

Apprentissage d'une mesure pour la stratification de patients atteints de la sclérose latérale amyotrophique



Guillaume Tejedor, Veronika Peralta, Nicolas Labroche,
Patrick Marcel, Hélène Blasco, Hugo Alarcón

contact : guillaume.tejedor@univ-tours.fr



Contexte

Problématique

- Maladie impactant le contrôle des mouvements musculaires volontaires.
- Stratification des patients pour identifier des signatures cliniques/biologiques.

Données

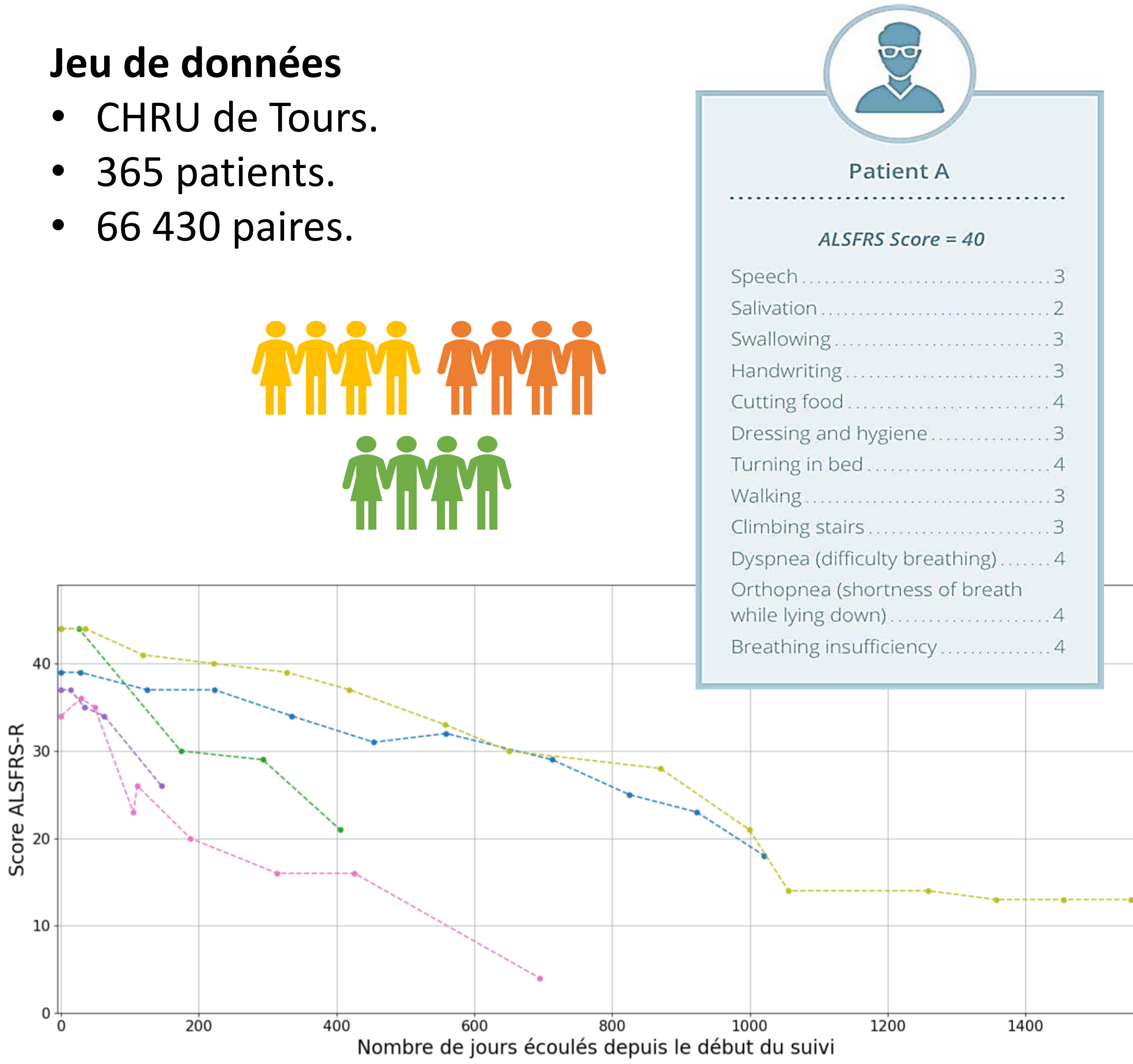
- Démographiques : Âge, localisation du début de la maladie, etc.
- Longitudinales : Séquences temporelles (cliniques, biologiques).
- Représentation progression maladie : Score ALSFRS-R
 - 12 questions (Chaque question notée entre 0 et 4)
 - Score total : entre 0 (Perte totale) et 48 (Aucune Perte).

État de l'art

- Critères de recherche :
 - Suivi de la maladie via le score ALSFRS-R.
 - Utilisation des déclin fonctionnels pour stratifier.
- Méthodes employées :
 - Seuils appliqués sur une variable extraite des séquences ALSFRS-R.
 - Mesure de similarité (DTW - Dynamic Time Warping).

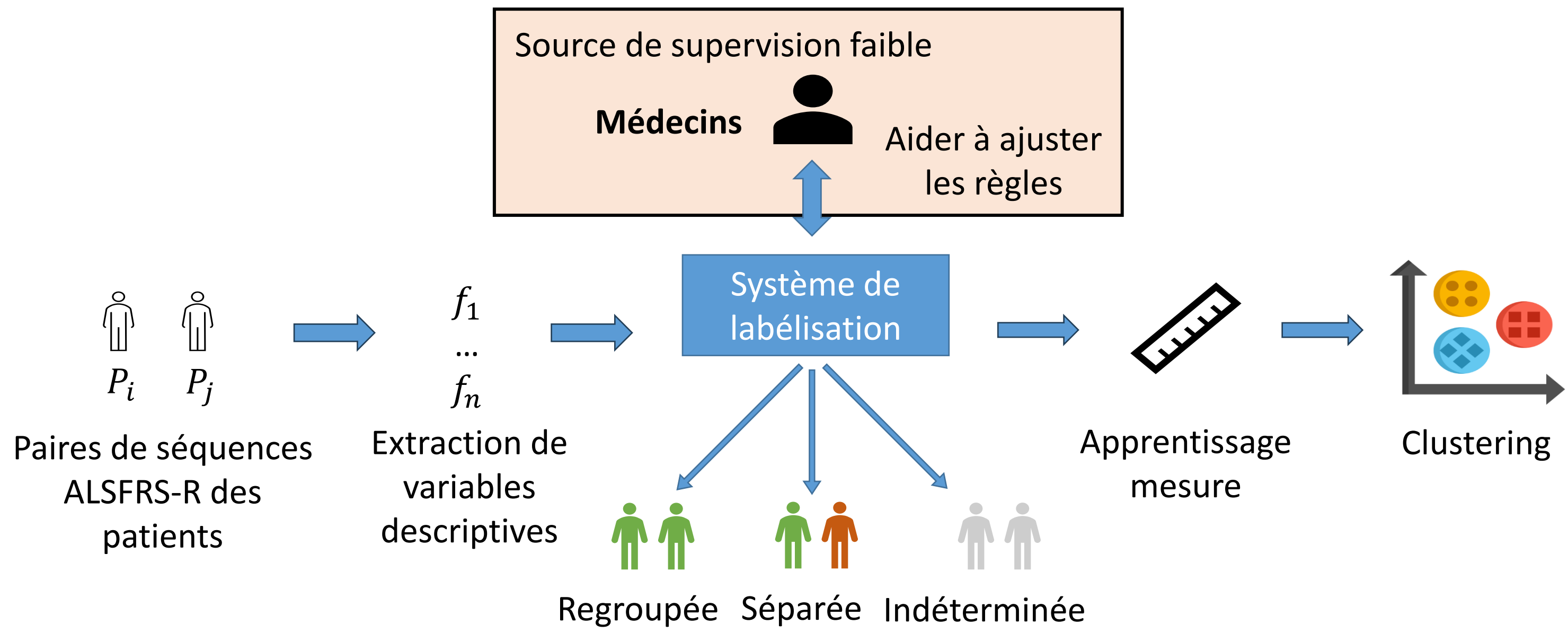
Jeu de données

- CHRU de Tours.
- 365 patients.
- 66 430 paires.

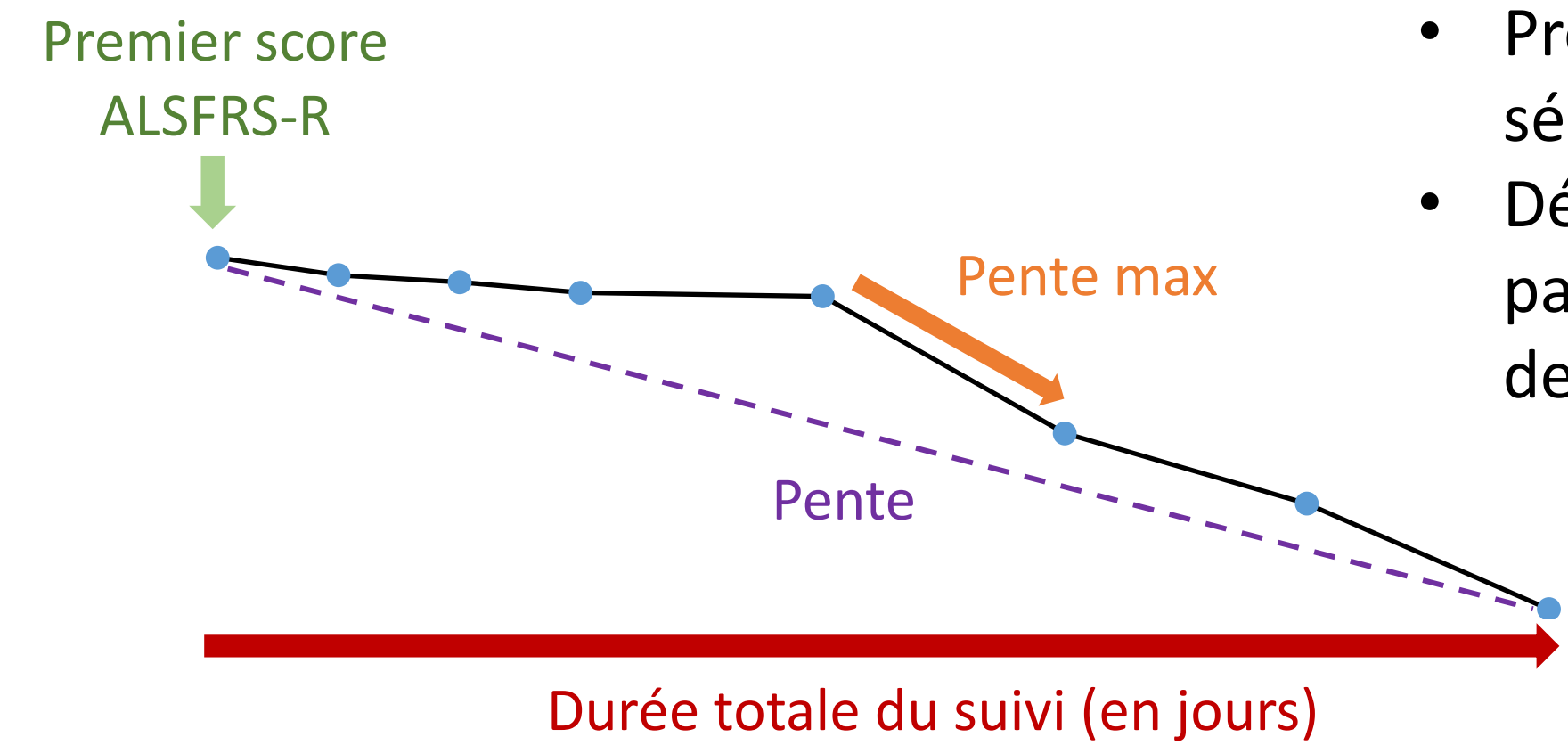


Approche

Méthodologie

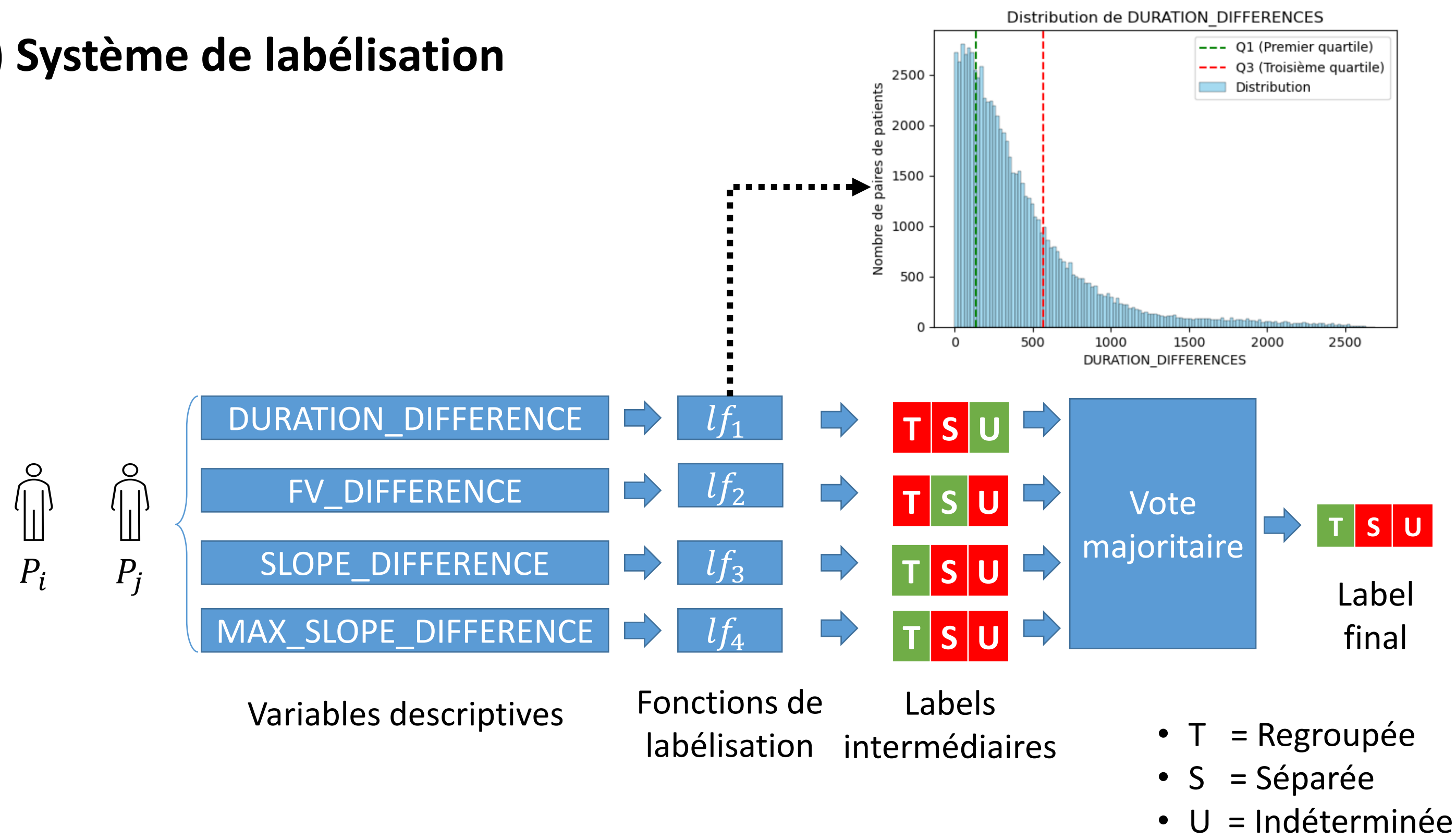


1) Extraction des variables descriptives



- Pré-calcul d'informations extraites des séquences ALSFRS-R.
- Définition de variables descriptives par paire de séquences (ex : différence de durée de suivi).

2) Système de labélisation



3) Apprentissage de la mesure

Support Vector Machine (SVM) linéaire pour apprendre les poids de chaque variable.

- P_i et P_j représentent respectivement les patients i et j .
- k est le nombre de variables descriptives.
- f_z désigne la fonction de labélisation z .
- w_{f_z} désigne le poids appris par classifieur.
- m_{f_z} la valeur de la variable descriptive.

$$mesure(P_i, P_j) = \sum_{z=1}^k w_{f_z} * m_{f_z}$$

Expérimentations

Mesure	UMAP	Algorithme de clustering	c=3	c=4	c=5
DTW	NO	AHC	0.304 ± 0.263	0.234 ± 0.241	0.22 ± 0.243
DTW	NO	KMEDOID	0.299 ± 0.251	0.211 ± 0.266	0.157 ± 0.254
DTW	YES	AHC	0.449 ± 0.243	0.43 ± 0.248	0.43 ± 0.192
DTW	YES	KMEANS	0.475 ± 0.192	0.456 ± 0.174	0.413 ± 0.204
DTW	YES	KMEDOID	0.473 ± 0.194	0.452 ± 0.184	0.426 ± 0.178
DSHF	NO	AHC	0.627 ± 0.182	0.442 ± 0.193	0.443 ± 0.198
DSHF	NO	KMEDOID	0.343 ± 0.223	0.274 ± 0.199	0.229 ± 0.2
DSHF	YES	AHC	0.494 ± 0.291	0.471 ± 0.263	0.373 ± 0.272
DSHF	YES	KMEANS	0.483 ± 0.183	0.436 ± 0.183	0.451 ± 0.169
DSHF	YES	KMEDOID	0.485 ± 0.181	0.411 ± 0.178	0.443 ± 0.191

- AHC : Agglomerative Hierarchical Clustering.
- Silhouette Moyenne ± écart-type variant le nombre de clusters c entre 3 et 5.

Conclusion et perspectives

Résumé

- Approche combinant système de labélisation et SVM pour stratifier les patients via les séquences ALSFRS-R.
- Comparaison DTW : Meilleure performance de notre méthode.

Perspectives

- Nouvelles caractéristiques pour affiner les clusters.
- Analyse statistique : Survie, différences entre clusters.
- Prédiction du cluster dès le diagnostic.