Deep Learning for Predicting Disease Progression of Clinical Endpoints in ALS

Lucas Silva, Nuno Garcia, Sara C. Madeira

LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal



Motivation

- Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease.
- No cure and a short life expectancy following the start of symptoms.
- Most patients die of respiratory failure within 3-5 years.

Problem

- ALS affects most muscles of the body.
- Several clinical conditions derive from it.
- Non-Invasive Ventilation showed the most promise in terms of extending life expectancy and improving quality of life.

St Ok

Methods

Objective

Our **main objective** is to increase life expectancy and improve quality of life. For that we propose the use of Deep Learning to:

 Predict ALS disease progression within a predefined time window (90, 180 and 365 Days).

Background

- Prognostic models based on predefined time windows to predict the disease progression and the need of NIV treatment in ALS.
- Categorization of patients depending on their stage of disease progression, stratifying them in 3 groups, slow, neutral and fast.
- Use of neural networks (mainly LSTM) to predict the disease progression of the patients in the various time windows.
- Shapley Additive Explanations (SHAP) explaining the deep model was helpful to determine the influence of certain features for the overall prediction.

Data

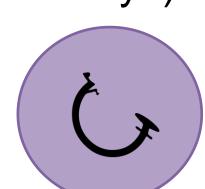
- Tabular Data from the ALS clinic of Centro Hospitalar Universitário de Lisboa Norte, Hospital Santa Maria [1].
- Static information
 - Demographic information about the patient.
- Temporal information Multivariate Time Series
 - Results of a series of clinical assessments of specific tests:
 - Respiratory tests;
 - Neurophysiological data;
 - Questionnaire answers to the ALS Functional Rating Scale.
- For each of the following clinical endpoints, there is predefined time windows (90, 180 and 365 days).



NIV



Aux Comm Device



PEG



Caregiver



Wheelchair

Class Distribution						
Dataset	Evolution	# Patients for the Time Windows in k Days				
		k=90	k=180	k=365		
C1 – Need for NIV	Yes	217	740	1394		
	No	4178	3655	3001		
C2 – Need for na Auxiliary Communication Device	Yes	159	454	859		
	No	5237	4942	4537		
C3 – Need for PEG	Yes	68	267	605		
	No	5983	5784	5446		
C4 – Need for a Caregiver	Yes	341	837	1437		
	No	2750	2254	1654		
C5 – Need for a Wheelchair	Yes	202	589	1242		
	No	4928	4541	3888		

	Preprocessing						
ALS Datasets	Label Encoding 1 • Encoding categorical features • Transforming them into numeric values Data Imputation • KNN Imputation • For features with less than 650 missing values • Remove columns above threshold						
Feature Selection 3 · Maximum Relevance- Minimum Redundancy Algorithm • Find the "minimal-optimal" subset of features • 15 features used as input for the models							
	Training Pipeline						
	ata Normalization Stratified Group Kfold CV						
4 • Rescales between 0 • Leads to a	of the data with MinMax Scaler from its original range to values and 1 a more eficient neural network training results in a better performance model • Divides the data into k groups • Selects one group for testing • The rest are used for training • Attempts to return a stratified fold • However in some cases is impossible, due to highly imbalanced data						
Outpu	Class Resampling SMOTE followed by Random Undersampling Use SMOTE until 60/40 distribution of majority and minority class Random Undersampling to remove majority class instances until 50/50 Only applied on the training sets						
	CNN A 3 layer Multilayer Perceptron 1 D Convolutional neural network Output is a continuous value between O (No Evolution) and 1 (Positive Evolution)						

Results						
Dataset	Model	AUC ROC Mean				
		k=90	k=180	k=365		
C1 – Need for NIV	MLP	0.761	0.746	0.792		
	CNN	0.768	0.766	0.762		
C2 – Need for na Auxiliary Communication Device	MLP	X	0.924	0.924		
	CNN	X	0.921	0.906		
C3 – Need for PEG	MLP	X	X	X		
	CNN	X	X	X		
C4 – Need for a Caregiver	MLP	0.834	0.797	0.824		
	CNN	0.850	0.782	0.800		
C5 – Need for a Wheelchair	MLP	X	0.840	0.833		
	CNN	X	0.827	0.823		







