

Optimisation des Random Forests pour le Diagnostic de Maladies Rares

HURET Alexandre

Année 2024-2025

Introduction

- L'intelligence artificielle est de plus en plus utilisée en médecine.
- Les maladies rares posent un problème de classification car elles sont sous-représentées dans les bases de données.
- Les Random Forests sont une méthode populaire, mais leur performance peut être limitée dans ce contexte.

Question centrale

Peut-on adapter la construction des arbres dans une Random Forest pour améliorer le diagnostic de maladies rares à partir de données médicales ?

- Comment équilibrer les classes rares et majoritaires ?
- Quelles modifications peut-on apporter aux arbres de décision ?
- Quels critères d'évaluation sont les plus pertinents ?

- **Introduction** : Problématique et enjeux médicaux.
- **État de l'art** : Random Forests en médecine et problème des classes rares.
- **Méthodologie** : Présentation du dataset, des outils et des techniques testées.
- **Résultats et analyses** : Comparaison des performances avec et sans optimisation.
- **Conclusion et perspectives** : Bilan et améliorations futures.

Faire une revue bibliographique

- **Objectif** : Comprendre ce qui existe déjà et identifier les limites des approches actuelles.

Concepts clés à explorer

- Comment fonctionne une Random Forest en général ?
- Pourquoi les maladies rares posent un problème de classification ?
- Quelles techniques sont déjà utilisées pour gérer les classes déséquilibrées en médecine ?
 - Sur-échantillonnage
 - Pondération des classes
 - Arbres spécialisés

Comment fonctionne une Random Forest ?

- Une **Random Forest** est un ensemble de plusieurs arbres de décision.
- Chaque arbre est entraîné sur un **sous-ensemble aléatoire** des données (*bootstrap sampling*).
- À chaque nœud, une sélection aléatoire de variables est utilisée pour déterminer la meilleure séparation.
- La prédiction finale est obtenue par **vote majoritaire** (classification) ou **moyenne** (régression).

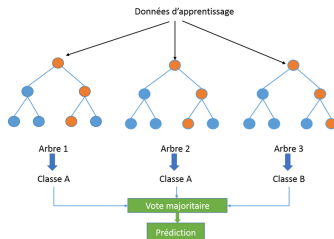


Figure: Schéma illustrant le fonctionnement d'une Random Forest.