

Une pluralité de méthodes, une seule partition

Contexte :
La maladie de Wegener et la polyangéite microscopique sont deux maladies auto-immunes rares qui appartiennent à une même famille : celle des vascularites associées aux ANCA (*Anti-Neutrophilic Cytoplasmic Antibodies*). Ces maladies possèdent des symptômes communs les rendant difficilement différenciables. Afin de faciliter leur diagnostic, une étude visant à discriminer les patients a été réalisée et publiée. Elle a abouti à une classification en 5 groupes cliniquement différents de ces patients.

Objectif :
Le but de ce projet est de valider cette partition en 5 groupes ou d'en trouver une plus robuste en utilisant d'autres méthodes de classification

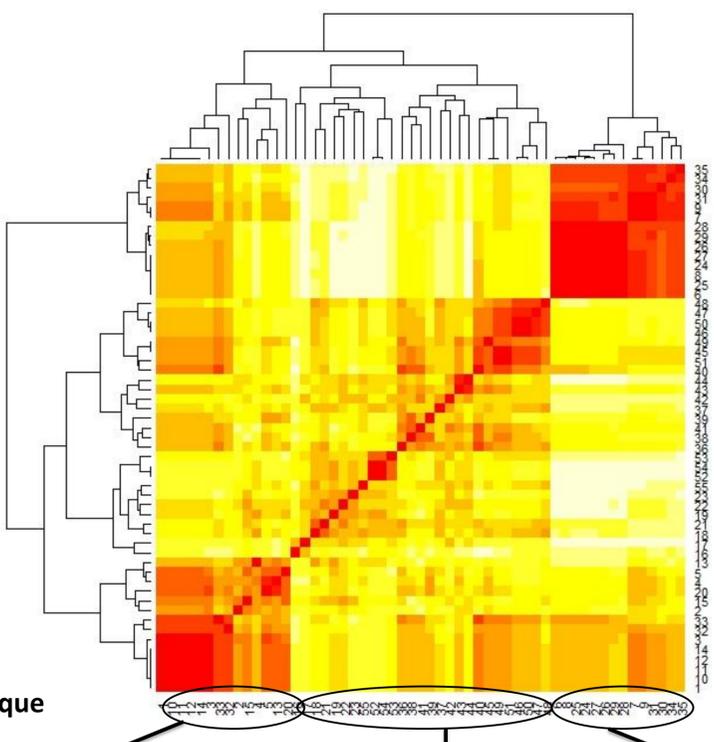
Données :

- La base de données se compose de 673 patients issus de plusieurs études européennes.
- Elle comporte 23 variables dont 11 seront utilisées.

Variables cliniques		Variables générales
Signes O.R.L.	Signes cutanés	Sexe
Signes oculaires	Signes pulmonaires	Variables biologiques
Signes neurologiques	Atteintes rénales	PR3-ANCA
Signes cardio-vasculaires	Signes digestifs	MPO-ANCA

Statistiquement parlant...

- Choix de départ :**
- Tableau binaire (données initiales)
 - Tableau quantitatif (obtenu après une ACM)
- Méthodes mises en œuvre :**
- Des méthodes hiérarchiques : ascendantes : AGNES, descendantes : DIANA
 - Des méthodes de partitionnement : Kmeans, CLARA, PAM
 - Des méthodes basées sur la densité ou sur une approche probabiliste : DBSCAN, algorithme EM



- Comparaison des 55 partitions :**
- Pour comparer les 55 partitions obtenues, un tableau de dissimilarité est construit. Il renvoie un indice compris entre 0 et 1. Plus 2 méthodes classent les patients de façon identique, plus cet indice est proche de 0.
 - Afin de regrouper les partitions semblables, une nouvelle classification est réalisée sur ce tableau de dissimilarité.
 - Dans le but de mieux visualiser les résultats de cette classification, une heatmap est construite à partir des 2 résultats précédents. (Il combine une représentation graphique du dendrogramme et des indices de dissimilarité.)
- Si 2 partitions sont proches (classent les patients de la même façon), elles sont représentées par une couleur rouge sur la carte de chaleur.
- 3 cas de distinguent ici.
- 3 groupes retenus

→ 55 partitions retenues d'un point de vue statistique

Cliniquement parlant...

Groupe 1

Méthodes contenues :

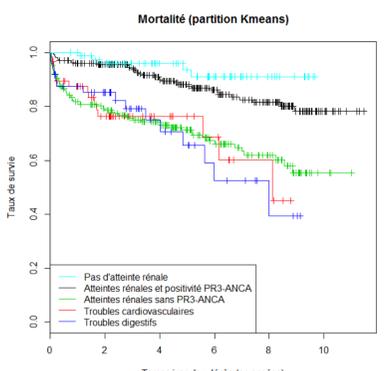
- Kmeans
- Agnes
- EM

Interprétation clinique :
Les partitions de ce groupe discriminent les patients en 5 classes caractérisées de la façon suivante :

- Pas d'atteinte rénale
- Atteintes rénales et positivité PR3-ANCA
- Atteintes rénales sans PR3-ANCA
- Troubles cardio-vasculaires
- Troubles digestifs

Variables discriminantes :
Troubles rénaux, PR3-ANCA, signes cardio-vasculaires, signes digestifs

Analyse de survie :
Ces 5 groupes se caractérisent par des survies différentes



Ce graphique représente les courbes de survies de chacun des 5 groupes. La survie est définie comme étant la probabilité que l'évènement (ici le décès), ne soit pas survenu à chaque temps (ici les années).

Groupe 2

Méthodes contenues :

- Clara et Pam
- Toutes les méthodes appliquées sur tableau qualitatif

Interprétation clinique :
Pas d'interprétation clinique pertinente

Interprétation générale :

- Méthodes du groupe 3 ne prennent pas en considération autant de variables discriminantes que celles du groupe 1
- Importance des troubles cardio-vasculaires et digestifs dans la différenciation des patients
- Les méthodes construites sur un tableau qualitatif ne donnent pas de partitions pertinentes. D'où l'importance de la réalisation d'une ACM en amont.
- Les méthodes du groupe 1 nous donnent des résultats similaires à la publication

Conclusion générale :
Validation de la classification en 5 groupes publiée

Groupe 3

Méthodes contenues :

- DBSCAN
- Diana
- EM

Interprétation clinique :
Les partitions discriminent les patients, en formant un « gros » groupe et plusieurs « petits », par exemple :

- Troubles cardio-vasculaires
- Troubles digestifs
- Troubles cardio-vasculaires et digestifs
- Aucun de ces 2 troubles (grande classe)

Variables discriminantes :
Signes cardio-vasculaires, signes digestifs

