

Classification de la maladie d'Alzheimer à partir des données tabulaires ADNI

Amadou Selle NDIAYE

Sous la supervision du professeur Maxime Descoteaux

Université de Sherbrooke

17 avril 2025

- 1 Introduction
- 2 Description des données
- 3 Expériences
- 4 Prétraitement
- 5 Modèles Utilisés
- 6 Résultats Comparatifs
- 7 Conclusion

- L'étude vise à détecter précocement la maladie d'Alzheimer.
- Les données tabulaires contiennent des informations cliniques, cognitives et biomarqueurs.
- Objectif : prédire le diagnostic clinique (CN, MCI, AD) à partir de ces données.

Description des données

- Extrait de la base de données ADNI (Alzheimers Disease Neuroimaging Initiative) contenant 744 patients et 2847 caractéristiques
- Variables clés :
 - Identifiants (PTID)
 - Démographiques (Âge)
 - Groupes d'interprétation (NC, MCI, AD)
 - Caractéristiques d'imagerie (volumes hippocampiques, etc.)

Biomarqueurs Pertinents

- amyloid_positive : Distinction NC vs MCI/AD
- PLASMA_NFL : Sévérité du déclin
- wm_mask_fw : Marqueur précoce
- entre autres ...

Premières expérimentations et difficultés rencontrées :

- Suppression des colonnes contenant plus de 50 % de valeurs manquantes.
- Imputation des données manquantes avec KNN.
- Utilisation de modèles classiques : SVM, KNN, Random Forest, etc.
- Résultats insuffisants : précision inférieure ou égale à 68 %.

Amélioration suite aux retours du superviseur

- Mise en place d'une sélection de caractéristiques basée sur la corrélation, suite à la proposition du superviseur.
- Amélioration des performances des modèles, comme le montre le tableau suivant :

Modèle	F1-score	Accuracy
Gradient Boosting	71.58 %	71.72 %
Voting Ensemble	71.05 %	71.12 %
Weighted Ensemble	70.70 %	70.78 %
SVM	65.44 %	65.55 %
Random Forest	65.09 %	65.09 %

Problèmes identifiés :

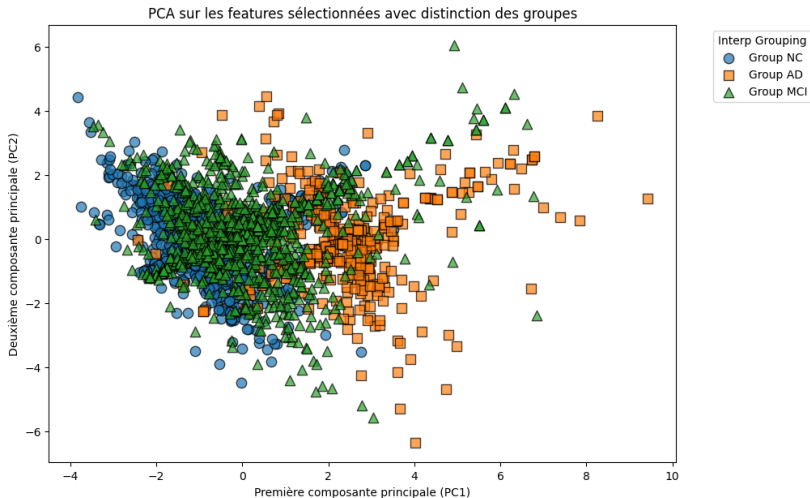
- Biais introduit par l'imputation sur un grand nombre de données manquantes.
- Trop d'approximation mauvaise qualité des données.
- Sélection basée uniquement sur la variance peu pertinente.
- Absence d'analyse approfondie des caractéristiques travail aléatoire.

- Analyse exploratoire :
 - Sélection raisonnée de 30 caractéristiques prépondérantes sur la base de mes recherches sur la maladie et sur les recommandations du professeur
 - Visualisation par boxplots
 - Calcul de la variance intra-groupe avec les features sélectionnés
- Sélection des 10 caractéristiques avec les plus fortes variances :

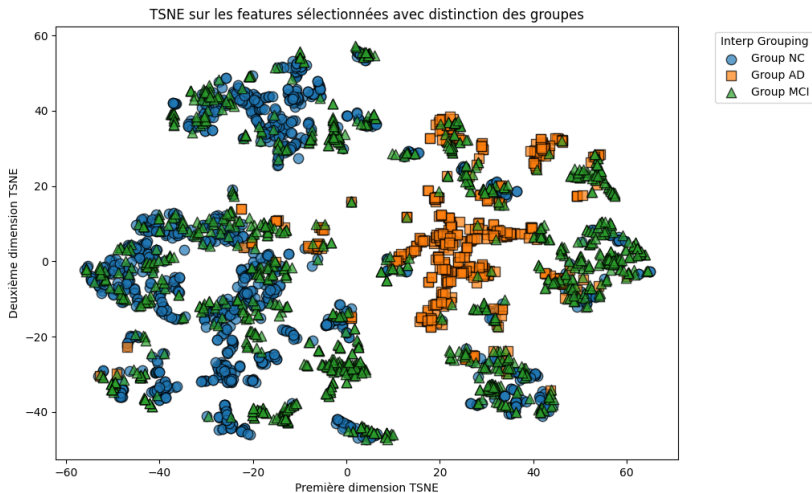
```
colonnes = ['fornix_L_safe__volume', 'fornix_R_safe__volume', 'safe_wm__volume', 'hippocampus_L__volume', 'hippocampus_R__volume', 'TAU', 'PLASMA_NFL', 'PTAU', 'APOE4', 'amyloid_positive']
```

- Nettoyage des données :
 - Gestion des valeurs manquantes (NaN) en utilisant la médiane par caractéristiques en fonction du groupe
- Normalisation des données

Visualisation avec les dimensions réduites : PCA à 2 dimensions



Visualisation avec les dimensions réduites : T-SNE à 2 dimensions



KNeighborsClassifier

- Classification par voisins
n_neighbors=5
- Accuracy modérée
- Sensible au déséquilibre

RandomForest

- Ensemble d'arbres
n_estimators=100
- Meilleure performance
- Interprétable

Neural Network

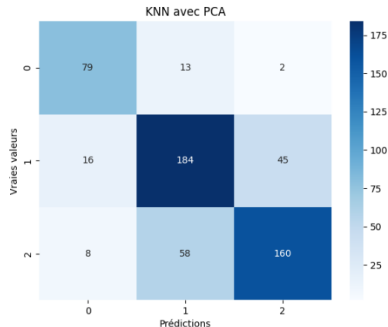
- Architecture :
 - Couche **lineaire** avec fonction d'activation **ReLu** et une fonction de sortie **softmax**
 - KFold validation (5 parties) + 150 epochs + learning rate (0.010)
- Performance modérée
- Coûteux en calcul

Résultats de classification par modèle

KNN sur les données tabulaires ADNI

Accuracy : 74.87%

Classe	Précision	Rappel	F1-score	Support
AD	0.77	0.84	0.80	94
MCI	0.72	0.75	0.74	245
NC	0.77	0.71	0.74	226
Macro avg	0.75	0.77	0.76	565
Weighted avg	0.75	0.75	0.75	565

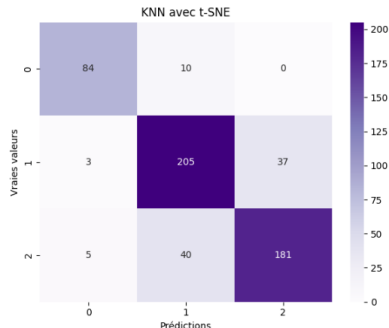


Résultats de classification par modèle

KNN sur les données tabulaires ADNI

Accuracy : 83.19%

Classe	Précision	Rappel	F1-score	Support
AD	0.91	0.89	0.90	94
MCI	0.80	0.84	0.82	245
NC	0.83	0.80	0.82	226
Macro avg	0.85	0.84	0.85	565
Weighted avg	0.83	0.83	0.83	565



- **Résultats** : Accuracy modérée, sensible au déséquilibre des classes.

Résultats de classification par modèle

Random Forest sur les données tabulaires ADNI

Accuracy : 77.52%

Classe	Précision	Rappel	F1-score	Support
AD	0.77	0.87	0.82	94
MCI	0.75	0.78	0.76	245
NC	0.81	0.73	0.77	226
Macro avg	0.78	0.79	0.78	565
Weighted avg	0.78	0.78	0.77	565

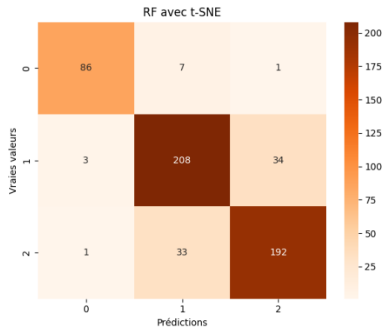


Résultats de classification par modèle

Random Forest sur les données tabulaires ADNI

Accuracy : 86.02%

Classe	Précision	Rappel	F1-score	Support
AD	0.96	0.91	0.93	94
MCI	0.84	0.85	0.84	245
NC	0.85	0.85	0.85	226
Macro avg	0.88	0.87	0.88	565
Weighted avg	0.86	0.86	0.86	565



● Résultats :

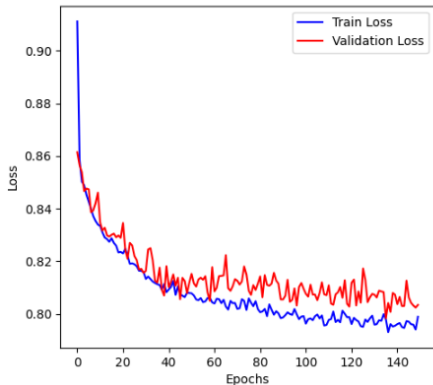
- Meilleure performance que KNN grâce à sa capacité à gérer des relations non linéaires.
- Importance des caractéristiques : amyloid_positive et PLASMA_NFL en tête.

Résultats de classification par modèle

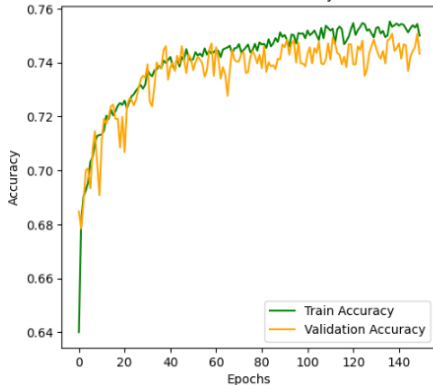
Neural Network sur les données tabulaires ADNI

Accuracy : 74.087%, Validation : 73.61%

Évolution de la Perte



Évolution de l'Accuracy



● Résultats :

- Performance intéressante, mais nécessite un tuning approfondi.

Conclusion

Améliorations Possibles

- Meilleure gestion des données manquantes
- SMOTE pour classes déséquilibrées
- Prédire la Progression MCI \rightarrow AD
- Optimisation hyperparamètres

Perspectives

- Intégration données multi-modales
- Déploiement outil clinique

Merci !