

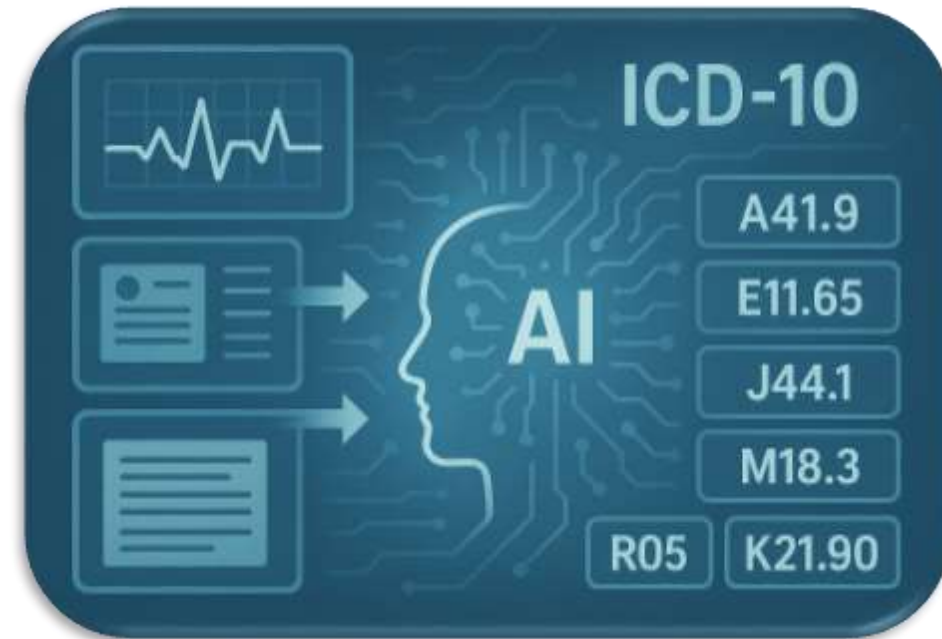
Développement d'un système de prédiagnostic médical automatique : la prédiction des codes ICD-10

Réalisé par : **Rihab Baccari**

Encadré par : **Pr. Yacine Yaddaden**

Introduction et contexte :

- Le codage médical à l'aide de la classification **ICD-10** est une étape cruciale dans les systèmes de santé modernes.
- Cependant, son traitement manuel reste **long, coûteux et sujet aux erreurs**.
- Avec l'explosion des données médicales et les progrès de l'intelligence artificielle, il est désormais possible de repenser cette étape **de manière automatisée et intelligente**.

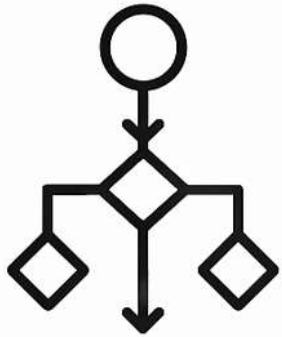


Problématique :



- Le codage manuel ICD-10 est chronophage et source d'erreurs.
- Les approches actuelles négligent les données numériques.
- Notre approche hiérarchique s'appuie sur la structure des codes ICD-10 pour améliorer la précision.

Objectifs :



Développer un système de
prédiction automatique des
codes ICD-10

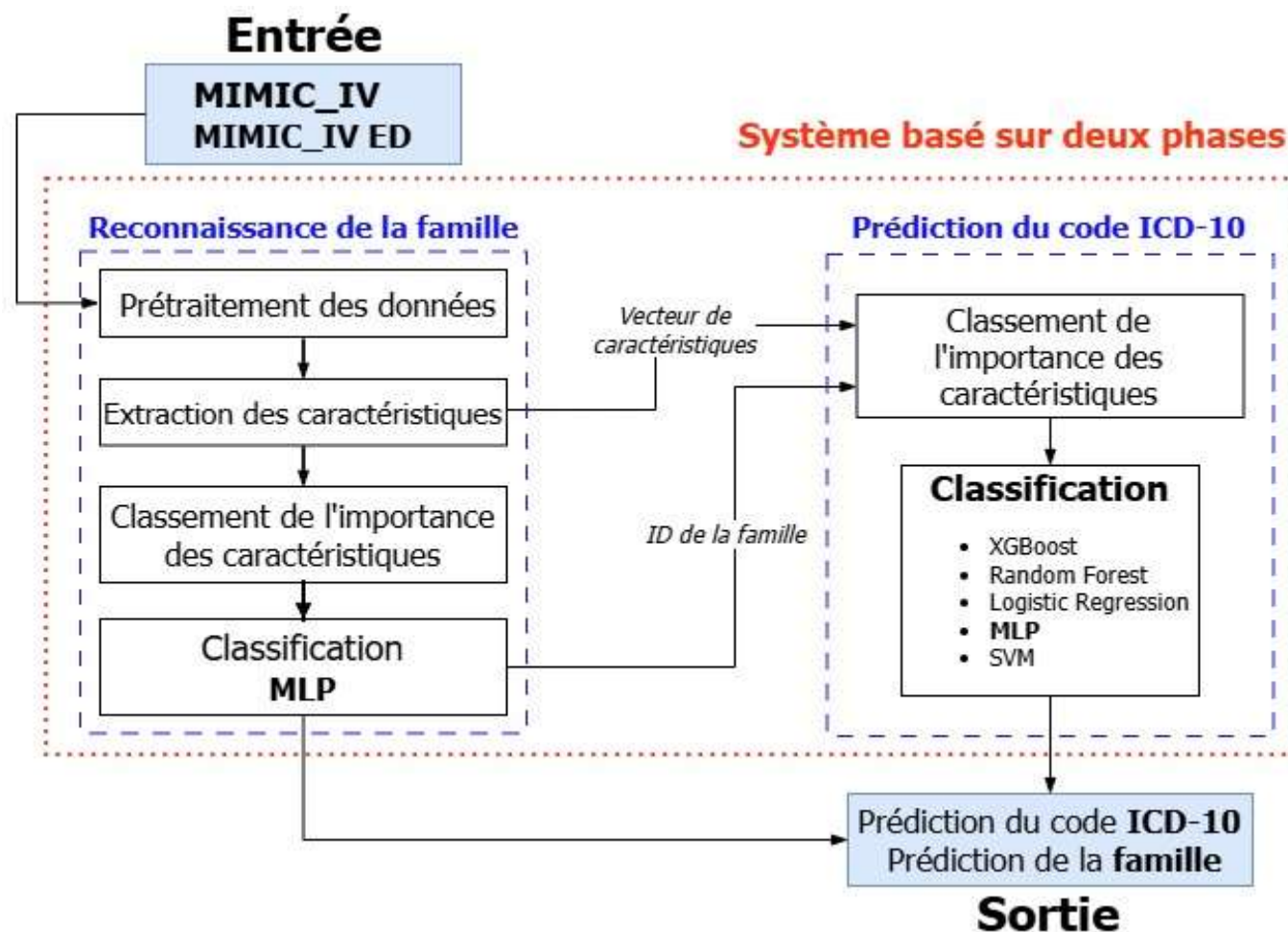


Exploiter uniquement les
données numériques
structurées



Optimiser précision et
efficacité du codage tout en
réduisant les erreurs

solution proposée :



Phase 1: Identification de la famille

- Données extraites de MIMIC-IV & MIMIC-IV ED
- Prétraitement : nettoyage, normalisation
- Regroupement selon la classification OMS (**Organisation mondiale de la Santé**)
- Extraction de caractéristiques par **autoencodeurs**
- Sélection de 120 variables via régression logistique
- Classification avec MLP

Phase 2: Prédiction du code ICD-10

- Réutilisation du vecteur + ID famille
- Nouvelle sélection de caractéristiques
- Modèles utilisés : XGBoost, MLP, SVM, RF, Régression logistique

Données utilisées :



MIMIC-IV

2008 – 2019



364 627 séjours
hospitaliers



Données
structurées :
signes vitaux,
labos, traitements



MIMIC-IV Emergency Department (ED)

2011 – 2019



448 972 séjours
aux urgences



Données d'urgences
symptômes initiaux,
diagnostics rapides

Résultats : *Reconnaissance de la famille pathologique*

Modèle	Exactitude	Précision	Rappel	Score F1	Macro AUC
MLP	96.98 %	96.04 %	97.32 %	97.14 %	99.73 %

Résultats : La prédiction du code ICD-10

Modèle	Exactitude	Précision	Rappel	Score F1	Macro AUC
XGBoost	93.01 %	92.48 %	95.87%	95.68%	97.91%
MLP	97.14 %	96.32 %	96.89 %	97.86 %	99.93%
Regression Logistique	95.99 %	94.39 %	95.74%	96.12 %	98.17 %
Foret Aleatoire	92.64 %	91.77 %	93.48 %	93.08 %	95.98 %
SVM	96.48 %	95.91 %	97.98 %	96.47 %	99.50 %

Résultats : Deux exemples de matrices de confusion

Exactitude : 97.96 %		
	M7989	Autres
M7989	100.0 %	00.00 %
Autres	03.56 %	96.44 %

Exactitude : 94.13 %		
	N186	Autres
N186	97.56 %	02.44 %
Autres	06.82 %	93.18 %

Conclusion et Perspectives :

- L'approche hiérarchique exploitant uniquement les données numériques s'est montrée très efficace.
- Intégrer les notes cliniques avec ClinicalBERT pourrait améliorer les résultats.

Références :

- [1] Alistair E.W. Johnson, Luiz Bulgarelli, Lu Shen, Steven Shammout, and Roger G. Mark. MIMIC-IV (version 2.2). 2023.
- [2] Xie, Jun Zhou, Jin Wee Lee, Mingrui Tan, Siqi Li, Logasan S/O Rajnthern, Marcel Lucas Chee, and et al. *Benchmarking prediction models with machine learning*. Scientific Data, 9(1) :658, 2022.
- [3] Yacine Yaddaden, Yacine Benahmed, Marc-Denis Rioux, and Mariem Kallel. *Machine learning-based pre-diagnosis tools in emergency departments : Predicting hospitalization, mortality and triage acuity*. pages 1–6, 2023.