Development of imaging biomarkers for ALS (Amyotrophic Lateral Sclerosis) using multivariate statistical techniques and machine learning

Carot-Sierra, J.M.[1], *Gil-Chong, P.*[2], *Vázquez-Barrachina, E.*[3], Cerdá-Alberich, L.[4]

[1]jcarot@eio.upv.es, Departamento de Estadística e Investigación Aplicadas y Calidad, Universitat Politècnica de València

[2]pgilchong@gmail.com, Departamento de Estadística e Investigación Aplicadas y Calidad, Universitat Politècnica de València

[3]evazquez@eio.upv.es, Departamento de Estadística e Investigación Aplicadas y Calidad, Universitat Politècnica de València

[4]leonor_cerda@iislafe.es, Grupo de Investigación Biomédica en Imagen, Instituto de Investigación Sanitaria La Fe

Amyotrophic Lateral Sclerosis is a degenerative motor neuron disease characterized by its diagnostic difficulty: more than 90% of cases are sporadic and there is no reliable paraclinical test capable of detecting it. The development of ALS biomarkers for diagnosis and monitoring is urgently needed.

This work has used a dataset of 211 patients (114 ALS, 45 mimic, 30 genetic carriers and 22 control) with radiomics attributes (morphometry, iron deposition) integrated with clinical variables and 6 semiquantitative visually-assessed indicators of iron deposition.

A binary classification task approach has been taken to classify patients with and without ALS. A sequential modelling methodology, understood from an iterative improvement perspective, has been followed. It has included variable filtering techniques, dimensionality reduction techniques (PCA, kernel PCA), oversampling techniques (SMOTE, ADASYN) and classification techniques (logistic regression, LASSO, Ridge, ElasticNet, Support Vector Classifier, K-neighbours, random forest). For each proposed architecture, several subsets of the available data have been used, proposing models with single datatypes and multimodal models.

The best results have been provided by a voting classifier composed of five classifiers: accuracy=0.896, AUC=0.929, sensitivity=0.886, specificity=0.929. The best results without the use of semiquantitative variables have been provided by Support Vector Classifier: accuracy=0.815, AUC=0.879, sensitivity=0.833, specificity=0.794. In both classifiers a filtering of variables by feature importance in LASSO has been used.

Keywords: biomarker, radiomics, iterative modelling