

# Classification automatique de séries chronologiques de patients souffrant de douleurs chroniques

Armel Soubeiga\*, Jessem Ettaghouti\*, Violaine Antoine\*, Alice Corteval\*\*, Nicolas Kerckhove\*\*\*, Sylvain Moreno\*\*\*\*

\* Université Clermont Auvergne, CNRS, LIMOS, ENSMSE LIMOS,  
F-63000 Clermont Ferrand France  
armel.soubeiga@uca.fr

\*\* Institut Analgesia, Clermont-Ferrand, France

\*\*\* Service de Pharmacologie médicale, CHU Clermont-Fd

\*\*\*\* Digital Health Hub, Université Simon Fraser, Vancouver, Canada

## 1 Introduction

La douleur chronique touche des millions de patients en France, soit environ 30% de la population générale. Les traitements disponibles sont anciens, ont une efficacité limitée et peuvent entraîner des effets indésirables importants (Kerckhove et al. (2022)). De plus, le parcours de santé des patients souffrant de douleur chronique est multiple ce qui entraîne des résultats médiocres en terme d'amélioration de leur santé. L'identification des profils de ces patients permettrait aux corps médical d'améliorer les résultats de soins et de mieux soutenir ces patients. Dans cette étude, nous proposons d'utiliser l'algorithme de clustering évidentiel des c-moyennes (ECM) pour identifier des parcours de soins et déterminer les divers profils de patients affectés par la douleur chronique.

## 2 Présentation des données et l'approche proposée

Toutes les données ont été collectées à l'aide de l'application mHealth eDOL (Kerckhove et al. (2022)), permettant aux patients et à leurs médecins de remplir des questionnaires cliniques, personnels et barométriques liés à la douleur chronique. Les six attributs barométriques (la douleur, la fatigue, le moral, le stress, le sommeil, le confort corporel, l'activité sportive et non-sportive), mesurés hebdomadairement ont en effet permis d'évaluer l'intensité de la douleur et ses répercussions. Au total, les données de 246 patients ont été analysées. L'objectif principal est de développer un cadre de clustering basé sur les caractéristiques des baromètres de la douleur. Pour cela, nous avons effectué d'abord, des transformations sur les baromètres pour extraire les caractéristiques et ensuite, nous avons appliqué des techniques de sélection de variables. Un grand nombre de caractéristiques de tendance, de dispersion, d'homogénéité, d'aplatissement et d'asymétrie ont été calculées. Ensuite, pour récupérer les caractéristiques relativement plus concises, la méthode de sélection proposée prend en compte la similarité

entre les données, par le calcul du score Laplacien (He et al. (2008)) et la corrélation de Pearson entre les attributs (Hall (2000)).

### 3 Application et résultats

Nous avons extrait 97 attributs des séries brutes des baromètres dans la phase de transformation et 8 attributs pertinents ont été choisis selon la similarité et la corrélation durant la phase de sélection. L'algorithme ECM (Masson et Denœux (2008)) est ensuite appliqué afin d'identifier des typologies de parcours de soins. Le nombre optimal de clusters, qui maximise la silhouette moyenne est deux clusters : un cluster caractérisé par les douleurs liées au stress et un cluster regroupant des patients dont la douleur est liée à la fatigue et des troubles de sommeil. Deux sous-groupes sont également identifiés : un groupe qui représente les patients incertains entre les deux clusters et un groupe de patients atypiques. L'explicabilité des clusters déterminés à travers une analyse descriptive et des tests statistiques nous a permis de déterminer le profil de ces patients douloureux chroniques.

### 4 Conclusion

Cette approche, nous a permis d'identifier deux typologies de parcours avec des profils différents. En plus de ces parcours, nous avons identifiés des groupes de patients incertains et atypiques. Il existe plusieurs améliorations possibles à ce travail. D'un point de vue technique, il serait intéressant d'étudier le résultat de différents algorithmes de sélection d'attributs, ou d'utiliser un algorithme de clustering évidentiel qui prend en entrée des données séquentielles.

### Remerciements

Ce projet a bénéficié du soutien de l'Europe dans le cadre du FEDER. Les auteurs remercient également le soutien reçu de l'Agence Nationale de la Recherche du gouvernement français à travers le programme Investissements d'Avenir (CAP 20-25).

### Références

- Hall, M. (2000). Correlation-based feature selection for discrete and numeric class machine learning. *undefined 2000*, 359–366.
- He, X., D. Cai, et P. Niyogi (2008). Laplacian score for feature selection. Technical report, Department of Computer Science, University of Chicago.
- Kerckhove, N. et al. (2022). eDOL mhealth app and web platform for self-monitoring and medical follow-up of patients with chronic pain : Observational feasibility study. *JMIR Form Res* 6(3), e30052.
- Masson, M. H. et T. Denœux (2008). ECM : An evidential version of the fuzzy c-means algorithm. *Pattern Recognition* 41, 1384–1397.