos_peaks[i]$
  $neg_peak \leftarrow neg_peaks[i]$
  **if** pos_peak < neg_peak **then** ▷ If the positive peak appears first, it is associated with the first point.
    $first_point \leftarrow pos_peak - 1$
    $second_point \leftarrow neg_peak + 1$
  **else if** neg_peak < pos_peak **then** ▷ If the negative peak appears first, it is associated with the first point.
    $first_point \leftarrow neg_peak - 1$
    $second_point \leftarrow pos_peak + 1$
  **end if**
  $amp \leftarrow peak_to_peak_amp[i]$
  $thres \leftarrow amp_frac \times amp$ ▷ Calculate the threshold.
  **for** left_value in MUAP[first_point : 0] **do**
    $start_point \leftarrow when(left_value > thres)$ ▷ Starting from the first point, the first value to the left that surpasses the threshold is found.
  **end for**
  **for** right_value in MUAP[second_point : -1] **do**
    $end_point \leftarrow when(right_value > thres)$ ▷ Starting from the second point, the first value to the right that surpasses the threshold is found.
  **end for**
  $start_points \leftarrow append(start_point)$ ▷ Add the obtained starting point to a list.
  $end_points \leftarrow append(end_point)$ ▷ Add the obtained ending point to a list.
**end for**

# 4

# A Framework for Amyotrophic Lateral Sclerosis Diagnosis based on SEMG

A framework was developed to distinguish between pathological and healthy SEMG signals. An overview of such framework is represented in Figure 4.1. The first stage of the framework begins with the study and preprocessing of the produced SEMG signals. Then, the signals are divided into windows, according to the activation moments of each subject. The features are then extracted and all the information is combined into one vector of features for each window. The parameters of the classifier to be used and the threshold for sample rejection are determined through tuning processes. Consequently, some features are removed and feature selection is performed, as well as a 10-fold cross

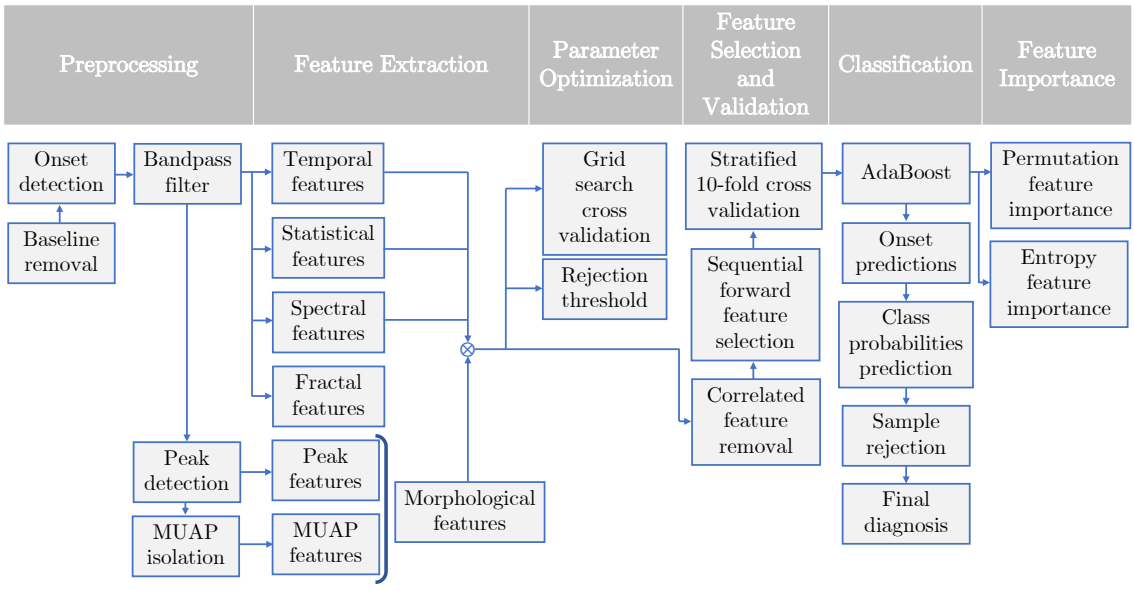



Figure 4.1: Schematic representing the framework's pipeline.

validation strategy, guaranteeing that samples from the same subject are only on the train or test set. At last, a supervised learning algorithm employs the features as input and after the learning process it is able to associate each window with one of the possible classes: healthy or pathological (ALS). Sample rejection makes use of the class probabilities, leading to a new set of onset predictions, with the rejected windows excluded from these results. Finally a voting system decides the final diagnosis of an entire signal.

## 4.1 Preprocessing

The SEMG signal was measured using multiple channels. The acquired data from all channels was then converted to Millivolts (mV), centered by removing the baseline, and then digitally filtered using a 3$^{rd}$ order Butterworth bandpass filter, between the frequencies of 10 and 300 Hz. Subsequently, each SEMG signal was split into windows according to its muscle contraction moments or bursts, which were unique for each subject and isolated with a semi-automatic method.

The sequential outcomes of the muscle onset detection method are shown in Figure 4.2. This method starts with the application of the TKEO to the signal, which measures instantaneous energy changes and has proven its utility in minimizing error when detecting onsets of SEMG signals [122]. This was followed by rectification, application of the first moving average filter, and smoothing of the signal. Lastly, according to a method proposed by [122], a threshold value was defined as:

$$threshold = \mu_{EMG} + h\sigma_{EMG} \tag{4.1}$$

where $\mu_{EMG}$ is the average of the signal, $\sigma_{EMG}$ its standard deviation, and $h$ a defining level of the threshold. The variable $h$ is defined in percentage relatively to the average value of the smoothed signal and it could vary for each subject.

To ensure that the threshold level was within the maximum and minimum of the smoothed signal, a normalisation regression function was applied to the threshold. For visualization purposes, a square wave reflecting the activation and inactivation periods was generated. The detected potential onsets (beginning of an activation), and idle moments (ending of an activation) were visually inspected to guarantee that all contractions were being correctly isolated. Signals from different channels were synchronized in time, so this segmentation was done for one channel and the onset and idle times were saved and used for all channels of that subject. This resulted in a group of samples for each subject that were considered either a muscular contraction or a rest moment.

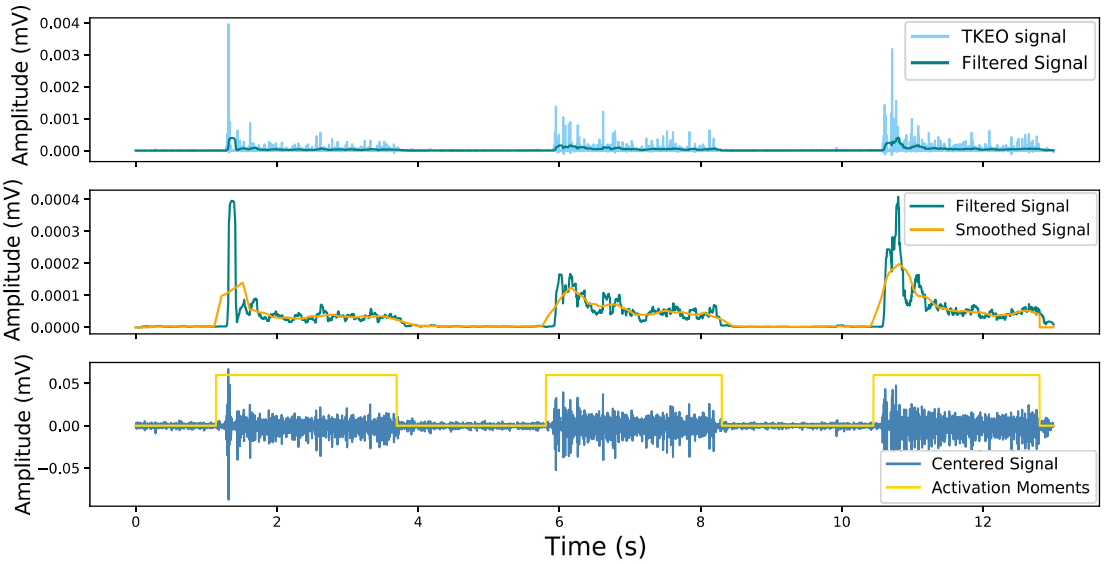



Figure 4.2: Signal processing for onset detection. The upper graph shows the signal after the application of TKEO, and after the application of the first moving average filter. The middle graph represents the filtered signal before and after its smoothing. The bottom graph shows the original signal and its detected onsets.

## 4.2 Feature Extraction

The complete set of features was extracted for each onset window separately. This process was repeated for all subjects, originating a feature matrix. Each row in this matrix represents a sample and each column refers to the features' values that describe each sample quantitatively. It is worth clarifying that sample refers to a signal window associated with a muscular contraction. Therefore each subject has one signal with multiple samples, or contraction windows. Two additional columns were added providing information about the subject's ID and the class of that subject (ALS or HC). There were as many rows for the same subject as activation moments detected in that subject's signal.

Since each subject had at least three channels of data acquired simultaneously, three files containing this table of features were extracted. In order to distinguish the information coming from different channels, a suffix was added to the feature's name indicating the channel from which that feature had been extracted.

Features were extracted exclusively for contraction moments, so the subject's rest instants were not considered.

## 4.3 Feature Selection

The feature selection process identifies the subset of features most effective in discriminating the two classes. A theoretical outline of the prevailing approaches for feature selection was described in Section 2.4.2. In this section, we will further describe the

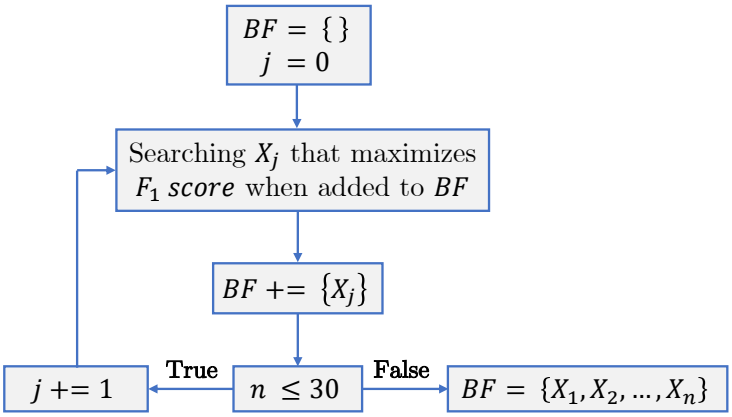



Figure 4.3: Schematic representation of SFS. $BF$ represents the vector of selected features that is being iteratively updated, $X_j$ represents the feature being analyzed in the current iteration $j$, and $n$ is the length of the selected features vector.

feature selection methods adopted in the context of the proposed framework.

Firstly, the pairwise correlation of features was computed using the Pearson correlation method. This method measures the linear correlation between data, and provides a coefficient value between -1 and +1, with 0 indicating no relationship present, and -1 and +1 meaning a negative and positive, respectively, perfect correlation. Highly correlated features, with a coefficient over 0.95, do not provide discriminating information simultaneously, and therefore one of them can be removed.

Figure 4.3 displays the SFS algorithm that was implemented, which searches the space of all feature subsets for the best predictive performance, evaluated through the $F_1$ score. It starts from an empty set, where the feature resulting in the highest $F_1$ score is added. Then the additional feature that provides the best $F_1$ score when combined with the previously selected subset is added into this feature subset. These process is repeated sequentially until a total of 30 features is obtained. The subset size was limited due to the large amount of evaluated features. A high computational cost prevented all features to be considered in feature selection until the model's predictive performance plateaued.

Since a 10-fold Shuffle-Group-Out Cross-Validation was implemented, detailed in Section 2.4.4, each fold will provide a different subset of features. The most frequent throughout all folds are considered the final selected features.

## 4.4 Supervised Learning

Supervised techniques were applied to discriminate between subjects with ALS and HCs, using multiple samples from each subject. The first step of this process was to optimize the classifier's parameters. Then, an AdaBoost classifier was trained with multiple train samples, guaranteeing that all samples from each subject were either on the training set

or the test set. After this, the classifier was able to attribute a class for each sample of the test set. To improve the robustness of the proposed method, class probabilities were used for sample rejection, which eliminated signal windows whose attributed class had a low probability value. The rejected windows were not considered for classification.

However, there was still missing a final decision that took into account all samples from one signal, acquired from one subject. Therefore, a voting system was implemented to provide a diagnosis for the subject.

These steps were performed in a 10-fold Shuffle-Group-Out Cross-Validation context, leading to 10 groups of results, each associated with a different fold. The average of these results led to the final considered metrics for the classification task.

The reason why the Leave-One-Group-Out Cross-Validation strategy was not implemented was due to the fact that it had a higher computational cost and it tended to *overfit* the model especially during the parameter optimization.

### 4.4.1 Classifier's Parameters Optimization

A Grid Search Cross-Validation technique obtained the best parameters for the chosen classifier before performing the actual classification of samples. This strategy iteratively experimented different groups of parameters for the model in the classification task when all available features were being used as input, throughout 10 folds of Cross-Validation using only the train data. The group of parameters that maximized the $F_1$ score metric was chosen as the ideal set of parameters for that fold. To avoid *overfitting*, the parameters were not updated for each fold, but instead the most common ones throughout experimenting folds were selected and used during the 10 folds of the actual classification task.

As explained in Section 2.4.3, AdaBoost needs a base estimator from which the boosted ensemble is built. The other parameters include: the number of estimators at which the ensemble is terminated; the weight applied to each classifier; and the boosting algorithm. Grid Search went through multiple possibilities of groups of parameters, obtaining an ideal fit for 50 estimators, with a learning rate of one, and the *Stagewise Additive Modeling using a Multi-class Exponential Loss Function (SAMME)* boosting algorithm, using Random Forest as the base estimator. SAMME is an adapted implementation of AdaBoost, with a difference in the estimator weight computation. Since our task is a binary classification problem, the SAMME algorithm is equal to the traditional AdaBoost implementation, firstly described in [123].

In this context, as Random Forest is applied as a weak classifier, its optimized maximum depth has a value of only one. The other enhanced parameter in Random Forest is the criterion for the creation of partitions in the trees of the forest, that are usually measures of impurity of a node. The information gain principle led to the best results. In this case, entropy is used as a measure of information indicating the disorder or impurity of the features with the possible classes. If the entropy of a feature is zero, all samples of

a node belong to the same class, and if it is maximal, the samples have a uniform class distribution. Therefore, the chosen feature for the split is the one that minimizes entropy and impurity, maximizing information gain.

### 4.4.2 Sample Rejection

The main objective of designing a rejection class was to consider only the predicted class results based on high probability values. If the probability for that prediction was less that an optimal threshold, then the respective sample was rejected and excluded from the overall classification performance. For this purpose, after classification, the probability results were obtained. The difference between these results and the normal classification is that instead of giving us a final class, it provides us with the probability value for each class. That is to say, if the labels are 0 and 1, a normal classification would output 0 or 1, whereas probability results provide the probability value of the sample being 0 and 1.

The optimal threshold was obtained by plotting the curves threshold (x-axis) versus accuracy (y-axis) and threshold (x-axis) versus rejection rate (y-axis) during the supervised learning process, considering only the training set. The goal was to maximize accuracy while minimizing the threshold, so that the percentage of rejected samples was also minimal. Therefore, the slope of non-equal consecutive points was calculated until the rejection rate was 35%. Then, the threshold corresponding to the maximum slope was chosen as optimal threshold. This process was done in all 10 folds of the learning task when using all available features. The average of the selected optimal thresholds for all folds was considered as the final optimal threshold, and the classification process was repeated taking into consideration only this average to reject samples. This was done to prevent *overfitting*.

### 4.4.3 Voting System or Diagnosis Criteria

As previously mentioned, each subject is associated with one signal which contains multiple samples, and each of these samples is classified individually. A voting system was implemented in order to reach an agreement on a final diagnosis for the subject, considering all attributed classes within the same signal as shown in Figure 4.4. This system was based on what is done in voting ensembles, which are algorithms that make a prediction by majority vote of multiple contributing models. This can be done either through hard voting, that predicts the class with the most votes from the other models, or soft voting, that predicts the class with the largest summed probability from the models. We performed the first approach but in the context of combining contraction windows information. This criteria counted the number of healthy classified samples as the number of votes for the HC class, and the same for the ALS class. The class with the most votes was considered the final decision of the classifier. Therefore, two different types of classification tasks were performed: **onset classification** and **subject classification**. The latter

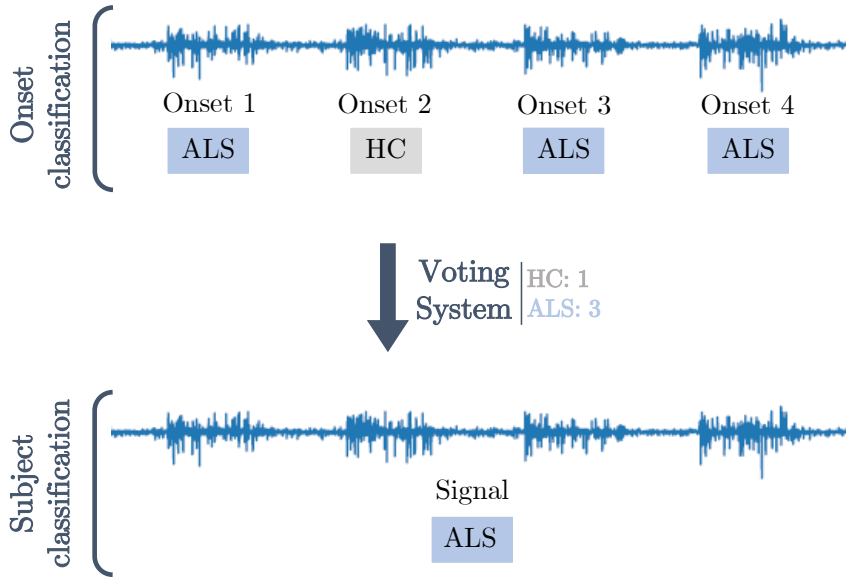



Figure 4.4: Voting system schematic, showing the contribution of each onset for the final subject's class.

consists of the implementation of this voting system, that takes into consideration the information given by the first task.

# 5

# Performance Evaluation

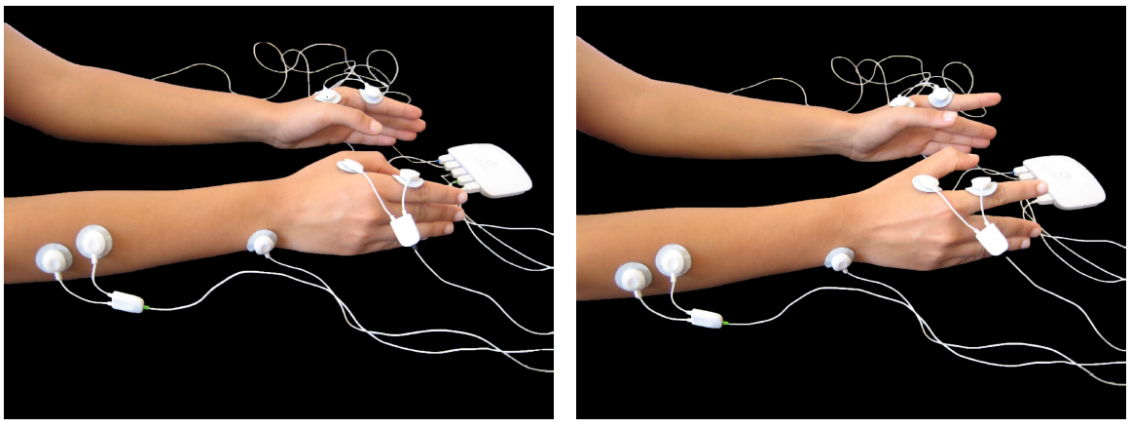
A case study was conducted aiming to validate the new SEMG features in a binary classification process by analyzing their performance. Following the steps delineated in Chapter 4, healthy and ALS signals were discriminated. A combination of the different groups of features described in Chapter 3 was explored to improve the classification performance. This chapter encompasses the description of the experimental conditions in which data was collected, the outcomes of the classification experiments, the discussion of these results, and the importance ranking and statistical analysis of the implemented features. These tasks were performed resorting to the computational tools described in Appendix A.

## 5.1 Dataset

The present study used a dataset that had been previously explored in [116], where a more detailed description of the experimental protocol can be found. Data was acquired from two different subsets of subjects: HC and patients diagnosed with ALS within the preceding 36 months, with a muscle strength greater than three according to the Medical Research Council scale (MRC scale), in the tested muscles. Patients could not present any other neurological disorders. All patients were medicated with Riluzole.

The ALS population was initially comprised of 24 subjects; however, three of these subjects' signals were imperceptible since moments of contraction could not be isolated, leading to a group of 21 subjects, eight men and 13 women, mean age 59.4 ± 9.4. One subject was also excluded from the HC subset for the same reason, resulting in a group of 24 healthy subjects, nine men and 15 women. The patients were segmented into two categories depending on their disease phenotype. Since the SEMG signal is not as pathologically expressive in bulbar subjects, spinal patients were the only ones considered for further analysis. Therefore we had a total of 41 subjects, 17 diagnosed with ALS and 24 HCs.

Subjects were seated with both hands and forearms on a desk in a parallel position, 10 cm away from each other with hand palms facing one another in 90 degrees flexion with



(a) (b)

Figure 5.1: Experimental setup showing the position of both limbs, and the Bioplux device with the four SEMG sensors and ground electrode. Acquisition protocol consisting of (a) rest period and (b) contraction period. Source: [116].

the elbow. Then, they were asked to perform the same movement on both left and right hands while listening to a programmed sound, which guided the task. This task was a coordinated abduction of both index fingers in the opposing direction of the remaining fingers, with maximum articular amplitude, hold that position with a certain degree of force for three seconds, return to the original position and remain in that position for three seconds while trying to relax as much as possible. Each having two surface electrodes connected, the sensors were fixed on the first dorsal interosseus muscle for both hands and the extensor digitorum communis muscle for both forearms. The ground was placed on ulna bone inferior extremity since no muscle activity is present in that region. This resulted in the recording of four time-synchronized signals for each subject [116].

The performed task, visible in Figure 5.1, was repeated for six minutes or less depending on the maximum time tolerated by the patients.

All acquisitions were performed with a BioPlux device with eight analog input channels converted to 12-bit signals and an external channel used as reference ground. The SEMG sensors include $2^{nd}$ order bandpass analog filters with 25 and 450 Hz cut-off frequencies adjacent to the electrodes. The SEMG signals were acquired with a gain of 1000 and a sampling frequency of 1000 Hz.

After inspecting the raw data, it was verified that not all subjects had information from four channels. Therefore the first three channels, which were present in all recordings, were selected for analysis. The first channel refers to the left hand, the second channel refers to the left forearm and the third channel refers to the right hand.

After the processing methods explained in Section 4.1 were applied to all signals, an average of 87.61 contractions was detected per subject, resulting in a total of 3592 onset moments for the 41 subjects. Of this total, 1923 were from healthy subjects, and 1669 were from ALS patients, resulting in a proportion of approximately 53.5% of healthy

samples and 46.5% of pathological samples. However, it is important to mention that some subjects had as little as four onset intervals, and a few presented an amount far higher than average.

## 5.2 Classification

For each muscular activation interval, a total of 196 features, previously described in Chapter 3, were extracted. These features were extracted for the three channels, resulting in 588 features. During the feature preprocessing step, six features were eliminated due to missing values, and 265 were also removed for being correlated. A total of 320 features per muscular activation interval was considered.

This chapter summarizes the average results of a 10-fold Shuffle-Group-Out Cross-Validation that split the number of subjects in half for train and test in each fold. The group of chosen subjects for train and test in a fold was the same throughout all experiments in that specific fold so that outcomes from different experiments could be compared. The calculated metrics, except accuracy, were macro-averaged since the dataset was approximately balanced, giving a more generalized performance measure irrespective of the class.

Additionally, as previously mentioned in Section 4.4.3, two classification strategies were considered: onset and subject classification. The voting criterion assumes that the model attributed the same class to all the samples of that subject, which is not always true. Therefore, the onset and subject classification results which will be presented below are not comparable, considering they represent distinct outcomes of the same signal.

Table 5.1 summarizes the metric scores for the onset classification task using each group of features individually. The group of spectral features produced the overall best results, followed by the group of proposed features. On the other hand, the temporal and statistical groups performed relatively poorly, with overall lower scores.

Table 5.2 summarizes the metric scores for the subject classification task using each group of features individually. The results are in agreement with the outcome for the onset classification task, with the group of features with the best overall scores being the spectral, followed by the proposed feature group.

Table 5.3 compares the onset classification performance when using the temporal, spectral, statistical, and fractal features and when using all features. We hypothesized that introducing the proposed features to the remaining feature group would increase the overall classification results. Table 5.3 shows evidence that the proposed features slightly improved all metrics when combined with the remaining feature groups. The hypothesis was further confirmed in the subject classification, whose results are presented in Table 5.4.

Table 5.1: Onset classification results obtained by using each feature group separately measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **Temporal** | 68.87 ± 7.84 | 77.03 ± 4.52 | 69.30 ± 5.71 | 73.37 ± 6.12 |
| **Spectral** | **81.18 ± 4.68** | **85.31 ± 3.71** | **80.65 ± 4.21** | **82.76 ± 4.61** |
| **Statistical** | 69.24 ± 8.55 | 74.60 ± 6.41 | 69.99 ± 6.85 | 73.04 ± 7.85 |
| **Fractal** | 71.52 ± 9.86 | 78.65 ± 4.14 | 72.27 ± 7.33 | 75.46 ± 6.91 |
| **Proposed** | 77.12 ± 3.19 | 80.67 ± 4.04 | 76.66 ± 2.74 | 78.90 ± 3.41 |

Table 5.2: Subject classification results obtained by using each feature group separately measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **Temporal** | 67.25 ± 10.05 | 74.12 ± 9.97 | 68.66 ± 7.34 | 70.00 ± 9.54 |
| **Spectral** | **75.37 ± 7.84** | **79.45 ± 7.57** | **76.14 ± 7.31** | **76.67 ± 7.51** |
| **Statistical** | 69.57 ± 8.35 | 74.62 ± 7.93 | 70.68 ± 7.48 | 71.43 ± 8.25 |
| **Fractal** | 69.65 ± 10.62 | 77.20 ± 5.88 | 71.53 ± 7.92 | 72.38 ± 8.98 |
| **Proposed** | 70.84 ± 4.68 | 75.53 ± 6.06 | 71.86 ± 4.12 | 72.38 ± 4.67 |

Table 5.3: Onset classification results obtained by using all feature groups and all features except the proposed ones, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **Temporal, Spectral, Statistical, Fractal** | 79.98 ± 3.00 | 85.17 ± 1.90 | 79.00 ± 2.95 | 82.56 ± 3.27 |
| **All features** | **83.11 ± 4.22** | **86.87 ± 2.83** | **82.11 ± 4.19** | **84.75 ± 3.65** |

Table 5.4: Subject classification results obtained by using all feature groups and all features except the proposed ones measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **Temporal, Spectral, Statistical, Fractal** | 76.64 ± 5.79 | 81.15 ± 6.01 | 76.93 ± 5.37 | 78.10 ± 5.71 |
| **All features** | **78.85 ± 5.96** | **82.27 ± 5.97** | **78.84 ± 5.29** | **80.00 ± 5.95** |

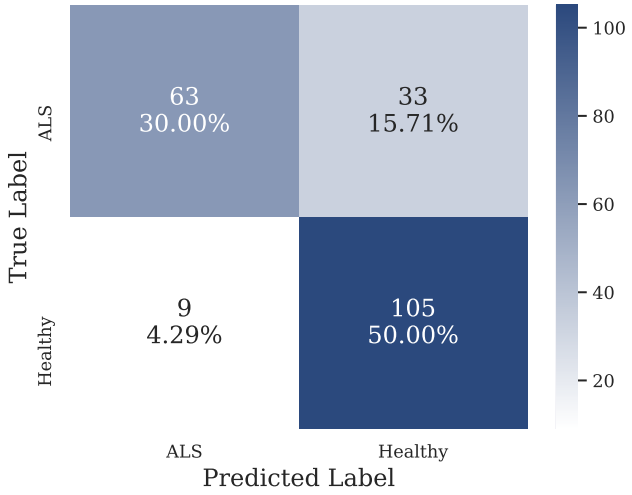



Figure 5.2: Confusion Matrix for the subjects classification task, when using all features. The values are the summation of the results of 10 folds. The predicted values for each class are on the columns and the actual values on the rows.

The confusion matrix displayed in Figure 5.2 showcases the experiment correspondent to Table 5.4 in a different view. This matrix is the end product of 10 folds of Cross-Validation when performing the subject classification with all available features. Considering ALS as the positive class, $T_P$ represents the number of patients who were properly classified as having ALS, and $T_N$ the number of correctly classified subjects who were healthy. We see that the number of $T_N$ is superior to the number of $T_P$, allowing us to understand that the model performs better when classifying healthy subjects. Furthermore, we verify that the number of $F_N$ is higher than the number of $F_P$, meaning that the classifier more often did not identify ALS as being present when it was than the opposite. Since the purpose of this task is to diagnose the disease, falsely classifying a subject as healthy is an error that should be prevented.

The importance of rejecting samples with lower predicted probability values is illustrated in Table 5.5, where it can be confirmed that all metric values improve after introducing sample rejection for onset classification. The calculated optimal threshold for rejecting a sample was approximately 0.5465, which meant that if the classifier had conferred a probability value lower than this for a particular onset, that onset would be excluded. The average probability value attributed by the classifier to the onsets throughout all folds was 0.5846. These values led to a mean rejection rate of 27.17% when using all feature groups, i.e., approximately 450 out of 1658 samples were rejected each fold. These results fluctuated in the additional experiments employing alternating feature groups since the probability value is calculated using the training set, which varied according to the adopted set of features.

To evaluate the relevance of acquiring SEMG data with multiple channels, Table 5.6 summarizes the results in classifying the onsets using a single channel of data at a time

Table 5.5: Onset classification results obtained by using all feature groups before and after sample rejection, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **No sample rejection** | 75.91 ± 2.74 | 79.66 ± 2.79 | 76.09 ± 2.69 | 77.39 ± 2.67 |
| **Sample rejection** | **83.11 ± 4.22** | **86.87 ± 2.83** | **82.11 ± 4.19** | **84.75 ± 3.65** |

Table 5.6: Onset classification results obtained by using all features from each channel and from all channels simultaneously, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **Channel 1** | 68.93 ± 5.73 | 75.96 ± 3.69 | 70.39 ± 5.61 | 71.90 ± 4.44 |
| **Channel 2** | 72.33 ± 7.86 | 79.72 ± 8.98 | 71.69 ± 6.88 | 76.12 ± 8.24 |
| **Channel 3** | 81.88 ± 5.09 | 84.90 ± 5.45 | 81.32 ± 4.87 | 82.99 ± 5.26 |
| **All channels** | **83.11 ± 4.22** | **86.87 ± 2.83** | **82.11 ± 4.19** | **84.75 ± 3.65** |

and using the three channels simultaneously. The results show that considering the three channels leads to improved results. Another pertinent finding is that channel 3 alone produces results that approximate what is achieved when using all three channels. Channel 3 refers to the first dorsal interosseus muscle of the right hand. On the one hand, using the complete set of information is more complete and less prone to *overfit* the model. On the other hand, using only one channel reduces the number of used features by two-thirds, which increases the computational processing speed, namely when performing feature selection, and facilitates the acquisition process.

For this reason, the information gathered from the right hand was explored further and applied in additional classification attempts. Table 5.7 compares the model's performance anew when using all features except the proposed ones and when using all feature groups, in this case, extracted exclusively from channel 3. It reveals that the results are similar to those obtained with features from the three channels. Additionally, the proposed features reach the highest metric values when individual feature groups are extracted from the right hand only, as Table 5.8 conveys. This outcome implies that with a smaller set of features as input, extracted from one channel only, the proposed features originate the best results, producing an efficient performance with fewer resources.

From the results presented above, we conclude that the classification task with the best outcome included the use of all feature groups extracted from the three channels after performing sample rejection, without applying feature selection. This experiment produced an $F_1$ score of 83.11 ± 4.22%. However, this result approximates to what was obtained for the proposed set of features extracted from the right hand, in the same experimental conditions, which resulted in an $F_1$ score of 81.94 ± 5.67%.

Table 5.7: Onset classification results obtained by using all feature groups and all features except the proposed ones extracted from channel 3 only, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **Temporal, Spectral, Statistical, Fractal** | 79.49 ± 4.45 | 82.48 ± 5.39 | 79.24 ± 3.91 | 80.58 ± 4.86 |
| **All features** | **81.88 ± 5.09** | **84.90 ± 5.45** | **81.32 ± 4.87** | **82.99 ± 5.26** |

Table 5.8: Onset classification results obtained by using each feature group separately extracted from channel 3 only measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **Temporal** | 73.16 ± 6.28 | 79.34 ± 5.62 | 73.09 ± 6.04 | 76.92 ± 4.65 |
| **Spectral** | 78.58 ± 4.82 | 81.47 ± 5.18 | 78.41 ± 4.83 | 79.66 ± 4.83 |
| **Statistical** | 76.50 ± 4.10 | 79.99 ± 3.66 | 76.69 ± 3.71 | 77.85 ± 3.93 |
| **Fractal** | 77.42 ± 7.14 | 80.86 ± 6.30 | 77.13 ± 7.18 | 79.58 ± 6.23 |
| **Proposed** | **81.94 ± 5.67** | **84.01 ± 4.92** | **81.90 ± 5.45** | **82.56 ± 5.59** |

The achieved classification accuracy is in agreement with other published works based on SEMG [34]. Quintão et al. [33] used the same dataset as this dissertation, and achieved an average accuracy of 77%. However, one should mention that the authors of [33] employed different classification strategies, considered one more channel of data, and implemented a different subset of features, including two time-domain features, seven frequency-domain features, two synchrony measures, and five fractal-domain features. When using all features extracted from all channels, we obtained an average accuracy of 85%, which is a slightly higher value.

Several additional experiments on onset and subject classification were conducted, varying the number of tested feature groups as the employed channels. For the results of the thorough analysis, please refer to Appendix B.

## 5.3 Feature Selection

As previously detailed in Section 4.3, 30 features were chosen for feature selection. This led to a worsened onset classification performance when compared to the onset classification results using all features, as Table 5.9 discloses. Nevertheless, the most common features throughout the 10 folds were identified, as illustrated in Figure 5.3. It displays a bar chart with the number of times each feature, extracted from any of the three channels, was selected during the 10 folds of the Cross-Validation scheme. Features from all feature groups, except the fractal, were chosen as part of the optimal subset of features, demonstrating the importance of extracting diversified information from the signal instead of

Table 5.9: Onset classification results obtained by using all feature groups and 30 features selected with SFS, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **No Feature Selection** | **83.11 ± 4.22** | **86.87 ± 2.83** | **82.11 ± 4.19** | **84.75 ± 3.65** |
| **Feature Selection** | 75.83 ± 5.78 | 82.80 ± 10.95 | 75.86 ± 4.80 | 77.57 ± 5.33 |

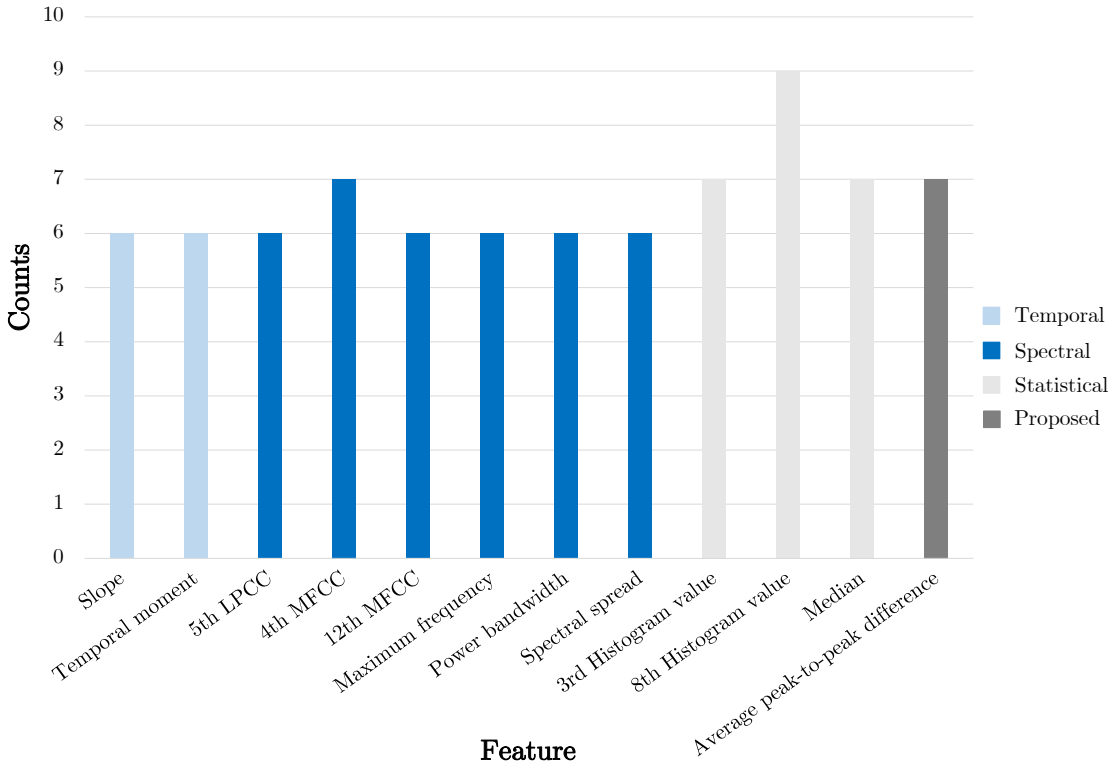



Figure 5.3: Most commonly selected features, when selecting the 30 best from all features extracted from the three channels. The variable counts represents the number of times each feature was selected. Each color represents a feature group.

using parameters from one domain exclusively. The average peak-to-peak difference was the only feature present in Figure 5.3 from the proposed group.

Feature selection was further explored for the features extracted from the right hand. A lesser amount of features needed to be evaluated during the SFS process by accounting for only one channel. Therefore, it was possible to conduct additional experiments in useful time. A specific assessment was performed for this channel's correspondent features, by changing the number of features to select since this parameter had an established value of 30 to reduce processing time when using all channels. The other tested parameter was 'best', which causes the feature selector to return the feature subset that leads to the best predictive performance of the model. When using this channel exclusively, both possibilities were tested: selecting 30 features or selecting the number of features that maximized accuracy. The latter led to an average number of selected features of 25.5 throughout

Table 5.10: Onset classification results obtained by using all feature groups but varying parameters for the number of features and the number of channels, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| Features to select | Channels | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|
| **30** | **All channels** | 75.83 ± 5.78 | 82.80 ± 10.95 | 75.86 ± 4.80 | 77.57 ± 5.33 |
| **30** | **Channel 3** | **79.24 ± 5.72** | **85.29 ± 10.23** | **79.15 ± 5.18** | **80.10 ± 5.89** |
| **'best'** | **Channel 3** | 78.75 ± 3.12 | 81.28 ± 3.18 | 78.63 ± 2.95 | 79.78 ± 3.38 |

the 10 folds. The outcomes of these tests are represented in Table 5.10, which evidences more efficient results for channel 3. A remark should be made regarding the increased accuracy for 30 features, as one would expect that the accuracy would be maximal for the 'best' parameter. This is probably due to the fact that these results are the average of multiple folds, so each fold is not directly compared. The increased standard deviation for the 30 features parameter indicates a higher variation within these results as well.

Subsequently, the most commonly selected features were studied for both parameters, as disposed in Figures 5.4 and 5.5. When performing feature selection for channel 3 individually, the number of times the proposed features are selected increases, which may be an indication that the right first dorsal interosseus muscle had more morphological-related information. Average MUAP rise time, standard deviation of MUAP turns, and peaks rate appear in both experiments, suggesting their importance in the classification task. Figure 5.4 indicates two additional features of the proposed group commonly chosen by the selector: average MUAP duration and standard deviation of peak-to-peak difference.

## 5.4 Feature Importance

For a deeper understanding of the significance of the proposed features, different scoring methods ranked the most important features when the classification task employed all features. These methods do not depend on the relationship between the feature and the respective class but instead try to understand how features influence the model's learning process and performance.

The first scoring method presented in this section is based on the mean decrease in the impurity of the split when the Random Forest trees are being generated. In this case, the features for internal nodes are selected according to the information gain criterion, i.e., the entropy of a feature is used as a measure of impurity of a node, as explained in Section 4.4.1. The feature with the highest decrease in impurity, calculated through the information gain, is selected for the internal node, so the higher the reduction in impurity that feature causes in a node, the more important it is. Thus, the average impurity decrease of a certain feature over all the forest's trees is the measure of importance of

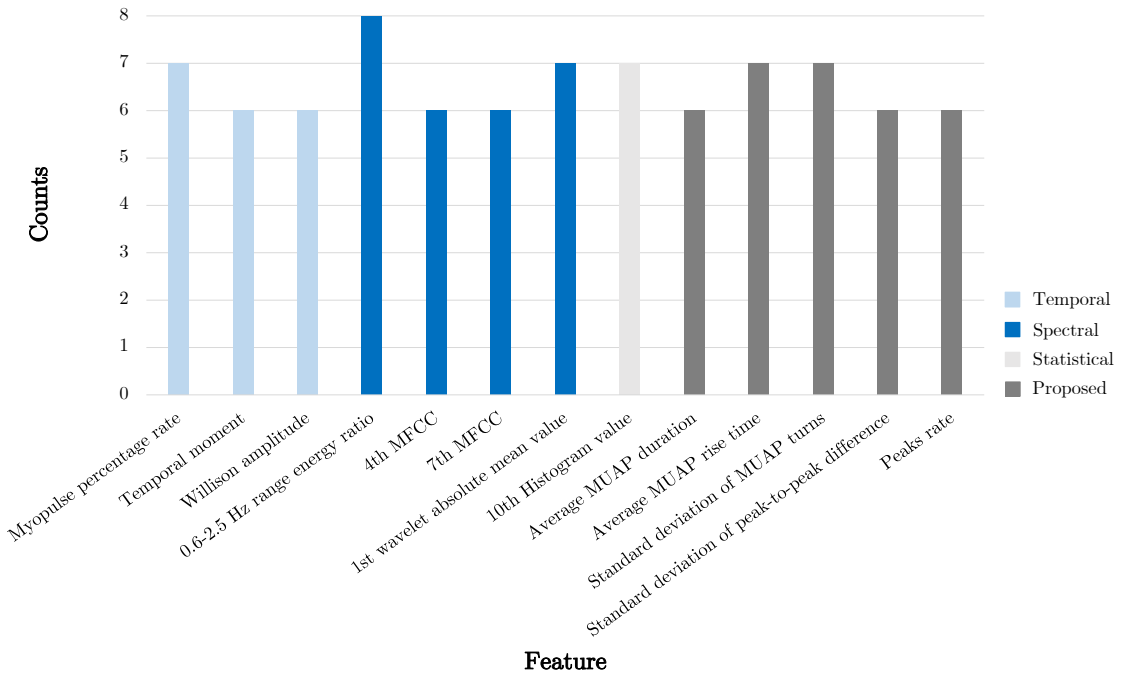



Figure 5.4: Most commonly selected features, when selecting the 30 best from all features extracted from channel 3 only. The variable counts represents the number of times each feature was selected. Each color represents a feature group.

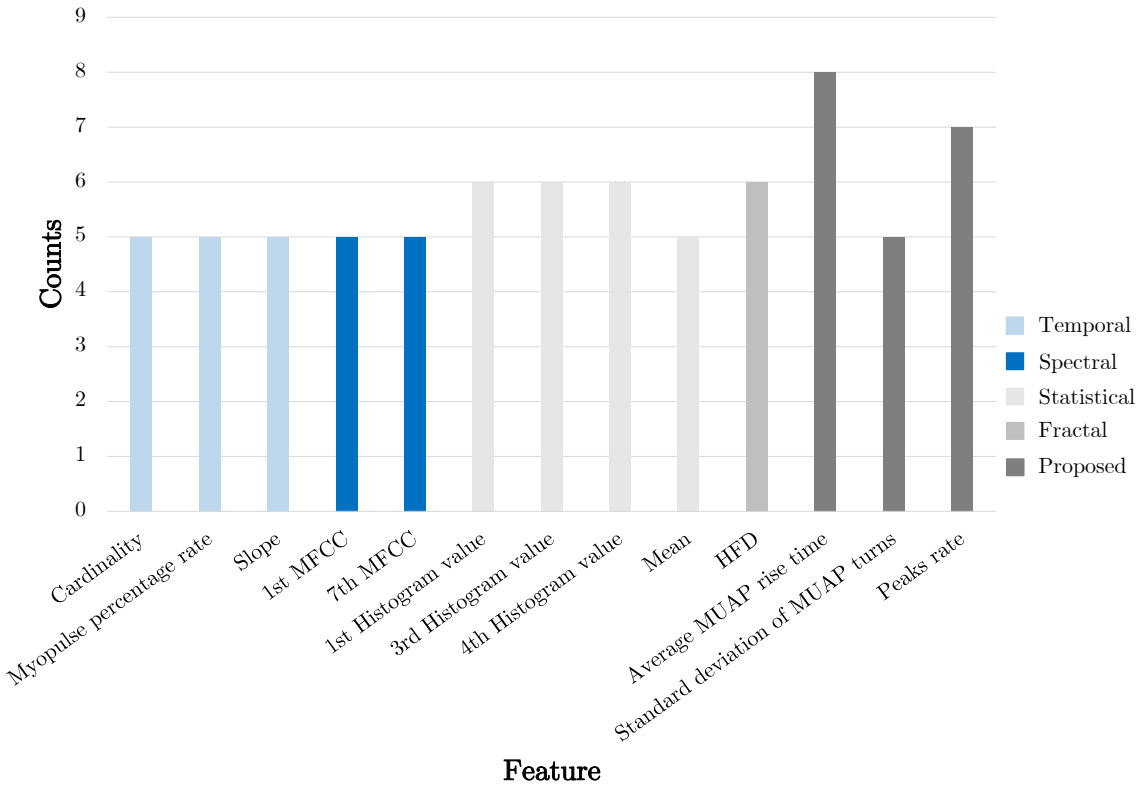



Figure 5.5: Most commonly selected features, when selecting the best feature subset from all features extracted from channel 3 only. The variable counts represents the number of times each feature was selected. Each color represents a feature group.

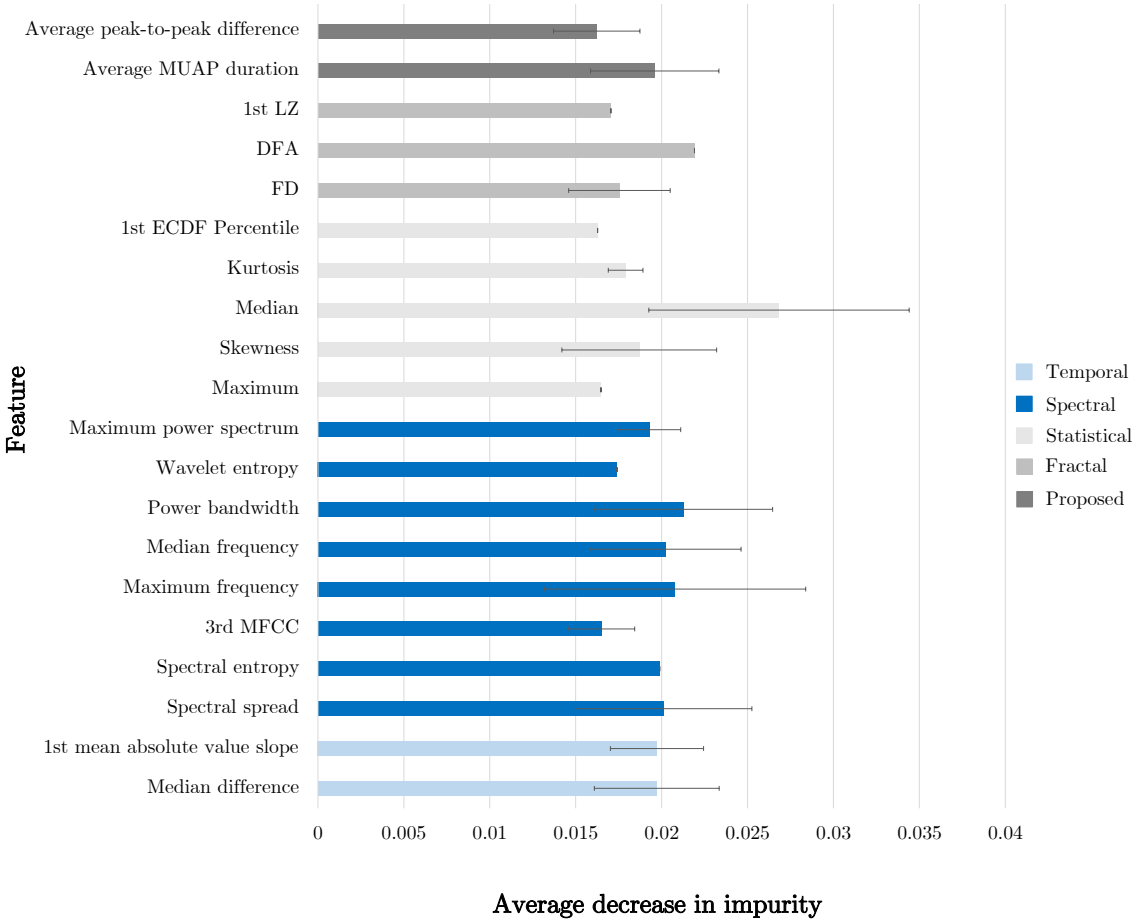



Figure 5.6: Average impurity decrease of the most important features, when using all features. The values are the average of the results of 10 folds, with the respective standard deviation.

that feature. Figure 5.6 comprehends the features that led to overall lower impurities in the nodes of the trees, showing three of the proposed features, peaks rate, average peak-to-peak difference and average MUAP duration, all of which chosen during feature selection experiments as well.

Permutation-based feature importance is a more effective method to evaluate the relevance of each feature, since it is computed on unseen data, whereas entropy-based methods depend on the training set. This method randomly shuffles each feature value and computes the change in the model's accuracy. The features which lead to a higher decrease in the model's performance are considered the most important ones. Thus significant changes in the classifier's score with a certain feature indicate how much the model depends on that feature. Peaks rate and the average MUAP duration appear once again as relevant features in the classifier's predictive power, as Figure 5.7 illustrates. Once again, average MUAP duration and peaks rate rank high amongst all features. Additionally, average peaks difference appears, which is understandable as it is a measure related to the increased MU firing rate, often present in the ALS condition and visible in NEMG.

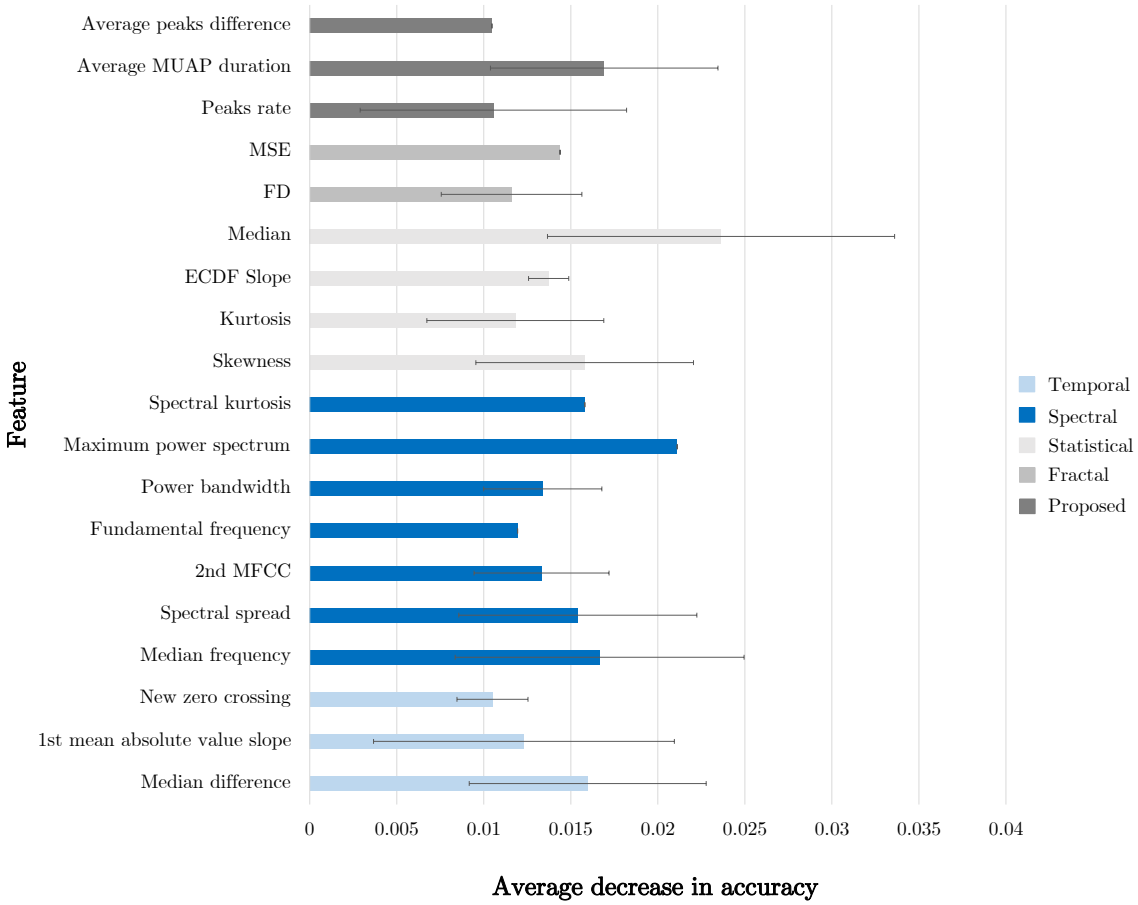



Figure 5.7: Average accuracy decrease of the most important features, when using all features. The values are the average of the results of 10 folds, with the respective standard deviation.

Lastly, features were descriptively analyzed through their boxplot representations for each class. The features peaks rate, average peak-to-peak difference, and average MUAP duration, whose boxplots are represented in Figures 5.8, 5.9 and 5.10 respectively, show promising results with relatively distinctive values for each class. Therefore, it is plausible that these parameters have the ability to distinguish the two types of subjects effectively. The mean value of peaks rate was 6.61 ± 0.91 peaks/s for the healthy class, and 7.51 ± 1.00 peaks/s for the ALS class. This goes in line with what is established for NEMG, which is a higher number of peaks in pathological samples due to the reduced recruitment pattern manifested in ALS patients. So it is expected that peaks rate shows promising results as differentiating measure. Regarding the average peak-to-peak difference, a mean value of 0.0044 ± 0.0005 ms was obtained for the healthy class, whereas 0.0040 ± 0.0007 ms was obtained for the ALS class, which are not particularly distinctive measures. However, we can visualize a high number of data outliers in this feature's boxplot, which could be influencing the results. Additionally, for the average MUAP duration, a mean value of 4.74 ± 0.68 ms was obtained for the healthy class, and 5.58 ± 0.83 ms was obtained for

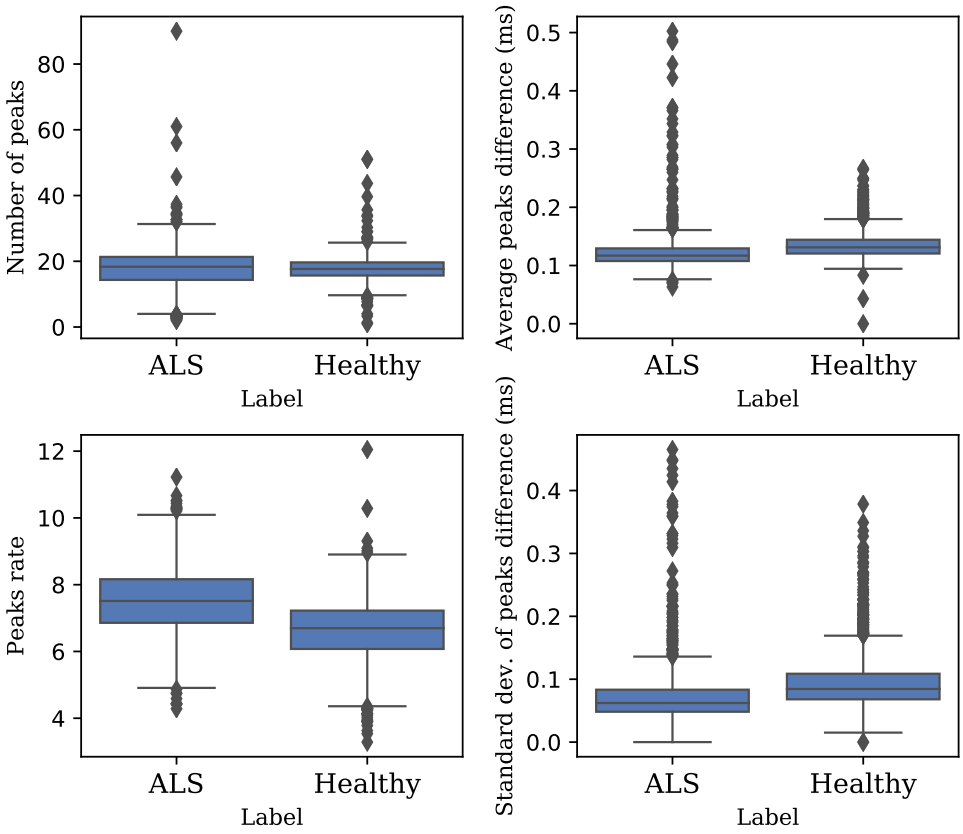



Figure 5.8: Boxplots for the peaks features, displaying the distribution of data based on the minimum, first quartile, median, third quartile, and maximum. The points lying beyond the minimum and maximum values represent outliers.

the ALS class. These results are in agreement with what was expected, since the MUAP duration tends to be higher for pathological NEMG signals, as mentioned in Section 2.3.2, and show, once again, the potential of this feature. These conclusions are consistent with what was observed during feature selection and when calculating feature importance through entropy and permutation, which is a preference of the model for specific features from the proposed set, including the features peaks rate, average peak-to-peak difference, and average MUAP duration.

On the contrary, from the observation of Figure 5.11, we reached the conclusion that some of the novel features present identical values for both groups, such as the measures related to the MUAP integrated area, MUAP phases, and MUAP turns. Most of the features related with the standard deviation did not present relevant results as well. Therefore, these parameters do not represent suitable measures of distinction between ALS patients and HC subjects.

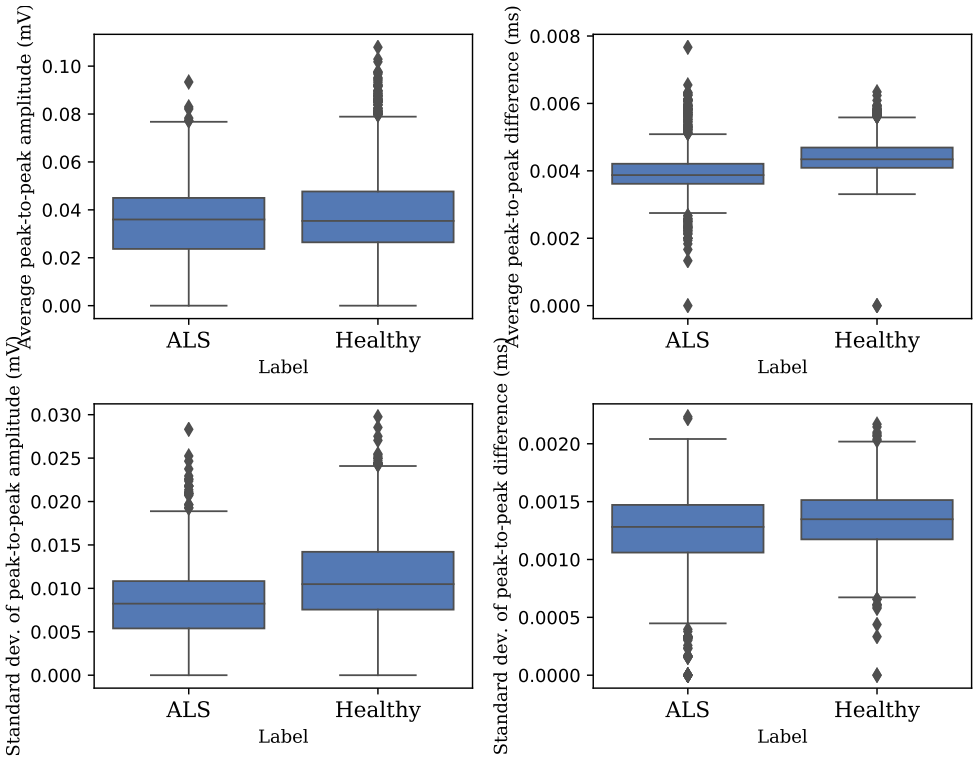



Figure 5.9: Boxplots for the peak-to-peak features, displaying the distribution of data based on the minimum, first quartile, median, third quartile, and maximum. The points lying beyond the minimum and maximum values represent outliers.

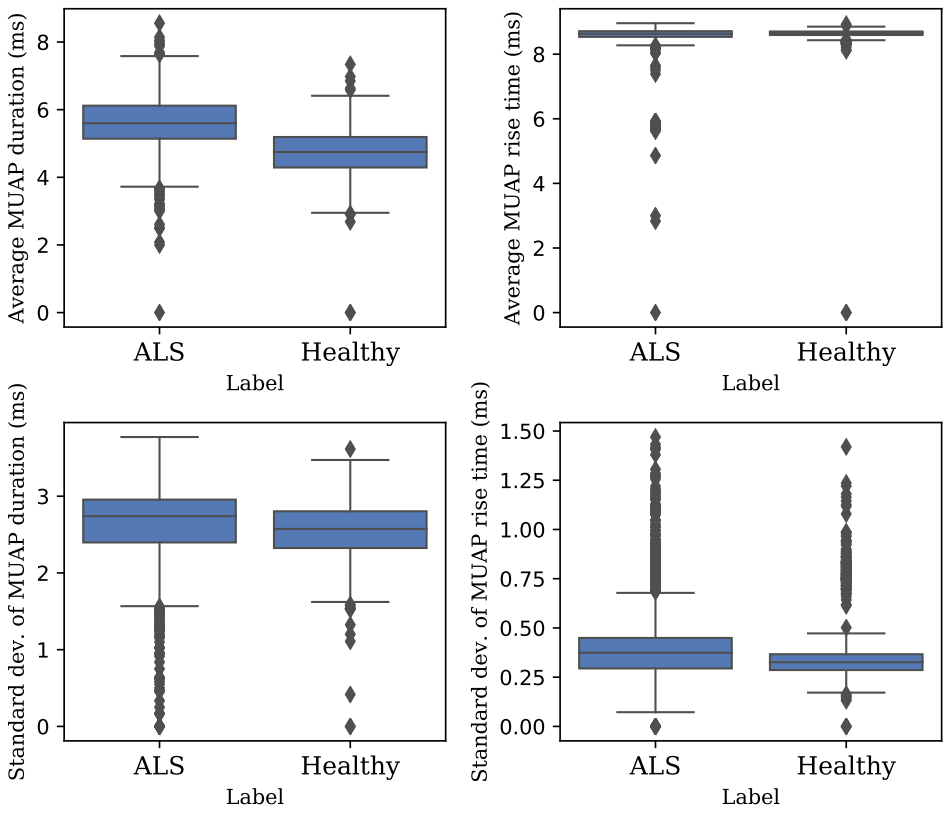



Figure 5.10: Boxplots for the rise time and MUAP duration features, displaying the distribution of data based on the minimum, first quartile, median, third quartile, and maximum. The points lying beyond the minimum and maximum values represent outliers.

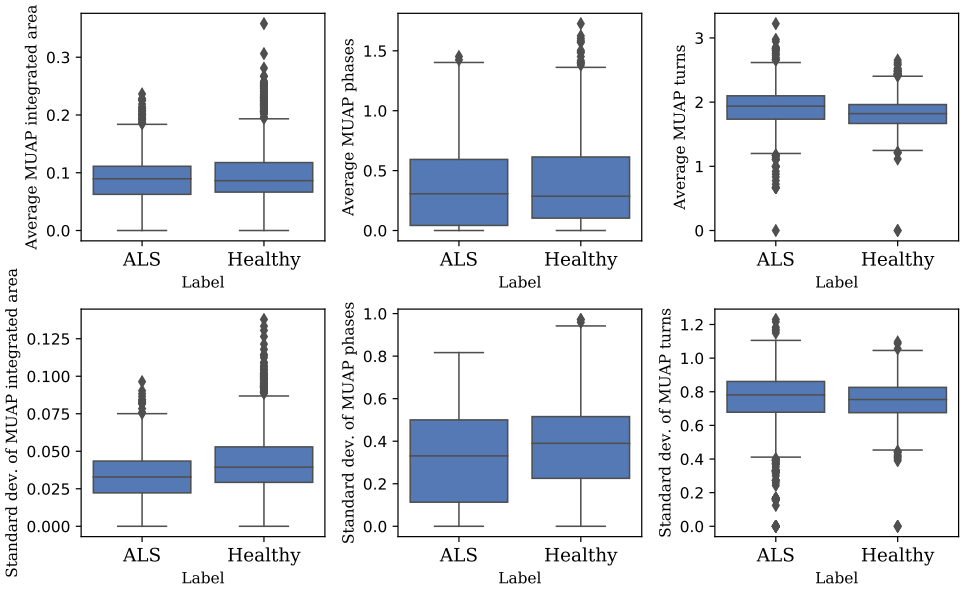



Figure 5.11: Boxplots for the MUAP integrated area, phases and turns, displaying the distribution of data based on the minimum, first quartile, median, third quartile, and maximum. The points lying beyond the minimum and maximum values represent outliers.

# 6

# Conclusions

## 6.1 Overall Conclusions

Nowadays, most patients with ALS die without reaching a concrete diagnosis. ALS is a fast-progressing disease and, therefore, new approaches must be created to provide an early and specific disease assessment. The sooner adequate healthcare is delivered, the better the chances of increasing survival, and patients can be promptly included in clinical trials when neurological defensive therapies are often most effective.

The presented research showed promising results, uncovering the benefits of introducing morphological information into machine learning algorithms classifying SEMG signals. The primary contribution of this study is its convenience and trustworthiness in differentiating ALS subjects from neurologically intact control subjects, using innovative SEMG features related to the signal's peaks and MUAPs' morphology.

The implemented approach for MUAP detection allowed a significant amount of surface MUAPs to be considered. Its low computational complexity make it an interesting approach to be applied in clinical settings.

The results demonstrated the efficiency of the proposed features in a diagnosis classification task, either by themselves or associated with other groups of features commonly used in a SEMG context. When all features were extracted from all the available channels, a mean $F_1$ score of 83.11% was obtained, evidencing the relevance of exploring features from different signal domains.

Another conclusion was drawn out of the results when it was noted that one single acquisition channel led to a similar overall performance when all channels were being used. The proposed features extracted from the right hand led to a mean $F_1$ score of 81.94%. This outcome suggests that it may be possible to obtain a reliable diagnosis using fewer resources during acquisition. This allows for faster computational speed when processing data. Additionally, from an applicable point of view, the use of one channel extracting information from a single muscle is a more convenient data collection setup, which makes the implementation of this solution more practical in an outpatient regime or in a medical appointment.

The features were additionally validated through feature selection, feature importance measures, and descriptive analysis. A few of the proposed features proved their discriminative power potential, reflecting the presence of ALS in the SEMG signal, namely peaks rate, and average MUAP duration. These feature's values are closely related to the physiology of the disease, which might contribute to increase the classification interpretability and foster the adoption in a medical context. The importance of implementing personalized features related to the disease's pathophysiology, lies in their utility for a better understanding of the model and its performance. As part of the medical community still avoids the use of AI techniques due to their lack of transparency, comprehensible features arise as an alternative to mitigate that concern.

In addition, feature selection on features extracted from the right hand demonstrated that the proposed morphological features are more regularly selected when only this one channel is used. Therefore, morphological features may be even more relevant when less information is acquired, or when information from a specific muscle is measured.

Our investigation aims to spark future research on automated analysis methods that are able to exploit the advantages of SEMG. The non-invasive identification of abnormalities in the functional structures of the muscle is of particular relevance for patients that do not tolerate needles or for repetitive monitoring of the muscle activity. The SEMG holds as a promising versatile technique for ALS diagnosis and prognosis and could be applied in a remote setting, increasing the clinician's available data towards better decision-making and understanding of ALS natural history.

## 6.2 Future Work

Firstly, we believe that the designed algorithms for peak and MUAP detection can be further improved, either by applying different acquisition methods or implementing more complex processing tools. The authors of [13], [14] stated that the number of identifiable MUAPs increases with the number of acquisition channels. Nevertheless, our work indicated that the information from one channel might be sufficient to provide a truthful diagnosis, which deepens the relevance of gaining a better understanding regarding the appropriate acquisition tools. The enhancement of these algorithms could further improve the obtained results and shed a light on why some of the designed features were not considered relevant for the classification task.

Additional data from a higher number of subjects could also enable the study of the features' performance in a larger population. Applying the detailed framework in other datasets, even outside the scope of ALS, is paramount to support the hypothesis that the morphological features obtain good results in detecting neuropathic abnormalities.

The progression level of a neuropathic condition can influence the results, as the more advanced the stage of the condition, the more noticeable it tends to be in the collected muscle signals. Combining this contextual information with the remaining data could provide better results.

In the context of disease identification, it is essential to avoid incorrectly identifying a subject as healthy. Therefore the results should be optimized as to minimize the number of falsely healthy misclassifications.

# Bibliography


[1] J. M. Lourenço, *The NOVAthesis LaTeX Template User's Manual*, NOVA University Lisbon, 2021. [Online]. Available: https://github.com/joaomlourenco/novathesis/raw/master/template.pdf (cit. on p. v).

[2] R. S. Snell, *Clinical Neuroanatomy*, 7th ed. Lippincott Williams & Wilkins, 2010, ISBN: 9780781794275 (cit. on pp. 1, 7–11).

[3] L. I. Grad, G. A. Rouleau, J. Ravits, and N. R. Cashman, "Clinical Spectrum of Amyotrophic Lateral Sclerosis (ALS)", *Cold Spring Harb. Perspect. Med.*, vol. 7, no. 8, a024117, Aug. 2017, ISSN: 2157-1422. DOI: 10.1101/cshperspect.a024117. [Online]. Available: http://perspectivesinmedicine.cshlp.org/lookup/doi/10.1101/cshperspect.a024117 (cit. on pp. 1, 11, 13).

[4] L. C. Wijesekera and P. N. Leigh, "Amyotrophic lateral sclerosis", *Orphanet J. Rare Dis.*, vol. 4, no. 1, p. 3, 2009. DOI: 10.1186/1750-1172-4-3. [Online]. Available: http://ojrd.biomedcentral.com/articles/10.1186/1750-1172-4-3 (cit. on pp. 1, 11–13).

[5] R. B. D. M. J. Aminoff, *Encyclopedia of the Neurological Sciences*, 2nd ed. Elsevier Science, Academic Press, 2014, ISBN: 978-0-12-385158-1 (cit. on pp. 1, 9–11, 13, 16).

[6] E. Longinetti and F. Fang, "Epidemiology of amyotrophic lateral sclerosis", *Curr. Opin. Neurol.*, vol. 32, no. 5, pp. 771–776, Oct. 2019, ISSN: 1350-7540. DOI: 10.1097/WCO.0000000000000730. [Online]. Available: http://journals.lww.com/00019052-201910000-00018 (cit. on pp. 1, 2, 11, 13).

[7] E. Tiryaki and H. A. Horak, "ALS and Other Motor Neuron Diseases", *Contin. Lifelong Learn. Neurol.*, vol. 20, no. 5, pp. 1185–1207, Oct. 2014, ISSN: 1080-2371. DOI: 10.1212/01.CON.0000455886.14298.a4. [Online]. Available: http://journals.lww.com/00132979-201410000-00010 (cit. on pp. 1, 11–13).

[8] M. de Carvalho, R. Dengler, A. Eisen, *et al.*, "Electrodiagnostic criteria for diagnosis of ALS", *Clin. Neurophysiol.*, vol. 119, no. 3, pp. 497–503, 2008, ISSN: 13882457. DOI: 10.1016/j.clinph.2007.09.143 (cit. on pp. 1, 16).

[9] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine Learning Interpretability: A Survey on Methods and Metrics", *Electronics*, vol. 8, no. 8, p. 832, Jul. 2019, ISSN: 2079-9292. DOI: 10.3390/electronics8080832. [Online]. Available: https://www.mdpi.com/2079-9292/8/8/832 (cit. on pp. 1, 19).

[10] M.-L. Ong, P. F. Tan, and J. D. Holbrook, "Predicting functional decline and survival in amyotrophic lateral sclerosis", *PLoS One*, vol. 12, no. 4, A. Musaro, Ed., e0174925, Apr. 2017, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0174925. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0174925%20https://dx.plos.org/10.1371/journal.pone.0174925 (cit. on p. 2).

[11] X. Zhang, P. E. Barkhaus, W. Z. Rymer, and P. Zhou, "Machine Learning for Supporting Diagnosis of Amyotrophic Lateral Sclerosis Using Surface Electromyogram", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 96–103, Jan. 2014, ISSN: 1534-4320. DOI: 10.1109/TNSRE.2013.2274658. [Online]. Available: https://ieeexplore.ieee.org/document/6587612/ (cit. on pp. 2, 4).

[12] J. Bashford, K. Mills, and C. Shaw, "The evolving role of surface electromyography in amyotrophic lateral sclerosis: A systematic review", *Clin. Neurophysiol.*, vol. 131, no. 4, pp. 942–950, Apr. 2020, ISSN: 13882457. DOI: 10.1016/j.clinph.2019.12.007. [Online]. Available: https://doi.org/10.1016/j.clinph.2019.12.007%20https://linkinghub.elsevier.com/retrieve/pii/S1388245719313732 (cit. on pp. 2, 16).

[13] M. Gazzoni, D. Farina, and R. Merletti, "A new method for the extraction and classification of single motor unit action potentials from surface EMG signals", *J. Neurosci. Methods*, vol. 136, no. 2, pp. 165–177, Jul. 2004, ISSN: 01650270. DOI: 10.1016/j.jneumeth.2004.01.002. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0165027004000299 (cit. on pp. 3, 68).

[14] R. Merletti, A. Holobar, and D. Farina, "Analysis of motor units with high-density surface electromyography", *J. Electromyogr. Kinesiol.*, vol. 18, no. 6, pp. 879–890, Dec. 2008, ISSN: 10506411. DOI: 10.1016/j.jelekin.2008.09.002. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1050641108001399 (cit. on pp. 3, 68).

[15] C. Spiewak, "A Comprehensive Study on EMG Feature Extraction and Classifiers", *Open Access J. Biomed. Eng. Biosci.*, vol. 1, no. 1, pp. 17–26, Feb. 2018, ISSN: 26374579. DOI: 10.32474/OAJBEB.2018.01.000104. [Online]. Available: https://www.lupinepublishers.com/biomedical-sciences-journal/fulltext/a-comprehensive-study-on-eMG-feature-extraction-and-classifiers.ID.000104.php (cit. on p. 3).

[16] N. F. Güler and S. Koçer, "Classification of EMG signals using PCA and FFT", *J. Med. Syst.*, vol. 29, no. 3, pp. 241–250, 2005, ISSN: 01485598. DOI: 10.1007/s10916-005-5184-7 (cit. on p. 3).

[17] B. Cesqui, P. Tropea, S. Micera, and H. I. Krebs, "EMG-based pattern recognition approach in post stroke robot-aided rehabilitation: A feasibility study", *J. Neuroeng. Rehabil.*, vol. 10, no. 1, p. 75, 2013, ISSN: 17430003. DOI: 10.1186/1743-0003-10-75. [Online]. Available: http://jneuroengrehab.biomedcentral.com/articles/10.1186/1743-0003-10-75 (cit. on p. 3).

[18] P. Karthick and S. Ramakrishnan, "Surface electromyography based muscle fatigue progression analysis using modified B distribution time–frequency features", *Biomed. Signal Process. Control*, vol. 26, pp. 42–51, Apr. 2016, ISSN: 17468094. DOI: 10.1016/j.bspc.2015.12.007. [Online]. Available: http://dx.doi.org/10.1016/j.bspc.2015.12.007%20https://linkinghub.elsevier.com/retrieve/pii/S1746809415002050 (cit. on pp. 3, 15, 33).

[19] T. Triwiyanto, O. Wahyunggoro, H. A. Nugroho, and H. Herianto, "An Investigation Into Time Domain Features of Surface Electromyography to Estimate the Elbow Joint Angle", *Adv. Electr. Electron. Eng.*, vol. 15, no. 3, pp. 448–458, Oct. 2017, ISSN: 1804-3119. DOI: 10.15598/aeee.v15i3.2177. [Online]. Available: http://advances.utc.sk/index.php/AEEE/article/view/2177 (cit. on p. 3).

[20] T. Tuncer, S. Dogan, and A. Subasi, "Surface EMG signal classification using ternary pattern and discrete wavelet transform based feature extraction for hand movement recognition", *Biomed. Signal Process. Control*, vol. 58, p. 101 872, Apr. 2020, ISSN: 17468094. DOI: 10.1016/j.bspc.2020.101872. [Online]. Available: https://doi.org/10.1016/j.bspc.2020.101872%20https://linkinghub.elsevier.com/retrieve/pii/S1746809420300288 (cit. on p. 3).

[21] I. Elamvazuthi, G. A. Ling, K. A. R. K. Nurhanim, P. Vasant, and S. Parasuraman, "Surface electromyography (sEMG) feature extraction based on Daubechies wavelets", in *2013 IEEE 8th Conf. Ind. Electron. Appl.*, IEEE, Jun. 2013, pp. 1492–1495, ISBN: 978-1-4673-6322-8. DOI: 10.1109/ICIEA.2013.6566603. [Online]. Available: http://ieeexplore.ieee.org/document/6566603/ (cit. on p. 3).

[22] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Fractal analysis features for weak and single-channel upper-limb EMG signals", *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11 156–11 163, Sep. 2012, ISSN: 09574174. DOI: 10.1016/j.eswa.2012.03.039. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2012.03.039%20https://linkinghub.elsevier.com/retrieve/pii/S0957417412005623 (cit. on pp. 3, 35).

[23] S. Anowarul Fattah, "Identifying the Motor Neuron Disease in EMG Signal Using Time and Frequency Domain Features with Comparison", *Signal Image Process. An Int. J.*, vol. 3, no. 2, pp. 99–114, Apr. 2012, ISSN: 22293922. DOI: 10.5121/sipij.2012.3207. [Online]. Available: http://www.aircconline.com/sipij/V3N2/3212sipij07.pdf (cit. on p. 3).

[24] A. R. Verma and B. Gupta, "Detecting Neuromuscular Disorders Using EMG Signals Based on TQWT Features", *Augment. Hum. Res.*, vol. 5, no. 1, p. 8, Dec. 2020, ISSN: 2365-4317. DOI: 10.1007/s41133-019-0020-7. [Online]. Available: https://doi.org/10.1007/s41133-019-0020-7%20http://link.springer.com/10.1007/s41133-019-0020-7 (cit. on p. 3).

[25] N. Sengar, M. K. Dutta, and C. M. Travieso, "Identification of amyotrophic lateral sclerosis using EMG signals", in *2017 4th IEEE Uttar Pradesh Sect. Int. Conf. Electr. Comput. Electron.*, vol. 2018-Janua, IEEE, Oct. 2017, pp. 468–471, ISBN: 978-1-5386-3004-4. DOI: 10.1109/UPCON.2017.8251093. [Online]. Available: http://ieeexplore.ieee.org/document/8251093/ (cit. on p. 3).

[26] M. de Carvalho and M. Swash, "Fasciculation potentials and earliest changes in motor unit physiology in ALS", *J. Neurol. Neurosurg. Psychiatry*, vol. 84, no. 9, pp. 963–968, Sep. 2013, ISSN: 0022-3050. DOI: 10.1136/jnnp-2012-304545. [Online]. Available: https://jnnp.bmj.com/lookup/doi/10.1136/jnnp-2012-304545 (cit. on p. 3).

[27] M. Kefalas, M. Koch, V. Geraedts, H. Wang, M. Tannemaat, and T. Back, "Automated Machine Learning for the Classification of Normal and Abnormal Electromyography Data", in *2020 IEEE Int. Conf. Big Data (Big Data)*, IEEE, Dec. 2020, pp. 1176–1185, ISBN: 978-1-7281-6251-5. DOI: 10.1109/BigData50022.2020.9377780. [Online]. Available: https://ieeexplore.ieee.org/document/9377780/ (cit. on p. 3).

[28] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)", *Neurocomputing*, vol. 307, pp. 72–77, Sep. 2018, ISSN: 09252312. DOI: 10.1016/j.neucom.2018.03.067. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925231218304843 (cit. on p. 3).

[29] A. Mokdad, S. M. El Amine Debbal, and F. Meziani, "Diagnosis of amyotrophic lateral sclerosis (ALS) disorders based on electromyogram (EMG) signal analysis and feature selection", *Polish J. Med. Phys. Eng.*, vol. 26, no. 3, pp. 155–160, Sep. 2020, ISSN: 18980309. DOI: 10.2478/pjmpe-2020-0018. [Online]. Available: https://www.sciendo.com/article/10.2478/pjmpe-2020-0018 (cit. on pp. 3, 31).

[30] A. Tamborska, J. Bashford, A. Wickham, *et al.*, "Non-invasive measurement of fasciculation frequency demonstrates diagnostic accuracy in amyotrophic lateral sclerosis", *Brain Commun.*, vol. 2, no. 2, pp. 1–9, Jul. 2020, ISSN: 2632-1297. DOI: 10.1093/braincomms/fcaa141. [Online]. Available: https://academic.oup.com/braincomms/article/doi/10.1093/braincomms/fcaa141/5902444 (cit. on p. 4).

[31] P. Zhou, X. Li, F. Jahanmiri-Nezhad, W. Rymer, and P. E. Barkhaus, "Duration of observation required in detecting fasciculation potentials in amyotrophic lateral sclerosis using high-density surface EMG", *J. Neuroeng. Rehabil.*, vol. 9, no. 1, p. 78, 2012, ISSN: 1743-0003. DOI: 10.1186/1743-0003-9-78. [Online]. Available: http://jneuroengrehab.biomedcentral.com/articles/10.1186/1743-0003-9-78 (cit. on p. 4).

[32] M. Camara, M. De Carvalho, T. Araújo, H. Gamboa, and C. Quintão, "Coherence and Phase Locking Disruption in Electromyograms of Patients with Amyotrophic Lateral Sclerosis", in *Proc. Int. Conf. Bio-inspired Syst. Signal Process.*, SCITEPRESS - Science, 2014, pp. 163–170, ISBN: 978-989-758-011-6. DOI: 10.5220/0004749901630170. [Online]. Available: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0004749901630170 (cit. on p. 4).

[33] C. Quintão, R. Vigário, M. M. Santos, *et al.*, "Surface electromyography for testing motor dysfunction in amyotrophic lateral sclerosis", *Neurophysiol. Clin.*, no. xxxx, Jun. 2021, ISSN: 09877053. DOI: 10.1016/j.neucli.2021.06.001. [Online]. Available: https://doi.org/10.1016/j.neucli.2021.06.001%20https://linkinghub.elsevier.com/retrieve/pii/S098770532100068X (cit. on pp. 4, 57).

[34] R. Istenič, P. A. Kaplanis, C. S. Pattichis, and D. Zazula, "Multiscale entropy-based approach to automated surface EMG classification of neuromuscular disorders", *Med. Biol. Eng. Comput.*, vol. 48, no. 8, pp. 773–781, Aug. 2010, ISSN: 0140-0118. DOI: 10.1007/s11517-010-0629-7. [Online]. Available: http://link.springer.com/10.1007/s11517-010-0629-7 (cit. on pp. 4, 36, 57).

[35] F. Fernandes, I. Barbalho, D. Barros, *et al.*, "Biomedical signals and machine learning in amyotrophic lateral sclerosis: a systematic review", *Biomed. Eng. Online*, vol. 20, no. 1, pp. 1–22, 2021, ISSN: 1475925X. DOI: 10.1186/s12938-021-00896-2. [Online]. Available: https://doi.org/10.1186/s12938-021-00896-2 (cit. on p. 4).

[36] D. Farina, F. Negro, M. Gazzoni, and R. M. Enoka, "Detecting the Unique Representation of Motor-Unit Action Potentials in the Surface Electromyogram", *J. Neurophysiol.*, vol. 100, no. 3, pp. 1223–1233, Sep. 2008, ISSN: 0022-3077. DOI: 10.1152/jn.90219.2008. [Online]. Available: https://www.physiology.org/doi/10.1152/jn.90219.2008 (cit. on pp. 5, 15).

[37] C. J. De Luca, A. Adam, R. Wotiz, L. D. Gilmore, and S. H. Nawab, "Decomposition of surface EMG signals", *J. Neurophysiol.*, vol. 96, no. 3, pp. 1646–1657, 2006, ISSN: 00223077. DOI: 10.1152/jn.00009.2006 (cit. on p. 5).

[38] V. Singh, *Textbook of Clinical Neuroanatomy*, 2nd ed. Elsevier, 2010, ISBN: 978-8131223079 (cit. on pp. 7–9).

[39] K. Punkt, *Fibre types in skeletal muscles*, 1st ed., ser. Advances in Anatomy Embryology and Cell Biology. Springer-Verlag, 2002, vol. 162 (cit. on p. 8).

[40] K. N. H. Elaine N. Marieb, *Human Anatomy & Physiology*, 10th, global. Pearson, 2015, ISBN: 9781292096971 (cit. on pp. 9, 10).

[41] J. Duchateau and R. M. Enoka, "Human motor unit recordings: Origins and insight into the integrated motor system", *Brain Res.*, vol. 1409, pp. 42–61, Aug. 2011, ISSN: 00068993. DOI: 10.1016/j.brainres.2011.06.011. [Online]. Available: http://dx.doi.org/10.1016/j.brainres.2011.06.011%20https://linkinghub.elsevier.com/retrieve/pii/S0006899311010584 (cit. on pp. 10, 15).

[42] C. K. Thompson, F. Negro, M. D. Johnson, *et al.*, "Robust and accurate decoding of motoneuron behaviour and prediction of the resulting force output", *J. Physiol.*, vol. 596, no. 14, pp. 2643–2659, Jul. 2018, ISSN: 00223751. DOI: 10.1113/JP276153. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1113/JP276153 (cit. on p. 10).

[43] T. E. Andreoli, I. Benjamin, R. C. Griggs, E. J. Wing, and J. G. Fitz, *Andreoli and Carpenter's Cecil Essentials of Medicine*, 8th ed. Saunders, 2010, pp. 1078–1079, ISBN: 9781416061090 (cit. on pp. 11–13).

[44] M. A. J. van den Bos, N. Geevasinga, M. Higashihara, P. Menon, and S. Vucic, "Pathophysiology and Diagnosis of ALS: Insights from Advances in Neurophysiological Techniques", *Int. J. Mol. Sci.*, vol. 20, no. 11, p. 2818, Jun. 2019, ISSN: 1422-0067. DOI: 10.3390/ijms20112818. [Online]. Available: https://www.mdpi.com/1422-0067/20/11/2818 (cit. on pp. 11, 12).

[45] B. R. Brooks, R. G. Miller, M. Swash, and T. L. Munsat, "El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis", *Amyotroph. Lateral Scler. Other Mot. Neuron Disord.*, vol. 1, no. 5, pp. 293–299, Jan. 2000, ISSN: 1466-0822. DOI: 10.1080/146608200300079536. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/146608200300079536 (cit. on pp. 12, 13, 16).

[46] M.-H. Soriani and C. Desnuelle, "Care management in amyotrophic lateral sclerosis", *Rev. Neurol. (Paris).*, vol. 173, no. 5, pp. 288–299, May 2017, ISSN: 00353787. DOI: 10.1016/j.neurol.2017.03.031. [Online]. Available: http://dx.doi.org/10.1016/j.neurol.2017.03.031%20https://linkinghub.elsevier.com/retrieve/pii/S0035378716302788 (cit. on p. 13).

[47] J. M. Cedarbaum, N. Stambler, E. Malta, *et al.*, "The ALSFRS-R: A revised ALS functional rating scale that incorporates assessments of respiratory function", *J. Neurol. Sci.*, vol. 169, no. 1-2, pp. 13–21, 1999, ISSN: 0022510X. DOI: 10.1016/S0022-510X(99)00210-5 (cit. on pp. 13, 14).

[48] M. Proudfoot, A. Jones, K. Talbot, A. Al-Chalabi, and M. R. Turner, "The ALSFRS as an outcome measure in therapeutic trials and its relationship to symptom onset", *Amyotroph. Lateral Scler. Front. Degener.*, vol. 17, no. 5-6, pp. 414–425, Aug. 2016, ISSN: 2167-8421. DOI: 10.3109/21678421.2016.1140786. [Online]. Available: https://www.tandfonline.com/doi/full/10.3109/21678421.2016.1140786 (cit. on p. 14).

[49] P. H. Gordon, R. G. Miller, and D. H. Moore, "ALSFRS-R", *Amyotroph. Lateral Scler. Other Mot. Neuron Disord.*, vol. 5, no. sup1, pp. 90–93, Sep. 2004, ISSN: 1466-0822. DOI: 10.1080/17434470410019906. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/17434470410019906 (cit. on p. 14).

[50] B. R. Brooks, J. A. Jorgenson, B. J. Newhouse, J. M. Shefner, and W. Agnese, "Edaravone in the treatment of amyotrophic lateral sclerosis: efficacy and access to therapy - a roundtable discussion", *Am. J. Manag. Care*, vol. 24, no. 9, S175–S186, 2018, ISSN: 19362692 (cit. on p. 14).

[51] P. Narayanaswami, "The Spectrum of Functional Rating Scales in Neurology Clinical Trials", *Neurotherapeutics*, vol. 14, no. 1, pp. 161–175, Jan. 2017. DOI: 10.1007/s13311-016-0488-5. [Online]. Available: http://dx.doi.org/10.1007/s13311-016-0488-5%20http://link.springer.com/10.1007/s13311-016-0488-5 (cit. on p. 14).

[52] H. Tankisi, D. Burke, L. Cui, *et al.*, "Standards of instrumentation of EMG", *Clinical Neurophysiology*, vol. 131, no. 1, pp. 243–258, 2020, ISSN: 1388-2457. DOI: https://doi.org/10.1016/j.clinph.2019.07.025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245719311782 (cit. on p. 14).

[53] A. Phinyomark, E. Campbell, and E. Scheme, "Surface Electromyography (EMG) Signal Processing, Classification, and Practical Considerations", in *Biomed. Signal Process. Ser. Bioeng.* Springer, Singapore, 2020, pp. 3–29, ISBN: 9789811390968. DOI: 10.1007/978-981-13-9097-5_1. [Online]. Available: http://link.springer.com/10.1007/978-981-13-9097-5%7B%5C_%7D1 (cit. on p. 14).

[54] G. D. Meekins, Y. So, and D. Quan, "American Association of Neuromuscular & Electrodiagnostic Medicine evidenced-based review: Use of surface electromyography in the diagnosis and study of neuromuscular disorders", *Muscle Nerve*, vol. 38, no. 4, pp. 1219–1224, Oct. 2008, ISSN: 0148639X. DOI: 10.1002/mus.21055. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/mus.21055 (cit. on p. 15).

[55] L. A. Green, A. Christie, and D. A. Gabriel, "Spike shape analysis for the surface and needle electromyographic interference pattern", *Biomed. Signal Process. Control*, vol. 36, pp. 1–10, Jul. 2017, ISSN: 17468094. DOI: 10.1016/j.bspc.2017.03.006. [Online]. Available: http://dx.doi.org/10.1016/j.bspc.2017.03.00

6%20https://linkinghub.elsevier.com/retrieve/pii/S1746809417300575 (cit. on p. 15).

[56] K. R. Mills, "The basics of electromyography", *J. Neurol. Neurosurg. Psychiatry*, vol. 76, no. suppl_2, pp. ii32–ii35, Jun. 2005. DOI: 10.1136/jnnp.2005.069211. [Online]. Available: https://jnnp.bmj.com/lookup/doi/10.1136/jnnp.2005.069211 (cit. on p. 16).

[57] G. R. Naik, S. E. Selvan, and H. T. Nguyen, "Single-Channel EMG Classification With Ensemble-Empirical-Mode-Decomposition-Based ICA for Diagnosing Neuromuscular Disorders", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 7, pp. 734–743, Jul. 2016, ISSN: 1534-4320. DOI: 10.1109/TNSRE.2015.2454503. [Online]. Available: https://ieeexplore.ieee.org/document/7153535/ (cit. on p. 17).

[58] M. Nikolic, "Detailed Analysis of Clinical Electromyography Signals EMG Decomposition, Findings and Firing Pattern Analysis in Controls and Patients with Myopathy and Amytrophic Lateral Sclerosis", Ph.D. dissertation, University of Copenhagen, 2001 (cit. on p. 17).

[59] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What do we need to build explainable AI systems for the medical domain?", no. January 2018, Dec. 2017, ISSN: 2331-8422. arXiv: 1712.09923. [Online]. Available: http://arxiv.org/abs/1712.09923 (cit. on p. 17).

[60] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*, 2nd ed., ser. Adaptive Computation and Machine Learning. The MIT Press, 2012, pp. 1–6, ISBN: 978-0262018258 (cit. on pp. 17, 18).

[61] X.-D. Zhang, "Machine learning", in *A Matrix Algebra Approach to Artificial Intelligence*. Singapore: Springer Singapore, 2020, pp. 223–440, ISBN: 978-981-15-2770-8. DOI: 10.1007/978-981-15-2770-8_6. [Online]. Available: https://doi.org/10.1007/978-981-15-2770-8_6 (cit. on pp. 17, 18).

[62] I. Guyon, *Feature extraction: Foundations and applications*, ser. Studies in Fuzziness and Soft Computing. Springer, 2006, vol. 207 (cit. on pp. 18–21).

[63] C. Molnar, *Interpretable Machine Learning, A Guide for Making Black Box Models Explainable*. 2021, https://christophm.github.io/interpretable-ml-book/ (cit. on pp. 18, 19).

[64] A. Dey, "Machine Learning Algorithms: A Review", *Int. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 3, pp. 1174–1179, 2016, ISSN: 2227524X. [Online]. Available: www.ijcsit.com (cit. on p. 18).

[65] P. Kubben, M. Dumontier, and A. Dekker, *Fundamentals of Clinical Data Science*. Springer International Publishing, 2019 (cit. on pp. 18, 19, 21, 22).

[66] I. G. Maglogiannis, *Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and ... in Artificial Intelligence and Applications)*. 2007, ISBN: 9781586037802. [Online]. Available: libgen.li/file.php?md5=8fe25eba46ec1e098fe97e0d43c078ed (cit. on pp. 18, 19, 21).

[67] X. Ying, "An Overview of Overfitting and its Solutions", *J. Phys. Conf. Ser.*, vol. 1168, no. 2, 2019, ISSN: 17426596. DOI: 10.1088/1742-6596/1168/2/022022 (cit. on p. 18).

[68] W. Samek, T. Wiegand, and K.-R. Müller, "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models", Aug. 2017, ISSN: 2331-8422. arXiv: 1708.08296. [Online]. Available: http://arxiv.org/abs/1708.08296 (cit. on p. 19).

[69] G. Chandrashekar and F. Sahin, "A survey on feature selection methods", *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014, ISSN: 00457906. DOI: 10.1016/j.compeleceng.2013.11.024. [Online]. Available: http://dx.doi.org/10.1016/j.compeleceng.2013.11.024%20https://linkinghub.elsevier.com/retrieve/pii/S0045790613003066 (cit. on pp. 20, 21).

[70] N. Sánchez-Maroño, A. Alonso-Betanzos, and M. Tombilla-Sanromán, "Filter methods for feature selection – a comparative study", in *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, H. Yin, P. Tino, E. Corchado, W. Byrne, and X. Yao, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 178–187, ISBN: 978-3-540-77226-2 (cit. on pp. 20, 21).

[71] A. Ben Brahim and M. Limam, "A hybrid feature selection method based on instance learning and cooperative subset search", *Pattern Recognit. Lett.*, vol. 69, pp. 28–34, Jan. 2016, ISSN: 01678655. DOI: 10.1016/j.patrec.2015.10.005. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2015.10.005%20https://linkinghub.elsevier.com/retrieve/pii/S0167865515003451 (cit. on pp. 20, 21).

[72] N. El Aboudi and L. Benhlima, "Review on wrapper feature selection approaches", in *2016 Int. Conf. Eng. MIS*, IEEE, Sep. 2016, pp. 1–5, ISBN: 978-1-5090-5579-1. DOI: 10.1109/ICEMIS.2016.7745366. [Online]. Available: http://ieeexplore.ieee.org/document/7745366/ (cit. on pp. 20, 21).

[73] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning", in *2014 Sci. Inf. Conf.*, IEEE, Aug. 2014, pp. 372–378, ISBN: 978-0-9893193-1-7. DOI: 10.1109/SAI.2014.6918213. [Online]. Available: https://ieeexplore.ieee.org/document/6918213 (cit. on p. 21).

[74] B. Xue, M. Zhang, and W. N. Browne, "A comprehensive comparison on evolutionary feature selection approaches to classification", *Int. J. Comput. Intell. Appl.*, vol. 14, no. 2, 2015, ISSN: 14690268. DOI: 10.1142/S146902681550008X (cit. on p. 21).

[75] C. W. Chen, Y. H. Tsai, F. R. Chang, and W. C. Lin, "Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results", *Expert Syst.*, vol. 37, no. 5, pp. 1–10, Oct. 2020, ISSN: 0266-4720. DOI: 10.1111/exsy.12553. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/exsy.12553 (cit. on p. 21).

[76] T. Jiang, J. L. Gradus, and A. J. Rosellini, "Supervised Machine Learning: A Brief Primer", *Behav. Ther.*, vol. 51, no. 5, pp. 675–687, Sep. 2020, ISSN: 00057894. DOI: 10.1016/j.beth.2020.05.002. [Online]. Available: https://doi.org/10.1016/j.beth.2020.05.002%20https://linkinghub.elsevier.com/retrieve/pii/S0005789420300678 (cit. on pp. 21, 22).

[77] B. F. Huang and P. C. Boutros, "The parameter sensitivity of random forests", *BMC Bioinformatics*, vol. 17, no. 1, p. 331, Dec. 2016, ISSN: 1471-2105. DOI: 10.1186/s12859-016-1228-x. [Online]. Available: http://dx.doi.org/10.1186/s12859-016-1228-x%20http://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-016-1228-x (cit. on pp. 21, 22).

[78] A. Priyam, R. Gupta, A. Rathee, and S. Srivastava, "Comparative Analysis of Decision Tree Classification Algorithms", *Int. J. Curr. Eng. Technol.*, vol. 3, no. 2, pp. 334–337, 2013 (cit. on p. 21).

[79] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random forests", in *Ensemble Machine Learning: Methods and Applications*, C. Zhang and Y. Ma, Eds. Boston, MA: Springer US, 2012, pp. 157–175, ISBN: 978-1-4419-9326-7. DOI: 10.1007/978-1-4419-9326-7_5. [Online]. Available: https://doi.org/10.1007/978-1-4419-9326-7_5 (cit. on p. 22).

[80] B. Schölkopf, Z. Luo, and V. Vovk, *Empirical Inference*, B. Schölkopf, Z. Luo, and V. Vovk, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 1–287, ISBN: 978-3-642-41135-9. DOI: 10.1007/978-3-642-41136-6. [Online]. Available: http://link.springer.com/10.1007/978-3-642-41136-6 (cit. on p. 22).

[81] Y. Shen, Y. Jiang, W. Liu, and Y. Liu, "Multi-class AdaBoost ELM", in vol. 2, 2015, pp. 179–188. DOI: 10.1007/978-3-319-14066-7_18. [Online]. Available: http://link.springer.com/10.1007/978-3-319-14066-7%7B%5C_%7D18 (cit. on p. 22).

[82] A. Shahraki, M. Abbasi, and Ø. Haugen, "Boosting algorithms for network intrusion detection: A comparative evaluation of Real AdaBoost, Gentle AdaBoost and Modest AdaBoost", *Eng. Appl. Artif. Intell.*, vol. 94, no. July, p. 103770,

Sep. 2020, ISSN: 09521976. DOI: 10.1016/j.engappai.2020.103770. [Online]. Available: https://doi.org/10.1016/j.engappai.2020.103770%20https://linkinghub.elsevier.com/retrieve/pii/S0952197620301706 (cit. on p. 22).

[83] *sklearn.model_selection.GridSearchCV — scikit-learn 0.24.2 documentation*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model%7B%5C_%7Dselection.GridSearchCV.html (visited on 09/23/2021) (cit. on p. 22).

[84] I. Syarif, A. Prugel-Bennett, and G. Wills, "SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance", *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 14, no. 4, p. 1502, Dec. 2016, ISSN: 2302-9293. DOI: 10.12928/telkomnika.v14i4.3956. [Online]. Available: http://journal.uad.ac.id/index.php/TELKOMNIKA/article/view/3956 (cit. on p. 22).

[85] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in Python", *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011 (cit. on pp. 22, 24).

[86] G. S. Handelman, H. K. Kok, R. V. Chandra, *et al.*, "Peering Into the Black Box of", *Am. J. Roentgenol.*, vol. 212, no. 1, pp. 38–43, 2019. DOI: 10.2214/AJR.18.20224 (cit. on pp. 24, 25).

[87] S. Tabe-Bordbar, A. Emad, S. D. Zhao, and S. Sinha, "A closer look at cross-validation for assessing the accuracy of gene regulatory networks and models", *Sci. Rep.*, vol. 8, no. 1, p. 6620, Dec. 2018, ISSN: 2045-2322. DOI: 10.1038/s41598-018-24937-4. [Online]. Available: http://dx.doi.org/10.1038/s41598-018-24937-4%20http://www.nature.com/articles/s41598-018-24937-4 (cit. on p. 23).

[88] D. Berrar, "Cross-Validation", in *Encycl. Bioinforma. Comput. Biol.* January 2018, vol. 1-3, Elsevier, 2019, pp. 542–545. DOI: 10.1016/B978-0-12-809633-8.20349-X. arXiv: 1703.03167. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B978012809633820349X (cit. on p. 23).

[89] *sklearn.model_selection.GroupShuffleSplit — scikit-learn 0.24.2 documentation*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model%7B%5C_%7Dselection.GroupShuffleSplit.html (visited on 09/24/2021) (cit. on p. 24).

[90] O. Caelen, "A Bayesian interpretation of the confusion matrix", *Ann. Math. Artif. Intell.*, vol. 81, no. 3-4, pp. 429–450, Dec. 2017, ISSN: 1012-2443. DOI: 10.1007/s10472-017-9564-8. [Online]. Available: http://link.springer.com/10.1007/s10472-017-9564-8 (cit. on p. 24).

[91] M. Fatourechi, R. K. Ward, S. G. Mason, J. Huggins, A. Schlögl, and G. E. Birch, "Comparison of evaluation metrics in classification applications with imbalanced datasets", in *2008 Seventh International Conference on Machine Learning and Applications*, 2008, pp. 777–782. DOI: 10.1109/ICMLA.2008.34 (cit. on p. 25).

[92] P. Domingos, "A few useful things to know about machine learning", *Commun. ACM*, vol. 55, no. 10, pp. 78–87, Oct. 2012, ISSN: 0001-0782. DOI: 10.1145/2347736.2347755. [Online]. Available: https://dl.acm.org/citation.cfm?id=2347755%20https://dl.acm.org/doi/10.1145/2347736.2347755 (cit. on p. 27).

[93] O. W. Samuel, H. Zhou, X. Li, *et al.*, "Pattern recognition of electromyography signals based on novel time domain features for amputees' limb motion classification", *Comput. Electr. Eng.*, vol. 67, pp. 646–655, Apr. 2018, ISSN: 00457906. DOI: 10.1016/j.compeleceng.2017.04.003. [Online]. Available: http://dx.doi.org/10.1016/j.compeleceng.2017.04.003%20https://linkinghub.elsevier.com/retrieve/pii/S0045790617303932 (cit. on pp. 27, 28).

[94] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification", *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012, ISSN: 09574174. DOI: 10.1016/j.eswa.2012.01.102. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2012.01.102%20https://linkinghub.elsevier.com/retrieve/pii/S0957417412001200 (cit. on pp. 27, 28).

[95] A. Waris and E. N. Kamavuako, "Effect of threshold values on the combination of EMG time domain features: Surface versus intramuscular EMG", *Biomed. Signal Process. Control*, vol. 45, pp. 267–273, 2018, ISSN: 17468108. DOI: 10.1016/j.bspc.2018.05.036. [Online]. Available: https://doi.org/10.1016/j.bspc.2018.05.036 (cit. on pp. 27, 28).

[96] K. S. Kim, H. H. Choi, C. S. Moon, and C. W. Mun, "Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions", *Curr. Appl. Phys.*, vol. 11, no. 3, pp. 740–745, May 2011, ISSN: 15671739. DOI: 10.1016/j.cap.2010.11.051. [Online]. Available: http://dx.doi.org/10.1016/j.cap.2010.11.051%20https://linkinghub.elsevier.com/retrieve/pii/S1567173910004153 (cit. on p. 27).

[97] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, and Y. Laurillau, "Feature extraction of the first difference of EMG time series for EMG pattern recognition", *Comput. Methods Programs Biomed.*, vol. 117, no. 2, pp. 247–256, 2014, ISSN: 18727565. DOI: 10.1016/j.cmpb.2014.06.013. [Online]. Available: http://dx.doi.org/10.1016/j.cmpb.2014.06.013 (cit. on p. 27).

[98] A. Subasi, "Classification of EMG signals using combined features and soft computing techniques", *Appl. Soft Comput.*, vol. 12, no. 8, pp. 2188–2198, Aug. 2012, ISSN: 15684946. DOI: 10.1016/j.asoc.2012.03.035. [Online]. Available: http://dx.doi.org/10.1016/j.asoc.2012.03.035%20https://linkinghub.elsevier.com/retrieve/pii/S1568494612001330 (cit. on p. 27).

[99] J. Too, A. R. Abdullah, and N. M. Saad, "Classification of Hand movements based on discrete wavelet transform and enhanced feature extraction", *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 6, pp. 83–89, 2019, ISSN: 21565570. DOI: 10.14569/ijacsa.2019.0100612 (cit. on pp. 27, 28, 33).

[100] R. N. Khushaba, A. H. Al-Timemy, A. Al-Ani, and A. Al-Jumaily, "A Framework of Temporal-Spatial Descriptors-Based Feature Extraction for Improved Myoelectric Pattern Recognition", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1821–1831, Oct. 2017, ISSN: 1534-4320. DOI: 10.1109/TNSRE.2017.2687520. [Online]. Available: https://ieeexplore.ieee.org/document/7886279/ (cit. on pp. 27, 28).

[101] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition", *J. Neuroeng. Rehabil.*, vol. 7, no. 1, p. 21, Dec. 2010, ISSN: 1743-0003. DOI: 10.1186/1743-0003-7-21. [Online]. Available: https://jneuroengrehab.biomedcentral.com/articles/10.1186/1743-0003-7-21 (cit. on pp. 27, 28, 31).

[102] D. C. Toledo-Perez, J. Rodriguez-Resendiz, and R. A. Gomez-Loenzo, "A Study of Computing Zero Crossing Methods and an Improved Proposal for EMG Signals", *IEEE Access*, vol. 8, pp. 8783–8790, 2020, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2964678. [Online]. Available: https://ieeexplore.ieee.org/document/8951119/ (cit. on p. 27).

[103] "A New Strategy for Multifunction Myoelectric Control", *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, 1993, ISSN: 1558-2531. DOI: 10.1109/10.204774 (cit. on pp. 27, 28).

[104] R. N. Youngworth, B. B. Gallagher, and B. L. Stamper, "An overview of power spectral density (PSD) calculations", in *Optical Manufacturing and Testing VI*, H. P. Stahl, Ed., International Society for Optics and Photonics, vol. 5869, SPIE, 2005, pp. 206–216. DOI: 10.1117/12.618478. [Online]. Available: https://doi.org/10.1117/12.618478 (cit. on p. 31).

[105] M. Barandas, D. Folgado, L. Fernandes, *et al.*, "TSFEL: Time Series Feature Extraction Library", *SoftwareX*, vol. 11, p. 100456, Jan. 2020, ISSN: 23527110. DOI: 10.1016/j.softx.2020.100456. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352711020300017 (cit. on pp. 31, 33).

[106] M. Yoshikawa, M. Mikawa, and K. Tanaka, "Hand pose estimation using EMG signals", *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, no. 1, pp. 4830–4833, 2007, ISSN: 05891019. DOI: 10.1109/IEMBS.2007.4353421 (cit. on p. 31).

[107] C. Jing, D. Chablat, F. Bennis, and L. Ma, "Estimating the emg response exclusively to fatigue during sustained static maximum voluntary contraction", in May 2016, vol. 489, pp. 29–39, ISBN: 978-3-319-41693-9. DOI: 10.1007/978-3-319-41694-6_4 (cit. on p. 33).

[108] K. Nazarpour, A. H. Al-Timemy, G. Bugmann, and A. Jackson, "A note on the probability distribution function of the surface electromyogram signal", *Brain Research Bulletin*, vol. 90, pp. 88–91, 2013, ISSN: 0361-9230. DOI: https://doi.org/10.1016/j.brainresbull.2012.09.012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0361923012002122 (cit. on p. 33).

[109] S. Poosapadi Arjunan, "Fractal features of surface electromyogram: A new measure for low level muscle activation", eng, Ph.D. dissertation, RMIT University, 2008 (cit. on p. 35).

[110] S. Poosapadi Arjunan and D. K. Kumar, "Computation of fractal features based on the fractal analysis of surface Electromyogram to estimate force of contraction of different muscles", *Comput. Methods Biomech. Biomed. Engin.*, vol. 17, no. 3, pp. 210–216, Feb. 2014, ISSN: 1025-5842. DOI: 10.1080/10255842.2012.675055. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/10255842.2012.675055 (cit. on p. 35).

[111] S. P. Arjunan and D. K. Kumar, "Fractal properties of surface electromyogram for classification of low-level hand movements from single-channel forearm muscle activity", *Journal of Mechanics in Medicine and Biology*, vol. 11, no. 03, pp. 581–590, 2011. DOI: 10.1142/S0219519411003867. eprint: https://doi.org/10.1142/S0219519411003867. [Online]. Available: https://doi.org/10.1142/S0219519411003867 (cit. on p. 35).

[112] A. Phinyomark, P. Phukpattaranont, C. Limsakul, and M. Phothisonothai, "Electromyography (EMG) signal classification based on detrended fluctuation analysis", *Fluct. Noise Lett.*, vol. 10, no. 03, pp. 281–301, Sep. 2011, ISSN: 0219-4775. DOI: 10.1142/S0219477511000570. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/S0219477511000570 (cit. on p. 35).

[113] L. Sarlabous, A. Torres, J. A. Fiz, J. Morera, and R. Jané, "Index for estimation of muscle force from mechanomyography based on the Lempel–Ziv algorithm", *J. Electromyogr. Kinesiol.*, vol. 23, no. 3, pp. 548–557, Jun. 2013, ISSN: 10506411. DOI: 10.1016/j.jelekin.2012.12.007. [Online]. Available: http://dx.doi.org/10.1016/j.jelekin.2012.12.007%20https://linkinghub.elsevier.com/retrieve/pii/S1050641113000199 (cit. on p. 35).

[114] M. Talebinejad, A. D. Chan, and A. Miri, "A Lempel–Ziv complexity measure for muscle fatigue estimation", *J. Electromyogr. Kinesiol.*, vol. 21, no. 2, pp. 236–241, Apr. 2011, ISSN: 10506411. DOI: 10.1016/j.jelekin.2010.12.003. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1050641110002002 (cit. on p. 35).

[115] Xu Zhang, Xiang Chen, P. E. Barkhaus, and Ping Zhou, "Multiscale Entropy Analysis of Different Spontaneous Motor Unit Discharge Patterns", *IEEE J. Biomed. Heal. Informatics*, vol. 17, no. 2, pp. 470–476, Mar. 2013, ISSN: 2168-2194. DOI: 10.1109/JBHI.2013.2241071. [Online]. Available: http://ieeexplore.ieee.org/document/6415231/ (cit. on p. 36).

[116] M. M. O. A. dos Santos, "Study of the electromyographic signal dynamic behavior in amyotrophic lateral sclerosis (als)", M.S. thesis, Dissertation (Master's) - FCT-NOVA, 2014, p. 76 (cit. on pp. 36, 51, 52).

[117] I. Rodriguez-Carreno, L. Gila-Useros, and A. Malanda-Trigueros, "Motor Unit Action Potential Duration: Measurement and Significance", in *Adv. Clin. Neurophysiol.* Figure 1, InTech, Oct. 2012. DOI: 10.5772/50265. [Online]. Available: http://www.intechopen.com/books/advances-in-clinical-neurophysiology/motor-unit-action-potential-duration-measurement-and-significance (cit. on p. 36).

[118] K. C. McGill, Z. C. Lateva, and H. R. Marateb, "EMGLAB: An interactive EMG decomposition program", *J. Neurosci. Methods*, vol. 149, no. 2, pp. 121–133, Dec. 2005, ISSN: 01650270. DOI: 10.1016/j.jneumeth.2005.05.015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0165027005001810 (cit. on p. 37).

[119] W. Sun, J. Zhu, Y. Jiang, H. Yokoi, and Q. Huang, "One-channel surface electromyography decomposition for muscle force estimation", *Front. Neurorobot.*, vol. 12, no. 5, pp. 1–12, 2018, ISSN: 16625218. DOI: 10.3389/fnbot.2018.00020 (cit. on p. 37).

[120] D. Dumitru and J. C. King, "Motor unit action potential duration and muscle length", *Muscle Nerve*, vol. 22, no. 9, pp. 1188–1195, Sep. 1999, ISSN: 0148-639X. DOI: 10.1002/(SICI)1097-4598(199909)22:9<1188::AID-MUS4>3.0.CO;2-I. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4598(199909)22:9%7B%5C%%7D3C1188::AID-MUS4%7B%5C%%7D3E3.0.CO;2-I (cit. on p. 39).

[121] I. Rodríguez, L. Gila, A. Malanda, *et al.*, "Motor unit action potential duration, I: Variability of manual and automatic measurements", *J. Clin. Neurophysiol.*, vol. 24, no. 1, pp. 52–58, 2007, ISSN: 07360258. DOI: 10.1097/01.wnp.0000236606.53923.0d (cit. on pp. 39, 40).

[122] S. Solnik, P. Rider, K. Steinweg, P. DeVita, and T. Hortobágyi, "Teager–Kaiser energy operator signal conditioning improves EMG onset detection", *Eur. J. Appl. Physiol.*, vol. 110, no. 3, pp. 489–498, Oct. 2010, ISSN: 1439-6319. DOI: 10.1007/s00421-010-1521-8. [Online]. Available: http://link.springer.com/10.1007/s00421-010-1521-8 (cit. on p. 44).

[123] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting", *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997, ISSN: 00220000. DOI: 10.1006/jcss.1997.1504. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S002200009791504X (cit. on p. 47).

# A

# Computational Tools

This section provides a list of the computational tools used throughout the course of this work. We selected Python as the primary programming language, which is a widely used language within the data science community. The tasks described in chapters 3, 4 and 5 were carried out in Python 3.8.3, using the Integrated Development Environment (IDE) of PyCharm 2020.2.2 (Community Edition), with the Anaconda 4.9.2 distribution.

## A.1 Python Packages

Next, the used python packages are listed.

### Data Structure

```
glob2==0.7
numpy==1.20.1
pandas==1.0.5
```

### Feature Design and Extraction

```
hfda==0.1.1
lempel-ziv-complexity==0.2.2
novainstrumentation==0.4
pyentrp==0.6.0
scipy==1.6.1
statsmodels==0.11.1
tsfel==0.1.4
tsfresh==0.18.0
```

### Data Visualization

```
matplotlib==3.1.2
seaborn==0.10.1
```

### Signal Processing

```
biosignalsnotebooks==0.6.5
neurokit2==0.1.1
```

### Machine Learning

```
mlxtend==0.18.0
scikit-learn==0.23.2
```

# B

# Additional Results

This section presents additional results obtained from experiments that were performed when testing the different groups of features and the available data channels. These include both onset and subject classification metrics.

Table B.1: Onset classification results obtained by each individual feature group combined with the proposed group, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| Temporal | Spectral | Statistical | Fractal | Proposed | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|---|---|---|
| × | | | | × | 75.82 ± 6.44 | 82.09 ± 3.67 | 75.56 ± 4.95 | 78.48 ± 5.74 |
| | × | | | × | **81.45 ± 4.23** | **85.31 ± 3.64** | **80.59 ± 3.82** | **83.25 ± 4.09** |
| | | × | | × | 76.21 ± 6.74 | 81.34 ± 5.16 | 75.90 ± 5.40 | 78.94 ± 6.67 |
| | | | × | × | 76.51 ± 10.16 | 82.57 ± 2.76 | 76.83 ± 7.73 | 79.57± 6.85 |

Table B.2: Onset classification results obtained by duos of feature groups, excluding the proposed group, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| Temporal | Spectral | Statistical | Fractal | Proposed | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|---|---|---|
| × | × | | | | 79.75 ± 4.13 | **85.90 ± 2.92** | 79.06 ± 3.86 | 82.04 ± 4.17 |
| × | | × | | | 72.18 ± 6.99 | 78.55 ± 4.42 | 72.31 ± 6.35 | 76.04 ± 5.62 |
| × | | | × | | 73.27 ± 9.36 | 81.18 ± 3.68 | 73.70 ± 7.92 | 77.46 ± 5.33 |
| | × | × | | | **81.11 ± 3.68** | 85.34 ± 2.81 | 80.15 ± 3.27 | **83.34 ± 3.57** |
| | × | | × | | 80.62 ± 5.02 | 84.79 ± 4.18 | **80.16 ± 4.45** | 82.26 ± 4.80 |
| | | × | × | | 70.03 ± 9.46 | 76.85 ± 6.10 | 71.03 ± 7.85 | 74.06 ± 7.12 |

Table B.3: Onset classification results obtained by duos of feature groups in combination with the proposed group, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| Temporal | Spectral | Statistical | Fractal | Proposed | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|---|---|---|
| × | × | | | × | 80.76 ± 5.11 | 85.38 ± 3.98 | 79.92 ± 5.06 | 82.81 ± 4.33 |
| × | | × | | × | 77.93 ± 5.97 | 83.34 ± 4.66 | 77.43 ± 4.62 | 80.41 ± 5.89 |
| × | | | × | × | 77.86 ± 6.34 | 84.00 ± 3.99 | 77.28 ± 6.02 | 80.76 ± 4.73 |
| | × | × | | × | 82.08 ± 3.67 | 85.43 ± 3.55 | 81.62 ± 3.69 | 83.48 ± 3.55 |
| | × | | × | × | **82.37 ± 3.71** | **85.81 ± 3.49** | **81.64 ± 3.30** | **83.89 ± 3.86** |
| | | × | × | × | 75.82 ± 8.78 | 82.27 ± 2.90 | 76.04 ± 7.39 | 79.04 ± 6.52 |

Table B.4: Onset classification results obtained by trios of feature groups, excluding the proposed group, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| Temporal | Spectral | Statistical | Fractal | Proposed | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|---|---|---|
| × | × | × | | | 80.53 ± 5.81 | 85.89 ± 4.47 | 79.49 ± 4.87 | 82.93 ± 5.41 |
| × | × | | × | | **81.12 ± 5.18** | **86.47 ± 4.54** | 80.04 ± 4.94 | **83.40 ± 4.88** |
| × | | × | × | | 73.25 ± 6.39 | 80.99 ± 3.46 | 73.35 ± 5.44 | 76.78 ± 4.94 |
| | × | × | × | | 80.83 ± 4.86 | 84.51 ± 3.40 | **80.31 ± 4.31** | 82.77 ± 4.73 |



Table B.5: Onset classification results obtained by trios of feature groups in combination with the proposed group, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| Temporal | Spectral | Statistical | Fractal | Proposed | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|---|---|---|
| × | × | × | | × | **81.95 ± 3.91** | **86.47 ± 3.04** | 80.84 ± 3.81 | **84.01 ± 3.73** |
| × | × | | × | × | 81.31 ± 4.34 | 85.74 ± 3.76 | 80.43 ± 4.37 | 83.23 ± 3.98 |
| × | | × | × | × | 76.30 ± 9.22 | 82.56 ± 4.89 | 76.10 ± 7.40 | 79.71 ± 6.85 |
| | × | × | × | × | **81.95 ± 3.73** | 86.20 ± 1.98 | **81.12 ± 3.48** | 83.90 ± 3.39 |

Table B.6: Subject classification results obtained by each individual feature group combined with the proposed group, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| Temporal | Spectral | Statistical | Fractal | Proposed | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|---|---|---|
| × | | | | × | 74.97 ± 6.43 | 80.00 ± 6.54 | 75.48 ± 5.75 | 76.52 ± 6.32 |
| | × | | | × | **76.09 ± 5.59** | **80.60 ± 5.23** | **76.71 ± 5.11** | **77.50 ± 5.28** |
| | | × | | × | 70.48 ± 8.42 | 76.81 ± 7.18 | 72.30 ± 7.12 | 72.38 ± 7.91 |
| | | | × | × | 73.26 ± 12.81 | 79.56 ± 5.15 | 75.17 ± 8.89 | 75.24 ± 10.61 |

Table B.7: Subject classification results obtained by duos of feature groups, excluding the proposed group, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| Temporal | Spectral | Statistical | Fractal | Proposed | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|---|---|---|
| × | × | | | | 74.92 ± 6.29 | 81.47 ± 5.60 | 76.06 ± 5.65 | 76.67 ± 5.81 |
| × | | × | | | 67.77 ± 10.30 | 73.34 ± 8.77 | 69.48 ± 9.24 | 70.48 ± 8.73 |
| × | | | × | | 70.41 ± 11.17 | 77.39 ± 5.91 | 72.38 ± 8.81 | 72.86 ± 9.54 |
| | × | × | | | **77.83 ± 5.46** | **81.74 ± 4.82** | **78.20 ± 4.37** | **79.05 ± 5.30** |
| | × | | × | | 72.58 ± 7.01 | 78.33 ± 6.28 | 73.89 ± 6.03 | 74.29 ± 6.46 |
| | | × | × | | 63.11 ± 12.04 | 68.07 ± 11.31 | 65.81 ± 10.28 | 66.19 ± 9.63 |

Table B.8: Subject classification results obtained by duos of feature groups in combination with the proposed group, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| Temporal | Spectral | Statistical | Fractal | Proposed | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|---|---|---|
| × | × | | | × | 76.48 ± 8.70 | 81.39 ± 6.94 | 77.19 ± 8.10 | 78.10 ± 7.74 |
| × | | × | | × | 73.32 ± 7.09 | 78.66 ± 5.71 | 74.25 ± 6.15 | 75.24 ± 6.32 |
| × | | | × | × | 74.34 ± 8.83 | 78.97 ± 7.20 | 75.10 ± 8.24 | 76.02 ± 8.18 |
| | × | × | | × | 77.75 ± 5.80 | 80.94 ± 6.68 | 78.46 ± 5.34 | 78.57 ± 5.73 |
| | × | | × | × | **79.34 ± 7.80** | **83.22 ± 4.29** | **79.89 ± 6.18** | **80.48 ± 7.21** |
| | | × | × | × | 69.33 ± 8.40 | 74.80 ± 6.66 | 70.97 ± 7.43 | 71.43 ± 7.38 |

Table B.9: Subject classification results obtained by trios of feature groups, excluding the proposed group, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| Temporal | Spectral | Statistical | Fractal | Proposed | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|---|---|---|
| × | × | × | | | 74.96 ± 6.49 | 80.13 ± 6.62 | 75.52 ± 5.93 | 76.67 ± 6.19 |
| × | × | | × | | **76.26 ± 10.41** | **81.96 ± 7.96** | **77.14 ± 9.13** | **78.00 ± 9.31** |
| × | | × | × | | 68.67 ± 7.30 | 76.24 ± 6.04 | 70.59 ± 6.50 | 71.43 ± 6.02 |
| | × | × | × | | 76.15 ± 10.04 | 78.87 ± 8.87 | 76.85 ± 8.74 | 77.00 ± 9.81 |



Table B.10: Subject classification results obtained by trios of feature groups in combination with the proposed group, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| Temporal | Spectral | Statistical | Fractal | Proposed | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|---|---|---|
| × | × | × | | × | **76.99 ± 6.96** | **81.95 ± 5.10** | **77.42 ± 5.98** | **78.57 ± 6.48** |
| × | × | | × | × | 74.55 ± 7.22 | 79.77 ± 6.26 | 75.45 ± 6.57 | 76.19 ± 6.73 |
| × | | × | × | × | 70.92 ± 9.38 | 77.26 ± 4.86 | 72.72 ± 6.60 | 72.86 ± 8.26 |
| | × | × | × | × | 76.43 ± 4.77 | 80.74 ± 3.06 | 77.18 ± 3.32 | 77.62 ± 4.79 |

Table B.11: Subject classification results obtained by using all feature groups before and after sample rejection, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **No sample rejection** | 77.75 ± 6.71 | **83.47 ± 5.00** | 77.98 ± 6.07 | 79.52 ± 6.04 |
| **Sample rejection** | **78.85 ± 5.96** | 82.27 ± 5.97 | **78.84 ± 5.29** | **80.00 ± 5.95** |

Table B.12: Subject classification results obtained by using all features from each channel and from all channels simultaneously, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **Channel 1** | 65.52 ± 10.45 | 71.92 ± 9.66 | 67.81 ± 9.44 | 68.10 ± 9.29 |
| **Channel 2** | 66.75 ± 10.80 | 76.11 ± 11.98 | 68.51 ± 9.88 | 69.52 ± 10.48 |
| **Channel 3** | **79.35 ± 8.42** | **83.72 ± 7.91** | **79.39 ± 7.80** | **80.88 ± 7.93** |
| **All channels** | 78.85 ± 5.96 | 82.27 ± 5.97 | 78.84 ± 5.29 | 80.00 ± 5.95 |

Table B.13: Subject classification results obtained by using all feature groups and all features except the proposed ones extracted from channel 3 only, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **Temporal, Spectral, Statistical, Fractal** | 77.31 ± 7.98 | 79.85 ± 8.50 | 77.53 ± 7.31 | 78.10 ± 8.02 |
| **All features** | **79.35 ± 8.42** | **83.72 ± 7.91** | **79.39 ± 7.80** | **80.88 ± 7.93** |

Table B.14: Subject classification results obtained by using each feature group separately extracted from channel 3 only measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **Temporal** | 74.27 ± 6.48 | 81.54 ± 5.27 | 75.21 ± 4.66 | 76.19 ± 6.39 |
| **Spectral** | 76.00 ± 5.87 | 81.14 ± 6.70 | 76.26 ± 5.39 | 77.62 ± 5.65 |
| **Statistical** | 77.56 ± 4.87 | 80.18 ± 5.39 | 77.90 ± 5.23 | 78.57 ± 4.88 |
| **Fractal** | 78.80 ± 9.44 | 83.05 ± 8.61 | 79.32 ± 8.62 | 80.00 ± 8.98 |
| **Proposed** | **83.61 ± 6.90** | **86.64 ± 6.45** | **83.67 ± 6.31** | **84.29 ± 6.75** |

Table B.15: Subject classification results obtained by using all feature groups and 30 features selected with SFS, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| **No Feature Selection** | **78.85 ± 5.96** | 82.27 ± 5.97 | **78.84 ± 5.29** | **80.00 ± 5.95** |
| **Feature Selection** | 75.71 ± 5.59 | **86.53 ± 13.77** | 76.25 ± 4.80 | 77.14 ± 5.13 |

Table B.16: Subject classification results obtained by using all feature groups but varying parameters for the number of features and the number of channels, measured through $F_1$ score, precision, recall, and accuracy. The scores are presented in percentage (%) and were calculated as the average of 10 folds with the respective standard deviation. The best scores per metric are highlighted in bold.

| Features to select | Channels | $F_1$ Score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|
| **30** | **All channels** | 75.71 ± 5.59 | **86.53 ± 13.77** | 76.25 ± 4.80 | 77.14 ± 5.13 |
| **30** | **Channel 3** | 76.91 ± 6.10 | 82.41 ± 12.76 | **77.42 ± 5.03** | 77.62 ± 6.41 |
| **'best'** | **Channel 3** | **77.03 ± 4.95** | 79.77 ± 5.65 | 77.29 ± 4.99 | **78.10 ± 4.86** |