ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Deep Temporal Consensus Clustering for Patient Stratification in Amyotrophic Lateral Sclerosis

Miguel Pego Roque[1]*, Andreia S. Martins[2]*, Marta Gromicho[3], Mamede de Carvalho[3]
Sara C. Madeira[2], Pedro Tomás[1] and Helena Aidos[2] †

1- INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal

2- LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

3- Instituto de Medicina Molecular and Faculdade de Medicina, Universidade de Lisboa, Portugal

Abstract. Amyotrophic Lateral Sclerosis (ALS) is a fast-acting neurodegenerative disease, characterized by loss of muscle movement and heterogeneity in disease evolution. This poses a challenge in predicting the best time for therapy administration. Here, we propose Deep Temporal Consensus Clustering (DTCC), a stratification method to uncover patient groups with similar disease progression. Using only the initial 6-month follow-up period, DTCC uncovered five clusters that were evaluated in terms of disease evolution and time-to-event. For three critical events (non-invasive ventilation, gastrostomy and death) the attained groups show distinct 10-year progressions, validating the approach.

1 Introduction

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease, characterized by a progressive degeneration of motor neurons. It leads to muscle atrophy and paralysis, with the leading cause of death being respiratory failure. The median life expectancy after the first symptoms ranges from 2 to 4 years [1]. However, 10-20% of patients have a slower disease progression and survival times longer than 10 years [1], and others have a very steep progression and low life expectancy. ALS has no cure and most therapeutic procedures are focused on alleviating symptoms and improving quality of life at later stages of the disease. Most common therapies include non-invasive ventilation (NIV) and percutaneous endoscopic gastrostomy (PEG). The first has been shown to increase life expectancy, while the effect of the latter on survival is still an open subject [1, 2].

Clinical presentations of ALS can vary in affected areas and displayed symptoms, as well as disease progression speed and severity. This heterogeneity poses

*These authors contributed equally to this work.

†This work was partially supported by Fundação para a Ciência e a Tecnologia (FCT) funding to project AIpALS ref. PTDC/CCI-CIF/4613/2020 (https://doi.org/10.54499/PTDC/CCI-CIF/4613/2020), LASIGE Research Unit, ref. UIDB/00408/2020 (https://doi.org/10.54499/UIDB/00408/2020) and ref. UIDP/00408/2020 (https://doi.org/10.54499/UIDP/00408/2020), INESC-ID Research Unit, ref. UIDB/50021/2020 (https://doi.org/10.54499/UIDB/50021/2020); and by the BRAINTEASER project funded by the European Union's Horizon 2020 research and innovation programme under the grant agreement No 101017598 (https://doi.org/10.3030/101017598).

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

a challenge when understanding the disease's stages and evolution, which are essential factors for prognosis prediction and treatment administration, such as whether a patient is eligible for NIV or PEG and when to start them. Therefore, a proper stratification of ALS patients is crucial for understanding the disease progression and developing insights into when critical events are reached.

While there have been several works on ALS patient stratification, they mostly rely on clinical criteria (based on static data), statistical methods and trajectory modelling [3]. However, the latter methods are based on a single feature and, due to disease heterogeneity, lead to several dozens of trajectory groups (e.g., [4]), which is unfeasible for clinical decision-making. Thus, while ALS patient stratification based on temporal data can be challenging, it is still under-explored but highly important for clinical practice.

This work aims to build a novel multivariate method for ALS patient stratification, using deep learning techniques, that takes advantage of the temporal evolution of patients' records. Using the Lisbon ALS Clinic dataset, it uncovers clinically relevant groups of patients, which show significant differences in clinical features and in time-to-event analysis of critical endpoints.

2 Deep Temporal Consensus Clustering

The proposed Deep Temporal Consensus Clustering method (DTCC) was developed to capture the temporal dimension of patients' records and use it to obtain medically relevant groups. The approach is presented in Fig. 1 and is composed of four main modules: (A) temporal autoencoder, (B) manifold learning, (C) hierarchical clustering, and (D) consensus clustering, as detailed next.

(A) Temporal Autoencoder. Consider $\mathcal{D} = \{p_1, ..., p_N\}$, a set of N patients with follow-up, $p_i = \{X_t, X_{t+1}, ...\}$ (X_t is the set of features at time t). This module learns a lower dimensionality space representation, $\mathcal{Z} = \{p'_1, ..., p'_N\}$, by capturing the most relevant information from the data. Both encoder and decoder consist of two Long-Short Term Memory (LSTM) layers, with tanh as the activation function. However, the decoder has an additional time-distributed dense layer at the end to match the original input shape. In the case of ALS data, the LSTM layers of the encoder have sizes 16 and 8 (size of the latent space), and the LSTM decoder layers have sizes 8 and 16. The autoencoder is trained by minimizing the mean-squared error between the reconstructed $\hat{\mathcal{D}}$ and

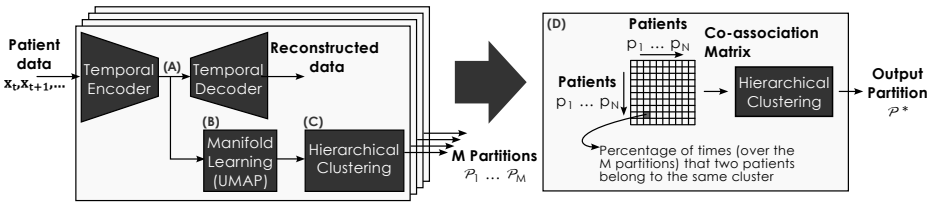

Fig. 1: Proposed stratification methodology.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

the original data $\mathcal{D}$. Here, training uses 1000 epochs with the Adam optimizer.
(B) Manifold Learning. Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) preserves the local and global structure of data which potentiates the learning of an optimized, more clusterable embedding manifold, improving the performance of clustering algorithms [5]. Here, it is used to map the latent space $\mathcal{Z}$ to a new representation of data $\mathcal{Y}$ with the same dimension, because the autoencoder already reduces the dimensionality of the data, and further reduction could lead to loss of information. In the case of ALS data, the parameter responsible for balancing local and global structure is set to 50, because with a lower number UMAP focuses on local structure, and a larger number reflects the global structure.
(C) Hierarchical Clustering. An Agglomerative Hierarchical Clustering is used to stratify the patients, considering Ward linkage and Euclidean distance as the affinity metric. The number of clusters is set by inspecting the dendrogram and computing the silhouette score. It leads to a partition $\mathcal{P} = \{C_1, ..., C_K\}$, where C_i is the i-th cluster in the data partition, and K the number of clusters.
(D) Consensus Clustering. Different runs of the Temporal Autoencoder can produce different data representations, which may lead to different clustering partitions $\mathcal{P}$. To take this into consideration, a clustering ensemble $\mathbb{P} = \{\mathcal{P}_1, ..., \mathcal{P}_M\}$ is obtained by running modules (A) to (C) M times. Then, a co-association matrix $\mathcal{C}$ [6] is computed by counting the number of times (n_{ij}) two patients p_i and p_j are assigned to the same cluster, i.e., $\mathcal{C}(i,j) = n_{ij}/M$. Finally, a consensus partition $\mathcal{P}^*$ is extracted by applying an agglomerative clustering algorithm over the co-association matrix $\mathcal{C}$. In the ALS data, the Ward linkage with Euclidean distance is used and the number of clusters is obtained by computing the silhouette score and the dendrogram inter-cluster distance.

3 ALS Data

The dataset is from the Lisbon ALS Clinic, which consists of Electronic Health Records from ALS patients who have been regularly monitored ($\approx$ every 3 months) from 1995 until May 2023. It includes 1677 patients with a set of static features collected at first appointment (such as demographics, disease severity, comorbidities, and others) and temporal features collected during follow-up evaluations (such as functional assessment with ALSFRS-R, and respiratory tests). The ALSFRS-R is a clinically validated instrument for monitoring ALS progression. It is comprised of 12 questions to assess different affected domains [7]: bulbar (Q1-Q3), upper limbs (Q4-Q5), trunk (Q6-Q7), lower limbs (Q8-Q9), and respiratory (Q10-Q12). To stratify patients with DTCC, only the ALSFRS-R (total score and subscores) and MiToS [8] were used, normalized between $[-1, 1]$.

Since a patient may take some days or weeks to perform all prescribed exams of the same appointment, it is relevant to be able to group all these into a single patient snapshot. This is achieved by following [9], which uses an Agglomerative Hierarchical Clustering algorithm, with the restriction that a snapshot cannot have two observations of the same test.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

4 Experimental Results

4.1 Group Characterization

DTCC was applied to the Lisbon ALS Clinic dataset, by considering the functional scores recorded in the first three appointments (initial assessment plus a 6-month follow-up period). Patients with fewer appointments were discarded, yielding N = 945 patients. The stratification with the best silhouette score (0.667) features five clusters (groups of patients). Their functional domain scores, for the first six evaluations, are presented in Fig. 2. The clusters show separation in affected functional domains and present varying levels of disease severity. This contrasts with the literature [4], which focuses on trajectory modeling, to find dozens of clusters, some associated to individual progressions.

Analyzing the attained clusters, group 1 (N = 192) has a purely motor disease presentation and is the slowest-evolving group overall. The lower limbs and trunk are the most affected domains in the observation period. Group 2 (N = 180) is also a pure motor group but with predominant upper limb and trunk symptoms at the first appointment. However, it quickly worsens in all motor domains. This group also begins to experience respiratory decline towards the end of the observation window. Group 3 (N = 154) is predominantly motor at the first appointment (with similar upper and lower limb scores), but quickly develops both respiratory and bulbar symptoms. It is the group with the fastest respiratory decline overall. Group 4 (N = 152) has a predominantly bulbar presentation at the first three appointments (with only some respiratory decline), but soon after that experiences generalized motor impairment. Group 5 (N = 267) has a fast and generalized disease progression, experiencing some level of functional decline in all domains by the first appointment. Patients in this group worsen in both bulbar and motor domains at similar rates, and exhibit

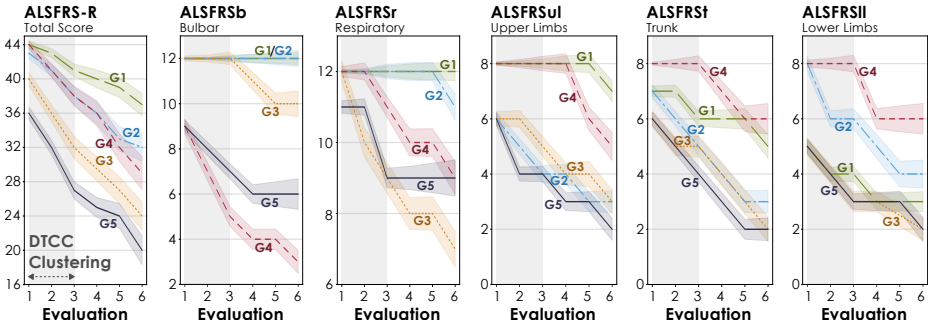

Fig. 2: Disease progression of ALS patient groups obtained by DTCC, by functional domain. Lower values of the scores correspond to the worsening of the disease. Lines represent median values and shaded areas correspond to the 95% confidence interval. Note that DTCC only uses the first 3 evaluations, corresponding to the grey area in each plot.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

respiratory worsening between appointments 2 and 3.

Characterization in terms of demographic and clinical features, collected at first appointment, was also performed through statistical analysis. Some differences were found in terms of age at onset, with groups 1 and 2 being slightly younger (G1: 60yr, 95% CI [51.5:68.0], G2: 61yr, 95% CI [52.0:69.0]), having predominant lower motor neuron involvement, (G1: 55%, G2: 53% of patients in the groups), lower rates of cardiac disease (G1: 5% and G2: 7%) and dyslipidemia (G1: 26% and G2: 21%). The remaining groups are older (G3: 63yr, G4: 65yr, G5: 65yr) and have more propensity for early-disease weight loss (G3: 24%, G4: 17%, G5: 24%) and for more comorbidities such as blood hypertension (G3: 38%, G4: 40%, G5: 37%). Group 4 is comprised of more females than the remaining groups (G1: 43%, G2: 27%, G3: 30%, G4: 61%, G5: 54%), which is consistent with bulbar involvement [1]. Group 5 presents comparing rates of limb and bulbar onset (51% and 42%, respectively), which could not be determined from the functional assessment only.

4.2 Time-to-event Analysis

To validate the obtained clusters, Kaplan-Meier estimators were fitted to each patient group and compared to the baseline curve (Kaplan-Meier estimator of the entire cohort), for three distinct events: NIV, PEG and death. The log-rank test was used to test the difference between each group curve and the baseline curve. Time-to-event curves are shown on a (up to) 10-year follow-up period. Median time-to-event was analysed for each curve and presented along with the 95% confidence interval (95% CI).

Fig. 3 presents the Kaplan-Meier curves for the three critical events. Median time to NIV is typically lower than to PEG, except in the bulbar group (G4) where the need for both therapies arises at similar times. All group curves report significant differences when compared to baseline (p-value ≤ 0.01), except survival in G3 (p-value $= 0.71$).

The motor groups G1 and G2 have the longest expected survival at over 3

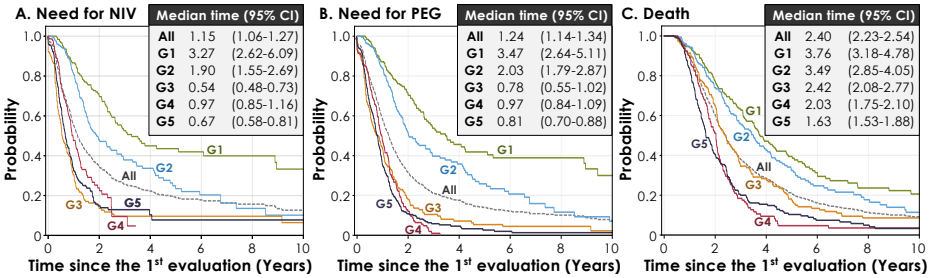

Fig. 3: Kaplan-Meier curves for predicting (A) need for non-invasive ventilation (NIV), (B) need for percutaneous endoscopic gastrostomy (PEG) and (C) death. *All* stands for the baseline curve (i.e. full cohort).

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

years, but subjects with predominant upper limb symptoms (G2) are expected to need NIV and PEG considerably sooner. The other motor group (G3), has the overall shortest times to NIV ($\approx$ 6 months) and PEG ($\approx$ 9 months). However, median survival falls in line with the baseline ($\approx$ 2.5 yr). The general group (G5) has the shortest expected survival and should require NIV and PEG soon after the third evaluation ($\approx$ 6 months). Finally, the bulbar group (G4) has similar times to NIV and PEG ($\approx$ 1 yr), and median survival at 2 years.

From these results, we conclude that critical events occur at vastly different times depending on the attained group. As such, stratification should be taken into account when developing more complex prognosis prediction models.

5 Conclusions

A temporal multivariate stratification method is proposed, Deep Temporal Consensus Clustering (DTCC), and applied to the Lisbon ALS Clinic dataset. Using the first three clinical evaluations, DTCC uncovers five clinically relevant patient groups, that we have evaluated according to (1) disease progression, and (2) time-to-event for three critical events: NIV, PEG and death. In both cases, the groups were vastly different, supporting the use of patient stratification for the development of specialized prognosis prediction models, which we propose to explore in future work.

References

[1] A. Chiò et al. Prognostic factors in ALS: A critical review. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 10(5-6):310–323, 2009.

[2] A. Castanheira, M. Swash, and M. De Carvalho. Percutaneous gastrostomy in amyotrophic lateral sclerosis: a review. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 23(3-4):176–189, 2022.

[3] E. Tavazzi et al. Artificial intelligence and statistical methods for stratification and prediction of progression in amyotrophic lateral sclerosis: A systematic review. *Artificial Intelligence in Medicine*, 142:1–17, 2023.

[4] D. Ramamoorthy et al. Identifying patterns in amyotrophic lateral sclerosis progression from sparse longitudinal data. *Nature Computational Science*, 2(9):605–616, 2022.

[5] M. Allaoui, M.L. Kherfi, and A. Cheriet. Considerably improving clustering algorithms using UMAP dimensionality reduction technique: a comparative study. In *Int. Conf. on Image and Signal Processing*, pages 317–325, 2020.

[6] A. Fred and A.K Jain. Combining multiple clusterings using evidence accumulation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(6):835–850, 2005.

[7] J.M. Cedarbaum et al. 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, 1999.

[8] A. Chiò, E.R. Hammond, G. Mora, et al. Development and evaluation of a clinical staging system for amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 86:38–44, 2015.

[9] A. Carreiro et al. Prognostic models based on patient snapshots and time windows: predicting disease progression to assisted ventilation in amyotrophic lateral sclerosis. *Journal of biomedical informatics*, 58:133–144, 2015.