

Stratification de patients atteints de la maladie de Charcot

Vincent Grollemund^{1,2}, Jean-François Pradat-Peyre^{1,3}, François Delbot^{1,3},
Gaétan Le Chat², Pierre-François Pradat^{4,5}

¹ Sorbonne Université, LIP6, Paris, France

`{vincent.grollemund, francois.delbot, jean-francois.pradat-peyre}@lip6.fr`

² FRS Consulting `gaetan.lechat@frsconsulting.fr`

³ Université de Paris Nanterre, Nanterre, France

⁴ LIB, INSERM Sorbonne Université

⁵ Département de neurologie, APHP `pierre-francois.pradat@aphp.fr`

Mots-clés : *stratification, pronostique, réduction de dimension non-linéaire, médecine personnalisée*

1 Introduction

La Sclérose Latérale Amyotrophique (SLA), ou plus communément appelée maladie de Charcot, est une pathologie neurodégénérative pour laquelle il n'existe actuellement aucun traitement. L'espérance de vie médiane à l'apparition des symptômes varie entre 3 et 5 ans. Établir un pronostique fiable est un enjeu majeur puisqu'il conditionne les modalités de prise en charge du patient et sa qualité de vie. En particulier, la stratification des patients, permet de regrouper les individus en fonction de leurs besoins et ainsi d'adapter le traitement du patient de manière plus pertinente. La mise à disposition récente de données de patients atteints de SLA pour la recherche[1] a permis l'étude de différents modèles de pronostique et de stratification[2]. Dans ce cadre, certaines méthodes d'apprentissage automatique ont réussi à exploiter les corrélations présentes dans les données pour mieux comprendre la progression de la maladie. En particulier, certains travaux font part d'une stratification intéressante[7]. Cependant, ces modèles ne donnent aucune information quant à l'incertitude du résultat fourni alors qu'il s'agit d'une information cruciale pour le clinicien. Cela empêche l'adoption de cette méthode dans un cadre clinique. Par ailleurs, les limites observées par ces modèles dépendent de la pertinence des attributs récoltés sur les patients, du nombre des patients et de la qualité des cas utilisés lors de la construction du modèle. Les pratiques médicales actuelles ne permettent pas d'envisager une amélioration à court terme de ces différents facteurs.

En réponse à ces contraintes, nous avons développé un modèle de prédiction de pronostique robuste, informant le médecin de l'incertitude du résultat obtenu et pouvant être facilement enrichi par des nouvelles données.

2 Méthodologie

Les données utilisées dans le cadre de ces travaux sont issues de 4 bases de données. 3 bases de données regroupent des patients issus d'essais cliniques, PRO-ACT avec 3 539 patients[1], Trophos avec 422 patients[3] et Exonhit avec 297 patients[6] pour un total de 4 258 patients. Ces données constituent notre ensemble d'apprentissage. La quatrième base de données contient 243 cas de patient de l'hôpital de la Pitié Salpêtrière¹. Ces données « vie réelle »² constituent notre

1. Base de donnée mise à notre disposition grâce à l'aide du Professeur de neurologie Vincent Meininger

2. Données obtenues sans intervention sur les modalités usuelles de prise en charge des malades et non collectées dans un cadre expérimental. Elles reflètent, a priori, la pratique courante.

échantillon de test. Les attributs considérés dans notre étude (sexe, âge, temps écoulé depuis l'apparition des premiers symptômes, site du début de maladie (spinal ou bulbaire), perte de motricité fonctionnelle initiale, poids initial du patient et estimation de la pente initiale de perte de motricité) dépendent directement de la variabilité des informations présentes dans les différentes bases de données. Notre méthode de prédiction consiste en la combinaison d'une méthode de réduction de dimension non-linéaire combinée à une approche probabiliste. En effet, les méthodes linéaires de réduction de dimension, telles que l'ACP, n'ont pas permis d'identifier une stratification pertinente des données. A contrario, l'algorithme UMAP³[5] a quant à lui, produit des résultats intéressants (plus appropriés que l'état de l'art actuel en matière d'apprentissage de variété : l'algorithme t-SNE[4]). En effet, la projection des données patients dans un espace bi-dimensionnel (ce que fait UMAP) fait apparaître une corrélation entre la projection des attributs d'entrée des patients et la survie à un an. Cet espace est ensuite divisé ad-hoc en trois zones avec des taux de survie propres avec un intervalle de confiance associé. Les données de test sont ensuite utilisées pour montrer la pertinence de ce modèle à un niveau clinique.

3 Résultats et perspectives

La probabilité de survie des patients atteints de SLA dans notre jeu de données d'apprentissage est de 78.9% (+/- 2.4%) à un an. Grâce à notre approche, trois zones sont identifiées avec respectivement des taux de survie à un an de 94.6% (+/- 2.4%) - taux de survie élevé - de 80.9% (+/- 3.4%) - taux de survie moyen - et de 54.8% (+/- 5.2%) - taux de survie faible. L'appartenance aux zones de taux de survie élevé et faible est informative pour le médecin et permet de segmenter les patients atteints de SLA. Notre approche passe facilement à l'échelle⁴, fournit l'incertitude du résultat, est robuste (car elle ne nécessite qu'un nombre limité d'attributs et ceux-ci sont couramment mesurés dans le cadre de la SLA) et son implémentation dans un cadre clinique est relativement aisée.

L'augmentation du nombre de données collectées permettra une stratification plus fine, et l'ajout d'attributs patients supplémentaires (paramètres respiratoires) pourront améliorer la projection des données.

Références

- [1] Nazem Atassi, James Berry et al. The PRO-ACT database : design, initial analyses, and predictive features. *Neurology*, 83(19) :1719–1725, 2014.
- [2] Vincent Grollemund, Pierre-François Pradat et al. Machine Learning in Amyotrophic Lateral Sclerosis : Achievements, Pitfalls and Future Directions. *Frontiers in Neuroscience*, 13, 2008.
- [3] Timothée Lenglet, Lucette Lacomblez et al. A phase II- III trial of olesoxime in subjects with amyotrophic lateral sclerosis. *European journal of neurology*, 21(3) :529–536, 2014.
- [4] Laurens van der Maaten, Geoffrey Hinton Visualizing data using t-SNE. *The Journal of Machine Learning Research*, 9 :2579–2605, 2008.
- [5] Leland McInnes, John Healy et al. UMAP : Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv :1802.03426*, 2018.
- [6] Vincent Meininger, Bernard Asselain et al. Pentoxifylline in ALS : a double-blind, randomized, multicenter, placebo-controlled trial. *Neurology*, 66(1) :88–92, 2006.
- [7] Henk-Jan Westeneng, Thomas PA Debray et al. Prognosis for patients with amyotrophic lateral sclerosis : development and validation of a personalised prediction model. *The Lancet Neurology*, 17(5) :423–433, 2018.

3. Algorithme non-linéaire de réduction de dimension (ou d'apprentissage de variété)

4. i.e. la précision de la prédiction s'améliore au fur et à mesure que de nouveaux patients sont ajoutés.