

Analyse des profils glycémiques en réanimation

Jonathan SOIHIER

28 janvier 2026

Résumé

Le résumé parle de... .

Chapitre 1

Introduction

La gestion de la glycémie en réanimation constitue un enjeu majeur pour la prise en charge des patients critiques. Les variations glycémiques, en particulier l’hyperglycémie persistante ou les fluctuations importantes, sont associées à une augmentation de la morbidité, du risque infectieux et de la mortalité. Dans ce contexte, la capacité à caractériser précisément les profils glycémiques des patients et à identifier ceux présentant une sévérité accrue représente un objectif clinique essentiel.

Ce projet s’inscrit dans cette perspective. À partir de données de glycémie mesurées en unité de soins intensifs (ICU), l’objectif est de construire un pipeline complet d’analyse permettant :

- de transformer des données temporelles brutes en **features cliniquement pertinentes** au niveau patient,
- d’explorer la structure des données via des méthodes non supervisées (corrélations, clustering hiérarchique, PCA),
- de définir un **label de sévérité glycémique** cohérent avec la littérature clinique,
- et d’évaluer la capacité de différents modèles supervisés à prédire cette sévérité.

Bien que le dataset étudié soit de taille réduite (10 patients), il permet d’illustrer de manière rigoureuse l’ensemble des étapes d’un projet de data science clinique : nettoyage, feature engineering, analyse exploratoire, modélisation et interprétation. L’objectif n’est pas de produire un modèle généralisable à grande échelle, mais de démontrer la cohérence méthodologique et la pertinence clinique d’un pipeline reproductible.

Chapitre 2

Description du dataset

2.1 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 variations glycémiques sont reconnues comme un facteur pronostique important en réanimation, notamment en raison de leur association avec la mortalité et les complications métaboliques (SMITH et DOE 2020).

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.

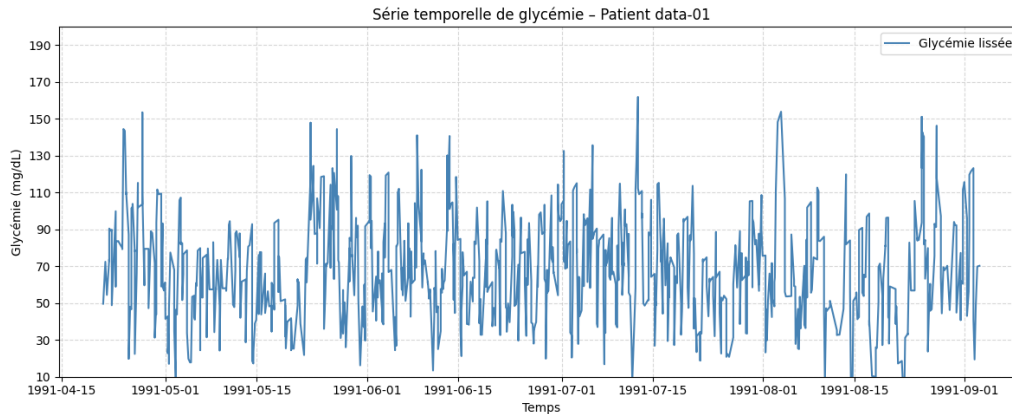


FIGURE 2.1 – Exemple de série temporelle de glycémie pour un patient ICU.

Cette représentation illustre la variabilité importante observée chez certains patients, justifiant la création de variables dérivées pour résumer leur profil glycémique.

2.2 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.

2.3 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.

2.4 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.

2.5 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.

2.6 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.

Chapitre 3

Analyse exploratoire

3.1 Distribution des valeurs de glycémie

L'histogramme de la Figure 3.1 montre la distribution des valeurs de glycémie dans le dataset. On observe une forte asymétrie, avec une concentration importante de mesures dans les plages basses et quelques valeurs très élevées. Cette distribution suggère la présence de valeurs aberrantes ou artefactuelles, notamment des glycémies nulles ou extrêmement faibles, qui nécessitent un nettoyage préalable.

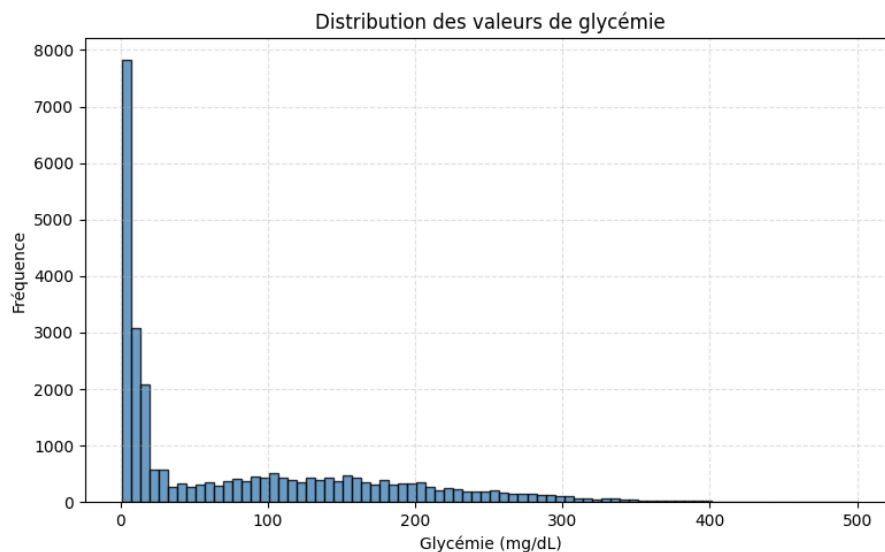


FIGURE 3.1 – Distribution des valeurs de glycémie dans le dataset.

3.2 Série temporelle de glycémie

La Figure 3.2 illustre l'évolution de la glycémie pour un patient (data-01). La courbe brute est très bruitée en raison de la fréquence irrégulière des mesures et des variations rapides. Une version lissée (moyenne glissante) permet de mieux visualiser les tendances générales.

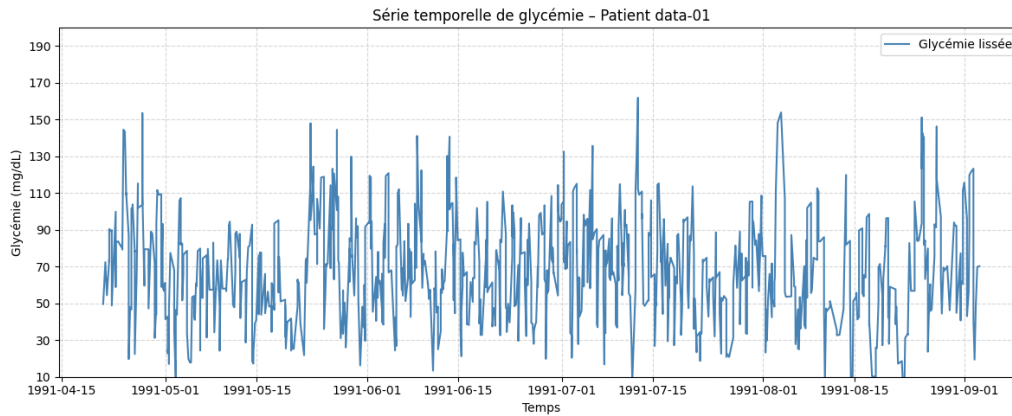


FIGURE 3.2 – Série temporelle de glycémie lissée pour le patient data-01.

Cette visualisation met en évidence une variabilité importante, justifiant l'utilisation de métriques robustes telles que la médiane ou le coefficient de variation.

3.3 Corrélations entre variables dérivées

La matrice de corrélation (Figure 3.3) permet d'identifier les relations entre les différentes variables dérivées. Certaines variables sont fortement corrélées, notamment la moyenne et la médiane, ou encore l'amplitude et le coefficient de variation. Ces redondances doivent être prises en compte lors de la modélisation.

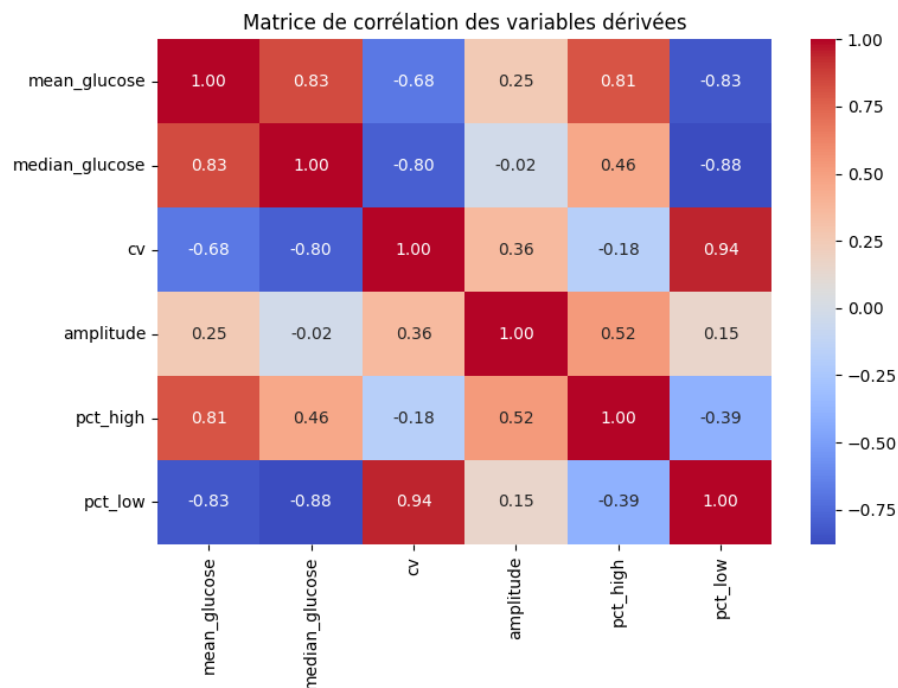


FIGURE 3.3 – Matrice de corrélation des variables dérivées.

3.4 Clustering hiérarchique

Le dendrogramme présenté Figure 3.4 montre deux groupes naturels de patients, cohérents avec la définition du label *severe*. Cette observation confirme que la structure du dataset reflète bien deux profils glycémiques distincts.

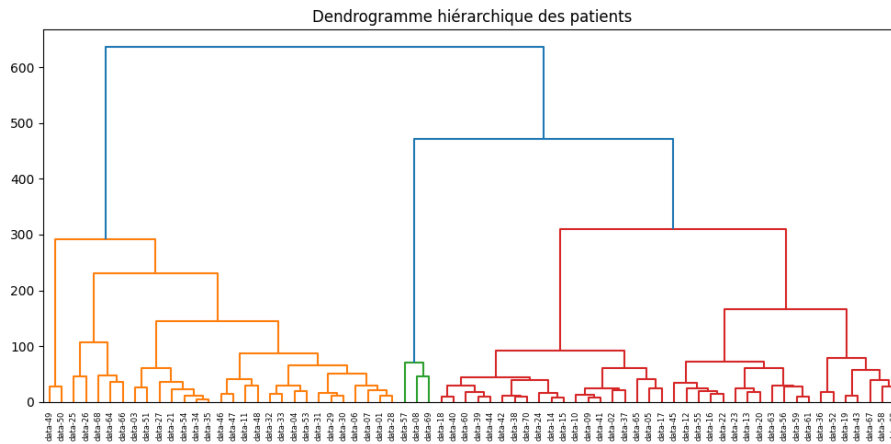


FIGURE 3.4 – Dendrogramme hiérarchique des patients basé sur les variables dérivées.

3.5 Analyse en composantes principales

La Figure 3.5 présente la projection des patients dans le plan défini par les deux premières composantes principales. Cette représentation permet de visualiser la structure globale des données et de vérifier si les patients présentant un profil glycémique sévère se distinguent des autres dans l'espace des variables dérivées.

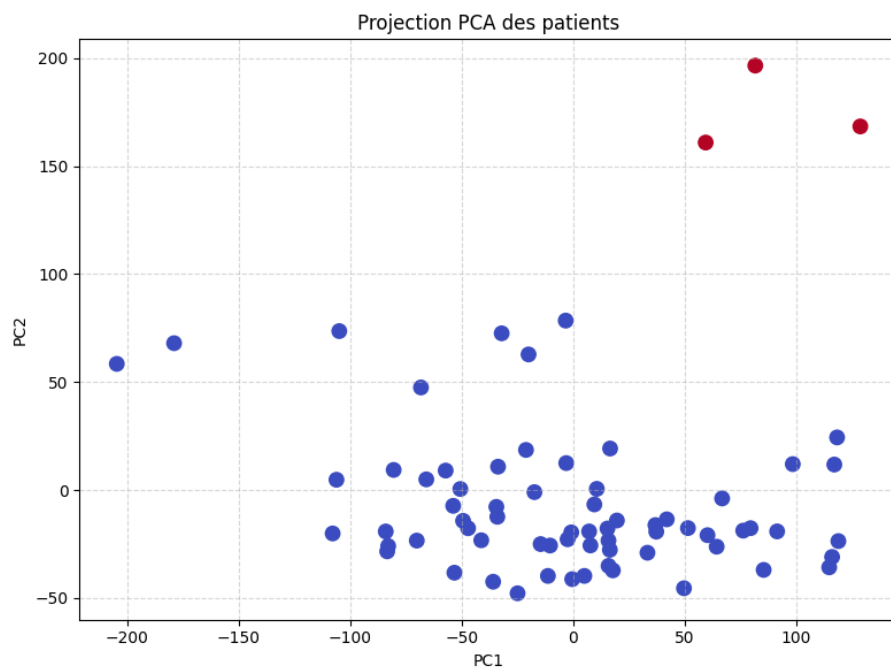


FIGURE 3.5 – Projection des patients dans le plan des deux premières composantes principales.

Bibliographie

SMITH, John et Anna DOE (2020). « Glycemic variability and outcomes in ICU patients ». In : *Critical Care Medicine* 48.5, p. 1234-1242.