

Problématique

- Augmentation importante des visites aux urgences au Canada (>16 millions/an) [1].
- Forte incertitude clinique lors de l'admission
- Difficulté à identifier rapidement les patients à haut risque de mortalité
- Limites des systèmes de triage actuels (subjectivité, lenteur)

Objectifs

- ✓ Développer un modèle d'apprentissage automatique pour prédire la mortalité hospitalière aux urgences.
- ✓ Utiliser uniquement les données disponibles à l'arrivée du patient.
- ✓ Garantir l'interprétabilité du modèle en intégrant des méthodes explicatives comme SHAP pour faciliter l'adoption clinique.
- ✓ Identifier les facteurs clés de risque pour mieux prioriser les patients à haut risque.

Méthodologie suivie

1. Extraction des données

→ Bases de données Medical Information Mart for Intensive Care **MIMIC-IV-IV** [2] et **MIMIC-IV-ED**[3].

→ Seules les données disponibles à l'arrivée du patient ont été conservées

2. Pré-traitement des données

→ Nettoyage : détecter et corriger les erreurs et les incohérences afin d'obtenir un jeu de données fiable et exploitable.

→ Valeurs manquantes : imputation itérative pour numériques, modalité la plus fréquente pour catégoriques.

3. Feature Engineering

→ Temporalités et mode d'arrivée : conversion de l'heure d'arrivée en périodes de service et codage numérique du mode de transport.

4. Feature Extraction : La plainte principale est prétraitée (nettoyage, tokenisation), transformée en vecteurs numériques via **TF-IDF**, puis résumée par un score global par patient basé sur la somme des poids TF-IDF.

5. Classification

→ Logistic Regression (LR), Random Forest (RF) et XGBoost, combinés dans une approche de **stacking ensemble** avec LR comme méta-classifieur.

6. Interprétabilité du modèle

→ Analyse via **SHAP** pour quantifier la contribution des variables et permettre une interprétation globale et locale des prédictions..

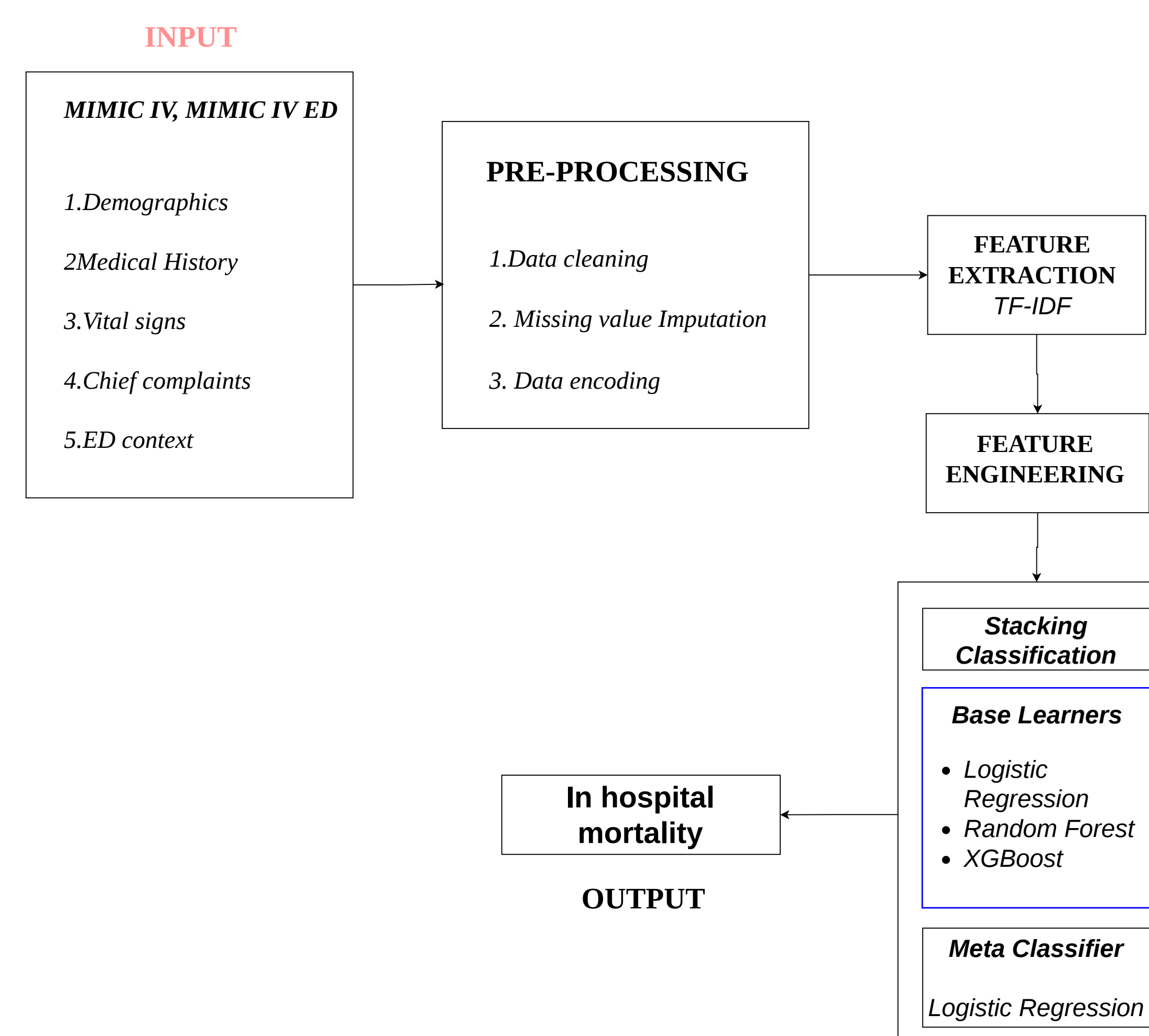


FIGURE 1 – Diagramme de prédiction de la mortalité hospitalière

Évaluation

→ Données issues de **MIMIC-IV** et **MIMIC-IV-ED** (Beth Israel Deaconess Medical Center, 2008–2019), offrant une couverture riche et diversifiée des patients

→ **MIMIC-IV** : données hospitalières complètes (séjours ICU, données physiologiques, historiques cliniques)

→ **MIMIC-IV-ED** : données des urgences (démographie, signes vitaux, plaintes principales, diagnostics initiaux)

Ces données permettent une **analyse fiable et une prédiction précoce** de la mortalité hospitalière.

Résultats

→ Le modèle **Stacking Ensemble** obtient les meilleures performances avec un **ROC-AUC de 0.94**

→ Amélioration significative par rapport aux modèles individuels (LR, RF, XGBoost)

TABLE 1 – Comparaison des performances (%).

Pipeline	Accuracy	Precision	Recall	F1-score	ROC-AUC
LR	0.82	0.80	0.84	0.82	0.90
XGBoost	0.85	0.83	0.86	0.85	0.92
RF	0.85	0.83	0.86	0.85	0.92
Ensemble	0.88	0.86	0.89	0.88	0.94

TABLE 2 – Matrice de confusion (%).

Actual / Predicted	Survival	Death
Survival	85.14	14.86
Death	10.85	89.15

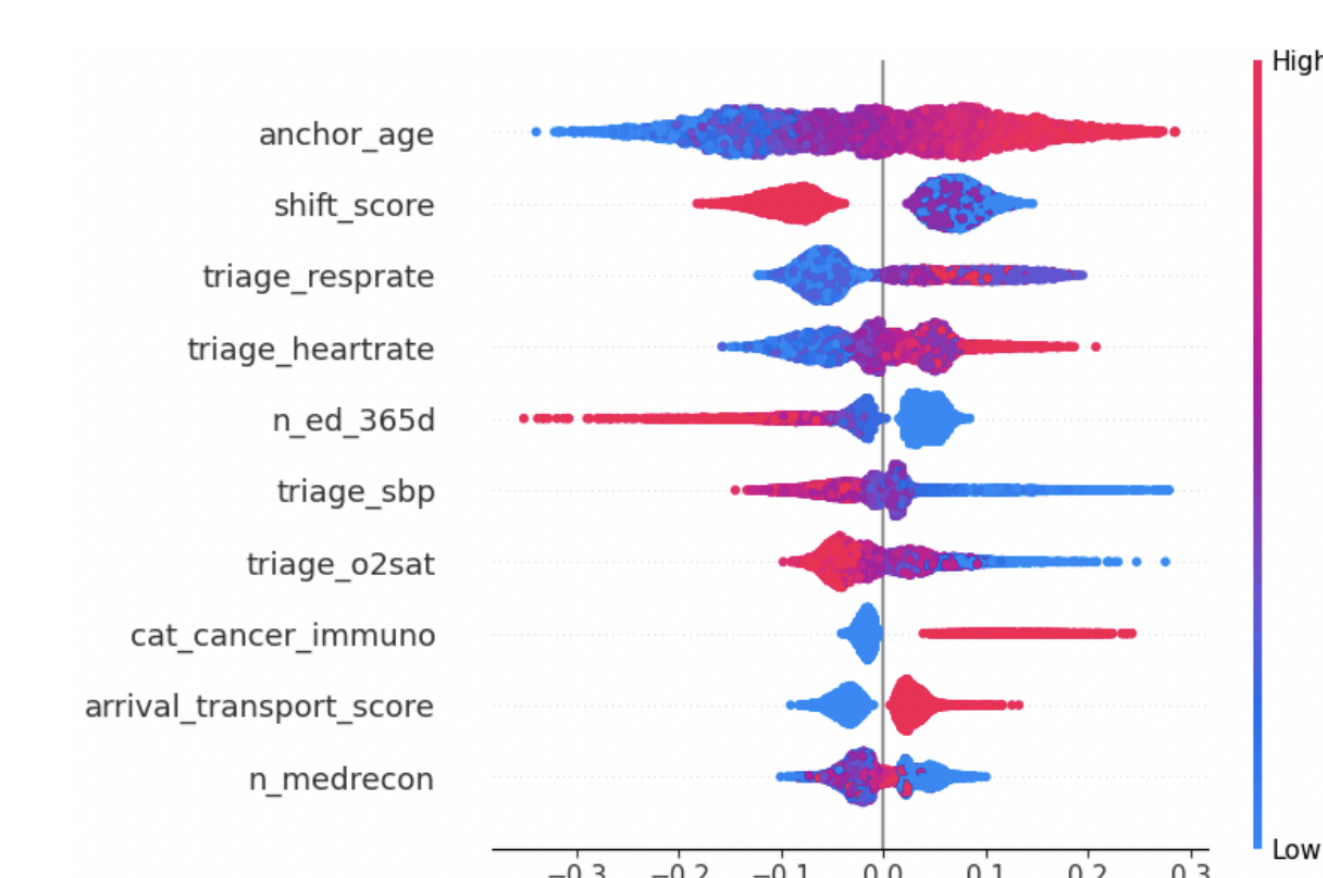


FIGURE 2 – Variables les plus influentes sur la mortalité (SHAP)

Conclusion

→ Le modèle proposé permet une **prédiction précoce et fiable** de la mortalité hospitalière

→ L'approche **ensemble** améliore significativement les performances

→ L'intégration de données textuelles et structurées renforce la précision du modèle

Références

- [1] Canadian Institute for Health Information. Nacrs emergency department visits and lengths of stay, 2026. [Consulté : 6 avril 2026].
- [2] A. Johnson, L. Bulgarelli, T. Pollard, L. A. Celi, R. Mark, and S. Horng. MIMIC-IV-ED (version 2.2), 2023.
- [3] A. Johnson, L. Bulgarelli, T. Pollard, B. Gow, B. Moody, S. Horng, L. A. Celi, and R. Mark. MIMIC-IV (version 3.1), 2024.