Opening the Personal Gate between Technology and Health Care
M. Giacomini et al. (Eds.)

doi:10.3233/SHTI260173

A Consensus Clustering Approach to Amyotrophic Lateral Sclerosis Phenotyping

Pilar M. FERRARO[a,1], Sara NARTENI[b,1,*], Marta LENATTI[b], Francesca OLIVERI[c], Chiara GEMELLI[a], Corrado CABONA[d], Antonio UCCELLI[a], Alessia PAGLIALONGA[b], Maurizio MONGELLI[b] and Angelo SCHENONE[a,c]
[a]*IRCCS Ospedale Policlinico San Martino, 16152, Genoa, Italy*
[b]*CNR-IEIIT, 10129, Turin, Italy.*
[c]*Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINOGMI), University of Genoa, 16131, Genoa, Italy*
[d]*Division of Clinical Neurophysiology and Epilepsy Center, IRCCS Ospedale Policlinico San Martino, 16131, Genoa, Italy*
ORCiD ID: Pilar M. FERRARO https://orcid.org/0000-0002-5008-2791, Marta LENATTI, https://orcid.org/0000-0003-0731-7505 Sara NARTENI https://orcid.org/0000-0002-0579-647X, Chiara GEMELLI https://orcid.org/0000-0003-1911-2726, Corrado CABONA https://orcid.org/0000-0002-4282-8805, Antonio UCCELLI https://orcid.org/0000-0002-2008-6038, Alessia PAGLIALONGA https://orcid.org/0000-0002-1341-2560

Abstract. Amyotrophic Lateral Sclerosis (ALS) phenotyping is a challenging task due to its heterogeneous nature and low prevalence. In this paper, we introduce a data-driven approach to support the characterization of ALS phenotypes based on clinical data from a battery of examinations. A consensus clustering method is proposed to identify stable clusters across multiple random data sub-samples, with the objective of discovering whether the retrieved patients' groups and related features align with clinical phenotypes and medical knowledge. Results suggest consistent profiles for bulbar onset ALS patients, driven by onset characteristics, whereas spinal onset ALS patients exhibit greater within-phenotype heterogeneity.

Keywords. consensus clustering, motoneuron diseases, disease phenotyping

1. Introduction

Motor neuron diseases, including Amyotrophic Lateral Sclerosis (ALS), have emerged as one of the most life-threatening neurological disorders. ALS is a multisystem neurodegenerative disease, characterized by heterogeneous clinical, pathological, and genetic background, making disease phenotype classification challenging [1,2]. In this study, we aim to determine whether multivariate clustering of ALS patients using clinical

[1] Equal Contribution
* Corresponding Author: Sara Narteni, CNR-IEIIT (sara.narteni@cnr.it)

data may enhance the identification of disease heterogeneities at the time of diagnosis, thereby building the basis for designing personalized treatment approaches.

2. Methods

2.1. Dataset

Retrospective data were collected from 219 recently diagnosed ALS patients at IRCCS Ospedale Policlinico San Martino, Genoa, Italy. The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (Comitato Etico Regionale della Liguria, LongALS, ID 11992). The dataset included consenting patients following a validated disease diagnosis according to the revised El Escorial Criteria [3]. Participants underwent a battery of neurophysiological, clinical, and extra-motor symptoms evaluations performed by experienced neurologists and neuropsychologists. Disease phenotyping was performed according to [1]. Flail arm, flail leg, and progressive muscular atrophy were grouped into pure/predominant lower motor neuron (pLMN) phenotype. Conversely, pyramidal and primary lateral sclerosis were grouped into pure/predominant upper motor neuron (pUMN) phenotype. Hence, clinical ALS phenotypes were distributed as follows: 66 spinal onset amyotrophic lateral sclerosis (S-ALS), 32 bulbar onset ALS (B-ALS), 31 pUMN, and 55 pLMN patients. Disease severity was evaluated using the ALS Functional Rating Scale-revised (ALSFRS-r), muscular weakness was assessed using the Medical Research Council (MRC) scale , and the severity of UMN involvement was graded using the UMN score [1, 2]. Demographics, family history of motoneuron diseases, onset site (spinal vs. bulbar vs. respiratory), type (UMN vs. LMN vs. both), side (no side vs. right vs. left vs. bilateral), limbs (no limbs vs. upper vs. lower vs. upper and lower) and muscular involvement (absence vs. proximal vs. distal vs proximal and distal), as well as disease duration, were recorded. After excluding 35 patients with missing values, pseudonymized data for 25 clinical features of 184 patients (107 males, 77 females, mean age: 64.7$\pm$12.1 years, mean disease duration: 31 months) were available for the analysis.

2.2. Consensus Clustering

Feature reduction was necessary due to the high number of features relative to the limited number of observations in the dataset. Two feature reduction strategies were evaluated to determine the most effective approach to reduce dimensionality while preserving meaningful information from both numerical and categorical features. These included: (i) applying Principal Component Analysis (PCA) to all features after converting categorical variables into one-hot encoded format, and (ii) using PCA for numerical features while applying Multiple Discriminant Analysis (MDA) for categorical variables. Following feature reduction, unsupervised machine learning was applied for patient stratification using a consensus clustering algorithm [4]. The algorithm begins by sampling a random 80% portion of the data as in [4,5]. Each portion is then clustered into k groups by a clustering algorithm, i.e., KMeans, hierarchical clustering or both (ensemble). Since the goal was to assess whether the intrinsic structure of the data aligned with the known phenotype labelling (S-ALS, B-ALS, pUMN, pLMN), we fixed $k = 4$.

This process was repeated for *n=100* runs [5] yielding a consensus matrix that stores the consensus values for each pair of data instances. The consensus value for a pair of instances represents the proportion of clustering runs in which the two samples are grouped together. Within-cluster consensus is defined as the average consensus value for all pairs of samples inside a cluster. A value close to 1 indicates high cluster stability. Average consensus with respect to a phenotype label is defined as the mean pairwise consensus score for all samples associated with that label. A value close to 1 indicates high cluster reproducibility across runs. Final cluster labels, reflecting samples' stability over clustering iterations, were obtained by applying hierarchical clustering on the consensus matrix as suggested in [4]. In the ensemble case, the final consensus matrix was obtained by averaging the consensus matrices obtained from both methods. The optimal combination of feature reduction and clustering method was selected based on external clustering metrics [6], assessing how well resulting clusters correspond to phenotype labels. In detail, Adjusted Rand Index (ARI), Fowlkes-Mallows index (FM), homogeneity (HM), and completeness (CM) were considered. Moreover, key features of each retrieved clusters were determined based on the standardized mean difference (SMD) [5] with respect to the whole study population (absolute SMD < 0.2: negligible effect; 0.2 ≤ absolute SMD <0.5: minor effect; absolute SMD ≥ 0.5: moderate effect).

3. Results

Dimensionality reduction via PCA yielded 15 components, whereas PCA-MDA resulted in 6 components. Table 1 shows that the combination of PCA feature reduction and ensemble consensus clustering performed better than the other combinations in terms of external metrics (higher ARI, FM and CM values). This method yielded four clusters, C1, C2, C3, and C4, containing 44, 65, 9 and 66 patients, respectively (Silhouette score=0.17). Phenotypes distributions across the identified clusters were as follows: C1: 73% B-ALS, 9% S-ALS, 9% pUMN, and 9% pLMN; C2: 60% S-ALS, 37% pUMN, and 3% pLMN; C3: 45% S-ALS, 33% pLMN, and 22% pUMN; C4: 69.5% pLMN, 29% S-ALS and 1.5% pUMN. The average within-cluster consensus was 0.87 for C1, 0.75 for C2, and 0.72 for C3 and C4. The average within-phenotype consensus suggested that B-ALS patients were better grouped by the clustering algorithm, with an average consensus of 0.92. S-ALS patients presented the lowest average consensus (0.42), while pUMN and pLMN showed a similar behavior (0.59 and 0.62, respectively).

Table 1. External clustering metrics for different combinations of feature reduction and clustering methods.

Clustering	Feature Reduction	ARI	FM	HM	CM
KMeans	**PCA**	0.37±0.06	0.54±0.05	0.44±0.05	0.45±0.05
	PCA-MDA	0.16±0.04	0.41±0.02	0.22±0.03	0.24±0.03
Hierarchical	**PCA**	0.33±0.04	0.53±0.03	0.38±0.03	0.40±0.04
	PCA-MDA	0.15±0.05	0.40±0.04	0.19±0.04	0.22±0.04
Ensemble	**PCA**	0.38	0.56	0.41	0.45
	PCA-MDA	0.16	0.44	0.20	0.25

The four clusters showed distinctive patterns, indicated by SMDs between cluster samples and the entire population across the clinical variables considered (Figure 1). A minor positive effect of diagnostic delay has been observed within C3, pointing to the relevance of disease progression analysis in understanding disease heterogeneity [7]. Onset variables appeared to have a moderate positive effect (SMD $\geq$ 0.5) within C1, compatibly with the prevalence of B-ALS patients in the cluster (key features: no side, bulbar onset site, no limbs and absence of muscular involvement), while the presence of ALS familiarity and definite familiar ALS El Escorial were key (interrelated) features for C3. All features in C2 and C4 showed a negligible or minor effect.

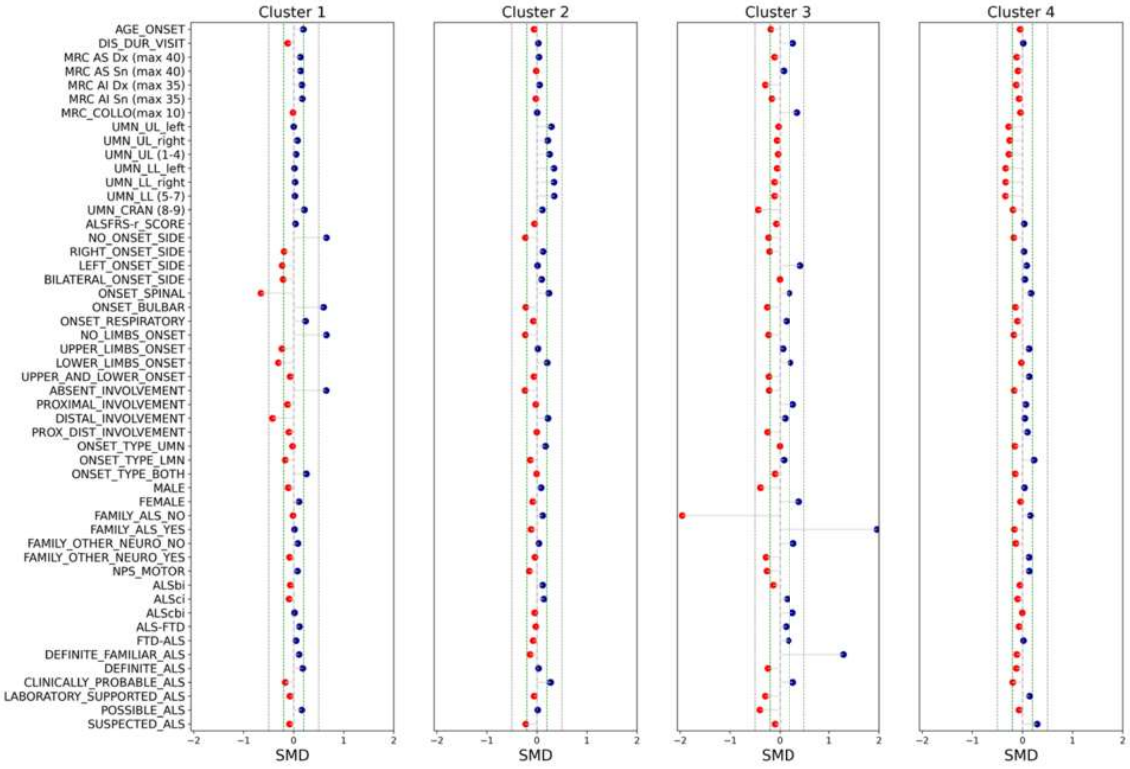

Figure 1 SMD values across the four clusters obtained with ensemble consensus clustering over PCA-reduced features. Blue and red dots denote a positive or negative effect of the considered variable on the clustering.

4. Discussion and conclusions

In this study, we used unsupervised machine learning, specifically consensus clustering, to examine a retrospective Italian cohort of ALS patients, integrating clinical, cognitive/behavioral, and neurophysiological measures. Different combinations of feature reduction and clustering techniques have been analyzed (Table 1) and evaluated based on their ability to reflect four clinical phenotypes (S-ALS, B-ALS, pUMN and pLMN). In the selected configuration, i.e., ensemble consensus clustering following PCA applied to all features, within-cluster consensus was above 0.7 for each cluster. This indicates consistent cluster stability across iterations, even given the relatively small sample size. Analysis of the within-phenotype consensus showed a clear and accurate profiling of B-ALS patients, in line with [8], while progressively weaker profiling was observed for pLMN, pUMN, and S-ALS, consistent with the onset characteristics of

these phenotypes [1]. The low consensus observed for S-ALS patients denotes that they are not consistently grouped across clustering runs. This behavior suggests that the phenotype label is not sufficiently supported by the underlying data and reflects substantial heterogeneity among S-ALS patients at the beginning of the disease, thus requiring longitudinal analysis to evaluate disease progression. To conclude, consensus clustering revealed subgroups of patients with markedly different baseline characteristics indicating that a deeper analysis of patient subgroups beyond traditional phenotyping may help drive advancements in personalized treatment. However, due to the limited sample size, the retrieved clusters may not generalize beyond the observed sample and results should be interpreted as exploratory. Further research is required to expand the cohort, perform a longitudinal assessment, and conduct external validation on available datasets, such as those mentioned in [9] for clinical profile stratification purposes.

Acknowledgements

This work was supported in part by the European Union- NextGenerationEU and by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.5, Project "RAISE- Robotics and AI for Socio-economic Empowerment" (Grant ECS00000035) and through Italian Ministry of Health - 5x1000 funds 2024. A.P. and M.M. are part of RAISE Innovation Ecosystem. M.L. is a PhD student enrolled in the National PhD in AI, XXXVIII cycle, course on Health and Life Sciences, organized by Università Campus Bio-Medico di Roma.

References

[1] Chiò A, Calvo A, Moglia C, Mazzini L, Mora G. Phenotypic heterogeneity of amyotrophic lateral sclerosis: a population based study. Journal of Neurology, Neurosurgery & Psychiatry. 2011; 82(7): 740–746. doi: 10.1136/jnnp.2010.235952.

[2] Meo G, Ferraro PM, Cillerai M, Gemelli C, Cabona C, Zaottini F, Roccatagliata L, Villani F, Schenone A, Caponnetto C. MND Phenotypes Differentiation: The Role of Multimodal Characterization at the Time of Diagnosis. Life (Basel). 2022; 12(10):1506. doi: 10.3390/life12101506.

[3] Brooks BR, Miller RG, Swash M, Munsat TL. El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders. 2000;1(5):293–299. doi:10.1080/146608200300079536.

[4] Monti S, Tamayo P, Mesirov J, Golub T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Machine Learning. 2003;52(1–2):91–118. doi: 10.1023/A:1023949509487.

[5] Qin Y, Xuan L, Wu Z, Deng Y, Liu B, Wang S. Use of consensus clustering to identify distinct subtypes of chronic kidney disease and associated mortality risk. Scientific Reports. 2024;14(1). doi: 10.1038/s41598-024-81208-1.

[6] Hassan BA, Tayfor NB, Hassan AA, Ahmed AM, Rashid TA, Abdalla NN. From A-to-Z review of clustering validation indices. Neurocomputing. 2024; 601:128198. doi:10.1016/j.neucom.2024.128198.

[7] Pires S., Gromicho M., Pinto S., de Carvalho M., Madeira SC. Patient Stratification Using Clinical and Patient Profiles: Targeting Personalized Prognostic Prediction in ALS. Lecture Notes in Computer Science. 2020, 12108:529-541. doi: 10.1007/978-3-030-45385-5_47.

[8] Faghri F, Brunn F, Dadu A, PARALS consortium, ERRALS consortium et al. Identifying and predicting amyotrophic lateral sclerosis clinical subgroups: a population-based machine-learning study. The Lancet Digital Health. 2022; 4(5): e359 - e369. doi: 10.1016/S2589-7500(21)00274-0.

[9] Tavazzi E, Longato E, Vettoretti M, Aidos H, Trescato I, Roversi C, et al. Artificial intelligence and statistical methods for stratification and prediction of progression in amyotrophic lateral sclerosis: A systematic review. Artificial Intelligence in Medicine. 2023; 142:102588. doi: 10.1016/j.artmed.2023.102588.