

Analyse des profils glycémiques en réanimation

Jonathan SOIHIER

13 janvier 2026

Résumé

Le résumé parle de... .

Chapitre 1

Description du dataset

1.1 Origine et nature des données

Le dataset utilisé dans ce projet provient d'un ensemble de mesures de glycémie réalisées chez dix patients hospitalisés en unité de soins intensifs (ICU). Les données initiales sont de nature temporelle : chaque patient dispose d'une série de mesures de glycémie enregistrées au cours de son séjour, avec une fréquence variable selon les pratiques cliniques (mesures capillaires, prélèvements sanguins, etc.).

Bien que la taille du dataset soit réduite, il s'agit de données réelles issues d'un contexte clinique complexe. L'objectif du projet n'est pas de produire un modèle généralisable à grande échelle, mais d'illustrer un pipeline complet d'analyse et de modélisation à partir de données ICU.

1.2 Structure des données brutes

Pour chaque patient, les données brutes comprennent :

- une série temporelle de valeurs de glycémie,
- la durée totale d'observation,
- des métadonnées cliniques minimales (identifiant patient, période d'enregistrement).

Ces données ont été transformées en un tableau patient-level, chaque ligne représentant un patient et chaque colonne une caractéristique dérivée de son profil glycémique.

1.3 Feature engineering

Afin de résumer les profils glycémiques individuels, plusieurs variables cliniquement pertinentes ont été calculées :

- **Moyenne glycémique** : indicateur global du niveau d'hyperglycémie.
- **Médiane glycémique** : plus robuste aux valeurs extrêmes.
- **Coefficient de variation (CV)** : mesure de la variabilité glycémique.
- **Amplitude glycémique** : différence entre les valeurs minimale et maximale.
- **Pourcentage de valeurs > 180 mg/dL** : indicateur d'hyperglycémie persistante.
- **Pourcentage de valeurs < 70 mg/dL** : indicateur d'hypoglycémie.

Ces variables sont couramment utilisées dans la littérature clinique pour caractériser la stabilité ou l'instabilité du contrôle glycémique en réanimation.

1.4 Définition du label de sévérité

Un label binaire *severe* a été défini afin de distinguer les patients présentant un profil glycémique sévère. Ce label repose sur des critères cliniques simples, notamment :

- une médiane glycémique élevée,
- un pourcentage important de valeurs supérieures à 180 mg/dL,
- une variabilité glycémique marquée.

Cette définition permet de séparer les patients en deux groupes : *modéré* et *sévère*. La distribution finale est équilibrée (5 patients dans chaque groupe), ce qui facilite l'analyse comparative.

1.5 Choix de ne pas fusionner avec des datasets externes

Plusieurs sources publiques de données sur le diabète existent (UCI, Kaggle, Mendeley Data, IEEE DataPort). Toutefois, ces datasets diffèrent fortement du contexte ICU :

- populations non critiques (ambulatoires ou généralistes),
- mesures non comparables (glycémie unique, imagerie, données synthétiques),
- objectifs différents (diagnostic du diabète, dépistage de rétinopathie, etc.).

Pour garantir la cohérence clinique et éviter les biais liés à la fusion de populations hétérogènes, il a été décidé de travailler exclusivement sur le dataset ICU.

1.6 Résumé

Le dataset final contient dix patients et un ensemble de variables dérivées permettant de caractériser précisément leur profil glycémique. Malgré sa taille réduite, il constitue une base pertinente pour illustrer un pipeline complet de data science clinique.