UNIVERSIDADE D
COIMBRA

Fabiano Papaiz

Machine Learning-based Decision Support System for Amyotrophic Lateral Sclerosis Prognosis

PhD Thesis in Informatics Engineering, Intelligent Systems, advised by Professors Joel Perdiz Arrais and Antonio Higor Freire de Morais and presented to the Informatics Engineering Department at the Faculty of Science and Technology of the University of Coimbra.

April 2025

DEPARTAMENTO DE
ENGENHARIA INFORMÁTICA
FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Fabiano Papaiz

Machine Learning-based Decision Support System for Amyotrophic Lateral Sclerosis Prognosis

PhD Thesis in Informatics Engineering, Intelligent Systems, advised by Professors Joel Perdiz Arrais and Antonio Higor Freire de Morais and presented to the Informatics Engineering Department at the Faculty of Science and Technology of the University of Coimbra.

April 2025

DEPARTAMENTO DE
ENGENHARIA INFORMÁTICA
FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Fabiano Papaiz

Sistema de Suporte à Decisão Baseado em Machine Learning Aplicado ao Prognóstico da Esclerose Lateral Amiotrófica

Tese de Doutoramento em Engenharia Informática, Sistemas Inteligentes, orientada pelos Professores Joel Perdiz Arrais e Antonio Higor Freire de Morais e apresentada ao Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Abril 2025

The work presented in this thesis was funded by the Brazilian Ministry of Health, the Laboratory of Technological Innovation in Health (LAIS/HUOL/UFRN), and the Federal Institute of Rio Grande do Norte (IFRN).

Acknowledgements

I want to express my gratitude to:

- God for the gift of life.
- My parents for all the love and education they provided me.
- My beloved wife, Zaira, for her unwavering encouragement and support during the completion of this work, even during the most challenging moments.
- My children, Renan and Caio, for being a constant source of joy in my life.
- Professors Joel Arrais, Ricardo Valentim, Mario Emilio Dourado, Fernanda Brito, and Higor Morais for their valuable advice throughout this project.
- My friends Leo, Alexandro, and Ivanilson.
- Fellow professors at DIATINF/IFRN.
- Fellow researchers at LAIS/UFRN.
- Fellow professors and researchers at CISUC.
- Funding sources:
  - Brazilian Ministry of Healthy
  - Laboratory of Technological Innovation in Health (LAIS/HUOL/UFRN)
  - Federal Institute of Rio Grande do Norte (IFRN)

Abstract

Amyotrophic Lateral Sclerosis (ALS) is a progressive neurodegenerative disease characterized by significant variability in patient outcomes. Reliable prognostic models are crucial for enhancing patient care, guiding treatment strategies, and informing clinical decision-making. This thesis investigates the use of Machine Learning (ML) algorithms and techniques in predicting ALS prognosis, concentrating on two main objectives: (i) identifying ALS patients who are likely to have a short survival period at the time of diagnosis and (ii) forecasting the functional decline of patients over time.

The initial contribution involved developing and evaluating Machine Learning (ML) algorithms that integrated Ensemble and Imbalance Learning techniques to classify patients based on their likelihood of short-term versus non-short-term survival at the time of diagnosis. The objective was to identify individuals at high risk of mortality within 24 months of symptom onset, utilizing patient data typically encountered in everyday clinical practice. Our Ensemble-Imbalance approach was assessed using six ML algorithms as base classifiers. Significantly, our results surpassed those of individual algorithms, achieving a Balanced Accuracy of 88% and a Sensitivity of 96%. Additionally, we employed the Shapley Additive Explanations framework to elucidate the decision-making process of the top-performing model, highlighting the most critical features and their correlations with the target prediction. Moreover, we introduced valuable tools for visualizing and comparing patient similarities, offering insightful perspectives.

The second contribution of this study introduces an innovative method for predicting functional disability over the next 12 months on a month-by-month basis, utilizing patient data collected during the initial three months. This research pioneers the application of autoregressive multi-step multi-output time series forecasting for ALS prognosis, aimed explicitly at forecasting functional decline over time. The study examined static and dynamic features to develop and evaluate deep learning models that employ Gated Recurrent Units and Long Short-Term Memory algorithms. Our method outperformed previous studies, achieving superior results with a significantly smaller set of input features, thus demonstrating enhanced effectiveness.

The final contribution is developing a Clinical Decision Support (CDS) system that delivers information on ALS prognosis based on the target predictions outlined earlier. This marks a significant initial step toward integrating the insights gained about ALS in this thesis into the clinical setting in Brazil. Additionally, this study established a comprehensive Brazilian ALS Database to bolster current and future research endeavors. The purpose of this database is to facilitate the analysis of data collected from Brazilian patients.

The promising findings presented in this thesis are essential in helping physicians deliver timely and appropriate information to patients and their families. They also have the potential to enhance the quality of end-of-life care while facilitating effective treatment and resource planning. These findings could help physicians develop personalized treatment plans and optimize resource management. Fur-

thermore, these results may serve as criteria for including or excluding patients in clinical trials.

Keywords

Amyotrophic Lateral Sclerosis, Disease Prognosis, Machine Learning, Deep Learning, Explainable Artificial Intelligence, Clinical Decision Support, Health Informatics, Time Series Analysis

Resumo

A Esclerose Lateral Amiotrófica (ELA) é uma doença neurodegenerativa progressiva caracterizada por uma significativa variabilidade nos desfechos dos pacientes. Modelos prognósticos confiáveis são fundamentais para aprimorar o atendimento ao paciente, orientar estratégias de tratamento e embasar a tomada de decisões clínicas. Esta tese investiga o uso de algoritmos e técnicas de *Machine Learning* (ML) na previsão do prognóstico da ELA, concentrando-se em dois principais objetivos: (i) identificar pacientes com maior probabilidade de ter um curto período de sobrevivência no momento do diagnóstico e (ii) prever o declínio funcional dos pacientes ao longo do tempo.

A primeira contribuição envolveu o desenvolvimento e a avaliação de algoritmos de ML que integraram técnicas de Aprendizado de Conjuntos (Ensemble) e Aprendizado com Dados Desbalanceados (Imbalance Learning) para classificar os pacientes quanto à probabilidade de sobrevivência de curto ou longo prazo no momento do diagnóstico. O objetivo foi identificar indivíduos com alto risco de mortalidade dentro de 24 meses após o início dos sintomas, utilizando dados clínicos comumente coletados na prática médica diária. Nossa abordagem baseada em Ensemble-Imbalance foi avaliada utilizando seis algoritmos de ML como classificadores base. Notavelmente, nossos resultados superaram os algoritmos individuais, atingindo uma Acurácia Balanceada de 88% e uma Sensibilidade de 96%. Além disso, empregamos o *framework* Shapley Additive Explanations para explicar o processo de tomada de decisão do modelo de melhor desempenho, destacando as características mais críticas e suas correlações com a previsão alvo. Também foram desenvolvidas ferramentas valiosas para visualizar e comparar semelhanças entre pacientes.

A segunda contribuição deste estudo introduz um método inovador para prever o declínio funcional ao longo dos próximos 12 meses (mês-a-mês) utilizando dados coletados nos três primeiros meses. Esta pesquisa é pioneira na aplicação da previsão de séries temporais multietapa e de múltiplas saídas (*autoregressive multi-step multi-output*) para o prognóstico da ELA, visando especificamente a previsão da perda funcional ao longo do tempo. O estudo analisou características estáticas e dinâmicas para desenvolver e avaliar modelos de aprendizado profundo baseados em *Gated Recurrent Units* e *Long Short-Term Memory*. Nossa abordagem superou estudos anteriores, alcançando melhores resultados com um conjunto significativamente menor de variáveis de entrada, demonstrando assim maior eficiência.

A contribuição final consiste no desenvolvimento de um Sistema de Suporte à Decisão Clínica (CDS) que fornece informações sobre o prognóstico da ELA com base nas previsões mencionadas. Isso representa um avanço inicial significativo na integração dos conhecimentos adquiridos nesta tese ao ambiente clínico no Brasil. Além disso, este estudo estabeleceu um Banco de Dados Brasileiro de ELA para apoiar pesquisas atuais e futuras, facilitando a análise de dados coletados de pacientes brasileiros.

Os achados promissores apresentados nesta tese são fundamentais para auxiliar

os médicos a fornecer informações oportunas e adequadas aos pacientes e seus familiares. Além disso, têm o potencial de melhorar a qualidade dos cuidados em final de vida e de otimizar o planejamento do tratamento e a alocação de recursos. Esses resultados também podem servir como critério para inclusão ou exclusão de pacientes em ensaios clínicos.

Palavras-Chave

Esclerose Lateral Amiotrófica, Prognóstico de Doenças, Aprendizado de Máquina, Aprendizado Profundo, Inteligência Artificial Explicável, Suporte à Decisão Clínica, Informática em Saúde, Análise de Séries Temporais

Contents

Acronyms

ALS Amyotrophic Lateral Sclerosis.

ALSFRS ALS Functional Rating Scale.

ALSFRS-R Revised ALS Functional Rating Scale.

Biomarker Biological Marker.

BMI Body Mass Index.

CDS Clinical Decision Support.

CSF Cerebrospinal Fluid.

CV Cross-Validation.

DT Decision Tree.

EC Exclusion Criteria.

EMR Electronic Medical Record.

FVC Forced Vital Capacity.

GRU Gated Recurrent Unit.

IC Inclusion Criteria.

IFRN Federal Institute of Rio Grande do Norte.

k-NN K-Nearest Neighbors.

LAIS Laboratory of Technological Innovation in Health.

LMN Lower Motor Neurons.

LSTM Long Short-Term Memory.

MeSH Medical Subject Heading.

MiToS Milano-Torino staging system.

ML Machine Learning.

MRI Magnetic Resonance Imaging.

NB Naïve Bayes.

NIV Non-Invasive Ventilation.

NN Neural Networks.

PRO-ACT Pooled Resource Open-Access ALS Clinical Trials Database.

R^2 Coefficient of Determination.

RF Random Forest.

RMSE Root Mean Squared Error.

RNN Recurrent Neural Networks.

RQ Research Question.

SHAP Shapley Additive Explanations.

SNIF Sniff Nasal Inspiratory Force.

SVC Slow Vital Capacity.

SVM Support Vector Machines.

UFRN Federal University of Rio Grande do Norte.

UMAP Uniform Manifold Approximation and Projection.

UMN Upper Motor Neurons.

WFN World Federation of Neurology.

XAI Explainable Artificial Intelligence.

List of Figures

List of Tables

Chapter 1

Introduction

In medicine, prognosis is a term related to predicting future events in patients' health. It includes tasks such as estimating the possibility of developing a disease, how quickly the symptoms worsen over time, associated complications, and how long a patient is likely to live (survival time) [1]. The prognosis is based on probability, which indicates it is likely (and not sure) that the disease will follow a determined course. Some complex and rare diseases, such as Amyotrophic Lateral Sclerosis (ALS), make it challenging to determine their prognosis mainly because they have a very heterogeneous development among their patients, leading to the need to research and develop more complex prognosis models to obtain more precise predictive outcomes.

The analysis of medical data usually involves dealing with high-dimensional data, covering a large number of features. It is clear that Artificial Intelligence techniques are vital for analyzing, processing, and extracting information and knowledge from patient data. Machine Learning (ML) has emerged as a powerful tool in improving disease prognosis and can help enhance the performance in the ALS prognosis field.

1.1 Motivation

ALS is a rare, incurable, and progressive disease that impacts the neurons of the human motor system. It is clinically heterogeneous, with varying disease onsets, extra-motor involvements, progression rates, and survival times among patients [2–5]. Despite being discovered over a century ago, accurately predicting the prognosis of ALS patients continues to challenge physicians, particularly in terms of survival time and disease progression. Therefore, finding solutions that can enhance patients' prognoses and help maintain their quality of life is crucial.

Studies on ALS prognosis using more complex and costly features have shown promising results, such as imaging, metabolomics, cerebrospinal fluid, proteomics, and genetics (more detail in Chapter 3). Nowadays, many countries present financial limitations within their healthcare systems. This fact makes it unfeasible to gather complex and costly data in primary care. Therefore, it is essential to

conduct studies considering these limitations and develop computational solutions that can be widely utilized in primary care. One approach to achieving this goal is to use commonly collected features from routine ALS clinical practice. These may include clinical evaluations, assessment of functional capabilities, and respiratory function measurements. These features are often derived from less expensive and less complex procedures, making them more accessible.

The Brazilian Ministry of Health has launched a research project called "Scientific and Technological Development Applied to ALS" to advance studies on ALS disease in 2020 [7]. This project involves funding for researching prognosis solutions, remote respiratory support monitoring, establishing a national ALS registry database, and providing distance learning courses to health workers. The study presented in this thesis was funded by this project.

1.2 Objectives

The main objective of this work is to propose advanced ML algorithms and computational data analysis methods to improve ALS prognosis. The specific main objectives are outlined below.

- **Patient Data Acquisition and Exploratory Data Analysis**: The focus is on the analysis of patient data commonly encountered in routine ALS clinical practice, obtained through a less complex and costly process.
- **Identifying Short-Survival ALS Patients at Diagnosis**: Propose advanced ML algorithms to classify patients into Short and Non-Short survival groups.
- **ALS Progression Forecasting**: Propose advanced ML algorithms and temporal data modeling to forecast the functional decline of patients over time.
- **Decision Support System Development**: Leverage the findings of this thesis to develop a CDS system that provides prognostic information for ALS patients.

1.3 Thesis Structure

This section briefly describes the chapters comprising this thesis to offer a comprehensive view of its structure.

- **Chapter 2** supplies background information related to the main topics explored in this thesis, providing essential context and understanding of the subject matter.
- **Chapter 3** summarizes the state-of-the-art ML approaches employed for ALS prognosis. It specifically focuses on studies using less complex and costly features. Furthermore, this chapter describes the different datasets and features analyzed, along with their availability for access.

- **Chapter 4** focuses on the investigation to accomplish the first target prediction: identifying short-survival ALS patients at diagnosis.
- **Chapter 5** focuses on the investigation to accomplish the second target prediction: ALS progression forecasting.
- **Chapter 6** describes the development of the CDS system to provide ALS prognosis information related to the target predictions.
- **Chapter 7** concludes this thesis by providing an overview of the main findings from this research and the contribution of this work to the ALS prognosis field. It also presents future research directions.

1.4 Scientific Contributions

The contributions of this thesis resulted in the following publications in international peer-reviewed journals and conferences:

Journal papers:

- Machine learning solutions applied to amyotrophic lateral sclerosis prognosis: a review.
  Frontiers in Computer Science. 2022. *DOI: 10.3389/fcomp.2022.869140*
  Authors: Fabiano Papaiz, Mario Emílio Teixeira Dourado Jr., Ricardo Alexsandro de Medeiros Valentim, Antonio Higor Freire de Morais, and Joel Perdiz Arrais.
- Ensemble-imbalance-based classification for amyotrophic lateral sclerosis prognostic prediction: identifying short-survival patients at diagnosis.
  BMC Medical Informatics and Decision Making. 2024. *DOI: 10.1186/s12911-024-02484-5*
  Authors: Fabiano Papaiz, Mario Emílio Teixeira Dourado Jr., Ricardo Alexsandro de Medeiros Valentim, Rafael Pinto, Antonio Higor Freire de Morais, and Joel Perdiz Arrais.
- Predicting ALS progression using autoregressive deep learning models.
  Intelligence-Based Medicine. 2025. *DOI: 10.1016/j.ibmed.2025.100247*
  Authors: Fabiano Papaiz, Mario Emílio Teixeira Dourado Jr., Ricardo Alexsandro de Medeiros Valentim, Felipe Ricardo dos Santos Fernandes, João Paulo Queiroz dos Santos, Antonio Higor Freire de Morais, Fernanda Brito Correia, and Joel Perdiz Arrais.

Conference papers:

- Clinical decision support system applied to amyotrophic lateral sclerosis prognosis.
  XXX Brazilian Conference of Neurology. 2022. *ISSN: 0004-282X*
  Authors: Fabiano Papaiz, Mario Emílio Teixeira Dourado Jr., Ricardo Alexsandro de Medeiros Valentim, Antonio Higor Freire de Morais, Anna Paula Paranhos Miranda Covaleski, Marcela Câmara Machado Costa, Isaac Holanda Mendes Maia, Francisco Marcos Bezerra da Cunha, Daniele Montenegro da Silva Barros, and Joel Perdiz Arrais.

Others:

- Health research project to collect data from Brazilian ALS patients.
- Clinical Decision Support system.

Chapter 2

Background

This chapter introduces the main concepts related to ALS disease, ML algorithms and techniques, and the significance of clinical decision-support systems.

2.1 Amyotrophic Lateral Sclerosis Disease

ALS is a rare, progressive neurodegenerative disease that affects the neurons of the human motor system. In ALS, both upper and lower motor neurons could be affected, interrupting the communication between the brain and muscles (Fig. 2.1), leading patients to paralysis and inevitably to death. It typically affects individuals between 40 and 70 years old, slightly more prevalent in men than women. Most patients succumb to the disease within 2-5 years after symptom onset. Diagnosing ALS is complex, as there is no single diagnostic test. Physicians typically rely on a comprehensive clinical analysis, symptom progression, and the exclusion of other diseases. On average, it takes up to 18 months from the onset of symptoms to reach a definitive diagnosis. The worldwide incidence of ALS is approximately 1.9 cases per 100,000 individuals per year [2–5, 8].

ALS exhibits significant clinical heterogeneity, manifesting diverse phenotypes among patients, such as motor neuron involvement, site of onset, symptoms, disease progression, and survival time (Fig. 2.2). Approximately 10% of patients have Familial ALS, which is a hereditary form of the disease. The majority of these cases are explained by mutations in the genes SOD1, C9orf72, TARDBP, and FUS. The remaining 90% of cases, known as Sporadic ALS, have an unknown cause. However, studies suggest that, in addition to a genetic component, there may be a possible relationship with environmental and lifestyle factors [5, 9, 10].

Riluzole is the predominant drug used in the treatment of ALS patients. Research has demonstrated that this medication has the potential to extend survival by approximately three months following 18 months of treatment [11, 12].

In light of the inherent complexity associated with ALS disease, it is imperative to allocate research efforts towards enhancing the diagnostic and prognostic processes. These efforts encompass endeavors such as identifying biological mark-

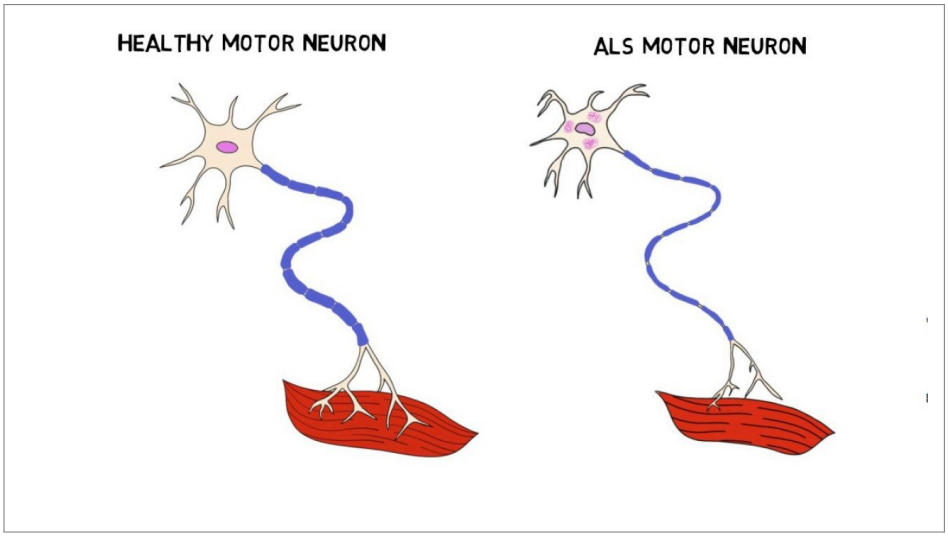


Figure 2.1: Comparison between a healthy motor neuron and a motor neuron affected by ALS disease. Source: internet.

ers specific to the disease, devising computational solutions that can be integrated into clinical settings to aid healthcare professionals, establishing a comprehensive database containing patient information to bolster future research endeavors, and other related initiatives aimed at advancing our understanding and management of ALS.

2.1.1 Biological Markers in ALS

The term biological marker (biomarker) was added to the Medical Subject Heading (MeSH) thesaurus in 1989. It refers to a biological parameter that represents a specific healthy or diseased state of a patient, which can be measured and quantified [13]. These biomarkers consist of clinical data and other indicators obtained from various sources in the human body, such as imaging and biological fluids (e.g., blood, urine, or saliva). When developing machine learning models (see section 2.2), the biomarkers will be used as input variables for training and will be referred to as *features*. The general application of biomarkers includes:

- Diagnosis: confirming the presence of a particular disease or condition;
- Prognosis/Monitoring: monitoring disease progression and assessing responses to various treatments;
- Risk: estimating the probability of disease development based on individual or population characteristics.

For diagnosing ALS disease, physicians generally use the criteria defined by the Revised *El Escorial* algorithm created by the World Federation of Neurology (WFN) [14, 15]. According to WFN criteria, the diagnosis of ALS requires the presence of (i) lower motor neuron degeneration, (ii) upper motor neuron degeneration,

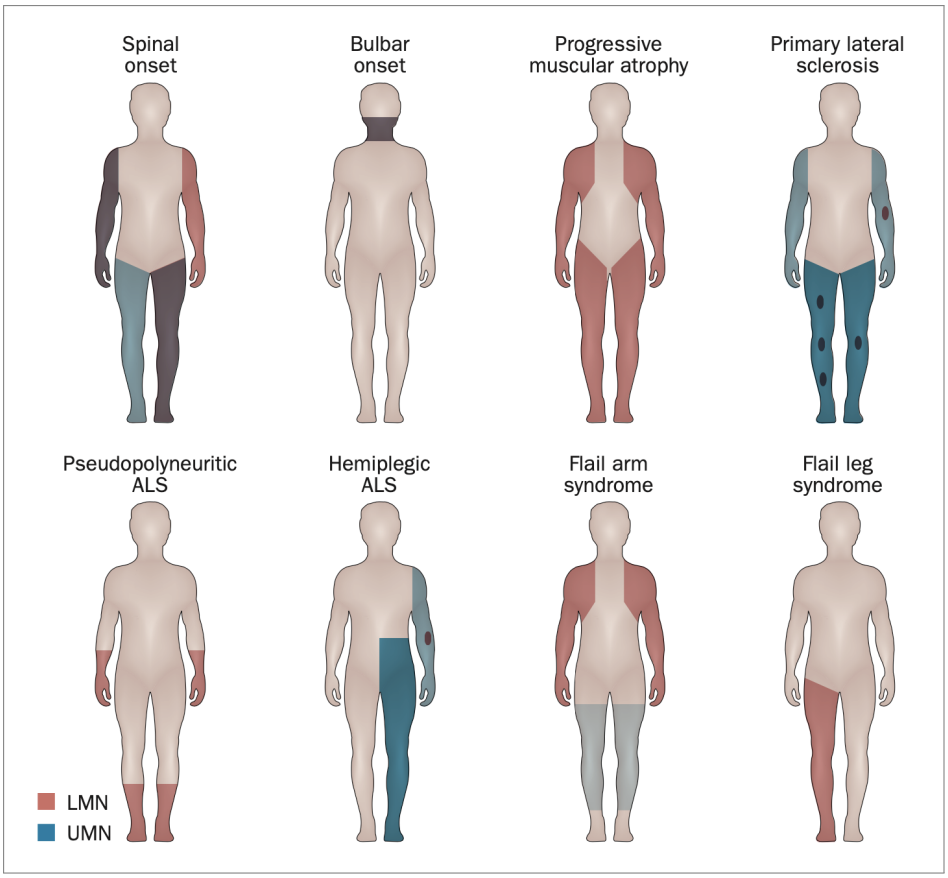


Figure 2.2: Upper (UMN) and lower (LMN) motor neurons involvement in different ALS phenotypes. Red color indicates LMN involvement, blue indicates UMN involvement. Darker shading indicates more severe involvement. Source: [4].

(iii) progressive spread of symptoms, (iv) the absence of imaging, electrophysiological or pathological evidence of other disease processes that might explain the motor neuron degeneration. As a result of this algorithm, physicians get the patient's classification biomarker, which could be one of the following: *Definite ALS*, *Probable ALS*, *Probable ALS Laboratory-supported*, and *Possible ALS*. The *El Escorial* can be used as a criterion to include or exclude patients from a research cohort (e.g., a study could exclude all patients classified as *Possible ALS*).

2.1.2 Prognosis

The common biomarkers used to predict ALS prognosis include the site of symptom onset, age at onset, patient gender, delay in diagnosis, presence of cognitive impairment, administration of the drug Riluzole, results from imaging tests, findings from respiratory exams, and patient examination using the ALS Functional Rate Scale. Studies have shown that certain biomarkers are associated with a worse prognosis [5, 16, 17]:

- Age $\geq$ 60 years old at symptom onset;
- Male gender;

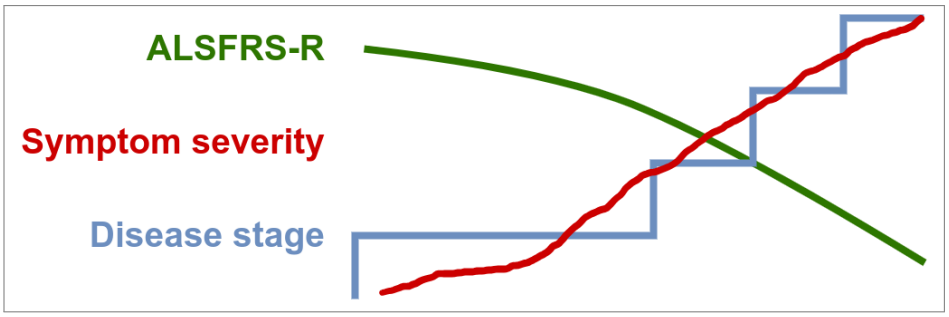


Figure 2.3: Correlation between ALS progression and the ALSFRS-R rates. Adapted from [26].

- Shorter diagnostic delay (the time from symptom onset to diagnosis);
- Presence of cognitive impairment, such as frontotemporal dementia;
- Site of onset in the bulbar region

The ALS Functional Rating Scale (ALSFRS) is an important tool for evaluating the functional disability of patients as the disease progresses. It is a questionnaire administered during clinic visits, consisting of ten questions about physical abilities such as speech, swallowing, writing, walking, and respiratory functions [18]. The ALSFRS was updated to provide more precise information about respiratory functions, leading to the development of the Revised ALS Functional Rating Scale (ALSFRS-R), which replaces the respiratory question in the original version with three new questions [19]. Each question is scored from 0 to 4, with 0 indicating a high level of disability and 4 indicating no disability. Depending on the version used, the sum of all questions (total score) can range from 0 to 40 (ALSFRS) or 0 to 48 (ALSFRS-R) points. Tracking changes in the total score over time is a common method for monitoring disease progression. An alternative approach for evaluating this predictor is to compute its slope value, which indicates the rate of functional decline within a specific time period, expressed in points per month (a higher slope corresponds to a faster disease progression). This scale has been utilized for patient diagnosis, monitoring symptom progression, and as a criterion for selecting participants in research studies [20–22]. The correlation between an accelerated ALSFRS-R score decline and a worse prognosis was demonstrated in Kollewe et al. [23]. Figure 2.3 provides a general overview of the correlation between ALSFRS-R scores, symptom severity, and disease stages. An example of the ALSFRS-R questionnaire is displayed in Figure 2.4. The ALSFRS-R has some known limitations, such as being influenced by medications, not including an assessment for cognitive impairment, and requiring several months to highlight progress for patients with slow disease progression. Due to its subjective nature, it may also be influenced by the mood and optimism of the evaluator [23–25].

Some staging methods have been proposed to define milestones for identifying phases during the progression of the disease. Two main approaches are the King's College and the Milano-Torino (MiToS) staging systems. The King's system is more sensitive to the earlier phases of ALS, while the MiToS system is more sensitive to the later phases [5, 27–29]. The King's system consists of a scale divided

ALSFRS-R Questionnaire

1. Speech
   - ❑ 4 - Normal speech processes
   - ❑ 3 - Detectable speech disturbance
   - ❑ 2 - Intelligible with repeating
   - ❑ 1 - Speech combined with nonvocal communication
   - ❑ 0 - Loss of useful speech

2. Salivation
   - ❑ 4 - Normal
   - ❑ 3 - Slight but definite excess of saliva in mouth; may have nighttime drooling
   - ❑ 2 - Moderately excessive saliva; may have minimal drooling
   - ❑ 1 - Marked excess of saliva with some drooling
   - ❑ 0 - Marked drooling; requires constant tissue or handkerchief

Figure 2.4: First two questions in the ALSFRS-R Questionnaire. Adapted from [19].

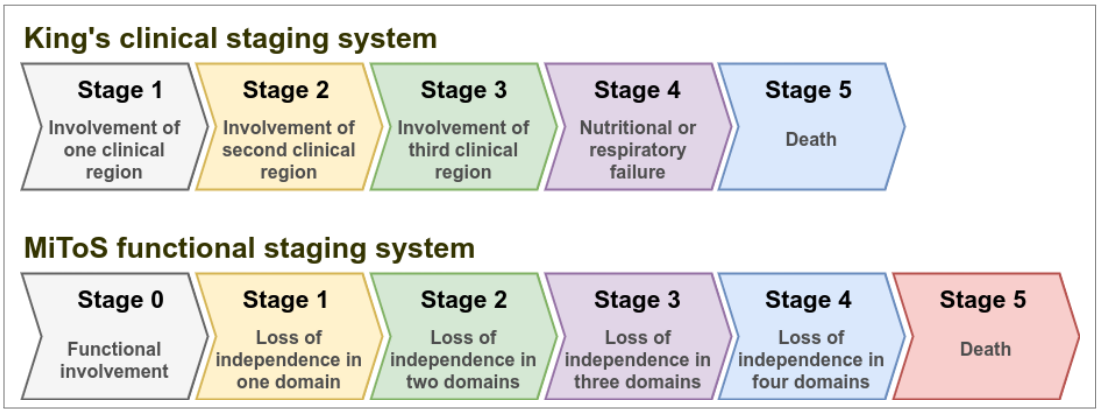


Figure 2.5: King's and MiToS stages and definiftions. Adapted from [29].

into five stages (1 to 5), with stage 1 representing symptom onset and stage 5 representing the patient's death. It is based on disease burden measured by clinical involvement, notable feeding, or respiratory malfunction. This system can be estimated using the ALSFRS-R score with a conformity of 92%. The MiToS system is divided into six stages (0 to 5), with stage 0 representing normal function and stage 5 representing the patient's death. It is based on functional ability assessed by the ALSFRS-R. Figure 2.5 illustrates all stages and definitions of these two systems. Staging systems have been used to gain a better understanding of disease progression, severity, prognosis (improvement or deterioration), and the timing of medical interventions such as respiratory support, gastrostomy, and medication, in order to enhance the quality of life for patients.

In ALS, the occurrence of respiratory failure serves as a critical indicator requiring evaluation by physicians. This manifestation is closely tied to shortened survival and may necessitate the implementation of respiratory support, such as noninvasive ventilation. Usually, the assessment of this condition involves the use of three biomarkers: Forced Vital Capacity (FVC), Slow Vital Capacity (SVC), and Sniff Nasal Inspiratory Force (SNIF) [20, 30, 31].

Imaging is also configuring an important field of study in ALS disease, with particular attention to the use of Magnetic Resonance Imaging (MRI) data as a valuable biomarker. In van der Burgh et al. [32], MRI data was combined with clinical information to improve predictions regarding the short, medium, and prolonged survival of patients. Schuster et al. [33] also demonstrated an improvement in the accuracy of survival prediction using MRI measures. Another MRI study revealed that volume decrease in the cervical spinal cord between diagnosis time and three months later could be used as a reference to predict the change in the SVC at 12 months [34].

2.1.3 More Complex and Costly Biomarkers

Researchers are currently studying additional biomarkers that are not commonly used in standard clinical practice. These biomarkers are more invasive and expensive compared to the ones mentioned previously. Some of them have been described below.

- **Metabolites:** In the study by Blasco et al. [35], an analysis of metabolites from ALS patients and a control group revealed significant differences in three out of seventeen analyzed metabolites. The ALS patient group showed higher concentrations of Ascorbate and Pyruvate and lower concentrations of Acetate. Moreover, there was a noticeable increase in the concentrations of Acetone. The authors suggested that these metabolites could be valuable for aiding in the early diagnosis and prognosis of ALS.
- **Cerebrospinal fluid:** The study related by Varghese et al. [36], suggests a correlation between neurodegeneration and elevated levels of four proteins in the cerebrospinal fluid (CSF). Quantitative mass spectrometry conducted on two sets of samples (Sporadic-ALS patients and Controls) revealed upregulation of proteins in the ALS group. Neurofilaments present in CSF were studied by Rossi et al. [37]. The study collected two types of NF through lumbar puncture: Neurofilament Light Chain (NF-L) and Phosphorylated Neurofilament Heavy Chain (pNF-H). The patient cohort was categorized into ALS patients and Controls (other neurological diseases), excluding those with familial ALS phenotype. Within the ALS group, further subgroups were identified: Spinal-onset ALS (ALS-S), Bulbar-onset ALS (ALS-B), and Frontotemporal Dementia ALS (ALS-FTD). The Control group was divided into Non-inflammatory Neurological Disorders (CTL-1) and Inflammatory Autoimmune Neurological Disorders (CTL-2). The findings supported NF-L and pNF-H as highly promising diagnostic and prognostic biomarkers for ALS. They demonstrated elevated levels of both neurofilaments in the CSF of ALS patients compared to Controls. Furthermore, both NF-L and pNF-H effectively distinguished ALS patients from those in CTL-1, with pNF-H exhibiting better sensitivity and specificity than NF-L. However, when compared to the CTL-2 group, these NF showed reduced specificity. The study also revealed that higher levels of NF-L and pNF-H at the time of diagnosis were associated with a more rapidly progressing disease and shorter survival.

- **MicroRNA:** Waller et al. [38] investigated the potential of microRNAs (miRNAs) in supporting the diagnosis and prognosis of ALS. The study involved the analysis of patients categorized into three groups: Sporadic ALS, ALS Mimics, and Controls. Blood samples were obtained from the patients at the time of diagnosis and prior to the initiation of Riluzole treatment. Analysis of 750 miRNAs extracted from each patient's serum revealed differential expression patterns between the groups. Notably, the expressions of miR-206 and miR-143-3p were found to be significantly higher in the Sporadic ALS group compared to the Controls group, while miR-374b-5p expression showed a significant decrease in the same comparison. Furthermore, the study demonstrated that over time, miR-143-3p exhibited a significant increase, whereas miR-374b-5p showed a significant decrease in patients with sporadic ALS.

2.2 Machine Learning Algorithms and Techniques

Machine Learning (ML) is a sub-area of Artificial Intelligence focused on creating computer programs that can learn from previous experience (training data) without being explicitly programmed for it. ML algorithms extract information from the training data, transform it into knowledge, and use this to solve different categories of problems [39, 40]. In 1997, Mitchell [41] provided a classical definition of ML: "*A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E*". He also defined the essential characteristics that a well-defined ML problem must have: (i) the class of tasks, (ii) the measure of performance to be evaluated and improved, and (iii) the source of experience for training. According to Mitchell, figure 2.6 illustrates an example of a well-defined ML problem and its features.

A robot driving learning problem:

- **Task *T***: driving on public four-lane highways using vision sensors
- **Performance measure *P***: average distance traveled before an error (as judged by human overseer)
- **Training experience *E***: a sequence of images and steering commands recorded while observing a human driver

Figure 2.6: A robot driving learning problem. Source [41].

ML consists of a collection of versatile algorithms capable of solving various problems, including classification, regression, and clustering. The key concept is that each algorithm can adapt its logic based on the training data, treating it as previous experience. For example, an algorithm designed to identify human faces in photographs can be easily repurposed to recognize cars instead of faces by making small changes to its source code and using different training data. Therefore, an ML algorithm must be capable of translating training data

into acquired knowledge, which it can then leverage to address problems [40]. In theory, the greater the volume of available training data, the more effective the algorithm's learning process. ML algorithms are typically categorized into three main groups, as explained below.

Supervised Learning: The algorithms are trained using a labeled dataset, where both the input and output (target) variables are known. For instance, we can have a dataset related to cancer that includes both patients and healthy individuals, with each sample labeled as "cancer" or "healthy." The algorithm learns from this dataset and can classify new samples into these classes (cancer/healthy) based on input variables. In another example, we can use a dataset about house characteristics and prices to train an algorithm that can estimate the price of a given house based on its characteristics (building area, number of rooms, number of bathrooms, neighborhood). Supervised Learning algorithms can be categorized into two types of problems: classification and regression. In a classification problem, the goal is to map the output value into discrete categories or classes, as explained in the first example above. In a regression problem, the aim is to make predictions using a continuous number as the output, as explained in the second example.

Unsupervised Learning: The algorithms attempt to learn from an unlabeled dataset, which does not include information about the classes associated with the input variables. The algorithm aims to uncover the hidden properties of the available data. It will attempt to group similar samples into classes (Clustering), discover existing patterns and relationships in the data (Association), or reduce the number of input variables to obtain a less complex algorithm (Dimensionality Reduction). For example, an unsupervised algorithm can derive insights from a dataset containing information about customers' orders and subsequently assist in identifying analogous customer behaviors, enabling the implementation of personalized marketing strategies.

Reinforcement Learning: The algorithms do not use any data for training and must figure out the best solution on their own. The system using the algorithm will provide "experience" through rewards and punishments. The algorithm's objective is to maximize the rewards and minimize the punishments. Reinforcement learning algorithms are commonly used in games, robotics, and navigation systems.

A general overview of the Supervised and Unsupervised categories and the types of problems they address is provided in Figure 2.7. As cited previously, the right choice of an ML algorithm is crucial to acquire a good performance. Table 2.1 summarizes some of the most known and used ML algorithms, organized by categories and the types of problems they address. ML algorithms have a wide range of applications and are successfully used in various industries such as manufacturing, healthcare, retail, travel, and finance.

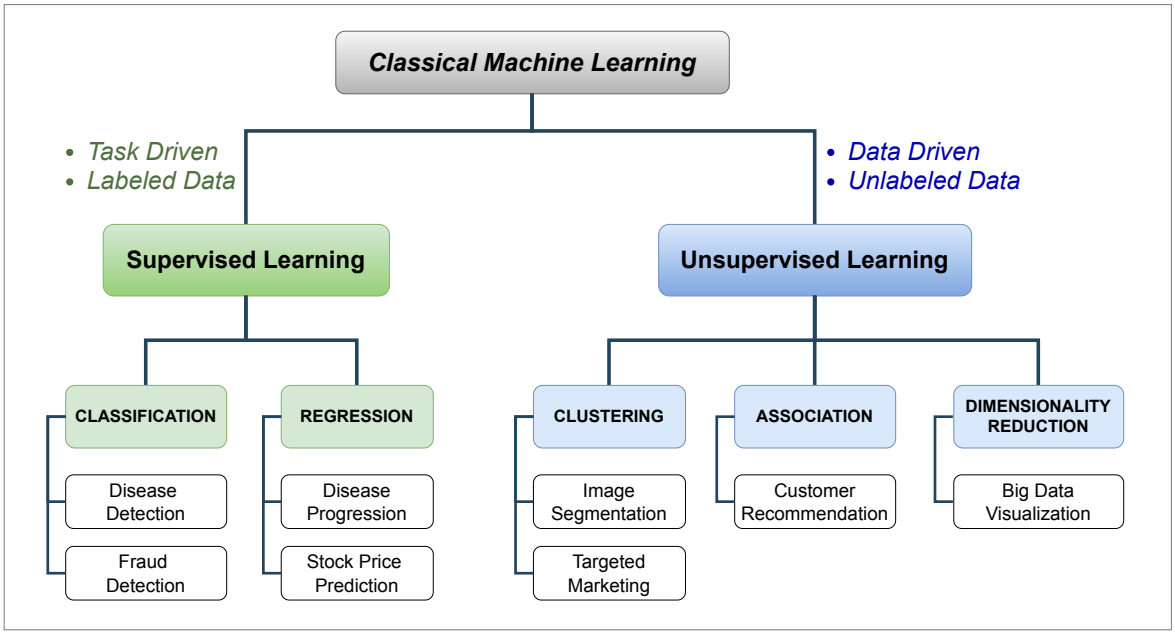


Figure 2.7: Supervised and Unsupervised Learning categories and their types of addressed problems.

2.2.1 ML Process

The ML process commonly has five main steps, as illustrated in Figure 2.8. First, we need to gather the data that will be used to train and test our system (model). This step is crucial because the quantity and quality of the raw data could directly affect the model's performance, determining its effectiveness.

The following step concerns preparing the data collected before making it available for use by the models. This is necessary because the raw data could be unorganized and contain missing and noisy information. At this moment, some data analysis and visualizations are also essential to verify the existence of outliers, data imbalances, the need for data normalization, and other problems that could affect the model's performance.

In the third step, the data is split into two smaller datasets: Train and Test. The Train dataset will be used to build and train the chosen ML model, and it generally contains the majority of the data (typically 70% or 80%). Choosing one suitable ML model to solve the addressed problem is also very important. This step could need to be repeated many times, and each iteration is called one training step.

The fourth step involves using some metrics to evaluate the performance of the ML model created. The ML model will be evaluated using the unseen data present in the Test dataset, simulating how it might perform with real-world data. Finally, if the ML model created does not perform satisfactorily, the last step is to refine it to improve its accuracy. After performing these steps, we will get an ML model ready to be used in real scenarios.

Table 2.1: ML algorithms organized by category and problem types.

SUPERVISED LEARNING		
Algorithm	**Problem Type**	
	Classification	**Regression**
Naïve Bayes	✓	
K-Nearest Neighbors (k-NN)	✓	
Support Vector Machines (SVM)	✓	
Logistic Regression	✓	
Linear Regression		✓
Decision Trees	✓	✓
Random Forests	✓	✓
Gradient Boosting Machines	✓	✓
XGBoost	✓	✓
Neural Networks *(Feedforward, Convolutional, Recurrent)*	✓	✓
UNSUPERVISED LEARNING		
Algorithm	**Clustering**	**Dimensionality Reduction**
K-Means	✓	
Hierarchical Clustering	✓	
Gaussian Mixture Model	✓	
DBSCAN	✓	
Principal Component Analysis (PCA)		✓
Uniform Manifold Approximation and Projection (UMAP)		✓
t-distributed Stochastic Neighbor Embedding (t-SNE)		✓
Linear Discriminant Analysis (LDA)		✓

2.2.2 Evaluation Metrics

Evaluation metrics are quantitative measures used to assess the performance of ML models. They evaluate how well a model performs a specific task, often by comparing its outputs against expected or ground-truth outcomes. These metrics offer distinct perspectives on a model's effectiveness and play a crucial role in guiding model selection and optimization. The choice of evaluation metric depends on the type of problem addressed (e.g., classification, regression) and the nature of the data.

Classification metrics often depend on using a confusion matrix. This matrix is commonly square with dimensions $n \times n$, where n represents the number of classes in the classification scenario. In the case of a binary classification problem involving two classes (positive and negative), the confusion matrix assumes the form as illustrated in Figure 2.9. Thus, a confusion matrix is a table that helps visualize the performance of a classification algorithm by summarizing the actual outcomes versus the predicted outcomes for each class. The most commonly used classification metrics and their calculations are detailed hereafter.

- **Accuracy**: Proportion of correct predictions out of the total number of pre-

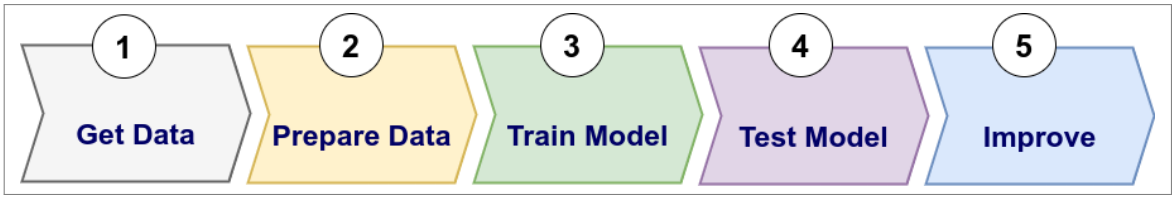


Figure 2.8: Main steps of a machine learning process.

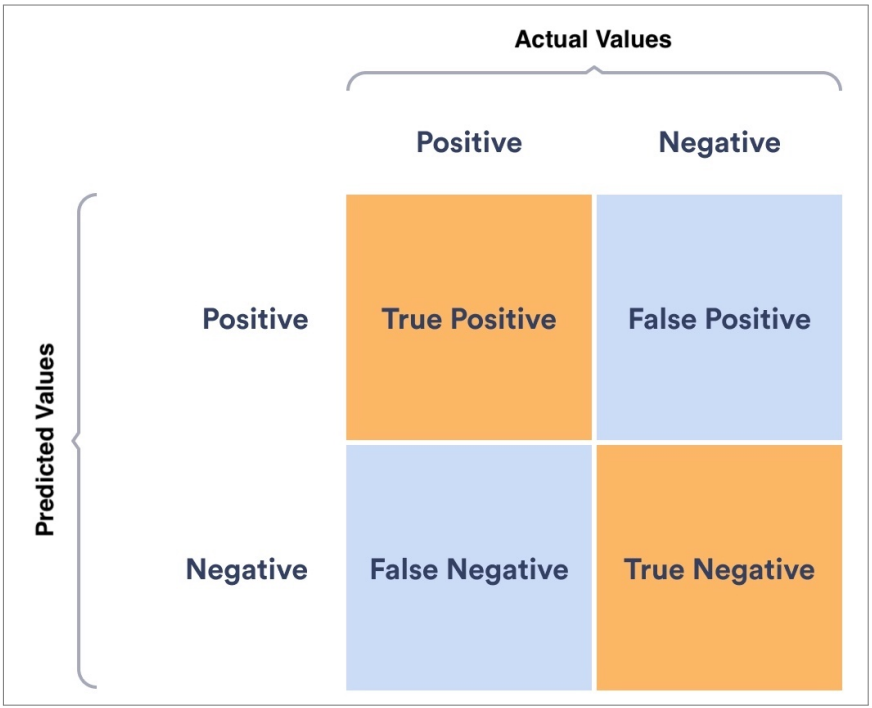


Figure 2.9: Confusion matrix for a binary classification problem.

dictions.

$$Accuracy = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Number of Samples}} \tag{2.1}$$

- **Precision**: Measures the correctness of positive predictions. It is the proportion of correctly predicted positives to the total predicted positives.

$$Precision = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{2.2}$$

- **Recall (Sensitivity)**: Measures the ability of a model to identify all relevant instances (positives). It is the proportion of true positives to all actual positives.

$$Recall = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{2.3}$$

- **Specificity**: Measures the ability of a model to identify all negative instances.

$$Specificity = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \tag{2.4}$$

- **F1 Score**: Harmonic mean of precision and recall. It balances both metrics, useful when the class distribution is uneven.

$$F1 = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.5}$$

- **Balanced Accuracy**: Represents the average of Sensitivity and Specificity, useful when dealing with imbalanced data.

$$Balanced\_Accuracy = \frac{Recall + Specificity}{2} \tag{2.6}$$

- **ROC-AUC**: Area under the Receiver Operating Characteristic curve, which plots the True Positive Rate against the False Positive Rate at various threshold settings (Fig. 2.10). It measures the ability of the model to distinguish between classes.

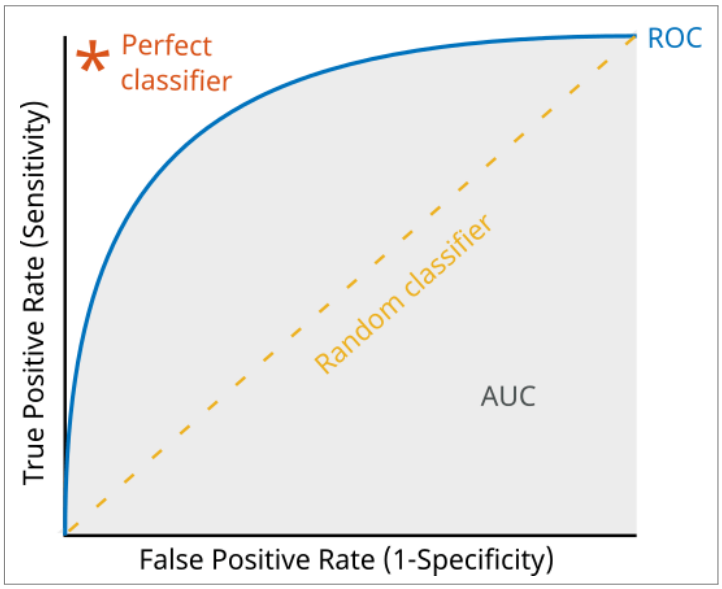


Figure 2.10: Area under the Receiver Operating Characteristic curve.

Regression metrics are utilized to evaluate tasks in which the ML model makes predictions of continuous values. Below are the most commonly used regression metrics and their calculations.

- **Mean Absolute Error (MAE)**: The average of absolute differences between the predicted and actual values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{2.7}$$

, where n is the number of samples, y is the actual value, and $\hat{y}$ is the predicted value.

- **Mean Squared Error (MSE)**: The average of squared differences between predicted and actual values. Penalizes larger errors more than MAE due to squaring the differences, useful when large errors are undesirable.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{2.8}$$

, where n is the number of samples, y is the actual value, and $\hat{y}$ is the predicted value.

- **Root Mean Squared Error (RMSE)**: The square root of MSE, providing a measure of error in the same units as the target variable.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{2.9}$$

, where n is the number of samples, y is the actual value, and $\hat{y}$ is the predicted value.

- **R-Squared (R^2)**: Indicates the proportion of variance in the dependent variable explained by the independent variables. An R^2 close to 1 implies a better fit of the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.10}$$

, where n is the number of samples, y is the actual value, $\hat{y}$ is the predicted values, and $\overline{y}$ is the mean of the actual values.

2.2.3 Techniques

When developing ML models, it is important to consider the issue of overfitting. Overfitting is a common problem in ML where a model performs well on the training data but struggles to perform well on new, unseen data. It occurs when the model becomes overly complex, capturing random noise and fluctuations in the training data as if they are significant patterns. Consequently, the model may display high accuracy on the training set but poor performance on the validation or test set. This section will outline some practical techniques to prevent overfitting.

Cross-Validation

Cross-validation is a resampling technique used to assess the performance of a ML model by dividing the data into multiple subsets (or folds). This approach helps prevent overfitting by testing the model's performance on different parts of the data. One of the most commonly used methods is the k-Fold Cross-Validation, where the dataset is divided into k equally sized subsets. The model is then

trained k times, with one fold left out for validation each time, and the remaining $k-1$ folds used for training. The final performance is determined as the average of the performances on all k validation folds [42]. Figure 2.11 illustrates an example of this method using 5 folds. Stratified Cross-Validation is a variation of k-Fold that ensures each fold is representative of the entire dataset in terms of class distributions, making it particularly useful for imbalanced datasets.

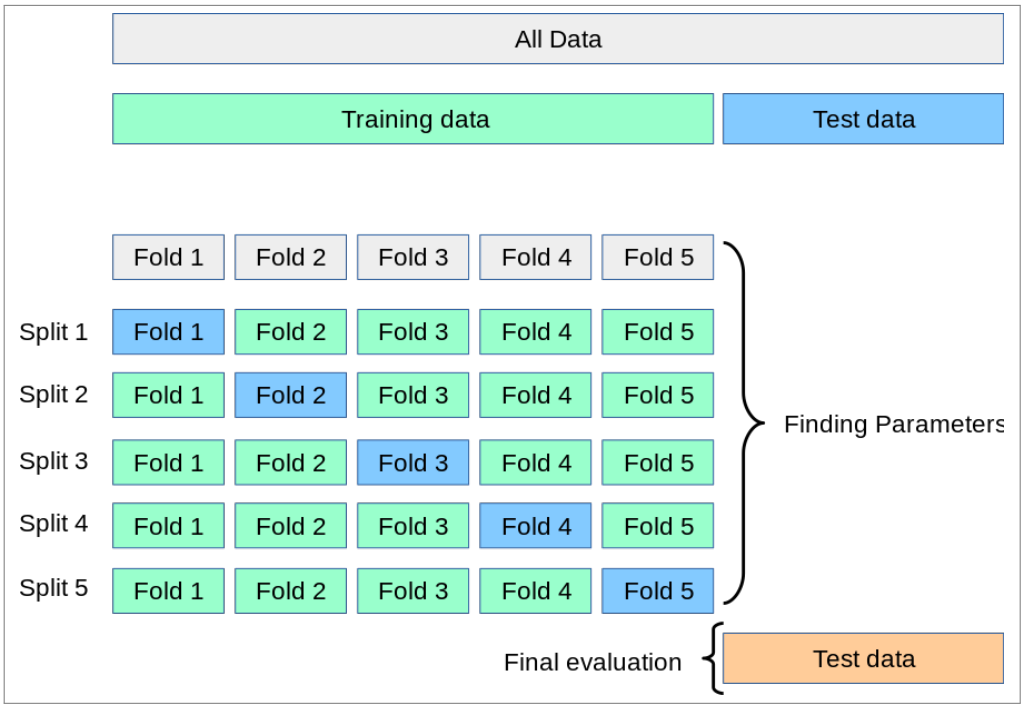


Figure 2.11: Example of a 5-Fold Cross Validation method. Source: [43].

Handling Imbalanced Data

Medical data analysis often involves dealing with imbalanced datasets. This issue arises when there is a significant imbalance in the number of samples between different classes, resulting in a substantially lower representation of one class compared to the others. Typically, the target prediction is linked to samples from the minority class, as in the case of detecting patients with lung cancer through the analysis of tomography images, where there are significantly more images of healthy patients (the majority class) than those depicting lung cancer cases (the minority class). In such scenarios, ML models frequently exhibit bias towards samples belonging to the majority class, resulting in an elevated misclassification rate within the minority class [44, 45]. To mitigate the issue of imbalance, resampling techniques such as Undersampling and Oversampling can be employed (see Fig. 2.12). Undersampling involves reducing the number of instances from the majority class to balance the dataset. The goal is to make the size of both the majority and minority classes roughly equal. Oversampling is the process of increasing the number of instances of the minority class to balance the dataset. Unlike undersampling, oversampling does not discard data; instead, it replicates or synthetically generates new instances for the minority class.

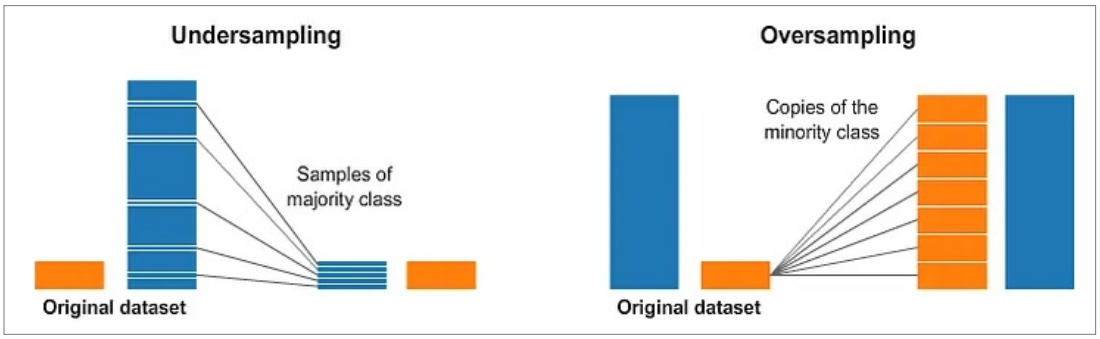


Figure 2.12: Undersampling and Oversampling methods overview. Source: internet.

Regularization

Regularization encompasses a collection of techniques utilized to mitigate overfitting by introducing a penalty term to the model's objective function. This penalty discourages the model from becoming excessively complex, thereby enhancing its ability to generalize to unseen data. Regularization commonly involves limiting or reducing the estimated parameters (weights) of the model to prevent it from fitting the noise or minor fluctuations in the training data [46]. By penalizing overly complex models (with large or numerous parameters), regularization forces the model to be simpler and more adaptable to new data. It also enhances generalization by encouraging the model to focus on the most relevant features and avoid noise in the data. Examples of this technique are listed below:

- **L1 Regularization (Lasso)**: introduces the absolute value of the coefficients as a penalty term into the objective function. This encourages sparsity, effectively driving some coefficients to become exactly zero, thereby performing feature selection. The objective function becomes:

$$\mathcal{L}(\mathbf{w}) = \mathcal{L}_{\text{original}} + \lambda \sum_{i=1}^{n} |w_i| \tag{2.11}$$

  , where $\mathcal{L}_{\text{original}}$ is the original loss function (e.g., Mean Squared Error), λ is the regularization parameter that controls the strength of the penalty, and w_i represents the model coefficients or weights.

- **L2 Regularization (Ridge)**: adds the square of the coefficients as a penalty term to the objective function. It effectively reduces the coefficients, preventing them from becoming excessively large. The objective function becomes:

$$\mathcal{L}(\mathbf{w}) = \mathcal{L}_{\text{original}} + \lambda \sum_{i=1}^{n} w_i^2 \tag{2.12}$$

  , where $\mathcal{L}_{\text{original}}$ is the original loss function, λ is the regularization parameter, and w_i represents the model coefficients or weights.

- **Elastic Net**: combines both L1 and L2 regularization, balancing between sparsity and shrinkage. The objective function becomes:

$$\mathcal{L}(\mathbf{w}) = \mathcal{L}_{\text{original}} + \lambda_1 \sum_{i=1}^{n} |w_i| + \lambda_2 \sum_{i=1}^{n} w_i^2 \tag{2.13}$$

, where $\mathcal{L}_{\text{original}}$ is the original loss function, λ_1 controls the L1 regularization strength, λ_2 controls the L2 regularization strength, and w_i represents the model coefficients or weights.

Dropout

Dropout is a technique for addressing overfitting in neural networks. During each training step, individual neurons are randomly omitted from the network with a probability of p, or retained with a probability of $1 - p$. This process effectively generates a new thinned network during each iteration, preventing the network from becoming excessively dependent on specific neurons and ultimately improving its ability to generalize [47]. Dropout is different from other regularization methods like L1 or L2. Instead of penalizing large weights, dropout reduces the model's capacity by randomly removing neurons. This forces the network to be more resilient to changes in the architecture. The dropout process is illustrated in Figure 2.13.

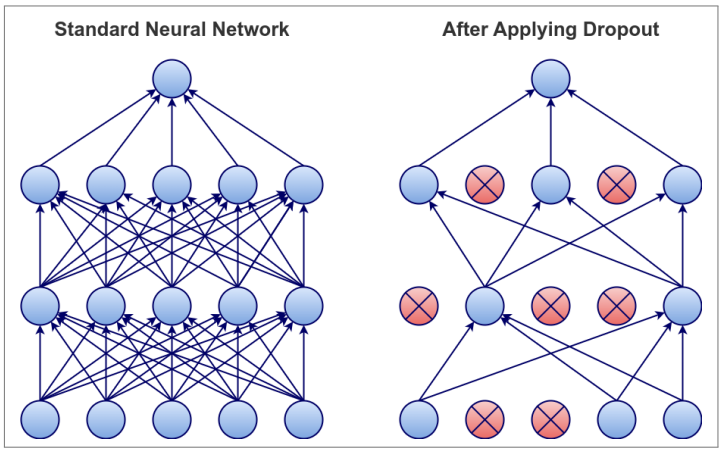


Figure 2.13: Dropout technique overview.

Early Stopping

Early stopping serves as a regularization technique to combat overfitting in ML models, especially in iterative learning algorithms such as neural networks. First, the model undergoes training across multiple iterations or epochs. Following each epoch, the model's performance, including metrics like loss or accuracy, is assessed using a separate validation dataset. If the validation performance ceases to improve and begins to deteriorate, early stopping halts the training process, indicating that the model has likely reached its optimal point for generalization.

This serves as an indicator that the model is starting to overfit the training data [48]. Figure 2.14 illustrates an overview of this technique.

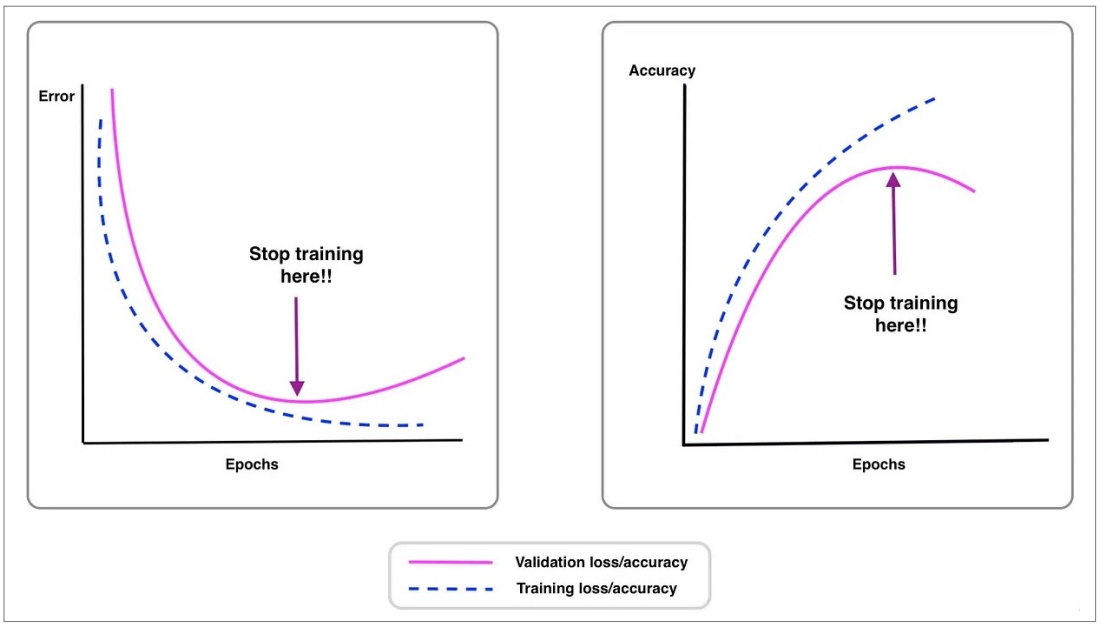


Figure 2.14: Early stop technique overview. Source: [49].

Grid Search

Grid search is a hyperparameter optimization technique used to determine the best combination of hyperparameters for a ML model. Many models have hyperparameters, such as the number of trees in a random forest or the regularization strength in logistic regression, and tuning these hyperparameters can significantly enhance model performance. To perform grid search, a grid of potential values for each hyperparameter is established, and the model is trained and evaluated for every possible combination of the hyperparameters in the grid [50]. However, it is important to note that a limitation of grid search is its potential computational expense, particularly for models with numerous hyperparameters or large datasets.

2.2.4 Time Series Forecasting

Time series prediction involves forecasting future values based on historical data that is ordered by time. This process is crucial in various fields, including finance, energy, healthcare, and retail, where understanding trends, seasonality, and anomalies can yield critical insights. In recent years, advancements in machine learning and deep learning models have significantly improved time series forecasting by effectively capturing non-linear dependencies and complex temporal relationships. Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs) are particularly well-known for their ability to model sequential data [51–53].

Time series forecasting models can be classified based on the following characteristics:

- Univariate vs. Multivariate: A univariate model forecasts a dependent variable (y) using a single independent variable (x). In contrast, a multivariate model predicts a dependent variable (y) based on multiple independent variables (x) (Fig. 2.15).

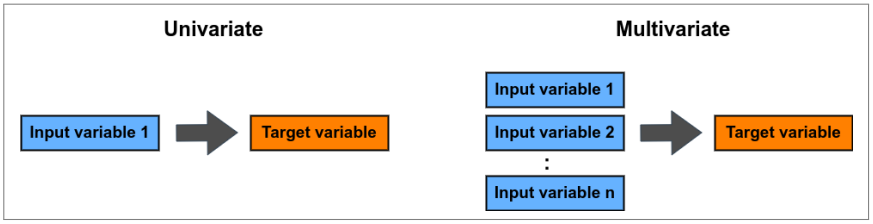


Figure 2.15: Univariate vs. Multivariate overview.

- Single-Output vs. Multi-Output: A single-output model forecasts the value of one feature for a specific future time step, whereas a multi-output model predicts values for several features simultaneously (Fig. 2.16).

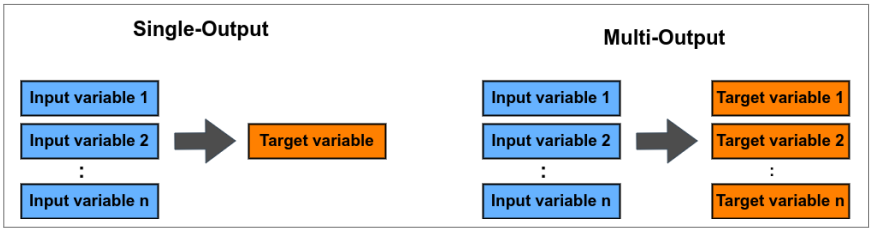


Figure 2.16: Single-Output vs. Multi-Output overview.

- Single-Step vs. Multi-Step: A single-step model generates a prediction for one specific time step ahead, whereas a multi-step model can project multiple future time steps (Fig. 2.17). Multi-step forecasting can be further divided into:
  - Single-Shot, which predicts the entire sequence of future time steps in one go, with a single model outputting all future values simultaneously.
  - Autoregressive, which predicts one step at a time, using previous predictions as inputs for subsequent steps.

Windowing Strategy

The windowing strategy is a method for organizing time series data into segments, which can be either overlapping or non-overlapping, to prepare it for machine learning or deep learning models. This method converts sequential data into input-output pairs, allowing the model to learn how to predict future values (outputs) based on a fixed set of past values (inputs) [54].

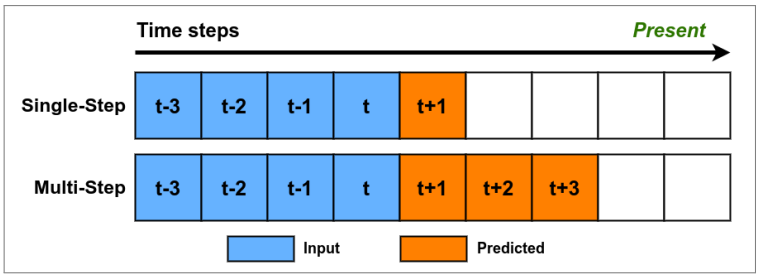


Figure 2.17: Single-Step vs. Multi-Step overview.

Two widely used windowing strategies in forecasting are the Fixing and Expanding windowing approaches. They differ in their selection and utilization of the input window, as illustrated in Figure 2.18. In the Fixing approach, the size of the input window remains constant throughout the forecasting process. Each input window has a fixed length and advances by one or more time steps to generate the next training instance. In contrast, the Expanding approach starts with an initial fixed-size window that gradually increases with each step. As it progresses forward, the model incorporates a growing amount of historical data into its analysis .

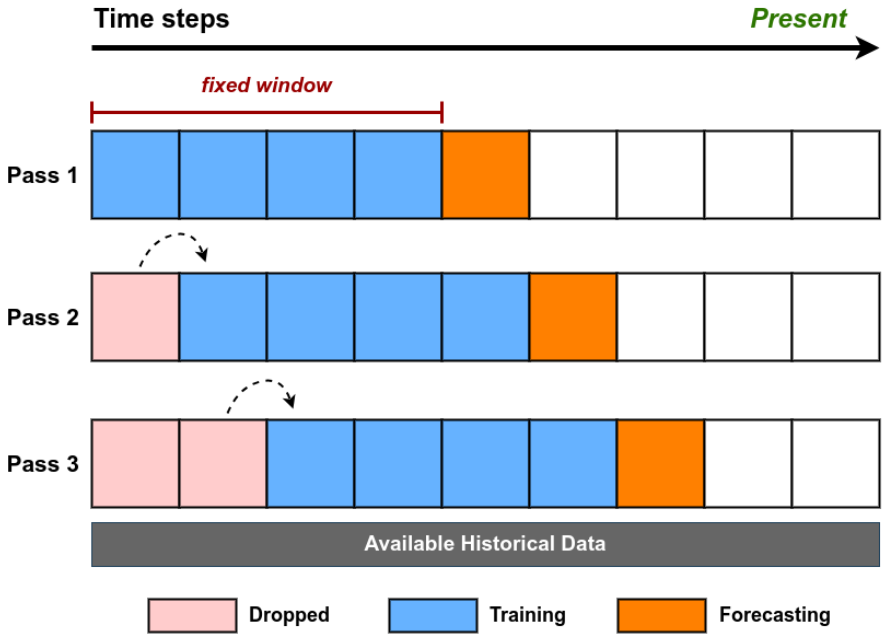
2.3 Clinical Decision Support Systems

Clinical Decision Support (CDS) systems are computer programs designed to assist healthcare workers in making more appropriate and timely decisions regarding patient care. These systems typically integrate a patient's current information with historical data to enhance decision-making and improve patient safety. As a result, CDS serves as a valuable tool to help healthcare workers make informed decisions based on accumulated information and experience over time. [55–57].

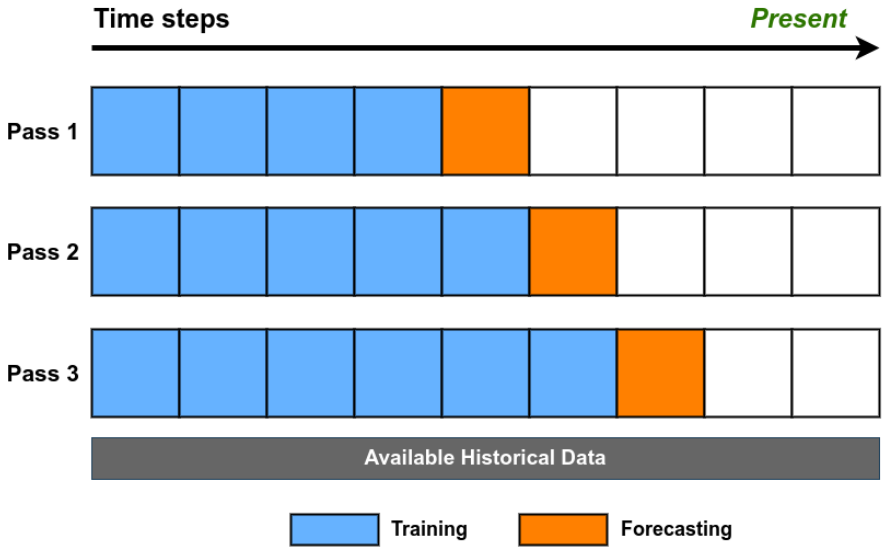
CDS can be categorized as Knowledge-based and Non-knowledge-based systems. Knowledge-based systems utilize established rules and best practices to provide recommendations to healthcare workers, while Non-knowledge-based systems employ ML algorithms to analyze past data and identify patterns to generate recommendations.

The advancement of ML algorithms has significantly enhanced the CDS application and its usefulness in the healthcare sector. However, some algorithms, like Neural Networks, can be challenging to interpret, explain, and evaluate in their decision-making process [58]. These are known as black-box solutions, which can restrict the acceptance and integration of CDS into clinical workflows. Consequently, the development of a CDS should prioritize interpretability and transparency to ensure that healthcare workers can comprehend the basis of any support provided.

A CDS must include the following essential components: (i) a Knowledge Base, (ii) an Inference Mechanism, and (iii) current patient information [59]. The Knowledge Base contains previous experience and typically includes data about relevant biomarkers from patients. The Inference Mechanism acts as the "brain of the



(a) Fixing



(b) Expanding

Figure 2.18: Comparison between the Fixing (a) and Expanding (b) windowing approaches.

system" and can be implemented using ML algorithms that utilize the Knowledge Base as training data. Finally, the end users input current information about a specific patient, and the Inference Mechanism generates predictions and recommendations related to that patient. Figure 2.19 illustrates the general architecture of a CDS.

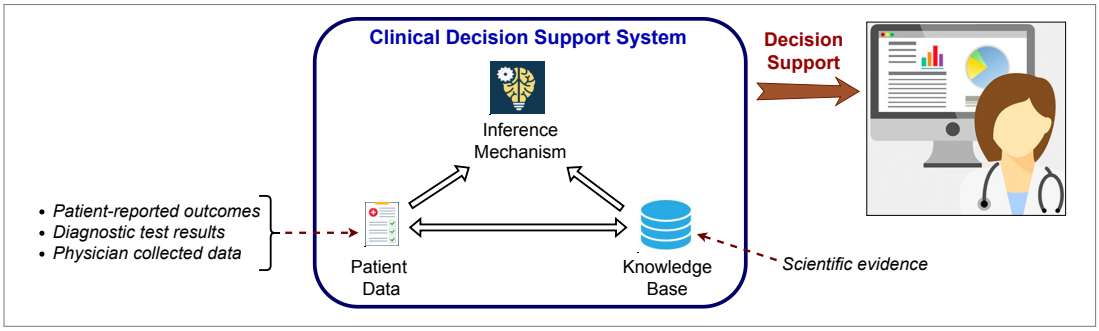


Figure 2.19: General architecture of a Clinical Decision Support system.

2.4 Brazilian Support for ALS Research

Since 2019, the Brazilian Ministry of Health initiated a research project titled "revELA: Scientific and Technological Development Applied to ALS" aimed at advancing studies and establishing guidelines for clinical practices related to ALS disease [7]. This comprehensive project encompasses support for research on prognostic solutions, remote respiratory support monitoring, the establishment of a national ALS database, and the development of distance learning courses for healthcare professionals (Fig. 2.20). The Laboratory of Technological Innovation in Health (LAIS) at the Federal University of Rio Grande do Norte is conducting this project, which has provided funding for this thesis. With initiatives like this project, it is expected that Brazilian researchers will have substantial support in the coming years to advance their studies on this disease and improve the quality of life for ALS patients.

Figure 2.20: Focus areas of the "revELA" project (language: Brazilian Portuguese). Adapted from [7].

Chapter 3

Machine Learning Applied to ALS Prognosis: State-of-the-Art

3.1 Introduction

This chapter provides a comprehensive overview of the state-of-the-art approaches used for predicting the ALS disease. It includes a thorough review of key studies that have utilized ML for ALS prognosis, describing the databases, relevant biomarkers, ML algorithms and techniques, and their findings. The focus is on studies that have analyzed commonly encountered biomarkers in clinical practice for ALS, such as demographic, clinical, laboratory, and imaging data. Hence, this chapter provides an overview of solutions that can be applied to develop decision support systems and be used by a higher number of ALS clinical settings. Furthermore, it discusses the open challenges, the limitations identified, and future research opportunities.

3.2 Methods

Six research questions (RQ) were elaborated to guide the conduct of this review, which are presented in Table 3.1. Next, the following stages were performed: (i) search articles related to ALS prognosis using ML in scientific databases, (ii) apply the inclusion criteria, (iii) apply the exclusion criteria, and (iv) analyze and summarize the selected articles.

In the first stage, the relevant literature was obtained from the *PubMed*, *Science Direct*, *IEEEXplore*, and *Web of Science* databases. The search was performed in April 2021 using the following search query: ("*artificial intelligence*" OR "*machine learning*" OR "*deep learning*") AND ("*amyotrophic lateral sclerosis*" OR "*motor neurone disease*") AND ("*predict*" OR "*prognosis*" OR "*progression*"). We used the Rayyan Web Application [60] to organize the resulting articles and also to perform the remaining stages.

In the second and third stages, we applied the Inclusion (IC) and Exclusion (EC)

Table 3.1: Research Questions.

RQ	Question
01	What are the ALS databases used in the study?
02	How many patients comprise the cohort of the study?
03	What are the types of prediction addressed by the study?
04	What are the ML algorithms and techniques used in the study?
05	What are the biomarkers evaluated and the most relevant identified by the study?
06	What are the performances of the used ML algorithms?

Criteria to filter the articles according to the scope of this paper (see Tables 3.2 and 3.3). We considered only articles published in Journals, written in English, and published between January 2011 and April 2021 (IC-01, IC-02, and IC-03). Articles that did not belong to the Information Technology, Computer Engineer, or Computer Science related areas were not included (IC-04). Next, we carried out the removal of the review articles (EC-01), the duplicate entries (EC-02), and articles not related to ML applied to ALS prognosis (EC-03). Then, the articles using *omics* data were removed (EC-04).

Finally, in the fourth stage, the select articles were thoroughly read, which allowed the final analysis and accomplishment of the objectives of this research.

Table 3.2: Inclusion criteria.

IC	Description
01	Articles published in Journals
02	Articles written in English
03	Articles published between January 2011 and April 2021
04	Articles in the Information Technology, Computer Engineer, or Computer Science related areas

Table 3.3: Exclusion criteria.

EC	Description
01	Review articles
02	Duplicate articles
03	Articles not related to Machine Learning applied to ALS prognosis
04	Articles using *omics* data (i.e., genomic, transcriptomic, proteomic, and metabolomic)

3.3 Results

Figure 3.1 illustrates the search and screening process for this systematic review. The search query and all inclusion criteria were used to perform the database searches. A total of 52 articles were retrieved, where 2 review articles were immediately excluded. After the removal of 15 duplicates, 35 articles were chosen for abstract review. A total of 25 studies were excluded due to the use of *omic* data (n=6) and not being related to ML applied to ALS prognosis (n=19). After completing the searching and screening process, 10 articles were selected to be analyzed and summarized. The following sections present the results that address the research questions defined in this study (Table 3.1).

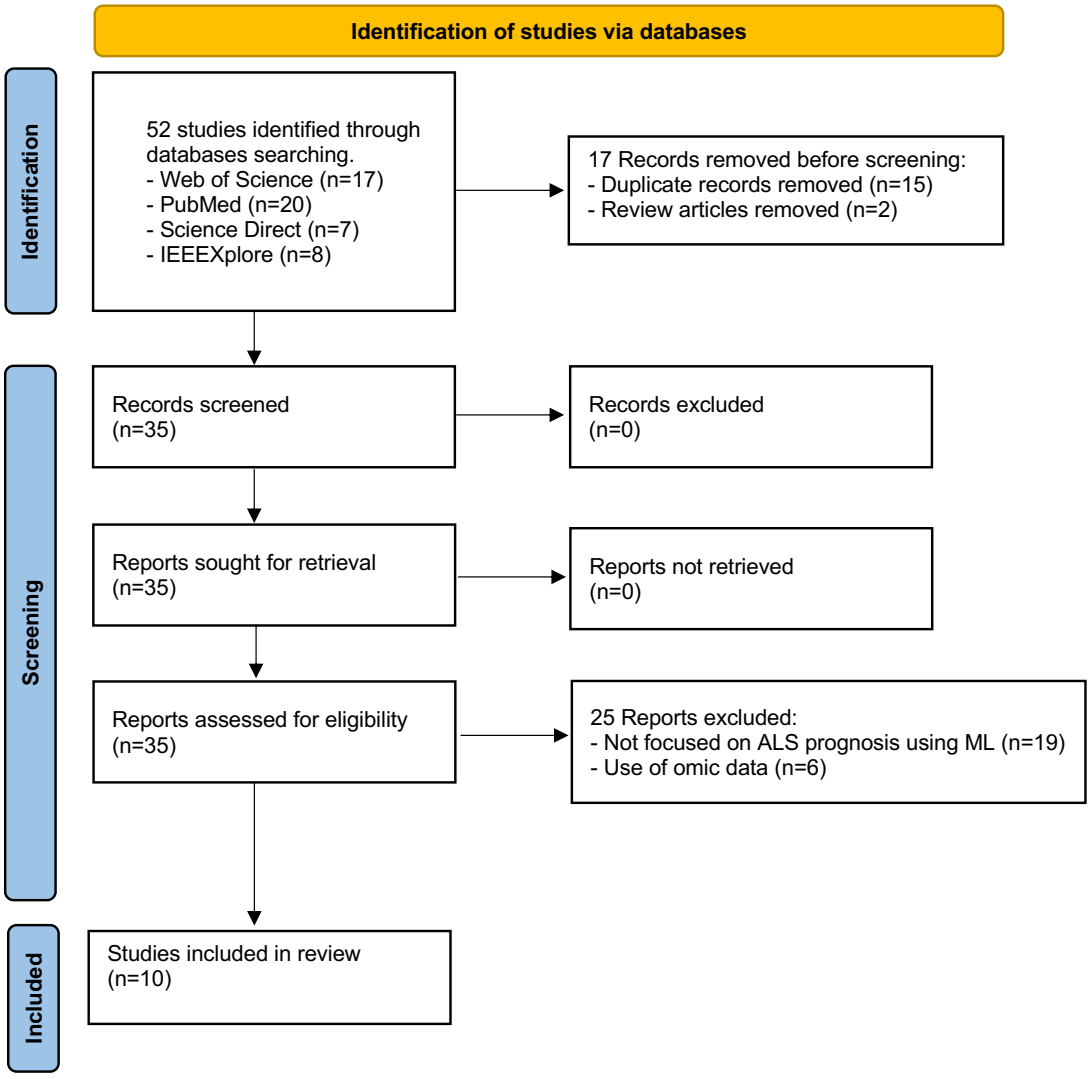


Figure 3.1: PRISMA flow chart of this study.

3.3.1 ALS Datasets and Sample Sizes

Different datasets were analyzed and their sample sizes ranged from 41 up to over 10,000 samples. Table 3.4 describes all the datasets analyzed. Most of the studies (60%) analyzed data from the PRO-ACT [61] dataset, probably because it was the only publicly available. The other datasets used were local or proprietary. The data formats analyzed included tabular (all studies) and image [62]. More detail about the sample size used by each study are described in Tables 3.6, 3.7, and 3.8.

Table 3.4: List of datasets used in the studies, inversely ordered by the sample size.

Dataset	Samples	References
PRO-ACT	+11,000	[6, 63–67]
Ireland-Italia	1,479	[64]
Tel Aviv - Sourasky Medical Center	1,328	[66]
Lisbon - Saint Mary's Hospital	1,214	[68, 69]
Paris - Tertiary Referral Centre for ALS	646	[67]
Trophos Company	431	[67]
Exonhit Pharma	172	[67]
Utrecht - University Medical Center	135	[62]
Italia	41	[70]

3.3.2 Types of Prediction Addressed

Based on the included studies, three distinct types of prediction were identified: *Disease Progression*, *Survival Time*, and *Need for Support* (more detail in Table 3.5). Kueffner et al. [64] addressed the *Disease Progression* and *Survival Time* types simultaneously.

The *Disease Progression* prediction aimed to estimate the patient's state at a given moment in the future and was the type most addressed by the studies included (70%). The *Survival Time* prediction aimed to estimate the occurrence of death from a baseline date to a point-time in the future, such as the probability of death after 12 months from symptoms onset. The *Need for Support* prediction aimed to estimate the moment when patients will need more specialized support.

3.3.3 Predictive Machine Learning Approaches

For *Disease Progression* prediction, most studies aimed to estimate changes in the ALS Functional Rating Scale (ALSFRS) or the Revised ALS Functional Rating Scale (ALSFRS-R) over time. Two other studies aimed to classify patients concerning their disease progression rates (Slow/Fast [64], Low/High [70]). Table 3.6 details the target predictions, best ML algorithm, performance, datasets, sam-

Table 3.5: Types of prediction addressed by the studies.

Type	Studies	References
Disease Progression	7	[6, 63–66, 69, 70]
Survival Time	3	[62, 64, 67]
Need for Support	1	[68]

ples size, techniques, validation strategies, and biomarkers evaluated for each study.

The studies that addressed the *Survival Time* prediction aimed to classify the patients into survival groups and estimate the probability of death after a specific time interval. van der Burgh et al. [62] aimed to classify patients into Short ($<$25 months), Medium (25$-$50 months), or Long ($>$50 months) survival groups. Kueffner et al. [64] aimed to estimate the probability of survival after 12, 18, and 24 months. Grollemund et al. [67] aimed to estimate the probability of patients being alive after 12 months. All three studies used the date of symptoms onset as the baseline date. The characteristics of each study are detailed in Table 3.7.

S. Pires et al. [68] was the unique study that addressed the *Need for Support* prediction, aiming to estimate the need for Non-Invasive Ventilation (NIV) support after 3, 6, and 12 months. The characteristics of this study are detailed in Table 3.8.

3.3.4 Biomarkers Evaluated and the Most Relevant Identified

As previously mentioned, we focused on the biomarkers commonly present in the ALS disease clinical practice, being obtained in a less costly and complex way. The biomarkers evaluated comprise clinical, demographic, vital signs, respiratory, functional, laboratory, imaging, neurophysiological, and medication data. For more detail, please see column *Biomarkers Evaluated* in Tables 3.6, 3.7, and 3.8. All the selected studies evaluated the ALS Functional Rating Scale (ALSFRS) or the Revised ALS Functional Rating Scale (ALSFRS-R) biomarkers. This fact highlights the importance of these biomarkers in monitoring ALS patients.

Table 3.9 depicts the most relevant biomarkers identified in the studies, with the information about their associated types of prediction. They comprised clinical, imaging, functional, respiratory, and laboratory data. The biomarkers identified as relevant in more than one study were the ALSFRS/ALSFRS-R (n=7), disease duration (n=5), Forced Vital Capacity (n=4), Body Mass Index (n=2), age at onset (n=2), and Creatinine (n=2).

Table 3.6: Overview of ML approaches on disease progression.

Study	Target Prediction	Best Algorithm	Performance	Biomarkers Evaluated	Dataset (Samples)	Techniques	Validation
[63]	Last patient state (ALSFRS score) recorded based on his past information	SPADE + DTW + Clustering Method	Accuracy: 73 F1 Score: 0.68 MAE: 0.30	*Tabular* : Clinical, demographic, laboratory, ALSFRS	PRO-ACT (2,590)		Hold-out
[6]	Last patient state (ALSFRS subscores) recorded based on his past information	CLM, and ODT	MAE (min-max): -CLM: 0.62−1.06 -ODT: 0.63−1.01	*Tabular* : Clinical, demographic, vital signs, laboratory, ALSFRS	PRO-ACT (3,772)	FS	10-Fold CV
[64]	-ALSFRS score at 12 months, using data from the first 3 months and only 6 biomarkers. -Patients classification into slow/fast progression groups.	GBM, and RF	GBM (PRO-ACT): -Z-score: ≈12 RF (Ireland-Italia) -Z-score: ≈6	*Tabular* : Clinical, demographic, vital signs, laboratory, FVC, SVC, ALSFRS	PRO-ACT (10,723) Ireland-Italia (1,479)		Hold-out
[65]	ALSFRS score and FVC at 12 months, using 1st visit and 3-month data	BART (ALSFRS), and RF (FVC)	BART: - R^2: 0.22 - RMSE: 0.55 - Corr: 0.47 RF: - R^2: 0.68 - RMSE: 14.27 - Corr: 0.83	*Tabular* : Clinical, demographic, pulse, BMI, FVC, laboratory, Riluzole medication, ALSFRS	PRO-ACT (2,424)	FS MI	5-Fold CV
[66]	ALSFRS score at several time intervals, varying from 6 up to 24 months	XGBoost	RMSE: 2.65−5.57 MAE: 1.98−4.42	*Tabular* : Clinical, demographic, vital signs, FVC, laboratory, ALSFRS	PRO-ACT (3,171) Tel Aviv (1,328)	FIA	Hold-out
[70]	Patients classification into low/high progression rates groups	SVM	Accuracy: 87.25	*Tabular* : Clinical, demographic, laboratory, ALSFRS-R	Italia (41)	FS	LOO CV
[69]	Changes in the ALSFRS-R score and subscores (before and after NIV)	Extension of DBN	Accuracy: 74−88 Sensitivity: 57−95 AUC: 75−98	*Tabular* : Clinical, demographic, El Escorial, BMI, C9orf72, FVC, MIP, MEP, PNRA, ALSFRS, ALSFRS-R	Lisbon (1,214)	MI	5-Fold CV

Abbreviations: ALSFRS: ALS Functional Rating Scale; ALSFRS-R: Revised ALS Functional Rating Scale; NIV: Non-Invasive Ventilation; BMI: Body Mass Index; FVC: Forced Vital Capacity; SVC: Slow Vital Capacity; MIP: Maximum Inspiratory Pressure; MEP: Maximum Expiratory Pressure; PNRA: Phrenic Nerve Response Amplitude; SPADE: Sequential Pattern Discovery using Equivalence Class; DTW: Dynamic Time Warping; CLM: Cumulative Link Models; ODT: Ordinal Decision Trees; GBM: Generalized Boosting Model; RF: Random Forest; BART: Bayesian Additive Regression Tree; SVM: Support Vector Machine; DBN: Dynamic Bayesian Network; AUC: Area Under the ROC Curve; MAE: Mean Absolute Error; MSPE: Mean Squared Prediction Error; R^2: Coefficient of Determination; RMSE: Root Mean Square Error; Corr: Pearson's Correlation Coefficient; FS: Feature Selection; FIA: Feature Importance Analysis; MI: Missing Data Imputation; CV: Cross-Validation; LOO: Leave-One-Out.

Table 3.7: Overview of ML approaches on survival time.

Study	Target Prediction	Best Algorithm	Performance	Biomarkers Evaluated	Dataset (Samples)	Techniques	Validation
[62]	Patients classification into Short (<25 months), Medium (25−50), and Long (>50) survival groups	Deep Neural Networks	Accuracy: 84	*Tabular* : Clinical, demographic, C9orf72, FTD, El Escorial, ALSFRS. *Image* : MRI.	Utretch (135)	FS MI	Hold-out
[64]	Probability of death within 12, 18, and 24 months	Gaussian Regression	PRO-ACT: -Z-score: ≈14.5 Ireland-Italia: -Z-score: ≈13	*Tabular* : Clinical, demographic, vital signs, laboratory, FVC, SVC, ALSFRS	PRO-ACT (10,723) Ireland-Italia (1,479)		Hold-out
[67]	1-year survival prediction, classifying patients into High, Intermediate, and Low survival rates groups	UMAP	BAcc: 91% F1 Score: 96%	*Tabular* : Clinical, demographic, ALSFRS	PRO-ACT (3971) Trophos (431) Exonhit (172) Paris (646)		Hold-out

Abbreviations: ALSFRS: ALS Functional Rating Scale; ALSFRS-R: Revised ALS Functional Rating Scale; MRI: Magnetic Resonance Image; FTD: Frontotemporal Dementia; FVC: Forced Vital Capacity; SVC: Slow Vital Capacity; UMAP: Uniform Manifold Approximation and Projection; BAcc: Balanced Accuracy; FS: Feature Selection; MI: Missing Data Imputation.

Table 3.8: Overview of ML approach on need for support.

Study	Target Prediction	Best Algorithm	Performance	Biomarkers Evaluated	Dataset (Samples)	Techniques	Validation
[68]	Patients need for NIV support at 3, 6, and 12 months for 3 progression groups (Slow, Neutral, and Fast)	RF	Slow: (3/6/12 months) - AUC: 81/87/91 - Sens: 70/72/78 - Spec: 76/83/86 Neutral: (3/6/12 months) - AUC: 76/82/86 - Sens: 58/62/79 - Spec: 78/83/77 Fast: (3/6/12 months) - AUC: 72/81/79 - Sens: 51/71/74 - Spec: 77/76/71	*Tabular* : Clinical, demographic, El Escorial, BMI, C9orf72, VC, FVC, P0.1, SNIP, MIP, MEP, NIV, PNRA, PNRL, CE, CF, ALSFRS, ALSFRS-R	Lisbon (1070)	FS DB	10-Fold CV

Abbreviations: ALSFRS: ALS Functional Rating Scale; ALSFRS-R: Revised ALS Functional Rating Scale; NIV: Non-Invasive Ventilation; BMI: Body Mass Index; FVC: Forced Vital Capacity; SVC: Slow Vital Capacity; VC: Vital Capacity; P0.1: Airway Occlusion Pressure; SNIP: Sniff Nasal Inspiratory Pressure; MIP: Maximum Inspiratory Pressure; MEP: Maximum Expiratory Pressure; PNRA: Phrenic Nerve Response Amplitude; PNRL: Phrenic Nerve Response Latency; CE: Cervical Extension; CF: Cervical Flexion; RF: Random Forest; AUC: Area Under the ROC Curve; Sens: Sensitivity; Spec: Specificity; FS: Feature Selection; DB: Data Balancing; CV: Cross-Validation.

3.3.5 Description of the Studies

van der Burgh et al. [62] demonstrated the positive impact of using Magnetic Resonance Images (MRI) along with clinical information to classify ALS patients into three survival groups: Short ($<$25 months), Medium (25$-$50 months), and Long ($>$50 months). The biomarkers evaluated were clinical information (e.g., site of onset, age at onset, ALSFRS slope, FVC) and MRI images (Structural Connectivity and Brain Morphology data) from 135 ALS patients. They developed Deep Neural Networks models and evaluate them in four scenarios using different biomarkers sets: (i) only Clinical Data, (ii) only Structural Connectivity MRI Data, (iii) only Brain Morphology MRI Data, and (iv) combining Clinical and MRI Data. The greater accuracy was obtained using the Clinical-MRI combined data (84%) compared to the other three strategies (Clinical: 69%; Structural Connectivity MRI: 63%; Brain Morphology MRI: 63%). They pointed out the power of Deep Neural Networks in making predictions using complex data. However, the relationships between input and output variables could not be easily recognized, needing more investigation to understand ALS progression better.

S. Pires et al. [68] developed a model to predict when a patient will need NIV support according to a given time window (3, 6, and 12 months). They used the Portuguese ALS Dataset (n=1,070), combining the static and temporal data into a data structure called snapshot, which contains all information about a patient at a specific date. The patients were divided into three disease progression groups (Slow, Neutral, and Fast) and, for each group, their respective snapshots were used as learning instances to evaluate several ML models. A Feature Selection Ensemble approach was used to select the relevant biomarkers for each group. The Random Forest model obtained the best performance for 3, 6, and 12 months time window values. The relevant biomarkers present in all groups were BMI, FVC, and VC. Other relevant biomarkers (present in 75% of the time) were age at onset, disease duration, and ALSFRS score. The authors reported the advantage of using specialized ML models for different patient groups (e.g., disease progression groups) rather than create generalized models treating all the patients similarly.

D. Halbersberg and B. Lerner [63] demonstrated the benefit of using temporal modeling, sequence clustering, and sequential pattern mining to predict the last patient state recorded (ALSFRS score) based on his past information. To find relevant deterioration patterns in temporal patients data they developed a framework consisting of three stages: (i) group patients with similar progression using hierarchical clustering based on Dynamic Time Warping, (ii) perform pattern mining to found out common functional deterioration patterns among patients based on the SPADE sequence mining algorithm, and (iii) develop a Random Forest model to classify patients into their most similar cluster to predict their next disease state. The performance obtained by the proposed framework (Accuracy: 73, F1 score: 0.68, Mean Absolute Error: 0.3) was superior related to two other benchmark models (Random Forest and Long Short-Term Memory, both using no temporal modeling). They used static (e.g., age at onset, time from onset, gender) and longitudinal (ALSFRS scores and subscores) data of 2,590 subjects from the PRO-ACT dataset. The most important predictors reported were the previ-

Table 3.9: Most relevant biomarkers identified, associated predictions, and references.

Type	Biomarker	Associated Predictions / References		
		Disease Progression	Survival Time	Need for Support
Clinical	Age at disease onset	[63]	[64]	-
	Body Mass Index (BMI)	[64, 69]	[64]	-
	Disease duration	[6, 63–65, 69]	[64]	-
	Site of onset	[6]	-	-
Imaging	Magnetic Resonance Imaging	-	[62]	-
Functional	ALSFRS	[63–66]	[64, 67]	[68]
	ALSFRS-R	[69]	-	[68]
Respiratory	Forced Vital Capacity (FVC)	[6, 64, 65]	[64]	[68]
	Maximal Expiratory Pressure (MEP)	[69]	-	-
	Maximal Inspiratory Pressure (MIP)	[69]	-	-
	Slow Vital Capacity (SVC)	[64]	[64]	-
	Vital Capacity (VC)	-	-	[68]
Laboratory	Absolute Monocyte Count	[65]	-	-
	Alanine Transaminase (ALT)	[65]	-	-
	Alkaline Phosphatase	[6]	-	-
	Calcium	[65]	-	-
	Chloride	[6]	-	-
	Cholesterol - Total	[70]	-	-
	Cholesterol - High-Density (HDL)	[70]	-	-
	Creatine Kinase (CK)	[6]	-	-
	Creatinine	[6, 64]	-	-
	Hematocrit	[65]	-	-
	Phosphorus	[6]	-	-
	Potassium	[65]	-	-
	Segmented Neutrophils	[64]	-	-
	Urine Ph	[64]	-	-
	Vitamin B12	[70]	-	-

ous ALSFRS score, the previous ALSFRS *Dressing* subscore, the previous *Climbing Stairs* subscore, the previous *Turning in Bed* subscore, the time from disease onset, and the deterioration pattern termed <*E,G,I*> (i.e., a sequential declining in the *Writing*, *Dressing*, and *Walking* ALSFRS subscores).

Gordon and Lerner [6] evaluated the capacity of ordinal classifiers to predict the functional decline of the patients. They used data about the first and last patient visits from the PRO-ACT dataset (n=3,772), analyzing the following biomarkers: clinical, demographic, ALSFRS, FVC, medication, vital signs, and laboratory tests. The target variables were all ten ALSFRS items (questions) separately. The patient states were mapped to the ALSFRS items, thus correlating patient state to disease progression for each point in time. Addressing the ordinal nature of the ALSFRS, they evaluated the following ordinal classifiers: Cumulative Link Models (CLM), Ordinal Decision Trees (ODT), and Cumulative Probability Tree (CPT). To evaluate their performances, they defined a penalizing system that accounts for various error severities differently. Thus, a classifier was less penalized when it predicted the value of 2 instead of 1 when the real value was 3. These three classifiers were compared with the Random Forest (RF), a non-ordinal classifier. The results showed that the CLM and ODT ordinal classifiers presented a similar performance and outperformed the RF classifier regarding the Mean Absolute Error measured in the best experiment scenario (CLM: $0.62-1.06$; ODT: $0.63-1.01$; RF: $1.01-1.61$). For feature selection, the authors implemented an algorithm based on the J_3 scattering matrix criterion for each ALSFRS item individually. The most relevant predictors were the FVC, the site of onset, the time from onset, and the laboratory tests Creatinine, CK, Chloride, Phosphorus, and Alkaline Phosphatase.

A crowdsourcing strategy was presented in Kueffner et al. [64], where were selected 30 teams around the world to participate in an ALS stratification challenge. They asked the participants to create ML models to perform prediction tasks using the PRO-ACT and the Irish-Italian Registries datasets. The teams used patient data from the first three months and were limited to evaluate only six of all biomarkers available. The target predictions were the Disease Progression at 12 months (decline of the Functional Rate Scale) and the Probability of Survival at 12, 18, and 24 months. Regarding the survival prediction, one team outperformed the others significantly using a Gaussian Process Regression model, presenting a better approach in leading with the right-censored patient outcome (dead or trial dropout). The best models related to the disease progression prediction used the Generalized Boosting Model and the Random Forest algorithms. The more relevant biomarkers were disease duration, age at onset, site of onset, gender, weight, BMI, respiratory exams (FVC and SVC), laboratory tests (Creatinine and Segmented Neutrophils), and ALSFRS scores and subscores. Based on the relevant biomarkers chosen by the teams, the authors have identified four distinct patient groups: Slow Progressing, Fast Progressing, Early Stage, and Late Stage. The main biomarkers related to each group were also detailed in this study, where the authors highlighted the importance of the ALSFRS Bulbar subscore (questions $1-3$) in discriminating between groups.

Tang et al. [65] addressed predictions in changing of the ALSFRS score and in the FVC percentage. They used static and longitudinal biomarkers from the PROC-

ACT dataset (n = 2,424), including only those patients with information about ALSFRS scores over time. The longitudinal data were transformed into signature vectors aggregating statistics values (minimum, median, maximum, and slope). Using data from the first visit and at the 3-month, the authors create models to predict the changes in the ALSFRS slope at 12-month. The evaluated models (Random Forest and Bayesian Additive Regression Tree) achieved modest results (Correlation: 0.47; RMSE: 0.55; R^2: 0.22), thus, indicating the difficulty in predicting 12-month ALSFRS slope using the only baseline and 3-months data. Feature Selection was performed using the Random Forest and the Knockoff Filter methods. After combining the top-ranked biomarkers returned by both methods, the best predictive biomarkers were the ALSFRS score, the disease duration, the FVC, and the Absolute Monocyte Count. To predict the FVC Percentage changes between 3 and 12 months, Random Forest models were tested in two scenarios (either including the baseline FVC or not). The best results were obtained using the FVC at baseline data, demonstrating the power of this biomarker, which increased the correlation from 0.67 to 0.83. The authors also applied unsupervised classification (K-Means) to find distinct phenotypes groups, founding four balanced clusters among the patients. However, it was considered impractical to clearly understand how the groups differ due to the high number of biomarkers defined for each group during the clustering process.

B. Hadad and B. Lerner [66] studied prediction of the ALSFRS score in several time intervals, varying from 6 to 24 months. Temporal (Long Short Term Memory - LSTM) and non-temporal (Random Forest, XGBoost, and Multilayer Perceptron) models were evaluated over the PRO-ACT dataset (n=3,171). To be used by the non-temporal models, the longitudinal data were transformed into vectors containing aggregated values (mean, standard deviation, slope, minimum, maximum). Each model was tested using 60 different randomly generated configurations, and their averaged performances were compared (Root Mean Square Error and Mean Absolute Error). The XGBoost model obtained superior performance for the most time intervals evaluated (RMSE: 2.65−5.57, MAE: 1.98−4.42), being more precise for shorter than longer intervals. The relevant predictive biomarkers were the ALSFRS subscores. In another experiment, these models were evaluated in two scenarios: (i) trained with the PRO-ACT and tested with the TASMC dataset (n=1,328), and (ii) trained and tested using only the TASMC data. The short-term predictions (up to 6 months) were more precise using models trained with the PRO-ACT, and the XGBoost obtained the best results again. The authors highlighted that the PROC-ACT contains data from clinical trials that may not reflect the reality presented by the clinical environment patients due to the inclusion/exclusion criteria used. Thus, their patients tend to be younger and to have a slower disease progression, in addition to having more visits registered than the usual clinical patients. To address this problem, they proposed a final experiment applying the Domain Adaptation approach to develop predictive models using the PRO-ACT data and improve their performances using patient clinical data. Firstly, LSTM and Multilayer Perceptron models were trained using only data from the PRO-ACT. Then, the training phase was complemented using the TASMC data to fine-tune the models to the clinical data. The results demonstrated that the use of domain adaptation improved the predictive performance for both models.

Grollemund et al. [67] presented a dimensionality reduction model to predict 1-year survival rates. The biomarkers analyzed were gender, site onset, age, weight, disease duration, ALSFRS scores, ALSFRS slopes, and if died or not after one year. They combined data from four datasets (PRO-ACT, Trophos, Exonhit, and Paris Tertiary Referral Centre), totaling 5,220 samples. The obtained dataset was further divided into development and validation sets. After, the high-dimensional data from the development set were reduced and projected onto 2D space through the Uniform Manifold Approximation and Projection (UMAP) algorithm. Thus, the authors were able to project information about the patients into a 2D graph. The 2D data were divided into three 1-year survival probability zones: High (90%), Intermediate (80%), and Low (58%). Then, the validation set was used to evaluate the proposed model, and the results were compared with the Random Forest and the Logistic Regression models. The UMAP model obtained better classification results (F1 score: 96%, Balanced Accuracy: 91%) when compared to the average results of the other models (F1 score: 50%, Balanced Accuracy: 60%). The adopted approach also helped identify the biomarkers with higher or lower correlation with the survival prediction. For example, the age and ALSFRS score presented a high correlation, while the gender and weight showed a low correlation. However, the total comprehension of the relationship between input and output variables cannot be obtained because the adopted model is considered a black-box approach, which degrades its interpretability.

Despite Greco et al. [70] aimed to find blood analytes to distinguish patients who have ALS from those with Lower Motor Neuron Disease (LMND), they also studied the classification of these patients with relation to their disease progression rates (High or Low). They analyzed clinic, demographic, and blood (108 analytes) data from 41 ALS patients. An SVM model was developed, and the Recursive-Feature-Elimination algorithm was used as a feature selection method. This model obtained an accuracy of 87.25% in classifying ALS patients into the High and Low groups using the first 16 ranked analytes, indicating the potential of using blood data as predictor biomarkers. Elevated levels of Vitamin-B12, Total Cholesterol, and HDL were related to a higher disease progression rate.

Leão et al. [69] proposed a predictive model based on Dynamic Bayesian Networks (DBN), including both static and longitudinal data. They accessed data from the Portuguese ALS dataset (n=1,214), and the target prediction was the disease progression (ALSFRS score and subscores) related to the need for NIV support. To be processed by the DBN model, the longitudinal data were converted into time-series data and then divided into Before NIV and After NIV subsets. Thus, they were able to determine the most relevant biomarkers related to these two essential disease stages. The authors developed a predictive model, termed stdDBN framework, which uses stationary DBNs to predict disease progression and non-stationary DBNs to determine how the biomarkers analyzed change over time in each subset. The average results for predicting disease progression were above 80% for both subsets regarding the Accuracy, Sensitivity, and AUC metrics, demonstrating the potential of the proposed methodology. Graphs were generated to visualize how the biomarkers change over time, displaying their values in different time steps for each stage (before and after NIV).

This approach allowed identifying some interesting relationships, as following mentioned. The Maximum Expiratory Pressure (MEP) was considered the most important respiratory exam to predict the patient ventilatory decline before the need for NIV support. The ALSFRS Bulbar subscore had more influence on disease progression after NIV than before NIV. The BMI and Disease Duration had a stronger influence than the other static biomarkers for both subsets.

3.4 Discussion

This study systematically reviewed the literature to identify relevant studies that used ML approaches to assist ALS disease prognosis. As explained before in Section 4.2, we focused on those studies comprising biomarkers commonly present in the daily ALS clinical practice. We identified 10 studies and detailed their the target predictions, best ML algorithm, performance, datasets, samples size, techniques, validation strategies, biomarkers evaluated, and the most relevant biomarkers identified.

3.4.1 ALS Datasets and Data Preprocessing

Notably, the studies accessed datasets that concentrate ALS patients from Europe and the United States of America. Data from other regions were not analyzed (e.g., South America, Africa, or Asia). We consider this analysis essential to confirm (or not) if the predictive ML solutions can be broadly generalized and if different datasets can be combined to compose an even more relevant ALS dataset. Most of the studies (60%) analyzed data from the PRO-ACT dataset. PRO-ACT is the largest public ALS dataset available, containing over 10,000 samples, serving as a basis for several studies on ALS disease, and suitable for developing ML solutions. However, some studies included advised that the PRO-ACT has limitations that can increase the risk of creating biased models [65–67]. Previous studies also reported these PRO-ACT limitations, and the risk of it does not represent the clinical patient population due to the inclusion and exclusion criteria used in the clinical trials [61, 71]. For instance, their patients tend to be younger and present fewer functional impairments. In this sense, using a validation strategy that includes an external dataset represents an alternative to decrease bias risk and achieve a more reliable ML algorithm evaluation. This strategy was utilized by B. Hadad and B. Lerner [66] and Grollemund et al. [67]. B. Hadad and B. Lerner [66] created a training dataset combining samples from the PRO-ACT (100%) and Tel Aviv (90%) dataset. The samples remaining (10%) of the Tel Aviv dataset were used to test the model. Grollemund et al. [67] performed the validation using the Paris dataset, which was not used in the training and testing stages. Preferably, the external dataset should contain data from the clinical patient population.

When designing ML solutions, we need to be aware of issues that can affect the performance and reliability of the model, such as missing values or data imbalance. The PRO-ACT dataset presented a considerable amount of missing values

what caused that only 32% of its samples could be used in practice. Thus, it is valuable to evaluate how the missing data imputation methods can help to increase the sample size. van der Burgh et al. [62] and Tang et al. [65] used a more straightforward imputation method, calculating the average for each feature and imputed it in the samples with missing values. Leão et al. [69] combined the results of Last Observation Carried Forward and Linear Interpolation missing imputation methods, eliminating posteriorly the samples that still presented some missing values. However, the authors did not detail the sample sizes increase by using these strategies. The data imbalance problem occurs when the training data presents an unequal distribution between samples regarding some class of interest. S. Pires et al. [68] combined Undersampling and Oversampling techniques to achieve a balance of 50% between the classes of interest. Grollemund et al. [67] reported that the data imbalanced related to the target prediction (1-year survival probability) influenced the choice of adequate evaluation metrics due to 75% of the patients had survived for more than one year.

3.4.2 Predictive Biomarkers Analysis

Although some biomarkers evaluated are collected longitudinally (e.g., ALSFRS, respiratory, laboratory), most studies modeled these temporal data as non-temporal by summarizing longitudinal data into single values (e.g., slope, minimum, maximum, mean, standard deviation). This approach is termed Summary Measures and has some advantages such as being simple to comprehend, can be applied with unequal time intervals between measurements, and being considered statistically robust and valid [72]. It allowed that longitudinal information could be processed by non-temporal ML algorithms (e.g., Random Forest, XGBoost) to develop predictive solutions. However, this approach can hide some details about the biomarker changes over time because the aggregated value represents a linear variation over time. For example, an ALSFRS slope decline of 10 in twelve months can be seen as a decline of 0,84 per month (i.e., a linear decline), but the decline may have been accentuated only in the last three months. Future ALS prognosis studies can address this subject by comparing the results obtained using Summary Measures and longitudinal data, depicting the advantages and disadvantages of each approach. Approaches using temporal ML algorithms were presented by D. Halbersberg and B. Lerner [63], B. Hadad and B. Lerner [66] and Leão et al. [69]. S. Pires et al. [68] used a strategy to create several snapshots representing the patient states over time by combining static and longitudinal data.

Regarding the ALSFRS/ALSFRS-R biomarker, we consider the approach of analyzing each subscore separately (e.g., swallowing, walking, writing, respiratory) should be preferred instead of analyzing the total score solely. A more precise analysis of the functional loss characteristics among patients can be performed. For example, two patients can have the same total score but with different values in their subscores, indicating a different disease progression for each patient. In the studies included, this approach helped to find distinct biomarkers associated with each subscore [6, 65, 69].

Different FS strategies were used by the studies included, which helped to find

the more relevant biomarkers related to ALS disease (see Table 3.9 for more detail). Some benefits reported were described hereafter. The FS strategy used by Greco et al. [70] helped to select the 16 best predictors among 108 blood analytes (a reduction of 85%). Two laboratory tests (Chloride and Alkaline Phosphatase) were first associated with ALS progression due to the FS strategy used by Gordon and Lerner [6].

3.4.3 Predictive Machine Learning Approaches

We identified three types of prediction addressed by the studies included (*Disease Progression*, *Survival Time*, and *Need for Support*). The studies evaluated and used different ML algorithms, techniques, datasets, sample sizes, biomarkers, and performance metrics. Consequently, a direct comparison of their performances is difficult, even within a specific type of prediction. In general, the results showed a considerable decrease in the predictive performance when using data from the first 3 months to predict long-term patient functional changes (e.g., at 12 or 24 months). Therefore, performing long-term predictions is still challenging due to ALS heterogeneity and complexity. The high accuracies reported by van der Burgh et al. [62] (87.25%) and Greco et al. [70] (84%) were overshadowed by the reduced number of samples analyzed (135 and 41 respectively), representing an elevated risk of model overfitting. Overfitting occurs when the algorithm presents good performance when using the training data but reduced performance when using the validation data, occurring a super adjust to the training data.

Both ML algorithms used by van der Burgh et al. [62] (Deep Neural Networks) and Grollemund et al. [67] (Dimensionality Reduction) presented interpretability issues by being considered black-box approaches. In these studies, the total comprehension of the relationship between input and output variables can not be easily explained. Physicians will desire to understand how the predictions were obtained to verify if they make sense and are trustworthy to be used for prognostication. The complexity of ALS disease makes a large number of biomarkers necessary to obtain good model performances. This fact also complicates the model interpretability when using black-box approaches. Thus, FS strategies can become an important allied to increase the model interpretability by reducing the number of biomarkers necessary. Some ML frameworks also can be explored to explain predictions obtained with black-box models, such as SHAP [73] and LIME [74]. These frameworks are part of a recent research field termed Explainable Artificial Intelligence (XAI) [75].

Finally, the research efforts analyzed in this review, which used only biomarkers commonly present in the ALS clinical practice, demonstrated promissory results that can be applied in developing CDSS. Unexpectedly, only Gordon and Lerner [6] reported the development of an information system based on their predictive approach and its deployment in an ALS clinical setting. This fact can indicate an absence of CDSS in the ALS prognostic area. Thus, the massive knowledge produced is not used to build decision support systems effective to assist physicians in their daily work. It is an essential step to verify if the results obtained by the

studies will be confirmed in a real-world clinical environment.

As the results are confirmed, the CDSS will become more reliable to be used as a support tool by the physicians, even when black-box approaches have been utilized. From a practical point of view, a CDDS to assist the ALS prognosis could provide numerous valuable predictions. For example, based on the current patient disease progression rate, the system can inform how much a functional condition is estimated to decline in the following months (e.g., speech, respiratory, walking, swallowing). With this information, physicians could plan adequate treatment for the patient and determine if additional support will be needed (e.g., wheelchair, non-invasive ventilation, gastrostomy, cough assist machine). It could also be helpful to keep patients and families informed to better prepare themselves for the changes resulting from the worsening of the disease.

3.4.4 Research Directions

After completing a literature review on the use of ML in predicting ALS prognosis using less complex and costly biomarkers, we held discussions with neurologists from the Department of Internal Medicine at the Federal University of Rio Grande do Norte in Brazil to outline the prognostic predictions to be explored in this thesis. These predictions constituted the basis of the decision support system developed to assist them in ALS prognosis tasks. The identified target predictions were as follows:

1. **Identification of ALS patients with short survival at the time of diagnosis**: Here, we investigated the use of ensemble and imbalance learning techniques (detailed in Chapter 4).
2. **Forecasting the ALSFRS score for one year using three-month patient data**: Here, we explored the use of deep learning and temporal data modeling (detailed in Chapter 5).

ALS is a rare disease, making it challenging to gather patient data. In this sense, we defined the Pooled Resource Open-Access ALS Clinical Trials Database (PRO-ACT) as a primary source of patient data. PRO-ACT is a comprehensive and publicly available resource consolidating data from multiple ALS clinical trials. It contains over 11,600 patient records, providing information on clinical, functional, respiratory, laboratory tests, death reports, and other exams [61]. Therefore, this database was considered suitable for the investigations outlined in this thesis.

3.5 Conclusion

This chapter reviewed relevant articles published between 2011 and 2021 that addressed the development of ML solutions to support the ALS prognosis. The

studies are promising, but some aspects need special attention. The datasets concentrated patients' data mainly from the USA and Europe. Thus, there is a need to collect and analyze data from other world regions to ensure that the ML solutions can be, in fact, generalized to all populations. When analyzing medical data, the Missing Values and Data Imbalance problems need to be addressed to avoid a negative impact on models' performance and reliability. The model interpretability issue is another important point to consider when using ML algorithms considered black-box, such as Neural Networks and Dimensionality Reduction. Despite the research advances, there is a probable lack of CDSS to assist the physicians in their daily work on ALS disease prognosis.

Chapter 4

Identifying Short-survival ALS Patients at Diagnosis

4.1 Introduction

Machine Learning (ML) has emerged as a powerful tool in improving disease diagnosis and prognosis. Learning from complex domains (e.g., ALS disease) is challenging, and ML techniques like Ensemble Learning can help improve predictive performance. Ensemble Learning combines single predictive models to build a more complex one, aiming to surpass the performance of each constituent model separately [76]. Furthermore, various problems and datasets within the biomedical and clinical domain have imbalanced datasets. This issue arises when there is a significant imbalance in the number of samples between different classes, resulting in a substantially lower representation of one class compared to the others. Typically, the target prediction is linked to samples from the minority class, as in the case of detecting patients with lung cancer through the analysis of tomography images, where there are significantly more images of healthy patients (the majority class) than those depicting lung cancer cases (the minority class). In such scenarios, ML models frequently exhibit bias towards samples belonging to the majority class, resulting in an elevated misclassification rate within the minority class [44]. To mitigate the issue of imbalance, resampling techniques such as Undersampling and Oversampling can be employed [45].

Efforts in ALS prognosis using ML should be directed toward the development of Clinical Decision Support (CDS) systems. We deemed it essential to develop CDS systems that are feasible for use on a large scale in primary care, taking into account financial limitations. Such biomarkers may include clinical evaluations, assessment of functional capabilities, and respiratory function measurements. These biomarkers are often derived from less expensive and complex procedures, making them more accessible.

Notably, some ML algorithms present results that humans cannot easily understand, decreasing their interpretability (e.g., Artificial Neural Networks or Support Vector Machines). Interpretability refers to the comprehensibility of the de-

cisions made by the ML algorithm [77]. Addressing this issue is crucial to ensure the acceptance of CDS systems in clinical practice, as physicians require explanations for patient classifications. Hence, the development of CDS systems must prioritize interpretability concerns. Existing frameworks can be explored to elucidate the predictions generated by ML models, one of which is the Shapley Additive Explanations (SHAP) framework [78]. SHAP employs a game-theoretic approach to clarify the prediction for a specific instance by quantifying the contribution (SHAP value) of each feature to the classification process. Consequently, SHAP values provide insights into the influence of individual features on the final prediction and their relative significance when compared to other features.

4.1.1 Related Work

van der Burgh et al. [62] demonstrated the positive impact of using Magnetic Resonance Images (MRI) and clinical information to classify ALS patients into survival groups (Short, Medium, and Long). They developed Deep Neural Networks models and obtained an accuracy of 84%. This study presented a high risk of model overfitting due to the reduced number of samples analyzed (n=135). Kueffner et al. [64] presented a crowdsourcing challenge involving more than 30 teams. One of the target predictions was the probability of survival at 12, 18, and 24 months using patient data from the first three months of records. A team using a Gaussian Process Regression model obtained the best performance compared to the others (Z-score $\approx$ 12). Grollemund et al. [67] presented a model based on Dimensionality Reduction to predict one-year survival probabilities. They used the Uniform Manifold Approximation and Projection (UMAP) algorithm to reduce the data. The resultant 2D projection was divided into three areas to classify the patients into Low, Intermediate, and High probabilities groups. The proposed classifier obtained superior performance (F1 score: 96%, Balanced Accuracy: 91%) than the Random Forest and Logistic Regression models.

These studies used different algorithms, techniques, databases, sample sizes, features, and performance metrics. Consequently, it was not possible to directly compare their performance with this study.

4.1.2 Our Contribution

In this study, we have delved into the utilization of Ensemble and Imbalance Learning techniques to enhance the prediction accuracy for ALS patients with short survival expectancy. Our primary aim was to classify patients into Short and Non-Short survival groups based on data collected at the time of diagnosis. The Short survival group comprises individuals who die within 24 months from the onset of symptoms, indicating a rapid disease progression rate. This 24-month threshold was chosen based on the typical life expectancy of ALS patients, which ranges from 2 to 5 years. Hence, our goal was to identify patients in critical condition during diagnosis. This classification is essential for providing timely information to patients and their families, improving the quality of end-of-life care, and facilitating treatment and resource planning. The analyzed dataset

showed a significant data imbalance, with 13% representing the minority class (Short) and 87% representing the majority class (Non-Short). Our focus was centered on the examination of biomarkers commonly encountered in routine ALS clinical practice.

The proposed solution combined Ensemble and Imbalance learning techniques to improve the prediction of critical ALS patients at diagnosis time. Our Ensemble-Imbalance approach obtained the best performance, achieving a Balanced Accuracy of 88% and a Sensitivity of 96% using a Neural Network model as the base classifier. Furthermore, we employed the SHAP framework to provide insights into how the best model conducted patient classifications.

The principal contributions of our study encompass: (i) offering an effective preprocessing methodology for ALS patient data that enabled the extraction of relevant ALS characteristics using biomarkers commonly encountered in clinical practice, (ii) the development and evaluation of models through an Ensemble-Imbalance-based approach, resulting in improved performance in identifying critically affected ALS patients at the time of diagnosis, and (iii) delivering both global and local explanations regarding the model's prediction mechanisms, including the identification of pivotal features and their correlations with the target variable.

4.2 Methods

To ensure the systematic execution of our experiments, we organized our models into two distinct scenarios: Single-Model and Ensemble-Imbalance. In the initial phase, we designed and evaluated models using state-of-the-art ML algorithms, including k-Nearest Neighbors (k-NN), Decision Tree (DT), Random Forest (RF), Support Vector Machines (SVM), Naïve Bayes (NB), and Neural Networks (NN). These models constituted the Single-Model scenario. Subsequently, we utilized the top ten models for each algorithm as base classifiers to develop and evaluate the Ensemble-Imbalance-based models. Following this, we selected the best-performing model for each algorithm and scenario and conducted a comparative analysis of their results. Finally, we employed the SHAP framework to elucidate how the overall best model executed patient classifications, offering insights into the significance of each feature, in addition to providing both global and local interpretability of the model. Supplementary information on variable distribution and model development is available in Appendix A.

4.2.1 Patient Data

All data used in this investigation were sourced from the PRO-ACT Database [61]. For the purposes of our study, we extracted a range of pertinent information, including demographic details (age, weight, height, and gender), the administration of the Riluzole drug, familial medical history, Forced Vital Capacity (FVC; expressed as a percentage of normal for a healthy individual, adjusted for gen-

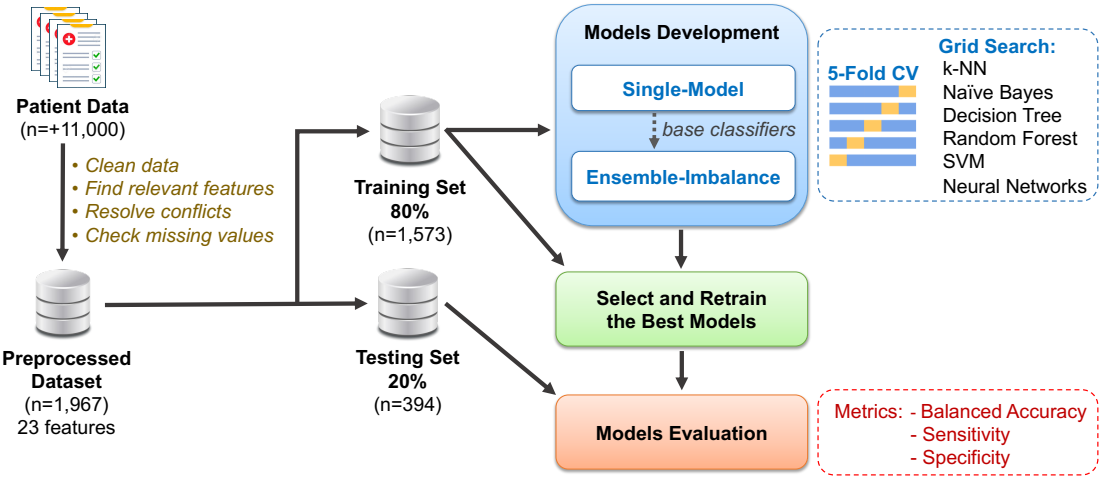


Figure 4.1: ML pipeline design to execute the experiments of this study.

der, age, and height), Slow Vital Capacity (SVC), Body Mass Index (BMI), El Escorial diagnostic criteria, ALS Functional Rate Scale (ALSFRS), and Revised ALS Functional Rate Scale (ALSFRS-R). The ALSFRS scale comprises ten inquiries focused on assessing various physical functionalities, such as speech, swallowing, handwriting, turning in bed, walking, climbing stairs, and respiratory [18]. The ALSFRS-R scale replaced the single respiratory function question with three more detailed questions [19].

4.2.2 Design of the Experiments

We put forward an ML pipeline divided into the stages as shown in Figure 5.1a, which were detailed hereafter. All experiments were performed using the Python programming language and packages for data analysis, Machine Learning, and data visualization, including Pandas [79], Scikit-Learn [43], Imbalance-Learning [80], Matplotlib [81], Seaborn [82], SHAP [83], and NumPy [84].

Data Preprocessing

Patient data collected during diagnosis was analyzed, as previously mentioned. To facilitate their analysis, temporal features were transformed into static data through a technique known as Summary Measures. This approach offers several advantages, including simplicity of interpretation, compatibility with uneven time intervals between measurements, and statistical robustness and validity [72]. We utilized values recorded on the date of diagnosis for temporal features such as FVC, SVC, and BMI. In instances where these values were unavailable, we selected the measurement closest to the diagnosis date for the respective samples.

As recommended in our previous study [85], we analyzed the slope of each ALSFRS question separately instead of the total slope. This approach enabled us to perform a more granular examination of functional loss characteristics among pa-

tients, aiding in the identification of the most pertinent ALSFRS questions for our target prediction. Gordon and Lerner [6] presented an approach to merge data from both ALSFRS and ALSFRS-R scales by combining the samples using only information about *Dyspnea* (question 10) for those assessed with the ALSFRS-R scale. This enabled them to convert the ALSFRS-R scale to ALSFRS, thereby expanding the sample size. In alignment with this approach, we adopted the same strategy in this study, as 51% of the PRO-ACT samples were assessed using the ALSFRS-R scale. Consequently, questions 11 and 12 of the ALSFRS-R scale were not included in our analysis. To model the ALSFRS questions as non-temporal variables, we summarized their data into single slope values. These slopes were calculated as depicted in equation (4.1), where 4 represents the maximum question score, *Question_Score_at_Diagnosis* denotes the score assessed at (or closest to) the time of diagnosis, and *Disease_Duration* is the time in months between symptom onset and the time of diagnosis.

$$Slope = \frac{4 - Question\_Score\_at\_Diagnosis}{Disease\_Duration} \tag{4.1}$$

Additional features were created to store information about the age at symptom onset, the BMI, and whether the patient deceased within 24 months from symptom onset (Survival Group). The age at onset was calculated using information about the age at diagnosis and the disease duration. The BMI was calculated using the patient weight and height collected at the diagnosis. The Survival Group feature was used to classify patients with respect to the target prediction, i.e., into the Short and Non-Short survival groups. The Short survival group included patients who died within 24 months from the onset of the symptoms. We excluded patients whose last visit was within 24 months of disease onset and who were not marked as deceased in the PRO-ACT database.

We performed a complete case analysis, whereby our preprocessed dataset comprised solely of samples with no missing feature values. Features exhibiting a significant percentage of missing values were excluded: SVC (87%) and El Escorial (71%). This action was imperative to prevent the loss of a substantial number of samples. Before being used by the ML models, the features were scaled to a range between 0 and 1, and the dataset was partitioned into Training and Validation subsets. We allocated 80% of the samples for training the models and reserved 20% for validation.

Models Development

This phase encompassed two key steps: (i) splitting the training data using a 5-fold Cross-Validation (CV) repeated three times and (ii) executing the models using a grid-search strategy. We developed models using the following ML algorithms: k-NN, NB, DT, RF, SVM, and NN. In the Single-Model scenario, the models were directly executed using the 5-fold CV strategy in conjunction with diverse hyperparameter configurations as part of the grid search.

In the Ensemble-Imbalance scenario, the models were executed using the follow-

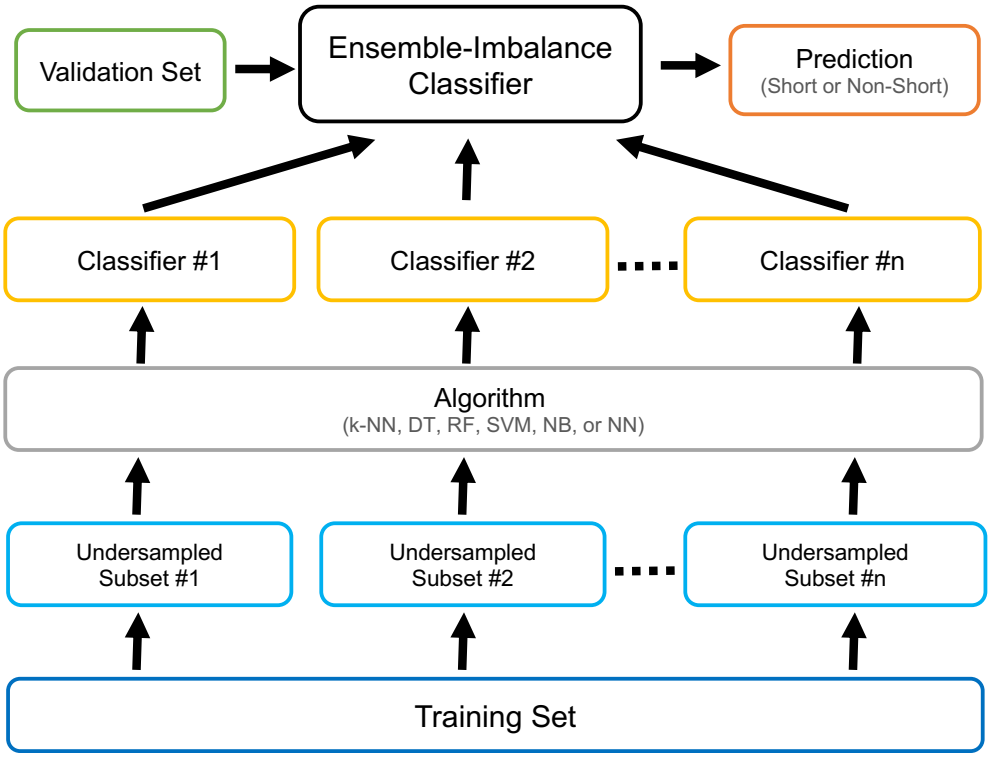


Figure 4.2: Overview of the Ensemble-Imbalance classifier proposed in this study. First, independent undersampled subsets are generated from the Training set using the Random Undersampling method. Then, classifiers created using a specific ML algorithm learn from these subsets (each classifier accesses only one subset). Finally, a majority voting strategy is used to classify patients into survival groups.

ing classifiers: *Balanced Bagging* (DT, SVM, NN, NB, and k-NN) and *Balanced Random Forest* (RF). These classifiers combine Ensemble and Resampling approaches. The models were trained using the ten best classifiers from the same algorithm (e.g., SVM), learning from different balanced subsets resampled from the Training set. Each subset was generated using the Random Undersampling [86] method, which equalizes the number of samples in the different classes (Short and Non-Short) by randomly removing samples from the majority class. Figure 4.2 provides an overview of the Ensemble-Imbalance classifier proposed in this study.

Selecting and Retraining the Best Models

The top-performing models were chosen based on their Balanced Accuracy achieved using the Training set in both scenarios. Afterward, the best models were retrained using the Training set and used to make predictions by accessing the Validation set. All obtained validation performance metrics were recorded and subjected to subsequent analysis and comparison.

Models Evaluation

In our evaluation and comparative analysis of all ML models, we employed the following metrics in this order: Balanced Accuracy, Sensitivity, and Specificity. Sensitivity and Specificity were utilized as they signify the proportion of correctly classified Short and Non-Short survival patients, respectively. It is worth noting that Sensitivity held greater significance than Specificity in our evaluation, given our priority was to correctly classify Short survival patients, as they represent the critical cases. Balanced Accuracy was selected as the appropriate metric for evaluating the experiments as it represents the arithmetic average of Sensitivity and Specificity. Consequently, a higher Balanced Accuracy signifies superior predictive performance concerning both groups of patients.

We applied the Bonferroni correction method to ascertain whether the performance attained in the Ensemble-Imbalance scenario significantly surpassed that of the Single-Model scenario for each algorithm. It is essential to counteract the multiple comparisons problem due to the number of executions using 5-Fold CV repeated three times.

Feature Importance and Model Explanation

Following the evaluation and identification of the best overall model (specifically, the Ensemble-Imbalance-based model utilizing NN as the base classifier), we conducted an in-depth analysis of how this model classifies patients using the SHAP framework. We detailed the significance of each feature for the classification process in the results section, providing comprehensive insights into global and local interpretability. To generate SHAP values and explanations, we employed the *Kernel-Explainer*. All SHAP graphs were produced using the functionalities provided by this framework.

4.3 Results

4.3.1 Data Preprocessing

This study accessed ALS patient data from the PRO-ACT database. Despite its large number of samples (over 11,600), we used only 17% of the available data. We opted to perform a complete case analysis, which reduced the number of samples that could be included due to a high percentage of missing values. The preprocessed dataset encompassed 1,967 patients, each characterized by 23 features. This dataset exhibited an Imbalance Ratio of 6.9 concerning the distribution of the minority and majority classes, with Short-survival comprising 13% and Non-Short constituting 87% of the cases. Table 4.1 provides a comprehensive overview of all features analyzed in this study, along with their respective values and distributions.

Table 4.1: Features analyzed in this study with details on overall distribution and by survival group.

Features		Values	All Samples (n=1967)	Short (n=250) (13%)	Non-Short (n=1717) (87%)	Temporal
Diagnostic Delay (Disease Duration)		• Short ($\leq$ 8 months)	39%	66%	35%	No
		• Average (9−18 months)	40%	29%	41%	
		• Long ($\geq$ 19 months)	21%	5%	24%	
Age at Onset (range)		• 0−39	13%	4%	15%	
		• 40−49	22%	16%	22%	
		• 50−59	31%	35%	31%	
		• 60−69	26%	31%	25%	
		• 70+	8%	14%	7%	
Sex		• Female	36%	30%	37%	
		• Male	64%	70%	63%	
Site of Onset		• Bulbar	19%	26%	18%	
		• Limb/Spinal	81%	74%	82%	
Riluzole		• No	69%	74%	68%	
		• Yes	31%	26%	32%	
Forced Vital Capacity (FVC)		• Abnormal (<80%)	21%	32%	20%	Yes
		• Normal ($\geq$80%)	79%	68%	80%	
Body Mass Index (BMI)		• Underweight ($\leq$ 18.4)	3%	3%	3%	
		• Normal (18.5 − 24.9)	37%	40%	37%	
		• Overweight (25.0 − 29.9)	37%	38%	37%	
		• Obesity ($\geq$ 30)	23%	19%	23%	
Gastrostomy		• No	96%	97%	96%	
		• Yes	4%	3%	4%	
Regions Involved	Quantity	• 1	13%	10%	14%	
		• 2	28%	25%	28%	
		• 3	34%	36%	33%	
		• 4	25%	29%	25%	
	Bulbar	• No • Yes	35%	29%	36%	
			65%	71%	64%	
	Upper Limb		19%	20%	19%	
			81%	80%	81%	
	Lower Limb		13%	13%	13%	
			87%	87%	87%	
	Respiratory		62%	54%	63%	
			38%	46%	37%	
ALSFRS Slopes by Question	Q1−Speech	• Rapid ($\geq$ 0.14) • Average (0.05 − 0.13) • Slow ($\leq$ 0.04)	3%	10%	2%	
			18%	34%	16%	
			79%	56%	82%	
	Q2−Salivation		2%	7%	1%	
			12%	26%	10%	
			86%	67%	89%	
	Q3−Swallowing		1%	4%	1%	
			13%	29%	10%	
			86%	67%	89%	
	Q4−Handwriting		3%	12%	1%	
			18%	30%	16%	
			79%	58%	83%	
	Q5−Cutting		4%	17%	2%	
			24%	31%	23%	
			72%	52%	75%	
	Q6−Dressing & Hygiene		5%	21%	3%	
			30%	42%	28%	
			65%	37%	69%	
	Q7−Turning in Bed		2%	11%	1%	
			18%	33%	16%	
			80%	56%	83%	
	Q8−Walking		3%	16%	2%	
			29%	43%	27%	
			68%	41%	71%	
	Q9−Climbing Stairs		11%	35%	8%	
			37%	36%	38%	
			52%	29%	54%	
	Q10−Respiratory		1%	4%	1%	
			8%	20%	6%	
			91%	76%	93%	

4.3.2 Performance Obtained by Algorithm and Scenario

Figure 4.3 visually depicts the top validation performances achieved by each algorithm and scenario. The "p" alongside each algorithm's name indicates the p-value calculated after employing the Bonferroni correction method. Algorithms that exhibited significantly improved performance in the Ensemble-Imbalance scenario were denoted by the ★ symbol. We applied the same method to compare the performance of algorithms in the Ensemble-Imbalance scenario. The Neural Networks outperformed significantly (p-value $<= 0.001$) the others (DT, RF, SVM, and k-NN).

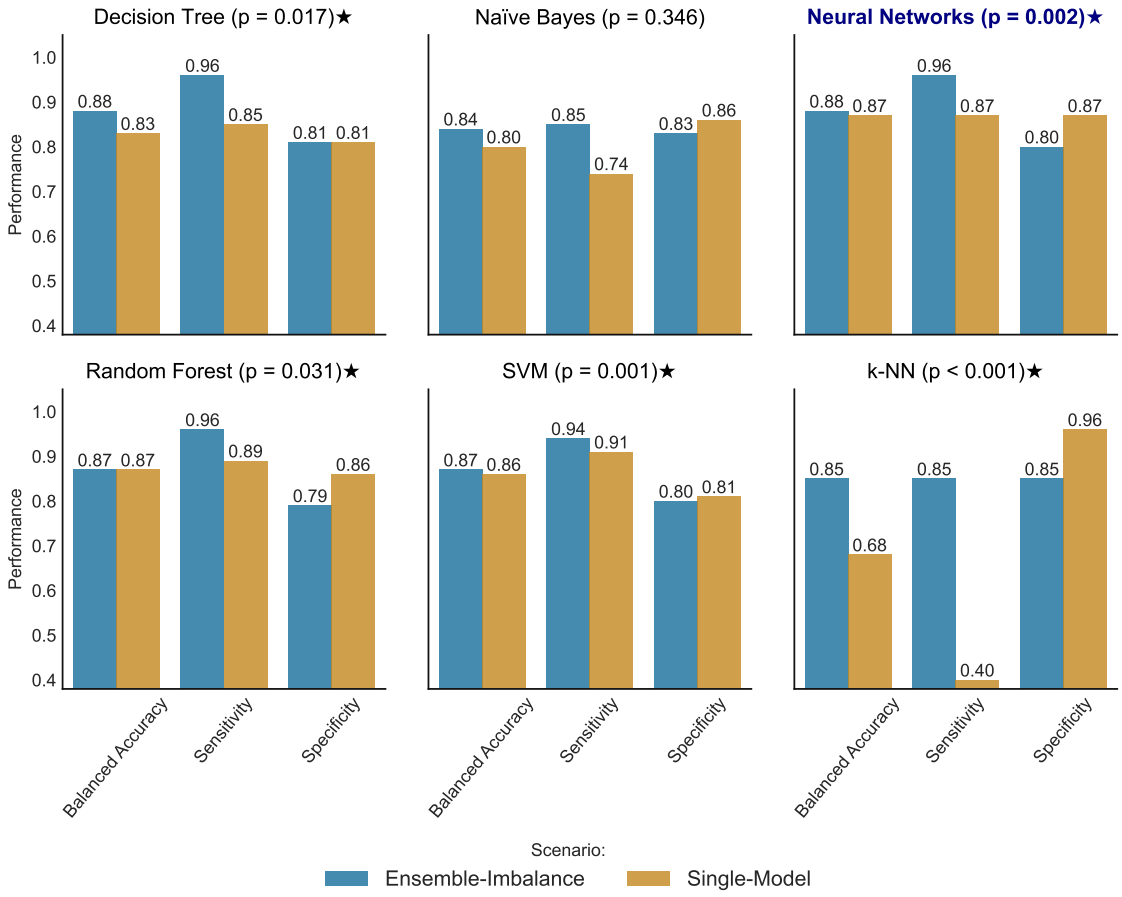


Figure 4.3: Comparison of the best performances obtained by each algorithm and scenario. The "p" represents the p-value obtained after applying the Bonferroni correction method to compare the performance in both scenarios for each algorithm. Algorithms presenting a significantly better performance in the Ensemble-Imbalance scenario were highlighted using the ★ symbol. Neural Networks performed significantly better (p-value ≤ 0.001) than the others in the Ensemble-Imbalance scenario (highlighted in bold and blue font).

4.3.3 Feature Importance and Model Explanation

Following the evaluation and selection of the overall best model (the Ensemble-Imbalance-based model using NN as a base classifier), we utilized the SHAP

framework to obtain insights into how this model conducted patient classifications. Figure 4.4 provides valuable information for comprehending the global interpretability of the model. The left graph displays the ranking of feature importances based on their average impact on the model's output. The right graph illustrates the correlations of feature values with the target prediction. SHAP values on the x-axis exceeding zero indicate that the feature value drove the prediction into the Short survival group, whereas those below zero into the Non-Short group. Figure 4.5 elucidates the global interpretability by detailing the impact on model prediction according to each feature value. Due to space constraints, we present the top ten most relevant features.

Figure 4.6 offers an illustration of how SHAP local interpretability can be leveraged to elucidate the classification of any given patient based on their feature values. This Figure displays information for two patients (A and B) extracted from the Validation set. While Patient A was classified into the Non-Short survival group, Patient B was placed into the Short group. Subfigures 'a' and 'b' show individualized classifications for both patients. The classification process was driven differently according to their feature values (displayed in gray font within parenthesis). Subfigure 'c' illustrates the classification process by comparing both patients on a feature-by-feature basis.

4.4 Discussion

This study assessed the application of Ensemble and Imbalance Learning to enhance the prediction of short-survival ALS patients at the time of diagnosis. Our focus was on the analysis of patient data commonly encountered in routine ALS clinical practice, obtained through a less complex process. We discuss the results obtained in the following subsections.

4.4.1 Predictive Performances

In the Ensemble-Imbalance scenario, most of the algorithms (5 out of 6) exhibited significantly improved performance when compared with the Single-Model scenario (Fig. 4.3). The proposed Ensemble-Imbalance approach notably increased Sensitivity without compromising Balanced Accuracy. This is crucial as it improves the classification of critical patients. The only exception was Naïve Bayes, where the difference between the scenarios was not statistically significant (p-value: 0.346). In the Single-Model scenario, k-NN was the most affected by the data imbalance problem, achieving a Balanced Accuracy of 0.68 and showing a tendency to favor the majority class (Non-Short).

The Ensemble-Imbalance-based model using Neural Networks as a base classifier (EI-NN) outperformed the others significantly (Balanced Accuracy: 0.88; Sensitivity: 0.96; Specificity: 0.80; p-value ≤ 0.001). The Decision Tree, SVM, and Random Forest models demonstrated similar performances to EI-NN. We assume these four models are proper for composing a CDS system inference mechanism

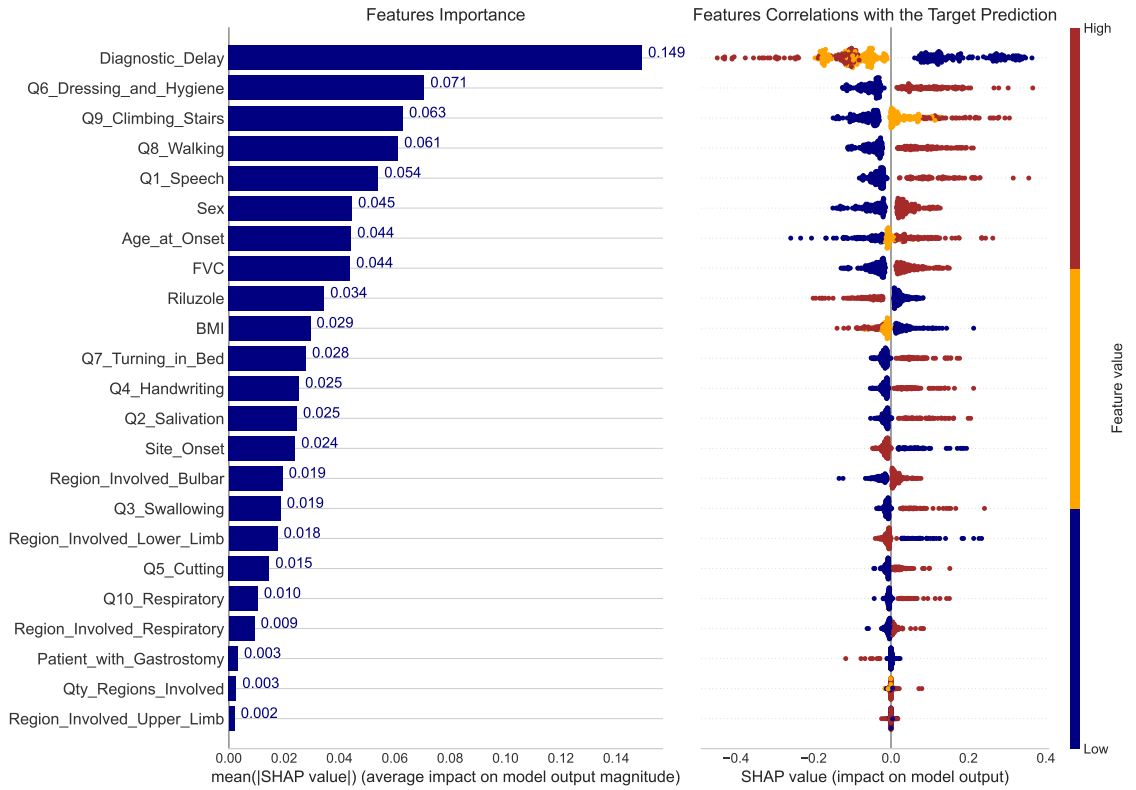


Figure 4.4: Ranking of feature importances and their correlations with the target variable (Short/Non-Short) for the best model in the Ensemble-Imbalance scenario. The x-axis on the left displays the average impact on the model prediction for each feature (mean absolute SHAP value). The x-axis on the right demonstrates the impact on the model prediction (SHAP value) concerning the feature values. Positive SHAP values indicate that the feature value led the model prediction toward the Short survival group, while negative SHAP values pushed it toward the Non-Short group.

for classifying critical ALS patients based on data collected at diagnosis. Our approach yielded promising results, but further validation with unseen data, preferably real-world patient data, is necessary to eliminate bias toward the minority class (Short). This step is essential for a more robust model comparison.

4.4.2 Data Preprocessing

The data preprocessing proposed and executed in this study proved to be highly efficient, enabling ML algorithms to gain a comprehensive understanding of ALS characteristics. Even in the Single-Model scenario, Neural Networks, Random Forest, and SVM models achieved good performance, considering the data imbalance and the complexity of ALS. In the context of ALS prognosis, a data categorization approach may be more effective than direct utilization of the actual feature values. Future studies could explore alternative definitions of categorical values to assess their impact on performance.

Our results also highlighted the feasibility of constructing ML solutions using less

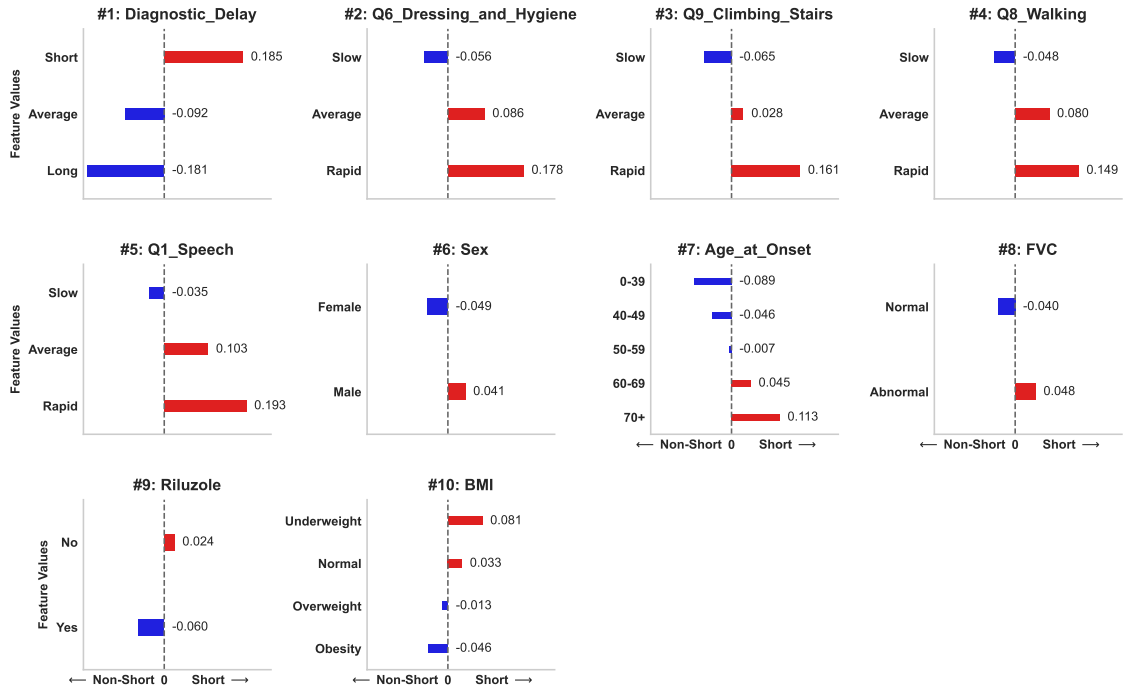


Figure 4.5: Average impact on model prediction for the top ten most important features detailed according to their values. Positive (red) and negative (blue) SHAP values drive the prediction into Short and Non-Short survival groups, respectively.

complex biomarkers. We consider it essential to develop feasible CDS systems for primary care, eliminating the need for more complex and costly biomarkers such as genetics.

4.4.3 Features Importance and Model Explanation

A comprehensive analysis of the results revealed valuable insights into understanding the global interpretability of the model, the importance of the features, and their correlations with target prediction (refer to Figures 4.4 and 4.5). Many features displayed substantial correlations with the target, underscoring their importance in identifying critical patients at the time of diagnosis. Table 4.2 provides details on the type of correlation (positive/negative) for the top ten ranked features based on their categorical-ordinal values. *Diagnostic Delay*, *BMI*, and *Riluzole* exhibited the most relevant negative correlations with the target. Conversely, *Q6-Dressing and Hygiene*, *Q9-Climbing Stairs*, *Q8-Walking*, *Q1-Speech*, *Sex*, *Age Range*, and *FVC* showed the most relevant positive correlations.

The *Diagnostic Delay* was the most relevant among the features. We can observe that 66% of patients in the Short survival group were diagnosed within the first eight months of the onset of the disease (Table 4.1). This is an important biomarker, although it is necessary to analyze the following features to understand which signs led to a faster diagnosis. Following the ranking, questions *Q6*,

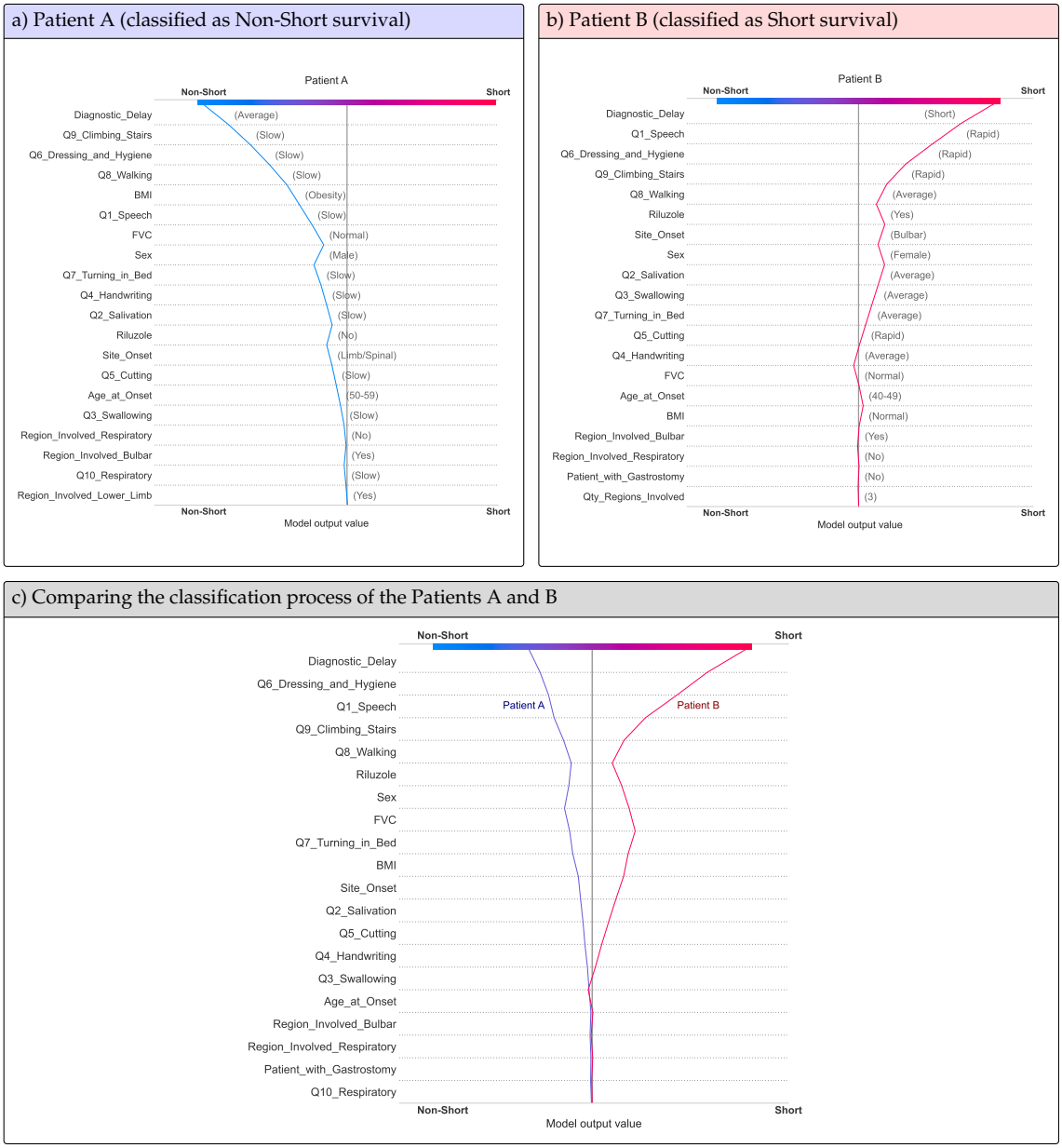


Figure 4.6: Examples of using the SHAP Decision plot to explain how the model classified patients into Short and Non-Short survival groups. Subfigures 'a' and 'b' show individualized classifications for each patient, where the process was conducted according to their feature values (displayed in gray font within parenthesis). Subfigure 'c' shows the classification process comparing feature by feature for both patients. These graphs must be read from bottom to top. The slope of the line within each feature area indicates when the feature value drove the prediction toward the Non-Short (left sloping) or the Short (right sloping) groups. The longer the line length within the feature area, the more significant the impact of its value on the model prediction.

Table 4.2: Features correlations with the target variable, ordered by the type and importance.

Correlation Type	Feature	Ordering of the Categorical Values
Negative	Diagnostic Delay	Short → Average → Long
	Riluzole	No → Yes
	BMI	Underweight → Normal → Overweight → Obesity
Positive	Q6−Dressing & Hygiene	Slow → Average → Rapid
	Q9−Climbing Stairs	
	Q8−Walking	
	Q1−Speech	
	Sex	Female → Male
	Age Range	[0−39] → [40−49] → [50−59] → [60−69] → [70+]
	FVC	Normal → Abnormal

Q9, and *Q8* of the ALSFRS scale appear as the most relevant, thus correlating the degree of lower motor neuron degeneration with a worse prognosis. This aligns with what was reported by Al-Chalabi et al. [87]. Other factors that represented a worse prognosis also conform with the literature on ALS, such as male gender, being older at diagnosis, and having an abnormal FVC. Previous studies using ML applied to ALS prognosis have also identified these features as survival predictors [6, 64, 67, 69]. Hence, we conclude that the proposed Ensemble-Imbalance approach effectively learned from patient data to extract crucial ALS disease characteristics.

The most significant characteristics for identifying critical ALS patients at the time of diagnosis were:

- Shorter diagnostic time ($\leq$ 8 months);
- Higher decline (*slope* $\geq$ 0.14) in ALSFRS questions *Q6* (Dressing and Hygiene), *Q9* (Climbing-Stairs), *Q8* (Walking), and *Q1* (Speech);
- Male gender;
- Age $\geq$ 60 years old;
- Abnormal FVC;
- Not treated with Riluzole;
- Underweight (BMI: $\leq$ 18.4).

Figure 4.6 illustrates the local interpretability of the model based on the SHAP results. Subfigures 'a' and 'b' provide personalized predictions for two patients extracted from the Validation set. Patient A was classified into the Non-Short survival group, whereas Patient B was classified into the Short. Please note that their feature values influenced the classification process (displayed in gray font within parentheses). Our approach enables the identification of the most influential features contributing to disease progression. Consequently, physicians can direct symptomatic treatment to enhance the patient's quality of life. For example, Patient B exhibited a significant functional decline in *Q1-Speech* and *Q9-Climbing Stairs*. This information could guide physicians in deciding that speech

and physical therapies are necessary. Moreover, this information may be employed as inclusion or exclusion criteria in clinical trials, facilitating the selection of patients with predefined characteristics. Subfigure 'c' details both patients feature-by-feature within the same graph, providing a valuable resource for visualizing and comparing two or more patients, thereby revealing their similarities and differences.

This study has certain limitations. The data analyzed were extracted from clinical trials rather than a population-based registry. Consequently, there is a risk that it may not fully represent the entire ALS population due to the inclusion and exclusion criteria applied. Furthermore, the dataset exclusively comprised ALS patients from the United States of America. It is essential to validate the results using data from other regions with different genetic backgrounds (e.g., South America, Africa, or Asia). Additionally, information about cognitive impairment at the time of diagnosis was absent in the PRO-ACT database, likely due to its use as an exclusion criterion in clinical trials. Nonetheless, this biomarker has been previously identified as a significant contributor to a worse prognosis, independent of specific motor impairments [88]. Therefore, it potentially holds relevance as a feature for classifying critical patients at the time of diagnosis.

4.5 Conclusion

This study evaluated the use of ML to predict short survival in ALS patients by analyzing biomarkers collected at the time of diagnosis. We focused on analyzing biomarkers commonly encountered in daily ALS clinical practice, thus avoiding the need for more complex and costly biomarkers such as genetics or imaging. Our findings demonstrate that the proposed Ensemble-Imbalance approach can significantly enhance predictive performance in classifying critical patients during diagnosis. Furthermore, we provided detailed insights into how the model generates predictions, emphasizing both global and local interpretability.

Data and Code Availability

The data used in this study can be obtained from the Pooled Resource Open-Access ALS Clinical Trials website (https://ncri1.partners.org/ProACT). It is important to note that data derived from this database cannot be shared due to restrictions. However, comprehensive details about the source code employed in this study, including data preprocessing, model development, hyperparameter settings, and software versions, are available at the public repository `https://github.com/fabianopapaiz/ensemble_imbalance_model_for_als_prognosis`. Please refer to the instructions in the README file.

Chapter 5

Predicting ALS Progression Using Autoregressive Deep Learning Models

5.1 Introduction

The ALS Functional Rating Scale (ALSFRS) serves as a crucial tool for evaluating patients' functional disability as their disease progresses over time. Administered during clinic visits, this questionnaire evaluates various physical functions such as speech, swallowing, writing, walking, and respiratory abilities [18]. Each question is assessed on a scale from 0 to 4, with 0 indicating severe disability and 4 indicating no disability. The total score, which can range from 0 to 40 (ALSFRS) or 0 to 48 (ALSFRS-R) points, is commonly used to monitor changes in functional capacity over time. An alternative method involves calculating the slope value, which quantifies the rate of functional decline over a specified time period in points per month, with higher values indicating faster disease progression.

In recent years, several studies have applied ML to predict the functional decline (ALSFRS/ALSFRS-R) over time with some success [6, 63–66, 69, 89–92]. However, only B. Hadad and B. Lerner [66] and Pancotti et al. [91] employed deep learning algorithms and temporal data modeling in their investigations. This limited number of studies highlights a noticeable gap in the adoption of these advanced techniques in ALS research, as noted by Tavazzi et al. [93].

5.1.1 Related Work

B. Hadad and B. Lerner [66] studied the prediction of the ALSFRS score in several time intervals, varying from 6 to 24 months. They combined temporal and static data to define the input variables related to each patient visit. The temporal features analyzed were the ALSFRS, vital signs, forced vital capacity (FVC), and five laboratory tests. The static features comprised the age, gender, family history, and the site where symptoms first appeared (site of onset). They compared

Long Short-Term Memory (LSTM) with the non-temporal models Random Forest, XGBoost, and Multilayer Perceptron. XGBoost outperformed the others, and the LSTM obtained the lowest performance.

Pancotti et al. [91] aimed to predict the decline of the ALSFRS slope at 12 months using patient static and temporal data from the first three months. The authors analyzed 202 temporal and static features, including ALSFRS, FVC, vital signs, laboratory tests, demographics, site of onset, and use of Riluzole medication. They developed models combining three neural network architectures: Convolutional, Recurrent, and Feed-Forward. The results were then compared with the non-temporal Random Forest and Bayesian Additive Regression Tree models. The performance of the deep learning-based models did not significantly outperform the non-temporal models.

Both studies used data from the PRO-ACT Database [61]. The deep learning-based models developed by Pancotti et al. [91] obtained a Root Mean Squared Error (RMSE) of 0.52 to predict the ALSFRS slope at 12 months. In the experiment where B. Hadad and B. Lerner [66] evaluated the LSTM using only data from the PRO-ACT, they achieved an RMSE of 4.02 and 5.70 to predict the ALSFRS score at 6 and 12 months, respectively.

Pancotti et al. [91] employed a time series forecasting approach called Single-Shot, where all time steps were forecasted at once to provide the target prediction, i.e., the slope at 12 months. B. Hadad and B. Lerner [66] employed an approach inspired by Nahon and Lerner [94], where the learning instances were created from multiple observations, each referring to a prediction of a specific future visit using different quantities of past visits. In this study, we employed a time series forecasting approach called Autoregressive, which is detailed in the following sections.

5.1.2 Our Contribution

We proposed a novel approach based on the Autoregressive Multi-Step Multi-Output Time Series Forecasting to model temporal features and train deep learning models [51–53]. Using patient data collected from the first three months, we predicted the ALSFRS score for the next 12 months in the future (month-by-month). To the best of our knowledge, this study is the first to employ this approach to ALS prognosis to predict the decline of ALSFRS over time. Similar to related studies, we analyzed patient data from the PRO-ACT database. We extracted the temporal features from the ALSFRS data (i.e., the total score, question subscores, and information on regions involved over time). The static features included age, gender, site of onset, disease duration, and the use of Riluzole medication. We implemented and evaluated temporal models using Gated Recurrent Unit (GRU) and LSTM. Our approach achieved superior performance compared to the related works using a much smaller set of input features (see section *5.3.3 Performance Comparison*). Hence, the results obtained suggest a greater effectiveness of our approach.

5.2 Methods

We implemented an ML pipeline, as shown in Fig. 5.1a, which will be detailed hereafter. Additional information on model development is available in Appendix B.

5.2.1 Patient Data

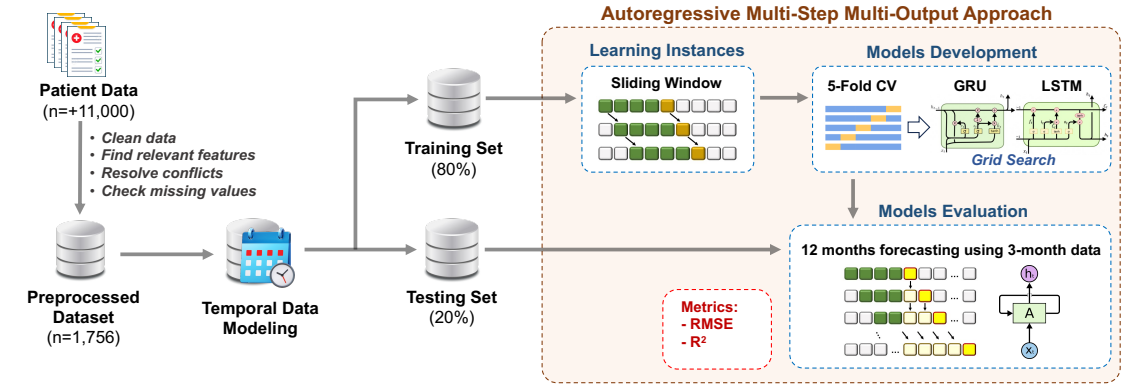
All data used in this investigation were sourced from the PRO-ACT Database [61]. We extracted both static and temporal features for this study, focusing on key patient data necessary for accurate ALS prognosis. The static features comprised the age, gender, disease duration, site of onset, El Escorial diagnostic criteria, and Riluzole treatment. We extracted temporal features from the ALSFRS and ALSFRS-R data.

5.2.2 Data Preprocessing

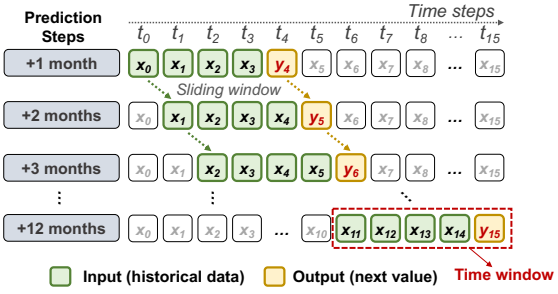
We performed a complete case analysis, and our preprocessed dataset consisted of samples with no missing values. The *El Escorial* feature exhibited a significant percentage of missing values (71%) and was excluded from this study.

Gordon and Lerner [6] converted data from the ALSFRS-R scale to ALSFRS by combining the samples using only information about *Dyspnea* (question 10) for those assessed with the ALSFRS-R scale. This strategy allowed them to expand the sample size. We adopted the same strategy in alignment with this approach because 51% of the PRO-ACT samples were assessed using the ALSFRS-R scale. Consequently, we excluded questions 11 and 12 of the ALSFRS-R scale from our analysis.

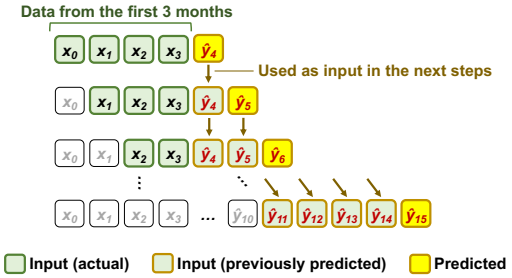
Additional features were created to store information about the age at symptom onset and gastrostomy support. The age at onset was calculated based on the age at the first visit and the disease duration. The gastrostomy feature indicates if a patient received gastrostomy feeding support over time and was defined based on the subitems 'a' and 'b' of question 5 (Q5-Cutting) from the ALSFRS scale. We also added features to indicate if a determined body region was affected or not over time. These features were created based on the ALSFRS question subscores and divided into the following groups: Bulbar, Upper Limb, Lower Limb, and Respiratory. The Bulbar group was defined based on the Q1-Speech, Q2-Salivation, and Q3-Swallowing subscores. If any points were lost in one of these subscores, this feature was marked as 'Yes'; otherwise, it was marked as 'No'. The other groups were defined similarly. The Upper Limb group was defined using the Q4-Handwriting and Q5-Cutting subscores. The Lower Limb group was defined using the Q8-Walking and Q9-Climbing-Stairs subscores. The Q10-Respiratory subscore was used to define the Respiratory group. Additionally, we created the Regions Involved feature based on these four groups to indicate the number of regions involved at a specific time (varying from 1 to 4). Table 5.1



(a) Machine learning pipeline.



(b) Learning instances generation.



(c) Autoregressive multi-step multi-output forecasting.

Figure 5.1: (a) Machine learning pipeline of our proposed forecasting approach. (b) Demonstrate how the learning instances were generated from the patient data (matrices X's). Each x_t is a vector containing information about temporal and static features of a patient measured at a specific time step t, where $t = 0, 1, ..., 15$, x_0 represented the value measured at the first visit, x_1 represented the value registered after one month; x_2 represented the value registered after two months, and so on. We defined each learning instance as a time window containing data of 5-time steps, where the first four represent the sequence of historical data used as input and the last represents the output to be predicted for the next month. The learning instance for the first prediction step (+1 month) was the time window constituted of the patient data for the initial three months (x_0, x_1, x_2, and x_3) and the output for the fourth month ($y_4 = x_4$). We utilized the sliding window approach to generate the subsequent learning instances, shifting the next time window by one month to the right until reaching the last prediction step (+12 months) formed by the time window $[x_{11}, x_{12}, x_{13}, x_{14}, y_{15}]$. (c) Illustrate the autoregressive multi-step multi-output forecasting approach employed in this study. The initial prediction utilized patient data registered in the first three months to predict values at month four ($\hat{y}_4$). After, the predicted values obtained in each step ($\hat{y}_t$) were fed back as input to the following steps to enable multi-step forecasting. This process was repeated until generating the predictions for the next 12 months ($\hat{y}_4 - \hat{y}_{15}$). Only the temporal features were predicted as an output ($\hat{y}_t$). Before being used by the next step, we concatenated the static features into each $\hat{y}_t$.

Table 5.1: List of the features analyzed in this study

Type	Data Type	Feature
Static	Integer	Age at Onset
	Categorical	Sex
		Site of Onset
	Boolean	Riluzole
Temporal	Integer	Disease Duration
		ALSFRS Total Score
		Q1 - Speech
		Q2 - Salivation
		Q3 - Swallowing
		Q4 - Handwriting
		Q5 - Cutting
		Q6 - Dressing & Hygiene
		Q7 - Turning in Bed
		Q8 - Walking
		Q9 - Climbing Stairs
		Q10 - Respiratory
		Regions Involved
	Boolean	Bulbar Involved
		Upper Limb Involved
		Lower Limb Involved
		Respiratory Involved
		Gastrostomy

describes all features (n=22) used in this study.

We aimed to predict the ALSFRS score and slope for the next 12 months using data collected within the first three months (a total of 15 months). Therefore, we included only patients with sufficient ALSFRS information, i.e., those with at least three registered visits and a minimum interval of 15 months between the first and last visit.

5.2.3 Temporal Data Modeling

For each patient and temporal feature, we defined a vector to store values for time steps varying from 0 to 15 ($t_0 - t_{15}$), where the time step t_0 represented the feature values measured at the first visit; t_1 represented the values measured after one month; t_2 represented the values measured after two months, and so on. This vector represented the disease variation over time month-by-month. Each time step was filled using the data available in the PRO-ACT database. We used the last visit value if more than one visit was registered for the same time step.

After, we employed a simple linear interpolation method to calculate the values for those time steps with no visit information, as illustrated in Fig. 5.2. Please note that some values for "Patient 2" (months 2-4) may not appear linearly aligned due to rounding since the ALSFRS Total Score is an integer value. The average number of imputed values per patient was 6 $\pm$2, with a minimum of 0 and a maximum of 10. For each static feature, we created a vector repeating their values across all time steps. Finally, we combined the vectors to create a matrix X (5.1)

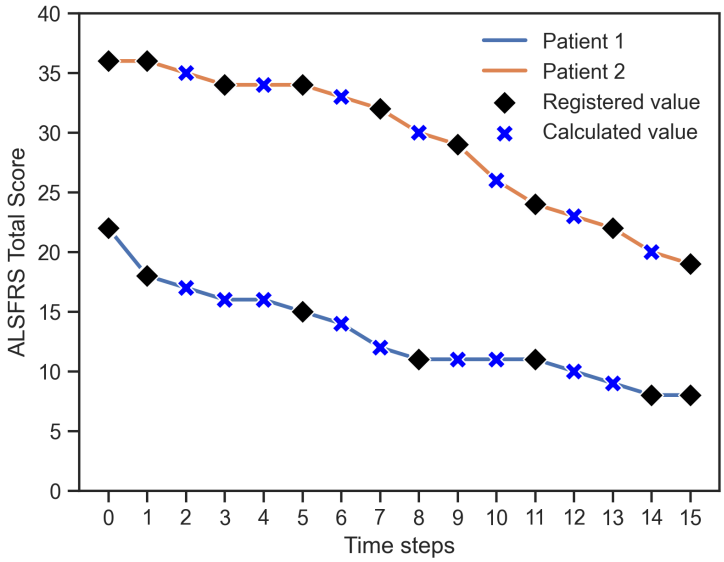


Figure 5.2: Illustration of the interpolation method applied to the ALSFRS Total Score measured over time for two patients extracted from the PRO-ACT database.

containing information about the variation of all features over time (multivariate time series) for a specific patient. This process was repeated for all patients to generate the matrices X's used to train and test the deep models.

$$X_{features \times timesteps} = \begin{bmatrix} x_{1,0} & x_{1,1} & \cdots & x_{1,t} \\ x_{2,0} & x_{2,1} & \cdots & x_{2,t} \\ \vdots & \vdots & \ddots & \vdots \\ x_{f,0} & x_{f,1} & \cdots & x_{f,t} \end{bmatrix} \tag{5.1}$$

where $f = 1, 2, ..., 22$ (features) and $t = 0, 1, ..., 15$ (time steps).

5.2.4 Deep Models Development

The model development was organized in two phases: (i) identifying the optimal hyperparameter settings and (ii) evaluating performance by retraining the best models employing repeated random train-test splits (100 times) in the entire dataset. We also examined a Naïve model that assumed the ALSFRS score remained constant after the third month. This model simply replicated the score from the third month for the subsequent months. This approach allowed us to establish a baseline for comparison with the proposed method employing deep models.

The learning instances were generated from the matrices X's. Each instance comprised a historical data sequence (inputs) and the next value to be predicted (output). Fig. 5.1b explains the process employed to generate the instances in more detail.

Identifying the Optimal Hyperparameter Settings

We divided the preprocessed dataset into training and testing sets, allocating 80% of the patients to train the models and reserving 20% for testing. Scalers were defined using only the training set, and the feature values were scaled to a range between 0 and 1. Categorical features were coded before being scaled. Before evaluating the models, we utilized these same scalers to scale the feature values in the testing set.

We developed GRU and LSTM models using Tensorflow [95] and the ADAM optimizer [96]. To prevent overfitting, we shuffled the learning instances before training and employed the following strategies: 5-fold Cross-Validation (CV), early stopping, regularization, and dropout. Several combinations of hyperparameters were applied using grid-search. Our hyperparameter search space included the number of hidden layers, the model depth (neurons), the initial learning rate, regularization, and the unidirectional and bidirectional architectures. The 5.5 provides comprehensive details about the source code employed in this study, including data preprocessing, model development, hyperparameter settings, and software versions.

Retraining the Best Models

The top-performing GRU and LSTM models were selected based on their average RMSE from the previous phase. These models were then retrained using a random train-test split strategy applied to the entire dataset, repeated 100 times. The dataset was divided into training (80%) and testing (20%) sets in each split. Scalers were defined using only the training set. Before evaluating the models, we employed these same scalers to scale the feature values in the testing set. As in the previous phase, we shuffled the learning instances before training to prevent overfitting.

5.2.5 Models Evaluation

We employed the RMSE metric to evaluate the predictive capacity of the suggested forecasting approach. The RMSE represents the sample standard deviation of the differences between predicted and actual values, and it is commonly used to measure the accuracy of regression problems. The RMSE is measured on the same scale as the data, which helped define the margin of error when visualizing predictions for each patient (see section *5.3.4 Individualized Predictions*). We evaluated the performance for each prediction step, i.e., from the next month up to 12 months in the future. The initial prediction step used patient data from the first three months. Subsequent steps utilized the predicted values obtained in the previous steps as an input (Fig. 5.1c). The best model for each algorithm (GRU and LSTM) was determined based on the overall RMSE obtained for all months.

As part of our analysis, we also computed the Coefficient of Determination (R^2) as a complementary metric. The R^2 measures the proportion of the variance in the

dependent variable explained by the independent variables. It varies from 0 to 1, where 1 means that the model perfectly explains the variance in the dependent variable, while a value of 0 means the model explains none of the variance. This metric helped to illustrate the impact on predictive performance for each time step, allowing comparison of short and long-term forecasting.

5.3 Results

5.3.1 Data Preprocessing

We accessed ALS patient data from the PRO-ACT database. Despite its large number of samples (over 11,000), we used only 15% of the available data. We performed a complete case analysis, which decreased the number of samples that could be included, due to a high percentage of missing values. The preprocessed dataset encompassed 1,756 patients, each characterized by 22 features (see Table 5.1).

5.3.2 Deep Models Performance

The performance of both the GRU and LSTM models was similar across all prediction time steps.

The Naïve model obtained inferior performance for all time steps compared to the deep models. Fig. 5.3 details the RMSE and R^2 obtained for predicting the *ALSFRS Total Score* for the next 3, 6, 9, and 12 months. The best performances were obtained using the following set of hyperparameters for both deep models: 2 recurrent layers with 1,024 neurons, L2 regularization, a dropout layer of 35%, and bidirectional architecture. We found GRU to be the better model due to its simpler architecture and faster training time compared to LSTM. Fig. 5.4 illustrates the neural network layers for the best deep model.

Table 5.2 presents the average performance obtained by each deep model for the months predicted, during the phase of identifying optimal hyperparameter settings. Table 5.3 depicts the average performance obtained during the retraining phase, which involved random train-test splits repeated 100 times. Both tables provide statistical data, including the mean and standard deviation. Additionally, Table 5.3 provides information on the confidence intervals.

All temporal features, not just the ALSFRS Total Score, were forecasted for each time step. Table 5.4 demonstrates the RMSE obtained for predicting each ALSFRS question separately.

The results demonstrated a considerable reduction in the predictive performance over time. This can be observed by analyzing the decrease in the R^2 metric according to the months predicted (Fig. 5.3). Using the GRU as a reference, our approach explained 85% of the ALSFRS Total Score for a 3-month prediction. However, our

approach explained only 60% for a prediction of 12 months. Therefore, long-term forecasting is still challenging due to ALS heterogeneity and complexity.

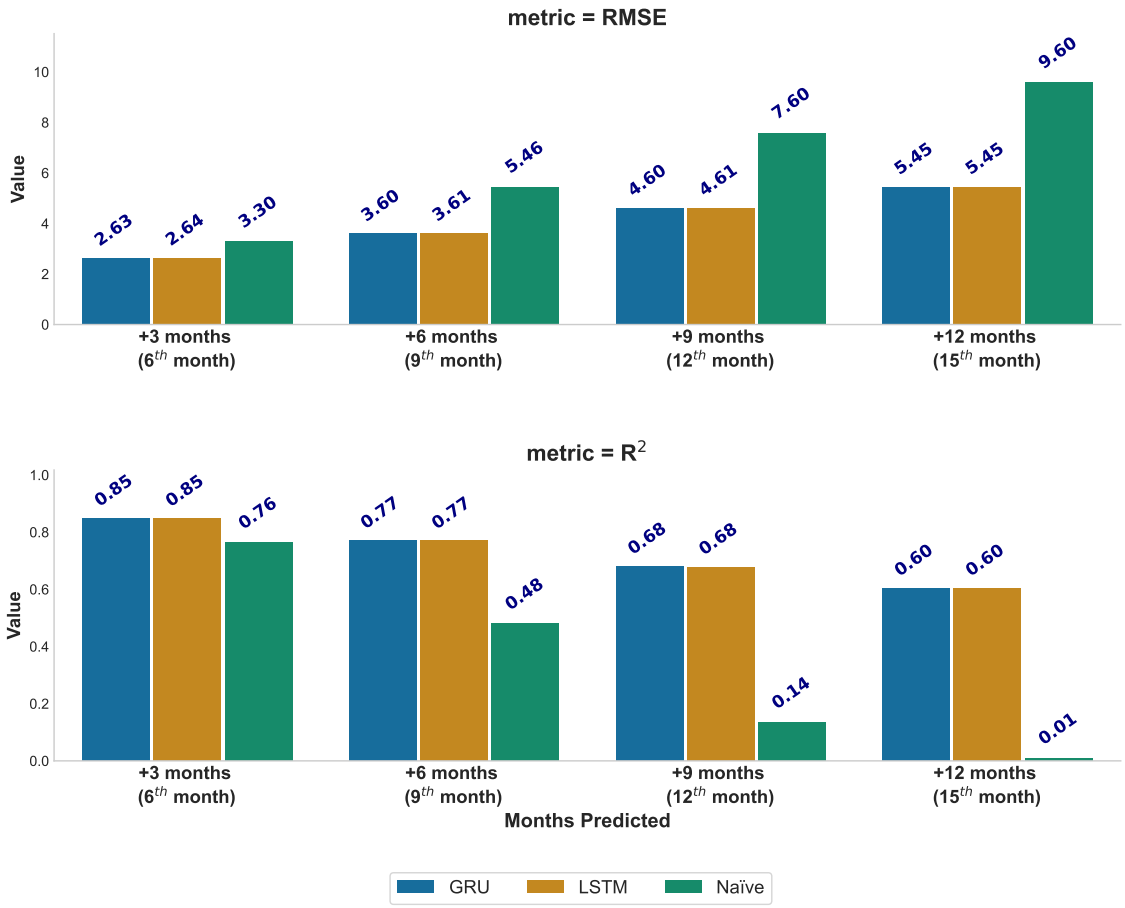


Figure 5.3: Best performance by deep model for predicting the *ALSFRS Total Score* for the next 3, 6, 9, and 12 months using patient data from the first three months. The metrics evaluated were the RMSE (top) and R^2 (bottom). The results were compared with those produced by the Naïve model.

Table 5.2: Best performance obtained by each deep model for the months predicted during the identification of optimal hyperparameter settings phase.

Months Predicted	Model	Mean ±SD	
		RMSE	R^2
+3 months (6^{th} month)	GRU	2.63 ±0.027	0.84 ±0.003
	LSTM	2.64 ±0.013	0.83 ±0.002
+6 months (9^{th} month)	GRU	3.43 ±0.015	0.76 ±0.002
	LSTM	3.43 ±0.033	0.76 ±0.005
+9 months (12^{th} month)	GRU	4.42 ±0.022	0.66 ±0.003
	LSTM	4.45 ±0.046	0.66 ±0.007
+12 months (15^{th} month)	GRU	5.47 ±0.017	0.55 ±0.003
	LSTM	5.48 ±0.054	0.54 ±0.009

Legend: SD: Standard Deviation.

Table 5.3: Best performance obtained by each deep model for the months predicted during the retraining phase, employing random train-test splits repeated 100 times.

Months Predicted	Model	RMSE		R^2	
		Mean ±SD	95% CI	Mean ±SD	95% CI
+3 months (6^{th} month)	GRU	2.63 ±0.143	2.60−2.66	0.85 ±0.017	0.84−0.85
	LSTM	2.64 ±0.145	2.61−2.67	0.85 ±0.018	0.84−0.85
+6 months (9^{th} month)	GRU	3.60 ±0.162	3.57−3.64	0.77 ±0.021	0.77−0.77
	LSTM	3.61 ±0.156	3.58−3.64	0.77 ±0.021	0.77−0.77
+9 months (12^{th} month)	GRU	4.60 ±0.217	4.56−4.64	0.68 ±0.029	0.67−0.69
	LSTM	4.61 ±0.211	4.57−4.65	0.68 ±0.029	0.67−0.68
+12 months (15^{th} month)	GRU	5.45 ±0.227	5.40−5.49	0.60 ±0.033	0.60−0.61
	LSTM	5.45 ±0.214	5.41−5.49	0.60 ±0.032	0.60−0.61

Legend: SD: Standard Deviation; CI: Confidence Interval.

Table 5.4: RMSE obtained by the best model (GRU) for predicting each ALSFRS question for the next 3, 6, 9, and 12 months.

ALSFRS Questions	RMSE by Month Predicted (Mean ±SD)			
	+3	+6	+9	+12
Q1 - Speech	0.50 ±0.037	0.62 ±0.034	0.73 ±0.038	0.84 ±0.042
Q2 - Salivation	0.54 ±0.030	0.65 ±0.032	0.74 ±0.033	0.81 ±0.039
Q3 - Swallowing	0.48 ±0.032	0.63 ±0.034	0.77 ±0.037	0.89 ±0.039
Q4 - Handwriting	0.61 ±0.030	0.73 ±0.031	0.82 ±0.033	0.91 ±0.036
Q5 - Cutting	0.61 ±0.025	0.74 ±0.033	0.83 ±0.034	0.94 ±0.042
Q6 - Dressing and Hygiene	0.61 ±0.029	0.70 ±0.032	0.81 ±0.037	0.89 ±0.039
Q7 - Turning in Bed	0.60 ±0.028	0.74 ±0.033	0.84 ±0.035	0.93 ±0.038
Q8 - Walking	0.50 ±0.023	0.61 ±0.022	0.70 ±0.026	0.78 ±0.026
Q9 - Climbing Stairs	0.66 ±0.028	0.74 ±0.035	0.81 ±0.035	0.85 ±0.031
Q10 - Respiratory	0.61 ±0.041	0.72 ±0.038	0.86 ±0.044	0.97 ±0.039

Legend: SD: Standard Deviation.

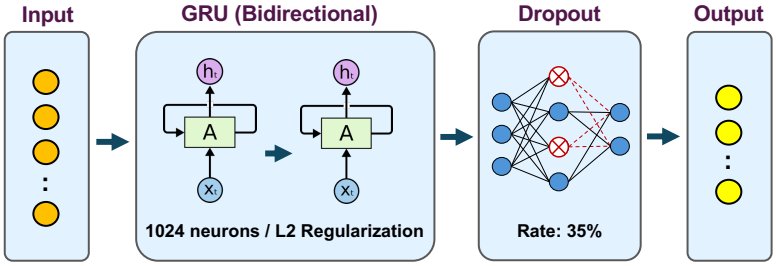


Figure 5.4: Neural network layers of the best deep model.

5.3.3 Performance Comparison

Table 5.5 shows a performance comparison between this study and the related works. We directly compared our results with Hadad and Lerner [66], where the RMSE at 6 and 12 months were obtained from the predictions for "+3" and "+9" months, respectively (see Fig. 5.3). To compare our study with Pancotti *et al.* [91], we calculated the ALSFRS Slope based on the score predicted at 12 months, and the score registered at three months using (5.2).

$$slope = \frac{score(t_{12}) - score(t_3)}{9} \tag{5.2}$$

where $score(t_3)$ is the ALSFRS Total Score registered at three months, $score(t_{12})$ is the score predicted at 12 months, and 9 represents the interval (in months) between the third and twelfth months. Our approach outperformed the method by Hadad and Lerner, achieving RMSE values of 2.63 and 4.60 for predicting the ALSFRS Total Score at 6^{th} and 12^{th} months (Table 5.5). We obtained better performance (RMSE: 0.50) than Pancotti *et al.* (RMSE: 0.52) for predicting the ALSFRS Slope at 12^{th} month.

Table 5.5: Comparison in relation to the RMSE obtained by predicting (i) the *ALSFRS Total Score* at 6^{th} and 12^{th} months (*Mean* $\pm$*SD*), and (ii) the *ALSFRS Slope* at 12^{th} month (*Mean [95% CI]*) from the third visit. The best performances are highlighted in bold.

Study	**ALSFRS Total Score**		**ALSFRS Slope (3^{rd} visit)**
	6^{th} month	**12^{th} month**	**12^{th} month**
Our study	**2.63** $\pm$0.143	**4.60** $\pm$0.217	**0.50** [0.49$-$0.51]
Hadad and Lerner [66]	4.02 $\pm$0.150	5.70 $\pm$0.180	-
Pancotti *et al.* [91]	-	-	0.52 [0.49–0.55]

Legend: SD: Standard Deviation; CI: Confidence Interval.

5.3.4 Individualized Predictions

We produced personalized predictions for three patients selected from the testing set using our best-performing deep model (GRU) (see Fig. 5.5). Using patient data from the first three months, we generated plots to compare the actual and

predicted information for the next 3, 6, 9, and 12 months. The information comprises the ALSFRS Total Score, slopes, and disease duration. The shadowed area denotes the margin of error for the predictions based on the corresponding RMSE for each month ($\pm$ RMSE value).

5.4 Discussion

In this section, we analyze the performance of deep learning models (GRU and LSTM) and compare them with previous studies. Additionally, we explore the implications of our research.

5.4.1 Deep Models Performance and Comparison

Both deep learning models (GRU and LSTM) achieved similar performance levels, outperforming the Naïve model across all time steps (Fig. 5.3). Despite its more elaborate architecture, the LSTM did not outperform GRU. It was contrary to our initial expectations, and one reason could be the time interval used to train the deep models (15-month patient data). However, if an ample time interval was analyzed, such as more than three years, the LSTM could perform better than the GRU. It is important to note that ALS disease is rare, and the acquisition of abundant patient data represents a significant challenge.

Our results demonstrated that the proposed approach outperforms prior studies, highlighting its potential for more effective ALS prognosis (see Table 5.5). Our approach needed a much smaller set of input features (n=22) than the other studies. They analyzed several temporal features, such as ALSFRS, vital signs, FVC, and laboratory tests. On the other hand, our temporal features were based solely on ALSFRS - a straightforward questionnaire commonly used in daily ALS clinical practice.

5.4.2 Implications

Fig. 5.5 provides individualized predictions for three patients, thus allowing visualization of the ALS progression based on their previous consultation data. These plots demonstrate how our findings can contribute to developing a Clinical Decision Support (CDS) [97] system to assist physicians in making more appropriate decisions based on prognosis information about their patients. This information can guide physicians in determining which resources or therapies will be necessary to improve the patient's quality of life in the following months. Moreover, our approach operates with a reduced number of input features (n=22) derived from the ALSFRS, which could facilitate the implementation of CDS systems in primary care settings.

Our approach is Multi-Output, which means that all temporal features (not only the ALSFRS Total Score) were forecasted for each time step. Table 5.4 shows the

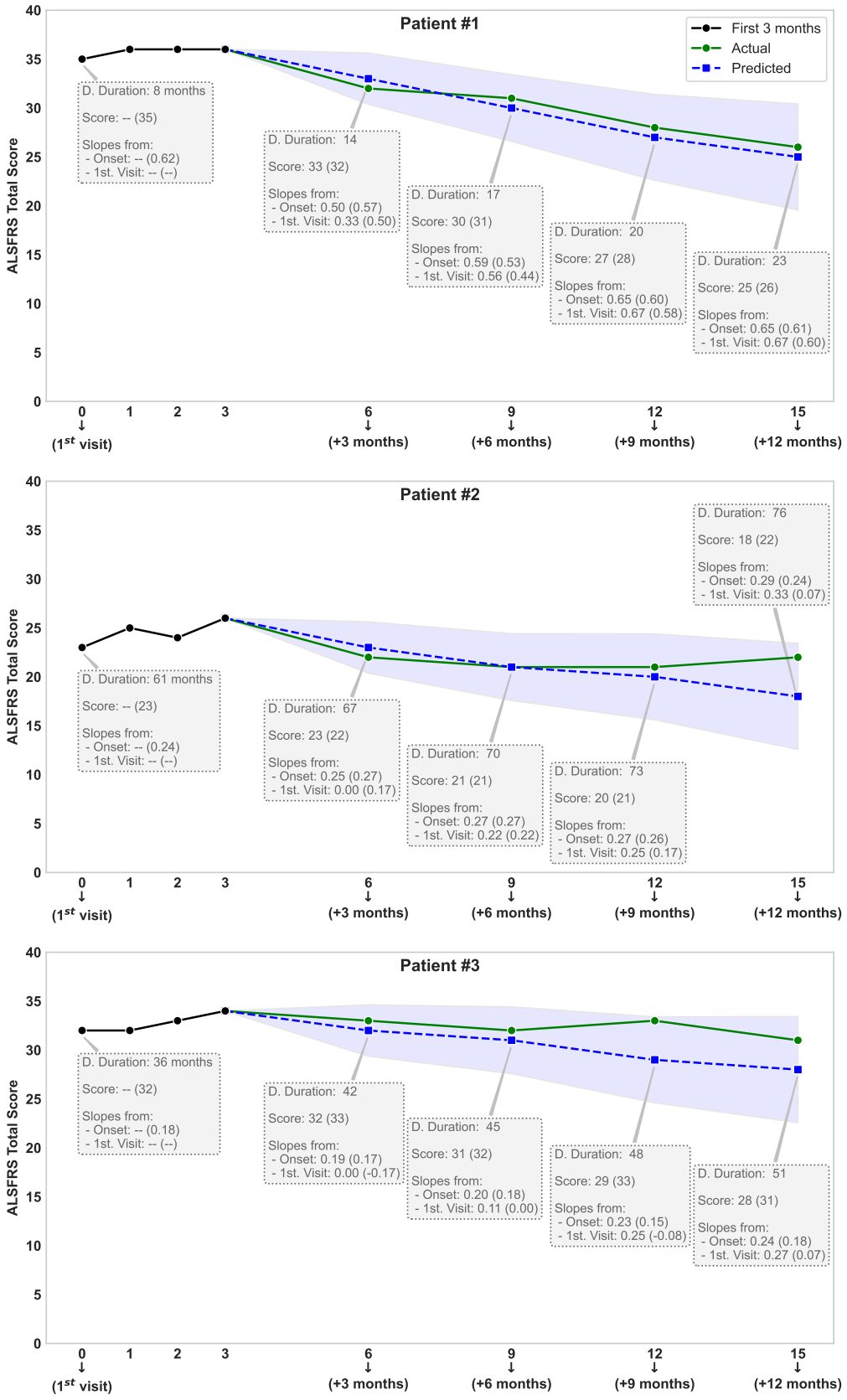


Figure 5.5: Examples using data from the first three months to make predictions for patients extracted from the Testing set. The plots compare the actual and predicted values for the next 3, 6, 9, and 12 months. The information comprises the disease duration, total score, and slopes from symptoms onset and first visit. The shadowed area represents the margin of error based on the RMSE. Score and slope values inside annotations are represented in the "predicted (actual)" format.

RMSE obtained for predicting all ALSFRS questions separately. It also represents valuable information to physicians. For example, based on the prediction for the next six months for questions 8 (walking) and 9 (climbing stairs), physicians can decide and plan if the patient will need a wheelchair in this time interval. Additionally, the provided predictions may be used as inclusion or exclusion criteria in clinical trials, facilitating the selection of patients with predefined characteristics.

5.5 Conclusion

In this study, we proposed a novel approach based on autoregressive multi-step multi-output time series forecasting to predict functional disability for the next 12 months, month-by-month, using patient data collected from the first three months. This study is the first to employ this approach to ALS prognosis to predict the functional decline over time. We extracted static and temporal features from the Pooled Resource Open-Access ALS Clinical Trials database. We developed and evaluated deep learning models using the Gated Recurrent Unit and Long Short-Term Memory algorithms. Our approach outperformed previous works with a significantly smaller set of input features, thus demonstrating greater effectiveness. With the promising results obtained, our approach could aid physicians in devising personalized treatment and resource planning or serve as an inclusion/exclusion criterion in clinical trials.

Data and Code Availability

The data used in this study can be obtained from the Pooled Resource Open-Access ALS Clinical Trials website (https://ncri1.partners.org/ProACT). It is important to note that data derived from this database cannot be shared due to restrictions. However, comprehensive details about the source code employed in this study, including data preprocessing, model development, hyperparameter settings, and software versions, are available at the public repository `https://github.com/fabianopapaiz/autoregressive_deep_network_for_predicting_als_progression`. Please refer to the instructions in the README file.

Chapter 6

Clinical Decision Support System for ALS Prognosis

6.1 Introduction

Despite research efforts, there is still a need to make the knowledge produced available in the clinical environment. It is essential to assist health professionals in their daily work. One way to bridge this gap is to develop a Clinical Decision Support (CDS) system focused on ALS issues. CDS are computer programs designed to help healthcare workers make more appropriate and timely decisions about patients by linking their current information with that previously gathered from other patients. In this way, a CDS represents a valuable tool to help healthcare workers make decisions based on information and experience acquired over time [55–57].

The final contribution of this thesis was the development of a CDS system that offers ALS prognosis information related to the target predictions mentioned in Chapters 4 and 5. These predictions are crucial in assisting physicians to provide more timely and appropriate information to patients and their families, improve the quality of end-of-life care, and facilitate treatment and resource planning.

6.2 Methods

The CDS system was designed to provide the following predictions: (i) identification of ALS patients with short survival at the time of diagnosis, and (ii) forecasting the ALSFRS score for one year using three-month patient data.

6.2.1 Acquiring Brazilian Patient Data

The first and crucial step in the process was to prepare and submit a comprehensive health research project to the National Ethics Committee. This was necessary

to obtain authorization to collect and access Brazilian patient data. This process was quite bureaucratic and required details on research objectives, eligibility criteria, biomarkers to be collected, risks to patients, benefits, and the data analysis methodology to be employed. In total, it took 15 months from the first submission to final authorization. The biomarkers to be collected during the patient visit were determined based on the findings described in Chapter 3 and are detailed in Table 6.1. The complete project document is available in Appendix C.

6.2.2 Integration with the National ALS Register Platform

The National ALS Register, shown in Figure 6.1, is an online platform designed for healthcare professionals to input detailed information about ALS patients. This platform includes an Electronic Medical Record (EMR) module that allows users to record all information gathered during patient visits. The platform can be accessed at the following URL: `https://revelanos.lais.ufrn.br`.

Figure 6.1: The National ALS Register platform (language: Brazilian Portuguese).

In this work, we expanded the capabilities of this platform by creating the Prognosis module. This module offers physicians prognostic information in a visual dashboard format, including tables and charts. Hence, it serves as a CDS system and was created based on the research findings outlined in Chapters 4 and 5. Figure 6.2 illustrates the general architecture of the Prognostic module proposed in this thesis.

Table 6.1: List of the biomarkers chosen to be collected from the Brazilian patients.

Demographic and Clinical		
Biomarker	**Time of Assessment**	**Data Type**
Date of Birth	1^{st} consultation	date
Date of Symptom Onset		
Date of Diagnosis		
Gender (Male/Female)		categorical
Ethnicity		
Site of Onset		
Revised El Escorial Criteria		
Motor Phenotype		
Family History		
Smoker		
Packs of Cigarettes per Day		numeric
Height (m)		
Weight (kg)	All consultations	numeric
Body Mass Index (BMI)		
Staging of Disease (*King's College*)		categorical
Genetic Mutation Identified	When it occurs	categorical
Frontotemporal Dementia Spectrum		
Date of Death		date
Medication		
Riluzole	When it occurs	categorical
Edaravone		
Functional, Respiratory and Nutritional		
Revised ALS Functional Rate Scale (ALSFRS-R)	All consultations	numeric
Forced Vital Capacity (FVC)		
Slow Vital Capacity (SVC)		
Maximum Inspiratory Pressure (MIP)		
Maximum Expiratory Pressure (MEP)		
Sniff Nasal Inspiratory Pressure (SNIP)		
Non-invasive Ventilation	When it occurs	categorical
Tracheostomy		
Cough Assist		
Gastrostomy		
Laboratory		
Uric Acid	All consultations	numeric
Alanine Aminotransferase (ALT or TGP)		
Albumin		
Aspartate Aminotransferase		
Calcium		
HDL Cholesterol		
LDL Cholesterol		
Total Cholesterol		
Chlorine		
Creatine Kinase (CK)		
Creatinine		
Ferritin		
Alkaline Phosphatase		
Phosphate (Phosphorus)		
Glucose		
Potassium		
Triglycerides		
Vitamin B12		
Urine Summary (pH)		
Hemogram: *- Hemoglobin* *- Hematocrit* *- Leukocytes* *- Segmented Neutrophils* *- Lymphocytes* *- Monocytes*		

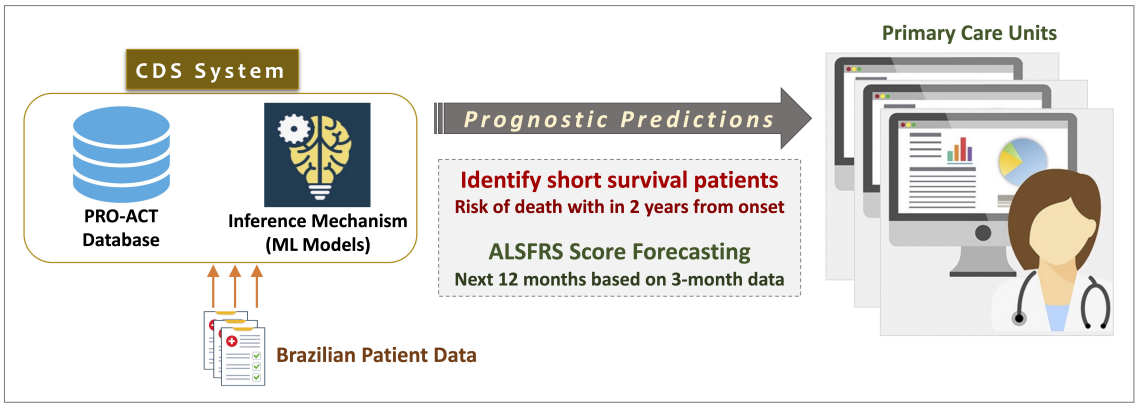


Figure 6.2: General architecture of the proposed Prognostic module.

6.2.3 System Architecture

The National ALS Register platform was developed using PHP and the following additional resources: Laravel, Node.js, React, and the PostgreSQL database system. In contrast, the predictive models developed in this thesis were created using the Python programming language and its packages for data analysis, machine learning, and data visualization (e.g., Pandas, Scikit-Learn, Imbalance-Learning, Tensor Flow, and SHAP). A web service was developed using the Django REST framework to facilitate interoperability between these different technologies. The user interfaces of the Prognostic module was developed using PHP. Figure 6.3 illustrates the architectural relationship involving the National ALS Register platform and the Prognostic web service.

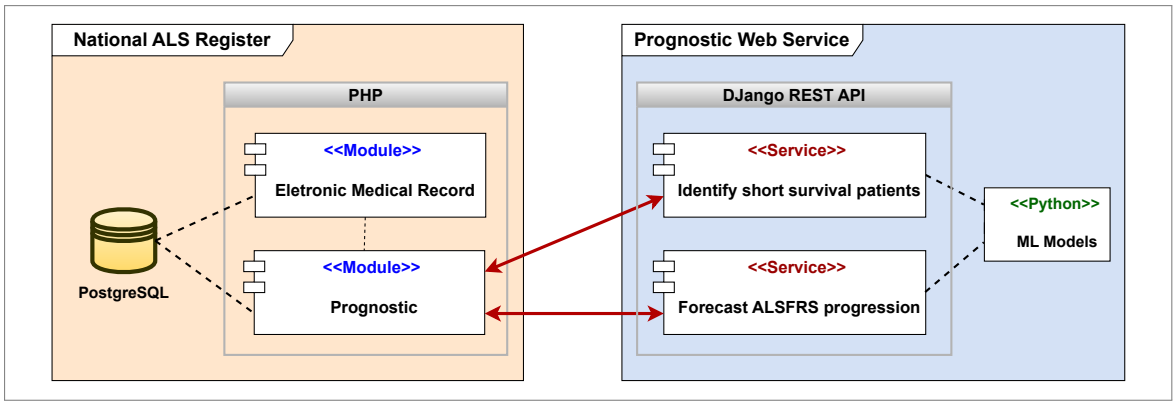


Figure 6.3: Architectural relationship between the National ALS Register platform and the Prognostic web service.

6.3 User Interface

This section presents the user interfaces designed to collect data from ALS patients and provide physicians with predictive information. It also outlines the process required to obtain the prognostic outcomes.

Collecting Patient Data

Figure 6.4 shows the user interface utilized to record patient follow-up, including information on ALSFRS questions, laboratory and respiratory exams.

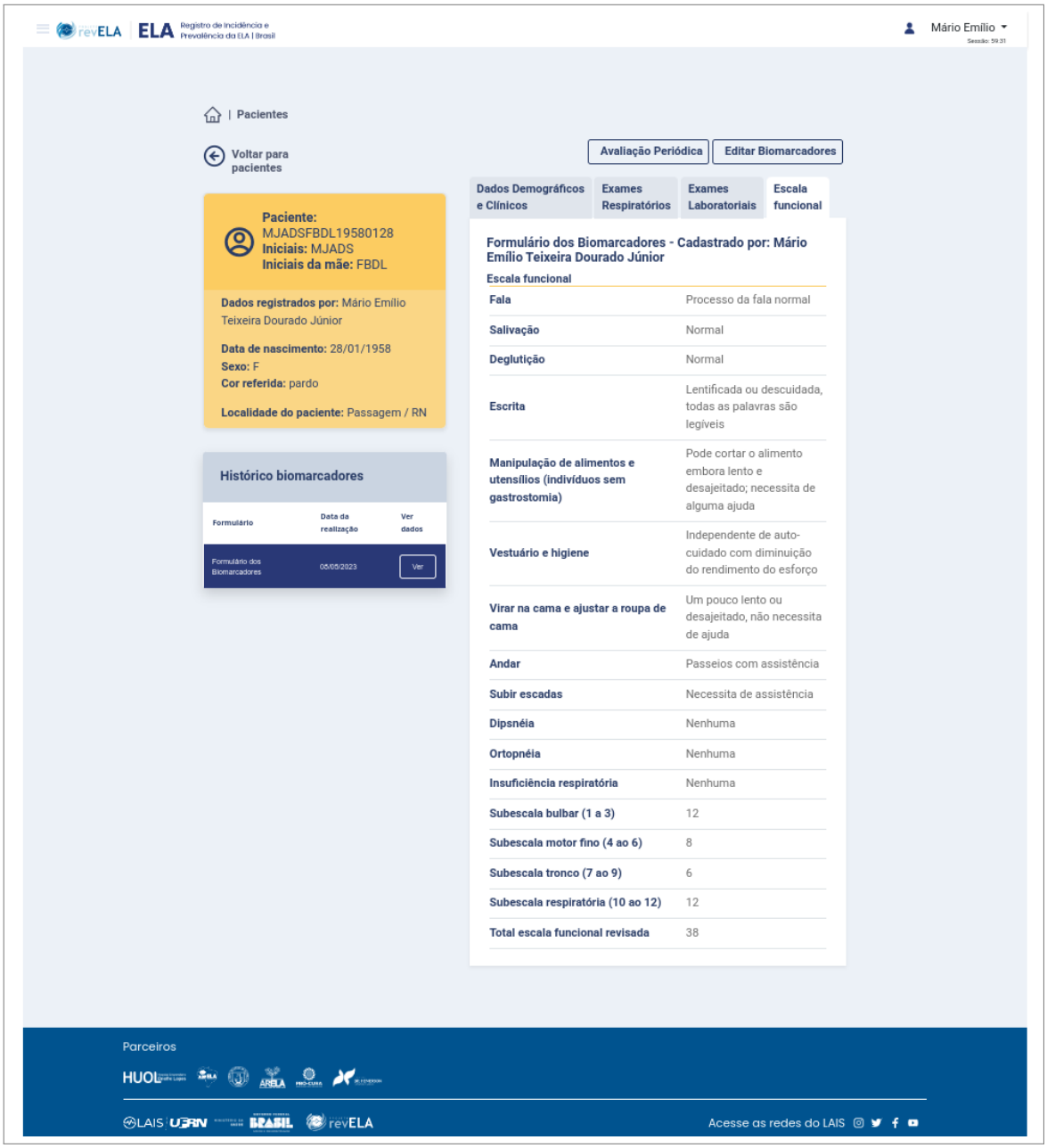


Figure 6.4: User interface for recording patient follow-up. (language: Brazilian Portuguese).

Identifying Short Survival ALS Patients at Diagnosis

The prediction process involves the following steps:

1. The user starts by selecting a specific patient and then requests the prediction, as depicted in Figure 6.5.
2. Next, the user chooses the option to perform the prediction, prompting the system to analyze the patient's current information and categorize them into either a short or non-short survival group.
3. Finally, the system displays the prediction results as shown in Figure 6.6, including the following information:
    - Predicted survival group (Short or Non-short).
    - Classification probability.
    - Chart depicting the influence of each biomarker on the prediction process.

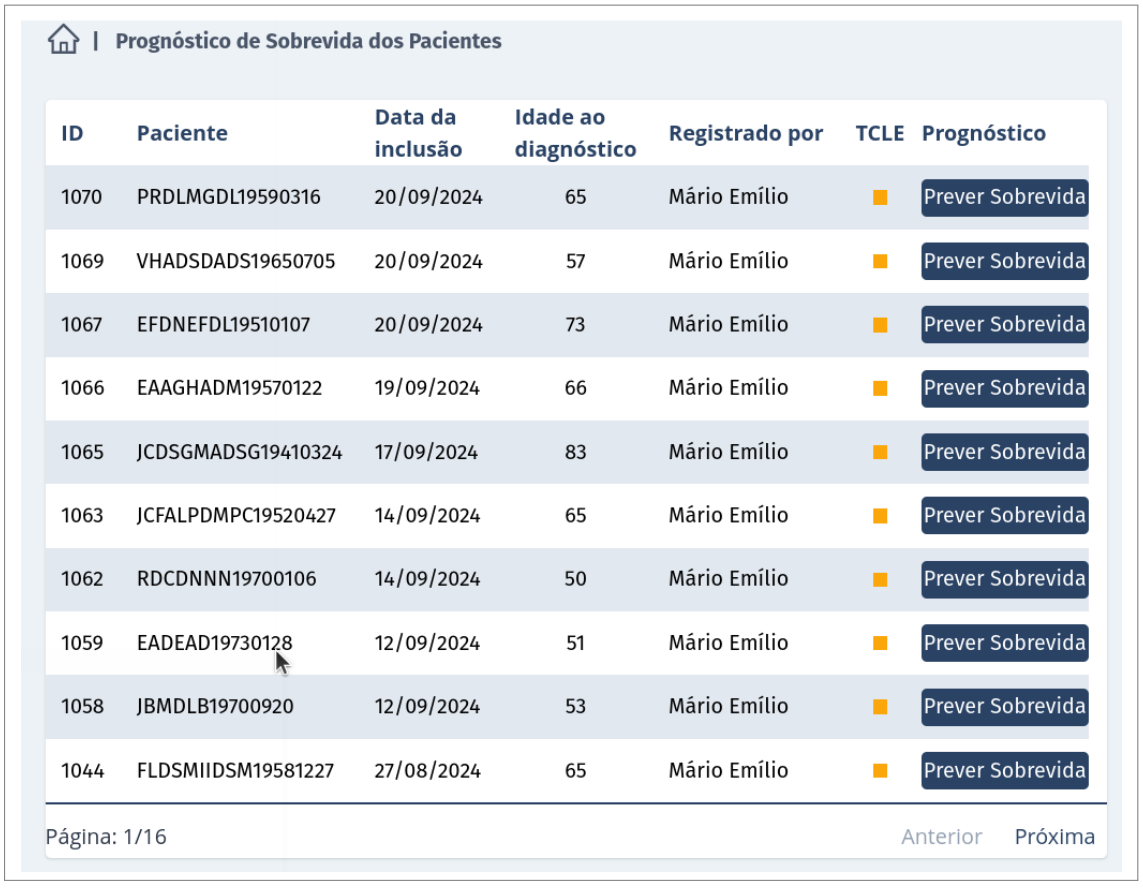


Figure 6.5: User interface used to select a patient and perform survival predictions (language: Brazilian Portuguese).

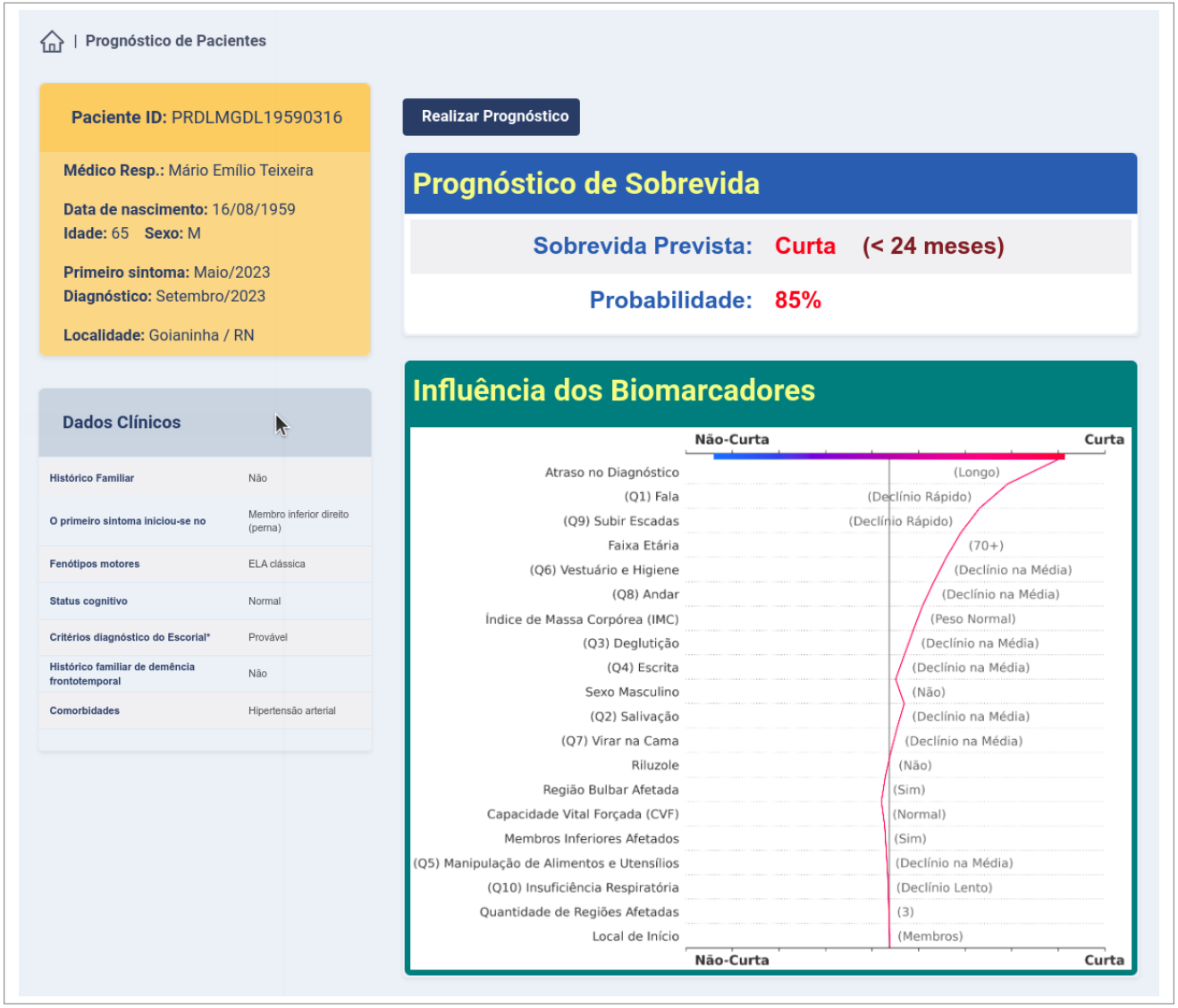


Figure 6.6: User interface used to classify patients into Short or Non-short survival groups (language: Brazilian Portuguese).

Forecasting the ALSFRS Score

The prediction process involves the following steps:

1. The user starts by selecting the specific patient and requesting the target prediction, as illustrated in Figure 6.7.
2. After, the user chooses the option to perform the prediction, prompting the system to utilize the patient's data from the last three months to predict the ALSFRS score for the next twelve months.
3. Finally, the system will display the prediction results through a chart containing information on the ALSFRS scores and slopes for the next 3, 6, 9, and 12 months (see Fig. 6.8). The information comprises the disease duration, ALSFRS score, and ALSFRS slopes from onset and first visit. The shadowed area in the chart denotes the margin of error for the predictions based on the corresponding RMSE for each month ($\pm$ RMSE value).

Figure 6.7: User interface used to select a patient and perform ALSFRS score forecasting (language: Brazilian Portuguese).

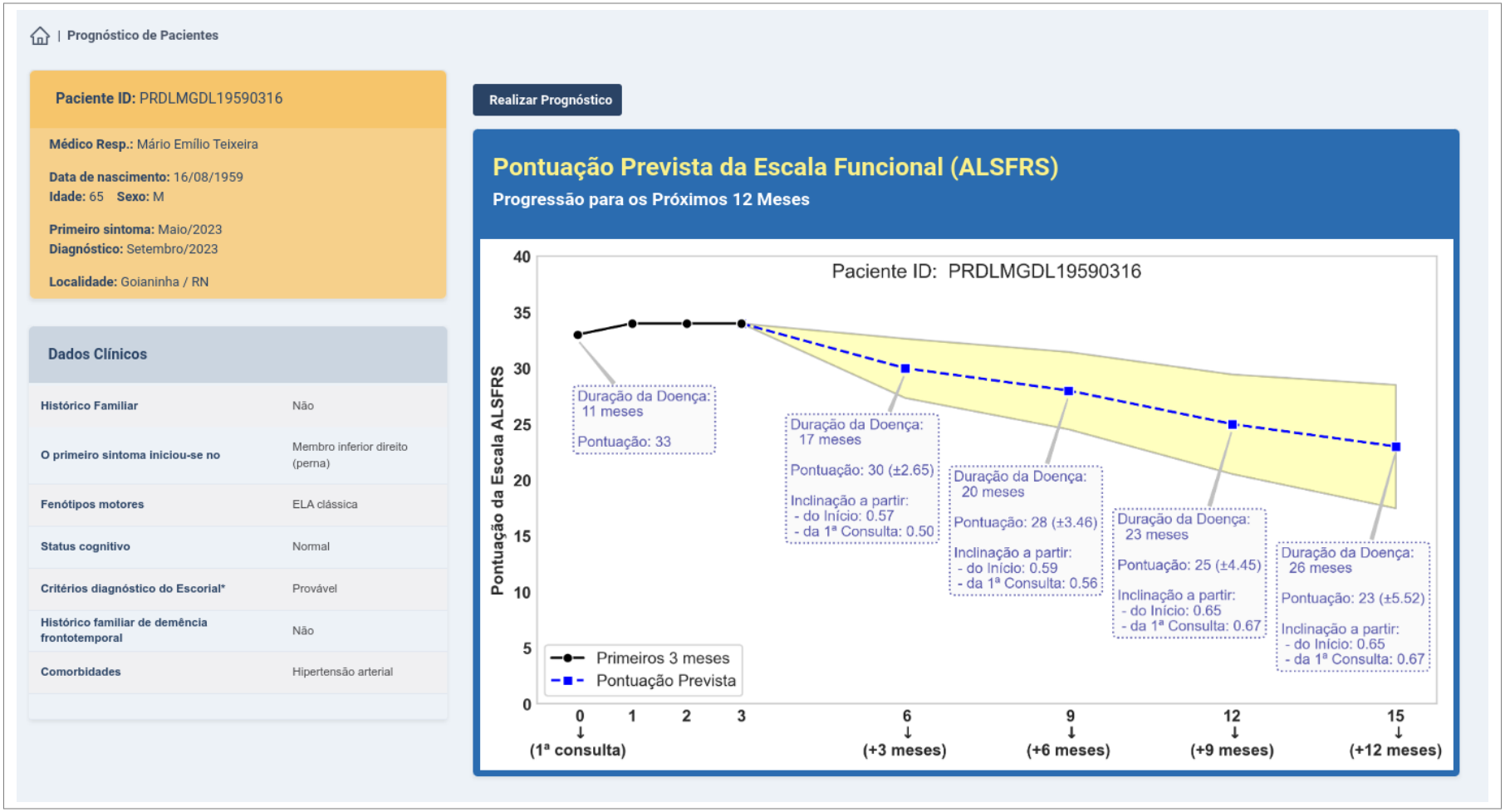


Figure 6.8: User interface for forecasting the ALSFRS score and slope for the next 12 months (language: Brazilian Portuguese). The yellow shaded area indicates the margin of error based on the RMSE for each forecasted month.

6.4 Discussion

In this chapter, we described the development of a CDS system designed to aid Brazilian health workers in their tasks related to ALS patient prognosis. Building on the research outlined in Chapters 4 and 5, the proposed system offers predictions for (i) identifying ALS patients with short survival at the time of diagnosis and (ii) forecasting the ALSFRS score for one year using three-month patient data.

We have established a comprehensive Brazilian ALS Database to support current and future research. This database is intended to facilitate the analysis of data gathered from Brazilian patients, offering a detailed understanding of ALS characteristics in Brazil. It will encompass information on incidence, prevalence, and regions with the highest number of cases. In the future, the results will be compared with those from studies conducted in other countries.

The CDS's inference mechanism is currently trained using data exclusively from the PRO-ACT database. As more patients are registered, the system will be evaluated using information from the population registry. Once we have a significant number of records, the system will be retrained using data exclusively from Brazilian patients. These steps will be crucial in confirming the results obtained in this thesis.

6.5 Conclusion

It is crucial to have access to a care and technological ecosystem to maintain the well-being, quality of life, and survival of people with ALS. Therefore, it is important to conduct research on the epidemiological aspects of the disease, assistive technologies, and preventive care. This includes validating prognostic models to improve the accuracy of predicting and managing ALS. The proposed CDS system provides predictions to assist physicians in making timely and appropriate decisions about their patients, with the goal of improving their quality of life.

Chapter 7

Conclusions

This chapter summarizes the key contributions and provides a thorough overview of the research conducted. Additionally, it explores potential future research directions.

7.1 Overview of the Main Contributions

This thesis presents an in-depth analysis of innovative ML-based solutions to enhance prognostic outcomes for patients with ALS disease. It concentrates on two primary areas of prediction: (i) identifying ALS patients who are likely to have a short survival period at the time of diagnosis and (ii) forecasting the functional decline of patients over time. The proposed solutions employ biomarkers commonly employed in routine ALS clinical practice, which are typically obtained through less expensive and simpler procedures. Consequently, these solutions are feasible for large-scale implementation in primary care, especially considering the financial constraints faced by some countries, such as Brazil.

In Chapter 3, we provided a summary of machine learning approaches for ALS prognosis, with a focus on studies utilizing less complex and more cost-effective features. As a result, we explored various studies to offer an overview of potential solutions that could be implemented to develop decision support systems, making them accessible in a wider range of ALS clinical settings. Additionally, we outlined the different datasets and features analyzed, along with their availability for access, which guided our decision to select the PRO-ACT database as our primary source of patient data. These findings helped define the predictions explored in this thesis and inspired a health research project aimed at collecting biomarkers from Brazilian patients.

In Chapter 4, we developed and evaluated ML algorithms that combined Ensemble and Imbalance Learning techniques to classify patients based on their likelihood of short versus non-short survival during diagnosis. Our goal was to identify individuals at high mortality risk within 24 months of symptom onset using patient data commonly encountered in daily clinical practice. We evaluated our Ensemble-Imbalance approach using six ML algorithms as base classifiers.

Notably, our results outperformed those of individual algorithms, achieving a Balanced Accuracy of 88% and a Sensitivity of 96%. Additionally, we used the Shapley Additive Explanations framework to explain the decision-making process of the top-performing model, identifying the most important features and their correlations with the target prediction. Furthermore, we presented helpful tools for visualizing and comparing patient similarities, providing valuable insights.

In Chapter 5, we introduced a novel approach for predicting functional disability over the next 12 months, month by month, using patient data collected during the first three months. This study is the first to use autoregressive multi-step multi-output time series forecasting for ALS prognosis, specifically to forecast functional decline over time. We extracted both static and temporal features from the PRO-ACT database to develop and evaluate deep learning models utilizing Gated Recurrent Units and Long Short-Term Memory algorithms. Our approach outperformed previous studies while using a significantly smaller set of input features, thereby demonstrating enhanced effectiveness.

In Chapter 6, we described the development of a clinical decision support (CDS) system that provides information on ALS prognosis based on the target predictions discussed in Chapters 4 and 5. This represents an initial step toward integrating the knowledge gained about ALS in this thesis into the clinical environment in Brazil. It is important to emphasize that the proposed CDS system is not designed to replace healthcare professionals but aims to serve as a supplementary resource to assist them in their daily work. Additionally, we have established a comprehensive Brazilian ALS Database to support both current and future research efforts. This database aims to facilitate the analysis of data collected from Brazilian patients.

The promising findings presented in this thesis are essential in helping physicians deliver timely and appropriate information to patients and their families. They also have the potential to enhance the quality of end-of-life care while facilitating effective treatment and resource planning. These findings could assist physicians in formulating personalized treatment plans and managing resources more efficiently. Furthermore, these results may serve as criteria for determining the inclusion or exclusion of patients in clinical trials.

7.2 Future Research Directions

The research presented in this thesis has opened up new avenues for future investigation and has highlighted several important issues to explore:

- Validate the observations and conclusions by evaluating the performance of the models with real-world patient data. This can be achieved by leveraging the Brazilian ALS database developed in this thesis and by collaborating with other research centers that maintain proprietary ALS databases.
- Investigate whether including laboratory tests as input variables will en-

hance the classification of critical patients at diagnosis and improve the forecasting of ALSFRS over time.

- Investigate different adaptations of the Fixing Windowing approach to forecast the ALSFRS for the upcoming 12 months and compare the resulting performance outcomes. For example, consider implementing a fixed time window of 5 or 6 months to predict the ALSFRS for the following year.
- Assess the Expanding Windowing approach to determine whether it improves forecasting accuracy.
- Employ advanced clustering methods to identify groups of patients with correlated clinical features. This approach has the potential to enhance the development of more effective models based on the identified clusters. Examples of such methods include FCAN-MOPSO [98], Patient Similarity Networks [99], and Biclustering [100].
- Investigate explainable frameworks for analyzing the importance of features in predicting the decline of the ALSFRS over time.

References

[1] MedlinePlus. Medlineplus. https://https://medlineplus.gov/genetics/understanding/consult/prognosis/, 2020. Accessed: 2020-09-13.

[2] Andersen, P.M., Abrahams, S., Borasio, G.D., de Carvalho, M., Chio, A., Van Damme, P., Hardiman, O., Kollewe, K., Morrison, K.E., Petri, S., Pradat, P.F., Silani, V., Tomik, B., Wasner, M., Weber, M., and The EFNS Task Force on Diagnosis and Management of Amyotrophic Lateral Sclerosis: . Efns guidelines on the clinical management of amyotrophic lateral sclerosis (mals) – revised report of an efns task force. *European Journal of Neurology*, 19(3):360–375, 2012. doi: 10.1111/j.1468-1331.2011.03501.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-1331.2011.03501.x.

[3] Chiò, A., Pagani, M., Agosta, F., Calvo, A., Cistaro, A., and Filippi, M. Neuroimaging in amyotrophic lateral sclerosis: insights into structural and functional changes. *The Lancet Neurology*, 13(12):1228 – 1240, 2014. ISSN 1474-4422. doi: https://doi.org/10.1016/S1474-4422(14)70167-X. URL http://www.sciencedirect.com/science/article/pii/S147444221470167X.

[4] Swinnen, B. and Robberecht, W. The phenotypic variability of amyotrophic lateral sclerosis. *Nature Reviews Neurology*, 10(11):661–670, 2014. ISSN 1759-4766. doi: 10.1038/nrneurol.2014.184. URL https://doi.org/10.1038/nrneurol.2014.184.

[5] Hardiman, O., Al-Chalabi, A., Chio, A., Corr, E.M., Logroscino, G., Robberecht, W., Shaw, P.J., Simmons, Z., and van den Berg, L.H. Amyotrophic lateral sclerosis. *Nature Reviews Disease Primers*, 3(1):17071, 2017. ISSN 2056-676X. doi: 10.1038/nrdp.2017.71. URL https://doi.org/10.1038/nrdp.2017.71.

[6] Gordon, J. and Lerner, B. Insights into Amyotrophic Lateral Sclerosis from a Machine Learning Perspective. *JOURNAL OF CLINICAL MEDICINE*, 8 (10), October 2019. doi: 10.3390/jcm8101578.

[7] LAIS. Scientific and technological development applied to als. https://www.ufrn.br/imprensa/noticias/31314/pesquisadores-discutem-ela-no-lais, 2020. Accessed: 2020-07-15.

[8] Nowicka, N., Juranek, J., Juranek, J.K., and Wojtkiewicz, J. Risk factors and emerging therapies in amyotrophic lateral sclerosis. *International Journal of Molecular Sciences*, 20(11), 2019. ISSN 1422-0067. doi: 10.3390/ijms20112616. URL https://www.mdpi.com/1422-0067/20/11/2616.

[9] Renton, A.E., Chiò, A., and Traynor, B.J. State of play in amyotrophic lateral sclerosis genetics. *Nature Neuroscience*, 17(1):17–23, January 2014. ISSN 1097-6256, 1546-1726. doi: 10.1038/nn.3584. URL http://www.nature.com/articles/nn.3584.

[10] Al-Chalabi, A. and Hardiman, O. The epidemiology of ALS: a conspiracy of genes, environment and time. *Nature Reviews Neurology*, 9(11):617–628, November 2013. ISSN 1759-4758, 1759-4766. doi: 10.1038/nrneurol.2013.203. URL http://www.nature.com/articles/nrneurol.2013.203.

[11] Miller, R.G., Mitchell, J.D., and Moore, D.H. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *Cochrane Database of Systematic Reviews*, March 2012. ISSN 14651858. doi: 10.1002/14651858.CD001447.pub3. URL http://doi.wiley.com/10.1002/14651858.CD001447.pub3.

[12] Kiernan, M.C., Vucic, S., Cheah, B.C., Turner, M.R., Eisen, A., Hardiman, O., Burrell, J.R., and Zoing, M.C. Amyotrophic lateral sclerosis. *The Lancet*, 377(9769):942–955, March 2011. ISSN 01406736. doi: 10.1016/S0140-6736(10)61156-7. URL https://linkinghub.elsevier.com/retrieve/pii/S0140673610611567.

[13] Group, B.D.W. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, 69(3):89–95, 2001. doi: 10.1067/mcp.2001.113989. URL https://ascpt.onlinelibrary.wiley.com/doi/abs/10.1067/mcp.2001.113989.

[14] Brooks, B.R., Miller, R.G., Swash, M., and Munsat, T.L. El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, 1(5):293–299, January 2000. ISSN 1466-0822. doi: 10.1080/146608200300079536. URL http://www.tandfonline.com/doi/full/10.1080/146608200300079536.

[15] Belsh, J.M. ALS diagnostic criteria of El Escorial Revisited: do they meet the needs of clinicians as well as researchers? *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, 1(sup1):S57–S60, March 2000. ISSN 1466-0822. doi: 10.1080/14660820052415925. URL https://www.tandfonline.com/doi/full/10.1080/14660820052415925.

[16] Chiò, A., Logroscino, G., Hardiman, O., Swingler, R., Mitchell, D., Beghi, E., Traynor, B.G., and On Behalf of the Eurals Consortium. Prognostic factors in ALS: A critical review. *Amyotrophic Lateral Sclerosis*, 10(5-6):310–323, January 2009. ISSN 1748-2968, 1471-180X. doi: 10.3109/17482960802566824. URL http://www.tandfonline.com/doi/full/10.3109/17482960802566824.

[17] Del Aguila, M., Longstreth, W., McGuire, V., Koepsell, T., and Van Belle, G. Prognosis in amyotrophic lateral sclerosis: A population-based study. 60(5):813–819, 2003. ISSN 0028-3878, 1526-632X. doi: 10.1212/01.WNL.0000049472.47709.3B. URL https://www.neurology.org/doi/10.1212/01.WNL.0000049472.47709.3B.

[18] Neurology, J. The Amyotrophic Lateral Sclerosis Functional Rating Scale: Assessment of Activities of Daily Living in Patients With Amyotrophic Lateral Sclerosis. *Archives of Neurology*, 53(2):141–147, 02 1996. ISSN 0003-9942. doi: 10.1001/archneur.1996.00550020045014. URL https://doi.org/10.1001/archneur.1996.00550020045014.

[19] Cedarbaum, J.M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., and Nakanishi, A. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences*, 169(1-2):13–21, October 1999. ISSN 0022510X. doi: 10.1016/S0022-510X(99)00210-5. URL https://linkinghub.elsevier.com/retrieve/pii/S0022510X99002105.

[20] Elamin, M., Bede, P., Montuschi, A., Pender, N., Chio, A., and Hardiman, O. Predicting prognosis in amyotrophic lateral sclerosis: a simple algorithm. *Journal of Neurology*, 262(6):1447–1454, June 2015. ISSN 0340-5354, 1432-1459. doi: 10.1007/s00415-015-7731-6. URL http://link.springer.com/10.1007/s00415-015-7731-6.

[21] Proudfoot, M., Jones, A., Talbot, K., Al-Chalabi, A., and Turner, M.R. The ALSFRS as an outcome measure in therapeutic trials and its relationship to symptom onset. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*, 17(5-6):414–425, August 2016. ISSN 2167-9223. doi: 10.3109/21678421.2016.1140786.

[22] Guedes, K., Pereira, C., Pavan, K., and Valério, B.C.O. Cross-cultural adaptation and validation of als Functional Rating Scale-Revised in Portuguese language. *Arquivos de Neuro-Psiquiatria*, 68:44 – 47, 02 2010. ISSN 0004-282X. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0004-282X2010000100010&nrm=iso.

[23] Kollewe, K., Mauss, U., Krampfl, K., Petri, S., Dengler, R., and Mohammadi, B. ALSFRS-R score and its ratio: A useful predictor for ALS-progression. *Journal of the Neurological Sciences*, 275(1-2):69–73, December 2008. ISSN 0022510X. doi: 10.1016/j.jns.2008.07.016. URL https://linkinghub.elsevier.com/retrieve/pii/S0022510X08003663.

[24] Rutkove, S.B. Clinical Measures of Disease Progression in Amyotrophic Lateral Sclerosis. *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics*, 12(2):384–393, April 2015. ISSN 1878-7479. doi: 10.1007/s13311-014-0331-9.

[25] Chipika, R.H., Finegan, E., Li Hi Shing, S., Hardiman, O., and Bede, P. Tracking a Fast-Moving Disease: Longitudinal Markers, Monitoring, and Clinical Trial Endpoints in ALS. *Frontiers in Neurology*, 10:229, March

2019. ISSN 1664-2295. doi: 10.3389/fneur.2019.00229. URL https://www.frontiersin.org/article/10.3389/fneur.2019.00229/full.

[26] Al-Chalabi, A., Hardiman, O., Kiernan, M.C., Chiò, A., Rix-Brooks, B., and van den Berg, L.H. Amyotrophic lateral sclerosis: moving towards a new classification system. *The Lancet Neurology*, 15(11):1182–1194, October 2016. ISSN 14744422. doi: 10.1016/S1474-4422(16)30199-5. URL https://linkinghub.elsevier.com/retrieve/pii/S1474442216301995.

[27] Roche, J.C., Rojas-Garcia, R., Scott, K.M., Scotton, W., Ellis, C.E., Burman, R., Wijesekera, L., Turner, M.R., Leigh, P.N., Shaw, C.E., and Al-Chalabi, A. A proposed staging system for amyotrophic lateral sclerosis. *Brain: A Journal of Neurology*, 135(Pt 3):847–852, March 2012. ISSN 1460-2156. doi: 10.1093/brain/awr351.

[28] Chiò, A., Hammond, E.R., Mora, G., Bonito, V., and Filippini, G. Development and evaluation of a clinical staging system for amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 86(1):38–44, January 2015. ISSN 0022-3050, 1468-330X. doi: 10.1136/jnnp-2013-306589. URL http://jnnp.bmj.com/lookup/doi/10.1136/jnnp-2013-306589.

[29] Fang, T., Al Khleifat, A., Stahl, D.R., Lazo La Torre, C., Murphy, C., Uk-Mnd LicalS, Young, C., Shaw, P.J., Leigh, P.N., and Al-Chalabi, A. Comparison of the King's and MiToS staging systems for ALS. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*, 18(3-4):227–232, 2017. ISSN 2167-9223. doi: 10.1080/21678421.2016.1265565.

[30] Hothorn, T. and Jung, H.H. RandomForest4Life: A Random Forest for predicting ALS disease progression. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 15(5-6):444–452, September 2014. ISSN 2167-8421, 2167-9223. doi: 10.3109/21678421.2014.893361. URL http://www.tandfonline.com/doi/full/10.3109/21678421.2014.893361.

[31] Morgan, R.K., McNally, S., Alexander, M., Conroy, R., Hardiman, O., and Costello, R.W. Use of Sniff nasal-inspiratory force to predict survival in amyotrophic lateral sclerosis. *American Journal of Respiratory and Critical Care Medicine*, 171(3):269–274, February 2005. ISSN 1073-449X. doi: 10.1164/rccm.200403-314OC.

[32] van der Burgh, H.K., Schmidt, R., Westeneng, H.J., de Reus, M.A., van den Berg, L.H., and van den Heuvel, M.P. Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis. *NeuroImage: Clinical*, 13:361–369, 2017. ISSN 22131582. doi: 10.1016/j.nicl.2016.10.008. URL https://linkinghub.elsevier.com/retrieve/pii/S2213158216301899.

[33] Schuster, C., Hardiman, O., and Bede, P. Survival prediction in Amyotrophic lateral sclerosis based on MRI measures and clinical characteristics. *BMC Neurology*, 17(1):73, December 2017. ISSN 1471-2377. doi: 10.1186/s12883-017-0854-x. URL http://bmcneurol.biomedcentral.com/articles/10.1186/s12883-017-0854-x.

[34] Grolez, G., Kyheng, M., Lopes, R., Moreau, C., Timmerman, K., Auger, F., Kuchcinski, G., Duhamel, A., Jissendi-Tchofo, P., Besson, P., Laloux, C., Petrault, M., Devedjian, J.C., Pérez, T., Pradat, P.F., Defebvre, L., Bordet, R., Danel-Brunaud, V., and Devos, D. MRI of the cervical spinal cord predicts respiratory dysfunction in ALS. *Scientific Reports*, 8(1):1828, December 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-19938-2. URL http://www.nature.com/articles/s41598-018-19938-2.

[35] Blasco, H., Corcia, P., Moreau, C., Veau, S., Fournier, C., Vourc'h, P., Emond, P., Gordon, P., Pradat, P.F., Praline, J., Devos, D., Nadal-Desbarats, L., and Andres, C.R. 1h-nmr-based metabolomic profiling of csf in early amyotrophic lateral sclerosis. *PloS one*, 5(10):e13223, Oct 2010. ISSN 1932-6203. URL https://www.ncbi.nlm.nih.gov/pubmed/20949041.

[36] Varghese, A.M., Sharma, A., Mishra, P., Vijayalakshmi, K., Harsha, H.C., Sathyaprabha, T.N., Bharath, S.M., Nalini, A., Alladi, P.A., and Raju, T.R. Chitotriosidase - a putative biomarker for sporadic amyotrophic lateral sclerosis. *Clinical proteomics*, 10(1):19–19, Dec 2013. ISSN 1542-6416. doi: 10.1186/1559-0275-10-19. URL https://www.ncbi.nlm.nih.gov/pubmed/24295388.

[37] Rossi, D., Volanti, P., Brambilla, L., Colletti, T., Spataro, R., and La Bella, V. Csf neurofilament proteins as diagnostic and prognostic biomarkers for amyotrophic lateral sclerosis. *Journal of Neurology*, 265(3):510–521, 2018. ISSN 1432-1459. URL https://doi.org/10.1007/s00415-017-8730-6.

[38] Waller, R., Goodall, E.F., Milo, M., Cooper-Knock, J., Costa, M.D., Hobson, E., Kazoka, M., Wollff, H., Heath, P.R., Shaw, P.J., and Kirby, J. Serum mirnas mir-206, 143-3p and 374b-5p as potential biomarkers for amyotrophic lateral sclerosis (als). *Neurobiology of Aging*, 55:123 – 131, 2017. ISSN 0197-4580. doi: https://doi.org/10.1016/j.neurobiolaging.2017.03.027. URL http://www.sciencedirect.com/science/article/pii/S0197458017301033.

[39] Samuel, A.L. Some Studies in Machine Learning Using the Game of Checkers. I. In Levy, D.N.L., editor, *Computer Games I*, pages 335–365. Springer New York, New York, NY, 1988. ISBN 978-1-4613-8718-3 978-1-4613-8716-9. doi: 10.1007/978-1-4613-8716-9_14. URL http://link.springer.com/10.1007/978-1-4613-8716-9_14.

[40] Kubat, M. *An Introduction to Machine Learning*. Springer International Publishing, Cham, 2017. ISBN 978-3-319-63912-3 978-3-319-63913-0. doi: 10.1007/978-3-319-63913-0. URL http://link.springer.com/10.1007/978-3-319-63913-0.

[41] Mitchell, T.M. *Machine Learning*. McGraw-Hill, Inc., USA, 1 edition, 1997. ISBN 0070428077.

[42] Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'95, page 1137–1143, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc. ISBN 1558603638.

[43] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830, 2011.

[44] Chawla, N.V. Data Mining for Imbalanced Datasets: An Overview. In Maimon, O. and Rokach, L., editors, *Data Mining and Knowledge Discovery Handbook*, pages 853–867. Springer-Verlag, New York, 2005. ISBN 978-0-387-24435-8. doi: 10.1007/0-387-25465-X_40. URL `http://link.springer.com/10.1007/0-387-25465-X_40`.

[45] Fernández, A., Galar, M., García, S., Herrera, F., Krawczyk, B., and Prati, R.C. *Learning from Imbalanced Data Sets*. Springer International Publishing : Imprint: Springer, Cham, 1st ed. 2018 edition, 2018. ISBN 978-3-319-98074-4. doi: 10.1007/978-3-319-98074-4.

[46] Ng, A.Y. Feature selection, l1 vs. l2 regularization, and rotational invariance. In *Proceedings of the Twenty-First International Conference on Machine Learning*, ICML '04, page 78, New York, NY, USA, 2004. Association for Computing Machinery. ISBN 1581138385. doi: 10.1145/1015330.1015435. URL `https://doi.org/10.1145/1015330.1015435`.

[47] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, jan 2014. ISSN 1532-4435.

[48] Prechelt, L. *Early Stopping - But When?*, pages 55–69. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998. ISBN 978-3-540-49430-0. doi: 10.1007/3-540-49430-8_3. URL `https://doi.org/10.1007/3-540-49430-8_3`.

[49] Nyandwi, J.d.D. Early stopping explained! 2024. Accessed: 2024-09-01.

[50] LaValle, S.M., Branicky, M.S., and Lindemann, S.R. On the relationship between classical grid search and probabilistic roadmaps. *The International Journal of Robotics Research*, 23(7-8):673–692, 2004.

[51] Ji, Y., Hao, J., Reyhani, N., and Lendasse, A. Direct and recursive prediction of time series using mutual information selection. volume 3512, pages 1010–1017, 06 2005. ISBN 978-3-540-26208-4. doi: 10.1007/11494669_124.

[52] Box, G.E., Jenkins, G.M., Reinsel, G.C., and Ljung, G.M. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[53] Brockwell, P.J. and Davis, R.A. *Introduction to time series and forecasting*. Springer, 2002.

[54] Box, G. Box and jenkins: time series analysis, forecasting and control. In *A Very British Affair: Six Britons and the Development of Time Series Analysis During the 20th Century*, pages 161–215. Springer, 2013.

[55] Berner, E.S., Hannah, K.J., and Ball, M.J., editors. *Clinical Decision Support Systems*. Health Informatics. Springer New York, New York, NY, 2007. ISBN 978-0-387-33914-6 978-0-387-38319-4. doi: 10.1007/978-0-387-38319-4. URL http://link.springer.com/10.1007/978-0-387-38319-4.

[56] Beeler, P.E., Bates, D.W., and Hug, B.L. Clinical decision support systems. 2014. doi: 10.5167/UZH-122774. URL https://www.zora.uzh.ch/id/eprint/122774. Publisher: EMH Swiss Medical Publishers.

[57] Winters-Miner, L.A., Bolding, P.S., Hilbe, J.M., Goldstein, M., Hill, T., Nisbet, R., Walton, N., and Miner, G.D. *Practical Predictive Analytics and Decisioning Systems for Medicine*. Elsevier, 2015. ISBN 978-0-12-411643-6. doi: 10.1016/C2012-0-02184-9. URL https://linkinghub.elsevier.com/retrieve/pii/C20120021849.

[58] Shortliffe, E.H. and Sepúlveda, M.J. Clinical Decision Support in the Era of Artificial Intelligence. *JAMA*, 320(21):2199, December 2018. ISSN 0098-7484. doi: 10.1001/jama.2018.17163. URL http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2018.17163.

[59] Agache, I. and Hellings, P. *Implementing Precision Medicine in Best Practices of Chronic Airway Diseases*. Elsevier, 2019. ISBN 978-0-12-813471-9. doi: 10.1016/C2016-0-05080-5. URL https://linkinghub.elsevier.com/retrieve/pii/C20160050805.

[60] Ouzzani, M., Hammady, H., Fedorowicz, Z., and Elmagarmid, A. Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1):210, 2016. ISSN 2046-4053. doi: 10.1186/s13643-016-0384-4.

[61] Atassi, N., Berry, J., Shui, A., Zach, N., Sherman, A., Sinani, E., Walker, J., Katsovskiy, I., Schoenfeld, D., Cudkowicz, M., and Leitner, M. The proact database: design, initial analyses, and predictive features. *Neurology*, 83(19):1719–1725, November 2014. ISSN 1526-632X. doi: 10.1212/WNL.0000000000000951.

[62] van der Burgh, H.K., Schmidt, R., Westeneng, H.J., de Reus, M.A., van den Berg, L.H., and van den Heuvel, M.P. Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis. *NEUROIMAGE-CLINICAL*, 13:361–369, 2017. ISSN 2213-1582. doi: 10.1016/j.nicl.2016.10.008.

[63] D. Halbersberg and B. Lerner. Temporal Modeling of Deterioration Patterns and Clustering for Disease Prediction of ALS Patients. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 62–68, December 2019. doi: 10.1109/ICMLA.2019.00019. Journal Abbreviation: 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA).

[64] Kueffner, R., Zach, N., Bronfeld, M., Norel, R., Atassi, N., Balagurusamy, V., Di Camillo, B., Chio, A., Cudkowicz, M., Dillenberger, D., Garcia-Garcia, J., Hardiman, O., Hoff, B., Knight, J., Leitner, M.L., Li, G., Mangravite, L., Norman, T., Wang, L., Alkallas, R., Anghel, C., Avril, J., Bacardit, J.,

Balser, B., Balser, J., Bar-Sinai, Y., Ben-David, N., Ben-Zion, E., Bliss, R., Cai, J., Chernyshev, A., Chiang, J.H., Chicco, D., Corriveau, B.A.N., Dai, J., Deshpande, Y., Desplats, E., Durgin, J.S., Espiritu, S.M.G., Fan, F., Fevrier, P., Fridley, B.L., Godzik, A., Golinska, A., Gordon, J., Graw, S., Guo, Y., Herpelinck, T., Hopkins, J., Huang, B., Jacobsen, J., Jahandideh, S., Jeon, J., Ji, W., Jung, K., Karanevich, A., Koestler, D.C., Kozak, M., Kurz, C., Lalansingh, C., Larrieu, T., Lazzarini, N., Lerner, B., Lesinski, W., Liang, X., Lin, X., Lowe, J., Mackey, L., Meier, R., Min, W., Mnich, K., Nahmias, V., Noel-MacDonnell, J., O'Donnell, A., Paadre, S., Park, J., Polewko-Klim, A., Raghavan, R., Rudnicki, W., Saghapour, E., Salomond, J.B., Sankaran, K., Sendorek, D., Sharan, V., Shiah, Y.J., Sirois, J.K., Sumanaweera, D.N., Usset, J., Vang, Y.S., Vens, C., Wadden, D., Wang, D., Wong, W.C., Xie, X., Xu, Z., Yang, H.T., Yu, X., Zhang, H., Zhang, L., Zhang, S., Zhu, S., Xiao, J., Fang, W.C., Peng, J., Yang, C., Chang, H.J., Stolovitzky, G., and Consortium, A.S. Stratification of amyotrophic lateral sclerosis patients: a crowdsourcing approach. *SCIENTIFIC REPORTS*, 9, January 2019. ISSN 2045-2322. doi: 10.1038/s41598-018-36873-4.

[65] Tang, M., Gao, C., Goutman, S.A., Kalinin, A., Mukherjee, B., Guan, Y., and Dinov, I.D. Model-Based and Model-Free Techniques for Amyotrophic Lateral Sclerosis Diagnostic Prediction and Patient Clustering. *NEUROINFORMATICS*, 17(3):407–421, July 2019. ISSN 1539-2791. doi: 10.1007/s12021-018-9406-9.

[66] B. Hadad and B. Lerner. Domain adaptation from clinical trials data to the tertiary care clinic – Application to ALS. In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 539–544, December 2020. doi: 10.1109/ICMLA51294.2020.00090. Journal Abbreviation: 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA).

[67] Grollemund, V., Chat, G.L., Secchi-Buhour, M.S., Delbot, F., Pradat-Peyre, J.F., Bede, P., and Pradat, P.F. Development and validation of a 1-year survival prognosis estimation model for Amyotrophic Lateral Sclerosis using manifold learning algorithm UMAP. *Scientific Reports*, 10(1):13378, December 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-70125-8. URL `http://www.nature.com/articles/s41598-020-70125-8`.

[68] S. Pires, M. Gromicho, S. Pinto, M. Carvalho, and S. C. Madeira. Predicting Non-invasive Ventilation in ALS Patients Using Stratified Disease Progression Groups. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 748–757, November 2018. ISBN 2375-9259. doi: 10.1109/ICDMW.2018.00113. Journal Abbreviation: 2018 IEEE International Conference on Data Mining Workshops (ICDMW).

[69] Leão, T., Madeira, S.C., Gromicho, M., de Carvalho, M., and Carvalho, A.M. Learning dynamic Bayesian networks from time-dependent and time-independent data: Unraveling disease progression in Amyotrophic Lateral Sclerosis. *Journal of Biomedical Informatics*, 117:103730, May 2021. ISSN 1532-

0464. doi: 10.1016/j.jbi.2021.103730. URL https://www.sciencedirect.com/science/article/pii/S1532046421000599.

[70] Greco, A., Chiesa, M.R., Da Prato, I., Romanelli, A.M., Dolciotti, C., Cavallini, G., Masciandaro, S.M., Scilingo, E.P., Del Carratore, R., and Bongioanni, P. Using blood data for the differential diagnosis and prognosis of motor neuron diseases: a new dataset for machine learning applications. *Scientific reports*, 11(1):3371, February 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-82940-8.

[71] Chio, A., Canosa, A., Gallo, S., Cammarosano, S., Moglia, C., Fuda, G., Calvo, A., Gabriele, M., and For the PARALS group. Als clinical trials: Do enrolled patients accurately represent the als population? *Neurology*, 77(15):1432–1437, October 2011. ISSN 0028-3878, 1526-632X. doi: 10.1212/WNL.0b013e318232ab9b. URL http://www.neurology.org/cgi/doi/10.1212/WNL.0b013e318232ab9b.

[72] Matthews, J.N., Altman, D.G., Campbell, M.J., and Royston, P. Analysis of serial measurements in medical research. 300(6719):230–235, January 1990. doi: 10.1136/bmj.300.6719.230. URL https://doi.org/10.1136/bmj.300.6719.230.

[73] Lundberg, S.M. and Lee, S.I. A unified approach to interpreting model predictions. In Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

[74] Ribeiro, M.T., Singh, S., and Guestrin, C. "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 1135–1144, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322. doi: 10.1145/2939672.2939778. URL https://doi.org/10.1145/2939672.2939778.

[75] Adadi, A. and Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6:52138–52160, 2018. ISSN 2169-3536. doi: 10.1109/ACCESS.2018.2870052. URL https://ieeexplore.ieee.org/document/8466590/.

[76] Rokach, L. Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1):1–39, Feb 2010. ISSN 1573-7462. doi: 10.1007/s10462-009-9124-7. URL https://doi.org/10.1007/s10462-009-9124-7.

[77] Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, February 2019. ISSN 00043702. doi: 10.1016/j.artint.2018.07.007. URL https://linkinghub.elsevier.com/retrieve/pii/S0004370218305988.

[78] Lundberg, S.M. and Lee, S.I. A unified approach to interpreting model predictions. In Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017.

[79] Team, T.P.D. pandas-dev/pandas: Pandas, April 2022. URL `https://zenodo.org/record/3509134`.

[80] Lemaître, G., Nogueira, F., and Aridas, C.K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5, 2017. URL `http://jmlr.org/papers/v18/16-365.html`.

[81] Hunter, J.D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

[82] Waskom, M.L. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021. doi: 10.21105/joss.03021. URL `https://doi.org/10.21105/joss.03021`.

[83] Rodríguez-Pérez, R. and Bajorath, J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*, 34(10):1013–1026, October 2020. ISSN 0920-654X, 1573-4951. doi: 10.1007/s10822-020-00314-0. URL `https://link.springer.com/10.1007/s10822-020-00314-0`.

[84] Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T.E. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL `https://doi.org/10.1038/s41586-020-2649-2`.

[85] Papaiz, F., Dourado, M.E.T., Valentim, R.A.d.M., de Morais, A.H.F., and Arrais, J.P. Machine Learning Solutions Applied to Amyotrophic Lateral Sclerosis Prognosis: A Review. *Frontiers in Computer Science*, 4, April 2022. ISSN 2624-9898. doi: 10.3389/fcomp.2022.869140.

[86] Fernández, A., Galar, M., García, S., Herrera, F., Krawczyk, B., and Prati, R.C. *Learning from Imbalanced Data Sets*. Springer International Publishing : Imprint: Springer, Cham, 1st ed. 2018 edition, 2018. ISBN 978-3-319-98074-4. doi: 10.1007/978-3-319-98074-4.

[87] Al-Chalabi, A., Hardiman, O., Kiernan, M.C., Chiò, A., Rix-Brooks, B., and Van Den Berg, L.H. Amyotrophic lateral sclerosis: moving towards a new classification system. *The Lancet Neurology*, 15(11):1182–1194, October 2016. ISSN 14744422. doi: 10.1016/S1474-4422(16)30199-5. URL `https://linkinghub.elsevier.com/retrieve/pii/S1474442216301995`.

[88] Manera, U., Calvo, A., Daviddi, M., Canosa, A., Vasta, R., Torrieri, M.C., Grassano, M., Brunetti, M., D'Alfonso, S., Corrado, L., Marchi, F.D., Moglia, C., D'Ovidio, F., Mora, G., Mazzini, L., and Chiò, A. Regional spreading of symptoms at diagnosis as a prognostic marker in amyotrophic lateral sclerosis: a population-based study. *Journal of Neurology, Neurosurgery & Psychiatry*, 91(3):291–297, 2020. ISSN 0022-3050. doi: 10.1136/jnnp-2019-321153. URL `https://jnnp.bmj.com/content/91/3/291`.

[89] Taylor, A.A., Fournier, C., Polak, M., Wang, L., Zach, N., Keymer, M., Glass, J.D., Ennist, D.L., and The Pooled Resource Open-Access ALS Clinical Trials Consortium. Predicting disease progression in amyotrophic lateral sclerosis. *Annals of Clinical and Translational Neurology*, 3(11):866–875, November 2016. ISSN 23289503. doi: 10.1002/acn3.348. URL `http://doi.wiley.com/10.1002/acn3.348`.

[90] Pires, S., Gromicho, M., Pinto, S., De Carvalho, M., and Madeira, S.C. Patient stratification using clinical and patient profiles: Targeting personalized prognostic prediction in ALS. In Rojas, I., Valenzuela, O., Rojas, F., Herrera, L.J., and Ortuño, F., editors, *Bioinformatics and Biomedical Engineering*, volume 12108, pages 529–541. Springer International Publishing, 2020. ISBN 978-3-030-45384-8 978-3-030-45385-5. doi: 10.1007/978-3-030-45385-5_47. URL `http://link.springer.com/10.1007/978-3-030-45385-5_47`. Series Title: Lecture Notes in Computer Science.

[91] Pancotti, C., Birolo, G., Rollo, C., Sanavia, T., Di Camillo, B., Manera, U., Chiò, A., and Fariselli, P. Deep learning methods to predict amyotrophic lateral sclerosis disease progression. *Scientific Reports*, 12(1):13738, 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-17805-9.

[92] Turabieh, H., Afshar, A.S., Statland, J., Song, X., and Pooled Resource Open-Access ALS Clinical Trials Consortium. Towards a machine learning empowered prognostic model for predicting disease progression for amyotrophic lateral sclerosis. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2023:718–725, 2023. ISSN 1942-597X.

[93] Tavazzi, E., Longato, E., Vettoretti, M., Aidos, H., Trescato, I., Roversi, C., Martins, A.S., Castanho, E.N., Branco, R., Soares, D.F., Guazzo, A., Birolo, G., Pala, D., Bosoni, P., Chiò, A., Manera, U., De Carvalho, M., Miranda, B., Gromicho, M., Alves, I., Bellazzi, R., Dagliati, A., Fariselli, P., and Madeira, B., Sara C. a nd Di Camillo. Artificial intelligence and statistical methods for stratification and prediction of progression in amyotrophic lateral sclerosis: A systematic review. *Artificial Intelligence in Medicine*, 142: 102588, 2023. ISSN 09333657. doi: 10.1016/j.artmed.2023.102588. URL `https://linkinghub.elsevier.com/retrieve/pii/S0933365723001021`.

[94] Nahon, A. and Lerner, B. Temporal modeling of als using longitudinal data and long-short term memory-based algorithm. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 25–27, Bruges, Belgium, 2018. ESANN 2018 proceedings. ISBN 9782875870476. URL `mhttp://www.i6doc.com/en/`.

[95] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

[96] Kingma, D.P. and Ba, J. Adam: A method for stochastic optimization, 2017.

[97] Beeler, P.E., Bates, D.W., and Hug, B.L. Clinical decision support systems. *Swiss Medical Weekly*, 144:w14073, December 2014. ISSN 0036-7672. doi: 10.4414/smw.2014.14073. URL `https://doi.org/10.5167/uzh-122774`.

[98] Hu, L., Yang, Y., Tang, Z., He, Y., and Luo, X. Fcan-mopso: An improved fuzzy-based graph clustering algorithm for complex networks with multiobjective particle swarm optimization. *IEEE Transactions on Fuzzy Systems*, 31(10):3470–3484, 2023. doi: 10.1109/TFUZZ.2023.3259726.

[99] Pai, S. and Bader, G.D. Patient similarity networks for precision medicine. *Journal of Molecular Biology*, 430(18, Part A):2924–2938, 2018. ISSN 0022-2836. doi: https://doi.org/10.1016/j.jmb.2018.05.037. URL `https://www.sciencedirect.com/science/article/pii/S0022283618305321`. Theory and Application of Network Biology Toward Precision Medicine.

[100] Matos, J., Pires, S., Aidos, H., Gromicho, M., Pinto, S., de Carvalho, M., and Madeira, S.C. Unravelling disease presentation patterns in als using biclustering for discriminative meta-features discovery. In Rojas, I., Valenzuela, O., Rojas, F., Herrera, L.J., and Ortuno, F., editors, *Bioinformatics and Biomedical Engineering*, pages 517–528, Cham, 2020. Springer International Publishing. ISBN 978-3-030-45385-5.

Appendices

Appendix A

Supplementary Information: Identifying Short-survival ALS Patients at Diagnosis

The following pages include the additional information related with Chapter 4.

Supplementary Information

Ensemble-Imbalance-based classification for amyotrophic lateral sclerosis prognostic prediction: identifying short-survival patients at diagnosis

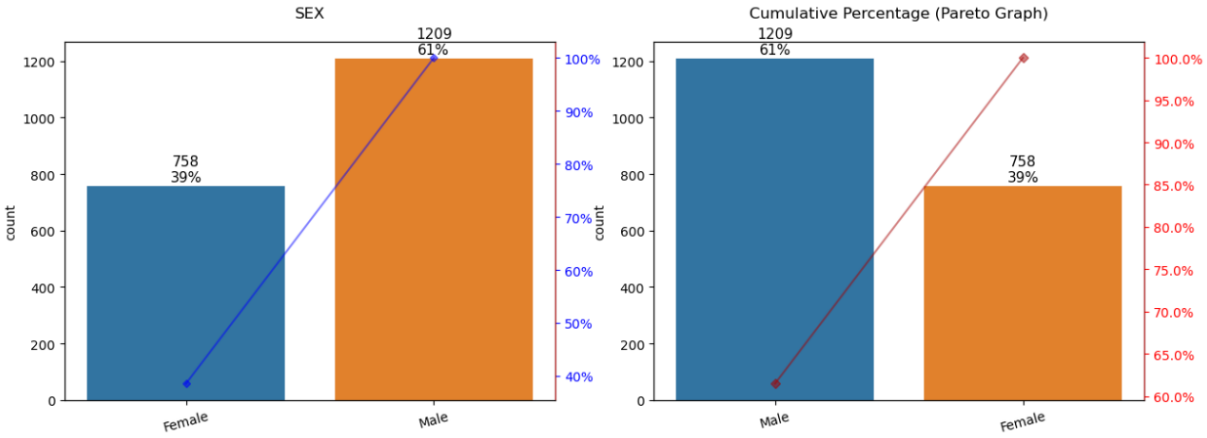
Fabiano Papaiz [1, 2, 3*], Mario Emílio Teixeira Dourado Jr. [1, 4],
Ricardo Alexsandro de Medeiros Valentim [1], Rafael Pinto [1, 5],
Antonio Higor Freire de Morais [1, 3], Joel Perdiz Arrais[2]

[1]Laboratory of Technological Innovation in Health (LAIS), Federal University of Rio Grande do Norte, Natal, Brazil. [2]Centre for Informatics and Systems, Department of Informatics Engineering, University of Coimbra, Coimbra, Portugal. [3]Advanced Nucleus of Technological Innovation (NAVI), Federal Institute of Rio Grande do Norte, Natal, Brazil. [4]Department of Internal Medicine, Federal University of Rio Grande do Norte, Natal, Brazil. [5]Information Systems Coordination, Federal Institute of Rio Grande do Norte, Natal, Brazil.

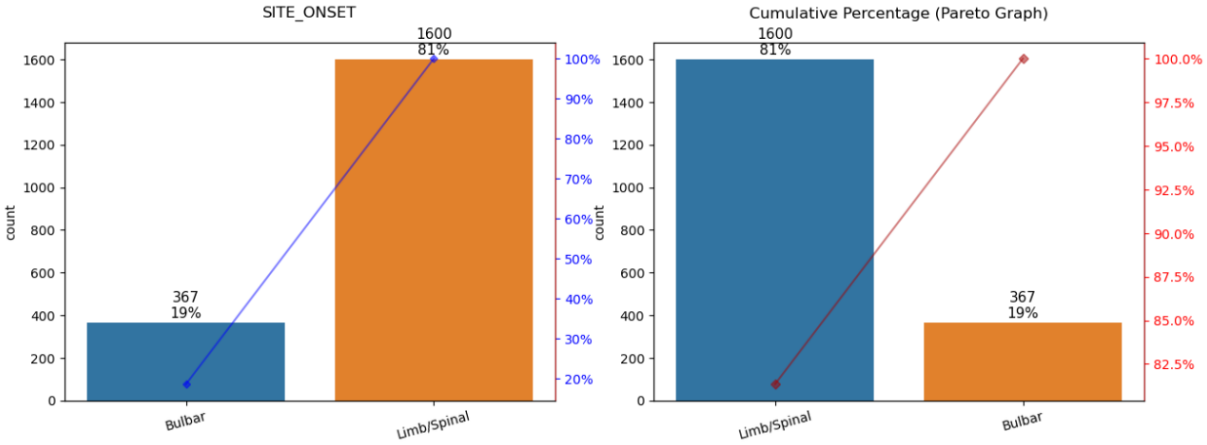
Table of Contents

Input and Output Variable Distributions.

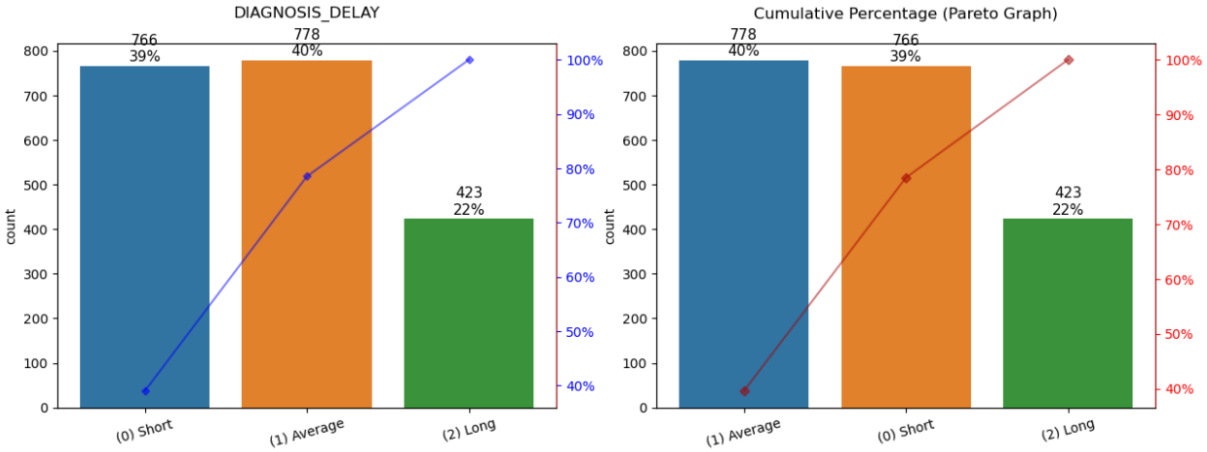
Column Sex



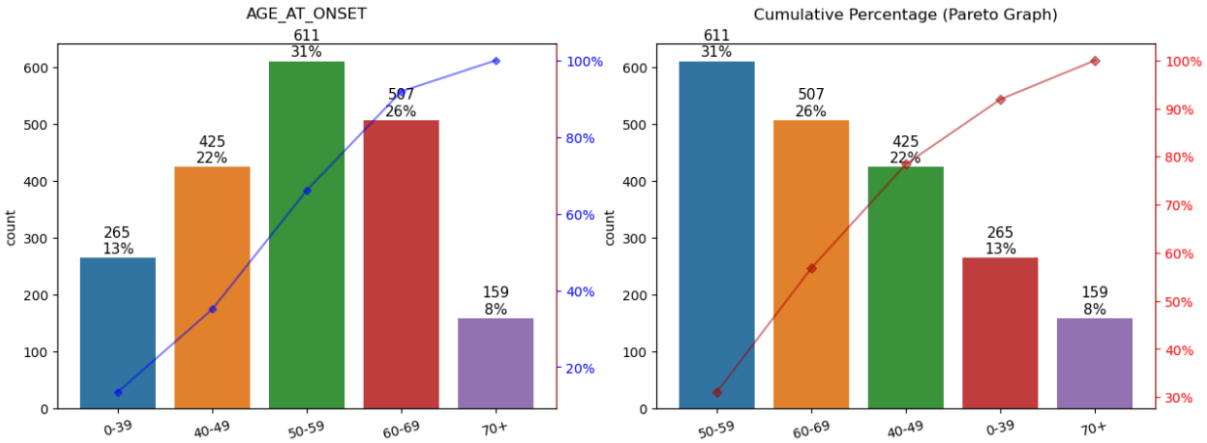
Column Site_Onset



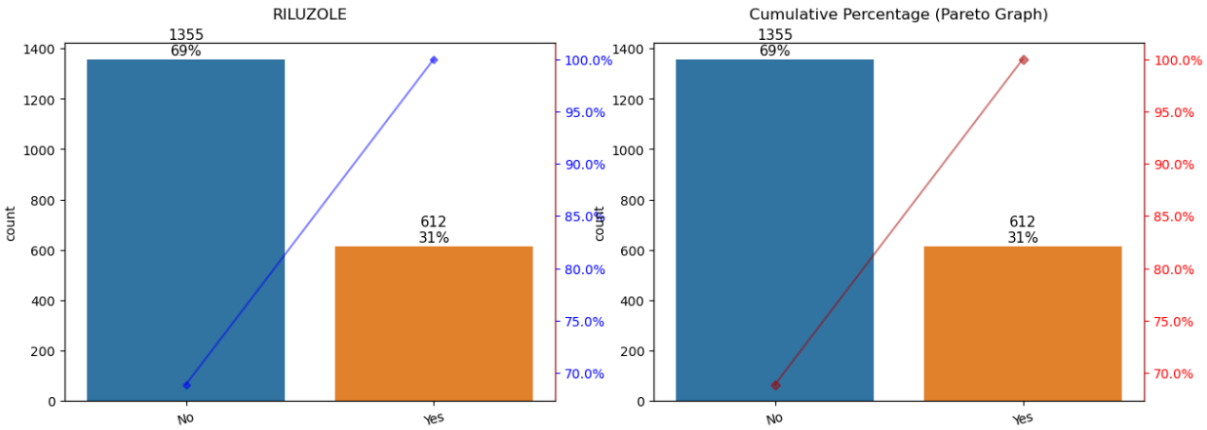
Column Diagnosis_Delay



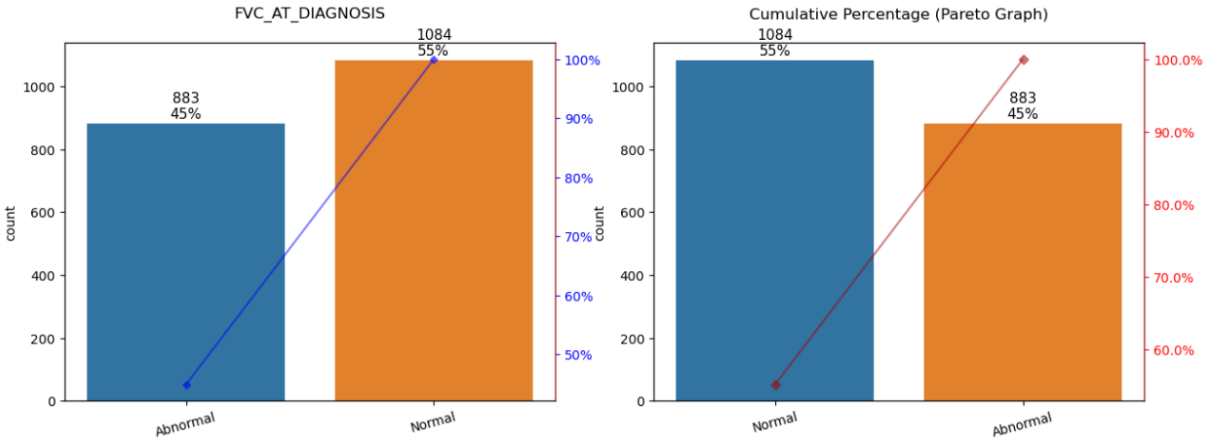
Column Age_at_Onset



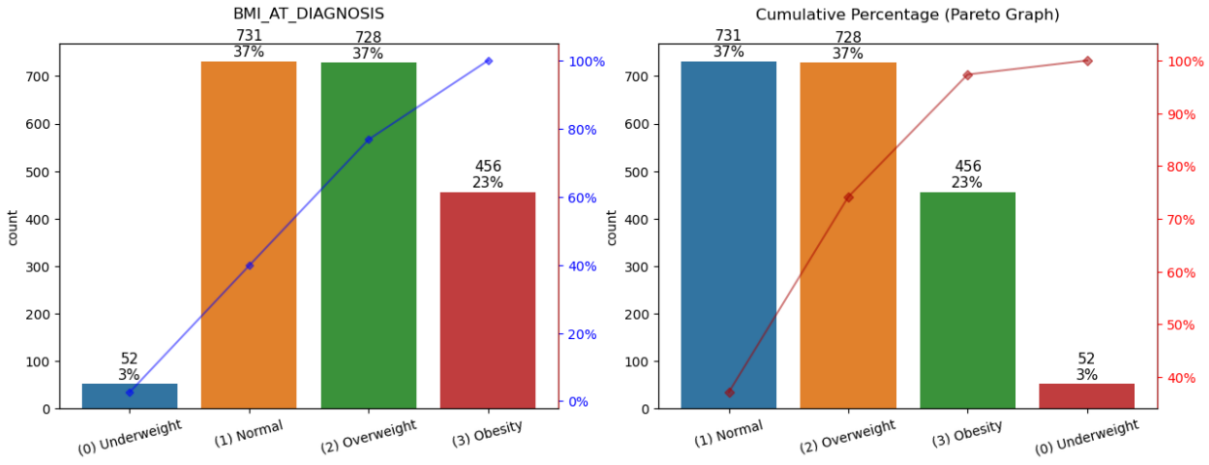
Column Riluzole



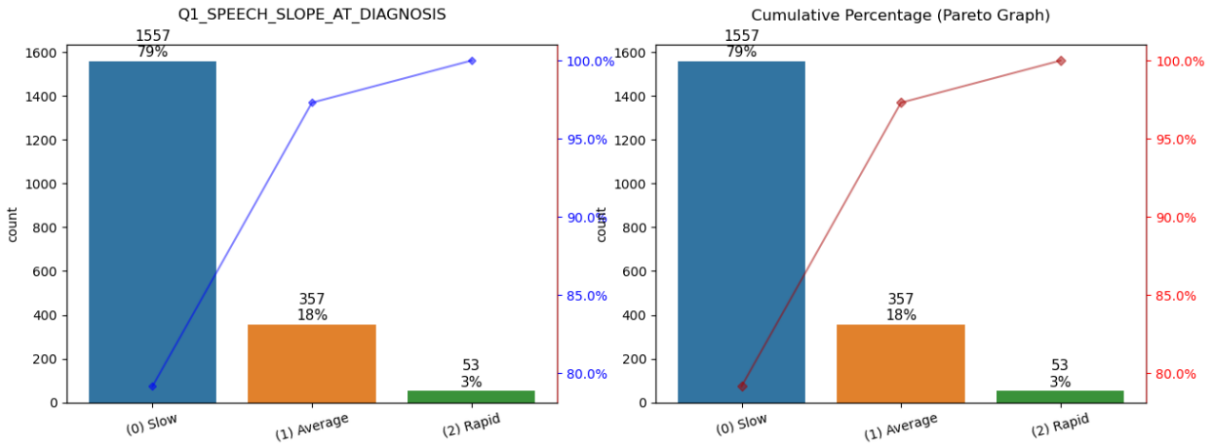
Column FVC_at_Diagnosis



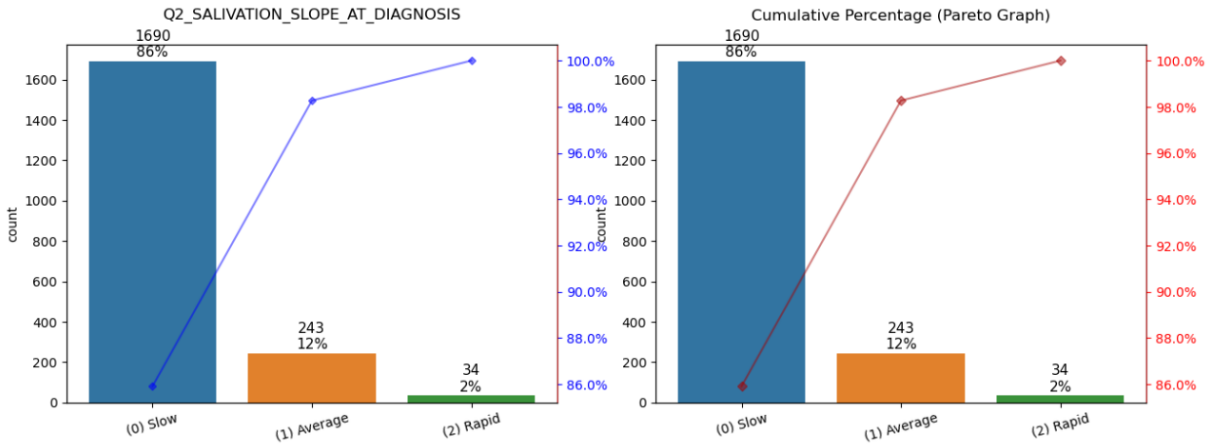
Column BMI_at_Diagnosis



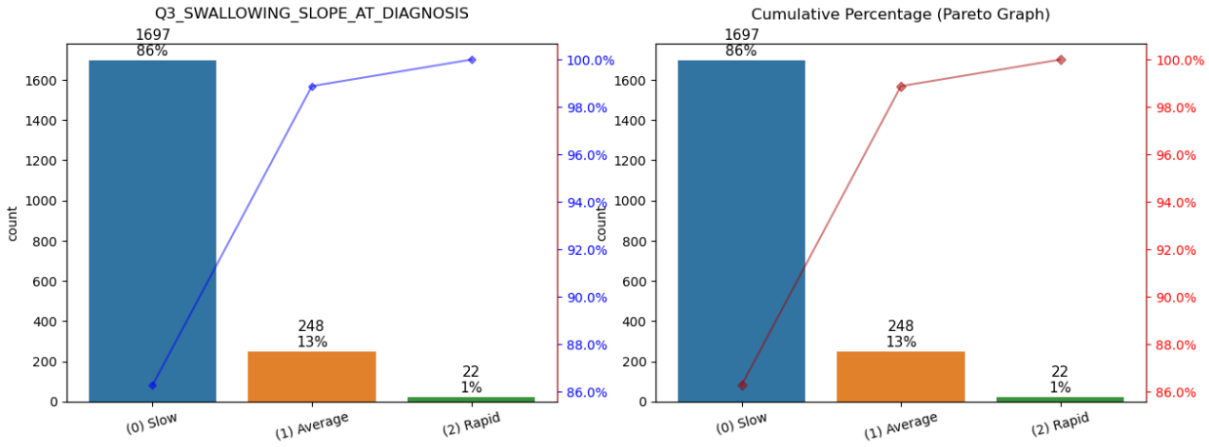
Column Q1_Speech_slope_at_Diagnosis



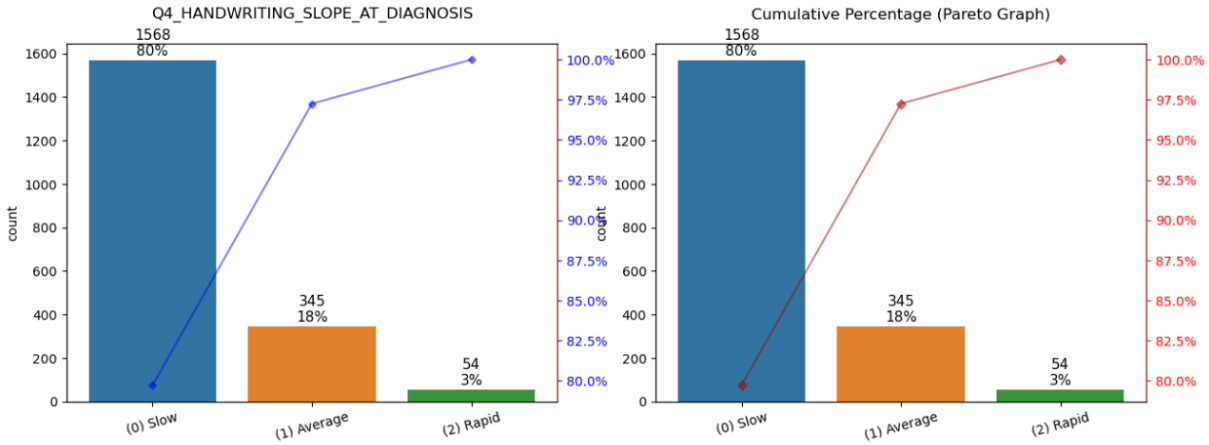
Column Q2_Salivation_slope_at_Diagnosis



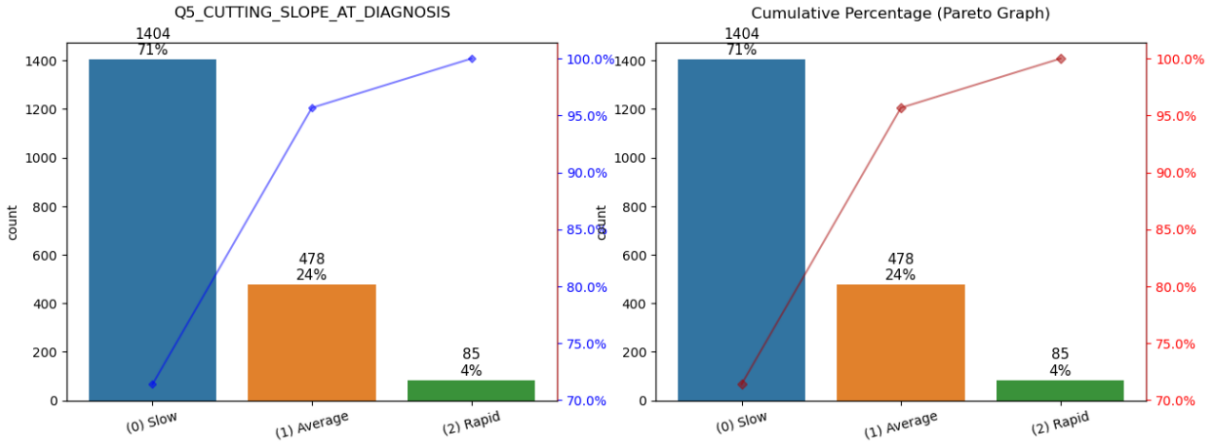
Column Q3_Swallowing_slope_at_Diagnosis



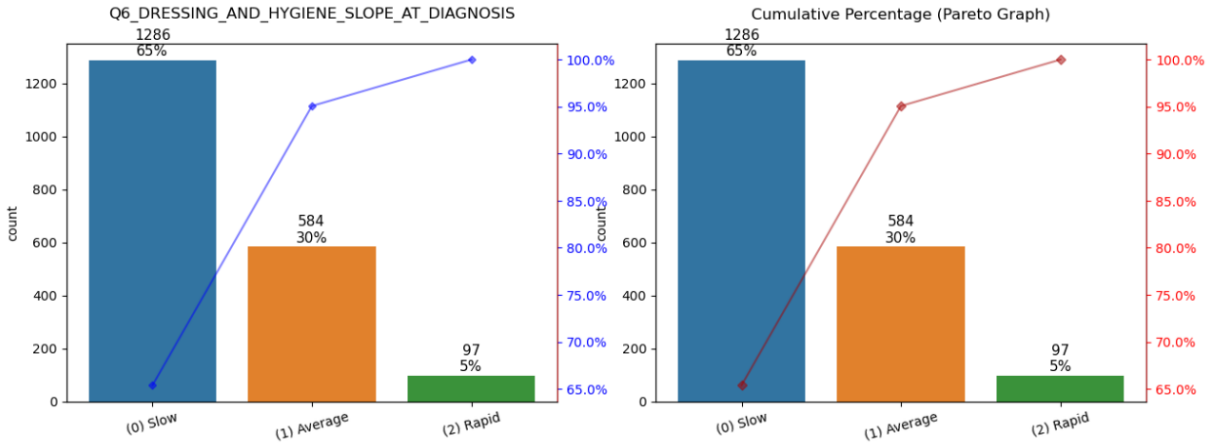
Column Q4_Handwriting_slope_at_Diagnosis



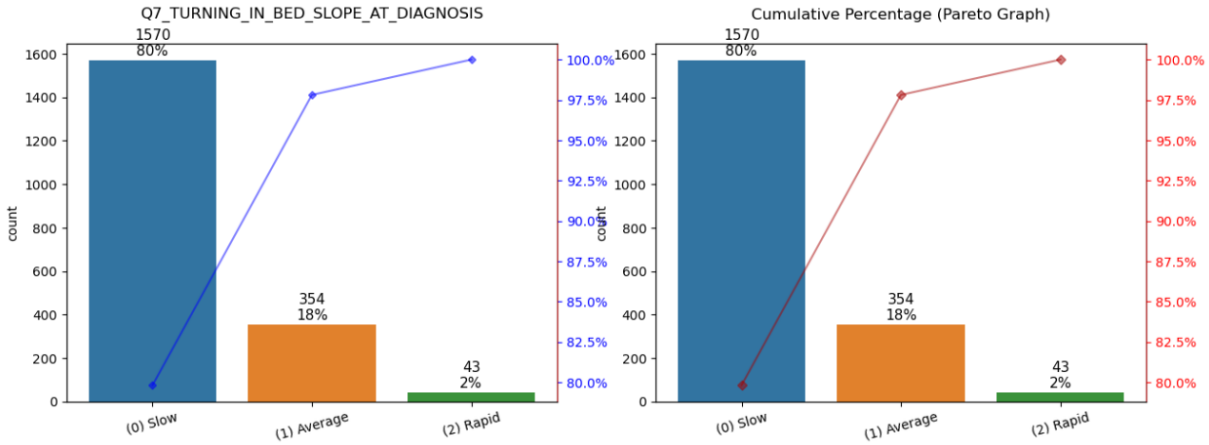
Column Q5_Cutting_slope_at_Diagnosis



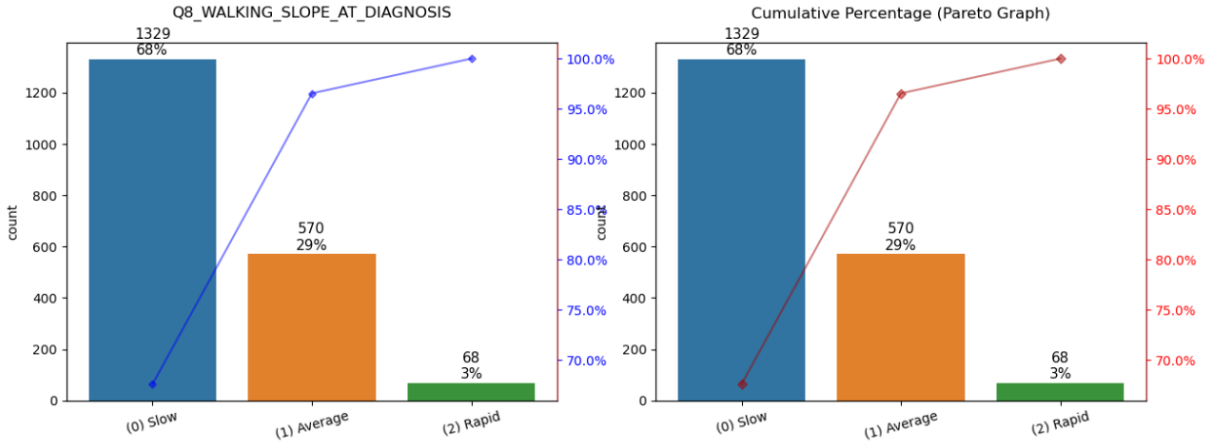
Column Q6_Dressing_and_Hygiene_slope_at_Diagnosis



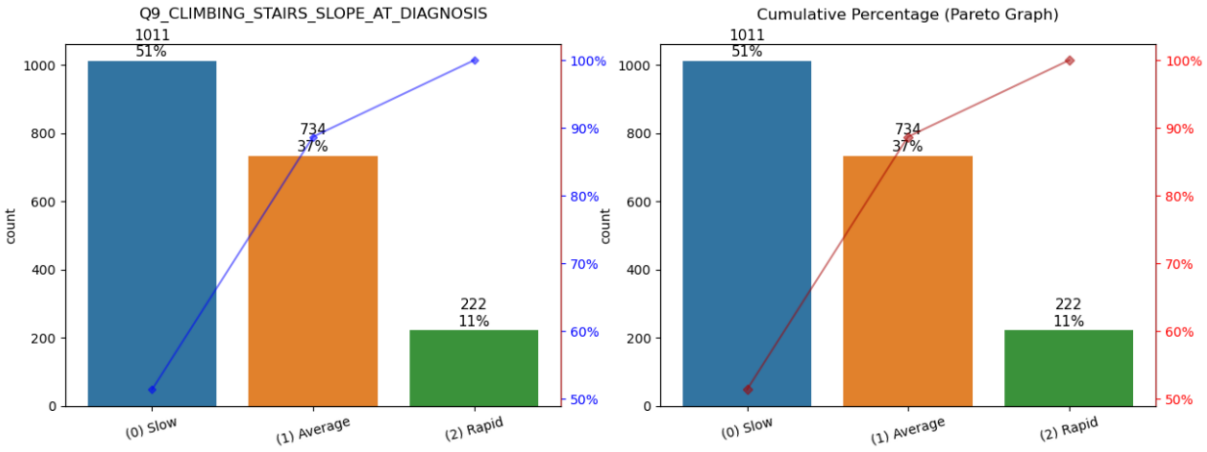
Column Q7_Turning_in_Bed_slope_at_Diagnosis



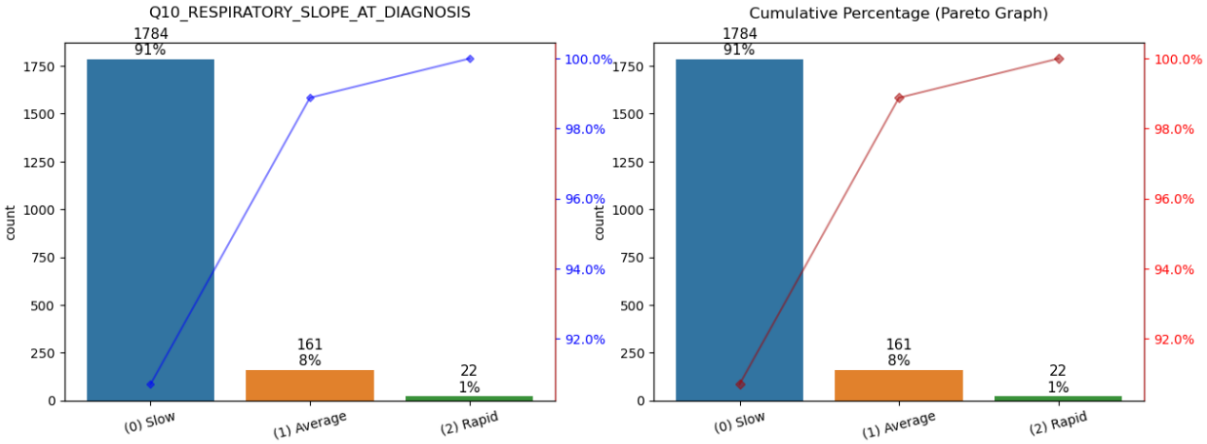
Column Q8_Walking_slope_at_Diagnosis



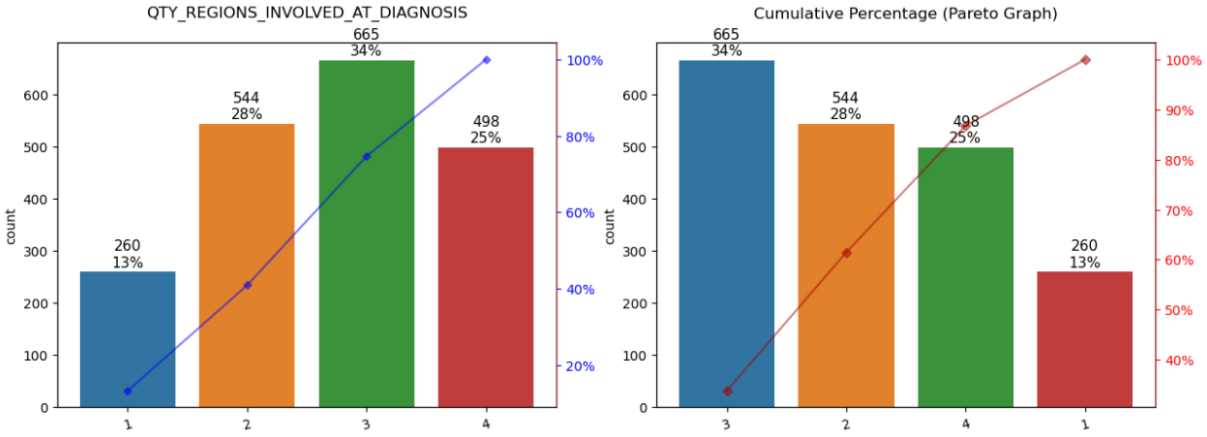
Column Q9_Climbing_Stairs_slope_at_Diagnosis



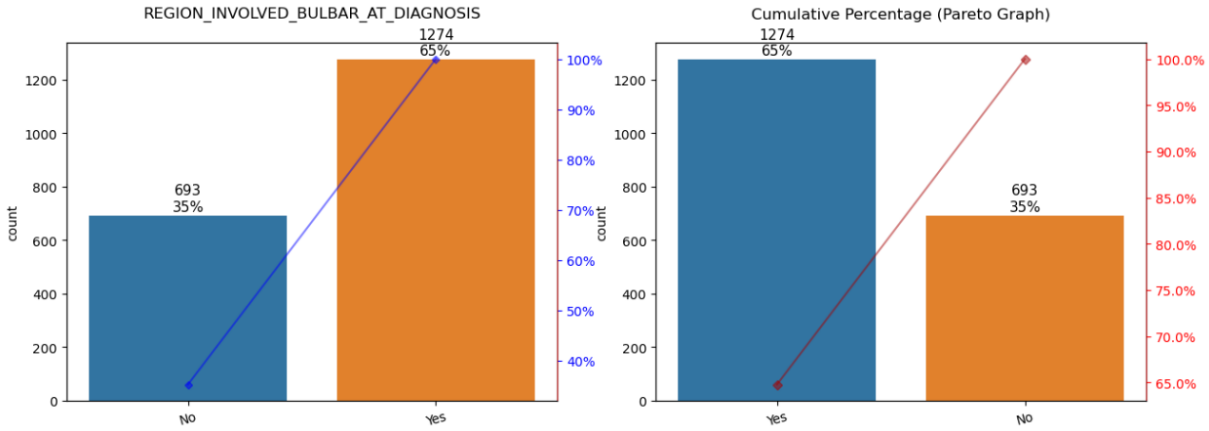
Column Q10_Respiratory_slope_at_Diagnosis



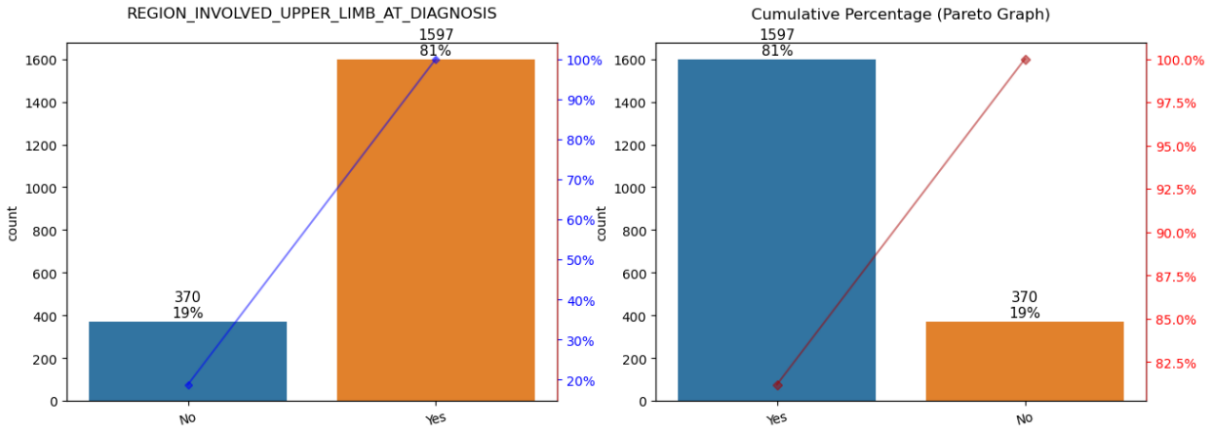
Column Qty_Regions_Involved_at_Diagnosis



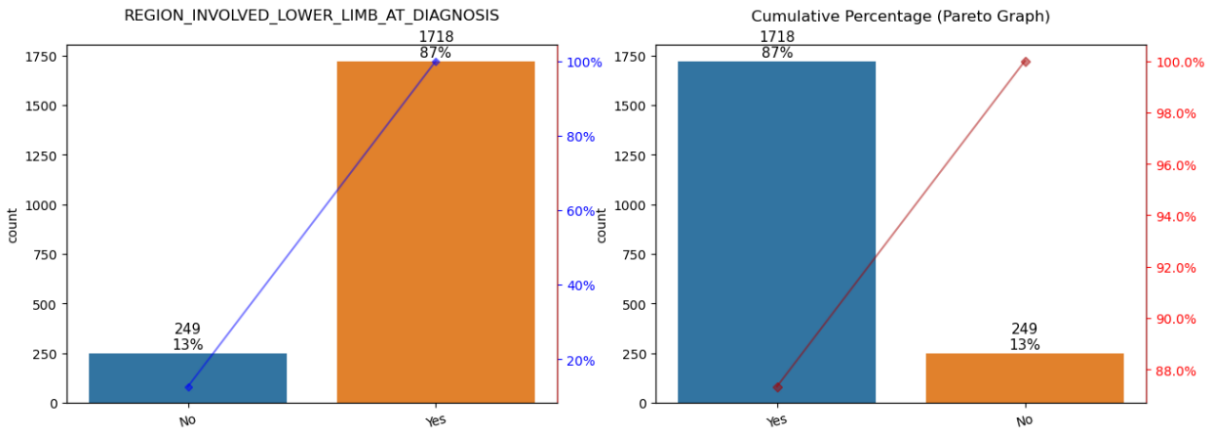
Column Region_Involved_Bulbar_at_Diagnosis



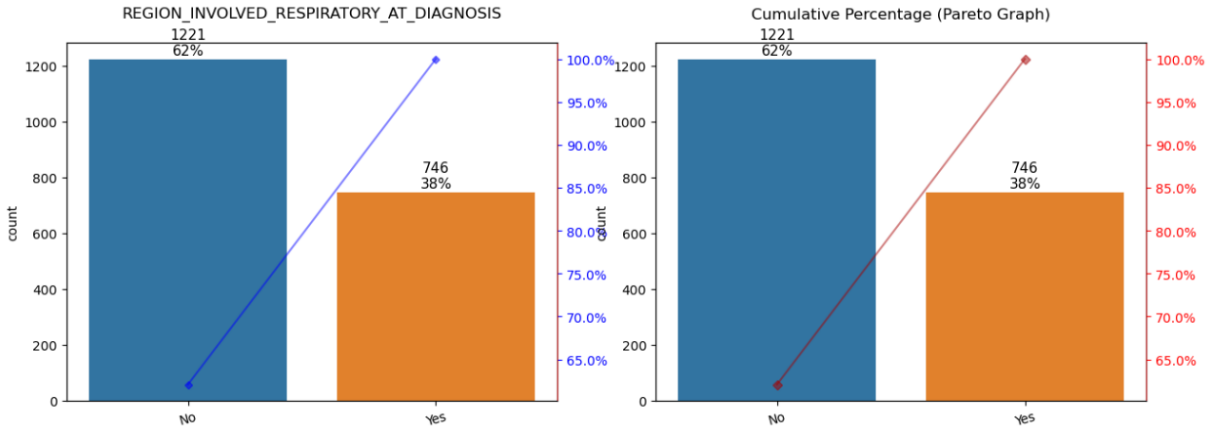
Column Region_Involved_Upper_Limb_at_Diagnosis



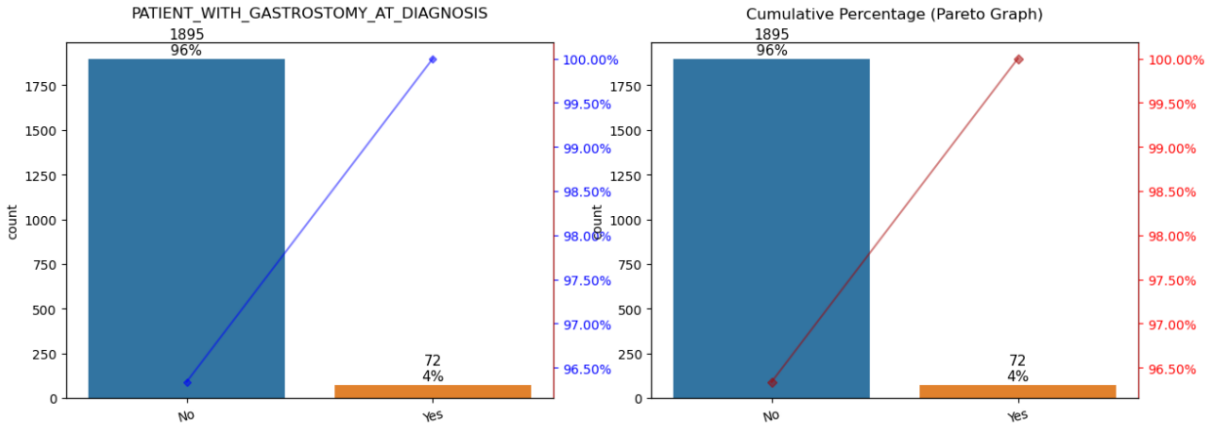
Column Region_Involved_Lower_Limb_at_Diagnosis



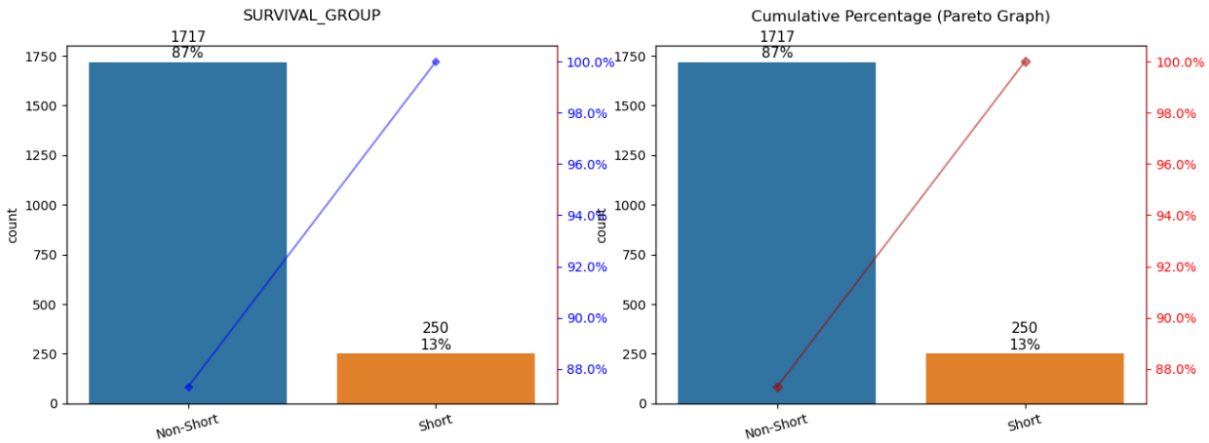
Column Region_Involved_Respiratory_at_Diagnosis



Column Patient_with_Gastrostomy_at_Diagnosis

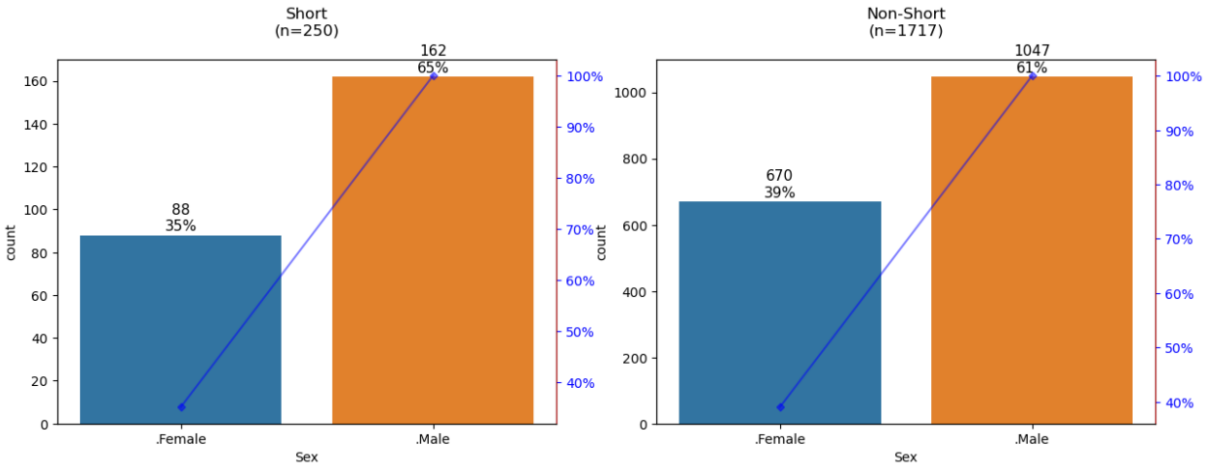


OUTPUT Column Survival_Group

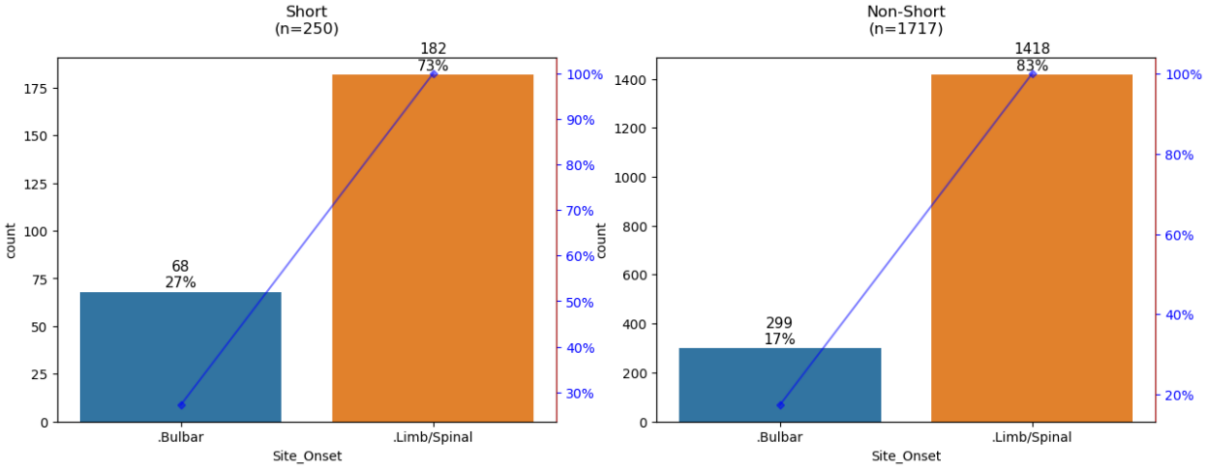


Compare the variable distributions for the 2 Survival Groups: *Short* and *Non-Short*.

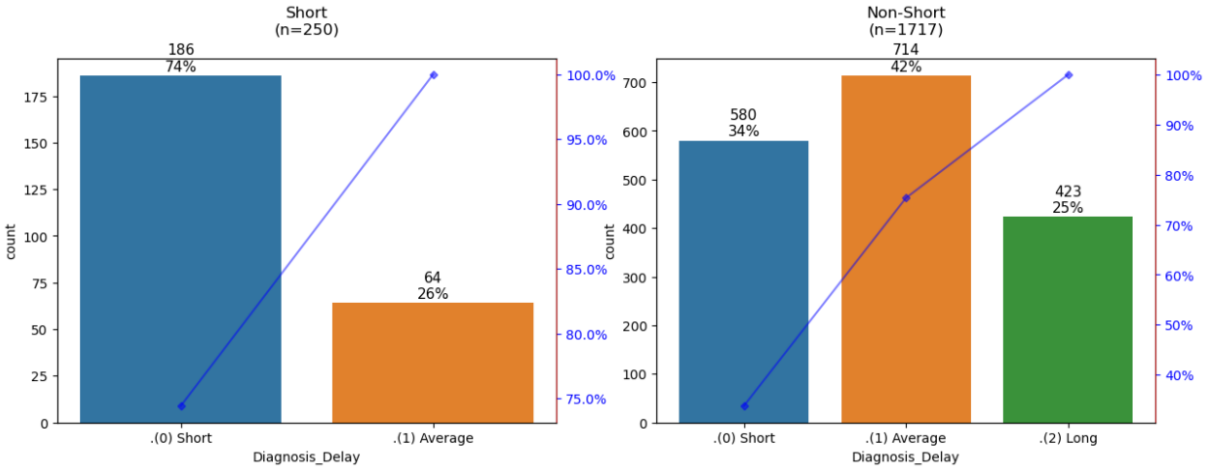
Column Sex



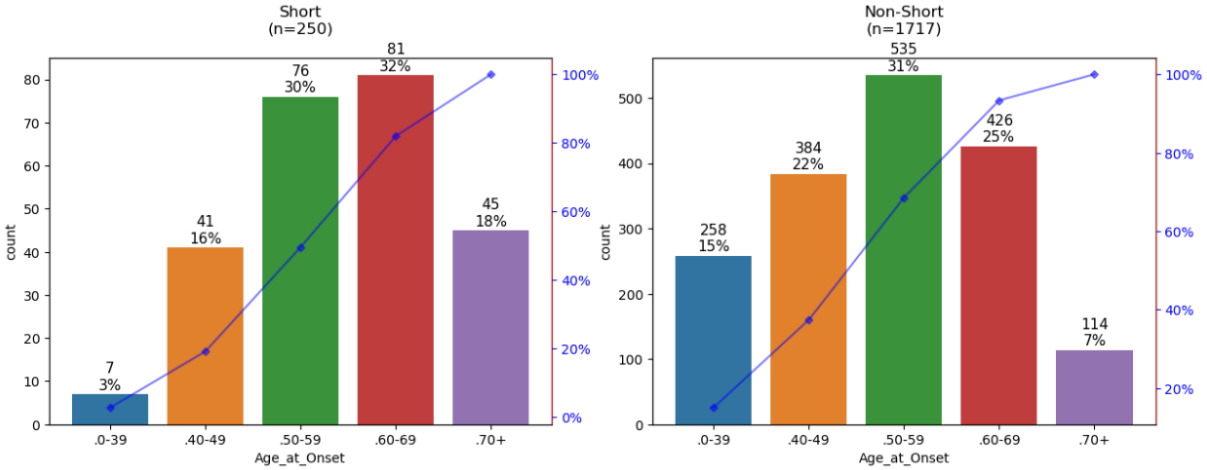
Column Site_Onset



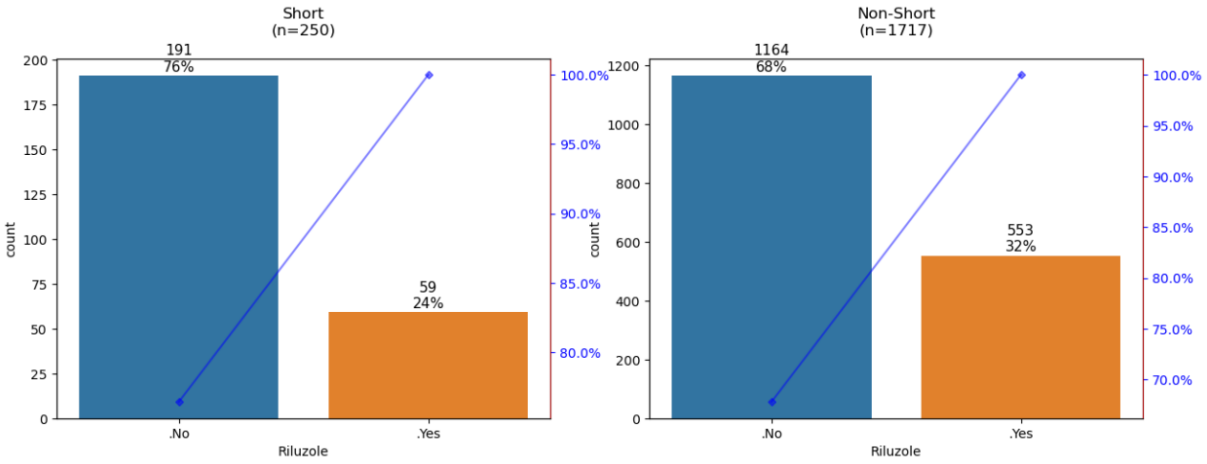
Column Diagnosis_Delay



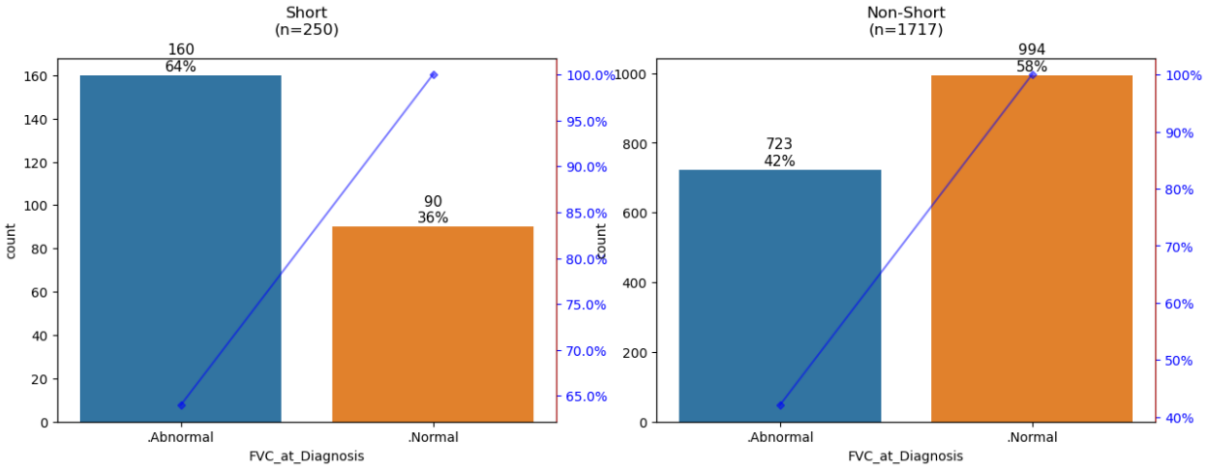
Column Age_at_Onset



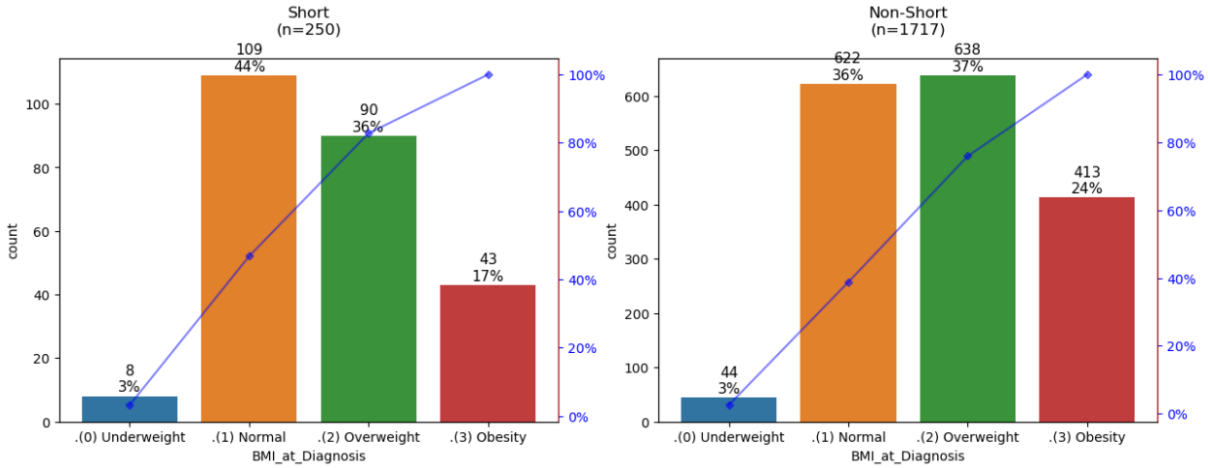
Column Riluzole



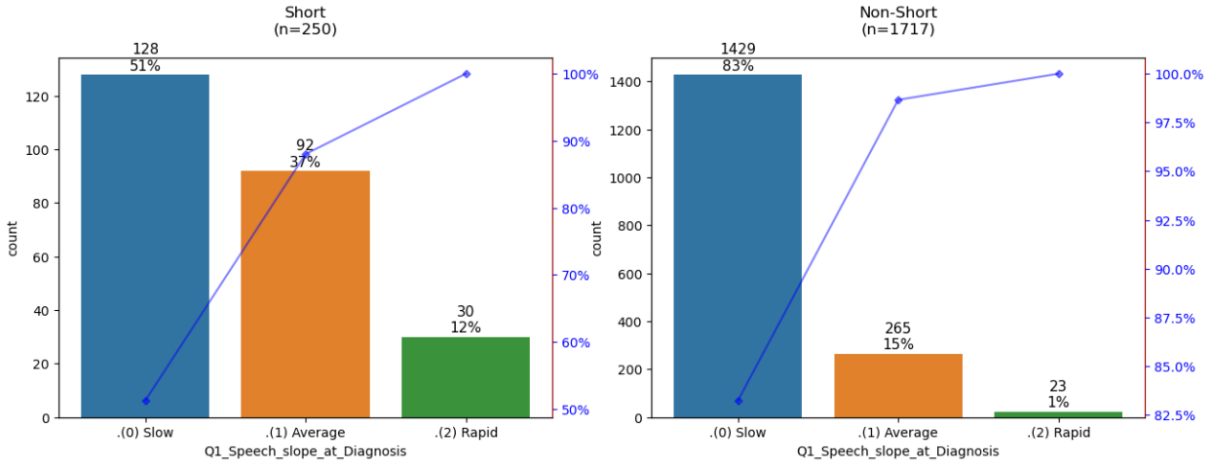
Column FVC_at_Diagnosis



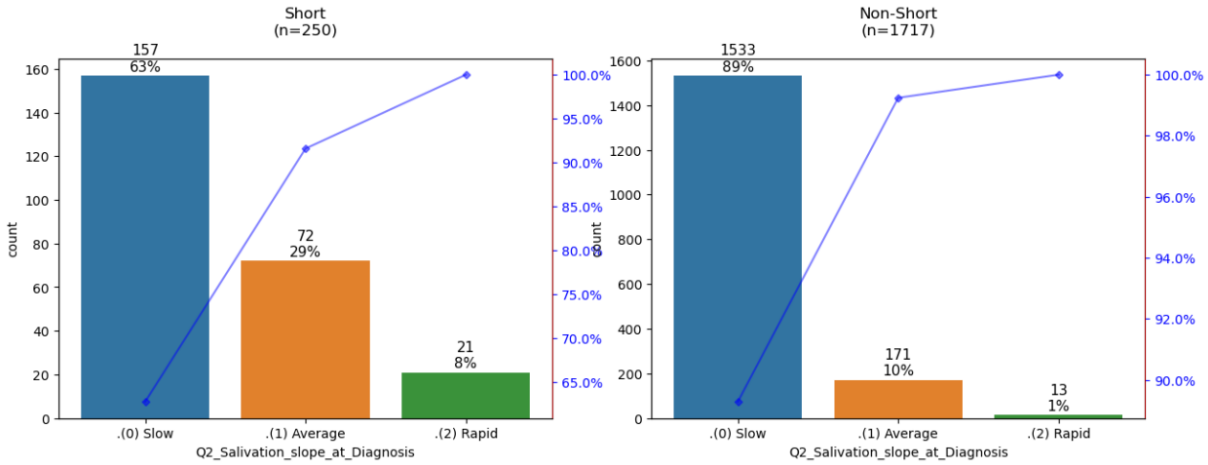
Column BMI_at_Diagnosis



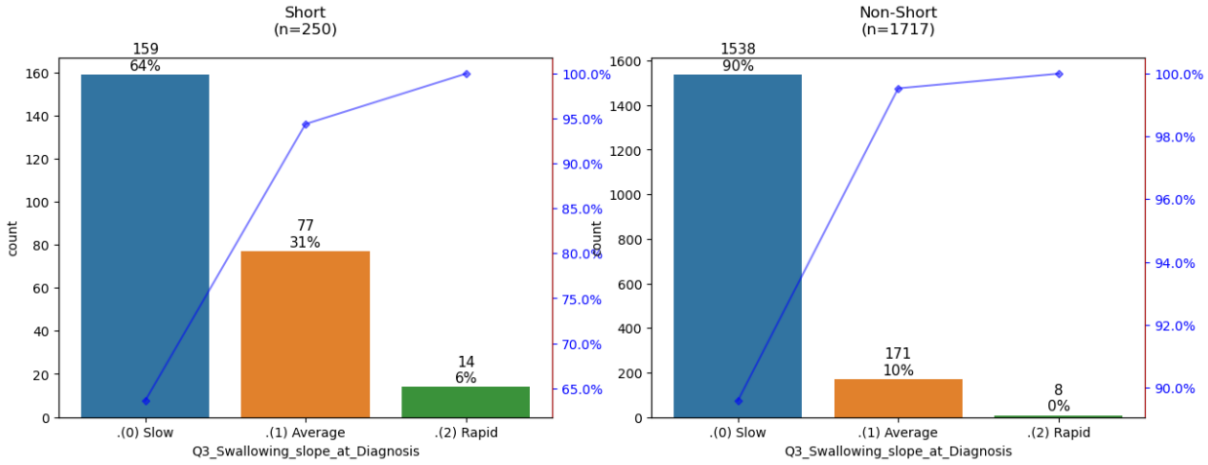
Column Q1_Speech_slope_at_Diagnosis



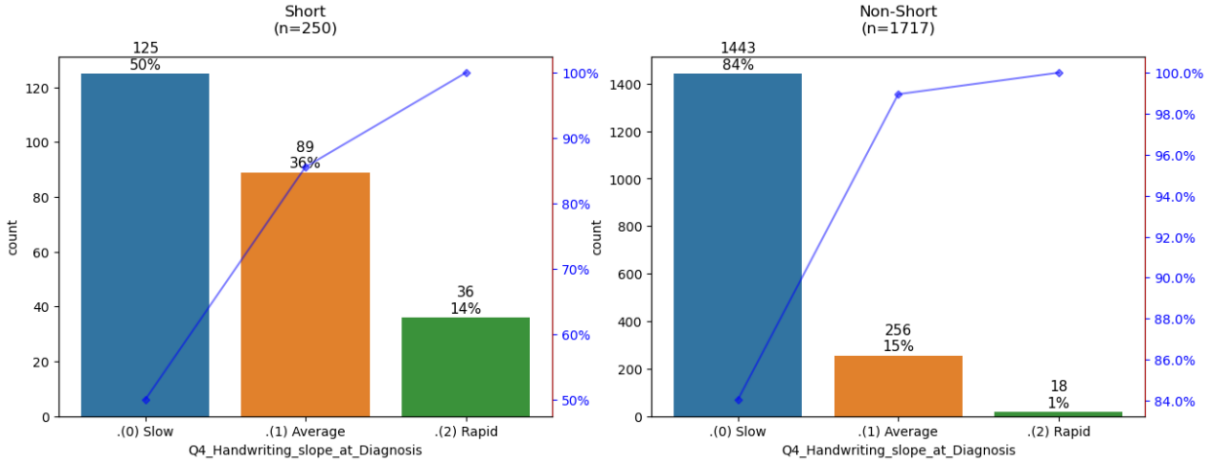
Column Q2_Salivation_slope_at_Diagnosis



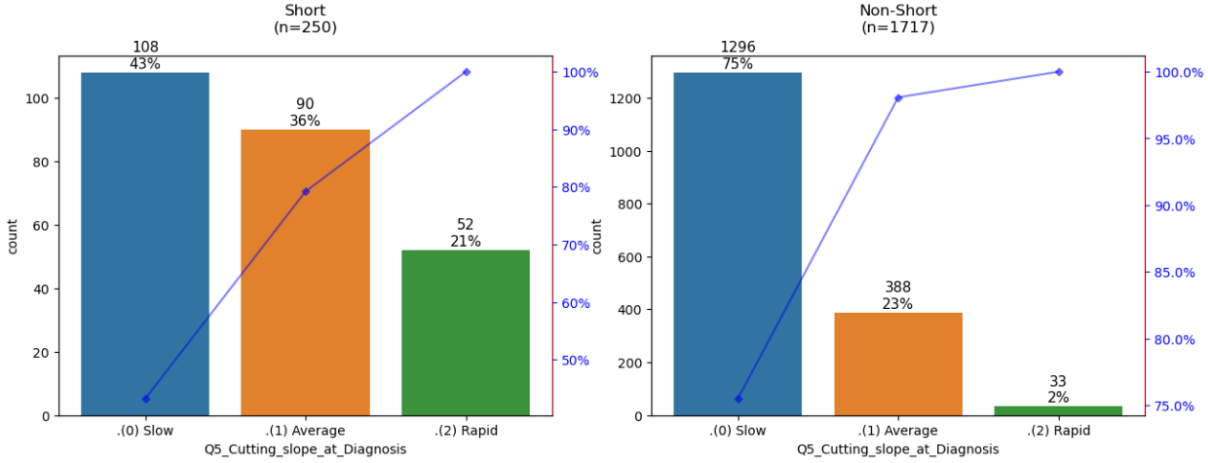
Column Q3_Swallowing_slope_at_Diagnosis



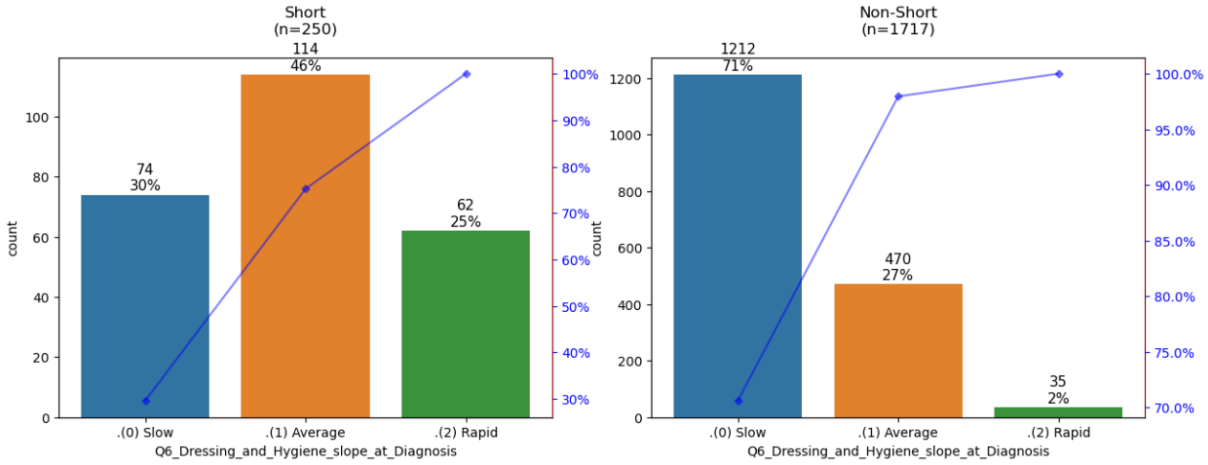
Column Q4_Handwriting_slope_at_Diagnosis



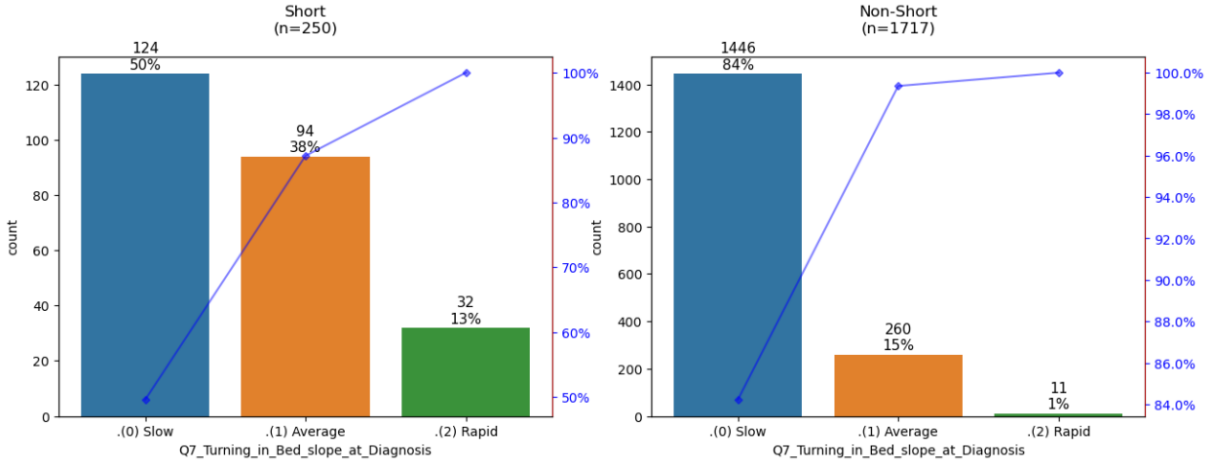
Column Q5_Cutting_slope_at_Diagnosis



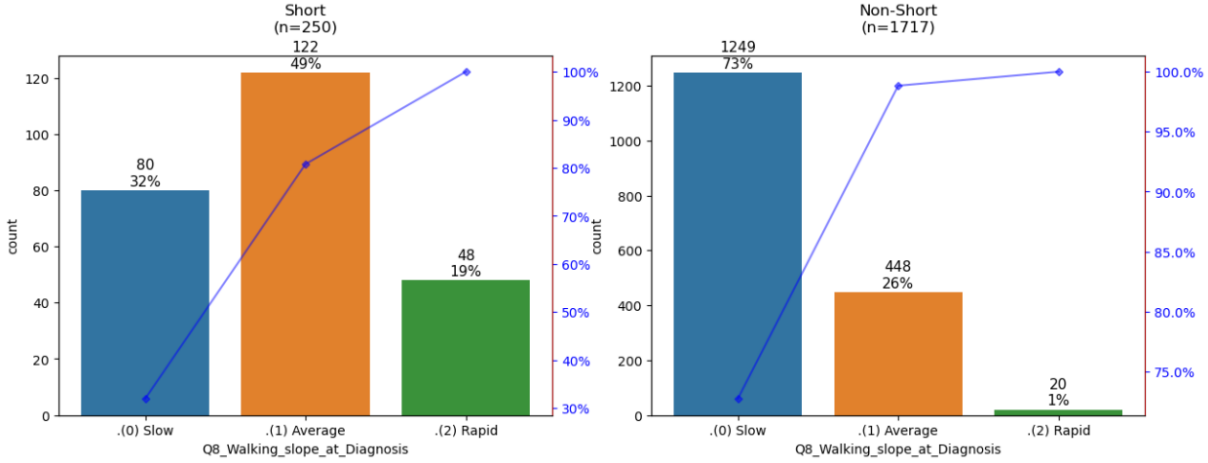
Column Q6_Dressing_and_Hygiene_slope_at_Diagnosis



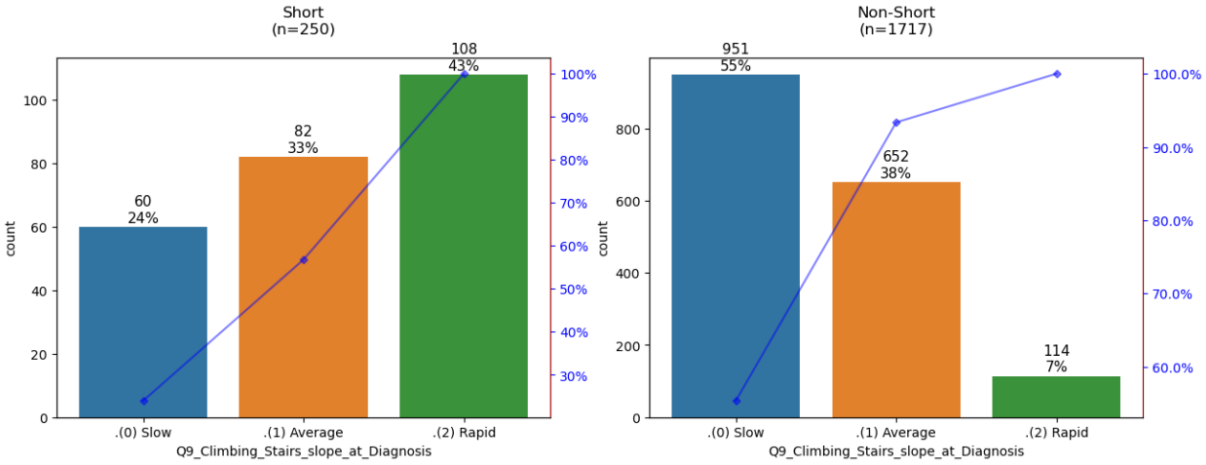
Column Q7_Turning_in_Bed_slope_at_Diagnosis



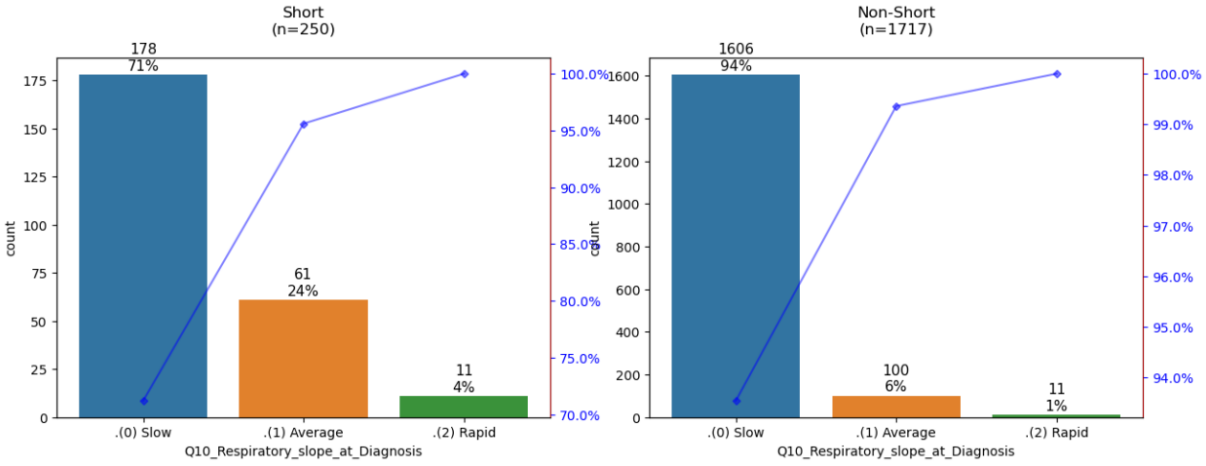
Column Q8_Walking_slope_at_Diagnosis



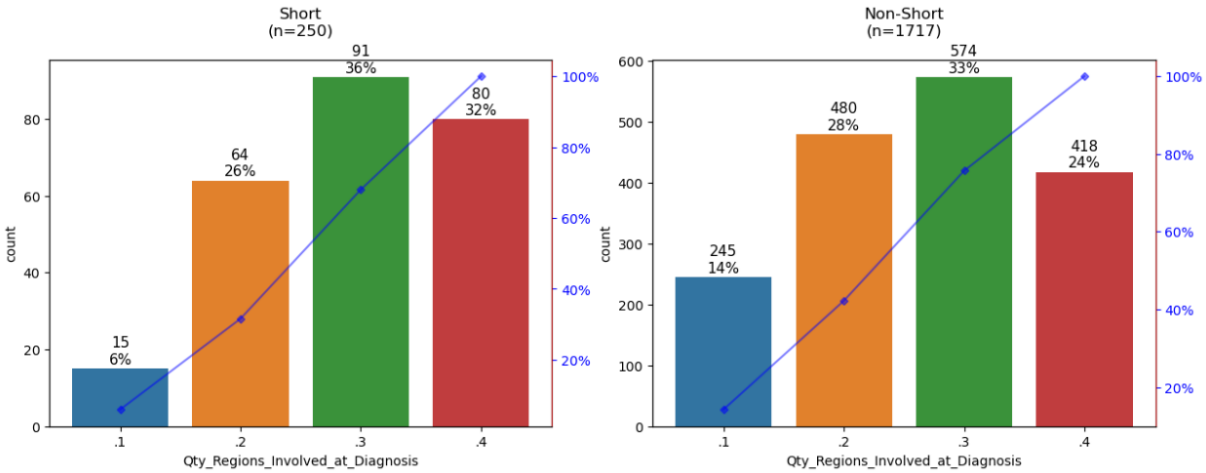
Column Q9_Climbing_Stairs_slope_at_Diagnosis



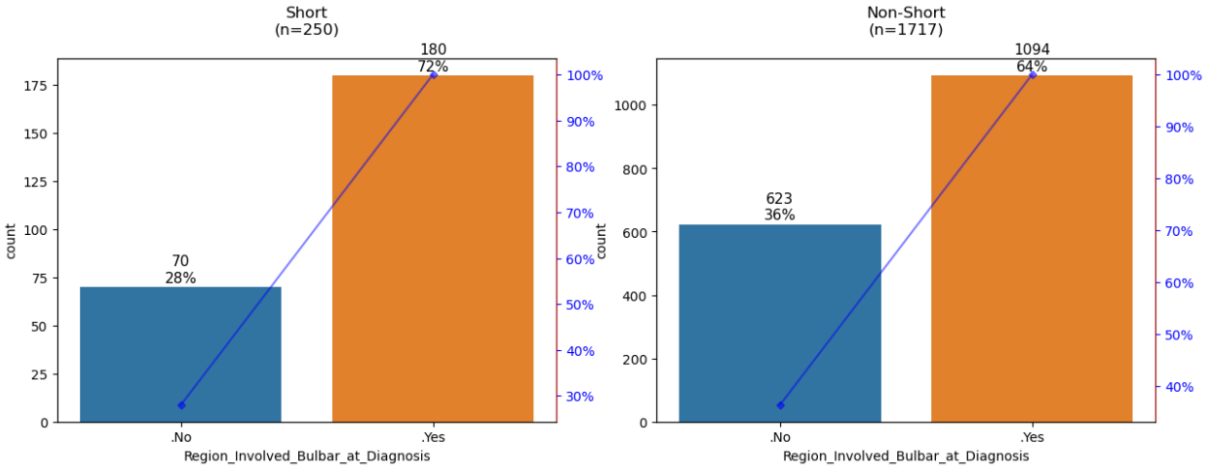
Column Q10_Respiratory_slope_at_Diagnosis



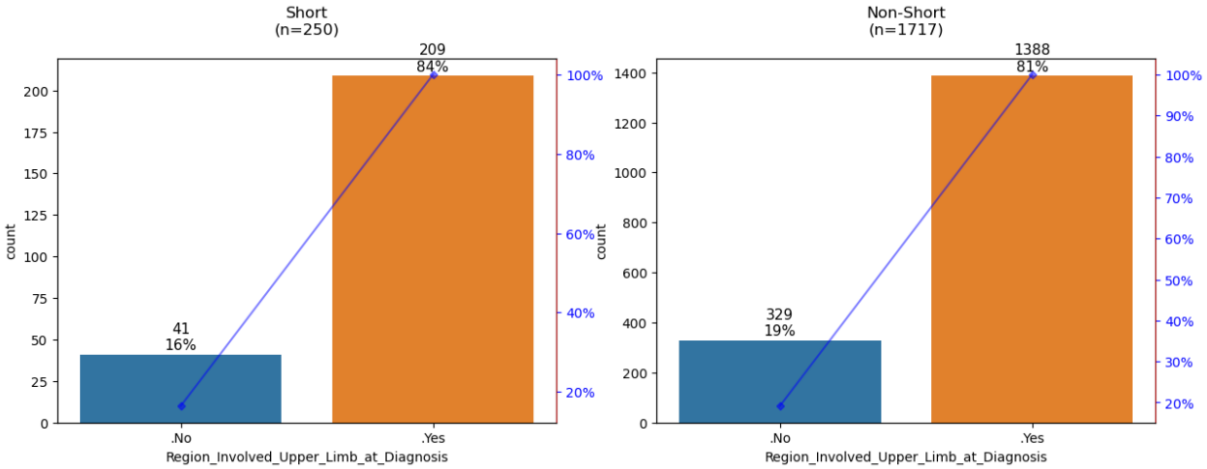
Column Qty_Regions_Involved_at_Diagnosis



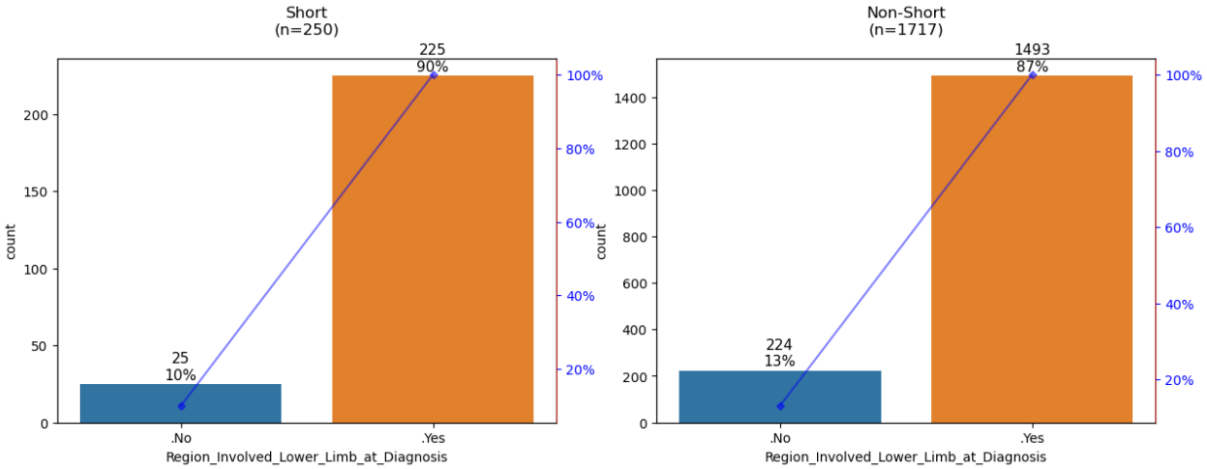
Column Region_Involved_Bulbar_at_Diagnosis



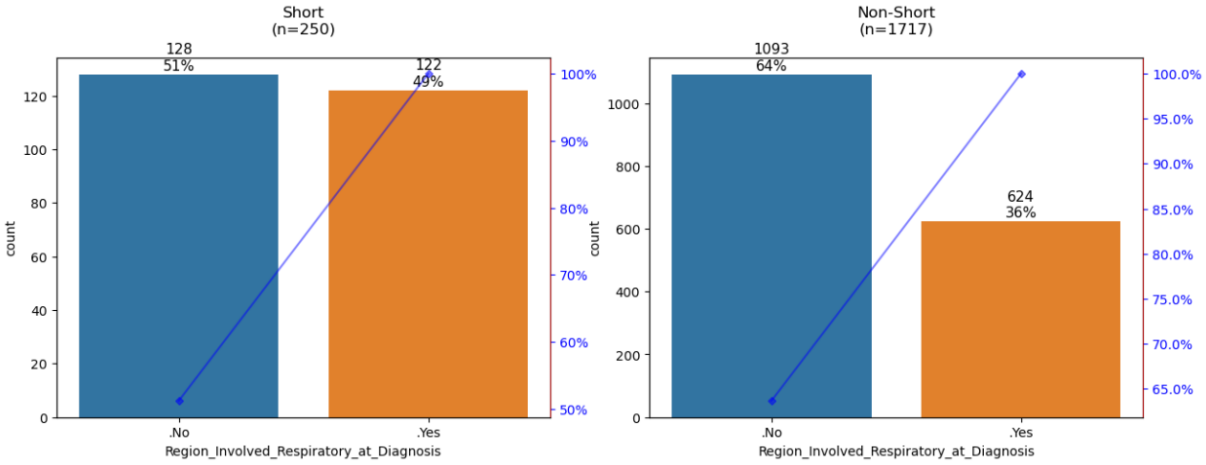
Column Region_Involved_Upper_Limb_at_Diagnosis



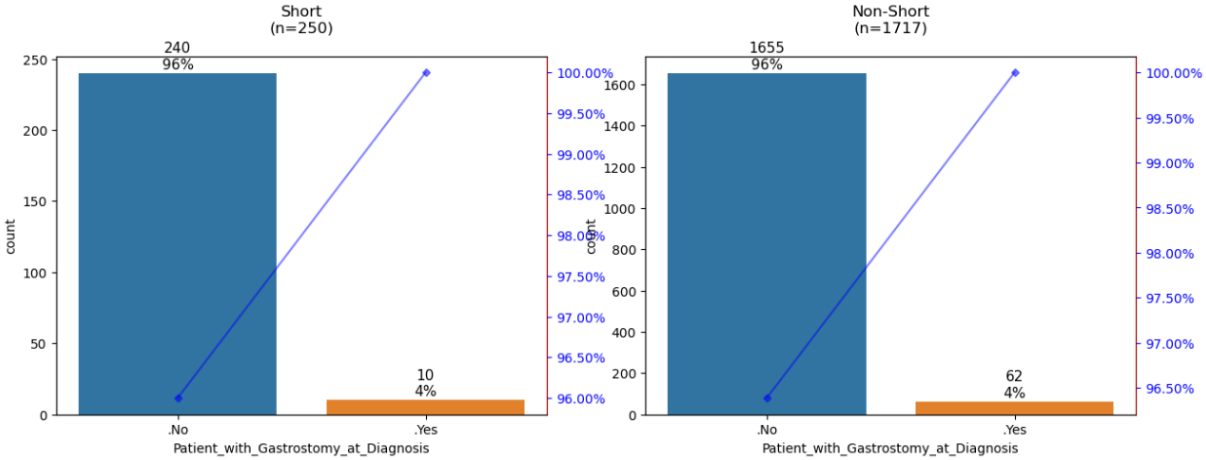
Column Region_Involved_Lower_Limb_at_Diagnosis



Column Region_Involved_Respiratory_at_Diagnosis

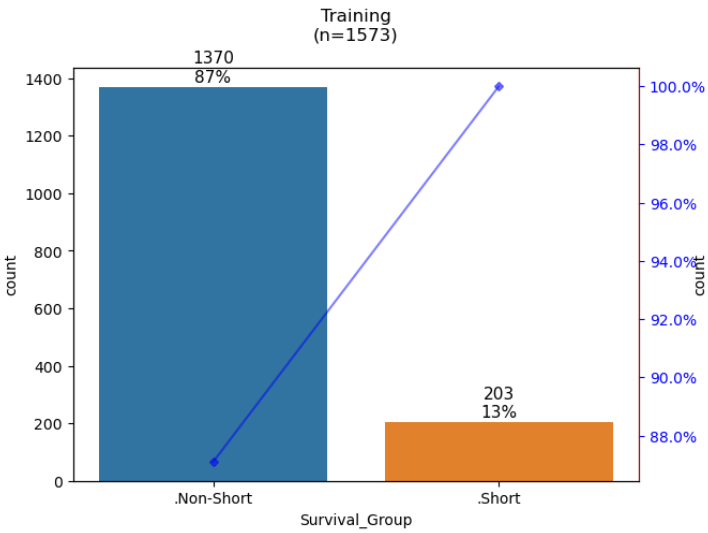


Column Patient_with_Gastrostomy_at_Diagnosis

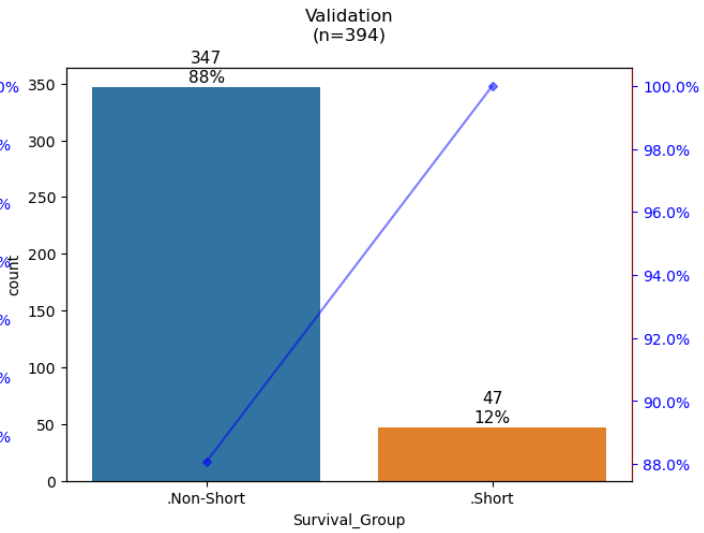


Compare the output variable distribution for the *Training* and *Validation* subsets used to train and validate the machine learning models

TRAINING



VALIDATION



Grid-Search hyperparameters used for each algorithm.

- Decision Tree

```
max_depths = [3, 4, 5, 7, 9, 10, 15, 25, 50]
criterions = ['gini', 'entropy']
class_weights = [None, 'balanced']
```

- k-Nearest Neighbors

```
weights = ['uniform', 'distance']
distance_metrics = ['euclidean', 'manhattan', 'chebyshev']
        k = [3, 5, 9, 15]
```

- Naïve Bayes

```
        alphas = [0.1, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0]
        norms = [False, True]
```

- Random Forest

```
        max_depths = [10, 15, 25, 50]
        num_estimators = [50, 75, 100, 200]
        criterions = ['gini', 'entropy']
        class_weights = [None, 'balanced', 'balanced_subsample']
```

- SVM

```
kernels = ['rbf', 'linear']
  gammas = ['scale', 'auto', ]
```

```
class_weights = [None, 'balanced', ]
Cs = [0.1, 0.3, 0.5, 0.7, 1, 3, 5, 10, 100, 200, 1000, 1500, 1700, 2000]
```

- Neural Networks

```
max_iter = [1000]
alphas = [0.0001, 0.00001, 0.05, 0.1, 0.3, 0.5]
activations = ['tanh', 'relu']
solvers = ['sgd', 'adam']
learning_rates = ['constant', 'adaptive']
learning_rate_init = [0.7]
layers = [
  (30),
  (30, 30),
  (30, 30, 30),
  (qty_features, ),
  (qty_features, qty_features),
  (qty_features, qty_features, qty_features),
  (qty_features, (qty_features*2)),
  (qty_features, (qty_features*2), qty_features),
  (qty_features, (qty_features*2), (qty_features*2), qty_features),
]
```

- Balanced Bagging

```
num_estimators = [11, 15, 51, 75, 101]
sampling_strategies = ['all', 'majority', 'auto']
warm_starts = [False, True]
```

- Balanced Random Forest

```
max_depths = [5, 7, 10, 15]
criterions = ['gini', 'entropy']
num_estimators = [7, 11, 15, 19, 21, 25, 31 , 51]
sampling_strategies = ['all', 'majority', 'auto']
warm_starts = [False, True]
replacements = [False, True]
```

Best models hyperparameters.

Scenario	Classifier	Hyperparameters
Ensemble-Imbalance	BalancedBaggingClassifier using Decision Tree	{'estimator': DecisionTreeClassifier (class_weight='balanced', max_depth=4, random_state=42), 'n_estimators':7, 'random_state':42, 'replacement':True, 'sampling_strategy':'all', 'warm_start':False}
	BalancedBaggingClassifier using Neural Networks	{'estimator': MLPClassifier(activation='tanh', alpha=0.1, hidden_layer_sizes=30, learning_rate='adaptive', learning_rate_init=0.7, max_iter=2000, random_state=42), 'n_estimators':101, 'random_state':42, 'replacement':True, 'sampling_strategy':'auto', 'warm_start':True}
	BalancedBaggingClassifier using SVM	"{'estimator': SVC (C=0.7, class_weight='balanced', gamma='auto', probability=True, random_state=42), 'n_estimators':31, 'random_state':42, 'replacement':True, 'sampling_strategy':'majority', 'warm_start':False}
	BalancedBaggingClassifier using k-NN	"{'estimator': KNeighborsClassifier (metric='euclidean', weights='distance'), 'n_estimators':101, 'random_state':42, 'replacement':True, 'sampling_strategy':'all', 'warm_start':True}
	BalancedBaggingClassifier using Naïve Bayes	"{'estimator': GaussianNB(), 'n_estimators':19, 'random_state':42, 'replacement':True, 'sampling_strategy':'all', 'warm_start':False}
	BalancedRandomForestClassifier	{'criterion':'entropy', 'max_depth':7, 'n_estimators':19, 'random_state':42, 'replacement':True, 'sampling_strategy':'auto', 'warm_start':False}
Single-Model	DecisionTreeClassifier	{'ccp_alpha':0.0, 'class_weight':'balanced', 'criterion':'gini', 'max_depth':4, 'max_features':None, 'max_leaf_nodes':None, 'min_impurity_decrease':0.0, 'min_samples_leaf':1, 'min_samples_split':2, 'min_weight_fraction_leaf':0.0, 'random_state':42, 'splitter':'best'}
	MLPClassifier	{'activation':'tanh', 'alpha':0.3, 'hidden_layer_sizes':(23,23,23), 'learning_rate':'constant', 'learning_rate_init':0.7, 'max_iter':2000, 'random_state':42, 'solver':'sgd'}
	RandomForestClassifier	{'class_weight':'balanced', 'criterion':'gini', 'max_depth':5, 'n_estimators':51, 'random_state':42}
	svm.SVC	"{'C':1, 'class_weight':'balanced', 'gamma':'auto', 'kernel':'rbf', 'probability':True, 'random_state':42}
	DecisionTreeClassifier	"{'class_weight':'balanced', 'criterion':'gini', 'max_depth':4, 'random_state':42}
	GaussianNB	{ }
	KNeighborsClassifier	{'metric':'manhattan', 'n_neighbors':3, 'weights':'uniform'}

Appendix B

Grid-Search hyperparameters used for each algorithm (GRU and LSTM)

This appendix includes the additional information related with Chapter 5.

Grid-Search hyperparameters:

1. **Optimizer**: ADAM
2. **Drop-out rates**: 10%, 20%, 30%, 35%, 40%, 45%, and 50%
3. **Regularizers**:
    - **Types**: LASSO, Ridge, and ElasticNet
    - **Values**: 0.1, 0.01, 0.001
    - **Initial learning rate**: 10^{-7}
4. **RNN architectures**: unidirectional and bidirectional

5. **Number of neurons and hidden layers**:
    - [64]
    - [64, 64]
    - [64, 64, 64]
    - [128]
    - [128, 128]
    - [128, 128, 128]
    - [256]
    - [256, 256]
    - [256, 256, 256]
    - [512]
    - [512, 512]
    - [512, 512, 512]
    - [1024]
    - [1024, 1024]
    - [1024, 1024, 1024]
    - [2048]
    - [2048, 2048]
    - [2048, 2048, 2048]

Appendix C

Research Project to Collect Brazilian Patient Data

The following pages include the health research project that was approved by the Brazilian National Research Ethics Committee.

Language: Brazilian Portuguese.

Universidade Federal do Rio Grande do Norte
Hospital Universitário Onofre Lopes
Laboratório de Inovação Tecnológica em Saúde

Projeto de Pesquisa

Desenvolvimento de um Sistema de Apoio à Decisão Clínica Aplicado ao Prognóstico da Esclerose Lateral Amiotrófica

Pesquisador
Fabiano Papaiz

Equipe
Ricardo Alexsandro de Medeiros Valentim
Mário Emílio Dourado
Joel Perdiz Arrais
Anna Paula Paranhos Miranda Covaleski
Isabella Araujo Mota
Marcela Câmara Machado Costa
Isaac Holanda Mendes Maia
Francisco Marcos Bezerra da Cunha
Antonio Higor Freire de Morais
Daniele Montenegro da Silva Barros
Emanuela Coriolano Fidelix
Paulo Santiago de Morais Brito
Ingridy Marina Pierre Barbalho
Felipe Ricardo dos Santos Fernandes

Natal/RN, 2021

Resumo

O prognóstico de doenças raras e complexas, como a Esclerose Lateral Amiotrófica (ELA), é um grande desafio e representa uma tarefa essencial para melhorar a qualidade de vida dos pacientes. Estudos recentes demonstraram que existem marcadores biológicos (biomarcadores) e algoritmos de Aprendizado de Máquina úteis para auxiliar no prognóstico da ELA, incluindo a predição do tempo de sobrevida e da progressão da doença. Apesar dos esforços de pesquisa neste campo, ainda existe a necessidade de disponibilizar o conhecimento produzido de forma mais célere para os profissionais de saúde, visando auxiliá-los no seu trabalho diário no ambiente clínico. Este estudo propõe o desenvolvimento de um Sistema de Apoio à Decisão Clínica (SADC) como uma forma de superar esta necessidade. O sistema proposto irá fornecer estimativas sobre eventos futuros da vida do paciente, como o tempo de sobrevida, o estágio da doença e quando o suporte respiratório e nutricional serão requeridos. Dados sobre os biomarcadores dos pacientes recrutados serão processados e então inseridos no SADC para compor a Base de Conhecimento do sistema. O Mecanismo de Inferência do SADC será composto por modelos de Aprendizado de Máquina, os quais serão desenvolvidos de modo a aprenderem a partir desta Base de Conhecimento. Os biomarcadores serão analisados para descobrir quais deles são os mais relevantes para auxiliar no prognóstico desta doença. No futuro, os profissionais de saúde irão interagir com o SADC inserindo as informações sobre seus novos pacientes e visualizando as predições em um Painel de Indicadores. Desta forma, esperamos criar uma valiosa solução computacional para auxiliar os profissionais de saúde a tomarem melhores decisões, melhorando a qualidade de vida dos pacientes. Também pretendemos compartilhar a base de dados criada neste estudo para auxiliar futuras pesquisas sobre esta doença.

Palavras-chave: informática em saúde, esclerose lateral amiotrófica, prognóstico, sistema de apoio à decisão clínica, aprendizado de máquina.

Conteúdo

Lista de Figuras

Lista de Tabelas

Lista de Acrônimos

ALSFRS-R Escala Funcional da ELA revisada

AM Aprendizado de Máquina

BIOMARCADORES Marcadores biológicos

ELA Esclerose Lateral Amiotrófica

HUOL Hospital Universitário Onofre Lopes

LAIS Laboratório de Inovação Tecnológica em Saúde

PRO-ACT Pooled Resource Open-access ALS Clinical Trials

SADC Sistema de Apoio à Decisão Clínica

UFRN Universidade Federal do Rio Grande do Norte (Brasil)

1 Introdução

Na Medicina, prognóstico é um termo relacionado com a predição da ocorrência de futuros eventos relacionados à saúde dos pacientes. Isto inclui realizar estimativas sobre: a possibilidade de uma pessoa vir a desenvolver uma certa doença, quão rápido os sintomas irão piorar ao longo do tempo, quais complicações estão associadas com a doença e o tempo de sobrevida do paciente [1]. O prognóstico baseia-se em estimativas e, portanto, ele representa uma probabilidade (e não uma certeza) de que a doença irá seguir um determinado curso. Doenças complexas e raras, como a Esclerose Lateral Amiotrófica (ELA), tornam muito desafiador a realização do prognóstico principalmente por apresentarem um desenvolvimento bastante heterogêneo entre seus pacientes. Tal fato aumenta a necessidade de realizar pesquisas e de desenvolver modelos de prognósticos mais complexos para se alcançar resultados de predição mais precisos. Este estudo está focado no campo de prognóstico da ELA.

A ELA é uma doença rara e progressiva que afeta os neurônios do sistema motor dos humanos. Até hoje não existe cura para esta doença e suas causas são desconhecidas. Gradualmente, ela leva seus pacientes à paralisia e, inevitavelmente, à morte [2] (Figura 1). Seu diagnóstico demora em média 18 meses e baseia-se na análise clínica, na progressão dos sintomas e na exclusão de outras doenças neurológicas. A maioria dos pacientes morrem num intervalo de 3 a 5 anos após o início dos sintomas [3, 4, 5]. A ELA é clinicamente heterogênea, com múltiplos fenótipos associados e apresentando diferentes conjuntos de sintomas iniciais, progressão da doença e tempos de sobrevida entre seus pacientes (Figura 2). Tal complexidade faz com que seja um grande desafio a realização do prognóstico, sendo fundamental a descoberta de marcadores biológicos (biomarcadores) para entender melhor esta doença e, assim, melhorar a qualidade de vida dos pacientes. Os biomarcadores compreendem um conjunto de elementos que podem ser medidos a partir de um ser humano, como dados clínicos, exames laboratoriais, imagens médicas etc. Alguns estudos recentes sobre a ELA demonstraram o potencial de se usar um conjunto de biomarcadores para realizar predições sobre o tempo de sobrevivência dos pacientes e da progressão desta doença [6, 7, 8].

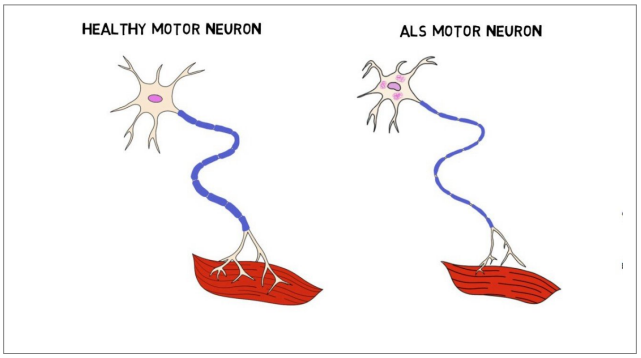


FIGURA 1: Comparação entre um neurônio saudável (esquerda) e um neurônio afetado pela doença da ELA (direita). Fonte [9].

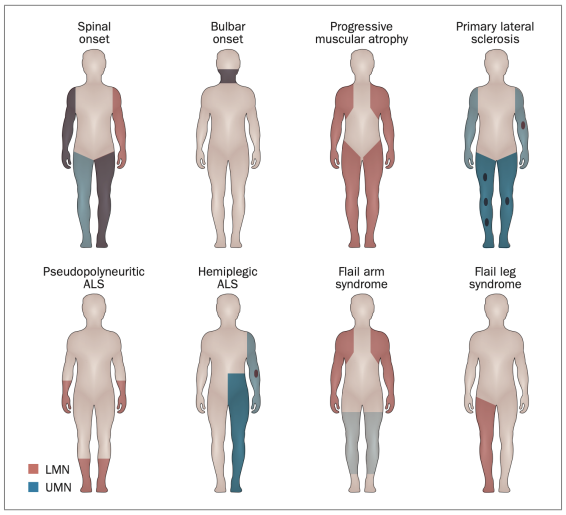


FIGURA 2: Representação dos diferentes conjuntos de sintomas inicias da ELA, com envolvimento do Neurônio Motor Superior (UMN) e Inferior (LMN). Fonte [5].

Pesquisas que aplicam técnicas de Inteligência Artificial, como algoritmos de Aprendizado de Máquina (AM), vêm obtendo bons resultados nas tarefas de diagnóstico e prognóstico de doenças, principalmente na área da Oncologia [10] [11]. A área de AM foca no desenvolvimento de programas de computador capazes de aprender utilizando informações previamente existentes e sem serem explicitamente programados para isso. Os algoritmos de AM podem extrair informações destes dados e transformá-los em conhecimento para solucionarem diferentes categorias de problemas, como classificação, regressão e agrupamento [12, 13]. Com respeito à ELA, diversos estudos estão utilizando diferentes biomarcadores e algoritmos de AM para realizarem predições sobre o tempo de sobrevivência e a progressão da doença com relativo sucesso [7, 14, 15, 16].

Apesar dos esforços de pesquisa sobre a ELA, ainda existe uma necessidade de tornar o conhecimento produzido disponível mais facilmente no ambiente clínico. O desenvolvimento de um Sistema de Apoio à Decisão Clínica (SADC) focado nos problemas da ELA é uma alternativa para superar esta lacuna. Os SADC's são soluções computacionais projetadas para ajudar os profissionais de saúde a tomarem decisões mais adequadas e oportunas em relação aos seus pacientes. Isto é realizado através da vinculação das informações atuais sobre um paciente com aquelas anteriormente coletadas de outros pacientes. Desta forma, um SADC representa uma valiosa ferramenta para auxiliar no processo de tomada de decisão baseando-se nas informações adquiridas ao longo do tempo [17, 18, 19]. O funcionamento de um SADC compreende os seguintes componentes: a Base de Conhecimento (SADC-BC), o Mecanismo de Inferência (SADC-MI) e as informações atuais do paciente (Figura 3). A SADC-BC representa a experiência prévia e normalmente contém dados sobre os biomarcadores mais relevantes dos pacientes. O SADC-MI representa o "cérebro do sistema" e pode ser desenvolvido utilizando-se algoritmos de AM que acessam as informações contidas na SADC-BC. Por fim, os usuários irão inserir informações

atuais sobre um paciente para visualizarem as predições e/ou recomendações relacionadas a ele. Até a data deste documento, nenhum estudo relacionado com o desenvolvimento de um SADC aplicado ao prognóstico da ELA foi encontrado na literatura.

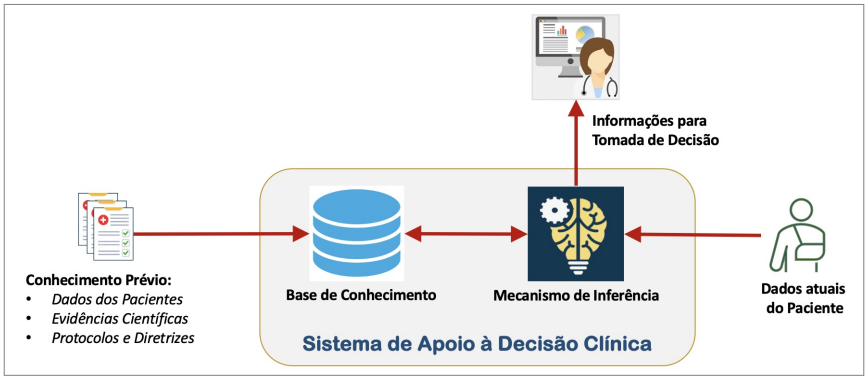


FIGURA 3: Arquitetura geral de um Sistema de Apoio à Decisão Clínica. Fonte: próprio autor.

O objetivo primário desta pesquisa é desenvolver um Sistema de Apoio à Decisão Clínica (SADC) para auxiliar os profissionais de saúde brasileiros em suas tarefas relacionadas ao prognóstico da ELA. Como objetivo secundário, esta pesquisa pretende analisar as informações obtidas sobre os pacientes para traçar uma visão geral e atualizada acerca das características da ELA no Brasil, comparando os resultados obtidos com aqueles já reportados por estudos de outros países. Consequentemente, pretende-se criar uma Base de Dados Nacional sobre biomarcadores prognósticos da ELA, a qual possa ser compartilhada e utilizada em futuras pesquisas brasileiras sobre esta doença.

Para guiar este estudo no sentido de cumprir os objetivos mencionados anteriormente, as seguintes perguntas de pesquisa e hipóteses foram definidas:

Pergunta 1: *É possível desenvolver um Sistema de Apoio à Decisão Clínica e torná-lo disponível no ambiente clínico para melhorar o processo de tomada de decisão sobre o prognóstico dos pacientes com ELA?*

> Hipótese 1: *Os biomarcadores de prognóstico atualmente coletados durante a prática clínica nos Centros de Pesquisa brasileiros são suficientes pra melhorar o processo de tomada de decisão sobre os pacientes com ELA.*

Pergunta 2: *Existem diferenças entre as características da ELA apresentadas no Brasil com aquelas já reportadas em estudos de outros países?*

> Hipótese 2: *As características da ELA no Brasil são similares às reportadas em outros países.*

2 Justificativa

Apesar das inúmeras pesquisas realizadas com relação ao prognóstico da ELA, ainda existe uma necessidade de disponibilizar o conhecimento produzido nestes estudos de forma mais célere, automatizada e simplificada para os profissionais de saúde no ambiente clínico. O desenvolvimento de um Sistema de Apoio à Decisão Clínica (SADC) representa uma alternativa para preencher esta lacuna e facilitar a disseminação do conhecimento e, dessa forma, facilitar o seu acesso pelos profissionais de saúde. Até a data deste documento, nenhuma pesquisa relacionada com o desenvolvimento de um SADC para auxiliar no prognóstico da ELA foi encontrada na literatura. Tal fato evidencia o caráter inovador do tema proposto por esta pesquisa.

Para uma melhor compreensão desta doença tão complexa é essencial possuir dados sobre os pacientes e, assim, poder realizar estudos relevantes sobre seu prognóstico. No Brasil, há uma necessidade de se obter informações mais detalhadas e atualizadas sobre a ELA, havendo deficiência de base de dados disponíveis para auxiliar as pesquisas neste campo. Esta pesquisa se propõe a acessar as informações sobre os biomarcadores prognósticos nos prontuários dos pacientes e armazená-las em uma base de dados centralizada e anonimizada (sem dados sensíveis), a qual possa ser compartilhada no futuro para auxiliar e alavancar as pesquisas relacionadas com o prognóstico desta doença.

Também não está claro se bases de dados de outros países podem ser combinadas com os dados dos pacientes brasileiros para formar uma base ainda mais significante. Os algoritmos de Aprendizado de Máquina (AM) são capazes de aprender a partir dos dados existentes e, dessa forma, quanto mais dados estiverem disponíveis, maior será o aprendizado. Portanto, julgamos importante descobrir se a combinação destas bases podem auxiliar na criação de um algoritmo de AM capaz de realizar predições prognósticas mais precisas.

Concluindo, ao auxiliar os profissionais de saúde a tomarem decisões mais precisas e oportunas sobre o prognóstico da ELA, espera-se que a principal contribuição desta pesquisa seja melhorar a qualidade de vida dos seus pacientes.

3 Objetivos

O objetivo primário desta pesquisa é desenvolver um Sistema de Apoio à Decisão Clínica (SADC) para auxiliar os profissionais de saúde brasileiros em suas tarefas relacionadas ao prognóstico da ELA. O SADC proposto irá fornecer informações úteis sobre predições prognósticas, incluindo o tempo de sobrevida e o momento quando os suportes respiratório e nutricional são requeridos (Figura 4). O sistema também irá permitir que os profissionais de saúde possam registrar informações sobre seus pacientes (sem inserir dados sensíveis) e, dessa forma, a sua Base de Conhecimento (SADC-BC) será constantemente alimentada ao longo do tempo com informações sobre novos pacientes e os já em acompanhamento. Serão armazenados os biomarcadores estáticos e longitudinais presentes nos prontuários dos pacientes, incluindo dados demográficos, clínicos, laboratoriais, respiratórios, nutricionais e demais biomarcadores que possam ser obtidos de acordo com a realidade apresentada pelos Centros de Pesquisa em Saúde no Brasil.

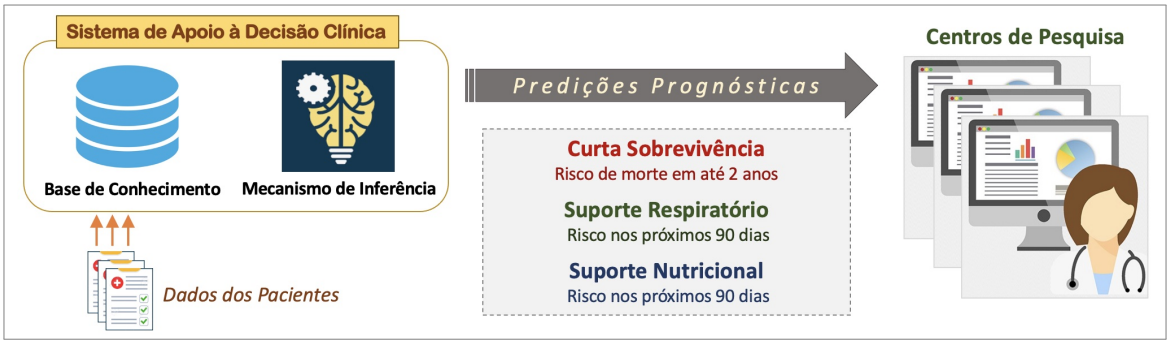


FIGURA 4: Visão geral do SADC proposto nesta pesquisa e as predições prognósticas que serão fornecidas aos profissionais de saúde dos Centros de Pesquisa participantes. Fonte: próprio autor.

Como objetivo secundário, esta pesquisa pretende analisar as informações obtidas sobre os pacientes para traçar uma visão geral e atualizada das características da ELA no Brasil em relação aos biomarcadores associados com um pior prognóstico. Os resultados obtidos serão então comparados com aqueles já reportados em estudos realizados por outros países.

Consequentemente, pretende-se criar uma Base de Dados Nacional sobre biomarcadores prognósticos da ELA, a qual possa ser compartilhada e utilizada em futuras pesquisas brasileiras sobre esta doença.

3.1 Objetivos Específicos

Para atingir os objetivos gerais desta pesquisa, os seguintes objetivos específicos foram definidos:

- Revisão da Literatura do Estado da Arte sobre:
  - Doença ELA e seus biomarcadores prognósticos;

  - Algoritmos de Aprendizado de Máquina (AM) aplicados ao prognóstico da ELA;
  - Sistemas de Apoio à Decisão Clínica (SADC);

- Definir o conjunto de biomarcadores prognósticos que serão analisados, os quais possam ser obtidos dos prontuários dos pacientes de acordo com a realidade apresentada pelos Centros de Pesquisa em Saúde brasileiros;
- Convidar os Centros de Pesquisa que realizam estudos sobre ELA para participarem desta pesquisa. Isto se faz necessário para que se atinja o número mínimo desejado de pacientes com ELA ($n >= 200$);
- Obter os biomarcadores prognósticos dos prontuários dos pacientes nos Centros de Pesquisa participantes;
- Projetar a Base de Conhecimento (SADC-BC) do sistema;
- Pesquisar sobre base de dados públicas que contenham informações sobre os biomarcadores escolhidos, as quais serão inseridas na SADC-BC e utilizadas como dados auxiliares para o desenvolvimento do Mecanismo de Inferência (SADC-MI) do sistema durante a etapa inicial deste estudo;
- Projetar e desenvolver o módulo de Prontuário Eletrônico para registrar as informações sobre os biomarcadores dos pacientes (sem inserir dados sensíveis);
- Conduzir estudos práticos utilizando a SADC-BC para determinar quais algoritmos de Aprendizado de Máquina atingem os melhores resultados na realização das predições prognósticas (curto tempo de sobrevida e necessidade de suporte respiratório e nutricional);
- Projetar e desenvolver o Mecanismo de Inferência (SADC-MI) do sistema;
- Projetar e desenvolver o módulo de Apoio à Decisão Clínica do sistema, contendo o Painel de Indicadores Prognósticos;
- Validar e avaliar o SADC desenvolvido;
- Realizar análise utilizando as informações obtidas para reportar as características gerais da ELA no Brasil em relação aos biomarcadores associados com um pior prognóstico.

4 Metodologia

4.1 Desenho do Estudo

A pesquisa proposta compreende um estudo clínico, prospectivo, observacional, longitudinal e com duração de aproximadamente 4 anos. Este estudo será coordenado e realizado pelo Laboratório de Inovação Tecnológica em Saúde (LAIS), do Hospital Universitário Onofre Lopes (HUOL), da Universidade Federal do Rio Grande do Norte (UFRN), tendo o auxílio dos Centros de Pesquisa em Saúde convidados para a obtenção dos biomarcadores prognósticos dos pacientes.

4.2 Participantes

Coorte

A coorte irá ser composta por pacientes de ambos os sexos, maiores de 18 anos e diagnosticados com ELA de acordo com o critério do *El Escorial Revisado* [20], onde serão incluídos aqueles classificados como: (i) *ELA Clinicamente Definida*, (ii) *ELA Clinicamente Provável*, (iii) *ELA Clinicamente Provável com Apoio Laboratorial* e (iv) *ELA Clinicamente Possível*. Pretende-se formar uma coorte contendo pelo menos 200 pacientes. Poderão ser incluídos os pacientes recentemente diagnosticados e aqueles que já estiverem sendo acompanhados, desde que estes últimos possuam informações suficientes sobre os biomarcadores prognósticos definidos (Tab. 1) e com data de diagnóstico não anterior a 02 anos da data do início desta pesquisa. Os pacientes poderão ser inseridos nesta pesquisa até junho/2024.

Todos os pacientes serão primeiramente convidados a colaborar com o estudo de forma individual e voluntária, conforme regulamentado pela Resolução CNS 466/2012. Após apresentação do projeto, objetivos e da forma como cada participante poderá contribuir, cada indivíduo inicialmente recrutado terá total liberdade e autonomia para decidir sobre a sua participação no presente estudo. Aqueles que vierem a aceitar, deverão assentir a sua participação mediante assinatura do Termo de Consentimento Livre e Esclarecido (TCLE), previamente aprovado pelo Comitê de Ética em Pesquisa do HUOL/UFRN. Para aqueles pacientes voluntários incapazes de assinar o TCLE, seja pela gravidade da doença ou por outro motivo, o termo poderá ser assinado pelos seus respectivos representantes legais. Os pacientes anuentes permitirão o acesso dos pesquisadores que os atendem aos seus prontuários, sendo isto necessário para a obtenção dos biomarcadores prognósticos definidos para este estudo (Tab. 1).

Todos os participantes e seus representantes legais poderão retirar o seu consentimento em qualquer fase do desenvolvimento desta pesquisa, sem que haja qualquer prejuízo ao seu acompanhamento médico e tratamento. De modo semelhante, nenhum paciente convidado que não concordar em participar sofrerá qualquer tipo de prejuízo ou discriminação ao seu acompanhamento médico e tratamento.

Centros de Pesquisa em Saúde

Como a ELA é uma doença rara, foram convidados outros Centros de Pesquisa em Saúde no Brasil que realizam estudos sobre esta doença e com interesse em participar deste estudo. Isto fez-se necessário para que o número mínimo esperado de pacientes seja alcançado.

Como pré-requisito para um Centro estar apto a participar desta pesquisa, este deverá poder obter os biomarcadores escolhidos a partir dos prontuários dos seus pacientes, sendo obrigatório a obtenção de pelo menos 80% dos biomarcadores descritos na Tabela 1. Para cada Centro de Pesquisa participante, serão solicitadas declarações de anuência do pesquisador responsável e do responsável pela Instituição, onde estes declaram concordar em participar da pesquisa.

Os Centros participantes desta pesquisa, juntamente com seus respectivos responsáveis pela obtenção do consentimento e esclarecimento dos participantes, estão listados a seguir:

- Hospital Universitário Onofre Lopes da Universidade Federal do Rio Grande do Norte (Natal/RN). Responsável: Dr. Mario Emílio Teixeira Dourado Jr. (CPF: 438.358.724-00);
- Fundação Bahiana para Desenvolvimento das Ciências (Salvador/BA). Responsável: Dra. Marcela Câmara Machado Costa (CPF: 941.348.985-87);
- Hospital Geral de Fortaleza (Fortaleza/CE). Responsável: Dr. Isaac Holanda Mendes Maia (CPF: 007.119.073-23);
- Hospital das Clínicas da Universidade Federal de Pernambuco (Recife/PE). Responsável: Dra. Anna Paula Paranhos Miranda Covaleski (CPF: 038.225.024-99);
- Hospital Universitário Walter Cantídio da Universidade Federal do Cariri (Barbalha/CE). Responsável: Dr. Francisco Marcos Bezerra da Cunha (CPF: 093.990.464-00).

Equipe de Pesquisa

A nossa equipe de trabalho conta com a colaboração de profissionais da área da saúde e cientistas da computação. Todos os integrantes listados na equipe desta pesquisa assentiram a sua participação, conforme explicitado nos respectivos termos de participação anexados a este projeto.

4.3 Coleta de Dados

Nesta pesquisa serão utilizados dados primários e secundários. Os dados primários compreenderão o conjunto de biomarcadores que serão obtidos a partir dos prontuários dos pacientes recrutados, conforme descrito na Tabela 1, os quais serão armazenados na Base de Conhecimento do SADC proposto. Estes biomarcadores foram definidos com base na literatura, no Protocolo Clínico e Diretrizes Terapêuticas da ELA (Ministério da Saúde) [21] e com a ajuda dos profissionais de saúde do HUOL que atendem pacientes com ELA. Na escolha destes biomarcadores

também foi levado em consideração a possibilidade de serem obtidos de acordo com a realidade dos Centros de Pesquisa de Saúde brasileiros e, assim, possibilitar a participação de uma quantidade mais significante de centros. O registro dos biomarcadores longitudinais durante as avaliações periódicas dos pacientes deverão ocorrer em intervalos não superiores a 4 meses. Dessa forma, a frequência em que haverá a consulta aos dados dos prontuários dos participantes será a cada 4 meses, podendo esta consulta se repetir até o mês de junho de 2024 (prazo final para a coleta de dados deste estudo). Todos os biomarcadores obtidos serão utilizados apenas na execução deste projeto. Os dados primários coletados serão registrados através de formulários que foram criados e organizados para atender aos objetivos desta pesquisa. Posteriormente, os dados coletados serão inseridos no sistema *Registro Nacional da ELA* [22], desenvolvido pelo LAIS/HUOL/UFRN, através do Módulo de Prognóstico.

Em hipótese alguma a obtenção dos biomarcadores, necessária para execução do presente estudo, deverá atrapalhar as consultas e exames de rotina de cada paciente, mesmo que este ou seu responsável legal tenha assentido a sua participação mediante assinatura do TCLE.

Os dados secundários serão extraídos de bases de dados públicas atualmente disponíveis para utilização, como a base denominada PRO-ACT [23], as quais possuam informações sobre os biomarcadores definidos. O principal motivo de utilização dos dados secundários é para poder permitir o desenvolvimento antecipado do SADC proposto, quando os dados primários ainda não estarão disponíveis ou em quantidade não significativa. Isto também permitirá a comparação de performance e precisão do Mecanismo de Inferência do sistema com base na utilização dos dados primários e secundários separadamente. Permitirá ainda a análise destas duas bases de dados em conjunto, verificando se estas podem ser combinadas para gerar uma base de dados ainda mais relevante.

Proteção dos Dados

Todas as pesquisas que envolvem dados médicos sobre pacientes devem ter uma atenção especial em relação às informações pessoais confidenciais (dados sensíveis). Quaisquer dados sobre um paciente que possam ser usados para identificá-lo devem estar protegidos contra acesso não autorizado ou divulgação acidental.

Os dados coletados serão inseridos no sistema *Registro Nacional da ELA* [22], desenvolvido pelo LAIS/HUOL/UFRN, através do Módulo de Prognóstico. Para cada paciente voluntário será gerado um código identificador único, sendo que apenas os pesquisadores responsáveis por atender o paciente poderão identificá-lo individualmente. Somente os pesquisadores participantes desta pesquisa poderão ter acesso aos dados e resultados, minimizando assim a possibilidade de quebra de sigilo ou extravio de documentos. Obrigatoriamente, qualquer arquivo com dados desta pesquisa que vier a ser disponibilizado para *download* pelos pesquisadores participantes não conterá informações sensíveis sobre os pacientes.

Os resultados provenientes deste estudo serão confidenciais e divulgados apenas em congressos ou publicações científicas, não havendo divulgação de nenhum tipo de dado que possa

Tabela 1: Biomarcadores que serão utilizados neste estudo, os quais serão coletados a partir dos prontuários dos pacientes.

Demográficos e Clínicos		
Biomarcador	**Tipo**	**Momento do Registro**
Data de Nascimento	data	1ª Avaliação
Gênero (Masculino/Feminino)	categórico	1ª Avaliação
Etnia	categórico	1ª Avaliação
Altura (m)	numérico	1ª Avaliação
Peso (kg)	numérico	Longitudinalmente
Índice de Massa Corporal (IMC)	numérico	Longitudinalmente
Data do Início dos Sintomas	data	1ª Avaliação
Local de Início	categórico	1ª Avaliação
Data do Diagnóstico	data	1ª Avaliação
Critério El Escorial Revisado	categórico	1ª Avaliação
Fenótipo Motor	categórico	1ª Avaliação
Histórico Familiar	categórico	1ª Avaliação
Fumante	categórico	1ª Avaliação
Maços de Cigarro por Dia	numérico	1ª Avaliação
Data do Óbito	data	Quando ocorrer
Estadiamento da Doença (*King's College*)	categórico	Longitudinalmente
Mutação Genética Identificada	categórico	Quando Ocorrer
Espectro da Demência Frontotemporal	categórico	Quando ocorrer
Medicação		
Riluzole	categórico	Quando ocorrer
Edaravone	categórico	Quando ocorrer
Funcionais, Respiratórios e Nutricionais		
Escala Funcional da ELA Revisada (ALSFRS-R)	numérico	Longitudinalmente
Capacidade Vital Forçada	numérico	Longitudinalmente
Capacidade Vital Lenta	numérico	Longitudinalmente
Pressão Inspiratória Máxima	numérico	Longitudinalmente
Pressão Expiratória Máxima	numérico	Longitudinalmente
Pressão Inspiratória Medida na Narina	numérico	Longitudinalmente
Ventilação Não-invasiva	categórico	Quando ocorrer
Traqueostomia	categórico	Quando ocorrer
Cough Assist (Máquina da Tosse)	categórico	Quando ocorrer
Gastrostomia	categórico	Quando ocorrer
Exames Laboratoriais		
Ácido Úrico	numérico	Longitudinalmente
Alanina Aminotransferase (ALT ou TGP)	numérico	Longitudinalmente
Albumina	numérico	Longitudinalmente
Aspartato Aminotransferase (AST ou TGO)	numérico	Longitudinalmente
Cálcio	numérico	Longitudinalmente
Colesterol HDL	numérico	Longitudinalmente
Colesterol LDL	numérico	Longitudinalmente
Colesterol Total	numérico	Longitudinalmente
Cloro	numérico	Longitudinalmente
Creatina Quinase (CK)	numérico	Longitudinalmente
Creatinina	numérico	Longitudinalmente
Ferritina	numérico	Longitudinalmente
Fosfatase Alcalina	numérico	Longitudinalmente
Fosfato (Fósforo)	numérico	Longitudinalmente
Glicose	numérico	Longitudinalmente
Potássio	numérico	Longitudinalmente
Triglicerídeos	numérico	Longitudinalmente
Vitamina B12	numérico	Longitudinalmente
Sumário de Urina - *pH*	numérico	Longitudinalmente
Hemograma - *Hemoglobina* - *Hematócrito* - *Leucócitos* - *Neutrófilos Segmentados* - *Linfócitos* - *Monócitos*	numérico	Longitudinalmente

identificar qualquer um dos participantes individualmente.

Os dados armazenados ficarão sob responsabilidade do pesquisador Fabiano Papaiz (CPF: 128.685.028-22) e do senhor Rodrigo Dantas da Silva (CPF: 059.989.104-19), Administrador de Banco de Dados, do Núcleo de Infraestrutura, do Laboratório de Inovação Tecnológica em Saúde, do Hospital Universitário Onofre Lopes, da Universidade Federal do Rio Grande do Norte. Esses dados ficarão armazenados por um período de 5 anos após o término desta pesquisa.

Considerando a Lei Geral de Proteção de Dados, informamos que os papeis de Controlador e Operador serão assumidos pelo Laboratório de Inovação Tecnológica em Saúde (LAIS). Os Encarregados de Dados nomeados pelo LAIS são: Luís Eduardo Germano Evangelista (CPF: 075.935.124-44), Rodrigo Dantas da Silva (CPF: 059.989.104-19) e Jadson Fábio Santos Júnior (CPF: 790.760.255-68). Sobre esse assunto, o participante/responsável poderá tirar suas dúvidas entrando em contato com o Núcleo de Segurança da Informação (NSI), do Laboratório de Inovação Tecnológica em Saúde (LAIS), do Hospital Universitário Onofre Lopes (HUOL), na Av. Nilo Peçanha, nº 620, Petrópolis, Natal/RN, e-mail nsi@lais.huol.ufrn.br, telefone (84) 3342-5249, horário de atendimento das 09h às 17h. O NSI tem como principal atribuição garantir que as informações tratadas no âmbito do LAIS estejam seguras, observados os pilares da confidencialidade, integridade, disponibilidade e autenticidade.

4.4 Análise de Dados

A análise dos dados obtidos será realizada seguindo as seguintes etapas: (i) Análise Preparatória, (ii) Análise Associativa e (iii) Análise Preditiva. Tais etapas serão aplicadas tanto nos dados primários como nos secundários. Durante a Análise Preparatória, primeiro será realizada a limpeza dos dados em busca de erros, inconsistências, dados faltantes e informações irrelevantes de modo que estes itens não venham afetar o resultado da análise. Em seguida serão analisadas as necessidades de transformação e expansão dos valores das variáveis (*features*), de remoção dos valores atípicos (*outliers*) e de balanceamento dos dados.

A etapa seguinte, Análise Associativa, terá como objetivo utilizar métodos estatísticos para identificar diferenças, correlações e relacionamentos existentes nos dados e que possam ser úteis para definir grupos distintos de pacientes, encontrar variáveis que expliquem a variação de uma outra variável e identificar padrões de relacionamento presentes nos dados. Neste ponto, é esperado que sejam identificados aqueles biomarcadores que mais contribuem (positiva ou negativamente) para o fornecimento das predições prognósticas definidas nesta pesquisa, sendo: curta sobrevivência e a necessidade de suporte respiratório e/ou nutricional.

Por fim, na etapa de Análise Preditiva serão criados algoritmos de Aprendizagem de Máquina que sejam capazes de realizar as predições prognósticas com base nos biomarcadores selecionados na etapa anterior. Tais algoritmos serão treinados e validados utilizando os dados passados (inseridos na Base de Conhecimento) para então serem capazes de fornecer predições a partir da entrada dos dados atuais de um paciente.

Métodos Estatísticos

Nos dados teremos a presença de variáveis estáticas e longitudinais podendo estas serem categóricas, contínuas ou *time-to-event* (tempo até um evento ocorrer). Os métodos estatísticos utilizados irão depender do tipo e da distribuição (normal ou não-normal) de cada variável, o que somente poderá ser identificado após obtermos um número significativo de amostras ($n \geq 50$).

Em relação à predição sobre *Curta Sobrevivência*, iremos aplicar os métodos visando identificar 02 grupos distintos de pacientes com base na data de início dos sintomas, sendo: (i) pacientes com sobrevida de até 2 anos e (ii) pacientes com sobrevida além de 2 anos. Neste caso, os métodos utilizados serão de acordo com o tipo e a distribuição de cada variável, conforme descrito na Tabela 2. Este mesmo procedimento será aplicado às demais predições (*Suporte Respiratório* e *Nutricional*).

Tabela 2: Métodos estatísticos utilizados para a predição prognóstica de *Curta Sobrevivência* de acordo com o tipo da análise, tipo de dado e distribuição das variáveis.

Análise	**Tipo**	**Distribuição**	**Métodos Estatísticos**
Diferença	Contínua	Normal	*Student's t-test* ou *One-Way ANOVA*
		Não-Normal	*Mann–Whitney test* ou *Kruskal-Wallis test*
	Categórica	- -	*Fisher's exact test* ou *Chi-square test*
	Time-to-event	- -	*Kaplan–Meier Estimate* e *Log-rank test*
Correlação	Contínua	Normal	*Pearson's correlation*
		Não-Normal	*Spearman's correlation*
	Categórica	- -	*Cohen's Kappa Correlation*
Explicativa	Contínua	Normal	*Linear Regression*
		Não-Normal	
	Categórica	- -	*Logistic Regression*
	Time-to-event	- -	*Cox Proportional Hazard Regression*

Tecnologias Computacionais

No desenvolvimento do SADC proposto nesta pesquisa serão utilizadas apenas tecnologias computacionais disponibilizadas como Software Livre, as quais não necessitam de aquisição de licenças para poderem ser utilizadas. O Sistema Operacional onde as tecnologias utilizadas serão hospedadas será o *Ubuntu Linux* (versão 20.04). Para a modelagem e criação da Base de Conhecimento do sistema será utilizado o Sistema Gerenciador de Banco de Dados *PostgreSQL* (versão 12.4). Para o desenvolvimento de todos os módulos do sistema será utilizada a linguagem de programação *Python* (versão 3.7), juntamente com as bibliotecas *Pandas*, *Scikit-Learn*, *NumPy*, *Statsmodel*, *Scipy*, *Matplotlib* e *Lifelines*. Estas bibliotecas são necessárias para a realização do processo de Extração-Transformação-Carregamento dos dados, para a execução dos métodos estatísticos, geração de gráficos analíticos e para construção dos algoritmos de Aprendizado de Máquina. As interfaces de usuário do sistema (módulo de Registro dos Biomarcadores e o Painel de Indicadores Prognósticos) serão desenvolvidas através do *Django Web Development*

Framework (versão 3.0.8), sendo que os gráficos dos indicadores serão criados com a biblioteca *Highcharts Javascript Charting Library*. Como Ambiente Integrado de Desenvolvimento (IDE, do inglês *Integrated Development Environment*) será utilizado o *PyCharm Community Edition* (versão 2020.3)

4.5 Riscos

Nesta seção estão descritos os riscos aos pacientes envolvidos nesta pesquisa, os quais consideramos como sendo de grau mínimo pois não há riscos físicos e/ou biológicos para o paciente, uma vez que o estudo é meramente observacional, não expondo os voluntários a nenhuma condição adicional nestes tipos de riscos.

Ressaltamos que o levantamento de dados em prontuários não interferirá nos cuidados recebidos pelos pacientes e, dessa forma, a sua concordância em colaborar com o presente estudo não implicará em qualquer prejuízo ou agravamento ao seu estado geral de saúde.

Para minimizar o risco de divulgação de dados confidenciais, o nome do paciente ou qualquer outro dado que possa identificá-lo individualmente será mantido em sigilo. Os pacientes serão identificados por códigos e, com isso, apenas os pesquisadores que os atendem serão capazes de identificá-los pessoalmente. Todas as informações armazenadas digitalmente nas bases de dados não conterão dados confidenciais sobre os pacientes. As informações coletadas serão compartilhadas entre os pesquisadores participantes e utilizadas somente para os fins desta pesquisa.

Poderá ocorrer algum constrangimento aos pacientes no momento do convite para participarem desta pesquisa. Esse constrangimento será minimizado com o convite sendo realizado de forma reservada, com a privacidade do paciente sendo respeitada. Nenhum paciente será coercido a participar desta pesquisa. Aqueles que forem convidados e que não concordarem em participar não sofrerão discriminação e nem qualquer prejuízo em seus acompanhamentos médicos e tratamentos.

Ressaltamos que todos os membros da equipe de pesquisa estão cientes da Lei Geral de Proteção de Dados e trabalharão o máximo para garantir a segurança e o armazenamento seguro de dados pessoais e prontuários, visando sempre evitar a perda ou vazamento desses documentos.

Por fim, os pacientes voluntários que vierem a sofrer qualquer tipo de dano (previsto ou não no TCLE) que seja resultante da sua participação nesta pesquisa, terão direito à indenização e à assistência necessária.

4.6 Benefícios

Os participantes desta pesquisa, assim como os futuros pacientes, poderão se beneficiar dos resultados obtidos à medida que as predições prognósticas forem sendo disponibilizadas aos profissionais de saúde, auxiliando-os no processo de tomada de decisão e, consequentemente, proporcionando uma melhora na qualidade de vida dos seus pacientes.

4.7 Desfecho Primário

Esperamos que esta pesquisa traga contribuições relevantes para o campo de prognóstico de doenças complexas, particularmente em relação à doença da ELA, através dos resultados obtidos e da escrita de artigos e participações em congressos. A seguir, estão melhor detalhados os resultados esperados.

Em primeiro lugar, esperamos que o desenvolvimento do Sistema de Apoio à Decisão Clínica (SADC) possa trazer melhorias no trabalho diário dos profissionais de saúde em relação ao prognóstico da ELA. Através da análise e descoberta dos biomarcadores mais relevantes, o sistema irá fornecer predições pronósticas para que os profissionais possam tomar melhores decisões baseados na ciência e, dessa forma, melhorar a qualidade de vida dos pacientes. Consideramos o SADC proposto como um passo inicial na direção de disponibilizar o conhecimento produzido nas pesquisas para o ambiente clínico no Brasil, permitindo assim a disseminação do conhecimento de forma mais célere, simplificada e abrangente. Ressaltamos que o sistema aqui proposto não tem a intenção de substituir o profissional de saúde, mas sim funcionar como uma ferramenta adicional para auxiliá-lo no processo de tomada de decisão.

Também esperamos fazer contribuições significativas no campo da Inteligência Artificial, aplicada na área da Saúde, através do desenvolvimento de algoritmos de Aprendizado de Máquina capazes de realizar predições mais precisas referentes ao prognóstico da ELA.

Ademais, esperamos facilitar e promover a pesquisa sobre ELA no Brasil através da criação da base de dados sobre biomarcadores prognósticos. Tal base poderá ser utilizada em futuras pesquisas da comunidade científica e, com isso, novos biomarcadores poderão ser adicionados e novas predições poderão ser fornecidas.

Por fim, espera-se que a metodologia utilizada nesta pesquisa possa ajudar outros pesquisadores e servir como base em futuros estudos envolvendo o desenvolvimento de um SADC para o prognóstico de doenças complexas.

5 Cronograma

6 Orçamento

O desenvolvimento desta pesquisa não acarretará custos adicionais para os Centros de Pesquisa participantes, já que os dados serão obtidos diretamente dos prontuários dos pacientes. A tabela a seguir apresenta o orçamento previsto para execução deste projeto. Todo financiamento será proveniente do projeto denominado *Desenvolvimento Científico e Tecnológico Aplicado a Esclerose Lateral Amiotrófica* (revELA), executado pelo LAIS/HUOL/UFRN.

Tabela 3: Orçamento previsto para o projeto.

Tipo de Despesa	**Valor Estimado**	**Observações**
Contratação de bolsista de iniciação científica para apoio ao projeto - Área da Computação	R$ 12.000,00	Bolsa no valor de R$ 1.000,00 mensais, durante 12 meses.
Despesas com passagens, deslocamentos e diárias para realização de reuniões	R$ 20.000,00	
Valor Total	**R$ 32.000,00**	

7 Instrumentos de Pesquisa

Nas páginas a seguir foram inseridos os formulários que foram definidos para coletar os dados primários dos pacientes de acordo com os objetivos desta pesquisa, sendo:

1. Formulário de Avaliação Inicial do Paciente: aplicado na primeira consulta após o aceite do paciente em participar desta pesquisa;
2. Formulário de Avaliação Periódica do Paciente: aplicado nas consultas posteriores do paciente, devendo ocorrer em intervalos não superiores a 4 meses.

Ressaltamos que somente serão coletados os biomarcadores indicados nestes formulários. Qualquer modificação visando incluir novos biomarcadores somente poderá ser adotada após uma nova submissão e aprovação por parte do sistema CEP/CONEP.

Formulário de Avaliação Inicial

• Dados de Identificação do Paciente

Nome: _____ Cód. Identificador: _____

• Dados da Avaliação

Resp. Informações: _____ Data de Realização: ____/____/_____

• Dados Demográficos e Clínicos

Idade: _____ Dt. Nascimento: ____/____/_____ Sexo: ○ Masculino ○ Feminino

Altura: _____ m Peso: _____ kg Etnia: ○ Branca ○ Preta ○ Parda ○ Indígena ○ Amarela

Data do Diagnóstico: ____/____/_____

Critério *El Escorial*: ○ Definida ○ Provável ○ Provável c/ Apoio Laboratorial ○ Possível

Data do Início dos Sintomas: ____/____/_____

Local de Início dos Sintomas:

- ○ Membro superior proximal direito (elevação do braço)
- ○ Membro superior proximal esquerdo (elevação do braço)
- ○ Membro superior distal direito (fraqueza na mão)
- ○ Membro superior distal esquerdo (fraqueza na mão)
- ○ Membro inferior direito (perna)
- ○ Membro inferior esquerdo (perna)
- ○ Falta de ar
- ○ Dificuldades para engolir e engasgos
- ○ Dificuldade para articular palavras e fonação
- ○ Não foi possível determinar

Possui Histórico de ELA na Família?

○ Não ○ Sim, parentesco: ☐ Pai ☐ Mãe ☐ Irmãos ☐ Filhos ☐ Outros: _____

É Fumante?

○ Não ○ Sim, quantidade de maços por dia: _____

- **Iniciou o Uso de Medicamentos?**

☐ RILUZOLE – iniciou em: ____/____/_____ ☐ EDARAVONE – iniciou em: ____/____/_____

- **Iniciou Suporte Respiratório e/ou Nutricional?**

☐ **Ventilação Não-Invasiva (BIPAP)**

☐ Menos de 4 horas p/ dia Iniciada em: ____/____/_____

☐ De 8 a 20 horas Iniciada em: ____/____/_____

☐ Mais de 20 horas Iniciada em: ____/____/_____

☐ Noturna, mais de 4 horas Iniciada em: ____/____/_____

☐ **Traqueostomia**

Iniciada em: ____/____/_____

Foi realizada de urgência? ○ Não ○ Sim

☐ **Cough Assist (Máquina da Tosse)**

Iniciada em: ____/____/_____

☐ **Gastrostomia**

Iniciada em: ____/____/_____

- **Exames Respiratórios**

CVF - Capacidade Vital Forçada……: _____ litros Percentual do Esperado: _____ %

CVL - Capacidade Vital Lenta……..: _____ litros Percentual do Esperado: _____ %

PImáx - Pressão Inspiratória Máxima……: _____ cmH_2O

PEmáx - Pressão Expiratória Máxima……: _____ cmH_2O

SNIP - Pressão Inspiratória Medida na Narina…: _____ cmH_2O

- **Exames Laboratoriais**

Ácido Úrico _____ mg/dL

Albumina _____ g/dL

Cálcio _____ mg/dL

Colesterol HDL _____ mg/dL

Colesterol LDL _____ mg/dL

Colesterol Total _____ mg/dL

Cloro _____ mmol/L

Creatina Quinase (CK) _____ U/L

Creatinina _____ mg/dL

Ferritina _____ ng/mL

Fosfatase Alcalina _____ U/L

Fosfato (Fósforo) _____ mg/dL

Glicose _____ mg/dL

Potássio _____ mEq/L

Triglicerídeos _____ mg/dL

Vitamina B12 _____ pg/mL

Sumário de Urina: pH _____

Transaminase Pirúvica (TGP) _____ U/L

Transaminase Oxalacética (TGO)_____ U/L

Hemograma:

- Hemoglobina _____ g/dL
- Hematócrito _____ %
- Leucócitos _____ mm^3
- Neutrófilos Segmentados _____ % _____ mm^3
- Linfócitos _____ % _____ mm^3
- Monócitos _____ % _____ mm^3

- **Avaliação Funcional** - **ALSFRS-R** (Escala Funcional da ELA Revisada)

1. FALA

- ○ 4 - Processo da fala normal
- ○ 3 - Distúrbio da fala detectável
- ○ 2 - Compreensível com repetição
- ○ 1 - Fala combinada com comunicação não-vocal
- ○ 0 - Perda da utilidade da fala

2. SALIVAÇÃO

- ○ 4 - Normal
- ○ 3 - Insignificante, mas notável o excesso de saliva na boca; podendo ter babas noturnas
- ○ 2 - Excesso de saliva moderada, podendo ter mínimas babas (durante o dia)
- ○ 1 - Excesso acentuado de saliva com alguma baba
- ○ 0 - Baba acentuada; exigindo constante uso de babador ou lenço para boca

3. DEGLUTIÇÃO

- ○ 4 - Normal
- ○ 3 - Problemas precoces para comer, com engasgos ocasionais
- ○ 2 - Alteração na consistência da dieta
- ○ 1 - Necessidade de suplemento alimentar pastoso
- ○ 0 - Nada pela boca, exclusivamente parenteral ou enteral

4. ESCRITA

- ○ 4 - Normal
- ○ 3 - Lentificada ou descuidada, todas as palavras são legíveis
- ○ 2 - Nem todas as palavras são legíveis
- ○ 1 - Capaz de segurar a caneta, mas incapaz de escrever
- ○ 0 - Não é capaz de segurar a caneta

5. MANIPULAÇÃO DE ALIMENTOS E UTENSÍLIOS

a) PACIENTES <u>SEM</u> GASTROSTOMIA

- ○ 4 - Normal
- ○ 3 - Um pouco lento e desajeitado, mas não necessita de ajuda
- ○ 2 - Pode cortar o alimento embora lento e desajeitado; necessita de alguma ajuda
- ○ 1 - Alimentos cortados por outra pessoa, mas alimenta-se sozinho lentamente
- ○ 0 - Precisa ser alimentado

b) PACIENTES <u>COM</u> GASTROSTOMIA (representando mais de 50% da alimentação do paciente)

- ○ 4 - Normal
- ○ 3 - Desajeitado, mas capaz de desempenhar todas as manipulações
- ○ 2 - Alguma ajuda necessária com tampas e fechos
- ○ 1 - Oferece assistência mínima ao cuidador
- ○ 0 - Incapaz de executar qualquer aspecto da tarefa

6. VESTUÁRIO E HIGIENE

- ○ 4 - Normal
- ○ 3 - Independente de auto-cuidado com diminuição do rendimento do esforço
- ○ 2 - Assistência intermitente ou substituição dos métodos
- ○ 1 - Necessita do cuidador para auto-cuidado
- ○ 0 - Dependência total

7. VIRAR NA CAMA E AJUSTAR A ROUPA DE CAMA

- ○ 4 - Normal
- ○ 3 - Um pouco lento ou desajeitado, não necessita de ajuda
- ○ 2 - Pode virar sozinho ou ajustar o lençol, mas com grande dificuldade
- ○ 1 - Tem iniciativa, mas não consegue virar ou ajustar o lençol sozinho
- ○ 0 - Incapaz

8. ANDAR

- ○ 4 - Normal
- ○ 3 - Deambulação precoce dificultada
- ○ 2 - Caminha com assistência
- ○ 1 - Somente movimento funcional não-deambulatório
- ○ 0 - Não apresenta movimentação voluntária das pernas

9. SUBIR ESCADAS

- ○ 4 - Normal
- ○ 3 - Lentidão
- ○ 2 - Ligeiro desequilíbrio ou fadiga
- ○ 1 - Necessita de assistência
- ○ 0 - Não realiza

10. DISPNÉIA

- ○ 4 - Nenhuma
- ○ 3 - Ocorre quando caminha
- ○ 2 - Ocorre quando come, toma banho e se veste
- ○ 1 - Ocorre no repouso, quando sentado ou deitado
- ○ 0 - Dificuldade significativa, considerando suporte mecânico

11. ORTOPNÉIA

- ○ 4 - Nenhuma
- ○ 3 - Alguma dificuldade de dormir, falta de ar, não se utiliza rotineiramente mais que 2 travesseiros
- ○ 2 - Necessita de travesseiros extras para dormir (mais que 2)
- ○ 1 - Pode dormir somente sentado
- ○ 0 - Não consegue dormir

12. INSUFICIÊNCIA RESPIRATÓRIA

- ○ 4 - Nenhuma
- ○ 3 - Uso intermitente do BIPAP
- ○ 2 - Uso contínuo do BIPAP à noite
- ○ 1 - Uso contínuo do BIPAP durante o dia e a noite
- ○ 0 - Ventilação mecânica invasiva por intubação

Formulário de Avaliação Periódica

• Dados de Identificação do Paciente

Nome: _____ Cód. Identificador: _____

• Dados da Avaliação

Resp. Informações: _____ Data de Realização: ____/____/_____

Houve Encerramento do Acompanhamento do Paciente nesta Pesquisa?

○ Não ○ Paciente faleceu em: ____/____/_____ ○ Desistência do paciente em: ____/____/_____

• Peso Atual do Paciente: _____ kg

• Iniciou o Uso de Medicamentos?

☐ RILUZOLE em: ____/____/_____ ☐ EDARAVONE em: ____/____/_____

• Iniciou Suporte Respiratório e/ou Nutricional?

☐ **Ventilação Não-Invasiva (BIPAP)**

☐ Menos de 4 horas p/ dia Iniciada em: ____/____/_____

☐ De 8 a 20 horas Iniciada em: ____/____/_____

☐ Mais de 20 horas Iniciada em: ____/____/_____

☐ Noturna, mais de 4 horas Iniciada em: ____/____/_____

☐ **Traqueostomia**

Iniciada em: ____/____/_____

Foi realizada de urgência? ○ Não ○ Sim

☐ **Cough Assist (Máquina da Tosse)**

Iniciada em: ____/____/_____

☐ **Gastrostomia**

Iniciada em: ____/____/_____

- **Exames Respiratórios**

CVF - Capacidade Vital Forçada……: _____ litros Percentual do Esperado: _____ %

CVL - Capacidade Vital Lenta……..: _____ litros Percentual do Esperado: _____ %

PImáx - Pressão Inspiratória Máxima……: _____ cmH_2O

PEmáx - Pressão Expiratória Máxima……: _____ cmH_2O

SNIP - Pressão Inspiratória Medida na Narina…: _____ cmH_2O

- **Exames Laboratoriais**

Ácido Úrico _____ mg/dL

Albumina _____ g/dL

Cálcio _____ mg/dL

Colesterol HDL _____ mg/dL

Colesterol LDL _____ mg/dL

Colesterol Total _____ mg/dL

Cloro _____ mmol/L

Creatina Quinase (CK) _____ U/L

Creatinina _____ mg/dL

Ferritina _____ ng/mL

Fosfatase Alcalina _____ U/L

Fosfato (Fósforo)_____ mg/dL

Glicose _____ mg/dL

Potássio _____ mEq/L

Triglicerídeos _____ mg/dL

Vitamina B12 _____ pg/mL

Sumário de Urina: pH _____

Transaminase Pirúvica (TGP)_____ U/L

Transaminase Oxalacética (TGO)_____ U/L

Hemograma:

-Hemoglobina _____g/dL

-Hematócrito _____%

-Leucócitos _____mm^3

-Neutróf. Segment._____% _____mm^3

-Linfócitos _____% _____mm^3

-Monócitos _____% _____mm^3

- **Avaliação Funcional** - **ALSFRS-R** (Escala Funcional da ELA Revisada)

1. FALA

- ○ 4 - Processo da fala normal
- ○ 3 - Distúrbio da fala detectável
- ○ 2 - Compreensível com repetição
- ○ 1 - Fala combinada com comunicação não-vocal
- ○ 0 - Perda da utilidade da fala

2. SALIVAÇÃO

- ○ 4 - Normal
- ○ 3 - Insignificante, mas notável o excesso de saliva na boca; podendo ter babas noturnas
- ○ 2 - Excesso de saliva moderada, podendo ter mínimas babas (durante o dia)
- ○ 1 - Excesso acentuado de saliva com alguma baba
- ○ 0 - Baba acentuada; exigindo constante uso de babador ou lenço para boca

3. DEGLUTIÇÃO

- ○ 4 - Normal
- ○ 3 - Problemas precoces para comer, com engasgos ocasionais
- ○ 2 - Alteração na consistência da dieta
- ○ 1 - Necessidade de suplemento alimentar pastoso
- ○ 0 - Nada pela boca, exclusivamente parenteral ou enteral

4. ESCRITA

- ○ 4 - Normal
- ○ 3 - Lentificada ou descuidada, todas as palavras são legíveis
- ○ 2 - Nem todas as palavras são legíveis
- ○ 1 - Capaz de segurar a caneta, mas incapaz de escrever
- ○ 0 - Não é capaz de segurar a caneta

5. MANIPULAÇÃO DE ALIMENTOS E UTENSÍLIOS

a) PACIENTES <u>SEM</u> GASTROSTOMIA

- ○ 4 - Normal
- ○ 3 - Um pouco lento e desajeitado, mas não necessita de ajuda
- ○ 2 - Pode cortar o alimento embora lento e desajeitado; necessita de alguma ajuda
- ○ 1 - Alimentos cortados por outra pessoa, mas alimenta-se sozinho lentamente
- ○ 0 - Precisa ser alimentado

b) PACIENTES <u>COM</u> GASTROSTOMIA (representando mais de 50% da alimentação do paciente)

- ○ 4 - Normal
- ○ 3 - Desajeitado, mas capaz de desempenhar todas as manipulações
- ○ 2 - Alguma ajuda necessária com tampas e fechos
- ○ 1 - Oferece assistência mínima ao cuidador
- ○ 0 - Incapaz de executar qualquer aspecto da tarefa

6. VESTUÁRIO E HIGIENE

- ○ 4 - Normal
- ○ 3 - Independente de auto-cuidado com diminuição do rendimento do esforço
- ○ 2 - Assistência intermitente ou substituição dos métodos
- ○ 1 - Necessita do cuidador para auto-cuidado
- ○ 0 - Dependência total

7. VIRAR NA CAMA E AJUSTAR A ROUPA DE CAMA

- ○ 4 - Normal
- ○ 3 - Um pouco lento ou desajeitado, não necessita de ajuda
- ○ 2 - Pode virar sozinho ou ajustar o lençol, mas com grande dificuldade
- ○ 1 - Tem iniciativa, mas não consegue virar ou ajustar o lençol sozinho
- ○ 0 - Incapaz

8. ANDAR

- ○ 4 - Normal
- ○ 3 - Deambulação precoce dificultada
- ○ 2 - Caminha com assistência
- ○ 1 - Somente movimento funcional não-deambulatório
- ○ 0 - Não apresenta movimentação voluntária das pernas

9. SUBIR ESCADAS

- ○ 4 - Normal
- ○ 3 - Lentidão
- ○ 2 - Ligeiro desequilíbrio ou fadiga
- ○ 1 - Necessita de assistência
- ○ 0 - Não realiza

10. DISPNÉIA

- ○ 4 - Nenhuma
- ○ 3 - Ocorre quando caminha
- ○ 2 - Ocorre quando come, toma banho e se veste
- ○ 1 - Ocorre no repouso, quando sentado ou deitado
- ○ 0 - Dificuldade significativa, considerando suporte mecânico

11. ORTOPNÉIA

- ○ 4 - Nenhuma
- ○ 3 - Alguma dificuldade de dormir, falta de ar, não se utiliza rotineiramente mais que 2 travesseiros
- ○ 2 - Necessita de travesseiros extras para dormir (mais que 2)
- ○ 1 - Pode dormir somente sentado
- ○ 0 - Não consegue dormir

12. INSUFICIÊNCIA RESPIRATÓRIA

- ○ 4 - Nenhuma
- ○ 3 - Uso intermitente do BIPAP
- ○ 2 - Uso contínuo do BIPAP à noite
- ○ 1 - Uso contínuo do BIPAP durante o dia e a noite
- ○ 0 - Ventilação mecânica invasiva por intubação

Referências

[1] MedlinePlus, "Medlineplus." `https://https://medlineplus.gov/genetics/understanding/consult/prognosis/`, 2020. Accessed: 2020-09-13.

[2] O. Hardiman, A. Al-Chalabi, A. Chio, E. M. Corr, G. Logroscino, W. Robberecht, P. J. Shaw, Z. Simmons, and L. H. van den Berg, "Amyotrophic lateral sclerosis," *Nature Reviews Disease Primers*, vol. 3, no. 1, p. 17071, 2017.

[3] P. M. Andersen, S. Abrahams, G. D. Borasio, M. de Carvalho, A. Chio, P. Van Damme, O. Hardiman, K. Kollewe, K. E. Morrison, S. Petri, P.-F. Pradat, V. Silani, B. Tomik, M. Wasner, M. Weber, and The EFNS Task Force on Diagnosis and Management of Amyotrophic Lateral Sclerosis: , "Efns guidelines on the clinical management of amyotrophic lateral sclerosis (mals) – revised report of an efns task force," *European Journal of Neurology*, vol. 19, no. 3, pp. 360–375, 2012.

[4] A. Chiò, M. Pagani, F. Agosta, A. Calvo, A. Cistaro, and M. Filippi, "Neuroimaging in amyotrophic lateral sclerosis: insights into structural and functional changes," *The Lancet Neurology*, vol. 13, no. 12, pp. 1228 – 1240, 2014.

[5] B. Swinnen and W. Robberecht, "The phenotypic variability of amyotrophic lateral sclerosis," *Nature Reviews Neurology*, vol. 10, no. 11, pp. 661–670, 2014.

[6] M. Elamin, P. Bede, A. Montuschi, N. Pender, A. Chio, and O. Hardiman, "Predicting prognosis in amyotrophic lateral sclerosis: a simple algorithm," *Journal of Neurology*, vol. 262, pp. 1447–1454, June 2015.

[7] A. A. Taylor, C. Fournier, M. Polak, L. Wang, N. Zach, M. Keymer, J. D. Glass, D. L. Ennist, and The Pooled Resource Open-Access ALS Clinical Trials Consortium, "Predicting disease progression in amyotrophic lateral sclerosis," *Annals of Clinical and Translational Neurology*, vol. 3, pp. 866–875, Nov. 2016.

[8] H.-J. Westeneng, T. P. A. Debray, A. E. Visser, R. P. A. van Eijk, J. P. K. Rooney, A. Calvo, S. Martin, C. J. McDermott, A. G. Thompson, S. Pinto, X. Kobeleva, A. Rosenbohm, B. Stubendorff, H. Sommer, B. M. Middelkoop, A. M. Dekker, J. J. F. A. van Vugt, W. van Rheenen, A. Vajda, M. Heverin, M. Kazoka, H. Hollinger, M. Gromicho, S. Körner, T. M. Ringer, A. Rödiger, A. Gunkel, C. E. Shaw, A. L. Bredenoord, M. A. van Es, P. Corcia, P. Couratier, M. Weber, J. Grosskreutz, A. C. Ludolph, S. Petri, M. de Carvalho,

P. Van Damme, K. Talbot, M. R. Turner, P. J. Shaw, A. Al-Chalabi, A. Chiò, O. Hardiman, K. G. M. Moons, J. H. Veldink, and L. H. van den Berg, "Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model," *The Lancet Neurology*, vol. 17, pp. 423–433, May 2018.

[9] "Amyotrophic lateral sclerosis (als)." `https://www.healthism.co/amyotrophic-lateral-sclerosis-als/`. Accessed: 2020-08-16.

[10] K. O'Shea, S. J. Cameron, K. E. Lewis, C. Lu, and L. A. Mur, "Metabolomic-based biomarker discovery for non-invasive lung cancer screening: A case study," *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1860, no. 11, Part B, pp. 2682 – 2687, 2016. Systems Genetics - Deciphering the Complex Disease with a Systems Approach.

[11] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8 – 17, 2015.

[12] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers. I," in *Computer Games I* (D. N. L. Levy, ed.), pp. 335–365, New York, NY: Springer New York, 1988.

[13] M. Kubat, *An Introduction to Machine Learning*. Cham: Springer International Publishing, 2017.

[14] Y. El Khoury, N. Collongues, J. De Sèze, V. Gulsari, C. Patte-Mensah, G. Marcou, A. Varnek, A. G. Mensah-Nyagan, and P. Hellwig, "Serum-based differentiation between multiple sclerosis and amyotrophic lateral sclerosis by random forest classification of ftir spectra.," *The Analyst*, vol. 144, pp. 4647–4652, Aug 2019.

[15] L. D'hulst, D. V. Weehaeghe, A. Chiò, A. Calvo, C. Moglia, A. Canosa, A. Cistaro, S. M. Willekens, J. D. Vocht, P. V. Damme, M. Pagani, and K. V. Laere, "Multicenter validation of [18f]-fdg pet and support-vector machine discriminant analysis in automatically classifying patients with amyotrophic lateral sclerosis versus controls," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 19, no. 7-8, pp. 570–577, 2018. PMID: 29862846.

[16] N. Bakkar, T. Kovalik, I. Lorenzini, S. Spangler, A. Lacoste, K. Sponaugle, P. Ferrante, E. Argentinis, R. Sattler, and R. Bowser, "Artificial intelligence in neurodegenerative disease research: use of ibm watson to identify additional rna-binding proteins altered in amyotrophic lateral sclerosis," *Acta neuropathologica*, vol. 135, pp. 227–247, Feb 2018.

[17] E. S. Berner, K. J. Hannah, and M. J. Ball, eds., *Clinical Decision Support Systems*. Health Informatics, New York, NY: Springer New York, 2007.

[18] P. E. Beeler, D. W. Bates, and B. L. Hug, "Clinical decision support systems," *Swiss Medical Weekly*, 2014. Publisher: EMH Swiss Medical Publishers.

[19] Elsevier, *Practical Predictive Analytics and Decisioning Systems for Medicine*. Elsevier, 2015.

[20] B. R. Brooks, R. G. Miller, M. Swash, and T. L. Munsat, "El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis," *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, vol. 1, pp. 293–299, Jan. 2000.

[21] "Clinical protocol and therapeutic guidelines for the als disease." `https://www.saude.gov.br/images/pdf/2020/August/19/Portaria-Conjunta-PCDT-ELA---VERSAO-COMPLETA.pdf`. Accessed: 2020-08-31.

[22] U. LAIS, HUOL, "Registro nacional da ela." `https://revelanos.lais.ufrn.br/`, 2021. Accessed: 2021-05-03.

[23] PRO-ACT, "Pro-act: Pooled resource open-access als clinical trials database." `https://nctu.partners.org/ProACT/Document/DisplayLatest/9`, 2020. Accessed: 2020-08-19.